République Algérienne Démocratique et Populaire

Ministère de l'enseignement supérieur et de la recherche scientifique

UNIVERSITE YAHIA FARES DE MEDEA

Faculté de Technologie

Département de Génie des Procédés et Environnement

Mémoire de fin d'études de Master

Filière : Génie des Procédés

Spécialité : Génie des Procédés Pharmaceutiques

THEME :

Modélisation computationnelle de l'inflammation neurogène médiée par les cellules T gamma delta dans la maladie d'Alzheimer

Proposé et dirigé par :

- ✓ Pr. BENKORTBI Othmane
- ✓ Dr. MADANI Achwak

Présenté par :

- ✓ Kelaiaia Raid
- ✓ Safa Mohamed

Promotion 2024_2025

Dédicace

Je dédie ce travail :

A Mes chers parents, ***autant de phrases et d'expressions aussi éloquentes soient-elles ne sauraient exprimer ma gratitude et ma reconnaissance envers vous, qu'Allah vous garde.***

A Mon frère Rami, je te souhaite une vie pleine de bonheur et de succès et qu'Allah te protège.

A ma grande-mère Messaouda.

A la mémoire de mes grand-parents (khedidija, yahia et sadek).

A tous mes oncles et toutes mes tantes.

A toute la famille.

A mon cher pays l'Algérie.

Raid

Dédicace

Je dédie ce travail :

A mes chers parents, à qui je dois ma réussite et tout mon respect.

A mes chers frères et mes chères sœurs, je vous souhaite tout le bonheur et qu'Allah vous protège.

A toute ma famille.

A tous mes amis.

Mohamed

REMERCIEMENTS

Nous tenons en premier lieu à remercier ALLAH le tout puissant, qui nous a donné la volonté et la patience d'accomplir ce travail.

Nos profondes gratitudes et nos sincères remerciements à monsieur le Professeur **Othmane Benkortbi,** notre encadreur de mémoire et à madame la **Dr Achwak Madani** notre co-encadreur de mémoire de master pour leurs patiences et leurs judicieux conseils qui nous ont permis de conclure ce travail.

Nous tenons à remercier les membres de jury d'avoir accepté d'examiner notre mémoire et de l'enrichir par leurs critiques et leurs suggestions.

Nos remerciements iront également aux personnels de notre faculté et surtout de notre département

ملخص:

يركز مشروع نهاية الدراسات هذا على استكشاف عقاقير جديدة لمرض الزهيمر، من خالل تطوير نموذج QSARو QSAR Inverseالربط بنية الجزيئات بنشاطها، وال سيما Pic50.تم تطبيق خوارزميةALO_SVRلتنفيذ هذه المهمة .ا لحداثتها تم استخدام الخوارزمية التدريجية للحصول على الواصفات التمثيلية لقاعدة البيانات المختارة وفقً ونشاطها المثبت. تم تقييم النموذج المركب ومقارنته باستخدام معايير تحقق داخلية وخارجية مختلفة ، مع تحديد المنطقة المحددة لتطبيق النموذج. أظهر نموذجALO_SVRقدرة وقوة تنبؤية استثنائية ، بمعايير (1 0.9997 0.9816)=R^2و RMSE = (0.0100)باإلضافة إلى ذلك ، تم استخدام ADME السويسري للحصول على المعلمات الدوائيةو لتقييم معلمات السمية. باإلضافة إلى ذلك ، تم إجراء االلتحام بواسطةschrodinger suiteمن أجل دراسة التفاعالت بين الجزيئات والهدف.

الكلمات الرئيسية: الزاهيمر، QSAR، SVM، اإلرساء، التنبؤ,cellules T Delta Gamma

Résumé :

Ce projet final porte sur la découverte de nouveaux médicaments contre la maladie d'Alzheimer. Il s'appuie sur le développement de modèles QSAR et QSAR inverses pour relier la structure des molécules à leur activité. L'algorithme ALO_SVR a été appliqué à cette tâche. Grâce à sa nouveauté, l'algorithme pas à pas a permis d'obtenir des descripteurs représentatifs de la base de données sélectionnée, en fonction de leur activité avérée. Le modèle construit a été évalué et comparé à l'aide de divers critères de validation internes et externes, identifiant ainsi son domaine d'application spécifique. Le modèle ALO_SVR a démontré une capacité prédictive et une puissance exceptionnelle, avec un R^2 = (0,9816 ± 0,9997 ± 1) et un RMSE = (0,0100). De plus, l'ADME suisse a été utilisé pour obtenir les paramètres pharmacocinétiques et évaluer les paramètres de toxicité. Enfin, un docking a été réalisé à l'aide de la suite Schrödinger afin d'étudier les interactions entre les molécules et la cible.

Mots-clés : Maladie d'Alzheimer, QSAR, SVM, Docking, Prédiction, T Delta Gamma,

Abstract:

This final project focuses on the discovery of new drugs for Alzheimer's disease, by developing QSAR and QSAR Inverse models to link the structure of molecules to their activity. The ALO_SVR algorithm was applied to this task. Due to its novelty, the stepwise algorithm was used to obtain representative descriptors from the selected database according to their proven activity. The constructed model was evaluated and compared using various internal and external validation criteria, identifying the specific area of application for the model. The ALO_SVR model demonstrated exceptional predictive ability and power, with R^2 = (0.9816 ± 0.9997 ± 1) and RMSE = (0.0100). In addition, the Swiss ADME was used to obtain pharmacokinetic parameters and to evaluate toxicity parameters. Furthermore, docking was performed using the Schrodinger suite to study the interactions between molecules and the target.

Keywords: Alzheimer's, QSAR, SVM, Docking, Prediction, T Delta Gamma celle,

Liste des Tableaux

Tableau I.1 : Cellules immunitaires entre immunité innée et adaptative……..8

Tableau III.1 : les descripteurs classés en fonction de la dimension chimique……43

Tableau IV1: la base dedonnées……...56

Tableau IV.2 : les descripteurs pertinents……..…..64

Tableau IV.3: descripteurs retenus…….64

Tableau IV.4: Critères statistiques pour le modèle ALO SVR……..66

Tableau IV.5 : Validation criteria globale……66

Tableau IV.6: descriptor_columns + 'SMILES'…….78

Tableau IV.7: chembl_smiles-9950 molécules…….81

Tableau IV.8: 9950 molécules with descriptors……82

Liste des Figures

Figure I.1 : Évolution des populations de lymphocytes T γδ humains dans le thymus, le sang périphérique et les tissus tout au long de la vie……11

Figure I.2: Ligands et récepteurs des LT γδ……12

Figure I.3 : Six des fonctions les mieux caractérisées des lymphocytes Tγδ…….....13

Figure II.1 – Exemple de perceptron à une couche cachée……21

Figure II.2 : Principe général d'une simulation de Dynamique Moléculaire…….....30

Figure III.1 : Methodologie generale d'une étude QSAR……..42

Figure III.2. : Concept d'optimisation : une fonction de régression f(x) ajuste les données d'entraînement pour déterminer de nouvelles coordonnées dans l'espace des descripteurs. Les coordonnées optimisées, basées sur f(x), se situent dans la plage des données d'entraînement mais donnent une valeur y plus élevée que tout autre point de données……50

Figure III.3: Simulation data sets. In three independent trials, simulation data sets were generated. For each trial, training (black dots) and test blue squares) data are shown with true y values produced by the bird function f(x_1, x_2)……51

Figure IV 1: capture d'écran Data Warriorr……..55

Figure IV. 2: Capture d'écran AlvaDesc……56

Figure IV.3 : Bar plot de l'importance relative des descripteurs……63
Figure IV.4 : Diagramme de corrélation……63
Figure IV.5 : Courbes de régression ALO-SVR……..68
Figure IV.6 : Domaine applicability……69
Figure IV.7 : Diagramme des interactions 2D de (65) avec A18 (CID:128861).....71
Figure IV.8 : Diagramme des interactions 3D de (65) avec A18 (CID:128861).....71
Figure IV.9 : Diagramme des interactions 2D de(64)avec A18 (CID:128861).....72
Figure IV.10 : Diagramme des interactions 3D de (64) avec A18 (CID:128861).....73
Figure IV.11 : Diagramme des interactions 2D de (60) avec A18 (CID:128861).....74
Figure IV.12 : Diagramme des interactions 3D de (60) avec A18 (CID:128861).....74
Figure IV.13 : Diagramme des interactions 2D de (58) avec A18 (CID:128861).....74
Figure IV.14: Diagramme des interactions 3D de (58) avec A18 (CID:128861).....75
Figure IV.15 : Diagramme des interactions 2D de (29) avec A18 (CID:128861).....76
Figure IV.16 : Diagramme des interactions 3D de (29) avec A18 (CID:128861).....76
Figure IV.17 : Olpadronic Acid……80
Figure IV.18 : Cabazitaxel molecule……83

Listes des abréviations

AD: Domaine d'applicabilité

AG: Antigène

ALO:Antlion Optimization algorithm

ANN:Réseaux de Neurones Artificiels (Artificial Neural Networks)

BPCO : Bronchopneumopathie Chronique Obstructive

CDR:région déterminante complémentaire(completementary-determiningregion)

CMH:Complexe Majeur d'Histocompatibilité

COX: Cyclooxygénase

DEVS:Spécification de Système à Événements Discrets (Discrete Event System Specification)

IA : Intelligence Artificielle

LT:lymphocytes T

MA: Maladie d'elzheimer

MAPK: Protéine Kinase Activée par les Mitogènes

MAIT:Lymphocytes T invariants associés aux muqueuses

MDCS: Monocytes dérivés des cellules dendritiques

PCA:Analyse en composantes principales

PDL: Ligand de la mort cellulaire programmée (Programmed cell Death Ligand)

PLS:Moindres Carrés Partiels(Partial Least Squares)

QSAR:Relations quantitative structure- activité

RLM:Régression Linéaire Multiple

RNLM:Réseaux de Neurones Linéaires Multicouches

SVC: Classification par vecteurs de supports

SVM:Machines à vecteurs de support

SVR:Régression à Vecteurs de Support

TCR:Récepteur des cellules T

TGF: Facteur de Croissance Transformant Beta

Treg:Lymphocytes T régulateurs

TRG:Récepteur des cellules T gamma (T CellReceptor Gamma)

TRD:Récepteur des cellules TDelta (T CellReceptor Delta)

UARNm:Acideribonucléiquemessager

Liste des symboles :

C : Paramètre de régularisation.

γ (gamma): Détermine l'impact des points de données individuels.

ε (epsilon): Représente la marge d'erreur acceptable.

R Récupéré.

R0 Numéro de reproduction.

I Infecté/infectant.

S Contagieux (sensible).

$\mathbf{R^2}$Coefficient de détermination.

$\boldsymbol{R_0^2}$ Coefficient de détermination sans intercept.

$\mathbf{Q^2}$ Leave one out.

Y_obs : valeurs observées.

Y_calc : valeurs calculées par le modèle.

$\bar{Y}$_obs : moyenne des valeurs observées.

Y_obs(train) : valeurs observées dans l'ensemble d'entraînement.

Y_calc(train) : valeurs calculées dans l'ensemble d'entraînement.

$\bar{Y}$_training : moyenne des valeurs observées dans l'ensemble d'entraînement.

k : coefficient de proportionnalité calculé.

n : nombre de points dans l'échantillon.

SOMMAIRE

Liste des tableaux

Liste des figures

Liste des abréviations

Liste des symboles

Introduction Générale 01

Chapitre I :Role des cellules T Delta Gamma dans la maladie d'Alzheimer

I.1 Introduction 03

I.2. Identification des cellules T 03

I.2 1. Cellules T types et caractéristiques 05

I.2 2. Caractéristiques des gènes codant le Tcr 06

I.2 3. Les T types es cellules T 06

I.2 4. Rôle et développement des Lymphocytes T γδ dans l'immunité Humaine 08

I.2 5. Découverte et place des Lymphocytes T γδ au sein de l'immunite 09

I.2 6. Développement et selection thymique des Lymphocytes T γδ 09

I.2 7. Caractéristiques des cellules T γδ 11

I.3 Rôle des Cellules T Delta Gamma dans la maladie d'Alzheimer 15

I.3 1. Concept de Névrite 15

I.3 2. Névrite et développement de la maladie d'Alzheimer 16

I.3 3. Impact des cellules T Gamma Delta sur la maladie d'Alzheimer 17

Chapitre II : Modélisation Computationnelle

II.1 Introduction 19

II.2. Définition 19

II.3 Modélisations computationnelles classiques 20

II.3.1 Méthodes génératives 20
II.3.1.1 Méthodes génératives et PCA 20
II.3.1.2 Méthodes analytiques et équations différentielles 20
II.3.1.3 Modèles de type boite noire et Markov Caches 21
II.3.2. Méthodes discriminantes 21
II.3.2.1 Méthodes discriminantes heuristiques 21
II.3.2 2. Méthodes discriminantes linéaires 22
II.3.2.3 Méthodes discriminantes stochastiques et arbres de décision 22
II.4. Applications dans les sciences biomédicales 23
II.4.1 Modélisation des maladies 23
II.4.1.1. Types des modèles 23
II.4.1.3 Modèles biologiques 26
II.4.1.4. Modèles statistiques 27
II.5. Conception des médicaments 28
II.5.1 Simulation des interactions 28
II.5.1.1 Modélisation moléculaire 28
II.5.1.2 Techniques utilisées 29
II.5.1.3. Outils et logiciels 31
II.5.1.4 Analyse de données 31
II.6. Outils et techniques modernes 31
II.6.1. Nouvelles techniques de modélisation 31
II.6.1 1. Réseaux de Neurones profonds 31
II.6.1 2. Modèles de simulation stochastique 32
II.6.1.3 Modélisation par équations différentielles partielles 32
II.6.1.4 Modélisation intégrative 32

II.6.1.5 Modélisation par apprentissage 33
II.6.2 Nouveaux outils logiciels 33
II.7. Avenir de la modélisation 35
II.7. Application actuelle 36
II.8. Conclusion 37

Chapitre III :QSAR, QSAR Inverse et Docking

III.1. Définition 39
III.2. Objectif 39
III.3. Principe 39
III.3. 1. Données biologiques 40
III.3. 2. Approche globale 40
III.3. 3. Descripteurs moléculaires 41
III.4. Méthodologie Générale d'une étude QSAR 43
III.4.1. Base de Données et paramètres biologiques 43
III.4.2. Descripteurs moleculaires 44
III.4.3. Techniques utilisées 46
III.4.3.1 Réseau de Neurones Artificiels 46
III.4.3.2 Support Vector Regression 46
III.4.3.3 Algorithme d'optimisation des fourmis-lions 47
III.5. Approche basé sur le ligand 47
III.5.1. ADMET 47
III.5.2. Docking 48
III.5.2.1. Principe du Docking 49
III.5.2.2. Les étapes du Docking 49

III.5.2.3. La fonction de Score 49
III.5.2.4. Les Types du Docking 50
III.6. QSAR Inverse 51
III.6.2. Principe du QSAR inverse 51
III.6.3. Les Etapes du QSAR Inverse 51
III.6.3. 1. Concept Méthodologique 52
III.6.3.2. Modèles de régression et domaine d'applicabilité 52
III.6.3.3. Simulation data 52
III.6.3.4. Compound data sets 53
III.6.3.5. Virtual screening 54
III.7. Conclusion 55

Chapitre IV : Procédure et Résultats

IV.1 Étude du QSAR 56
IV.1.1. Méthodologie de calcul 56
IV.1.2 Collecte de la base de données 57
IV.1.3 Sélection des descripteurs 57
IV.1.4 Division de la base de données et développement du modèle 57
IV.1.5 Support Vector Regression 61
IV.1.6 Domaine d'applicabilité 63
IV.1.7 Sélection des descripteurs représentatifs 63
IV.1.8 Développement des modèles QSAR 65
IV.1.9 Validation externe 69
IV.1.10 Domaine d'applicabilité 70

IV.2 Étude du Docking 70
IV.2.1 ADMET 70
IV.2.2 Préparation des structures 71
IV.3 Étude du QSAR inverse 78
IV.3.1 Objectifs 78
IV.3.2 Modèle Ant Lion-SVR 78
IV.3.3 Extraction de SMILES 81
IV.3.4 Traitement des données 82
IV.3.5 Exécution d'Alo avec de nouvelles données 83
Conclusion Générale 85
Références bibliographiques 86

INTRODUCTION GÉNÉRALE

Introduction Générale

La maladie d'Alzheimer est l'une des maladies les plus courantes à notre époque, Plutôt, c'est l'un des plus dangereux car il s'agit d'une maladie complexe qui affecte le système nerveux humain, entraînant une perte progressive de la mémoire au fil du temps et pouvant aller jusqu'à la mort. C'est aussi l'une des maladies chroniques qui nécessitent un suivi et une communication constante avec la personne atteinte, qui est à la fois présente et absente dans l'environnement dans lequel elle vit. [22]

Pour ces raisons, les chercheurs dans ce domaine n'ont cessé de rechercher des traitements appropriés pour cette maladie ou, du moins, de minimiser les effets secondaires qu'elle provoque. Les chercheurs s'accordent à dire que la neuroinflammation dans le cerveau joue un rôle clé dans l'apparition et la progression de la maladie. La modélisation informatique de la neuroinflammation par les cellules T.gamma-delta est l'un des moyen utilisés dans la recherche de solutions thérapeutiques, qui fait l'objet de cette étude. [1]

Cette étude vise à mettre en lumière le rôle de la modélisation informatique de la neuroinflammation par les cellules T gamma-delta dans la démonstration de la relation entre l'activité de ces cellules et l'apparition et la progression de la maladie d'Alzheimer, puis à essayer de concevoir des traitements appropriés pour cette maladie par l'utilisation d'un ensemble d'applications, à savoir QSAR, QSAR INVERSE et DOCKING.

La modélisation informatique pourrait contribuer à trouver de nouvelles pistes thérapeutiques pour la maladie d'Alzheimer en ciblant les cellules T gamma delta dans leur contexte d'activité.

Le mémoire est divisé en quatre chapitres :

Le premier chapitre est consacré à une analyse critique sur le rôle des cellules T. gamma delta dans la neuro inflammation et la maladie d'Alzheimer. Une attention particulière est accordée aux interactions complexes entre les lymphocytes T γδ, et les cellules microgliales qui sont considérées comme des acteurs clés de la réponse inflammatoire dans le cerveau.

Le troisième chapitre aborde un domaine en pleine expansion : la conception de médicaments assistée par ordinateur (QSAR). Ce chapitre ne se limite pas à une description des méthodes et des algorithmes utilisés dans ce domaine. Il propose une réflexion épistémologique sur les fondements de la QSAR, en explorant les liens entre la structure moléculaire et l'activité biologique. Une analyse critique des limitations et des biais potentiels des modèles QSAR est également présentée, ainsi qu'une discussion des stratégies permettant de valider et d'améliorer la robustesse et la prédictivité de ces modèles. L'objectif est de fournir une vision claire des potentialités et des défis de la QSAR dans le contexte de la découverte de médicaments.

Le quatrème chapitre a trait à l'exploration des modèles computationnels en sciences biologiques. Ce chapitre s'efforce de mettre en évidence les principes fondamentaux de la modélisation des systèmes complexes, en se basant sur l'importance de la validation expérimentale et de l'interprétation des résultats.

Le chapitre quatre décrit les méthodes et les procédures utilisées pour la conception et le développement du modèle pour prédire les différents types d'interactions.

L'interprétation des résultats obtenus sont discutés en détail dans le chapitre cinq afin de permettre de contribuer, dans une certaine mesure, à trouver des moyens d'atténuer les symptômes de la maladie d'Alzheimer sur la base d'une modélisation informatique de la neuroinflammation induite par les lymphocytes T gamma delta.

Le travail se termine par une conclusion générale qui regroupe tous les résultats trouvés afin de permettre de tracer les perspectives qui pourront être envisagés dans le future.

Chapitre I : Role des cellules T Delta Gamma dans la maladie d'Alzheimer

Chapitre I :

Role des cellules T Delta Gamma dans la maladie d'Alzheimer

I.1 INTRODUCTION

La maladie d'Alzheimer est une pathologie neurodégénérative complexe caractérisée par une détérioration progressive des fonctions cognitives. Bien que les mécanismes sous-jacents ne soient pas entièrement élucidés, il semblerait que la neuroinflammation chronique joue un rôle central dans sa progression. Les cellules immunitaires, notamment les lymphocytes T, sont de plus en plus impliqués dans cette réponse inflammatoire.

Ce chapitre a pour objectif d'effectuer une revue sur le rôle des lymphocytes T gamma delta (γδ) dans la neuroinflammation associée à la maladie d'Alzheimer. Ces cellules T particulières, dotées d'un récepteur TCR unique, présentent des caractéristiques fonctionnelles distinctes des lymphocytes T alpha beta conventionnels.

Les mécanismes par lesquels les cellules γδ contribuent à la pathogénèse de la maladie D'Alzheimer seront expliqués dans ce chapitre ainsi que leurs implications dans :

- La reconnaissance des antigènes : comment les cellules γδ détectent-elles les molécules associées à la maladie d'Alzheimer,
- La régulation de la réponse inflammatoire : quel est l'impact des cellules γδ sur l'activation des autres cellules immunitaires,
- Les effets neurotoxiques : dans quelle mesure les cellules γδ contribuent-elles directement ou indirectement à la mort neuronale,

I.2. IDENTIFICATION DES CELLULES T

Les lymphocytes T se définissent comme : « des cellules lymphoides d'origine hématopoiétique qui se développent dans le thymus » [1]**,** elles représentent un élément indispensable de la réponse immunitaire [2]. On peut dire aussi que les LT sont de petites cellules arrondies d'un diamètre de 7-15 µm, avec un rapport nucléocytoplasmique élevé [1].

La source principale des lymphocytes T est le thymus, et ce rôle de thymus a été observé pour la première fois après des études de thymectomies chez des souris nouveaux nées [3]**,** Alors

que glotathion (GSH) est la principale molécule anti-oxydante synthétisée par ces lymphocytes T et cela pour contrer l'accumulation des radicaux libres [4].

« Afin de faire la distinction entre le "soi" et le "non soi", le système immunitaire a élaboré un processus de reconnaissance antigénique. Un antigène (Ag) peut-être une protéine, une nucléoprotéine, un polysaccharide ou un glycolipidproduit par le pathogène. Cependant, les lymphocytes T ne peuvent reconnaître que des fragments d'antigène peptidique » [5].

Les lymphocytes T sont :

Des composantes majeures de l'immunité adaptative cellulaire, les lymphocytes T ont pour origine les progéniteurs lymphoïdes de la moelle osseuse. A ce stade, les progéniteurs lymphocytaires n'expriment pas de marqueurs lymphocytaires (CD3-, CD4-, CD8-, TCR-) et migrent vers le thymus pour y réaliser leurs maturations. En arrivant dans la région corticale thymique, elles vont exprimer l'ensemble des marqueurs lymphocytaires cités précédemment.

La maturation de ces thymocytes « double positifs » va consister en deux étapes successives, réalisées lors de la migration dans le thymus, de la région corticale à la région médullaire.

- La première étape consiste en la sélection positive des thymocytes capable de reconnaître les molécules du CMH des cellules épithéliales thymique. L'interaction TCR-CMH doit être d'une affinité suffisante pour permettre aux thymocytes « double positif » de recevoir les signaux cellulaires de survie. Si l'interaction est de trop forte affinité, cela signifie que le TCR reconnaît fortement le CMH du soi, et est donc délétère pour l'organisme. Le thymocyte ne recevra alors pas les signaux de survie, entrera en apoptose et sera éliminé par les macrophages résidents.

La sélection positive s'achève par l'engagement du thymocyte dans un lignage lymphocytaire « simple positif » CD4+ ou CD8+. La sélection du cluster de différentiation dépendra du type de CMH reconnu par le TCR du thymocyte : CD8 s'il a reconnu le CMH I, CD4 s'il a reconnu le CMH II. A cette étape, le TCR est capable de reconnaitre n'importe quel peptide présenté par un CMH, qu'il soit du soi ou du non soi.

- La seconde étape consiste à l'acquisition de la tolérance vis-à-vis du soi, par sélection négative des thymocytes qui reconnaissent les peptides du soi sur le CMH. Le peptide du soi est présenté aux thymocytes, par des cellules dendritiques de la médullaire thymique sous contrôle du facteur de transcription AIRE (Auto-immune regulator). Les gènes sous le contrôle d'AIRE codent pour des antigènes restreints à des organes spécifiques à l'exception du thymus.

Les thymocytes « simple positif » qui interagissent avec une trop forte affinité vont recevoir des signaux de mort cellulaire et entreront en apoptose. A la sortie du thymus, les lymphocytes T ont acquis :

- i) un TCR fonctionnel,
- ii) un lignage spécifique (CD4 ou CD8),
- iii) la tolérance au soi. Ils sont dits naïfs tant qu'ils n'ont pas rencontré un antigène. L'expression du TCR permet la reconnaissance spécifique de l'antigène par les lymphocytes matures. Pour répondre efficacement à la grande diversité des antigènes étrangers, des réarrangements somatiques aléatoires des segments V(D)J des gènes du TCR ont lieu tout au long de la maturation thymique**.** L'ensemble des TCR formé dans ce processus et retrouvé dans l'organisme constitue le répertoire lymphocytaire T. [6].

I.2 1. Cellules T : types et caractéristiques

On peut aussi identifier ces cellules en passant en revue leurs différents types (a) et en montrant également la totalité de leurs caractéristiques (b).

a. Caractéristiques des cellules T :

La structure globale se distinguent par leurs TCR. Elles expriment un certain nombre de protéines membranaires : des immunoglobulines, des intégrines, des sélectines L, des récepteurs de cytokines, ainsi qu'un certain nombre de cluster de différentiation (CD2, CD28 et CD45) [7].

a.1 Le TCR :

Les TCR sont des récepteurs membranaires caractéristiques des lymphocytes T, on ne les trouve donc nulle part ailleurs. Ils procurent aux LT la propriété de reconnaitre des fragments peptidiques antigéniques associés aux molécules du CMH et ceci de manière spécifique [7].

- **Structure**

Les différentes chaines polypeptidiques appartiennent à la superfamille des immunoglobulines. Chaque chaine peut être divisée en trois parties :

- **Partie extracellulaire**

Elle porte l'extrémité N-terminale, et comporte deux domaines globulaires.

Les domaines variables d'un TCR donné sont spécifiques d'un seul antigène. Le site de contact avec l'antigène sur le domaine variable est appelé paratope.

- **Partie transmembranaire**

Elle contient des acides aminés basiques et permet l'ancrage du TCR dans la membrane cytoplasmique. Ces acides aminés permettent l'interaction avec le complexe CD3.

- **Partie intracellulaire**

Constituée d'une courte séquence d'acides aminés contenant l'extrémité C-terminale. La partie intracellulaire du TCR étant très courte, elle ne permet pas de transmettre des signaux à l'intérieur de la cellule. Ainsi, le TCR se trouve toujours associé au complexe CD3, qui lui, possède une longue partie intracytoplasmique capable d'envoyer les signaux activateurs vers le noyau. De ce fait, le TCR est considéré comme le module de reconnaissance de l'antigène et le complexe CD3 comme le module de transmission du signal [1].

I.2 2. Caractéristiques des gènes codant le TCR

Les TCR ont pour rôle de reconnaître les fragments peptidiques antigéniques présentés par le complexe majeur d'histocompatibilité (CMH). Il est important de préciser qu'un lymphocyte ne possède qu'un seul type de TCR exprimé en plusieurs exemplaires.

Chaque chaîne α, β, γ, δ, sont codées par plusieurs gènes qui ne sont fonctionnels que dans la lignée T.

- ✓ Les chaînes α et γ sont codées chacune par 3 gènes : un gène V (pour variable), un gène J (pour jonction) et un gène C (pour constante),
- ✓ Les chaînes β et δ vont être codées par 4 gènes, les même que précédemment plus un gène D (pour diversité) [8].

I.2 3. Les types des cellules T

On distingue plusieurs types de cellules T:

A. Lymphocytes T auxiliaires (CD4+) :

Les lymphocytes T CD4 sont des cellules du système immunitaire efficaces pour défendre l'organisme contre les menaces extérieures, comme les virus, les bactéries ou les parasites. Elles jouent un rôle crucial dans la surveillance immunitaire, empêchant ainsi l'apparition et la progression des tumeurs.

L'efficacité des lymphocytes T CD4 repose sur leur capacité à ajuster leur phénotype et leurs fonctions en fonction du type de menace détectée par les cellules immunitaires innées. Selon la nature du danger, les cellules présentes dans les tissus endommagés émettent différents

signaux moléculaires, orientant ainsi les lymphocytes T CD4 naïfs vers des lignées fonctionnelles distinctes appelées T helper (Th) [8].

Les lymphocytes T CD4+ sont des cellules mononuclées ovales, avec un noyau central et un cytoplasme abondant, caractérisées par l'expression de la molécule CD4 à leur surface [9]. Elles possèdent une morphologie distinctive par rapport aux autres cellules immunitaires. Ils sont ovales ou sphériques, mesurant en moyenne 10-15 μm de diamètre. Leur noyau est centré, grand, et occupe environ 60% du volume cellulaire. Le cytoplasme est riche en mitochondries et ribosomes variés. [9]

B. Lymphocytes T cytotoxiques (CD8+) :

Les cellules T CD8+ sont considérées comme des lymphocytes adaptatifs grâce à leur capacité à reconnaître des antigènes étrangers spécifiques et à développer des réponses de mémoire. Cependant, des recherches récentes montrent que certains lymphocytes T CD8+ naïfs peuvent répondre à des cytokines innées sans stimulation des récepteurs des lymphocytes T, un phénomène appelé activation spontanée. [10]

C. Lymphocytes T régulateurs:

Les lymphocytes Treg sont les principaux acteurs de l'immunomodulation. Dans des conditions normales, ils garantissent la tolérance du soi et la résolution de la réponse immunitaire. Les lymphocytes Treg naturels expriment le facteur de transcription FoxP3 et le CD25, une sous-unité du récepteur de l'IL-2. Ils représentent 5 à 10 % des lymphocytes T CD4+ et sont induits par le TGF-β et l'IL-10. [6]

D. Lymphocytes T γδ (gamma delta):

Les lymphocytes T gamma-delta (γδ) sont un sous-type de lymphocytes T qui assurent la surveillance épithéliale et orchestrent des réponses pro- et anti-inflammatoires. [11]

L'activation des cellules T γδ peut-être induite par les molécules de type butyrophilin et de type butyrophilin qui sont structurellement similaires aux membres de la famille B7 immunosuppressive, mais on ignore comment elles régulent et coordonnent les réponses des cellules T αβ et γδ. [12]

Les cellules T γδ sont l'un des trois types de cellules immunitaires qui expriment des récepteurs d'antigènes. Elles jouent un rôle crucial dans la surveillance antitumorale lymphoïde et relient l'immunité innée et adaptative. Les cellules T γδ ont la capacité de sécréter des cytokines en grande quantité et d'exercer une cytotoxicité puissante contre une variété de cellules

cancéreuses. Elles sont donc essentielles dans l'immune surveillance et la défense immunitaire contre les tumeurs, devenant des effectrices prometteuses pour l'immunothérapie du cancer…Récemment, les cellules T γδ ont été intégrées dans des essais cliniques. Certains de ces essais ont montré que l'immunothérapie basée sur les cellules T γδ est bien tolérée et efficace. Malgré ces avantages, certains obstacles doivent être surmontés pour développer les immunothérapies à base de cellules T γδ. [13]

I.2 4. Rôle et le Développement des Lymphocytes T γδ dans l'Immunité Humaine:

L'immunité est divisée en deux compartiments : l'immunité innée qui confère une protection immédiate et non spécifique contre les pathogènes et l'immunité adaptative qui est plus tardive, spécifique d'antigène et acquise à long terme. À l'interface entre ces deux compartiments se trouvent les lymphocytes T γδ, qualifiés de lymphocytes T non conventionnels. Ces cellules possèdent des caractéristiques appartenant à la fois à l'immunité innée, comme leur localisation épithéliale et à l'immunité adaptative car elles présentent à leur surface un récepteur T à l'antigène (TCR). [14]

Les lymphocytes T (LT) sont des composants essentiels du système immunitaire, pour la défense contre les organismes étrangers et l'auto-régulation. Ils peuvent être divisés en deux catégories différentiées par leur TCR : α, β, δ, et γ. Les LT α β dits conventionnels et les LT γδ. Les LT γδ, sont connus pour leur fonction entre immunité innée et adaptative (Tableau I.1). Ces LT γδ ont des caractéristiques particulières liées à la reconnaissance directe de l'antigène, à leur tropisme tissulaire, leur double fonction cytotoxique et régulatrice qui leurs permet de combattre des infections microbiennes et de reconnaitre et d'éliminer des cellules cancéreuses. [15]

Tableau I.1 : Cellules immunitaires entre immunité innée et adaptative.

Immunité innée		Immunité adaptative
Cellule NK	Cellule NKT	Lymphocyte B et anticorps
Macrophage	Lymphocyte T γδ	Lymphocyte T CD8+
Mastocyte		Lymphocyte T CD4+
Cellule dendritique		
Basophile		
Eosinophile		
Neutrophile		
Protéine du complément		

I.2 5. Découverte et place des lymphocytes T γδ au sein de l'immunité :

À la différence des autres populations de lymphocytes T non conventionnels (NKT et MAIT), qui ont été découvertes relativement récemment, la première identification d'une population lymphocytaire exprimant un « T cell receptor » (TCR) différent du TCRαβ des lymphocytes T conventionnels remonte à plus de 30 ans. Initialement désignée sous le nom de TCR T3, cette population a été définie pour la première fois par le laboratoire du Pr. M. Krangel sous le terme de « lymphocytes T γδ », exprimant le TCR T3 sans l'expression d'un TCRαβ. [16]

Le système immunitaire est souvent présenté comme une entité séparée en un système immunitaire inné et un système immunitaire adaptatif. Sur cette base archaïque, il est difficile de concevoir que la réponse immunitaire puisse être un continuum. Si les acteurs cellulaires propres à l'immunité innée ou adaptative possèdent des caractéristiques distinctes et non redondantes les unes des autres, il existe une troisième « famille » de leucocytes « hybrides » partageant à la fois des caractéristiques et phénotypes de cellules innées et adaptatives. Ainsi ces cellules permettent d'établir un lien concret entre l'immunité innée et adaptative. Ces lymphocytes T dits « non conventionnels » (NKT, γδT et MAIT) forment ainsi un troisième « système » parfois appelé immunité transitionnelle. [16]

I.2 6. Développement et sélection thymique des lymphocytes T γδ

Le développement des LTγδ suit un processus très précis caractérisé par des vagues et dépendant de différents critères : la nature des progéniteurs au niveau de thymus fœtal, l'environnement thymique et l'intensité du signal du TCR. Chaque vague correspond à une population partageant une chaîne Vδ commune, qui colonise des tissus périphériques précis où elle exerce des fonctions spécifiques. [17]

A. Développement thymique des lymphocytes Tγδ : [14]

Bien qu'aujourd'hui les LTγδ soient un type cellulaire fortement étudié pour ses capacités thérapeutiques et fonctionnelles, son développement reste très peu connu. De nombreuses études ont été mises en place à l'aide de modèles murins pour améliorer cette compréhension mais il existe de grandes différences inter-espèces et les loci TRG et TRD sont très peu conservés ce qui se reflète dans l'apparition de sous-populations différentes de LTγδ.

Les LT se développent à partir des cellules souches hématopoïétiques et de progéniteurs (HSPC) provenant du foie fœtal lors du développement prénatal ou de la moelle osseuse lors du

développement postnatal. Au moment où ce réarrangement est effectué deux types de diversité des TCR sont générées :

E. La diversité combinatoire : qui consiste en l'association d'un gène J avec l'un des divers gènes D (pour le locus qui en contiennent) et l'un des divers gènes V (gènes de variation).
F. La diversité jonctionnelle : cette dernière consiste en l'ajout aléatoire de nucléotides au niveau de la jonction des différents segments associés, ce qui crée la région CDR3 des chaînes TRG et TRD, correspondant au domaine hypervariable et responsable de la reconnaissance de l'antigène.

Ces associations de gènes sont ensuite transcrites et associées à la région C au moment de l'épissage de l'ARNm. La traduction en protéine de l'ARNm épissé aboutit au TCR final exprimé à la surface des LTγδ. Ces cellules sont divisées en sous-populations en fonction du type de région V utilisée dans le TCR γδ exprimé en surface.

B. Evolution vers deux populations distinctes majoritaires : lymphocytes TVδ1 et Vδ2 [15]

On peut distinguer deux populations principales de LTγδ humains : les LT Vγ9Vδ2 et les LT non-Vγ9Vδ2, ces derniers étant principalement composés de cellules exprimant les chaînes Vδ1 et Vδ3 qui peuvent s'associer à n'importe qu'elle chaîne Vg. Les sous-populations de LTγδ humains apparaissent en plusieurs vagues développementales distinctes (Figure I.1). Ces LT Vγ9Vδ2 sont également les premières à quitter le thymus pour rejoindre le sang périphérique où ils présentent des caractéristiques dites innées. Il a été démontré que ces LT Vγ9Vδ2 étaient progressivement remplacés par des LT Vδ1, qui augmentent en nombre jusqu'à devenir le sous-ensemble dominant dans le sang du cordon ombilical lors de l'accouchement, où ils constituent un répertoire hautement polyclonal et privé.

Après la naissance, une seconde vague de LT Vγ9Vδ2 est générée à partir des thymocytes postnataux avec des caractéristiques différentes de la première.

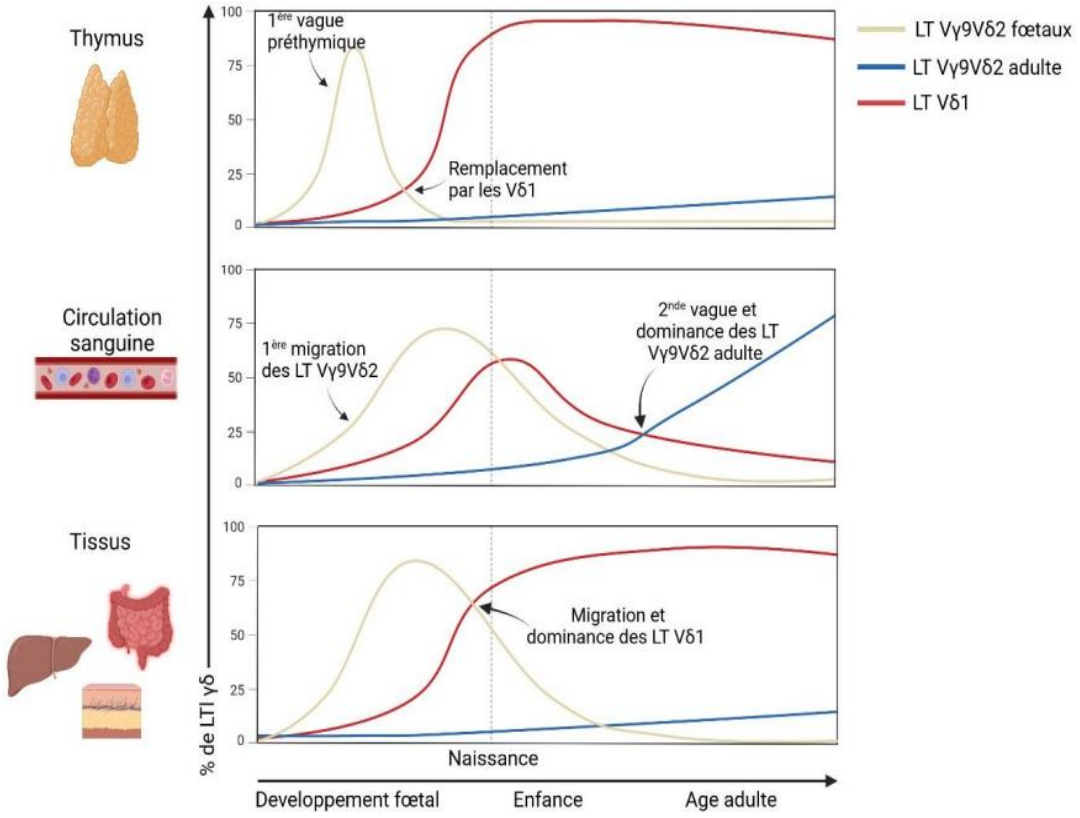


Figure I.1 : Évolution des populations de lymphocytes T γδ humains dans le thymus, le sang périphérique et les tissus tout au long de la vie.

I.2 7. Caractéristiques des cellules T gamma Delta :

A. Récepteur à l'antigène ou TCR :[17]

Dans les années 1980, la recherche des gènes codants les chaînes α et β du récepteur à l'antigène des lymphocytes T αβ a permis d'identifier les gènes γ et δ. Ces gènes codent pour les chaînes γ et δ qui s'associent pour former le TCR γδ, récepteur hétérodimérique caractéristique des lymphocytes T γδ.

B. Activation des LT γδ :

Les lymphocytes T γδ jouent un rôle non redondant dans l'immunosurveillance des tissus grâce à leur capacité unique de reconnaître des antigènes non conventionnels comme des métabolites phosphorylés ou des lipides, qui sont généralement invisibles aux lymphocytes T αβ. Cette reconnaissance s'effectue de manière indépendante du CMH, contrairement aux lymphocytes T αβ dont le TCR se lie aux peptides présentés par les molécules du complexe majeur d'histocompatibilité (CMH).

L'activation des lymphocytes T γδ implique plusieurs signaux qui dépendent à la fois de récepteurs activateurs, partagés avec des cellules de l'immunité innée, et de leur récepteur à l'antigène qui définit leur spécificité. Cependant, les mécanismes précis permettant cette activation ne sont pas encore totalement élucidés à l'heure actuelle. [17]

B. Les récepteurs :

Les LT γδ expriment à leur surface différents récepteurs (Figure I.2). Très peu expriment les co-récepteurs CD4 et CD8 à la surface des LT αβ, seuls les lymphocytes intra-épithéliaux (LIE) peuventexprimer CD8αα83. En revanche, les LT γδ expriment des récepteurs « Natural Killer » (NKR) pouvant être activateur ou inhibiteur. Comme le récepteur NKG2D, présent à la surface des LT γδ qui a une fonction activatrice qui va agir en synergie avec le TCR comme co-stimulateur. Il y a aussi des « Toll Like Receptor » (TLR). Ces récepteurs vont aussi agir comme co-stimulateurs avec le TCR80. [15]

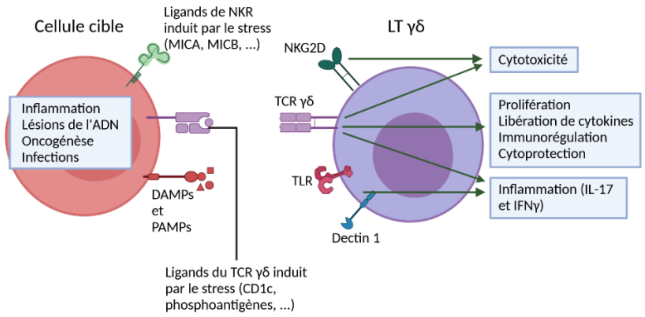


Figure I.2: Ligands et récepteurs des LT γδ

C. Les ligands :

Différents récepteurs ont été mis en évidence à la surface des LT γδ. Pour permettre leur activation, il faut maintenant comprendre avec quel(s) ligand(s) ceux-ci vont rentrer en contact (Figure I.2). Au vu de la diversité du répertoire du CDR3 du TCR, il semble qu'une grande variété d'antigène pourrait être détecté. Cependant, cette variété semble finalement restreinte.

D'autres ligands que le TCR ont été mis en évidence, comme les ligands apparentés au CMH. Parmi eux les membres de la famille des CD1, les ligands « MHC class I polypeptide-related sequence A and B » (MICA et MICB) pouvant être reconnus à la fois par les récepteurs

NKR et par le TCR γδ. Les ligands impliqués dans la réactivité des LT γδ sont liés à un stress cellulaire pouvant être d'origine infectieux ou tumoral et pouvant provenir du Soi. [15]

D. Fonctions des lymphocytes T γδ :

Les lymphocytes T γδ sont les premiers lymphocytes T circulants fonctionnels générés que ce soit chez l'homme ou la souris. Leurs fonctions sont diverses (Figure I.3) et partiellement non redondantes de celles des lymphocytes T αβ du fait de leur localisation, de leurs récepteurs innés et adaptatifs et de leur cinétique d'activation. Au niveau des épithéliums et des muqueuses, ils forment la première ligne de défense de l'organisme ce qui leur permet d'intervenir très rapidement pour limiter la dissémination des pathogènes et d'initier les réponses immunitaires innées et adaptatives. Lors des phases plus tardives, ils peuvent avoir un rôle anti-inflammatoire et favoriser la réparation tissulaire. À l'heure actuelle, le mode d'acquisition de ces fonctions pléïotropiques fait débat et de nombreuses études sont réalisées afin de déterminer comment ces cellules, à la fois adaptatives et innées, sont programmées. [14]

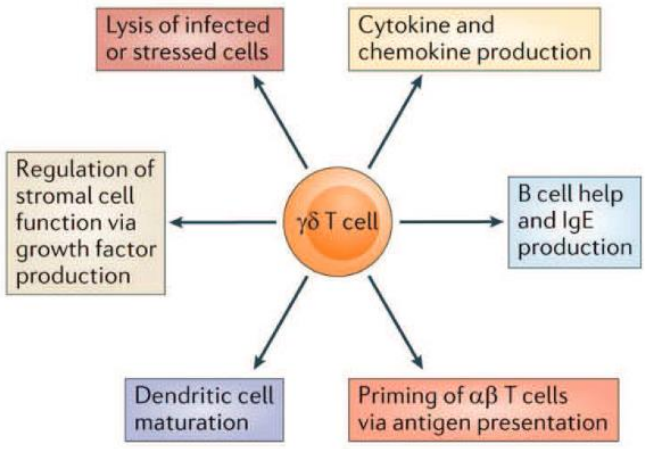


Figure I.3 : Six des fonctions les mieux caractérisées des lymphocytes Tγδ

Dans un contexte tumoral, l'immuno surveillance joue un rôle crucial dans l'expansion tumorale. Ce mécanisme permet de conserver un équilibre entre la réponse immunitaire antitumorale et les mécanismes d'échappement au système immunitaire. Les LT γδ sont

impliqués dans l'immuno surveillance. Ils sont capables de percevoir un changement au niveau de la synthèse de l'antigène et d'être activés pour maintenir l'équilibre.

Les LT γδ présents dans le sang ainsi que ceux présents dans les tissus vont pouvoir agir et maintenir l'intégrité des tissus par leur activité antitumorale ou protumorale. Les fonctions antitumorales des LT γδ sont communes à celles des LT polarisés Th1. Ces LT γδ, activés, peuvent relarguer des molécules effectrices cytotoxiques comme la perforine et les granzymes. Les LT γδ peuvent produire des cytokines immunomodulatrices, comme le TNF-α et IFN γ.
De plus, la mise en contact des récepteurs, présents à la surface des LT γδ, avec leurs ligands va entraîner une activation rapide des LT γδ.Les LT γδ activés vont pouvoir agir de plusieurs façons, permettant de « tuer » la cellule infectée ou transformée par l'engagement de récepteur induisant la mort cellulaire, comme FAS et «Tumor-Necrosis-Factor Related Apoptosis Inducing Ligand Receptor » (TRAILR).

En plus de leur capacité à être activé directement, les LT γδ vont interagir avec d'autres cellules immunitaires comme les LT αβ CD4+/CD8+, les cellules dendritiques, les neutrophiles et les MDCS85. Sur les cellules dendritiques, les LT γδ par la production l'interféron de type 1 vont favoriser leur maturation. Les LT γδ vont jouer le rôle de cellule présentatrice d'antigène (CPA) pour les LT αβ CD4+/CD8+86.

Les lymphocytes T γδ sont connus pour leur fonctions antitumorales comme montré précédemment. Cependant, il existe une dualité de fonction pour ces cellules. En effet, il a été démontré que les LT γδ pouvaient avoir des fonctions protumorales, directes ou indirectes via le recrutement de cellules immunosuppressives et la modification du MET en environnement immunosuppressif.

Les effets protumoraux des LT γδ semblent varier selon les types du cancer, ou du moins ils ont, pourl'instant, été mis en évidence dans certains cancers, comme le cancer du sein, le cancer colorectal etle cancer du pancréas. Sous l'action de cytokines, comme le TGF-β, l'IL-4, l'IL-17 et l'IL-21, les LT γδ vont changer dephénotypes et acquérir des fonctions protumorales.
Un autre mécanisme suppresseur a été mis en avant pour les LT γδ, chez les patients atteints d'uncancer du pancréas. Il a aussi été montré que les LT γδ, du sang périphérique ou de la tumeur, peuventexprimer PDL1. Grâce à la présence de PDL1 à leur surface, les LT γδ vont se lier à PD1 exprimé à lasurface des LT αβ, inhibant ainsi la réponse immunitaire des LT αβ91.

Certaines de ces fonctions anti ou pro tumorales sont reliées à un sous-type particulier de LT γδ : lesLT Vγ9Vδ2. Pour ce projet, nous allons nous intéresser plus particulièrement à ces LT γδ. [15]

I.3 ROLE DES CELLULES T DELTA GAMMA DANS LA MALADIE D'ALZHEIMER :

La Maladie d'Alzheimer (MA) est une maladie neurodégénérative causant une altération progressive des fonctions cognitives. La MA est caractérisée par les plaques amyloïdes, les dégénérescences neurofibrillaires, et une neuroinflammation chronique médiée principalement par les cellules microgliales et les astrocytes. De nombreux travaux suggèrent un rôle complexe de l'immunité adaptative et notamment des lymphocytes T dans la physiopathologie de la MA, qui semblent avoir des effets bénéfiques et/ou délétères selon le type et l'amplitude de la réponse lymphocytaire. Des études récentes suggèrent un effet bénéfique des **cellules T gamma Delta** au cours de la MA, qui semblerait notamment s'exercer via une modulation de la neuroinflammation associée aux lésions neuropathologiques.

Le premier objectif de ce travail a été d'étudier l'impact des **cellules T gamma Delta** sur la modulation de la fonctionnalité astrocytaire. Les résultats montrent que la déplétion transitoire des **cellules T gamma Delta** augmente la fréquence d'astrocytes réactifs neurotoxiques. Ces résultats suggèrent que les **cellules T gamma Delta** régulent la neuroinflammation dans la MA en modulant la fonctionnalité des astrocytes à proximité des lésions neuropathologiques.

Le second objectif de ce travail a été d'évaluer la pertinence clinique du rôle bénéfique des **cellules T gamma Delta** dans la MA chez l'homme. Les résultats mettent en évidence des corrélations entre certaines sous-populations fonctionnelles de **cellules T gamma Delta** et la rapidité de progression de la maladie. L'ensemble de ces travaux permet de mieux comprendre le rôle protecteur des **cellules T gamma Delta** au cours de la MA, et souligne leur intérêt clinique potentiel comme marqueur pronostic et/ou cible thérapeutique.

I.3 1. Concept de névrite :

« Les neuropathies inflammatoires regroupent des neuropathies associées à des infiltrats inflammatoires péri- ou intranerveux, observées ou non dans les maladies de système, et pour lesquelles il existe le plus de possibilités thérapeutiques ». [18]

La névrite du nerf optique (NNO) est une inflammation primaire du nerf optique dont la cause la plus fréquente est une affection démyélinisante du système nerveux central. [19]

On distingue plusieurs types de névrites, par exemple la névrite du nerf optique (NNO) qui « est une inflammation primaire du nerf optique dont la cause la plus fréquente est une affection démyélinisante du système nerveux central ». [19]

On peut citer aussi parmi ces types, « La névrite vestibulaire ou neuronite vestibulaire (NV) est un déficit vestibulaire périphérique unilatéral brusque isolé, caractérisé par l'apparition soudaine d'un vertige rotatoire intense et prolongé (plusieurs jours) associé à un déséquilibre postural, des nausées et des vomissements, sans atteinte auditive ni autre symptôme neurologique ». [20] « La névrite vestibulaire touche préférentiellement la branche supérieure du nerf vestibulaire (55100 %), mais peut aussi réaliser une atteinte de la branche supérieure et inférieure du nerf (1530 %). Plus rarement, elle implique uniquement la branche inférieure du nerf vestibulaire (3,715 %) ». [21]

I.3 2. Névrite et développement de la maladie d'Alzheimer :

L'infection du cerveau par divers types d'agents pathogènes, et les réponses inflammatoires qui s'en suivent, occupent une place grandissante dans notre compréhension de l'étiologie de la maladie d'Alzheimer (MA) [22]. La maladie d'Alzheimer, un trouble neurodégénératif caractérisé par l'accumulation d'amyloïde bêta, représente un défi de santé mondial significatif. Cette maladie est une cause majeure de démence, dont la prévalence augmente avec l'âge. Elle se définit par l'hyperphosphorylation de la tau et le dépôt d'amyloïde bêta, qui altèrent gravement les fonctions de mémoire. L'accumulation d'amyloïde bêta déclenche un spectre de neuroinflammation activée par les microglies, où les microglies et les astrocytes libèrent divers cytokines inflammatoires et anti-inflammatoires. La neuroinflammation est une caractéristique de la maladie d'Alzheimer, avec des signaux de cytokines pro-inflammatoires jouant des rôles cruciaux dans la neurodégénérescence et la neuroprotection. L'activation des voies pro-inflammatoires entraîne la libération de médiateurs immunitaires qui perturbent les fonctions neuronales et peuvent conduire à la mort cellulaire. De plus, une réponse anti-inflammatoire lente aggrave la neuroinflammation. [23]

De plus, l'infection du cerveau par divers agents pathogènes et les réponses inflammatoires qui en résultent influencent considérablement notre compréhension de l'étiologie de la maladie d'Alzheimer. Bien que la maladie d'Alzheimer entraîne des modifications de l'intellect, du comportement et de l'affect, elle n'est pas classée comme un trouble psychiatrique. Au contraire, elle résulte de changements structurels dans le cerveau qui peuvent être observés

tant macroscopiquement que microscopiquement lors de l'examen post-mortem, confirmant ainsi sa classification en tant que maladie neurologique. [24]

I.3 3. Impact des cellules T gamma delta sur la maladie d'Alzheimer:

Les cellules T gamma delta (γδ T) jouent un rôle complexe et significatif dans la progression de la maladie d'Alzheimer (MA). Voici les principaux points détaillés concernant leur impact : [25]

1. Rôle dans l'inflammation Neur inflammatoire

Les cellules γδ T sont impliquées dans la régulation des réponses inflammatoires dans le système nerveux central (SNC). Elles peuvent produire des cytokines pro-inflammatoires comme l'IL-17 et l'IFN-γ, qui contribuent à la neuroinflammation associée à la MA.

2. Infiltration dans le tissu cérébral

- Augmentation de l'infiltration : Les études montrent une augmentation de la présence de cellules γδ T dans le parenchyme cérébral des patients atteints de MA, ce qui est associé à une aggravation de la pathologie.
- Cytotoxicité : Les γδ T peuvent induire la mort neuronale par des mécanismes dépendants du contact cellulaire, tels que l'interaction avec le Fas ligand (FasL).

3. Dysfonction des cellules T

- Hyperactivité : Les γδ T des patients atteints de MA montrent une hyperactivité en réponse à des stimuli tels que les peptides amyloïdes, entraînant une production accrue de cytokines pro-inflammatoires
- Modulation de l'inflammation : Les γδ T peuvent moduler les réponses des cellules gliales par la production de cytokines, ce qui peut exacerber la neuroinflammation et la dégénérescence neuronale.

4. Réponse aux facteurs de risque de la MA

Bêta-amyloïde (Aβ) : Les cellules γδ T réagissent aux dépôts d'Aβ, ce qui peut induire une réponse immunitaire qui aggrave la pathologie.

5. Cytokines et mécanismes d'Action:

- Production d'IL-17 : Les cellules γδ T produisent de l'IL-17, qui est associé à la promotion de l'inflammation et à la mort neuronale

- Interaction avec les cellules T régulatrices : Les γδ T peuvent inhiber les Tregs, ce qui contribue à une réponse immunitaire déséquilibrée, aggravant ainsi les symptômes de la MA

Les cellules T gamma delta ont un rôle significatif dans l'aggravation des troubles neurologiques associés à la maladie d'Alzheimer. Leur capacité à moduler les réponses inflammatoires et à interagir avec d'autres cellules immunitaires en fait des cibles potentielles pour de nouvelles stratégies thérapeutiques visant à atténuer la progression de la maladie.

Chapitre III : 2sar , 2sar inverse , doking

Chapitre II:
Modélisation Computationnelle

II. INTRODUCTION :

La modélisation computationnelle est devenue une technique établie pour encoder des représentations mathématiques des processus cellulaires et acquérir des insights mécanistes qui conduisent à des prévisions testables. Ces modèles sont souvent construits à l'aide d'interfaces graphiques ou de langages spécifiques à un domaine, avec des normes communautaires utilisées pour l'échange. Les modèles subissent une analyse d'état stationnaire ou dynamique, ce qui peut inclure la simulation et l'étalonnage au sein d'une seule application, ou le transfert entre divers outils. Nous nous concentrons sur Python - un langage de programmation populaire avec un vaste écosystème de packages scientifiques. La modélisation programmatique est une technologie clé pour permettre le développement collaboratif de modèles et améliorer la diffusion, la transparence et la reproductibilité. [26]

L'augmentation de la puissance de calcul et la disponibilité croissante des données ont permis le développement de nouveaux types de modèles computationnels représentant davantage de détails des systèmes cibles. Cela permet de réaliser des expériences virtuelles sur le thème "Que se passerait-il si ?" — y compris la modification des règles de fonctionnement de ces détails — avant d'expérimenter dans la réalité. L'analyse et l'explication ne représentent que le point de départ de l'utilité des modèles. Ceux-ci peuvent aider à visualiser, prédire, optimiser, réguler et contrôler des systèmes complexes. La modélisation computationnelle, une technologie pouvant informer ou induire en erreur. La modélisation peut être appliquée correctement ou mal appliquée. C'est pourquoi une meilleure compréhension des processus de modélisation computationnelle et une sensibilisation accrue à quand et comment les modèles peuvent être utilisés de manière fiable sont essentielles. [27]

II.2. DEFINITION:

La modélisation computationnelle consiste à utiliser des ordinateurs pour simuler et analyser des systèmes complexes en s'appuyant sur les mathématiques, la physique et l'informatique. Un modèle computationnel comprend de nombreux paramètres qui caractérisent

le système étudié. La simulation se réalise en ajustant ces paramètres, seuls ou en combinaison, puis en observant les résultats. La modélisation par ordinateur permet aux scientifiques de réaliser des milliers d'expériences simulées. Ces expériences informatiques permettent d'identifier les quelques expériences en laboratoire les plus susceptibles de résoudre le problème étudié. Les modèles computationnels d'aujourd'hui peuvent examiner un système biologique à plusieurs niveaux. Les modèles relatifs à l'évolution des maladies incluent des processus moléculaires, des interactions entre cellules, ainsi que l'impact de ces changements sur les tissus et les organes. [28]

II.3 MODELISATIONS COMPUTATIONNELLES CLASSIQUES :

II.3.1 Méthodes génératives: [29]

II.3.1.1 Méthodes génératives et PCA :

Les méthodes génératives permettent d'extraire des informations à partir du mouvement en les comparant à leurs mesures. L'analyse en composantes principales (PCA) repose sur les principes de l'algèbre linéaire et postule que les données mesurées suivent une fonction linéaire, représentable par une matrice. La première étape consiste à déterminer les paramètres de cette matrice pour qu'elle reflète correctement la relation entre les entrées et les sorties. La stratégie de la PCA implique de classer les vecteurs propres de cette matrice selon leur impact sur la précision de la fonction, mesurée par l'erreur résultant de l'application du vecteur sur les données d'entrée. Le vecteur propre qui capture le principal comportement des données devient le premier composant, permettant ainsi de réduire la dimensionnalité tout en conservant un nombre limité de composants principaux.

II.3.1.2 Méthodes analytiques et équations différentielles :

Les méthodes génératives analytiques, telles que les polynômes et les équations différentielles, formalisent le système observé et permettent d'approximer une fonction à l'aide d'outils mathématiques classiques. La méthode la plus simple consiste à modéliser une fonction à l'aide d'un polynôme, considéré comme un approximateur universel. Cela signifie qu'en théorie, n'importe quelle fonction peut être représentée par un polynôme, dont les paramètres à identifier incluent l'ordre et les coefficients. Toutefois, cette approche présente des limites, particulièrement pour intégrer la dimension temporelle dans la représentation d'une trajectoire.

II.3.1.3 Modèles de type boîte noire et Markov cachés :

Les méthodes génératives de type boîte noire, telles que les réseaux de neurones artificiels (ANN) et les modèles de Markov cachés (HMM), sont également envisageables. L'approximation d'un système peut être réalisée à travers ces modèles computationnels, qui consistent en algorithmes générant une fonction de transfert à partir des mesures. **La figure II.1** illustre la structure du réseau de neurones artificiels, qui se compose d'un graphe de nœuds interconnectés par des liaisons pondérées. Chaque nœud représente une fonction d'intégration, souvent non linéaire. Un reproche souvent fait aux réseaux de neurones est leur manque de transparence, rendant difficile l'extraction des raisons derrière une décision. Il est donc envisageable d'utiliser des algorithmes offrant une meilleure lisibilité, comme la programmation génétique. Enfin, un modèle computationnel peut également être construit à l'aide de modèles de Markov cachés, représentant le système comme ayant une probabilité conditionnelle sur l'état actuel et recherchant les paramètres stochastiques des transitions entre différents états.

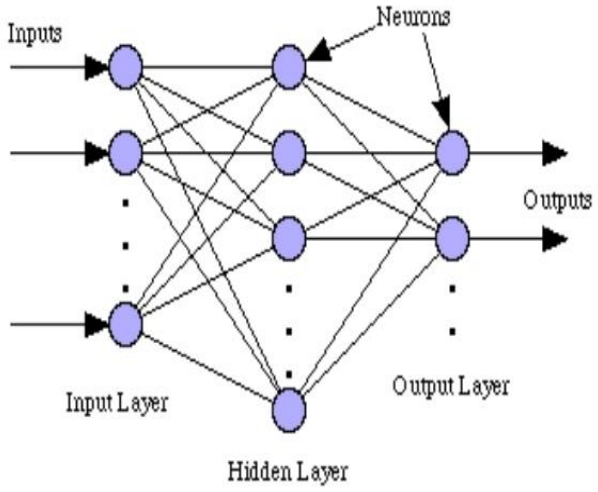


Figure II.1 – Exemple de perceptron à une couche cachée

II.3.2. Méthodes discriminantes : [29]

II.3.2.1 Méthodes discriminantes heuristiques :

Les méthodes discriminantes heuristiques (machines à états) déterminent la catégorie des données. Une approche simple pour déterminer la catégorie d'un vecteur de données par une heuristique basique. On peut définir un "raisonnement algorithmique" donnant comme résultat

l'appartenance d'un jeu de données à une catégorie ou une autre. Il existe différentes manières de représenter l'algorithme, que ce soit par un diagramme de traitement des données ou par une machine à états. Cet algorithme effectue des tests et des mesures sur les données fournies pour déterminer la catégorie à laquelle elles appartiennent.

II.3.2 2. Méthodes discriminantes linéaires :

Les méthodes discriminantes linéaires (KNN, SVM, LDA) abordent le problème de classification en recherchant une frontière dans les données. Cette approche travaille avec une base de données. Une nouvelle donnée est comparée aux données déjà enregistrées ou à une abstraction de cet ensemble. Une abstraction possible consiste à représenter un ensemble de données bruitées suivant une loi gaussienne par sa moyenne et son écart type.

La méthode KPPV (K-plus proches voisins) fonctionne selon un critère de distance euclidienne. Un vote d'appartenance est effectué par les k voisins les plus proches. L'appartenance majoritaire détermine l'appartenance de la nouvelle donnée. Une optimisation consiste à définir un centre des données d'une catégorie et à mesurer la distance de la donnée par rapport aux moyennes (k-means). La méthode "support vector machine" (SVM) se concentre plutôt sur la frontière. Pour cela, les données les plus proches de la frontière sont identifiées et appelées vecteurs de support, et une frontière est calculée pour maximiser la distance à ces vecteurs. Ainsi, l'ajout de données ne modifie la frontière que si la donnée est plus proche que les vecteurs de support actuels.

II.3.2.3 Méthodes discriminantes stochastiques et arbres de décision :

Les méthodes discriminantes stochastiques (réseaux bayésiens, logique floue) s'appuient sur une modélisation probabiliste des données. La logique floue consiste à définir un degré d'appartenance à une variable. L'association des ensembles flous permet d'obtenir un comportement dynamique de gestion des seuils. Ces fonctions d'appartenance remplacent une information binaire par une continuité. Si on formalise cela comme une distance, on peut retrouver une forme de frontière utilisée pour séparer les catégories, semblable à une distribution gaussienne autour du problème.

Les réseaux bayésiens construisent un réseau d'unités de probabilités avec la même topologie que les réseaux de neurones. Les nœuds présentent des fonctions de probabilités basées sur la loi de Bayes. Les méthodes discriminantes par arbre de décision sont également possibles

pour classer les données. De la même manière, les arbres de décision peuvent être vus comme des modèles génératifs, construisant un ensemble de frontières qui se suivent. On peut ainsi utiliser des méthodes basées sur les arbres de décision pour obtenir une méthode de vote, où les décisions sont discriminantes entre les solutions possibles.

II.4. APPLICATIONS DANS LES SCIENCES BIOMEDICALES :

La modélisation computationnelle joue un rôle crucial dans diverses applications, telles que les sciences biomédicales :

II.4.1 Modélisation des maladies :

La modélisation de maladies est un élément fondamental de la recherche biomédicale qui englobe la création de systèmes représentatifs imitant le comportement des maladies dans un environnement contrôlé. Ces modèles aident les chercheurs à mieux comprendre les mécanismes sous-jacents des maladies, à tester l'efficacité de traitements potentiels et finalement, à ouvrir la voie à de meilleurs soins aux patients.

II.4.1.1. Types des modèles

A. Modèles épidémiologiques :

Les modèles épidémiologiques utilisent des équations mathématiques pour décrire comment les maladies infectieuses se propagent au sein des populations, en tenant compte des dynamiques complexes d'interaction entre les individus. Ces modèles permettent d'analyser les tendances de transmission, d'identifier les facteurs influençant la propagation et d'évaluer l'impact des interventions de santé publique.

B. Modèle SIR: [30]

Le modèle SIR, introduit par Kermack et McKendrick en 1927, est un cadre mathématique utilisé pour prédire la propagation des maladies infectieuses au sein d'une population. Il s'agit d'un modèle déterministe qui utilise des équations différentielles pour diviser la population en trois compartiments :

- **Susceptibles (S)** : Les individus qui peuvent contracter la maladie.
- **Infectés (I)** : Les individus actuellement infectés et pouvant transmettre la maladie.
- **Rétablis (R)** : Les individus qui se sont rétablis de la maladie et qui sont supposés avoir une immunité.

B.1. Caractéristiques clés du modèle SIR :

- **Transitions** : Les individus passent d'un compartiment à l'autre en fonction des interactions et du taux de transmission du virus. L'infection se produit par contact entre des individus susceptibles et infectés.
- **Hypothèses** : Le modèle suppose une population homogène de taille constante, où tous les individus partagent le même niveau de susceptibilité à l'infection. Une fois rétablis, les individus acquièrent une immunité à vie.
- **Équations** : Le modèle est gouverné par un ensemble d'équations différentielles non linéaires qui décrivent les changements dans chaque compartiment au fil du temps :
- Changement dans les Susceptibles:

$$\frac{dS}{dt} = \beta . I . S \qquad (II.1)$$

- Changement dans les Infectés :

$$\frac{dI}{dt} = \beta . I . S - \gamma . I = I . (\beta . S - \gamma) \qquad (II.2)$$

- Changement dans les Rétablis :

$$\frac{dR}{dt} = \gamma . I \qquad (II.3)$$

- Nombre de reproduction (R0) :

$$R_0 = \frac{\beta}{\gamma} \qquad (II.4)$$

Ce nombre est crucial pour évaluer la transmissibilité de la maladie et le potentiel de flambées.

Explications :

R Récupéré.

R0 Numéro de reproduction.

I Infecté/infectant.

S Contagieux (sensible)

t temps (T)

β nombre de contacts (infectieux avec infectés) par jour (nombre moyen de personnes qui entrent en contact étroit avec un individu infecté donné par jour) [T-1].

γ inverse de la durée moyenne de la maladie [T-1]

B.2. Extensions et applications

Le modèle de base peut être étendu pour incorporer l'hétérogénéité de la population, permettant une représentation plus réaliste de facteurs tels que l'âge, la localisation et le comportement. En faisant cela, les chercheurs peuvent simuler diverses stratégies d'intervention (par exemple, quarantaine, vaccination) et évaluer leur efficacité dans le contrôle de la propagation des maladies.

II.4.1.2 Importance et limitations

Le modèle SIR fournit un outil précieux pour comprendre la dynamique épidémique et planifier des interventions de santé publique. Cependant, ses simplifications et hypothèses doivent être interprétées avec prudence, surtout dans le contexte des complexités du monde réel.

- **Qualité des données** : L'estimation précise des paramètres, en particulier du taux de contact (β), peut être difficile et nécessiter des ajustements basés sur des données observées ou des interventions.
- **Utilité** : Bien qu'il s'agisse d'un modèle déterministe, le cadre SIR est utile lorsque les données détaillées sont rares, offrant une approche simple pour modéliser la dynamique des maladies infectieuses.

C. Modèle SEIR [31]

Les modèles mathématiques peuvent être conçus pour simuler l'effet de la maladie à plusieurs niveaux. Ces modèles peuvent être utilisés pour évaluer la maladie depuis l'intérieur du modèle de l'hôte, c'est-à-dire l'influence des interactions au sein des cellules de l'hôte, jusqu'au modèle de métapopulation, c'est-à-dire comment la maladie se propage dans des populations géographiquement séparées. La partie la plus importante de ce modèle est de calculer la valeur de R_0.

Valeur de R_0 :

- R_0 (nombre de reproduction de base) exprime le degré de contagiosité de la maladie. C'est l'objectif fondamental des épidémiologistes lors de l'étude d'un nouveau cas.
- En termes simples, R_0 détermine le nombre moyen de personnes pouvant être affectées par une personne infectée sur une période donnée.
    - Si $R_0 < 1$: cela signifie que la propagation est censée s'arrêter.
    - Si $R_0 = 1$: cela signifie que la propagation est stable ou endémique.
    - Si $R_0 > 1$: cela signifie que la propagation augmente en l'absence d'intervention.

Pour le calcul de la vaccination requise, l'équation suivante est utilisée pour calculer le pourcentage de la population qui doit être vaccinée pour stabiliser la propagation de la maladie :

$$\textbf{Population requise pour la vaccination} = 1 - \left(\frac{\textbf{Objectif R0}}{\textbf{R0 actuel}}\right) * \textbf{Population}$$

Le modèle SEIR aide à comprendre comment les maladies infectieuses se propagent, et estimer la valeur de R0 est crucial pour planifier des stratégies d'intervention, telles que la vaccination, afin de réduire la propagation des maladies**.**

II.4.1.3 Modèles biologiques

La simulation des systèmes biologiques repose sur deux paradigmes principaux de modélisation : l'approche population-centrée et l'approche individu-centrée. Le couplage de ces deux paradigmes donne naissance à une approche mixte, qui combine leurs avantages tout en atténuant leurs inconvénients. La littérature regorge de publications comparant ces approches dans divers domaines de la biologie, tels que la biochimie, la biologie cellulaire, la biologie animale, la microbiologie, l'immunologie et l'écologie. Cette section introduit les concepts fondamentaux des approches population et individu-centrées, ainsi que leur utilisation pour la simulation des systèmes biologiques complexes. [32]

A.1. Types de modèles

Modèles d'agents : Simulent le comportement d'individus autonomes (agents) dans un environnement, permettant d'étudier les dynamiques de populations ou d'écosystèmes.

Modèles de dynamique des populations : Analysent la croissance et l'interaction des populations d'espèces au fil du temps.

A.2. Applications

- **Recherche sur le cancer** : Utilisés pour modéliser la progression tumorale, les réponses aux traitements et les mécanismes de résistance.
- **Immunologie** : Aident à comprendre les interactions entre les cellules immunitaires et les pathogènes.
- **Biotechnologie** : Application dans le développement de nouveaux médicaments et thérapies géniques.

II.4.1.4. Modèles statistiques :

Les modèles statistiques sont des outils puissants utilisés pour analyser et interpréter des données. Ils permettent de faire des prédictions, d'étudier des relations entre variables et de tirer des conclusions à partir d'échantillons de données. Voici quelques concepts clés :

A.1. Types de Modèles Statistiques

A.1.1. Modèles de Régression :

La régression logistique est une technique statistique utilisée pour modéliser une variable dépendante dichotomique, permettant ainsi de prédire un résultat discret, tel que l'appartenance à un groupe, à partir d'un ensemble de variables indépendantes.

- **Régression Linéaire**

L'analyse linéaire de la régression, appelée aussi plus simplement régression linéaire, est l'une des méthodes statistiques les plus utilisées dans les sciences appliquées et dans les sciences de l'homme et de la société. Son objectif est double : il consiste tout d'abord à décrire les relations entre une variable privilégiée, appelée variable expliquée (ou dépendante), et plusieurs variables jouant un même rôle par rapport à la première, appelées variables explicatives (ou indépendantes). [33]

- **Régression Logistique**

La régression logistique est une méthode statistique utilisée pour prédire un résultat qui n'a que deux options (par exemple, oui ou non) à partir de plusieurs variables. Contrairement à d'autres techniques, elle ne nécessite pas que les données suivent certaines règles, comme une distribution normale. Cette méthode permet d'analyser comment différentes variables influencent la probabilité d'un événement spécifique, comme la présence ou l'absence d'une caractéristique**.** [34]

A.1.2. Modèles ANOVA :

L'analyse de variance (ANOVA) est la procédure standard utilisée pour l'inférence statistique dans les plans factoriels. Les valeurs p déterminent la significativité statistique en termes de « tout ou rien ». L'adoption généralisée des statistiques bayésiennes, et en particulier de l'ANOVA bayésienne, est limitée par le fait que les concepts bayésiens sont rarement enseignés dans les cours de statistiques appliquées. Par conséquent, les utilisateurs peuvent ne pas savoir comment effectuer une ANOVA bayésienne et en interpréter les résultats. Un guide a

été fourni pour réaliser et interpréter une ANOVA bayésienne avec JASP, un logiciel statistique open-source ayant une interface utilisateur graphique. [35]

A.1.3. Modèles de Séries Temporelles :

Une série temporelle se compose d'un ensemble d'observations, chaque observation étant associée à un instant de temps unique, ce qui permet de la représenter par un vecteur de n variables aléatoires, où les réalisations correspondent aux observations effectives. Il est crucial de considérer plusieurs séries temporelles simultanément, car les variables modélisées par ces séries sont souvent liées à d'autres variables. Une série temporelle multivariée X = {X1, . . ., Xn} est un ensemble fini de séries uni variées dont les valeurs évoluent ensemble, toutes indexées par un même vecteur temporel. [36]

II.5. CONCEPTION DES MEDICAMENTS :

La conception de médicaments représente une étape cruciale dans le développement de traitements médicaux efficaces. Elle consiste à créer des composés capables de cibler des maladies spécifiques et d'améliorer les résultats de santé des patients. L'importance de cette discipline réside dans sa capacité à transformer des idées novatrices en solutions thérapeutiques tangibles. La conception de médicaments assistée par ordinateur (Computer-Aided Drug Design ou CADD en anglais) est utilisée depuis de nombreuses années dans le secteur de la recherche pharmaceutique. La CADD apporte un soutien aux tests tout au long du processus de recherche d'un candidat-médicament, depuis l'identification des cibles biologiques jusqu'aux premières études précliniques. L'objectif premier de la CADD est d'accélérer le processus de recherche en aidant les chimistes médicinaux à orienter les choix stratégiques des candidats-médicaments.

II.5.1 Simulation des Interactions

II.5.1.1 Modélisation Moléculaire : [37]

La **modélisation moléculaire** est cruciale pour l'optimisation des structures obtenues par des techniques expérimentales telles que la cristallographie aux rayons X ou la résonance magnétique nucléaire. Elle englobe un ensemble de méthodes théoriques et techniques informatiques visant à simuler virtuellement le comportement des molécules. Elle se divise en deux axes principaux : le **graphisme moléculaire** et la **chimie computationnelle**. Le graphisme moléculaire est utilisé pour concevoir et représenter des modèles en 2D ou 3D, permettant ainsi

de visualiser les propriétés physico-chimiques telles que les zones hydrophobes et les sites actifs. D'autre part, la chimie computationnelle permet de reproduire ces propriétés en appliquant différents algorithmes selon la taille du système.

La modélisation moléculaire permet de prédire les conformations, d'anticiper les interactions entre ligands et récepteurs, d'étudier les chemins conformationnels associés à l'activité de macromolécules complexes et de mener des études structure/activité des complexes ligand-récepteur. L'une des applications majeures de la modélisation moléculaire est la **conception de nouvelles molécules**, ou drug design. Pour cela, les structures de ligands connus du récepteur servent de base. Les calculs de dynamique moléculaire identifient un maximum de conformations favorables à la fixation de la molécule. Les techniques de drug design orientent ainsi efficacement la conception de nouveaux médicaments**.**

II.5.1.2 Techniques Utilisées:

- **Dynamique Moléculaire** :

Une simulation de dynamique moléculaire consiste à calculer l'évolution d'un système de particules au cours du temps. Dans ce cas, le temps évolue de manière discrète et l'on souhaite connaître la position et la vitesse des particules à chaque pas de temps. Le calcul des forces d'interaction entre les particules permet de déterminer l'évolution des vitesses, et donc des positions, en utilisant les lois de la dynamique classique de Newton. L'énergie est constituée de l'énergie potentielle et cinétique du système considéré. Les forces correspondantes sont obtenues à partir des dérivées premières de l'énergie. Les équations du mouvement de Newton (F = ma) doivent être résolues, elles peuvent s'écrire sous la forme :

$$\frac{-dV}{dr} = m\frac{d^2r}{dt^2} \qquad \text{(II.5)}$$

Où V est l'énergie potentielle à la position r. Le vecteur r contient les coordonnées de toutes les particules du système. Plusieurs algorithmes peuvent être invoqués pour résoudre les équations du mouvement. En général les positions à un temps $t+\Delta t$ sont obtenues par extension en une série de Taylor impliquant la vitesse et l'accélération des particules.

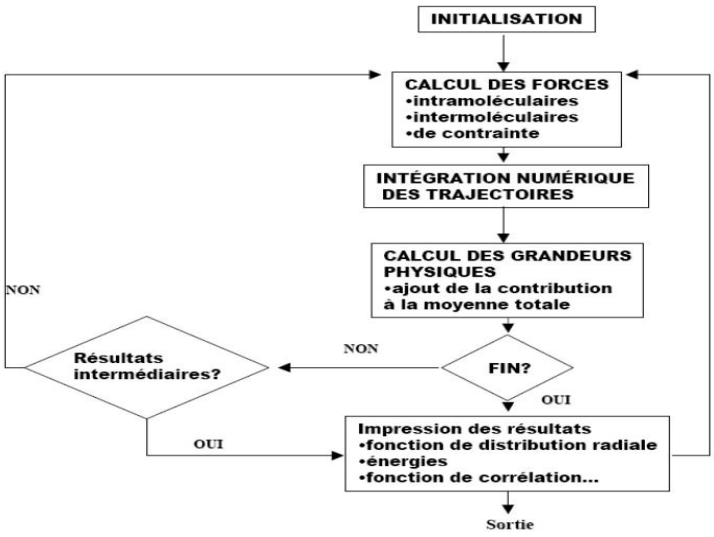


Figure II.2 : Principe général d'une simulation de Dynamique Moléculaire.

La dynamique moléculaire s'applique aussi bien à l'étude structurale des molécules qu'à des systèmes en interaction de grande taille. Néanmoins, les capacités de calcul étant limitées, le nombre de particules dans une simulation l'est aussi. Afin de réduire le nombre de molécules étudiées, tout en simulant un système le plus proche possible d'une phase condensée, la boîte contenant le système moléculaire est répliquée dans les trois directions de l'espace, formant ainsi un réseau infini (Conditions Limites Périodiques ou PBC en anglais). Si une particule quitte la boîte « par le haut », elle est immédiatement introduite par le bas. De plus, les interactions entre particules sont tronquées à une distance égale au maximum à la moitié de la taille de la boîte, afin qu'une particule n'interagisse jamais avec une seule de ses images (convention de l'image minimum), on parle alors de distance de troncature (cut-off). [38]

- **Modélisation de Docking** :

Il existe plusieurs algorithmes utilisés dans la modélisation de docking, et le choix de l'algorithme dépend souvent du nombre de molécules à tester et de la précision souhaitée, ces critères ayant un impact sur le temps de calcul et les ressources nécessaires. Les méthodes rapides, connues sous le nom de **docking rigide**, sont parmi les premières développées, où les molécules (le récepteur et le ligand) sont structurées de manière fixe. Avec l'amélioration des capacités de calcul, les méthodes **semi-flexibles**, qui permettent aux partenaires d'interaction d'être partiellement déformables (généralement un ligand flexible et un récepteur rigide), se sont

imposées, car elles offrent des résultats plus précis, même dans le cadre de criblages virtuels à haut débit.

Le docking semi-flexible se compose de deux étapes principales. La première, appelée **docking**, permet aux petites molécules testées d'explorer le site d'interaction avec la protéine. Au cours de cette phase, la molécule peut adopter plusieurs conformations possibles, et le programme génère différentes positions d'amarrage du ligand dans la protéine ainsi que plusieurs positions autour de celle-ci. La deuxième étape, nommée **scoring**, évalue les poses rencontrées lors de la première étape afin de retenir uniquement la meilleure, ce qui donne lieu à un calcul de score. [39]

II.5.1.3. Outils et Logiciels :

Il existe plusieurs logiciels de modélisation moléculaire, tels que :

- **Molecular Dynamics Software**: GROMACS, AMBER.
- **Logiciels de Mécanique Quantique** : Gaussian, VASP.
- **Outils de Docking** : AutoDock, Vina.

II.5.1.4 Analyse de Données :

- Big Data : Analyse de grandes quantités de données biologiques et chimiques pour identifier des patterns et des relations entre médicaments et cibles.
- Intelligence Artificielle : Utilisation d'algorithmes d'apprentissage automatique pour prédire les résultats des interactions et identifier de nouvelles cibles potentielles.

II.6. OUTILS ET TECHNIQUES MODERNES:

II.6.1. Nouvelles techniques de modélisation :

Dans le domaine de la modélisation computationnelle, plusieurs nouvelles techniques émergent pour améliorer la précision, l'efficacité et l'applicabilité des modèles. Voici quelques-unes des techniques les plus prometteuses :

II.6.1 1. Réseaux de neurones profonds:

Nous décrivons ici le modèle de réseau profond proposé par Hinton et ses collègues en 2006, qui repose sur un empilement de machines de Boltzmann restreintes (RBM). Ce modèle se caractérise par une architecture comprenant des unités visibles et des unités cachées, toutes interconnectées. L'apprentissage de ce réseau s'effectue par le biais d'évaluations statistiques

utilisant la méthode de recuit simulé. Ce processus se déroule en deux phases distinctes : une phase d'éveil, où le réseau est stimulé par des entrées externes, et une phase de repos, qui permet au système de se stabiliser en l'absence de nouvelles entrées. [40]

II.6.1 2. Modèles de simulation stochastique :

Les modèles de simulation stochastique sont des outils essentiels pour analyser des systèmes complexes où l'incertitude et le hasard influencent les résultats. Ils intègrent des éléments aléatoires pour simuler des processus tels que la propagation des épidémies, l'analyse des essais cliniques et les interactions biologiques. En exécutant de nombreuses itérations, ces modèles fournissent des estimations probabilistes qui aident à évaluer les risques et à optimiser les ressources dans le domaine de la biomedicine. Malgré leur complexité et les défis d'interprétation, leur capacité à représenter l'incertitude en fait des instruments précieux pour la prise de décision en santé publique. [41]

II.6.1.3 Modélisation par équations différentielles partielles (EDP):

La modélisation des phénomènes physiques repose généralement sur la résolution d'équations aux dérivées partielles (EDP), qui traduisent mathématiquement les lois fondamentales de la physique. Il est essentiel de valider le modèle en le comparant, lorsque cela est possible, aux données expérimentales. Souvent, les EDP sont non linéaires nécessitant l'utilisation d'outils de calcul numérique pour les résoudre. Lorsque les équations sont linéaires, il est bénéfique de trouver des solutions analytiques, car cela permet d'observer l'influence des différents paramètres et de réaliser une première étude d'optimisation du dispositif étudié. [42]

II.6.1.4 Modélisation integrative :

La modélisation intégrative est une approche systématique qui combine diverses échelles et modalités de données pour représenter des systèmes complexes. En intégrant des informations provenant de sources variées, telles que des données expérimentales, des modèles théoriques et des simulations numériques, elle vise à améliorer la compréhension des interactions au sein des systèmes, qu'ils soient biologiques, environnementaux ou technologiques. Cette méthode est particulièrement utile dans des domaines comme la biomedicine et l'écologie, où elle permet de prédire le comportement futur des systèmes tout en faisant face à des défis liés à la gestion de données hétérogènes et à l'interprétation des résultats.

II.6.1.5 Modélisation par apprentissage par renforcement : [43]

Les systèmes intégrant l'intelligence artificielle, tels que les chaînes d'approvisionnement numériques et les usines intelligentes, doivent obligatoirement intégrer l'IA dans leurs modèles de simulation. Par exemple, dans les systèmes d'analyse de simulation, les composants d'IA peuvent être directement intégrés pour permettre des tests et des prévisions. Dans certaines études, un algorithme d'apprentissage par renforcement comme le Q-Learning a été utilisé pour combiner des algorithmes d'équilibrage de charge dynamique avec des simulations à événements discrets des circuits VLSI.

L'optimisation par l'IA représente une opportunité d'intégration dans la modélisation de simulation. Les systèmes basés sur des agents présentent souvent de nombreux paramètres nécessitant des temps d'exécution importants pour explorer toutes les configurations possibles. L'IA peut accélérer cette phase de configuration, rendant l'optimisation plus efficace. De plus, la simulation peut également accélérer l'apprentissage et la configuration des algorithmes d'IA. Par exemple, des simulations DEVS sont utilisées pour tester diverses configurations des hyper-paramètres d'un réseau de neurones.

Dans le cadre de l'apprentissage par renforcement, des composants peuvent remplacer les modèles basés sur des règles, en tenant compte du comportement humain et de la prise de décision. En observant un comportement souhaité, il est possible de construire un modèle DEVS comportemental équivalent. Ces composants d'apprentissage peuvent être intégrés dans des modèles de simulation pour refléter le système réel ou pour former des composants d'IA. En générant les ensembles de données nécessaires, les modèles de simulation deviennent un outil puissant pour déployer des algorithmes d'apprentissage par renforcement.

II.6.2 Nouveaux outils logiciels :

II.6.2.1. Nouveaux outils logiciels :

Dans le domaine de la modélisation, plusieurs logiciels et outils récents ont été développés pour améliorer la précision et l'efficacité des modèles. Voici un aperçu des outils les plus récents et de leur impact :

1. TensorFlow:

TensorFlow est un Framework open source développée par Google pour la machine learning et le Deep learning. Il permet de construire et de déployer des modèles d'apprentissage

automatique à grande échelle. TensorFlow est particulièrement apprécié pour ses capacités de calcul distribué et son écosystème riche, qui inclut des bibliothèques pour le traitement de données, la visualisation et l'optimisation.

2. PyTorch:

PyTorch est un autre Framework open source, développé par Facebook, qui est largement utilisé pour le Deep learning. Il se distingue par sa flexibilité et sa simplicité d'utilisation, ce qui en fait un choix populaire parmi les chercheurs. PyTorch permet des calculs dynamiques, ce qui facilite le prototypage et l'expérimentation avec des modèles d'apprentissage automatique.

3. GROMACS :

GROMACS est un logiciel open source dédié à la dynamique moléculaire, utilisé principalement pour simuler des systèmes biologiques complexes tels que des protéines et des acides nucléiques. Il est particulièrement reconnu pour sa performance et son efficacité, permettant aux chercheurs de réaliser des simulations à grande échelle et d'analyser les interactions moléculaires avec précision.

4. Simul8 :

Simul8 est un logiciel de simulation conçu pour modéliser et analyser des systèmes complexes, notamment dans les secteurs de la fabrication et de la logistique. Il permet aux utilisateurs de visualiser le flux de travail et d'optimiser les processus en testant différentes configurations sans perturber les opérations réelles. Simul8 est apprécié pour sa facilité d'utilisation et ses capacités d'analyse.

5. Net Logo :

Net Logo est un environnement de programmation open source utilisé pour modéliser des systèmes complexes, en particulier dans les domaines des sciences sociales et de l'environnement. Il permet de créer facilement des modèles basés sur des agents, où chaque agent peut interagir avec son environnement et d'autres agents. Net Logo est idéal pour l'enseignement et la recherche grâce à sa simplicité et sa communauté active.

6. AnyLogic :

AnyLogic est un logiciel de simulation avancé qui supporte plusieurs paradigmes de modélisation, y compris la simulation basée sur les agents, la simulation de systèmes discrets et la simulation continue. Il est utilisé dans divers secteurs, tels que la gestion des opérations, la santé et la logistique. AnyLogic permet d'intégrer différents types de modèles dans un seul

projet, offrant ainsi une approche complète pour l'analyse et l'optimisation des systèmes complexes.

II.7. AVENIR DE LA MODELISATION :

La modélisation évolue rapidement grâce à l'augmentation de la puissance de calcul, à l'explosion des données disponibles et à la capacité croissante des modèles à traiter des systèmes extrêmement complexes. À l'avenir, il y aura un besoin accru de modèles fiables et prédictifs pertinents pour les systèmes complexes à grande échelle que nous souhaitons comprendre ou construire. Bien que des modèles plus grands et plus sophistiqués augmentent la capacité prédictive, ils nous permettront également de mieux saisir les limites de la prédiction, les incertitudes fondamentales, ainsi que la capacité aux points de basculement et au verrouillage.

Certains modèles fonctionneront en étroite collaboration avec des systèmes opérationnels, tirant des données potentiellement en temps réel. Ces données proviendront de nombreux capteurs et actionneurs désormais intégrés aux systèmes, et nous verrons de nouvelles formes de modélisation émerger en conséquence. Un aperçu des changements, des défis et des potentielles récompenses est donné pour la prochaine décennie :

- **Disponibilité des données** : La disponibilité de données à grande échelle sur les individus transformera la modélisation. Plutôt que d'utiliser des modèles agrégés basés sur des membres « moyens » de la population, il sera plus facile de modéliser les individus directement.
- **Données liées** : Les modèles nécessiteront des données plus largement interconnectées. Certaines données pourraient provenir d'autres modèles, nécessitant des liens supplémentaires vers des données dérivées.
- **Échelles multiples** : La modélisation couvrira de nombreuses échelles et niveaux de détail. En réunissant des communautés de modélisation de différentes disciplines, nous verrons des moyens plus sophistiqués de lier les modèles pour décrire des systèmes à plusieurs niveaux de détail.
- **Automatisation de la modélisation** : De plus en plus de modèles seront construits par des ordinateurs, qui pourront générer des modèles à partir des données par inférence automatisée ou semi-automatisée.
- **Entraînement des ordinateurs** : Les modèles aideront à former des ordinateurs en générant des données représentant des échecs, ce qui peut être difficile à recueillir dans le monde réel.

- **Intégration des systèmes** : De plus en plus de systèmes deviendront partie intégrante des modèles, et inversement, des modèles seront intégrés dans des systèmes. Cela changera la dynamique entre la modélisation et le déploiement des systèmes.
- **Nouvelles technologies** : Des simulateurs quantiques spécialisés deviendront disponibles, permettant de développer des modèles prédisant les propriétés des matériaux ou des médicaments.
- **Capteurs omniprésents** : Les capteurs et actionneurs intelligents nécessiteront de nouvelles formes de modélisation, prenant en compte la dégradation et la décalibration au fil du temps.
- **Utilisation stratégique** : La modélisation sera de plus en plus utilisée pour des questions stratégiques et de politique, ajoutant des détails aux scénarios futurs potentiels.
- **Implication des décideurs** : Les décideurs senior participeront plus souvent à la construction et à l'utilisation de modèles, ce qui sera vu comme une approche solide pour gérer la complexité.
- **Orientation vers l'humain** : Certains modèles seront plus orientés vers les caractéristiques personnelles des individus, offrant la possibilité de fournir des données personnelles pour la modélisation.
- **Entraînement des humains** : Les simulateurs seront utilisés plus largement pour former des professionnels dans divers domaines, y compris la médecine et la sécurité.
- **Compréhension des systèmes complexes** : Les modèles deviendront essentiels pour comprendre les propriétés de systèmes de plus en plus complexes, comme l'Internet, qui est continuellement modélisé et surveillé pour explorer ses comportements.

II.7. APPLICATION ACTUELLE :

Aujourd'hui, avec l'avancement technologique, nous voyons déjà des applications de ces concepts. Par exemple, les données massives permettent de mieux cibler les traitements médicaux, tandis que les simulateurs quantiques commencent à transformer la recherche sur les matériaux. L'intégration de l'intelligence artificielle dans la modélisation contribue également à des prévisions plus précises et à une meilleure prise de décision dans divers secteurs, y compris la santé, la finance et l'environnement. Les défis éthiques autour des données personnelles et de la responsabilité des algorithmes demeurent cruciaux dans ce contexte en évolution rapide.

II.8. CONCLUSION:

En conclusion, la modélisation computationnelle joue un rôle essentiel dans la compréhension et la gestion des systèmes complexes. À travers ce travail, nous avons exploré comment ces modèles, qu'ils soient basés sur des méthodes génératives, discriminantes ou des approches spécifiques aux domaines tels que la biomedicine, offrent des outils puissants pour prévoir et optimiser divers processus. La capacité à simuler des scénarios variés et à analyser des interactions complexes permet non seulement d'économiser du temps et des ressources, mais aussi d'élargir notre compréhension des phénomènes que nous étudions.

L'intégration des technologies modernes, comme l'intelligence artificielle et les algorithmes d'apprentissage, renforce encore davantage l'efficacité des modèles computationnels. L'utilisation de logiciels spécialisés, tels que GROMACS pour la dynamique moléculaire ou des outils de modélisation statistique, démontre l'importance croissante de ces outils dans des secteurs variés, allant de la recherche biomédicale à l'urbanisme.

Cependant, il est crucial de reconnaître que la modélisation computationnelle n'est pas exempte de défis. Une compréhension approfondie des processus de modélisation et de leurs limites est indispensable pour garantir des résultats fiables et pertinents. Cela nécessite une collaboration étroite entre les modélistes, les décideurs et les utilisateurs finaux afin d'assurer l'utilisation appropriée et éthique des modèles.

À l'avenir, la modélisation computationnelle continuera à évoluer, offrant des opportunités pour des découvertes et des innovations majeures. En exploitant les avancées technologiques et en favorisant une culture de la transparence et de la collaboration, nous pouvons espérer transformer nos interactions avec le monde et aborder des problèmes complexes avec des solutions plus efficaces et créatives.

Chapitre III : 2sar , 2sar inverse , doking

Chapitre III :

Qsar et Qsar inverse

III.1. DEFINITION :

Le terme QSAR désigne des méthodes qui mettent en corrélation la structure moléculaire avec des propriétés telles que l'activité biologique in vitro ou in vivo. Parfois, les méthodes QSAR prennent des noms plus spécifiques comme : QSPR (relation quantitative structure-propriété) ou QSTR (relation quantitative structure-toxicité). Les QSPR sont utilisées pour les propriétés physicochimiques, comme le point d'ébullition, la solubilité, logP. Par définition, un QSAR est un modèle mathématique qui relie un ou plusieurs paramètres issus de la structure chimique à une mesure quantitative d'une propriété ou d'une activité. [44]

III.2. OBJECTIF:

Les objectifs pratiques d'une QSAR sont variés, et ces techniques trouvent des applications dans de nombreuses situations. Ainsi, les études in silico visent notamment à atteindre les objectifs suivants :

- Prédire l'activité biologique et les propriétés physico-chimiques par des moyens rationnels.
- Comprendre les mécanismes d'action au sein d'une série de produits chimiques.
- Economies sur les couts de développement de produits.
- Les prévisions pourraient réduire la nécessité de tests sur les animaux longs et Couteux. [45]

III.3. PRINCIPE :

Le QSAR se rapporte à la construction de modèles prédictifs d'activités biologiques, en fonction de l'information de structure moléculaire d'une série de composés. C'est un procédé qui relie de manière quantitative les propriétés moléculaires aussi bien électroniques que géométriques, appelées descripteurs, avec une activité biologique, suivant un modèle mathématique : **[Activité biologique] = f [Descripteurs]**

Dans l'équation, l'activité biologique est normalement exprimée comme log [1/C], où C'est généralement la concentration minimum requise pour causer une réponse biologique définie.

Au départ, un groupe d'entités chimiques ou de molécules sont identifiées qui présentent l'activité biologique souhaitée. Une relation quantitative est établie entre les caractéristiques physico-chimiques des molécules actives et l'activité biologique.

Le modèle QSAR développé est ensuite utilisé pour optimiser les composés actifs afin de maximiser l'activité biologique pertinente. Les composés prédits sont ensuite testés expérimentalement pour l'activité souhaitée. La méthode QSAR peut donc être utilisée comme un outil de guide pour l'identification des modifications de composés présentant une meilleure activité.

Avant de construire un modèle par la QSAR, il faut, à l'avance, avoir des connaissances multidisciplinaires de base qui interviennent dans ce processus. Ils se divisent en 3 piliers principaux, à savoir les données biologiques, les descripteurs moléculaires et la modélisation. [44]

III.3. 1. Données biologiques :

Les données biologiques utilisées dans le QSAR proviennent d'essais en laboratoire sur des systèmes vivants (animaux ou végétaux). Ils sont mesurables et quantifiables (concentration d'enzymes ou d'hormones spécifiques, distribution phénotypique de gènes spécifiques dans une population, etc.) et servent d'indices pour les évaluations physiologiques liées à la santé, au risque de maladie, aux troubles psychiatriques, à l'exposition environnementale et à ses effets, aux processus métaboliques, à la toxicomanie, etc. Les données biologiques sont généralement exprimées sur une échelle logarithmique. De plus, le logarithme inverse de l'activité (log 1/C) est également utilisé pour obtenir des valeurs mathématiques plus élevées lorsque les structures sont biologiquement très efficaces. [44]

III.3. 2. Approche Globale :

Pour créer un modèle, il est primordial de commencer par rassembler un maximum de données expérimentales fiables. Par la suite, il faut développer une série de descripteurs qui décrivent les structures moléculaires des composés présents dans la base de données, afin de les relier à l'activité ou à la propriété expérimentale en question. Une fois le modèle élaboré, il doit être validé sur la base de la corrélation. L'impact des composés dans le jeu d'entraînement sur le modèle est évalué à l'aide de techniques de validation interne. Enfin, pour évaluer la capacité

prédictive du modèle, il est essentiel de disposer de données expérimentales supplémentaires, ce qui constitue la validation externe. [46]

III.3. 3. Descripteurs moléculaires : [44]

Les propriétés chimiques, telles que la réactivité des substances, nous disent que plus nous avons de substances plus il y a de nouveaux produits chimiques qui pourraient être développés. Les descripteurs moléculaires représentent la forme informatisée des propriétés chimiques des molécules. Ils transforment la structure du composé en un ensemble de valeurs numériques ou binaires représentant diverses propriétés moléculaires qui sont jugées importantes pour expliquer l'activité.

Chaque descripteur moléculaire ne prend en compte qu'une petite partie de l'ensemble des informations chimiques contenues dans la molécule réelle et, par conséquent, le nombre de descripteurs ne cesse d'augmenter avec la demande croissante d'études approfondies sur les systèmes chimiques et biologiques. Les différents descripteurs ont des méthodes ou des perspectives indépendantes pour visualiser une molécule, en tenant compte des diverses caractéristiques de la structure chimique.

Les descripteurs moléculaires font désormais partie des variables les plus importantes utilisées dans la modélisation moléculaire et, par conséquent, gérées par les statistiques et la chimie informatique.

La dimensionnalité des descripteurs moléculaires peut identifier le type de modèle QSAR comme décrit ci-dessous : [47]

0D QSAR : ce sont des descripteurs dérivés de formules moléculaires, par ex. poids moléculaire, nombre et type d'atomes, etc.…

1D QSAR : une représentation de liste de sous-structures d'une molécule peut être considérée comme une représentation moléculaire unidimensionnelle (1D) et consiste en une liste de fragments moléculaires par exemple, des groupes fonctionnels, des anneaux, des liaisons, des substituant, etc.…

2D QSAR : un graphique moléculaire contient des informations topologiques ou bidimensionnelles (2D). Il décrit comment les atomes sont liés dans une molécule, à la fois le type de liaison et l'interaction d'atomes particuliers (par exemple, le nombre total de chemins, les indices de connectivité moléculaire, etc.).

3D QSAR : ceux-ci sont calculés à partir d'une représentation géométrique ou 3D d'une molécule. Ces descripteurs incluent la surface moléculaire, le volume moléculaire et d'autres propriétés géométriques. Il existe différents types de descripteurs 3D, par ex. électronique, stérique, forme etc.

4D QSAR : des informations à quatre dimensions sont décrites dans ce type de modèles, et la quatrième dimension est un ensemble de conformation de chaque ligand

5DQSAR : des informations en cinq dimensions sont décrites dans ce type de modèles, et la cinquième dimension est la possibilité de représenter un ensemble de jusqu'à six modèles d'ajustement induit différents.

Les descripteurs se répartissent en quatre classes : topologique, géométrique, électronique et hybride. Les descripteurs topologiques en chimie sont des invariants de graphes générés en appliquant les théorèmes de la théorie des graphes. Des exemples de descripteurs topologiques sont le nombre d'atomes, le nombre d'anneaux, le poids moléculaire, les chemins pondérés, les indices de connectivité moléculaire, les nombres de sous-structures, les descripteurs de bord de distance moléculaire, les indices kappa, les indices d'état électrotopologique et certains autres invariants.

Les aspects des structures liés aux électrons sont codés en calculant des descripteurs électroniques. Des exemples de descripteurs électroniques sont : les charges atomiques partielles, les énergies HOMO ou LUMO, le moment dipolaire. Des descripteurs géométriques sont utilisés pour coder les aspects 3-D de la structure moléculaire tels que les moments d'inertie, la surface accessible au solvant, les rapports longueur/largeur, les zones d'ombre, l'indice gravitationnel.

Une classe de descripteurs hybrides appelés descripteurs de surface partielle chargée code la propension des composés à s'engager dans des interactions polaires. L'ensemble des descripteurs est basé sur les charges atomiques partielles et la surface partielle de chaque atome. Les deux listes d'attributs sont mixtes et un ensemble d'environ 25 descripteurs peut être généré en faisant correspondre les deux listes mixtes avec des schémas de pondération différents. Des exemples de descripteurs peuvent inclure une surface positive fractionnée, une surface négative pondérée chargée.

III.4. METHODOLOGIE GENERALE D'UNE ETUDE QSAR :

Généralement une étude QSAR débute par la constitution de la base des données à partir des mesures expérimentales (propriété ou de l'activité de chaque composé) fiables et en nombre le plus important possible. Il s'agit ensuite de calculer un certain nombre de descripteurs car les paramètres qui décrivent l'activité ou la propriété étudiée sont mal connus, puis nous sélectionnons parmi les descripteurs calculés ceux qui sont pertinents par construction des modèles de RLM, PLS et RNLM et en divisant la base de données, aléatoirement, en une série d'apprentissage (training set) qui contient généralement les 3/4 de la base de données et une série de test (test set) constituée par le 1/4 restant. Ainsi les modèles mathématiques en utilisant la série d'apprentissage sont obtenus, et les modèles élaborés sont validés en utilisant la série de test et leurs paramètres statistiques sont calculés, et pour tester la fiabilité des résultats obtenus la méthode de validation interne est mise en jeu. Enfin les modèles élaborés peuvent être exploités pour comprendre les mécanismes et les modes d'action de l'activité étudiée et aussi pour prédire de nouvelles molécules avant même que celle-ci ne soient synthétisées. [48]

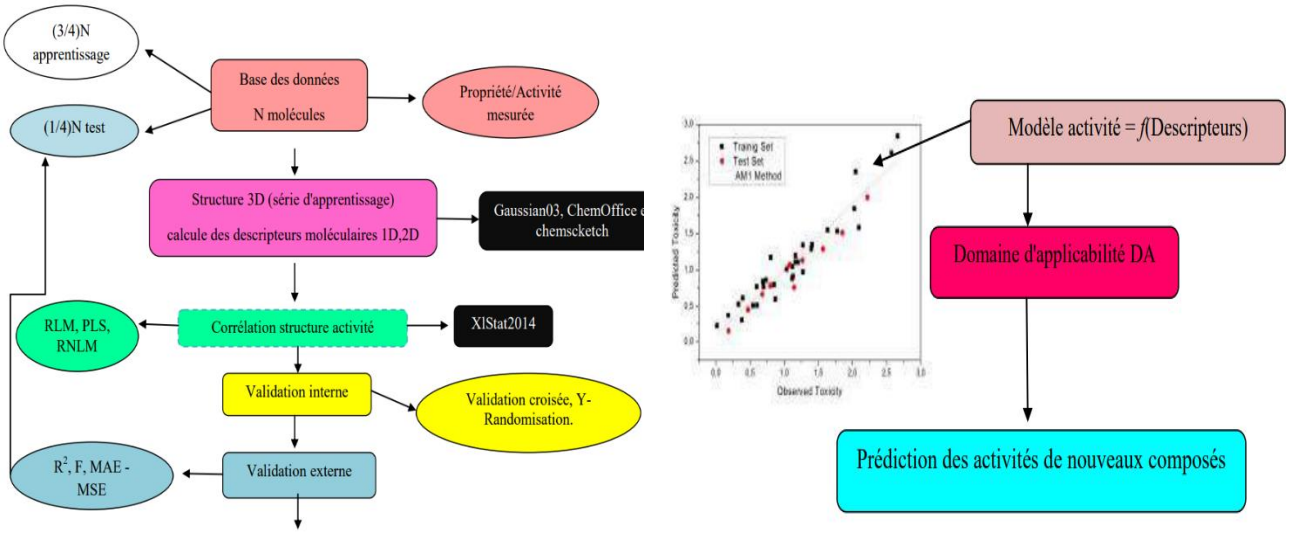


Figure III.1 : Methodologie generale d'une étude QSAR

III.4.1. Base de données et paramètres biologiques :

Le choix de la base de données est une étape très importante dans le développement des modèles QSAR, puisque ses constructions sont dépendantes des données expérimentales utilisées. Les données devraient, idéalement, être de grande qualité, ce qui signifie qu'elles

devraient être fiables et cohérentes. Il est donc impératif de choisir ceux ayant des incertitudes faibles afin de limiter les barres d'erreurs expérimentales.

De plus, le modélisateur doit s'assurer que les données expérimentales utilisées ont été obtenues selon le même protocole. En effet les conditions expérimentales ont, généralement, une forte influence sur les valeurs obtenues. Il faut également que la distribution des données soit la plus homogène et normale que possible, car la plupart des méthodes statistiques sont basées sur ce type de distribution. L'efficacité d'un modèle QSAR dépend également du type de molécules qui y sont incluses, plus le modèle présentera des composés de structures proches et similaires, plus il aura de la chance d'être performant. Les données biologiques sont généralement exprimées en logarithmes inverses (log1/C) afin d'obtenir des valeurs mathématiques plus significatives lorsque les structures sont biologiquement très efficaces. [48]

III.4.2. Descripteurs moléculaires

Les descripteurs sont des valeurs numériques qui encodent les caractéristiques physicochimiques des molécules à partir de leurs représentations structurales. La génération des descripteurs constitue l'étape cruciale lors de l'analyse de QSAR et la qualité d'un modèle dépendra essentiellement de la nature des descripteurs générés. Le Tableau III.1 présente une des manières de catégoriser les descripteurs classés en fonction de la dimension chimique utilisée pour les définir : [49]

Tableau III.1 : les descripteurs classés en fonction de la dimension chimique.

Dimension (D)	**Niveau de description**	**Exemple de descripteurs**
0-D	Formule chimique, propriété	Nombre d'atomes, poids moléculaire
1-D	Formule chimique développée	Dénombrement de fragments
2-D	Graphe moléculaire	Indices topologiques
3-D	Représentation 3-D	Surface moléculaire, surface polaire
4-D	Conformation des molécules	Energie potentielle des conformations

III.4.2.1. Les descripteurs 1-D :

Sont accessibles à partir de la formule brute de la molécule (par exemple C_6H_60H pour le phénol), et décrivent des propriétés globales du composé. Il s'agit par exemple de sa composition, c'est-à-dire les atomes qui le constituent, ou de sa masse molaire. On peut remarquer que ces descripteurs ne permettent pas de distinguer les isomères de constitution. [50]

III.4.2.2. Les descripteurs 2-D :

On peut obtenir ce type de descripteurs à partir de la structure développée de la molécule. On distingue : [51]

➢ **Les indices 2D constitutionnels :** Qui caractérisent les différents composants de la molécule. Il s'agit par exemple du nombre de liaisons simples ou multiples, du nombre de cycles, …etc.

➢ **les indices 2D topologiques :** Ce sont des quantificateurs numériques de topologie moléculaire qui sont mathématiquement dérivés de manière directe et sans ambiguïté du graphe structurel d'une molécule. Ils peuvent être sensibles à une ou plusieurs caractéristiques structurelles de la molécule telles que la taille, la forme, la symétrie, la ramification et la cyclicité Les indices topologiques sont principalement basés sur les distances entre les atomes calculées par le nombre de liaisons intermédiaires et sont donc considérés comme des indices de liaison traversante.

III.4.2.3. Les descripteurs 3-D :

Ces descripteurs sont évalués à partir des positions relatives des atomes dans l'espace, et décrivent des caractéristiques plus complexes. Leurs calculs nécessitent donc de connaître la géométrie 3D de la molécule. [52]

➢ **Les descripteurs géométriques :** Les plus importants sont le volume moléculaire, la surface accessible au solvant et le moment principal d'inertie. [53]

Ces descripteurs peuvent être obtenus soit par des méthodes expérimentales, soit par modélisation moléculaire. Ils reposent sur l'arrangement spatial des atomes qui composent la molécule et sont déterminés par les coordonnées des noyaux atomiques ainsi que par la taille de la molécule représentée. Ces descripteurs fournissent également des informations sur la surface moléculaire, calculée à partir des aires de Van der Waals et de leur superposition. [49]

➢ **Les descripteurs quantiques/électroniques**

Ces descripteurs permettent de quantifier les divers types d'interactions inter- et intramoléculaires, qui influencent fortement l'activité biologique des molécules. Le calcul de la plupart de ces descripteurs requiert la détermination de la géométrie correspondant à l'énergie minimale et fait souvent appel à la chimie quantique. [54]

III.4.3. Techniques utilisées :

III.4.3.1 réseau de neurones artificiels (ANN) :

Les réseaux de neurones artificiels (ANN) sont des programmes informatiques inspirés du cerveau humain, conçus pour simuler le traitement de l'information en apprenant à partir des données. Composés de neurones artificiels interconnectés par des poids organisés en couches, ils utilisent des fonctions de transfert pour produire des sorties non linéaires. Les ANNs optimisent les connexions entre unités durant l'entraînement pour minimiser les erreurs de prédiction. Ils sont particulièrement utiles dans le domaine pharmaceutique pour des applications telles que la classification, la reconnaissance de motifs, et l'interprétation de données analytiques, sans nécessiter de connaissances préalables sur la source des données. [60]

III.4.3.2 Support Vector Regression (SVR) :

L'analyse de régression par machines à vecteurs de support (SVR) est une technique d'apprentissage automatique supervisée utilisée pour résoudre des problèmes de régression, en analysant la relation entre les variables indépendantes et dépendantes. SVR formule un problème d'optimisation pour apprendre une fonction de régression reliant les entrées aux valeurs observées. Il équilibre la complexité du modèle et l'erreur de prédiction, tout en étant efficace avec des données de haute dimension.

Contrairement à l'algorithme SVM de classification qui produit des sorties binaires, SVR permet d'estimer des fonctions à valeur réelle. Il utilise des noyaux pour traiter les problèmes de régression non linéaire en projetant les caractéristiques dans un espace où les données peuvent être discrimées linéairement. [61]

III.4.3.3 L'algorithme d'optimisation des fourmis-lions (ALO)

L'algorithme d'optimisation des fourmis-lions (ALO) est une méthode inspirée de la nature qui simule le comportement de chasse des lions et des fourmis. Il repose sur la création d'une population de lions, représentant des agents d'optimisation, et de fourmis, symbolisant des solutions potentielles. Le processus commence par la génération aléatoire de lions qui cherchent des fourmis dans leur environnement. Les solutions sont évaluées selon une fonction objective, et les lions mettent à jour leurs positions en fonction des meilleures solutions trouvées. Ce processus est répété jusqu'à atteindre un critère d'arrêt. L'ALO se distingue par sa capacité à équilibrer l'exploration de nouvelles solutions et l'exploitation des meilleures, et trouve des applications dans divers domaines tels que l'ingénierie et la logistique. [62]

III.5. APPROCHE BASE SUR LE LIGAND

III.5.1. ADMET

ADMET est un acronyme qui désigne les étapes essentielles de la pharmacocinétique, c'est-à-dire l'analyse de l'absorption, de la distribution, du métabolisme et de l'élimination des médicaments dans l'organisme. Voici la définition de chaque composante : []

- **Absorption :** Il s'agit du mécanisme par lequel un médicament entre dans le corps, généralement par des voies telles que l'administration orale, intraveineuse, intramusculaire, transdermique ou inhalée. L'absorption détermine la quantité de médicament qui sera disponible pour une distribution dans l'organisme.
- **Distribution :** Cela se réfère à la manière dont le médicament est réparti dans les divers tissus de l'organisme, après son absorption. Ce processus est influencé par plusieurs facteurs, tels que la solubilité du médicament, le degré de liaison aux protéines plasmatiques et la taille moléculaire.
- **Métabolisme :** C'est le processus par lequel le médicament est transformé par des enzymes hépatiques ou d'autres tissus en métabolites qui peuvent être éliminés par les reins ou via la bile. Le métabolisme joue un rôle crucial dans l'efficacité et la sécurité du médicament.
- **Élimination :** Ce terme désigne le processus par lequel le médicament ou ses métabolites sont excrétés de l'organisme, principalement par l'intermédiaire des reins ou de la bile.

L'élimination est fondamentale pour maintenir des niveaux thérapeutiques appropriés du médicament et prévenir les effets toxiques.

- **Toxicité :** La toxicité fait référence à la capacité d'une substance à provoquer des effets nocifs chez les organismes vivants. Une substance toxique peut entraîner des dommages ou des dysfonctionnements au sein des cellules, des tissus, des organes ou des systèmes biologiques.

III.5.2. Docking :

Le Docking, également connu sous le nom de "Docking moléculaire", est une approche de modélisation informatique qui vise à prédire la manière dont deux molécules interagissent ou s'associent, généralement une petite molécule (comme un médicament) et une grande molécule (comme une protéine). L'objectif principal du Docking est de déterminer la géométrie et l'énergie de liaison des complexes formés par ces molécules.

Cette technique constitue un outil précieux dans le domaine de la découverte de médicaments, car elle permet d'anticiper comment un composé chimique interagira avec une cible biologique, tel qu'une protéine liée à une pathologie. Les informations obtenues sont ensuite utilisées pour concevoir de nouveaux composés chimiques qui présentent une affinité et une spécificité accrues envers leur cible, ce qui peut aboutir au développement de médicaments plus efficaces et plus sûrs. [55]

III.5.2.1. Principe du Docking :

Le docking moléculaire est une technique visant à prédire les interactions entre des ligands (tels que des substrats, activateurs ou inhibiteurs) et les acides aminés d'une protéine réceptrice. Ce processus se déroule en deux étapes distinctes.

- Docking : Cette première phase consiste à positionner le ligand dans le site actif de la protéine, en échantillonnant diverses conformations, positions et orientations possibles. Seules les poses représentant les interactions les plus favorables sont retenues.
- Scoring : La seconde étape évalue les interactions énergétiques potentielles (affinité) entre le ligand et la protéine, attribuant un score aux poses générées lors de la phase de docking. Ce score permet de sélectionner la meilleure pose parmi celles proposées.

Ces deux étapes varient en fonction du programme de docking utilisé... [55]

III.5.2.2. Les étapes du Docking :

Le processus de Docking se déroule en plusieurs étapes générales :

1. **Préparation des structures moléculaires** : Les structures moléculaires du ligand (la petite molécule) et du récepteur (la grande molécule) sont préparées à l'aide de techniques telles que la cristallographie aux rayons X, la résonance magnétique nucléaire (RMN) ou la modélisation informatique.
2. **Définition de l'espace de recherche** : Cet espace est déterminé en sélectionnant les zones de la surface du récepteur où le ligand peut potentiellement se lier. Cette sélection peut être fondée sur des données expérimentales ou des simulations de dynamique moléculaire.
3. **Exploration des poses** : Les différentes poses possibles du ligand sont explorées à l'aide d'algorithmes de recherche de conformation, tels que les méthodes de Monte Carlo ou les algorithmes génétiques.
4. **Évaluation des poses** : Chaque pose est analysée en termes de faisabilité géométrique et d'énergie, qui est calculée en utilisant des fonctions d'énergie tenant compte des forces de van der Waals, des liaisons hydrogène et des interactions électrostatiques.
5. **Raffinement des poses** : Les poses les plus prometteuses sont retenues pour un raffinement plus approfondi de leur conformation. Cela peut impliquer des techniques de minimisation d'énergie ou de dynamique moléculaire.
6. **Analyse des résultats** : Les poses finales sont classées selon leur affinité ou leur énergie de liaison prévisible. Ces informations sont utiles pour anticiper l'activité biologique ou pour orienter la conception de nouveaux composés. [47]

III.5.2.3. La fonction de score :

La fonction de score représente un élément clé du processus de Docking, permettant d'évaluer la qualité des différentes poses proposées pour l'interaction entre un ligand et un récepteur. Ce score mesure l'affinité entre le ligand et le récepteur, et est généralement calculé à l'aide d'une fonction d'énergie spécifique. Cette fonction de score prend en compte divers termes énergétiques qui décrivent les interactions entre les atomes du ligand et ceux du récepteur. Cela inclut notamment les forces de van der Waals, les liaisons hydrogène, les interactions électrostatiques, ainsi que des termes relatifs à l'entropie. En combinant ces

termes, la fonction de score détermine une énergie globale pour chaque pose du ligand dans son récepteur.

L'objectif principal est d'identifier la pose présentant l'énergie la plus basse, ce qui indique une affinité plus forte et une probabilité accrue que cette pose reflète la conformation réelle de la liaison. Les fonctions de score sont souvent développées et optimisées sur la base de données expérimentale concernant des liaisons connues, afin d'améliorer leur capacité à prévoir des liaisons potentielles. Il existe de nombreuses fonctions de score différentes utilisées dans le Docking, chacune ayant ses propres avantages et limites en termes de précision et de temps de calcul. Le choix d'une fonction de score dépend généralement de la nature des ligands et des récepteurs concernés, ainsi que des objectifs spécifiques de la recherche. [56]

III.5.2.4. Les types du Docking : [78]

Les trois principaux types du Docking moléculaire sont les suivants :

1. **Docking rigide (Rigid Docking)** : Dans cette approche, tant le ligand que le récepteur sont considérés comme des entités rigides, ce qui signifie qu'aucun changement de conformation n'est permis durant la simulation. Les positions et orientations optimales du ligand sont recherchées afin de maximiser l'ajustement dans le site actif du récepteur. Ce type de Docking est le plus simple et rapide, mais il ne prend pas en compte les changements conformationnels potentiels du récepteur lors de l'interaction.
2. **Docking semi-flexible (Semi-flexible Docking)** : Dans ce cas, le récepteur est considéré comme rigide, tandis que le ligand a la possibilité de subir des changements conformationnels limités. Cela permet à la molécule du ligand de s'ajuster à la cible en adoptant différentes conformations, mais le récepteur reste fixe. Cette méthode est plus réaliste que le Docking rigide, car elle offre une certaine flexibilité au ligand, tout en négligeant les éventuels ajustements conformationnels de la cible.
3. **Docking flexible (Flexible Docking)** : Cette approche permet à la fois au récepteur et au ligand de subir des changements de conformation durant la simulation. Cela permet de tenir compte des ajustements conformationnels possibles des deux molécules pendant leur interaction. Le Docking flexible est plus complexe et exige davantage de ressources de calcul, mais il fournit une représentation plus réaliste de l'interaction ligand-cible. Cette méthode est souvent privilégiée lorsque des changements conformationnels significatifs de la cible sont attendus lors de la liaison.

Ces trois types de Docking offrent divers niveaux de complexité et de réalisme dans la modélisation des interactions moléculaires. Le choix du type de Docking dépend des objectifs de l'étude, de la flexibilité supposée des molécules impliquées et des ressources disponibles en termes de temps de calcul et de capacité informatique.

III.6. QSAR INVERSE

III.6.1. Définition

QSAR inverse fait référence à l'utilisation inverse des techniques QSAR, où l'on analyse les données relatives à l'activité biologique d'un ensemble de composés pour déduire les structures chimiques pouvant exhiber cette même activité. Cette méthode est souvent employée pour la conception de médicaments et le développement de nouveaux composés. [57]

III.6.2. Principe du QSAR inverse

Illustré par le développement récent d'un nouveau cadre pour les relations quantitative structure-activité inverses, utilisant des réseaux de neurones artificiels et de la programmation linéaire mixte. Ce modèle, appelé modèle à deux couches, divise chaque graphe chimique en une partie extérieure et une intérieure, où l'extérieur se compose de sous-graphes induits acycliques maximaux, et l'intérieur représente le sous-graphe connexe. Cette approche permet une plus grande flexibilité dans l'inférence de divers types de graphes chimiques, offrant des perspectives prometteuses pour étendre les possibilités de modélisation QSAR inverse. [58]

III.6.3. Les étapes du QSAR inverse

Les étapes du QSAR inverse (Quantitative Structure-Activity Relationship inverse) peuvent être résumées comme suit : [59]

III.6.3. 1. Concept méthodologique :

QSAR inverse repose sur la dérivation de coordonnées de descripteurs pour un modèle et un ensemble de données donnés. La méthodologie présentée ici vise à identifier des coordonnées optimales au sein d'un espace de descripteurs prédéfini (espace x) basé sur une fonction de régression f(x) représentant un modèle QSAR. En restreignant la recherche au domaine d'applicabilité du modèle (AD), cette tâche devient un problème d'optimisation contraint (COP). Comme l'illustre la Figure III.2, l'objectif est de dériver de nouvelles coordonnées qui satisfont les critères d'évaluation prédéfinis de manière plus efficace que tout

point de données existant utilisé dans le modèle de régression original. Ainsi, le COP est formellement défini comme suit :

Minimize f(x)

s.t. gj (x) ≤ 0, j = 1, …, q

hj (x) = 0, j = q+1, …, m

x = (x1, …, xd), li ≤ xi ≤ ui, i = 1, …, d

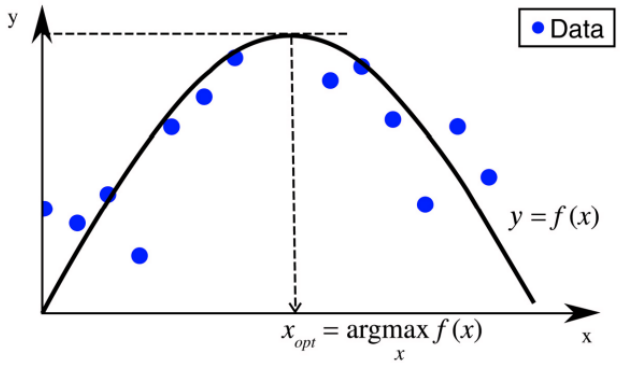


Figure III.2. : Concept d'optimisation : une fonction de régression f(x) ajuste les données d'entraînement pour déterminer de nouvelles coordonnées dans l'espace des descripteurs. Les coordonnées optimisées, basées sur f(x), se situent dans la plage des données d'entraînement mais donnent une valeur y plus élevée que tout autre point de données.

III.6.3.2. Regression and applicability domain model

Pour l'optimisation par évolution différentielle ε-contrainte (εDE), toute fonction de régression peut être employée. Dans cette étude, une régression par machines à vecteurs de support (SVR) avec un paramètre ν a été utilisée, tandis que le domaine d'applicabilité (AD) a été défini à l'aide d'un classifieur One-Class SVM (OCSVM) avec un paramètre ν. Le paramètre ν, variant entre 0 et 1, détermine la borne supérieure de la fraction d'erreurs de marge et la borne inférieure de la fraction de vecteurs de support. L'AD comprend les régions où la sortie de l'OCSVM est non négative.

Les modèles SVR et OCSVM ont tous deux utilisé un noyau gaussien (RBF) :

k (xi, xj) =exp(−γ‖xi−xj‖2)

Pour le ν-SVR, les hyperparamètres {C,ν,γ} {C,ν,γ} ont été optimisés par validation croisée sur les données d'entraînement, en se basant sur le Q2Q2. Pour l'OCSVM, γγ a été choisi de manière à maximiser la variance de la matrice de Gram issue de la fonction noyau des données d'entraînement, tandis que νν a été fixé à 0,01.

III.6.3.3. Simulation data:

Cette procédure concerne la génération de données aléatoires dans l'espace (x_1, x_2) en définissant des plages spécifiques pour x_1 et x_2, aboutissant à un ensemble de 50 points pour l'entraînement et 20 pour les tests ; les valeurs associées à y sont calculées à l'aide d'une fonction mathématique complexe appelée fonction oiseau, avec l'ajout d'un bruit aléatoire pour simuler des conditions réelles ; plusieurs essais indépendants sont réalisés en utilisant différents générateurs de nombres aléatoires pour renforcer la fiabilité des résultats, et les données sont ensuite tracées dans l'espace de sortie de la fonction, facilitant ainsi la compréhension des motifs et des relations entre les variables, ce qui illustre comment une fonction mathématique peut être utilisée pour générer des données et tester des modèles, aidant ainsi à analyser l'impact des variables et à comprendre les relations qui les unissent.

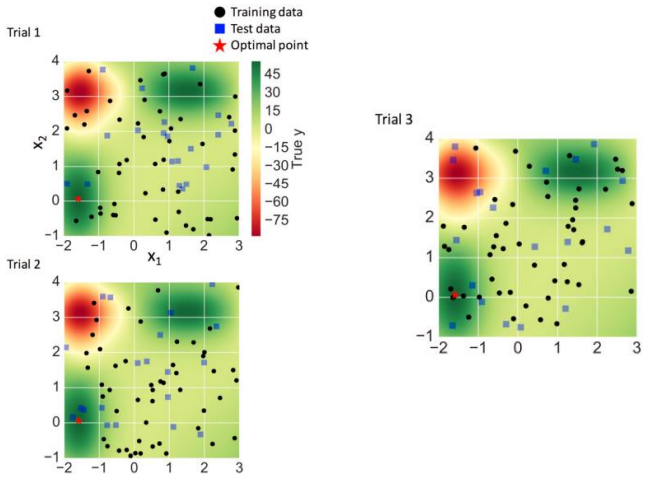


Figure III.3: Simulation data sets. In three independent trials, simulation data sets were generated. For each trial, training (black dots) and test blue squares) data are shown with true y values produced by the bird function f(x_1, x_2)

III.6.3.4. Compound data sets:

Des ensembles de composés ont été sélectionnés à partir de ChEMBL en appliquant les critères suivants :

Type de relation d'interaction : 'D'(directe)

- Unité standard d'activité: 'nM'
- Type standard d'activité: 'Ki'
- Relation standard d'activité: '='

Pour les composés présentant plusieurs valeurs de Ki, la moyenne géométrique a été calculée pour déterminer la valeur finale de puissance, à condition que toutes les mesures appartiennent au même ordre de grandeur (sinon, le composé était exclu).

Des filtres de sous-structure implémentés en interne ont été appliqués pour éliminer :

- Les composés interférents dans les tests (PAINS) (Pan Assay Interference Compounds)
- Les molécules réactives potentiellement problématiques.

III.6.3.5. Virtual screening:

Pour tester la capacité du criblage virtuel (VS) à identifier de nouveaux composés actifs à partir de coordonnées optimisées, ChEMBL (version 22) a été utilisé comme source de criblage. À partir de 1 414 176 composés uniques passant les filtres de sous-structure, les molécules d'entraînement utilisées pour la modélisation de chaque classe d'activité ont été supprimées. Tous les composés restants de ChEMBL ont fourni une grande source de criblage pour VS. Pour les composés de criblage, les descripteurs ont été calculés comme décrit ci-dessus et les composés entrant dans l'AD de chaque modèle spécifique à la classe ont été présélectionnés. Les composés actifs de chaque classe d'activité non utilisés pour l'entraînement représentaient des instances de test vraiment positives, quelles que soient leurs valeurs de puissance. Le calcul des descripteurs pour plus de 1,4 million de composés de criblage était exigeant en calcul et dépassait les exigences d'optimisation des coordonnées. Pour les composés de criblage ChEMBL, y compris les instances de test, deux classements VS ont été générés. Tout d'abord, les distances euclidiennes aux coordonnées optimisées ont été calculées. Dans ce cas, la puissance du composé n'a pas été prise en compte pour le classement. Deuxièmement, les valeurs de l'iCp ont été prédites pour tous les composés de criblage à l'aide des modèles SVR spécifiques à la classe. Ces derniers calculs ont été effectués pour déterminer si les vrais positifs étaient bien classés sur

la base des prédictions d'activité. L'aire sous les courbes caractéristiques de fonctionnement du récepteur (AUC) a été calculée comme critère d'évaluation.

III.7. CONCLUSION:

Le QSAR, ou relations quantitatives entre la structure et l'activité, représente une approche fondamentale dans le domaine de la chimie médicinale et de la biologie moléculaire. Grâce à des modèles mathématiques et statistiques, le QSAR permet de prédire l'activité biologique de composés chimiques à partir de leurs structures moléculaires. Cette méthode est particulièrement précieuse pour la découverte de nouveaux médicaments, car elle permet de rationaliser et d'accélérer le processus de conception.

En intégrant des données expérimentales sur l'activité de divers composés et en les liant à leurs caractéristiques structurales, le QSAR offre des outils puissants pour identifier des candidats médicamenteux prometteurs, optimiser des composés existants et réduire le besoin d'expérimentation in vitro et in vivo.

Cependant, malgré ses avantages, le QSAR présente des défis, notamment en matière de sélection des descripteurs appropriés, de validation des modèles et de compréhension des mécanismes d'action. L'évolution des techniques de modélisation, y compris l'apprentissage machine et l'intelligence artificielle, promet de surmonter certaines de ces limites, élargissant ainsi l'application du QSAR dans des domaines variés tels que la toxicologie, la bio-informatique et la pharmacologie.

En conclusion, le QSAR demeure une méthodologie essentielle qui lie la chimie et la biologie, facilitant le développement de nouveaux médicaments et contribuant ainsi à l'avancement des sciences de la vie.

CHAPITRE IV :
PROCEDURE
ET
RESULTATS

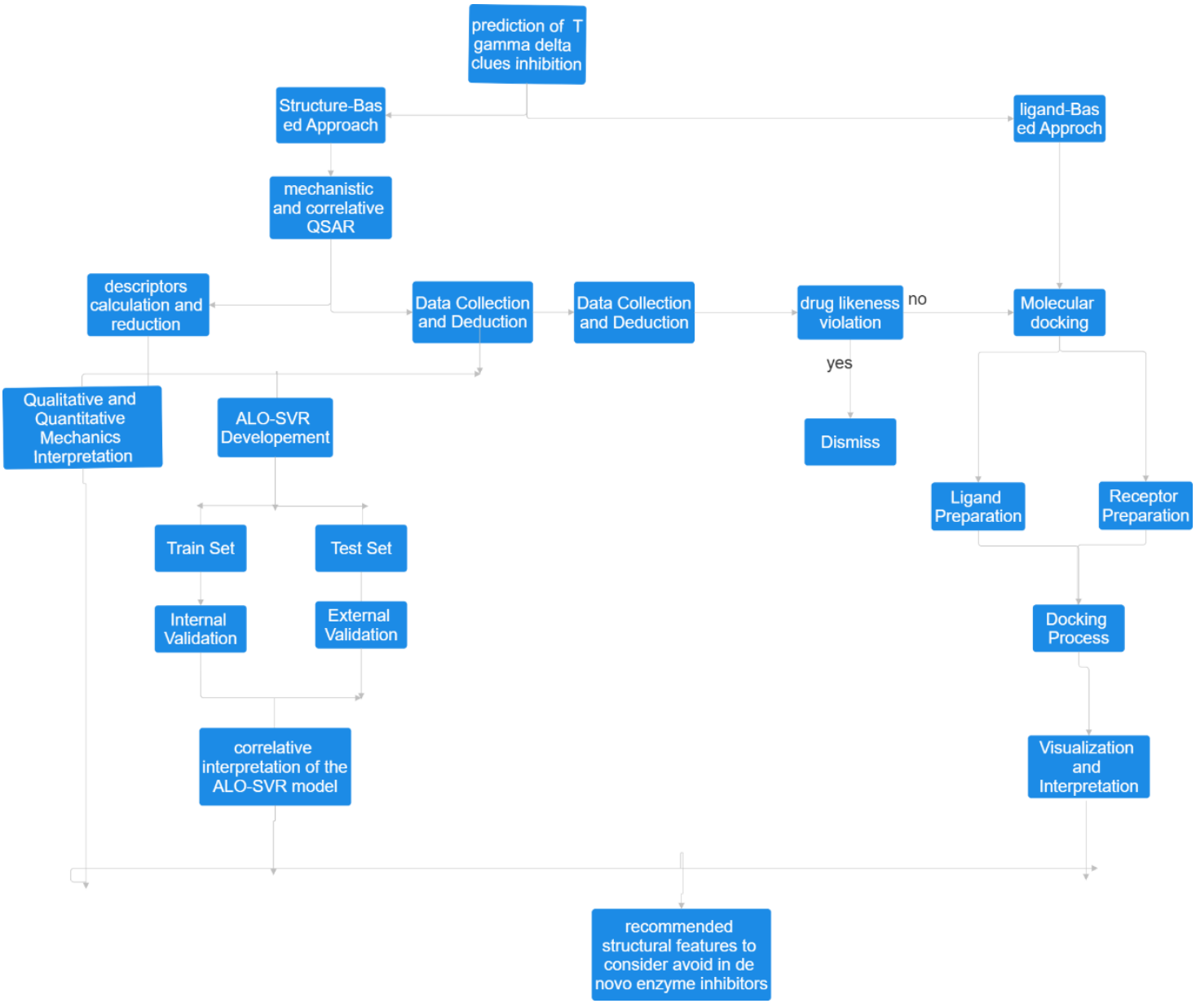
Chapitre IV :RESULTATS ET DISCUSSION

EXPLOITATION DES RESULTATS :

IV.1. Étude du QSAR :

1.1.METHODOLOGIE DE CALCUL :

La méthodologie adoptée pour le calcul de la concentration inhibitrice (PIC_{50}) par la technique ALO-SVR peut être schématisée suivant l'organigramme ci-dessus :



1.2 Collecte de la base de données:

Les données ont été recueillies au moyen du Pubchem et de la littérature [82] sur la base de leur activité prouvée Pour activer les cellulesT γδ Les données ont ensuite été visualisées et prétraitées à l'aide du logiciel Data-Warrior.

Ainsi que les molécules dupliquées ont été éliminées et seules celles ayant une concentration connue à sont disponibles pour leur effet stimulant. À des fins de linéarisation. Les valeurs expérimentales des IC_{50} (nM) ont été converties en échelle logarithmique décimale négative PIC_{50}=−$\log_{10}$ (IC_{50}).

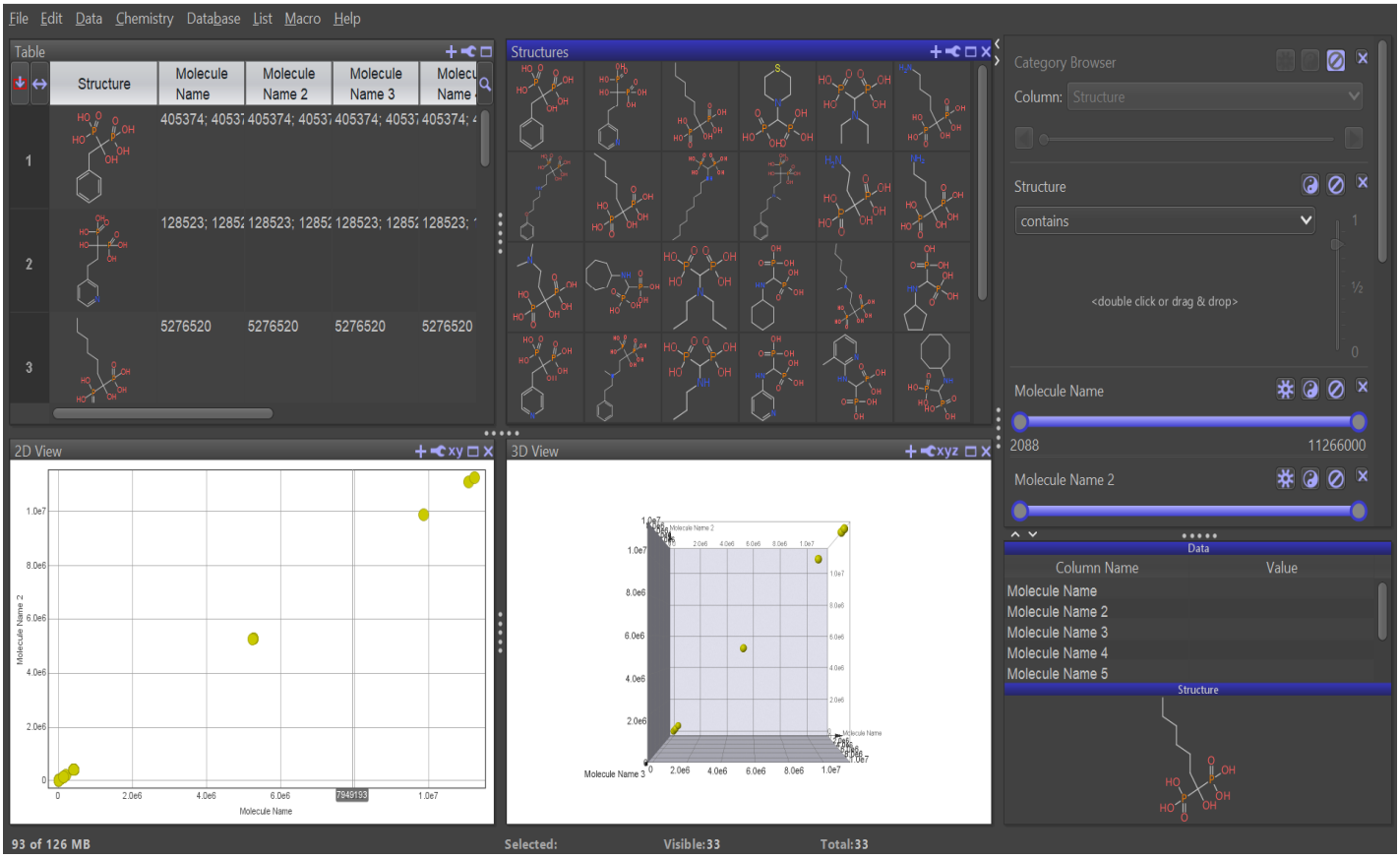


Figure IV 1: capture d'écran Data Warriorr

1.3 Sélection des descripteurs :

Les descripteurs moléculaires désignent la représentation numérique des caractéristiques structurelles des molécules ; ils ont été calculés par le logiciel AlvaDesc pour obtenir les descripteurs. L'algorithme GP-Tree a été utilisé pour sélectionner uniquement les descripteurs représentatifs réduisant le nombre de descripteurs à 5 et formant ainsi une matrice de 33x5.

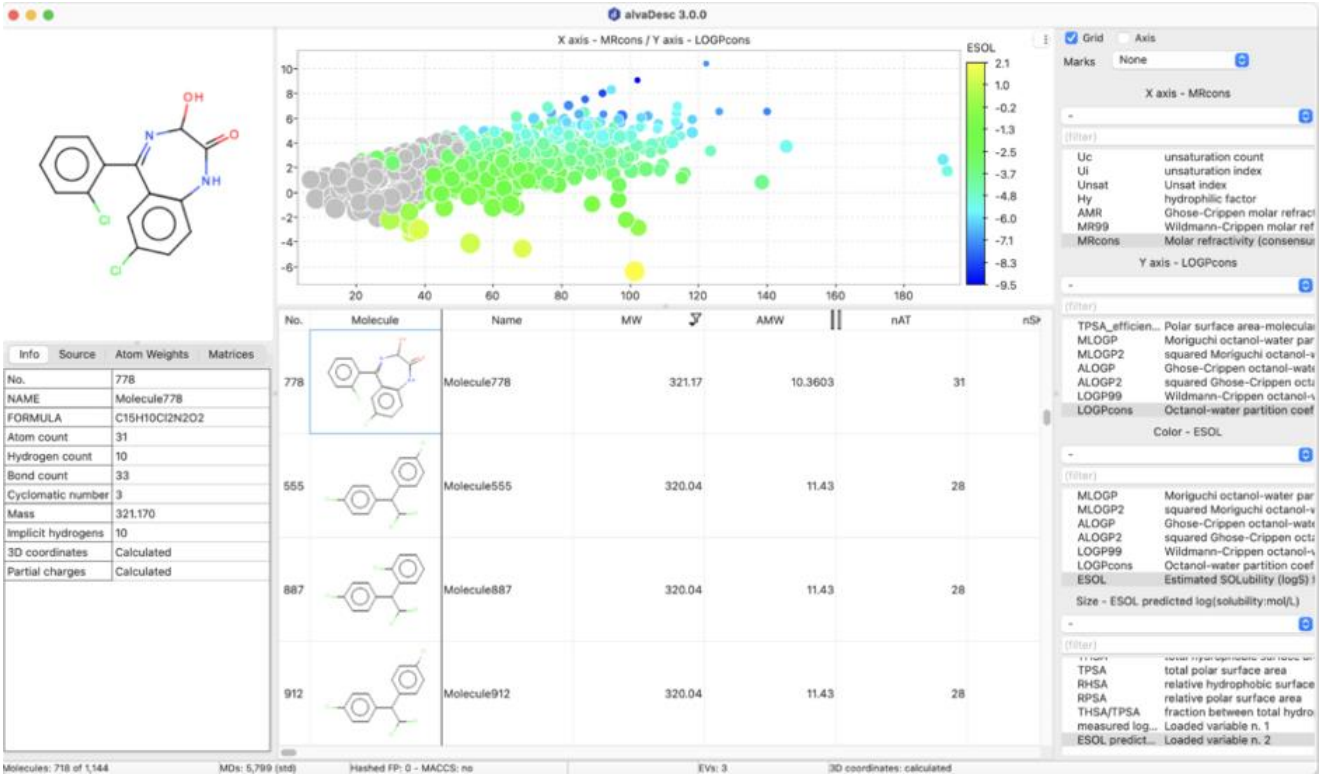
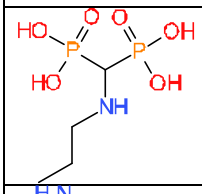
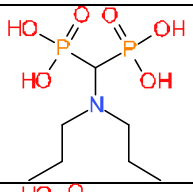
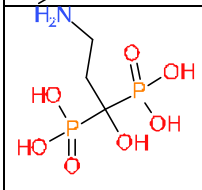
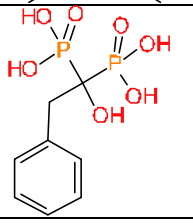


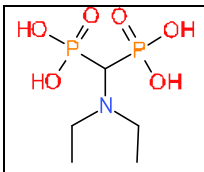
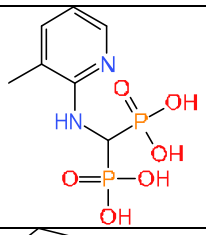
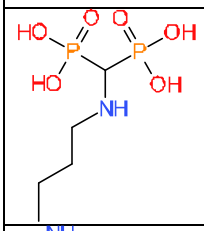
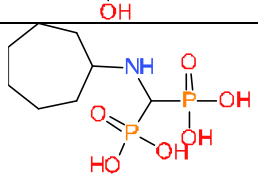
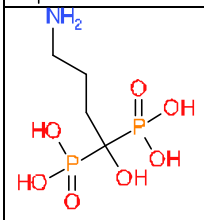
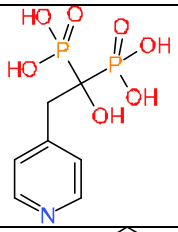
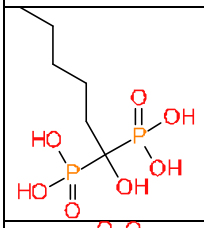
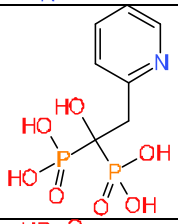
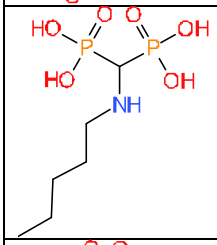
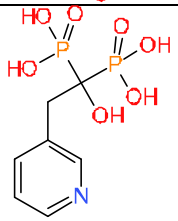
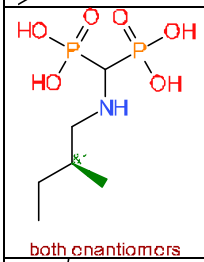
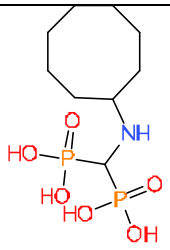
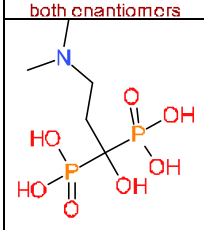
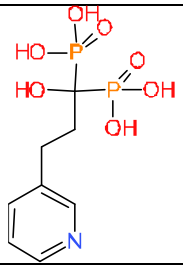
Figure IV. 2: Capture d'écran AlvaDesc

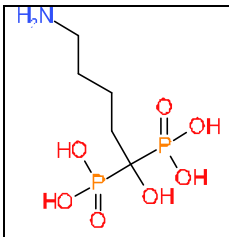
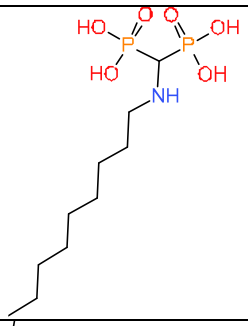
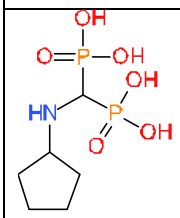
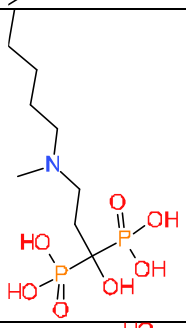
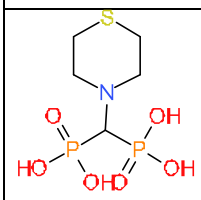
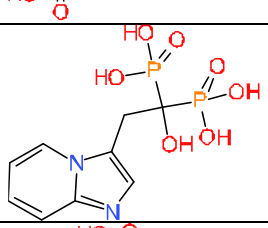
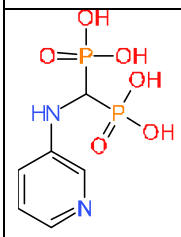
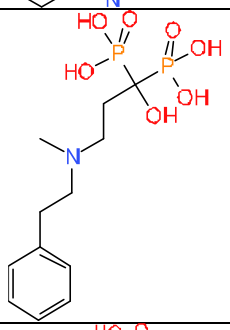
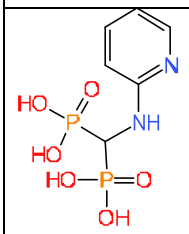
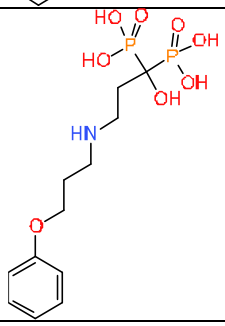
1.4 Division de la base de données et développement du modèle :

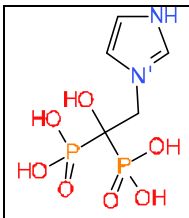
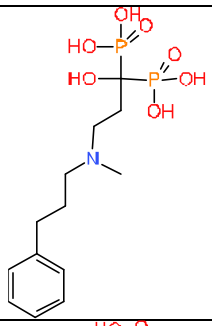
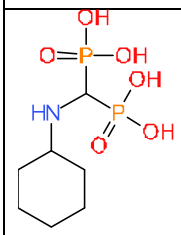
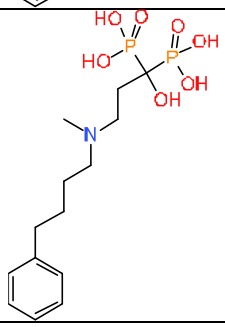
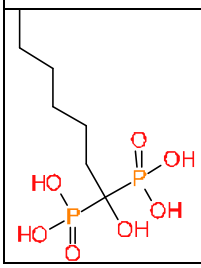
L'ensemble de données a été distribué de manière optimale en utilisant une technique de validation croisée à K -Fold. Dans chaque pli, 80% de l'ensemble de données a été utilisé pour l'apprentissage et 20% pour le test. Les modèles QSAR ont été développés en utilisant l'approche ALO-SVR, réalisés au moyen de Python.

Tableau IV1: la base de données

	PIC_{50}		PIC_{50}
	3,4		4,6
	3,54		4,66

	3,67		4,72
	3,72		4,74
	3,88		4,85
	3,99		4,92
	4,01		4,96
 both enantiomers	4,05		5
	4,18		5

	4,2		5,05
	4,21		5,11
	4,28		5,16
	4,34		5,22
	4,36		5,24

	4,39		5,28
	4,4		5,39
	4,51		

1.5 Support Vector Regression (SVR):

Les machines à vecteurs de support pour la régression (SVR) sont une technique d'apprentissage supervisé utilisée pour prédire des variables continues. Basées sur le principe des machines à vecteurs de support, les SVR visent à établir une fonction de régression qui minimise l'erreur tout en tolérant certaines erreurs inférieures à un seuil donné. Elles sont capables de modéliser des relations linéaires et non linéaires grâce à l'utilisation de noyaux. Les SVR sont largement appliquées dans divers domaines tels que la finance, la biologie, et l'ingénierie, offrant des solutions efficaces pour des tâches de prévision et d'analyse de données.

Validation :

Une fois développé. Le modèle doit être interprété en analysant tous les paramètres statistiques de ce modèle.sa qualité doit être aussi étudiée. Cette qualité est vérifiée par ce que l'on appelle validation.

Validation interne :

La validation interne d'un modèle QSAR a été réalisée en utilisant la validation croisée K-Fold, qui est quantifiée par le coefficient de moyenne d'erreur quadratique (MSE). Ce processus consiste

à diviser le jeu de données en plusieurs sous-ensembles, permettant ainsi d'entraîner le modèle SVR avec des paramètres optimisés (C, epsilon, gamma) à l'aide d'une approche d'optimisation par essaim, comme l'Optimiseur de Lion Ant (ALO). À chaque itération, un sous-ensemble est utilisé pour tester le modèle, tandis que les autres sont utilisés pour l'entraînement. Ce processus est répété pour chaque pli, garantissant ainsi une évaluation robuste du pouvoir prédictif du modèle. La validation interne est devenue une étape essentielle dans la modélisation QSAR, permettant d'assurer la fiabilité des résultats obtenus.

Validation externe

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

Récemment plusieurs études ont montré l'insuffisance du paramètre R^2 pour vérifier le pouvoir prédictif des modèles QSAR.

Par conséquent, d'autres paramètres doivent être vérifiés pour cet objectif. Ces paramètres sont connus sous le nom « critères de validation externe » ou souvent appelés « critères de Trophsa ».

$\mathbf{R^2}$: coefficient de détermination

$$1 - \frac{\sum(Y_{obs} - Y_{calc})^2}{\sum(Y_{obs} - \bar{Y}_{obs})^2} \qquad (1)$$

$\boldsymbol{R_0^2}$: coefficient de détermination sans intercept

$$1 - \frac{\sum(Y_{obs} - k.Y_{calc})^2}{\sum(Y_{obs} - \bar{Y}_{obs})^2}, where: k = \frac{\sum(Y_{obs}.Y_{calc})}{\sum(Y_{calc})^2} \qquad (2)$$

$\mathbf{Q^2}$:Leave one out

$$1 - \frac{\sum(Y_{obs(train)} - Y_{calc(train)})^2}{\sum(Y_{obs(train)} - \bar{Y}_{training})^2} \qquad (3)$$

Scaled Avg.Rm^2 : Rm² moyen normalisé

$$\frac{r_m^2 + r_m'^2}{2}, where: r_m^2 = r^2 \times \left(1 - \sqrt{(r^2 - r_0^2)}\right); r_m'^2 = r^2 \times \left(1 - \sqrt{(r^2 - r_0'^2)}\right) \qquad (4)$$

Scaled DeltaRm^2 : moyenne de rm²

$$|r_m^2 - r_m'^2| \quad (5)$$

RMSE : Erreur quadratique moyenne

$$\sqrt{\frac{\sum(Y_{obs)} - Y_{calc})^2}{n}} \quad (6)$$

SD: Standard deviation

$$\frac{\sum(Y_{obs} - Y_{calc})}{n} \quad (7)$$

MAE: Mean absolute error

$$\frac{\sum(Y_{obs(train)} - Y_{calc(train)})}{n} \quad (8)$$

1.6 Domaine d'applicabilité :

Un modèle QSAR ne peut pas être considéré comme un modèle universel parce qu'il est développé sur un nombre limité de composés qui ne couvrent pas tout l'espace chimique. Par conséquent, l'activité/propriété prédite d'un composé chimiquement dissimilaire au jeu d'apprentissage ne pourra pas être considérée fiable.

Le domaine d'applicabilité (DA) permet de définir la zone dans laquelle un composé pourra être prédit avec confiance. Le DA correspond donc à la région de l'espace chimique incluant les composés de la série d'apprentissage et les composés similaires proches dans ce même espace.

RESULTATS ET DISCUSSION :

1.7 Sélection des descripteurs représentatifs :

La sélection a été effectuée à l'aide du l'algorithme GP-Tree révélant les cinq descripteurs

Les plus pertinents et représentatifs : CATS3D_03_DA, CIC2, T (N. N)**,** B03[C-N], GATS4v.

- **CATS3D_03_DA**

CATS3D signifie Chemically Advanced Template Search en 3D.

Ce descripteur quantifie les interactions donneur–accepteur (Donor-Acceptor) entre deux atomes séparés par une distance de 3 Ångströms dans l'espace tridimensionnel. Il est utilisé pour identifier les possibilités d'interactions électrostatiques dans une molécule

- **CIC2**

CIC est l'abréviation de Complementary Information Content.

CIC2 représente le contenu d'information complémentaire d'ordre 2 basé sur la matrice de connectivité. Il mesure la complexité des liaisons dans une molécule.

- **T(N..N)**

Représente le nombre de chemins topologiques entre deux atomes d'azote (N) dans une molécule. Les points (..) indiquent une certaine longueur de chaîne ou de liaison entre les deux atomes. Cela sert à décrire les relations internes spécifiques entre les atomes similaires.

- **B03[C-N]**

B03 fait référence à des paires d'atomes séparées par 3 liaisons.

[C-N] signifie que ce descripteur compte les paires carbone–azote à cette distance. Cela permet d'étudier la connectivité chimique et la topologie moléculaire.

- **GATS4v**

GATS est une autocorrélation de type Geary, pondérée.

Le chiffre 4 indique un décalage de 4 liaisons chimiques, et v signifie pondération par le volume de van der Waals des atomes.

Cela permet d'évaluer la distribution des volumes atomiques dans la structure moléculaire.

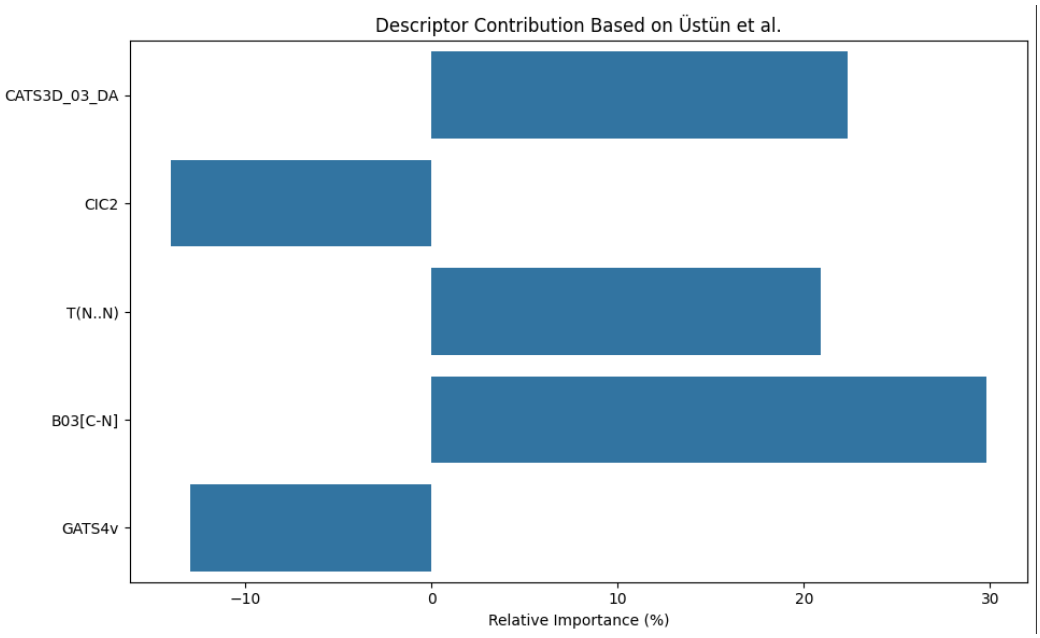


Figure IV.3 : Bar plot de l'importance relative des descripteurs

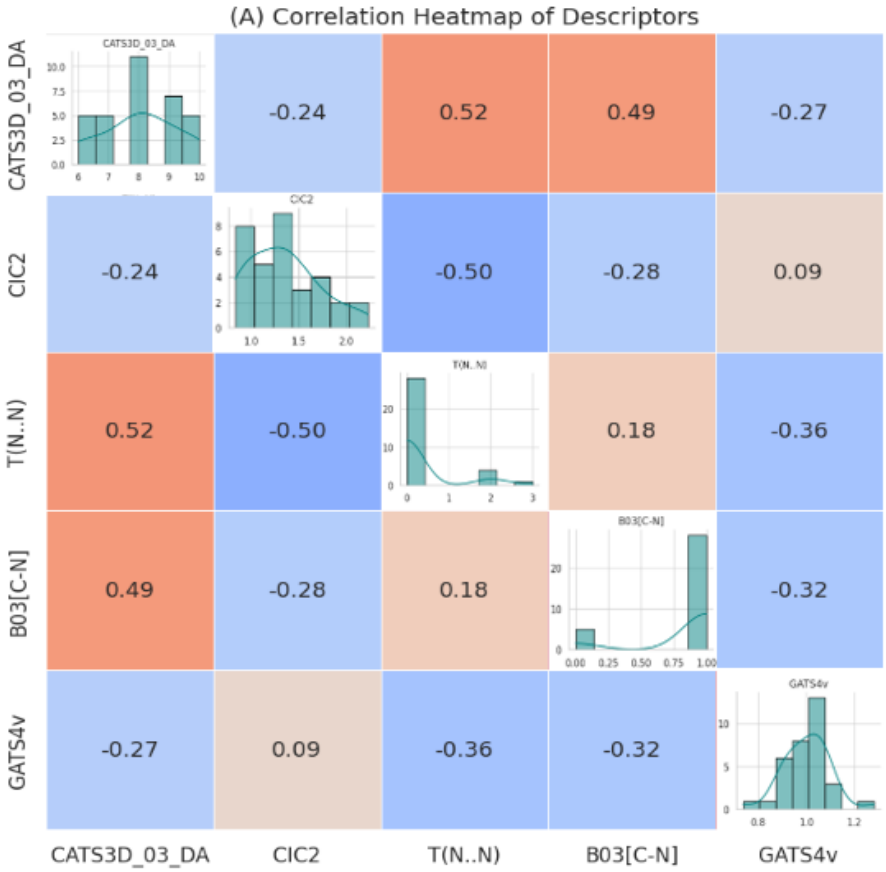


Figure IV.4 : Diagramme de corrélation

Des valeurs positives de R indiquent une corrélation positive, tandis que des valeurs négatives de R indiquent une corrélation négative. Cela peut être utile pour comprendre les relations et les interactions entre les descripteurs chimiques et leurs influences sur les propriétés ou les réponses des composés étudiés.

Tableau IV.2 : les descripteurs pertinents

Nom de descripteurs	Dimension	Noms complets
CATS3D_03_DA	3D	CATS3D Donor-Acceptor BIN 03 (3.000 – 4.000 Å)
CIC2	2D	Complementary Information Content index (neighborhood symmetry of 2-order)
T(N.. N)	2D	sum of topological distances between N.. N
B03[C-N]	2D	Presence/absence of C – N at topological distance 3
GATS4v	2D	Geary autocorrelation of lag 4 weighted by van der Waals volume

Tableau IV.3: descripteurs retenus

No	CATS3D_03_DA	CIC2	T (N... N)	B03[C-N]	GATS4v	pIC50
1	7	1.39188	0	0	0.956383	3.4
2	7	0.903226	0	1	1.105473	3.54
3	6	1.844972	0	0	1.035435	3.67
4	6	1.428571	0	0	1.068373	3.72
5	8	1.638268	0	0	1.283916	3.88
6	6	1.316993	0	1	1.063492	3.99
7	9	1.326467	0	1	1.074092	4.01
8	6	1.637254	0	0	1.041895	4.05
9	8	2.083352	0	1	0.965864	4.18
10	9	1.372161	0	1	1.074476	4.2
11	8	1.083333	0	1	1.146147	4.21
12	6	1.185185	0	1	1.075866	4.28
13	7	1.383659	0	1	0.995805	4.34
14	9	2.028321	0	1	1.025743	4.36
15	8	1.643136	0	1	0.989623	4.39
16	8	1.817399	0	1	0.879599	4.4
17	7	1.512138	0	1	1.051045	4.51
18	8	1.6	0	1	0.735902	4.6
19	7	1.071429	0	1	1.010198	4.66
20	9	1.281847	0	1	1.105256	4.72
21	10	1.105957	0	1	0.91735	4.74
22	10	0.923077	3	1	0.863169	4.85
23	10	0.922582	2	1	0.911797	4.92
24	9	2.228528	0	1	0.996937	4.96
25	9	1.441412	0	1	1.066473	5
26	8	1.198444	0	1	0.949348	5

27	8	1.23468	0	1	0.944516	5.05
28	8	1.320771	0	1	0.953946	5.11
29	8	0.928571	0	1	0.929443	5.16
30	10	0.952111	2	1	0.88944	5.22
31	9	0.955532	0	1	1.011354	5.24
32	8	0.888889	2	1	0.896831	5.28
33	10	0.83609	2	1	1.042722	5.39

L'analyse des descripteurs obtenus indique que :

(CATS3D_03_DA): Les valeurs de ce descripteur, qui varient entre 0,836 et 2,228, mesurent la distance maximale entre deux atomes dans la molécule. Des valeurs élevées signifient que la structure est plus flexible, ce qui peut aider l'inhibiteur à s'adapter à la forme de la cible, comme une enzyme ou un récepteur. Cette flexibilité est cruciale pour permettre des ajustements pendant l'interaction, augmentant ainsi la probabilité d'une liaison efficace.

CIC2 : Ce descripteur a des valeurs allant de 0 à 3 et évalue la distribution des groupes fonctionnels dans la molécule. Des valeurs faibles indiquent une bonne variété dans la disposition de ces groupes, ce qui est essentiel pour favoriser différentes interactions chimiques avec la cible. Une diversité dans les groupes fonctionnels augmente les chances que l'inhibiteur puisse interagir de manière optimale avec la cible, améliorant ainsi son efficacité.

T (N...N) : Les valeurs pour ce descripteur vont de 0 à 1 et représentent la force des liaisons entre les atomes d'azote. Des valeurs élevées montrent que ces liaisons sont solides, ce qui est important car elles stabilisent la structure de l'inhibiteur lorsqu'il se lie à sa cible. Une interaction plus forte entre les atomes d'azote peut également renforcer l'affinité de l'inhibiteur pour la cible, augmentant son efficacité.

B03[C-N] : Ce descripteur, dont les valeurs varient de 0 à 1, évalue les interactions entre les atomes de carbone et d'azote. Des valeurs élevées indiquent que ces interactions sont efficaces, renforçant ainsi la stabilité de l'inhibiteur lorsqu'il se fixe à la cible. Une bonne interaction entre le carbone et l'azote est souvent cruciale pour maintenir la structure de l'inhibiteur et favoriser une liaison solide.

GATS4v : Les valeurs de ce descripteur, qui vont de 0,735 à 1,283, mesurent les interactions électriques au sein de la molécule. Des valeurs élevées indiquent que les interactions électrostatiques sont fortes, ce qui peut considérablement améliorer l'affinité de l'inhibiteur pour sa cible. Ces interactions électriques jouent un rôle clé dans la stabilisation de l'inhibiteur lorsqu'il se lie, ce qui contribue à son efficacité globale.

1.8 Développement des modèles QSAR :

Modèle Ant Lion-SVR:

Dans cette étude, un modèle ALO-SVR a été développé pour prédire le pouvoir stimulant de croissance des cellulesT γδ L’objectif de la combinaison des algorithmes ALO et SVR était de

trouver la meilleure combinaison d'hyper paramètres en testant différentes valeurs. La meilleure combinaison obtenue était C=467.129 , γ= 0.9999, et ε= 0.00998 , avec l'utilisation d'une fonction gaussienne comme fonction Kernel. L'hybridation de SVR et ALO permet d'éviter le sur-ajustement du modèle, en obtenant des coefficients R^2 élevés respectivement (0.9816 0.9997 1) et une valeur de RMSE faible (0.0100). Le tableau 4 présente les paramètres statistiques du modèle ALO-SVR, tandis que la figure 5 illustre la corrélation entre les valeurs prédites du PIC_{50} et les valeurs observées pour l'apprentissage, le test et l'ensemble des données.

Tableau IV.4: Critères statistiques pour le modèle ALO-SVR

Metric	Name	Training	Testing
R^2:	Coefficient of determination	0.9997	0.9993
MAE	Mean Absolute Error	0.0100	0.0101
RMSE	Root Mean Squared Error	0.0100	0.0101
MSE	Mean Squared Error	0.0001	0.0001
Pearson Correlation	Pearson Correlation	0.9999	

Tableau IV.5 : Validation criteria globale

R_0^2	Determination coefficient without intercept	0.9999
$R_0'^2$	Reversed axes determination coefficient without intercept	0.9999
Q^2	Leave one out	0.9996
$\bar{r}_m^2$	Average r_m^2	0.9997
Δr_m^2	Delta r_m^2	0.000086
RMSE	Root mean square error	0.01000
SD	Standard deviation	0.5401

$\boldsymbol{R^2}$ également connu sous le nom de coefficient de détermination multiple, il varie de 0 à 1 et représente la proportion de variance expliquée par le modèle par rapport à la variance totale des données. Le coefficient de détermination R^2 est souvent utilisé pour évaluer la capacité d'un modèle à prédire l'activité biologique d'une molécule en fonction de ses descripteurs moléculaires. Un R^2 élevé indique une bonne corrélation entre les descripteurs moléculaires et l'activité biologique, ce qui suggère que le modèle est capable de capturer et de prédire de manière précise les relations quantitatives entre les structures chimiques et les activités biologiques.

$\mathbf{R0^2}$Test : est généralement calculé en utilisant un ensemble de données de test indépendant qui n'a pas été utilisé pour construire le modèle. Il représente la proportion de la variance totale des données de test expliquée par le modèle. Un $R0^2$ test élevé indique une bonne capacité de prédiction du modèle, tandis qu'un $R0^2$ test faible suggère une moins bonne performance prédictive.

$\boldsymbol{RO^2}$´test : Ce sont les coefficients de corrélation au carré (squared correlation coefficients) mesurent la similarité ou l'ajustement entre les valeurs observées et les valeurs prédites dans un modèle QSAR. Dans ce cas, il s'agit des coefficients de corrélation au carré entre les valeurs observées et les valeurs prédites obtenues en utilisant la méthode "leave-one-out" (lorsqu'un composé est exclu à la fois) pour évaluer la performance du modèle QSAR, avec et sans intercept (une constante dans l'équation du modèle).

Q² (Coefficient de détermination croisée) : est une métrique utilisée pour évaluer la capacité d'un modèle à prédire des données nouvelles, en particulier dans le cadre de la validation croisée. Il mesure la proportion de la variance dans les données qui est prédite par le modèle.

Scaled Avg.Rm^2: est une métrique qui mesure l'erreur quadratique moyenne normalisée d'un modèle QSAR. Une valeur plus faible de la métrique indique une meilleure précision de prédiction du modèle.

Scaled DeltaRm^2: est une métrique utilisée pour évaluer la performance d'un modèle QSAR Il est également connu sous le nom de "Scaled Delta R-squared" ou "Scaled ΔRm^2". Ce critère mesure l'écart entre la variance totale des valeurs expérimentales et la variance totale des valeurs prédites par le modèle QSAR. Un Scaled DeltaRm^2 plus faible indique une meilleure adéquation du modèle aux données expérimentales, c'est-à-dire une meilleure capacité à reproduire les variations des activités biologiques

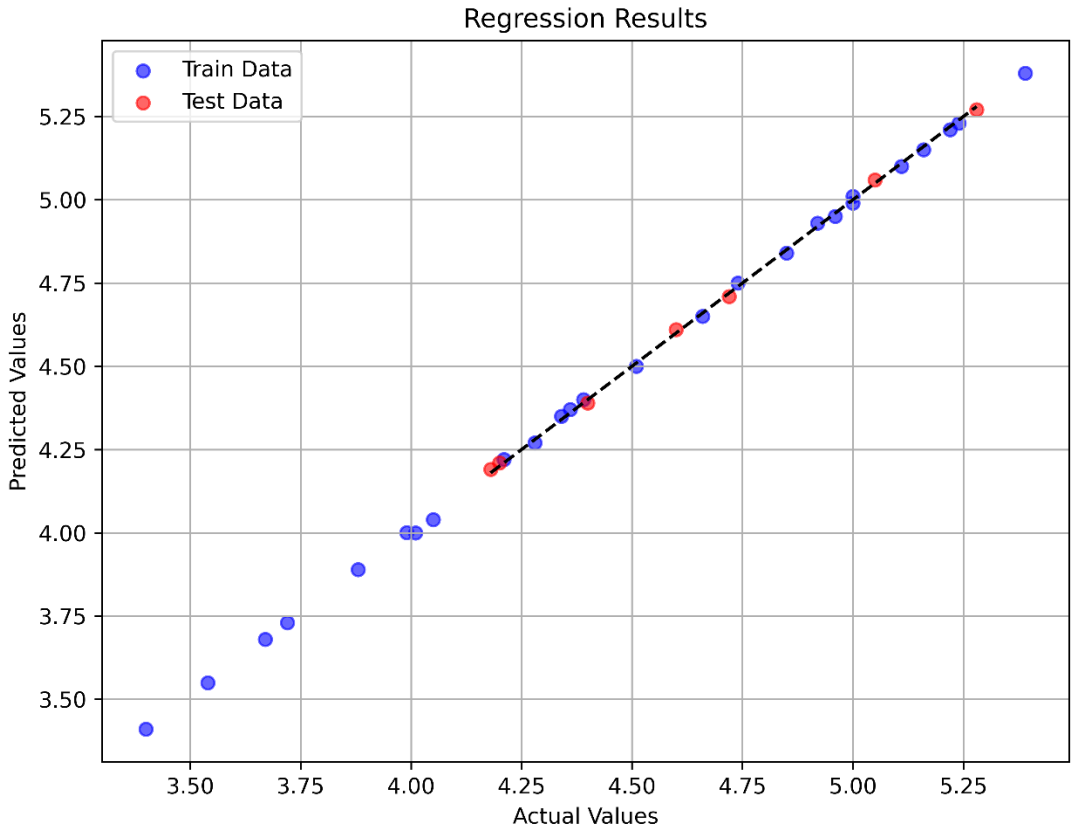


Figure IV.5 : Courbes de régression ALO-SVR

1.9 VALIDATION EXTERNE :

Le calcul des critères de validation externe sur le jeu de test de modèle montre une bonne performance pour le ALO-SVR (R^2 : 0.999) et des valeurs faibles des RMSEP (0.100). Les critères de validation externe permettent d'évaluer la précision et la capacité des modèles à prédire une nouvelle base qui n'appartient pas à la base initiale. Les résultats obtenus ont montré que le modèle ALO-SVR a une meilleure performance en termes de mesure de validation externe telles que les valeurs des RMSE indiquant que le modèle ALO-SVR est plus précis. En conséquence le modèle ALO-SVR a montré une bonne signification statistique et une bonne robustesse.

1.10 DOMAINE D'APPLICABILITE :

Algorithm Ant Lion-SVR :

Le diagramme de Williams (figure présentée) illustre la relation entre les résidus standardisés et la portance pour le modèle ALO-SVR. L'analyse révèle que la plupart des points sont regroupés dans la fourchette (3,-3), ce qui indique une bonne adéquation entre les valeurs prédites et observées.

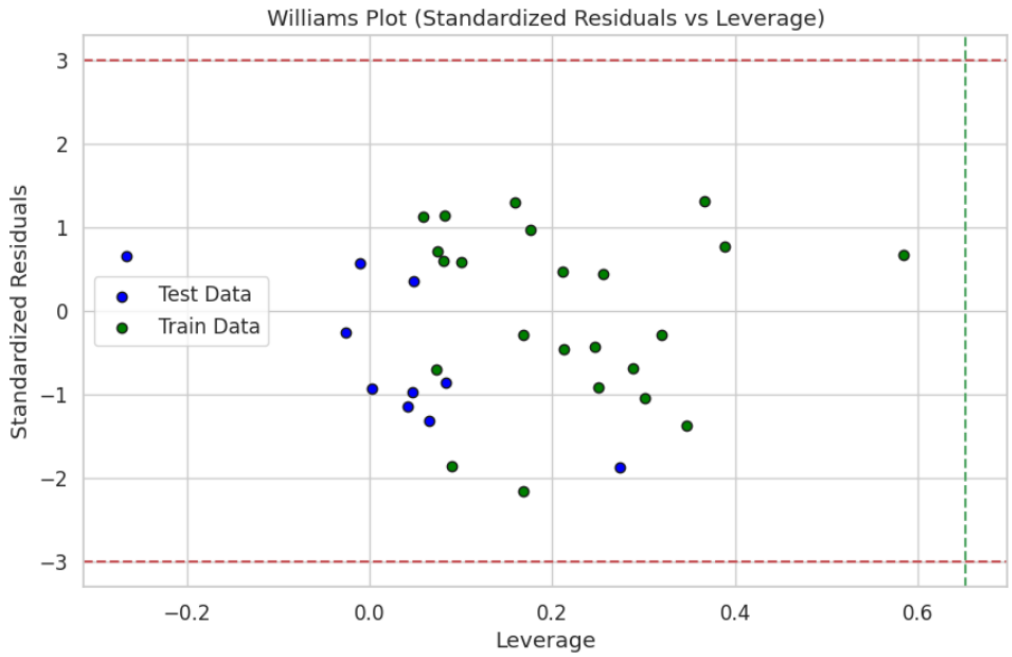


Figure IV.6 : Domaine applicability

IV.2 Étude du Docking :

2.1 ADMET

Afin d'évaluer la compatibilité des composés avec les propriétés médicamenteuses, la règle des cinq de Lipinski a été appliquée sur 65 composés. Les résultats du filtrage ont montré que l'ensemble des composés possèdent des propriétés physico-chimiques conformes aux critères requis pour les médicaments, ce qui indique leur potentiel en tant que candidats médicamenteux efficaces. De plus, les résultats de l'analyse des propriétés de toxicité et des paramètres ADMET, réalisés à l'aide de l'outil ADMET Lab 3.0, ont révélé que ces composés ne présentent aucun signe de toxicité ni d'effets secondaires potentiels, ce qui renforce leur innocuité et leur efficacité potentielle dans les applications pharmaceutiques.

2.2 Préparation des structures :

La structure cristallographique du récepteur A18(Cyanidin CID :128861) en complexe avec le ligand IL-17RA (PDB :5N9B) a été obtenue depuis la banque de données PDB

Préparation du ligand et de la protéine et étude du Docking :

L'expérience de docking moléculaire a été réalisée à l'aide du logiciel Schrödinger 2024, via l'interface graphique Maestro 13.9. Dans un premier temps, la protéine IL-17RA a été préparée à l'aide de l'outil Protein Preparation, en supprimant les molécules d'eau, les ions et les résidus non liés (ligands, cofacteurs), tout en corrigeant les états de protonation ainsi que les états d'oxydation et de valence des atomes. Par la suite, le site actif a été défini en fonction de la position de liaison du ligand A18 (CID:128861), et une grille de récepteur (Receptor Grid Generation) a été générée autour de ce site.

Le docking a ensuite été effectué à l'aide de l'outil Glide, avec une bibliothèque de 65 composés préalablement préparés via LigPrep, garantissant des structures 3D appropriées avec des états tautomériques et de protonation adaptés. Les résultats du docking ont été analysés sur la base des valeurs d'énergie de liaison et des interactions moléculaires au sein du site actif, afin d'identifier les composés présentant le plus grand potentiel inhibiteur envers la protéine**.**

Interactions entre le récepteur et le ligand :

Les inhibiteurs optimaux du récepteur A18 (CID:128861) ont été sélectionnés en fonction du score de Docking le plus faible, indiquant une forte affinité de liaison. Parmi eux, le composé (65) a montré une énergie de liaison libre de −5.967 kcal/mol, traduisant une interaction stable au sein du site actif de la protéine. Une analyse détaillée de la conformation liée a révélé la formation de quatre liaisons hydrogène conventionnelles entre le ligand et les résidus HIS148, SER147, THR145 et ARG143. De plus, les résidus aromatiques PHE144 et PHE146 établissent vraisemblablement des interactions π–π empilées avec les cycles aromatiques du ligand, contribuant ainsi à la stabilisation du complexe ligand-récepteur. Ces interactions multiples suggèrent que ce composé possède un bon potentiel en tant qu'inhibiteur efficace du récepteur A18.

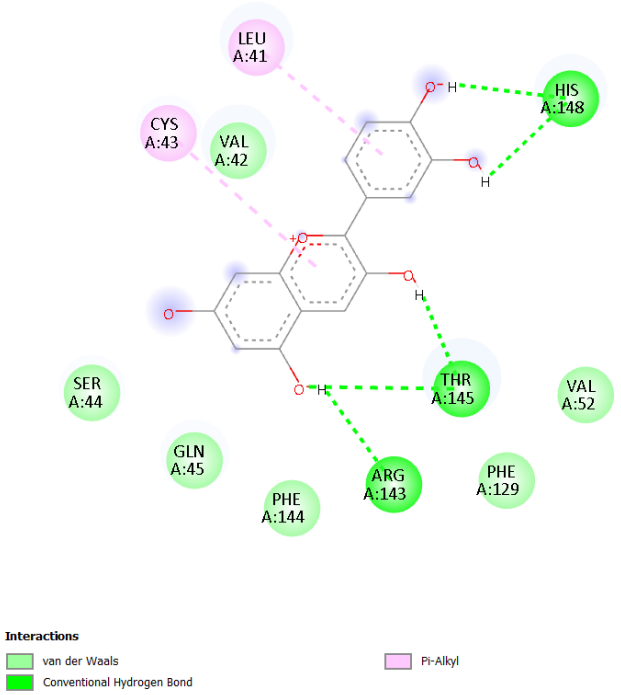


Figure IV.7 : Diagramme des interactions 2D de (65) avec A18 (CID:128861)

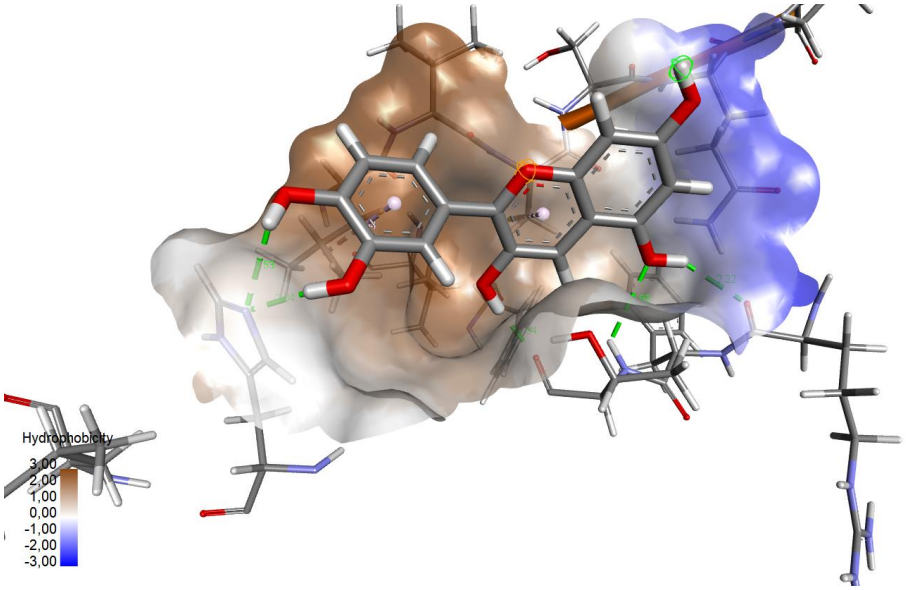


Figure IV.8 : Diagramme des interactions 3D de (65) avec A18 (CID:128861)

Le molécule (64) présente une affinité de liaison modérée envers le domaine A18 (CID:128861), avec une énergie de liaison libre mesurée à −5.536 kcal/mol, tel qu'il est illustré dans la figure ci-dessus. Cette interaction est caractérisée par la présence de plusieurs liaisons hydrogène conventionnelles entre le récepteur et le ligand. Plus spécifiquement, les résidus THR145, GLN45, VAL42, et SER147 sont impliqués dans ces liaisons, stabilisant efficacement la molécule dans le site actif. De plus, le résidu PHE146 interagit par proximité avec le groupe phosphate du ligand, suggérant une contribution possible par des interactions électrostatiques ou π-anion. Enfin, la présence des résidus LEU41 et PHE144 à proximité du ligand pourrait renforcer l'interaction par des contacts hydrophobes ou empilements π–π, contribuant ainsi à la stabilisation globale du complexe.

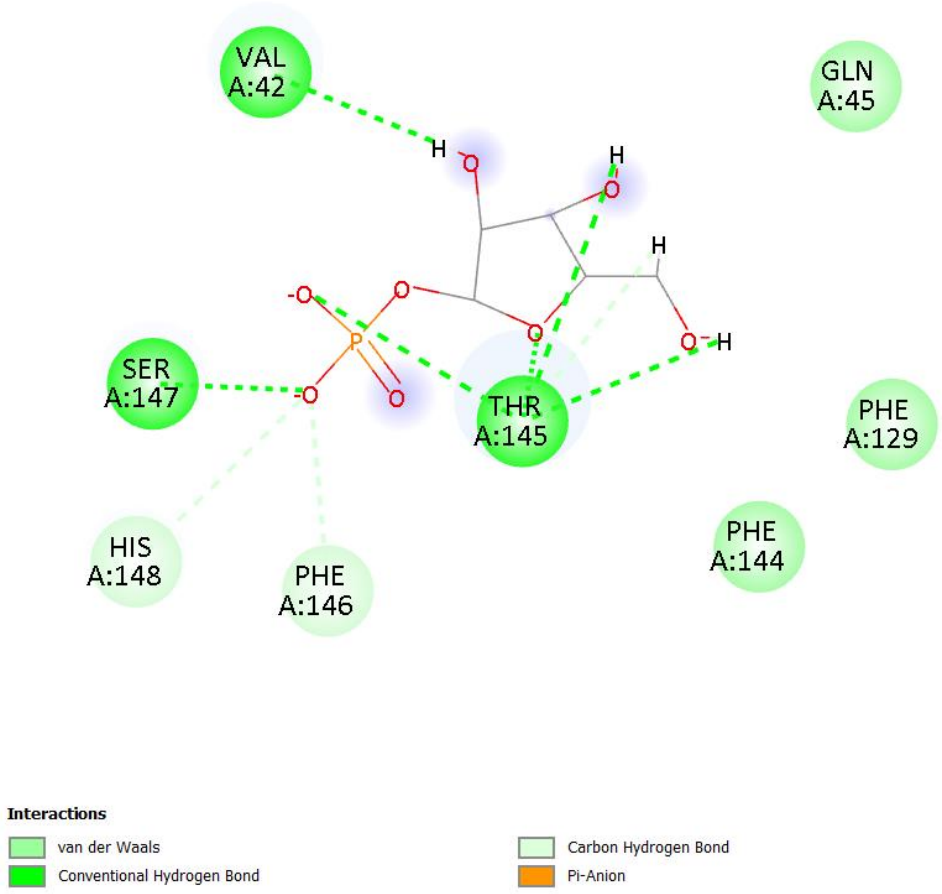


Figure IV.9 : Diagramme des interactions 2D de(64)avec A18 (CID:128861)

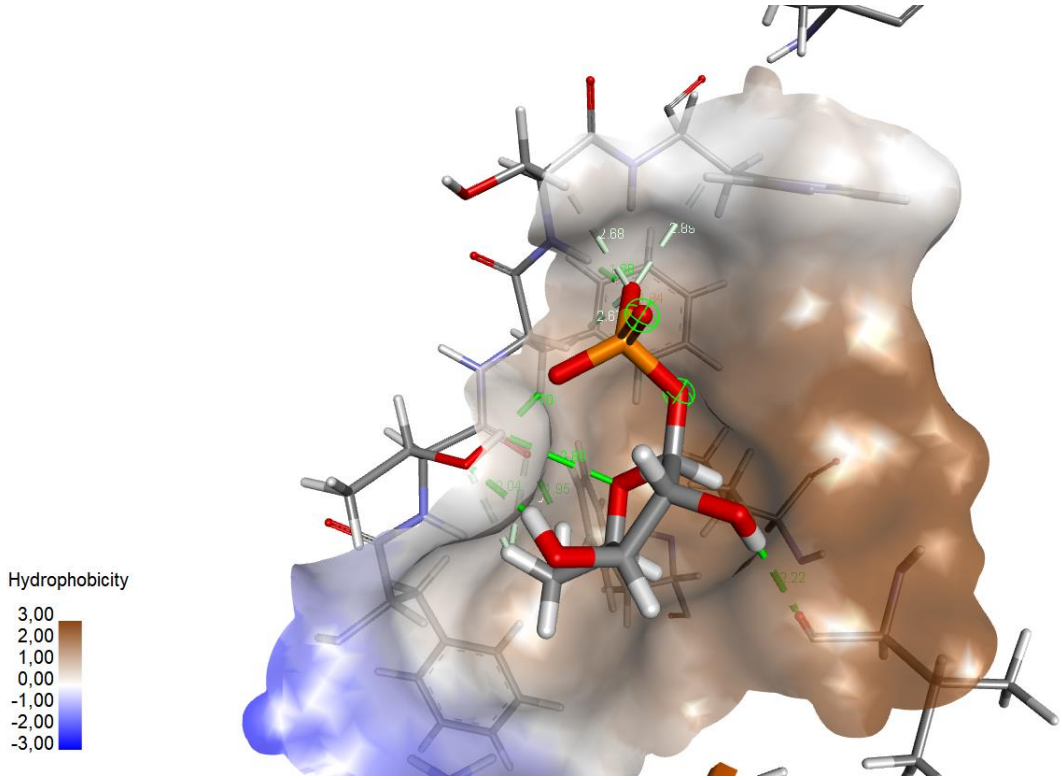


Figure IV.10 : Diagramme des interactions 3D de (64) avec A18 (CID:128861)

La Molécule (60) montre une affinité de liaison modérée avec le domaine A18, avec une énergie libre de liaison de −5.419 kcal/mol. Cette interaction est principalement stabilisée par plusieurs liaisons hydrogène conventionnelles impliquant les résidus THR145 et SER147, assurant une fixation efficace du ligand dans le site actif. De plus, les résidus CYS43, GLN45 et HIS148 participent à la stabilisation via des interactions de van der Waals et des liaisons hydrogène supplémentaires. Le groupe phosphate du ligand interagit également avec le résidu PHE146 à travers une interaction π-anion, suggérant une contribution électrostatique importante. Enfin, la disposition spatiale des résidus tels que PHE129 et VAL52 peut renforcer davantage la stabilité du complexe via des interactions de type van der Waals, faisant de ce complexe un candidat potentiellement stable au niveau structurel.

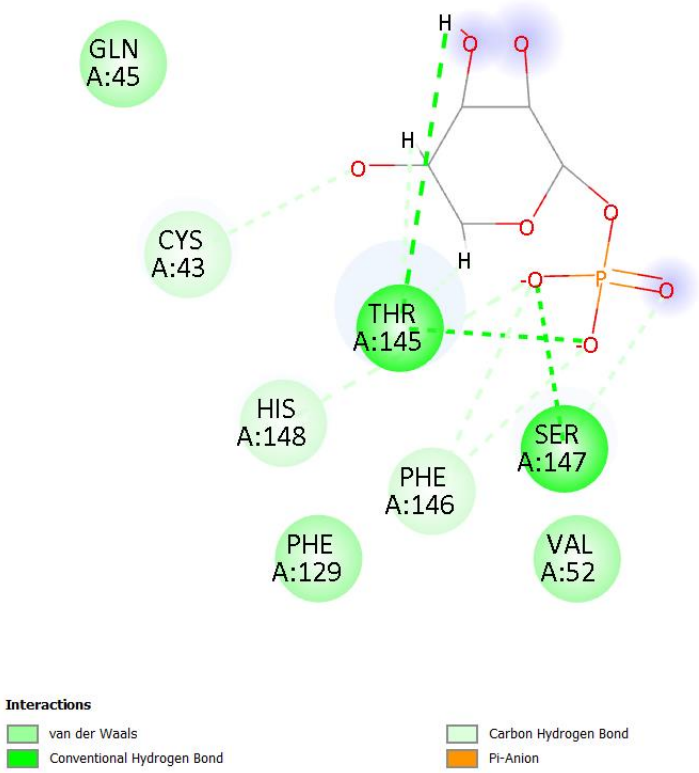


Figure IV.11 : Diagramme des interactions 2D de (60) avec A18 (CID:128861)

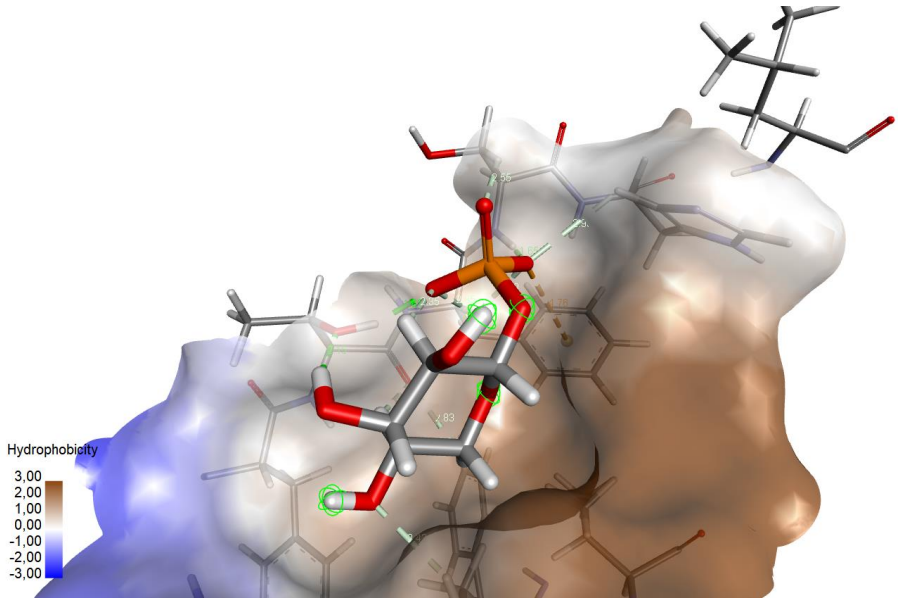


Figure IV.12 : Diagramme des interactions 3D de (60) avec A18 (CID:128861)

La molécule (58) présente une affinité de liaison modérée avec le domaine A18, avec une énergie libre de liaison de −5.005 kcal/mol, comme illustré dans la figure ci-dessus. Elle établit plusieurs liaisons hydrogène conventionnelles avec les résidus THR145 et SER147, assurant un ancrage stable dans le site actif. Une interaction de type π-anion est également observée avec le résidu PHE146, indiquant une contribution électrostatique significative. Par ailleurs, LEU41 interagit via une liaison alkyle, tandis que HIS148 forme une liaison hydrogène de type carbone-hydrogène, renforçant la stabilité globale du complexe. Les résidus VAL52 et PHE129, bien que plus distants, peuvent également contribuer à la stabilisation via des interactions van der Waals.

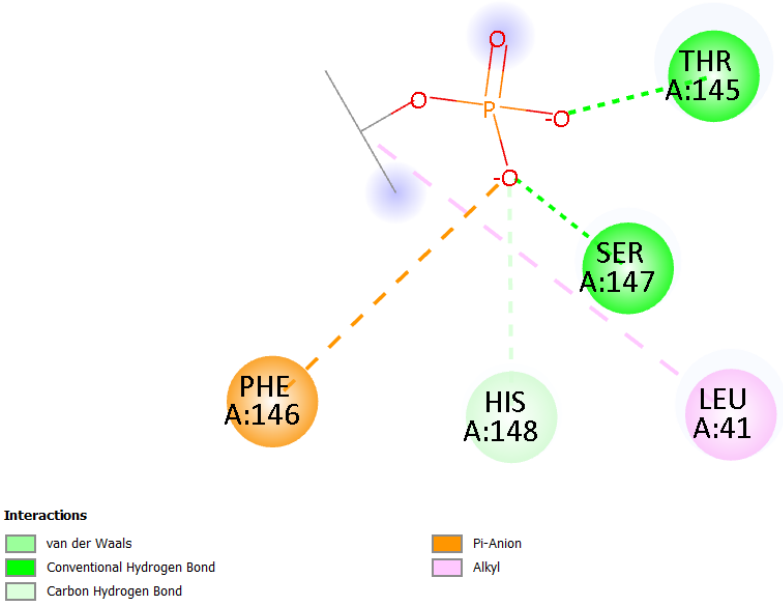


Figure IV.13 : Diagramme des interactions 2D de (58) avec A18 (CID:128861)

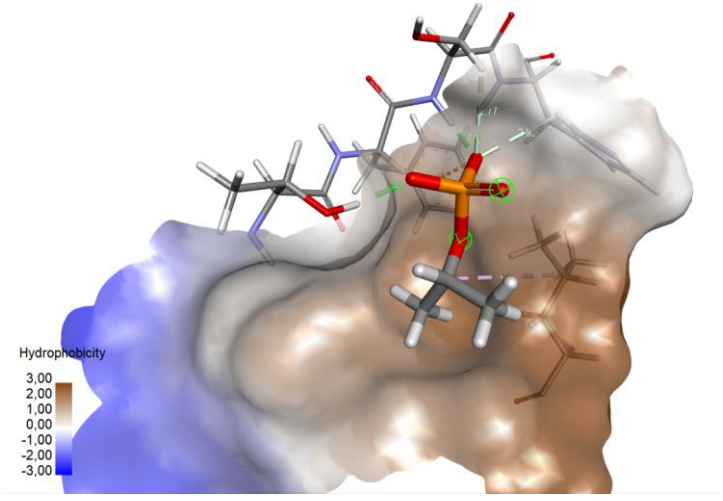


Figure IV.14: Diagramme des interactions 3D de (58) avec A18 (CID:128861)

La molécule (29) présente une affinité de liaison modérée avec le domaine A18 (CID:128861), avec une énergie libre de liaison de −4.957 kcal/mol, comme illustré dans la figure ci-dessus. Cette stabilisation est attribuée à la présence de plusieurs liaisons hydrogène conventionnelles impliquant les résidus VAL42, THR145, SER147 et HIS148, ce qui renforce l'ancrage du ligand dans le site actif. De plus, le résidu PHE146 participe à l'interaction via une liaison hydrogène de type carbone-hydrogène, tandis qu'une interaction de type π-anion entre le groupe phosphate et un résidu aromatique suggère un rôle potentiel des forces électrostatiques dans la stabilisation globale du complexe.

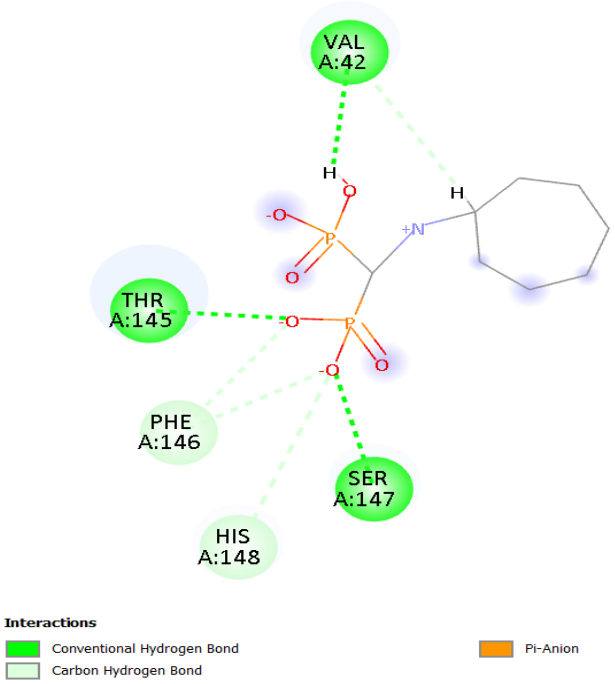


Figure IV.15 : Diagramme des interactions 2D de (29) avec A18 (CID:128861)

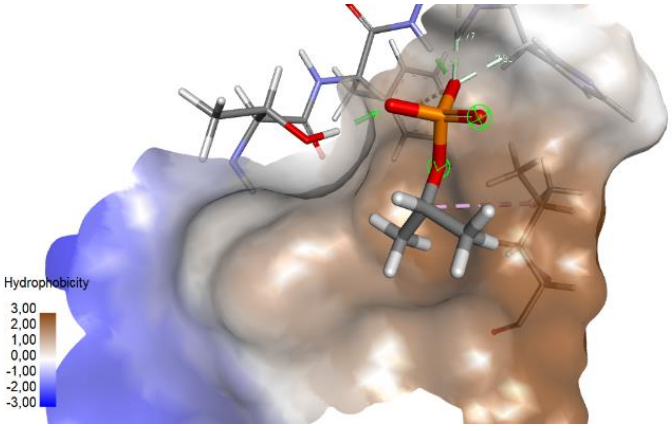


Figure IV.16 : Diagramme des interactions 3D de (29) avec A18 (CID:128861)

IV.3. Étude du QSAR inverse

3.1 Objectifs

Ce travail vise à :

- Développer un modèle capable de prédire l'activité biologique (pIC50) d'un ensemble de composés chimiques.
- Améliorer les propriétés chimiques des composés en utilisant l'algorithme ALO pour atteindre des valeurs cibles de pIC50.
- Identifier les composés les plus proches dans une base de données de composés chimiques, facilitant ainsi le processus de découverte de nouveaux médicaments.

3.2 Modèle Ant Lion-SVR

Nous avons utilisé plusieurs bibliothèques Python, telles que joblib pour le chargement des modèles, numpy et pandas pour la manipulation des données, sklearn pour les algorithmes d'apprentissage automatique, et rdkit pour la manipulation et la visualisation des structures chimiques. Ensuite, nous avons chargé un modèle préalablement entraîné"final_svr_model" ainsi que les données des composés chimiques"PIC50" à partir d'un fichier Excel, ce qui nous permet d'effectuer des prédictions sur l'activité biologique.

Tableau IV.6: descriptor_columns + 'SMILES'

CATS3D_03_DA	CIC2	T(N..N)	B03[C-N]	GATS4v	pIC50	SMILES
7	1.39188	0	0	0.956383	3.4	P(=O)(O)(O)C(P(=O)(O)O)NCCC
7	0.90326	0	1	1.105473	3.54	P(=O)(O)(O)C(P(=O)(O)O)(O)CCN
6	1.84972	0	0	1.035435	3.67	P(=O)(O)(O)C(P(=O)(O)O)N(CC)CC
6	1.42871	0	0	1.068373	3.72	P(=O)(O)(O)C(P(=O)(O)O)NCCCC
8	1.63868	0	0	1.283916	3.88	P(=O)(O)(O)C(P(=O)(O)O)(O)CCCN

6	1.316993	0	1	1.063492	3.99	P(=O)(O)(O)C(P(=O)(O)O)(O)CCCCC
9	1.326467	0	1	1.074092	4.01	P(=O)(O)(O)C(P(=O)(O)O)NCCCCC
6	1.637254	0	0	1.041895	4.05	P(=O)(O)(O)C(P(=O)(O)O)NCC(CC)C
8	2.083352	0	1	0.965864	4.18	P(=O)(O)(O)C(P(=O)(O)O)(O)CCN(C)C
9	1.372161	0	1	1.074476	4.2	P(=O)(O)(O)C(P(=O)(O)O)(O)CCCCN
8	1.083333	0	1	1.146147	4.21	P(=O)(O)(O)C(P(=O)(O)O)NC1(CCCC1)
6	1.185185	0	1	1.075866	4.28	S1(CCN(C(P(=O)(O)O)P(=O)(O)O)CC1)
7	1.383659	0	1	0.995805	4.34	P(=O)(O)(O)C(P(=O)(O)O)NC1(=CN=CC=C1)
9	2.028321	0	1	1.025743	4.36	P(=O)(O)(O)C(P(=O)(O)O)NC1(=NC=CC=C1)
8	1.643136	0	1	0.989623	4.39	P(=O)(O)(O)C(P(=O)(O)O)(O)C[N+]1(=CNC=C1)
8	1.817399	0	1	0.879599	4.4	P(=O)(O)(O)C(P(=O)(O)O)NC1(CCCCC1)
7	1.512138	0	1	1.051045	4.51	P(=O)(O)(O)C(P(=O)(O)O)(O)CCCCCC
8	1.6	0	1	0.735902	4.6	P(=O)(O)(O)C(P(=O)(O)O)N(CCC)CCC
7	1.071429	0	1	1.010198	4.66	P(=O)(O)(O)C(P(=O)(O)O)(O)CC1(=CC=CC=C1)
9	1.281847	0	1	1.105256	4.72	P(=O)(O)(O)C(P(=O)(O)O)NC1(=NC=CC=C1C)
10	1.105957	0	1	0.91735	4.74	P(=O)(O)(O)C(P(=O)(O)O)NC1(CCCCCC1)

10	0.923077	3	1	0.863169	4.85	P(=O)(O)(O)C(P(=O)(O)O)(O)CC1(=CC=NC=C1)
10	0.922582	2	1	0.911797	4.92	P(=O)(O)(O)C(P(=O)(O)O)(O)CC1(=NC=CC=C1)
9	2.228528	0	1	0.996937	4.96	P(=O)(O)(O)C(P(=O)(O)O)(O)CC1(=CN=CC=C1)
9	1.441412	0	1	1.066473	5	P(=O)(O)(O)C(P(=O)(O)O)NC1(CCCCCCC1)
8	1.198444	0	1	0.949348	5	P(=O)(O)(O)C(P(=O)(O)O)(O)CCC1(=CN=CC=C1)
8	1.23468	0	1	0.944516	5.05	P(=O)(O)(O)C(P(=O)(O)O)NCCCCCCCCC
8	1.320771	0	1	0.953946	5.11	P(=O)(O)(O)C(P(=O)(O)O)(O)CCN(CCCCC)C
8	0.928571	0	1	0.929443	5.16	P(=O)(O)(O)C(P(=O)(O)O)(O)CC=2(N1(C(C=CC=C1)=NC=2))
10	0.952111	2	1	0.88944	5.22	P(=O)(O)(O)C(P(=O)(O)O)(O)CCN(CCC1(=CC=CC=C1))C
9	0.955532	0	1	1.011354	5.24	P(=O)(O)(O)C(P(=O)(O)O)(O)CCNCCCOC1(=CC=CC=C1)
8	0.888889	2	1	0.896831	5.28	P(=O)(O)(O)C(P(=O)(O)O)(O)CCN(CCCC1(=CC=CC=C1))C
10	0.83609	2	1	1.042722	5.39	P(=O)(O)(O)C(P(=O)(O)O)(O)CCN(CCCCC1(=CC=CC=C1))C

Nous avons défini des limites minimales et maximales pour chaque propriété chimique afin de garantir que les composés générés par notre modèle respectent les contraintes chimiques connues. Parallèlement, nous avons conçu une fonction objective visant à minimiser l'écart entre les valeurs prédites par le modèle et la valeur cible de pIC50, orientant ainsi l'optimisation vers des solutions conformes aux critères d'activité biologique souhaités. Pour ce faire, nous avons mis en place l'algorithme ALO, qui s'inspire du comportement des lions dans la nature pour explorer l'espace de recherche et identifier des solutions optimales.

Résultats:

Descripteur optimisé	CATS3D_03_DA	CIC2	T(N..N)	B03[C-N]	GATS4v
	6.6086	1.997	0.1303	0.7322	0.790

- **pIC50 prédit** : **4,18**
- **Molécule correspondante (SMILES)** : P(=O)(O)(O)C(P(=O)(O)O)(O)CCN(C)C
- **Similarité cosinus** : 99%

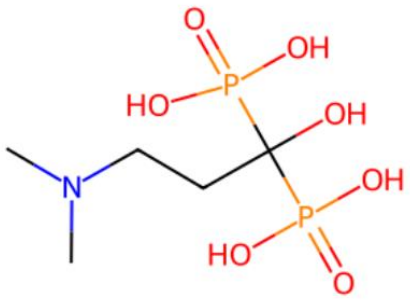


Figure IV.17 : Olpadronic Acid

Les résultats incluent un descripteur optimisé, une valeur de pIC50 prédite, la structure SMILES de la molécule correspondante, et une mesure de similarité cosinus, qui indique à quel point la molécule générée est similaire à d'autres composés connus.

3.3 Extraction de SMILES

Nous avons utilisé la bibliothèque chembl_webresource_client pour accéder à la base de données ChEMBL, qui contient des informations sur les composés chimiques et leurs activités biologiques. En définissant un critère spécifique, nous avons extrait un certain nombre de molécules et ces données ont été converties en un DataFrame pour faciliter leur manipulation.

Nous avons enregistré un total de 9950 molécules, avec des leurs représentations SMILES, qui sont des notations textuelles des structures chimiques.

	A	B
1	ChEMBL_I	SMILES
2	CHEMBL6:	Cc1cc(-n2ncc(=O)[nH]c2=O)ccc1C(=O)c1ccccc1Cl
3	CHEMBL6:	Cc1cc(-n2ncc(=O)[nH]c2=O)ccc1C(=O)c1ccc(C#N)cc1
4	CHEMBL2(	Cc1cc(-n2ncc(=O)[nH]c2=O)cc(C)c1C(O)c1ccc(Cl)cc1
5	CHEMBL6:	Cc1ccc(C(=O)c2ccc(-n3ncc(=O)[nH]c3=O)cc2)cc1
6	CHEMBL2(	Cc1cc(-n2ncc(=O)[nH]c2=O)ccc1C(=O)c1ccc(Cl)cc1
7	CHEMBL6:	Cc1cc(-n2ncc(=O)[nH]c2=O)ccc1C(=O)c1ccccc1
8	CHEMBL6:	Cc1cc(Br)ccc1C(=O)c1ccc(-n2ncc(=O)[nH]c2=O)cc1Cl
9	CHEMBL2(	O=C(c1ccc(Cl)cc1Cl)c1ccc(-n2ncc(=O)[nH]c2=O)cc1Cl
10	CHEMBL2(	CS(=O)(=O)c1ccc(C(=O)c2ccc(-n3ncc(=O)[nH]c3=O)cc2Cl)cc1
11	CHEMBL2(	c1cc2cc(c1)-c1cccc(c1)C[n+]1ccc(c3ccccc31)NCCCCCCCCCCNc1cc[n+](c3ccccc13)C2
12	CHEMBL6:	CSc1ccc(C(=O)c2ccc(-n3ncc(=O)[nH]c3=O)cc2Cl)cc1
13	CHEMBL6:	O=C(c1ccc(Cl)cc1)c1ccc(-n2ncc(=O)[nH]c2=O)cc1Cl
14	CHEMBL6:	O=C1O/C(=C/Br)CCC1c1cccc2ccccc12
15	CHEMBL6:	O=C1O/C(=C/I)CCC1c1cccc2ccccc12
16	CHEMBL2(	C#C/C=C1\CCC(c2cccc3ccccc23)C(=O)O1
17	CHEMBL6:	O=C1O/C(=C\I)CCC1c1cccc2ccccc12
18	CHEMBL2(	O=C(c1ccccc1)c1ccc(-n2ncc(=O)[nH]c2=O)cc1
19	CHEMBL6:	O=c1cnn(-c2ccc(C(O)c3ccc(Cl)cc3)c(Cl)c2)c(=O)[nH]1
20	CHEMBL6:	O=C(c1ccccc1)c1cc(-n2ncc(=O)[nH]c2=O)ccc1Cl

chembl_smiles (2)

Tableau IV.7 : chembl_smiles-9950 molécules

3.4 Traitement des données

Nous avons utilisé AlvaDesc pour recalculer les Descripteurs des 9 950 composés. Nous avons ensuite vérifié les colonnes disponibles afin de nous assurer qu'elles étaient incluses dans notre modèle [CATS3D_03_DA , CIC2 , T(N..N) , B03[C-N] , GATS4v]. Ce processus incluait également l'identification des colonnes manquantes afin de garantir l'intégrité des données utilisées dans l'analyse. Nous avons ensuite importé ces données au format CSV.

	A	B	C	D	E	F
1	CATS3D_0	CIC2	GATS4v	T(N..N)	B03[C-N]	SMILES
2	0	0.789066	1.064988	5	1	Cc1cc(-n2ncc(=O)[nH]c2=O)ccc1C(=O)c1ccccc1Cl
3	0	0.842196	1.108603	41	1	Cc1cc(-n2ncc(=O)[nH]c2=O)ccc1C(=O)c1ccc(C#N)cc1
4	0	1.000477	1.133475	5	1	Cc1cc(-n2ncc(=O)[nH]c2=O)cc(C)c1C(O)c1ccc(Cl)cc1
5	1	1.187636	1.028194	5	1	Cc1ccc(C(=O)c2ccc(-n3ncc(=O)[nH]c3=O)cc2)cc1
6	0	0.865591	1.098238	5	1	Cc1cc(-n2ncc(=O)[nH]c2=O)ccc1C(=O)c1ccc(Cl)cc1
7	0	1.062907	1.041109	5	1	Cc1cc(-n2ncc(=O)[nH]c2=O)ccc1C(=O)c1ccccc1
8	0	0.618463	1.041201	5	1	Cc1cc(Br)ccc1C(=O)c1ccc(-n2ncc(=O)[nH]c2=O)cc1Cl
9	0	0.795899	1.037662	5	1	O=C(c1ccc(Cl)cc1Cl)c1ccc(-n2ncc(=O)[nH]c2=O)cc1Cl
10	0	0.850289	1.116064	5	1	CS(=O)(=O)c1ccc(C(=O)c2ccc(-n3ncc(=O)[nH]c3=O)cc2Cl)cc1
11	0	2.435633	0.972763	54	1	c1cc2cc(c1)-c1cccc(c1)C[n+]1ccc(c3ccccc31)NCCCCCCCCCCNc1cc
12	0	0.842196	1.103669	5	1	CSc1ccc(C(=O)c2ccc(-n3ncc(=O)[nH]c3=O)cc2Cl)cc1
13	0	0.860799	1.070921	5	1	O=C(c1ccc(Cl)cc1)c1ccc(-n2ncc(=O)[nH]c2=O)cc1Cl
14	0	1.036289	1.088383	0	0	O=C1O/C(=C/Br)CCC1c1cccc2ccccc12
15	0	1.036289	1.066868	0	0	O=C1O/C(=C/I)CCC1c1cccc2ccccc12
16	1	0.975331	1.073515	0	0	C#C/C=C1\CCC(c2cccc3ccccc23)C(=O)O1
17	0	1.036289	1.066868	0	0	O=C1O/C(=C\I)CCC1c1cccc2ccccc12
18	0	1.372249	0.999292	5	1	O=C(c1ccccc1)c1ccc(-n2ncc(=O)[nH]c2=O)cc1
19	0	0.811611	1.131982	5	1	O=c1cnn(-c2ccc(C(O)c3ccc(Cl)cc3)c(Cl)c2)c(=O)[nH]1
20	0	0.931966	0.985575	5	1	O=C(c1ccccc1)c1cc(-n2ncc(=O)[nH]c2=O)ccc1Cl

chembl_with_descriptors

Tableau IV.8 : 9950 molécules with descriptors

Nous avons identifié les colonnes utilisables pour notre modèle, ce qui est essentiel pour garantir que les données sont adéquates pour l'analyse.

- **Colonnes du fichier CSV** :

['CIC2', 'GATS4v', 'T(N..N)', 'B03[C-N]', 'CATS3D_03_DA', 'SMILES']

- **Colonnes enregistrées dans descriptor_columns.pkl :**

['CATS3D_03_DA', 'CIC2', 'T(N..N)', 'B03[C-N]', 'GATS4v']

- **Colonnes partagées utilisables :**

['CATS3D_03_DA', 'CIC2', 'T(N..N)', 'B03[C-N]', 'GATS4v']

- **Colonnes manquantes ou manquantes dans le fichier CSV : NON**

III.5 Exécution d'Alo avec de nouvelles données

Nous avons d'abord rechargé le modèle et les descripteurs nécessaires pour l'optimisation, puis préparé les données en vérifiant la présence des colonnes requises et en les transformant au format approprié pour l'algorithme. Ensuite, nous avons lancé l'algorithme ALO afin d'obtenir des

propriétés optimisées, ce qui nous permet de générer de nouveaux composés avec les caractéristiques souhaitées**.**

Résultats de l'algorithme

Descripteur optimisé	**CATS3D_03_DA**	**CIC2**	**T(N..N)**	**B03[C-N]**	**GATS4v**
	7.5647	2.9352	3.0793e+03,	4.568e-01,	1.7404e-01

- Predicted PIC50 from optimized descriptor: 4.66
- Cosine similarity: 99%

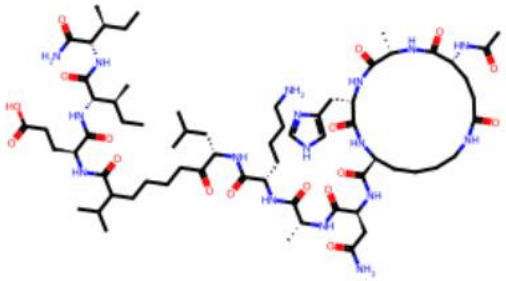


Figure IV.18 : Cabazitaxel molecule

Les résultats indiquent la performance du modèle à atteindre ses objectifs. La précision de la prédiction de l'activité biologique était élevée, reflétant l'efficacité des algorithmes SVR et ALO dans le traitement de données complexes. En conclusion, cette étude a permis de fournir un modèle solide pour l'application du QSAR inverse dans la découverte des médicaments. Les résultats ont montré que l'utilisation de l'algorithme ALO avec SVR peut être un outil efficace pour concevoir de nouveaux composés ayant une activité biologique ciblée.

CONCLUSION GÉNÉRALE

Conclusion Générale

L'objectif de cette étude était d'examiner l'importance de l'analyse quantitative de la relation structure-activité (QSAR) et des techniques de modélisation computationnelle, en mettant l'accent sur leurs applications dans le domaine de la découverte de médicaments en l'occurrence la maladie d'Alzheimer. Les modèles QSAR et QSAR inverse ont contribué de manière significative à la compréhension des mécanismes biologiques et à l'amélioration de l'efficacité des composés chimiques. Le rôle des cellules T gamma delta dans l'inflammation nerveuse associée à la maladie d'Alzheime a été mis en exergue, soulignant l'importance de ces cellules dans la régulation de la réponse immunitaire et leur impact sur la progression de la maladie.

Des techniques avancées telles que l'ALO et la SVR ont été utilisées pour prédire l'activité biologique d'un ensemble de composés chimiques. L'analyse des données et l'utilisation d'outils sophistiqués tels que PubChem, Data-Warrior et AlphaDesc nous ont permis d'identifier les descripteurs moléculaires représentatives des propriétés structurales des composés.

Les résultats obtenus ont montré que le modèle ALO-SVR présentait une grande précision de prédiction, avec des valeurs de R^2 de [0,9997, 0,9999, 1] et un RMSE fable de [0,0100], indiquant une forte corrélation entre les descripteurs moléculaires et l'activité biologique. La validation interne et externe du modèle a confirmé sa fiabilité et son applicabilité pour prédire l'activité inhibitrice de nouveaux composés.

De plus, le docking et l'ADMET ont contribué à une meilleure compréhension des interactions entre les composés cibles et les protéines, augmentant ainsi les chances de développer des médicaments efficaces et sûrs. L'intégration de ces outils ouvre de nouvelles perspectives pour la conception de médicaments innovants destinés au traitement de maladies incurables.

En perspective, cette recherche pourrait être élargie à l'exploration de nouvelles techniques de modélisation et d'optimisation, telles que les algorithmes génétiques, afin d'améliorer l'efficacité des modèles et la précision des prédictions. Ces travaux soulignent l'importance croissante des techniques d'IA dans la découverte des médicaments, qui contribuent à accélérer le développement de nouveaux traitements et à améliorer les résultats cliniques.

[1] Y. Zemouli, Les lymphocites T, faculté de médecine université constantine3, 2021, p. 1.

[2] Lauren Jachimowicz, Ming Li, Peifang Ye, Yan Lu et Garret Guenther, 'Analyse complète de l'état des lymphocytes T après activation avec un panel d'immunophénotypage à 16 couleurs sur le cytomètre en flux NovoCyte Advanteon', Agilent, Etats-Unis, 9 octobre 2019, p. 2.

[3] Rafik Terra, Développement des lymphocytes T : Mécanismes de différentiation extrathymique, thèse de Doctorat du Département de Sciences Biomédicales, Faculté de Médecine, université de Montréal, janvier 2005, p. 2

[4] Michelle D'Elia, Régulation des effets des glucocorticoides sur les lymphocytes T lors de la brûlure sévère, thèse de doctorat, Université du Québec, soutenue en 2009, p. 1.

[5] Marie -Hélène Lacombe, Etude du développement des lymphocytes T mémoires, mémoire de maîtrise en sciences biomédicales, Département de médecine, Faculté de médecine, Université de Montréal, soutenue Octobre 2003, p. 3.

[6] Yoan Velut, Caractérisation des clusters de lymphocytes T CD8+, T régulateurs FoxP3+ et des neutrophiles dans le cancer broncho-pulmonaire et son impact dans l'immunothérapie, Thèse de doctorat d'Immunologie, université de Surbonne, soutenue le 12 Décembre 2022 .

[7] Matthieu Simon, Les lymphocytes T, Cours Pharmacie (https://www.cours-pharmacie.com/immunologie/les-lymphocytes-t.html?form=MG0AV3, 07 septembre 2009.

[8] Véronique Adoue, Olivier Joffre, Les rétrovirus endogènes Un rôle clé dans la programmation des lymphocytes T CD4, revue médecine/sciences, Vol. 36, février 2020, p. 253.

[9] Zach Murphy, Lymphocytes T CD4 : structure, fonctions, maturatio, Olliewood, 03/07/2024.

[10] Neva B. Watson, Ravi K. Patel, The gene regulatory basis of bystander activation in $CD8^+$ T cells, ScienceImmunology, 23/02/2024, p. 1.

[11] Bernardo S. Reis et autres, TCR-Vγδ usage distinguishes protumor from antitumor intestinal γδ T cell subsets, PMC PubMed, 27/07/2022, p. 2.

[12] Kyle K. Payne, Jessica A. Mine, BTN3A1 governs antitumor responses by coordinating αβ and γδ T cells, Science, 21/08/2020, p. 1.

[13] Chang Zou, Pan Zhao, Zhangang Xiao, Xianghua Han, Fan Fu and Li Fu, in γδ T cells in cancer immunotherapy, Impactjounals, Volume 08, Numéro 05, 03/11/2016, p. 8900.

[14] Camille Khairallah, Infection par le Cytomégalovirus murin : réponse des lymphocytes T gamma delta et impact sur le développement tumoral, thèse de doctorat, soutenue le 02/11/2015, Université de Bordeaux .

[15] Bénédicte Buttard, Caractérisation des lymphocytes T gamma delta et de L'expression de la protéine BTN3A dans le cancer colorectal, thèse soutenue le 18/12/2023, Université de Sorbonne.

[16] Maya Hassane, Etude des lymphocytes T Gamma-delta producteurs d'interleukine-17 au cours des infections respiratoires, thèse soutenue le 14/06.2017, Université Lille 2, p. 20.

[17] Selma Cornillot-Clément, Développement d'un biohybride conjuguant lymphocytes T γδ et polymersomes : combinaison de la thérapie cellulaire et du drug delivery, thèse soutenue le 25/11/2024, Université de Bordeaux, pp. 42-45.

[18] T. Kuntzer et M. Chofflon, Neuropathies inflammatoires et mononévrites, Revue Médicale Suisse, décembre 2009, p. 2469.

[19] Heimo Steffen et David Tabibian, Névrite du nerf optique : diagnostic, traitement et implications cliniques, Revue Médicale Suisse, 2015, p. 2388.

[20] Antoine Reinhard Raphael Maire,'Névrite vestibulaire : traitement et pronostic, Revue Médicale Suisse, octobre 2013, p. 1775.

[21] Stephen Jacques Alzuphar et Raphaël Maire, Névrite vestibulaire inférieure, Revue Médicale Suisse, 2017, p. 1690.

[22] Alberto L. Epstein, Maladie d'Alzheimer, neuro-inflammation et virus herpétiques une piste qui trace son chemin, médecine/sciences, Vol. 36, mai 2020, p. 479.

[23] Thakur, S., Dhapola, R., Sarma, P. *et al.* Neuroinflammation in Alzheimer's Disease: Current Progress in Molecular Signaling and Therapeutics, Springer Nature, 20/08/2022.

[24] Hélène Amieva, Sandrine Andrieu, Claudine Berr, Luc Buée, Frédéric Checler, Serge Clément, Jean-François Dartigues, Béatrice Desgranges, Bruno Dubois, Charles Duyckaerts, et al., Maladie d'Alzheimer : enjeux scientifiques, médicaux et sociétaux, Institut national de la santé et de la recherche médicale (Inserm), paris, octobre 2007, p. 3.

[25] Linbin Dai, Yong Shen, Insights into T-cell dysfunction in Alzheimer's disease, Aging Cell. 2021.

[26] Alexander L. R. Lubbock, Carlos F. Lopez, Programmatic modeling for biological systems, PubMed, 27/09/2021, p. 1.

[27] Muffy Calder, Claire Craig, Computational modellingfor decision-making: where, why, what, who and how, The Royal Society Publishing, 2018, p. 2.

[28] Computational Modeling, National Institute of Biomedical Imaging and Bioengineering, mai 2020, p. 1.

[29] Ludovic Saint-Bauzel, Modélisation computationnelle de la coopération physique patient-robot : De la prédiction du mouvement pathologique pour la commande de robots d'assistance à l'étude de l'interaction humain-robot, soutenue le 08/12/2021, Université de Sorbonne.

[30] David Svoboda, Ondˇrej Havelka, Julie Holendova´ , Jiˇrí Kraft , An epidemiological model of SIR in a nanotechnological innovation environment, Heliyon 11 2025, p. 6.

[31] Gaurav Pandey, Poonam Chaudhary, Rajan Guptab , Saibal Palc, SEIR and Regression Model based COVID-19 outbreak predictions in India, arXiv, 2020m p. 2.

[32] Laurent Crépin, Couplage de modèles population et individu-centrés pour la simulation parallélisée des systèmes biologiques. Application à la coagulation du sang, thèse soutenue le 28/10/2013, Université de Bretagne occidentale, p. 5.

[33] Thierry Foucart, Colinéarité et régression linéaire Collinearity and linear regression analysis, Mathematics and Social Sciences, n° 173, 2006, pp. 5-6.

[34] Julie Desjardins, L'analyse de régression logistique, Tutorial in Quantitative Methods for Psychology, 2005, Vol. 1, p. 35.

[35] Don van den Bergh, Johnny van Doorn, A Tutorial on Conducting and Interpreting a Bayesian ANOVA in JASP, L'année psychologique/Topics in Cognitive Psychology, mars 2020, p. 74.

[36] Youssef HMAMOUCHE, Prédiction des Séries Temporelles Larges, thèse soutenue le 13/12/2018, Université de Aix Marseille, p. 14.

[37] S. Brillouet, S. Dorbes b,c, S. Silvente-Poirot b, B. Mestre-Voegtlé c, C. Picard c, M. Poirot b, F. Courbon, Ciblage peptidique en oncologie nucléaire : intérêt de la modélisation moléculaire, Médecine Nucléaire 34, 2010, p. 291,

[38] SEBASTIEN FIORUCCI, Activités biologiques de composés de la famille des flavonoïdes : approches par des méthodes de chimie quantique et de dynamique moléculaire, Soutenue publiquement le 6 Octobre 2006, p. 206-207.

[39] Anis Bessadok, La multiplicité de transport de la P-glycoprotéine : Etudes de modélisation comparative et de docking au sein de la famille des protéines ABC, thèse soutenue le 11/07/ 2011, Université de Sorbonne, p. 155.

[40] Ludovic Arnold, Hélène Paugam-Moisy, Michèle Sebag, Optimisation de la Topologie pour les Réseaux de Neurones Profonds, 17e congrès francophone AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle - RFIA 2010, Jan 2010, Caen, France, p. 3.

[41]

[42] Thierry Lubin, Equations aux dérivées partielles (EDP), Méthode de résolution des EDP par séparation de variables ; Applications, Cours, Faculté des Sciences et Technologies, Université de Lorraine, 2017, pp. 6-7.

[43] E. Barbieri1, L. Capocchi1, J.F. Santucci1, Modélisation et simulation DEVS des effets de levier financier en apprentissage par renforcement, Campus Grimaldi, Université de Corsica, pp. 2-3.

[44] HAMMOUMI Feriel, Modélisation Moléculaire : Classification Et Intérêt Dans La Conception Des Médicaments, mémoire de doctorat en pharmacie, soutenu le 16/10/2021, Faculté de médecine, Université Abou Bekr Belkaid.

[45] Kairouani Malika, Bachouche Nadjet, Etude QSAR sur les dérivés de l'imidazopyridazine comme inhibiteurs de la PfPK7, mémoire de Master, soutenu en juillet 2021, Faculté des Sciences, Université de Médéa, p. 14.

[46] Chanin Nantasenamat, Chartchalerm Isarankura-Na-Ayudhya, Thanakorn Naenna, Virapong Prachayasittikul, A Practical Overview of Quantitative Structure-Activity Relationship, EXCLI Journal 2009.

[47] Metidji Chourouk, Djabali Fella, Utilisation de la méthode QSAR –SVMDA et le Docking pour la prédiction de l'activité inhibitrice des inhibiteurs et l'évaluation des interactions entre le récepteur et le ligand du Cancer de la peau, mémoire de Master, Université Yahia Fares de Médea, 2022-2023.

[48] H a d a j i e l g h a l i a , Corrélation - (structure - activité anticancéreuse) par les méthodes QSAR des molécules hétérocycliques précurseurs de médicaments , thèse de doctorat soutenue le 26/12/2018, Université de Fès, p. 20.

[49] Dabouz Radhia Bouhemhem Sarra, Etude comparative de l'efficacité prédictive des modèles 2D-QSAR et 3D-QSAR pour la conception de molécules anti-malaria, mémoire de master, soutenue le 20/06/2023, Faculté des Sciences de la Nature et de la Vie, Université Frères Mentouri Constantine 1, pp.

[50] Aurélie Goulon-Sigwalt-Abram, Une nouvelle méthode d'apprentissage de données structurées : applications à l'aide à la découverte de médicaments, Université Pierre et Marie Curie, Paris 6, p. 12.

[51] Benchekor Hanane Aichouche Farida, Développement de modèle de type RQSA-RNA pour la production des quelques paramètres pharmaceutiques humain des composés médicamenteuses administré par la voie intraveineuse, mémoire de master, soutenue juin 2020, Faculté de Technologie, Université Yahia Fares de Médea, p. 34.

[52] Dabouz Radhia, Bouhemhem Sarra, Etude comparative de l'efficacité prédictive des modèles 2D-QSAR et 3D-QSAR pour la conception de molécules anti-malaria, mémoire de master, soutenue le 20/06/2023, Faculté des Sciences de la Nature et de la Vie, Université Frères Mentouri Constantine 1, p. 20.

[53] Benchekor Hanane Aichouche Farida, Développement de modèle de type RQSA-RNA pour la production des quelques paramètres pharmaceutiques humain des composés médicamenteux administré par la voie intraveineuse, mémoire de master, soutenue juin 2020, Faculté de Technologie, Université Yahia Fares de Médea, p. 35

[54] Yousfi Youcef, Etude QSAR de l'activité anti-oxydante d'une série de composes phénoliques, mémoire de Master soutenue le 19/06/2017, Faculté des sciences, Université Abou-Bekr Belkaid, Tlemcen, p. 23.

[55] Meddour Moufida, Rouaissia Hamida, Etude par Docking Moléculaire de l'activité inhibitrice des Furanocoumarines Contenues dans le Jus de Pamplemousse sur les Cytochromes CYP3A4, mémoire de Master, Université 8 mai 1945 Guelma, 2020.

[56] Brut M. Nouvelle approche méthodologique pour la prise en compte de la flexibilité dans les interactions entre molecules biologiques : les Modes Statiques, thèse de doctorat soutenue Le 5/03/2009, Université Toulouse III - Paul Sabatier ; p. 43.

[57] William WL Wong, Forbes J Burkowski, A constructive approach for discovering new drug leads: Using a kernel methodology for the inverse-QSAR problem, Journal of Cheminformatics, 28 April 2009, p. 1.

[58] Yu Shi, Jianshen Zhu, An Inverse QSAR Method Based on a Two-Layered Model and Integer Programming, International Journal of Molecular Sciences, 11 March 2021, p. 1.

[59] Tomoyuki Miyao, Kimito Funatsu, Jürgen Bajorath, Exploring differential evolution for inverse QSAR analysis, F1000Research, 31 Jul 2017, p. 3-5.

[60] S. Agatonovic-Kustrin, R. Beresford, Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research , Journal of Pharmaceutical and Biomedical Analysis, 17 September 1999,p1

[61] Fan Zhang, Lauren J. O'Donnell, Methods and Applications to Brain Disorders 2020, Pages 123-140,Support vector regression CHAPTER 7,P 123

[62] Mirjalili, S., & Lewis, A. (2016). "The Ant Lion Optimizer." Advances in Engineering Software