REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE YAHIA FARES DE MEDEA



Faculté de Technologie

Mémoire de fin d'études de Master

Filière : Génie des Procédés

Spécialité : Génie des Procédés Pharmaceutiques

THEME

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

Proposé et Dirigé par :

Pr. Benkortbi Othmane

Dr. Madani Achwak

Présenté par :

Metidji Chourouk

Djabali Fella

Promotion 2022-2023

Dédicaces

Je tiens à dédier ce projet de fin d'études à plusieurs personnes qui ont contribué de manière significative à sa réalisation .Tout d'abord, je tiens à exprimer ma gratitude envers **Allah** pour la force et le courage qu''il m'a accordé. Mes remerciements vont également envers ma famille, mes proches et mes amis qui m'ont soutenu et encouragé tout au long de cette aventure. Leur présence et leurs encouragements ont été une source de motivation inestimable, Et ce n'est que le début d'une aventure qui mérite un dédicace spéciale **à M. O. Benkortbi, Professeur à l'université de Medea** et au **Dr A. Madani**, qui ont joué un rôle crucial dans la réussite de ce projet. Leurs expertises, leurs conseils avisés, leurs soutiens indéfectible et leurs disponibilités tout au long de ce projet. Ont été des éléments clés de cette réalisation. Je leurs suis profondément reconnaissants pour leurs contributions précieuses. Enfin, Ce dédicace est un humble expression de gratitude envers toutes les personnes qui ont généreusement apporté leurs contributions à la réalisation de ce projet.

Metidji chourouk

Dédicaces

Parfois, dans notre parcours de vie, nous atteignons un point où nous oublions les étapes qui nous ont conduits jusqu'ici. Tout d'abord Je suis reconnaissante envers Allah pour la force et le courage qu'Il m'a accordé ; Je tiens à dédier ce projet de fin d'études à tous ceux qui ont contribué à mon parcours académique et personnel :

À mes chers parents, qui ont été mes plus grands soutiens tout au long de mes études.

À mes frères et sœurs, qui ont été mes compagnons de route, mes confidents et mes inspirations.

À mes amis fidèles, qui ont été présents à mon côté pendant toutes ces années

À mes précieux promoteurs **Mr, O. Benkortbi** et au **Dr A. Madani** qui ont été des piliers essentiels dans la réalisation de ce projet.

Djabali fella

Remerciements

En premier lieu, nous exprimons notre gratitude à Allah, le Tout-Puissant, pour nous avoir guidés et avoir été la source de courage et de force qui nous a permis de mener à bien ce modeste travail.

Nous tenons à exprimer notre profonde gratitude envers tous ceux qui nous ont aidés et soutenus dans les moments difficiles. .

Nous adressons nos profonds remerciements envers notre promoteur, **M. BENKORTBI Othmane**, Professeur à l'université de Médéa, pour sa patience, ses encouragements et ses précieuses orientations. Sa vision critique a été d'une grande valeur pour structurer ce travail et en améliorer la qualité.

Nos remerciements vont à notre Co-promotrice **Melle MADANI Achwak** pour avoir eu la patience de répondre à nos innombrables questions.

Notre respect aux membres du jury qui ont bien voulu accepter de juger ce modeste travail, d'apporter leurs réflexions et leurs critiques scientifiques.

En conclusion, nous tenons à exprimer notre profonde gratitude envers tous les employeurs et administrateurs de notre université.

À toutes les personnes qui nous ont encouragé et soutenu, que ce soit de près ou de loin, tout au long de la réalisation de ce travail.

ملخص:
يركز مشروع نهاية الدراسات هذا على استكشاف عقاقير جديدة للورم الميلاني ، من خلال تطوير نموذج QSARلربط بنية الجزيئات بنشاطها ، ولا سيما .pGI50 تم تطبيق خوارزمية SVM_DA لتنفيذ هذه المهمة. تم استخدام الخوارزمية التدريجية للحصول على الواصفات التمثيلية لقاعدة البيانات المختارة وفقًا لحداثتها ونشاطها المثبت. تم تقييم النموذج المركب ومقارنته باستخدام معايير تحقق داخلية وخارجية مختلفة ، مع تحديد المنطقة المحددة لتطبيق النموذج. أظهر نموذج SVM_DA قدرة وقوة تنبؤية استثنائية ، بمعايير R2: 0.9613 0.8230و 0.9428 وRMSE = 0.37
بالإضافة إلى ذلك ، تم استخدام ADME السويسري للحصول على المعلمات الدوائية ، بينما تم استخدام ADMET Lab 2.0لتقييم معلمات السمية. بالإضافة إلى ذلك ، تم إجراء الالتحام بواسطة Ubuntu من أجل دراسة التفاعلات بين الجزيئات والهدف.
الكلمات الرئيسية: السرطان ، QSAR، SVM، الإرساء ، التنبؤ

Résumé:

Ce projet de fin d'études se concentre sur l'exploration de nouveaux médicaments pour le mélanome, en développant un modèle QSAR permettant de corréler la structure des molécules avec leur activité, notamment le pGI50. L'algorithme SVM_DA a été appliqué pour réaliser cette tâche. L'algorithme Stepwise a été utilisé pour obtenir les descripteurs représentatifs de la base de données sélectionnée en fonction de sa nouveauté et de son activité avérée. Le modèle construit a été évalué et comparé en utilisant différents critères internes et de validation externe, tout en déterminant le domaine d'application spécifique du modèle. Le modèle SVM_DA a démontré une capacité prédictive et une robustesse exceptionnelle, avec des critères de R^2:[0,9613 0,8230, 0,9428] et RMSE=0,37
Par ailleurs, le Swiss ADME a été utilisé pour obtenir les paramètres Pharmacologiques, tandis que l'ADMET Lab 2.0 a été employé pour évaluer les paramètres de toxicité. En complément, le docking par Ubuntu a été effectué afin d'étudier les interactions entre les molécules et la cible.

Mots clés : Cancer, QSAR, SVM, Docking, Prediction

Abstract :

This final year project focuses on exploring new drugs for melanoma by developing a QSAR model to correlate the structure of molecules with their activity, particularly pGI50. The SVM_DA algorithm was applied to accomplish this task. The Stepwise algorithm was used to obtain representative descriptors from the selected database based on novelty and proven activity. The constructed model was evaluated and compared using various internal and external validation criteria, while determining the specific applicability domain of the model. The SVM_DA model demonstrated excellent predictive capacity and robustness, with criteria values of R2: [0,9613 0,8230, 0,9428] et RMSE=0,37 Additionally, SwissADME was utilized to obtain pharmacological parameters, while ADMET Lab 2.0 was employed to evaluate toxicity parameters. Furthermore, docking using Ubuntu was performed to study the interactions between the molecules and the target.

Keywords: Cancer, QSAR, SVM, Docking, Prediction

Liste des tableaux

Liste des figures

Liste des abréviations

Introduction générale 01

Chapitre I: Le cancer

I.1. Définition 04
I .2. Les types 04
I .2.1. L'organe 04
I .2.2. La nature 04
I. 2.2.1. Les cancer solides 04
I. 2.2.2. Les cancer métastatiques ou « disséminés » 04
I .2.2.3. Les cancer secondaires 05
I .3. Les causes 05
I .3.1. Le tabac 05
I .3.2. L'alimentation 05
I .3.3. L'alcool 05
I .4. Les symptômes 06
I .4.1. Les douleurs 06
I .4.2. Saignement 06
I .4.3. Perte de poids et fatigue 06
I .4.4. Gonflement ganglions lymphatique 07
I .4.5. Symptôme respiratoire 07
I .4.6. Symptôme neurologique et musculaire 07
I .5. Le traitement 07
I .5.1. Le but 07
I .5.2. Le choix 07
I .5.3. Les types 08
I .5.3.1. La churigie 08
I .5.3.2. La chimiothérapie 08
I .5.3.3. La radiothérapie 09

I .5.3.4. Les thérapies ciblées 11
I .5.3.5. L'hormonothérapie 11
I .5.3.6. L'immunothérapie 12
I .6. Conclusion 12

Chapitre II: Le cancer de la peau (La Mélanome)

II.1. Définition 13
II.2. Les types 13
II.2.1. Le carcinome basocellulaire 13
II.2.2. Le carcinome épidermoïde 13
II.2.3. Le mélanome 13
II.2.3.1. Les causes et les facteurs 14
II.2.3.2. Les stades 15
II.2.3.3. Les symptômes 16
II.2.3.4. Le dépistage du mélanome 16
II.2.3.5. Le diagnostic du mélanome 17
II.2.3.6. Le traitement du mélanome 17
II.3.Conclusion 18

Chapitre III : Conception des médicaments assisté par ordinateur

III.1. Approche Basé sur la structure 19
III .1.1. Généralités 19
III .1.2. Définition 19
III .1.3. Historique 19
III .1.4. Objectif 21
III .1.5. Principe 21
III .1.6. Stratégie globale 21
III .1.7. Modèles QSAR/QSPR 22
III .1.8. Les descripteurs moléculaires 22
III .1.9. Méthodologie générale 24
III .1.9.1. Base de données 24
III .1.9.2. Descripteurs moléculaires 24
III .1.9.3. Sélection des descripteurs 25
III .9.3.1 Sélection objective 25

III .9.3.2 Sélection subjective 26
III .1.9.4. Division de la base de données 26
III .1.9.5. Méthode d'analyse de données 26
III .1.9.6. Interprétation et validation d'un modèle 27
III .9.6.1. Validation interne 27
III .9.6.2. Validation externe 27
III .1.9.7. Domaine d'applicabilité 27
III .2. Approche Basé sur le ligand 28
III .2.1. Définition : 28
III .2.1.1. ADMET 28
III .2.1.2. Docking 29
III .2.2. Le principe du Docking 29
III .2.3. Les étapes de Docking 29
III .2.4. Les types 30
III .3. Conclusion 31

Chapitre IV : Techniques d'apprentissage supervises

IV.1. Introduction 32
IV. 2. Principe générale 32
IV. 3. But 32
IV. 4. Régression par machines à vecteur de support 33
IV.4.1. Problème linéaire 33
IV.4.2. Problème non-linéaire 36
IV.5. Fonction de Base Radiale RBF 37
IV.6. Prise en charge de la régression vectorielle de la machine 38
IV.6.1. SVM et SVMr 38
IV.6.2. Dragonfly algorithme 40
IV.7.Conclusion 41

Chapitre V : Exploitation des résultats

V.1. METHODOLOGIE DE TRAVAIL 42
V.1.1. Collecte de la base de données 42
V.1.2. Sélection des descripteurs 43
V.1.3. Division de la base de données et développement du modèle 44

V.1.3.1. SVM 44
V.1.3.2. Validation 45
V.1.4. Domaine d'applicabilité 46
V.1.5. ADMET 46
V.1.6. Prédiction des propriétés pharmacocinétiques (ADMET) et de la Similitude avec les médicaments 47
V.1.7. Préparation des structures 47
V.1.7.1 Préparation des ligands 47
V.1.7.2 Préparation de la protéine et étude de Docking 47
V.2. RESULTATS ET DISCUSSION 48
V.2.1. Sélection des descripteurs représentatifs 48
V.2.2. Développement des modèles QSAR 56
V.3. VALIDATION EXTERNE 61
V.4. DOMAINE D'APPLICABILITE 61
V.4.1. Algorithme SVR-Dragonfly 61
V.5. EXPLOITATION DES RESULTATS ' DOCKING 62
V.5.1. ADMET 62
V.5.2. Les interactions entre le récepteur et le ligand 62

Conclusion générale 70

Référence bibliographique

Liste des tableaux

Chapitre V: Exploitation des résultats

Tableau V.1 : Les descripteurs pertinents

Tableau V.2 : Descripteurs retenus

Tableau V.3 : Critères statistiques pour le modèle DA_SVR

Liste des figures

Chapitre I : Le cancer

Figure I.1: La radiothérapie externe

Figure I.2: La curiethérapie

Figure I.3: La radiothérapie métabolique

Chapitre IV: Technique d'apprentissage supervisé

Figure IV.1: Fonction d'erreur ε-insensible

Figure IV.2: Tube d'insensibilite de rayon ε

Chapitre V : Exploitation des résultats

Figure V.1: Capture d'écran Data Warrior

Figure V.2: Capture d'écran Alva Desc

Figure V.3: Word cloud de l'importance des descripteurs

Figure V.4: Bar plot de l'importance relative des descripteurs

Figure V.5: Diagramme de corrélation

Figure V.6: Courbes de Régression SVM-DA

Figure V.7 : Domaine d'applicabilité du modèle SVM –DA

Figure V.8 : Diagramme des interactions 2D d'Amsacrine avecV600E-BRAF

Figure V.9 : Vue en 3D du composé d'Amsacrine dans le site de liaison de V600-BRAF

Figure V.10 : Diagramme des interactions 2D d'Arnebin 1 avecV600E-BRAF

Figure V.11: Vue en 3D du composé d'Arnebin 1 dans le site de liaison de V600-BRAF.

Figure V.12: Diagramme des interactions 2D de Camptothecin Butylglycinate Ester Hydrochloride avecV600E-BRAF

Figure V.13: Vue en 3D du composé de Camptothecin Butylglycinate Ester Hydrochloride dans le site de liaison de V600-BRAF

Figure V.14: Diagramme des interactions 2D de Hydroxyurea avec V600E-BRAF

Figure V.15: Vue en 3D du composé hydroxyurea dans le site de liaison de V600-BRAF

Figure V.16: Diagramme des interactions 2D de Hycanthone MesylateavecV600E-BRAF

Figure V.17: Vue en 3D du composé hycantone mesylate dans le site de liaison de V600-BRAF.

Figure V.18: Diagramme des interactions 2D de thiotepa avecV600E-BRAF

Figure V.19: Vue en 3D du composé thiotepa dans le site de liaison de V600-BRAF.

Liste des abréviations

Chapitre I : le cancer

OMS : Organisation mondiale de la santé

Chapitre III: Conception des médicaments assistés par ordinateur

QSAR : Relations quantitative structure- activité

QSPR : Relations quantitative structure- propriété

QSRR : Régression quantitative des relations structure-propriétés

CoMFA : Analyse comparative du champ moléculaire

CoMSIA : Analyse comparative des indices de similarité moléculaire

HOMO : Orbitale moléculaire occupée la plus élevée

LUMO : Orbitale moléculaire inoccupée la plus basse

LOO: (Leave-One-Out) Laissez-Un-Out

LMO : (Leave-Many-Out) Laisser-beaucoup-dehors

DA : Domaine d'applicabilité

OCDE : Organisation de coopération et de développement économique

Chapitre IV: Techniques d'apprentissage supervises (SVM)

SVM : Machines à vecteurs de support

SVR : Régression à Vecteurs de Support

SVMR : Régression par Machines à Vecteurs de Support

SLT : La théorie d'apprentissage statistique

SRM : Minimisation des risques de structures

Liste des symboles

W, x	Produit scalaire entre le vecteur de poids w et le vecteur x.
b	Constantes à déterminer.
||w||	La norme du vecteur de poids w.
ξi,ξi*	Variable d'écart.
ε	tolérance dans le contexte donné
ηi, ηi*	Les multiplicateurs de Lagrange
x, xi	Produit scalaire entre le vecteur d'entrée x et le vecteur xi.
C	Paramètre de régularisation dans le SVM
αi, $\alpha i*$	Coefficient associé à l'échantillon i dans le SVM.
$\xi i*$	Vecteur d'entrée associé à l'échantillon i.
$i*$	Variable associée à l'échantillon i
Φ	La fonction mathématique définie ici.
F(x)	La valeur de la fonction pour une entrée x.
k (xi, x)	Une fonction noyau (kernel)
T	Un vecteur transposé.
C1	Un coefficient multiplicatif.
ξ	Un vecteur.
Ø(x)	fonction de transformation des caractéristiques de l'entrée x.
T	Le transposé du vecteur de poids ω.
Nd	Le nombre total d'échantillons dans l'ensemble de données.
W^T	Le vecteur transposé des poids ou coefficients du modèle SVM.
Yi	La valeur cible ou l'étiquette associée à l'échantillon i.

Vj	Les termes individuels que vous additionnez.
Jl	L'indice de la variable dans la sommation

INTRODUCTION GENERALE

Le développement de nouveaux médicaments est un processus complexe et coûteux qui nécessite une compréhension approfondie des interactions entre les composés chimiques et leurs cibles biologiques. Dans ce contexte, les techniques de modélisation in Silico telles que le QSAR (Quantitative Structure-Activity Relationship) en se basant sur le SVM (support victor Machine) et le Docking moléculaire ont émergé comme des outils puissants pour accélérer et rationaliser la découverte de médicaments. Le domaine de modélisation a connu une croissance significative ces dernières années, et joue un rôle crucial dans de nombreux secteurs, tels que l'ingénierie, science de la santé, l'informatique et technologie …etc avec des avancées technologiques et des défis complexes qui nécessitent une recherche approfondie.

Dans ce contexte, ce mémoire se concentre sur la recherche d'un modèle QSAR également connu sous le nom de relation quantitative structure-activité , c'est une équation qui établit une relation mathématique entre la structure chimique des composés et leurs activités biologiques en se basant sur l'AI et parmi les approches les plus utilisables le SVM, en utilisant des descripteurs moléculaires et des algorithmes de modélisation statistique, qu'il a une grande importance dans la découverte de nouveaux médicaments en raison d'améliorer la fiabilité et la précision du modèle et de faciliter et accélérer le processus de la découverte, se concentre aussi sur la méthode de modélisation 'Docking' qui permet d'étudier et de prédire les interactions entre une petite molécule, généralement un médicament candidat, et une macromolécule cible, telle qu'une protéine.

Cependant, malgré les progrès réalisés dans ce domaine, il existe encore des lacunes et des défis à relever. Parmi ceux-ci, nous nous intéressons

particulièrement à la prédiction précise et fiable de l'activité biologique des composés chimiques en utilisant des descripteurs moléculaires. Cette problématique présente des implications significatives de l'association du QSAR et du Docking moléculaire, permet d'avoir une approche intégrée et complémentaire dans la découverte de médicaments. Le QSAR fournit des informations prédictives sur l'activité et les propriétés des composés chimiques, tandis que le Docking moléculaire permet d'explorer les interactions moléculaires détaillées et de comprendre les mécanismes d'action, elle nécessite une analyse approfondie pour apporter des solutions innovantes et durables. L'objectif principal de ce mémoire est donc de fournir une vision globale et approfondie des applications du QSAR et du Docking moléculaire dans la découverte de médicaments et de développer des relations quantitatives entre la structure chimique des composés et leur activité biologique, physicochimique ou toxicologique. Ces relations permettent de prédire et d'évaluer de manière rapide, précise et économique les propriétés et les effets des composés chimiques. Pour atteindre cet objectif, nous avons fixé des objectifs secondaires qui permettent de guider la conception de nouveaux composés plus actifs, plus sélectifs, plus sûrs et plus durables dans divers domaines tels que la découverte de médicaments, l'agrochimie, la chimie environnementale, etc.

Le mémoire est subdivisé en quatre chapitres Le premier chapitre est consacré à une étude bibliographique sur le cancer, le chapitre deux trait au cancer de la peau ‘ le mélanome ‘Des généralités sur les méthodes QSAR et Docking sont présentés dans le chapitre trois. Le chapitre quatre décrit la méthode utilisée en l’occurrence la modélisation par machines à vecteurs supports (SVM) pour la conception du modèle, La procédure utilisée pour le développement du modèle pour prédire les interactions entre les ligands et le récepteur, ainsi que l’interprétation des résultats obtenus sont discutés en détail dans le chapitre cinq 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.

LE CANCER

I.1. DEFINITION :

Le Cancer est une maladie provoquée par la transformation de cellules qui deviennent anormales et prolifèrent de façon excessive. Ces cellules déréglées finissent par former une masse qu'on appelle tumeur maligne. Les cellules cancéreuses ont tendance à envahir les tissus voisins et à se détacher de la tumeur. [2]

Selon l'OMS : Est un terme générique désignant un large groupe de maladies pouvant toucher n'importe quelle partie de l'organisme. On parle aussi de tumeurs malignes et de néoplasmes. L'un des traits caractéristiques du cancer est la multiplication rapide de cellules anormales à la croissance inhabituelle, qui peuvent ensuite envahir des parties voisines de l'organisme, puis migrer vers d'autres organes. On parle alors de métastases. La présence de métastases étendues est la principale cause de décès par cancer.[3]

I.2. LES TYPES :

Chaque organe et chaque type de cellule peut être à l'origine d'une tumeur cancéreuse donc les types de cancer sont classés selon :

I.2.1 L'organe :

Cancer de peau, cancer du sein, cancer de la prostate, cancer de la colorectale (gros intestin) …etc

I.2.2. La nature :

Il n'existe pas moins de 200 types de cancers différents. On peut les répartir en 3 grandes catégories :

I.2.3.1. Les cancers solides :

Les carcinomes sont issus de cellules épithéliales (peau, muqueuses, glandes). Exemples : cancers du sein, des poumons, de la prostate, de l'intestin, etc. Les sarcomes, moins fréquents, sont issus de cellules des tissus conjonctifs (dits tissus de Soutien). Exemples : cancers de l'os, du cartilage, etc…

I.2.3.2. Les cancers métastatiques ou «disséminés » :

Un cancer métastatique ou disséminé signifie que les cellules cancéreuses ont migré à travers le corps, par la voie des vaisseaux sanguins ou lymphatiques et ont colonisé un ou plusieurs autre(s) tissu(s), loin de la tumeur d'origine. Différents organes peuvent être colonisés par ces métastases : os, foie, cerveau, poumon, etc. Mais la nature d'un cancer reste toujours déterminée par son point de départ.

I.2.3.3. Les cancers secondaires :

Il arrive que certains traitements anti cancer, indispensables pour guérir la maladie, puissent eux-mêmes provoquer l'apparition ultérieure de nouvelles cellules cancéreuses. Ce sont les cancers secondaires, consécutifs à un traitement. Le type de cancer, sa localisation, son degré d'agressivité, ses manifestations cliniques et l'état de santé du patient déterminent la façon dont le cancer est traité. [4]

I.3. LES CAUSES :

Les trois principales causes du cancer :

I.3.1. Le tabac :

Le tabac est une plante (Nicotiana tabac et Nicotiana rustiqua) qui contient de la nicotine. Les feuilles de ces plantes font souvent des nombreux produits comme :Le tabac à rouler, le tabac à pipe, le tabac à pipe à eau, les cigares, les cigarillos, le tabac à mâcher, le tabac à priser ainsi que les nouveaux produits à base de tabac chauffé [5] Mais la forme la plus connu et consommé au monde est la cigarette, défirent substance entre à la composition de cette aérosol causent le tumeur tel que le goudron, mais l'on trouve aussi des gaz toxiques comme le monoxyde de carbone, divers agents de saveur et des métaux lourds tels que le mercure ou le plomb(1).[6] en guise de solution les médecins recommande de réduire la consommation de tabac .

I.3.2. L'alimentation :

Le régime alimentaire est défini comme la consommation d'aliments pour répondre aux besoins nutritionnels ou pharmaceutique de tous les organismes pour leurs fonctions vitales, y compris la croissance, la locomotion, la reproduction et le maintien de la température corporelle. Les trois facteurs des risques sont :

- ❖ L'excès de viande rouge
- ❖ l'excès du sel et d'aliments salés expose à un sur risque de cancer de l'estomac.
- ❖ Des apports en fibres, en général équivalents à au moins 25g par jour (fruits et légumes, céréales complètes, oléagineux…) [7]

I.3.3. Alcool :

L'alcool est le nom populaire du principe actif psychotrope des boissons alcoolisées, l'éthanol. Il est utilisé par l'homme depuis des siècles et est obtenu par la fermentation de plantes riches en sucre comme les betteraves et les pommes. L'Organisation mondiale de la santé ne considère que le cancérigène est aussi remarquablement additif qu'une drogue illicite. [8]

I.4. LES SYMPTOMES :

I.4.1. Les douleurs :

Le plus souvent, le cancer se développe en silence, sans faire mal. C'est ce qui le rend si difficile à diagnostiquer. La tumeur n'a donc aucune sensibilité et n'est pas douloureuse. [9] Bien que la douleur puisse être un symptôme précoce de certains cancers, comme les tumeurs cérébrales qui causent des maux de tête et les cancers qui affectent la tête et le cou, et les cancers de l'estomac et de l'œsophage qui causent des problèmes de déglutition. Au fur et à mesure qu'ils grandissent, les premiers symptômes sont généralement un léger inconfort, qui s'aggrave ensuite jusqu'à devenir une douleur intense, transformant l'invasion en une invasion tumorale. Les douleurs causées par la tumeur sont mécaniques, inflammatoires ou neuropathiques.[10]

I.4.2. Saignement :

Au départ, le cancer peut saigner légèrement car ses vaisseaux sanguins sont fragiles. Ensuite, à mesure que le cancer se développe et envahit les tissus environnants, il peut se développer dans les vaisseaux sanguins adjacents et provoquer des saignements. Cela peut être subtil et indétectable ou détectable uniquement par examen, comme c'est le cas dans les premiers stades du cancer du côlon. Mais ces saignements, surtout en cas de cancer avancé, peuvent être plus importants, plus graves, voire mortels.

Le siège du cancer détermine le site de l'hémorragie. Un cancer qui se développe sur un segment du tube digestif, quel qu'il soit, peut être à l'origine d'une présence de sang dans les selles (méléna [sang noir] ou rectorragies [sang rouge]). Un cancer de l'appareil urinaire peut induire la présence de sang dans les urines (hématurie). D'autres types de cancers peuvent saigner à l'intérieur de l'organisme. Le saignement dans les poumons peut provoquer l'émission de sang lors de la toux. [11]

I.4.3. Perte de poids et fatigue :

Souvent, les personnes atteintes de cancer subissent une perte de poids et une faiblesse (fatigue) qui s'aggrave à mesure que la maladie progresse. Certains patients perdent du poids malgré un bon appétit, tandis que d'autres souffrent de perte d'appétit voire d'anorexie ou de difficulté à avaler. Ils peuvent devenir très fins. Les personnes atteintes d'un cancer avancé sont souvent très fatiguées, La fatigue liée au cancer est beaucoup plus importante. Cette sensation d'épuisement, non liée à un effort particulier, voire d'affaiblissement général de l'organisme (asthénie) n'est pas comparable aux ressentis d'une fatigue "normale". Elle perdure même après une nuit de sommeil. [12]

I.4.4. Gonflement ganglions lymphatique :

Lorsque le cancer commence à se propager dans tout le corps, il peut initialement affecter les ganglions lymphatiques voisins, les faisant grossir. Les ganglions lymphatiques gonflés sont souvent indolores et peuvent prendre une consistance ferme ou caoutchouteuse. Ils peuvent être mobiles ou, à un stade plus avancé, ils peuvent se coller aux tissus adjacents ou entre eux. [13]

I.4.5. Symptôme respiratoire :

Le cancer peut comprimer ou bloquer les voies respiratoires dans les poumons, provoquant des difficultés respiratoires, de la toux ou une pneumonie.la gêne respiratoire, une toux ou une pneumonie. La gêne respiratoire peut également se produire lorsque le cancer provoque un important épanchement pleural, un saignement dans les poumons ou une anémie.[14]

I.4.6. Symptôme neurologique et musculaire :

Le cancer peut infiltrer ou comprimer les fibres nerveuses ou la moelle épinière et produire des symptômes neurologiques et musculaires, comme une douleur [15]dans le dos, la nuque, le long d'un nerf, de céphalées…etc [16] une faiblesse ou une altération de la sensibilité [17] une sensibilité augmentée au toucher, la perte de la sensation du toucher, du froid, du chaud ou de la douleur [18] Lorsque le cancer se développe dans le cerveau, les symptômes sont difficiles à cerner, mais comprennent de la confusion, des étourdissements, des maux de tête, des nausées, une vision floue et des convulsions.

I.5. LE TRAITEMENT :

I.5.1. Le but :

Selon le cas, les traitements ont pour objectif de :

- Supprimer la tumeur ou les métastases.
- Réduire le risque de récidive.
- Ralentir le développement de la tumeur ou des métastases.
- Prévenir et traiter les symptômes et les complications engendrés par la maladie et les traitements pour assurer la meilleure qualité de vie possible.[19]

I.5.2.Le choix :

Le choix de traitement est adapté en fonction de chaque situation, Les médecins décident du meilleur traitement pour le cancer d'une personne en se basant sur :

- La chance de guérison (le cancer est éliminé et ne revient pas)
- L'espérance de vie la plus longue, la possibilité de mener une vie confortable

- ❖ L'efficacité du traitement sur les symptômes
- ❖ Les effets secondaires du traitement
- ❖ Les souhaits de la personne concernant le traitement

Le médecin doit discuter avec la personne des avantages et des inconvénients des différents traitements. Les protocoles de traitement sont des méthodes standardisées développées par les médecins pour traiter certains types de cancer. Ces programmes ont été testés avec rigueur au cours d'études de recherche appelées essais cliniques. Dans les essais cliniques, les médecins comparent les nouveaux médicaments ou traitements avec des traitements habituels pour savoir si les nouveaux traitements sont plus efficaces. De nombreuses personnes atteintes de cancer ont la possibilité de participer à un essai clinique. Le médecin peut expliquer les avantages et les inconvénients d'une participation à un essai clinique. [20]

I.5.3.Les types :

I.5.3.1. La chirurgie :

La chirurgie de cancer est le principal traitement de cancer, Elle est utilisée dans environ 80 % des cas [21] uniquement quand Le cancer ne s'est pas propagé (métastasé) ailleurs. L'organisme de la personne est suffisamment fort pour subir la chirurgie [22]. Ce traitement consiste à opérer le malade à fin d'éliminer les tumeurs cancéreuses solides Cette stratégie a pour objectif :

- ❖ **Curative :** lorsqu'elle permet de retirer 100 % des cellules tumorales
- ❖ **Diagnostique :** le tissu retiré est analysé pour préciser la nature de la lésion et faciliter le choix de traitements curatifs complémentaires
- ❖ **Palliative :** pour soulager la douleur liée à la taille de la tumeur, pour faciliter le fonctionnement de l'organe atteint… [23]

I.5.3.2. La chimiothérapie :

La chimiothérapie (appelée aussi chimio) est un traitement du cancer [24] qui repose sur l'administration d'un ou plusieurs médicament, dit « cytotoxique » [25] Le choix des médicaments est adapté en fonction de chaque situation ,chaque cancer est particulier et nécessite un traitement approprié.[26] Elle vise à combattre et éliminer les cellules tumorales présentés dans l'organisme quel que soit ses localisations, Y compris celles qui n'ont pas été repérées par les examens d'imagerie. La chimiothérapie agit soit en les détruisant directement, soit en les empêchant de se multiplier. [27]

Cette stratégie peut être proposé comme seul traitement ou être associé aux autres traitements [28], elle est proposée généralement dans les cas suivants :

❖ **Avant une chirurgie :** On parle alors de chimiothérapie néoadjuvante. Cette chimiothérapie a pour but de diminuer la taille de la tumeur.

❖ **Après une chirurgie complète de la tumeur** : c'est-à-dire lorsque le chirurgien a enlevé toutes les cellules cancéreuses visibles. C'est alors une chimiothérapie adjuvante. La chimiothérapie a pour but de diminuer les risques de récidive locale ou à distance. La chimiothérapie complète alors la chirurgie.

❖ **Pour traiter des métastases :** c'est-à-dire des cellules cancéreuses qui se sont propagées dans d'autres parties du corps. On dit que c'est une chimiothérapie métastatique.[29]

I.5.3.3. La radiothérapie :

La radiothérapie consiste à délivrer localement des rayons, ou radiations ionisantes, qui provoquent des dégâts majeurs au niveau de l'ADN. Comme les cellules cancéreuses ne parviennent pas à réparer ces lésions aussi bien que les cellules saines, elles ne peuvent plus se multiplier et/ou meurent.[30] La radiothérapie peut être le seul traitement mis en place pour traiter un cancer, mais elle est généralement couplée à une ou plusieurs autres thérapies.

❖ **Radiothérapie néo-adjuvante :** radiothérapie administrée avant l'opération pour réduire la taille d'une tumeur afin qu'elle soit plus facile à retirer par intervention chirurgicale.

❖ **Radiothérapie adjuvante :** après l'ablation chirurgicale d'une tumeur, la radiothérapie peut être utilisée pour éliminer les éventuelles cellules cancéreuses subsistant localement.

❖ **Chimio radiothérapie :** la radiothérapie peut être pratiquée en même temps que la chimiothérapie pour en renforcer l'effet.[31] qui rend les cellules cancéreuses plus sensibles aux rayons.[32], cette stratégie est administrée sous trois défirent formes :

1. ***La radiothérapie externe*** : les rayons sont émis par une machine appelée accélérateur linéaire de particules, située à proximité du patient et dirigée vers la région du corps à traiter. Ces rayons traversent la peau pour atteindre la zone à traiter.

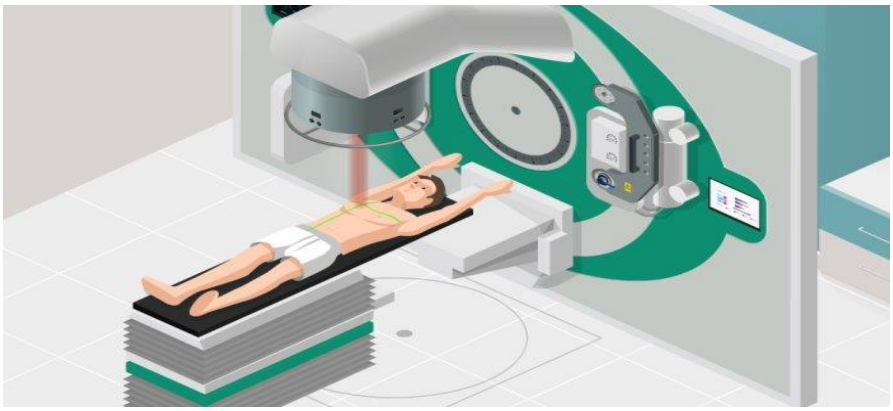


Figure I.1 : La radiothérapie externe

2. ***La curiethérapie*** : des sources radioactives sont mises en contact direct avec la zone à traiter à l'intérieur du corps afin d'épargner les tissus sains environnants. C'est un traitement le plus souvent dédié à des cancers localisés avec des indications spécifiques.

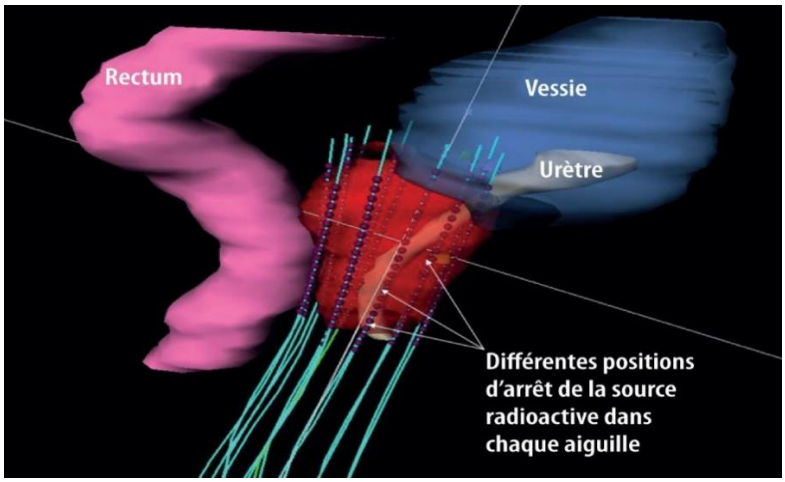


Figure I.2 : La curiethérapie

3. ***La radiothérapie métabolique :*** les sources radioactives sont administrées par voie orale (boisson ou capsule) ou par injection intraveineuse. Ces sources radioactives se fixent ensuite sur les cellules cancéreuses pour les détruire.[33]



Figure I.3 : La radiothérapie métabolique

I.5.3.4. Les thérapies ciblées :

Les thérapies ciblées constituent une autre famille de traitements du cancer le plus souvent disponibles par voie orale. En ciblant spécifiquement certaines molécules de l'organisme, elles bloquent des mécanismes qui sont indispensables à la prolifération des cellules cancéreuses ou, plus globalement, au développement de la tumeur.[34]

Les médicaments ciblés visent des molécules à l'intérieur ou à la surface des cellules cancéreuses pour ralentir la croissance du cancer [35], Selon les cancers, les thérapies ciblées peuvent être prescrites seules, en association entre elles ou avec d'autres traitements.[36]

I.5.3.5. L'hormonothérapie :

La croissance de certains cancers est favorisée par les hormones sexuelles produites par l'organisme : ainsi, certaines tumeurs du sein ou de l'utérus croissent sous l'action des œstrogènes ou de la progestérone, et certains cancers de la prostate progressent sous l'action de la testostérone. Les hormones sexuelles agissent sur les cellules tumorales en se fixant à leur surface au niveau de récepteurs spécifiques. Les médicaments d'hormonothérapie bloquent la synthèse de ces hormones ou empêchent leur fixation aux récepteurs. Ils sont administrés essentiellement par voie orale.

Un traitement par hormonothérapie peut être prescrit lorsque le cancer est dit « hormonodépendant », c'est-à-dire que sa croissance dépend de l'action d'hormones sexuelles. Avant le traitement, les médecins recherchent donc la présence de récepteurs aux hormones sur les cellules tumorales. Cette vérification passe par une analyse moléculaire, conduite à partir d'échantillons de la tumeur. [37]

I.5.3.6. L'immunothérapie :

L'immunothérapie regroupe un ensemble de stratégies visant à mobiliser ou à renforcer les défenses immunitaires des patients de manière à ce qu'elles s'attaquent aux cellules tumorales.

A. L'immunothérapie « spécifique »

Consiste à bloquer spécifiquement des protéines à la surface des cellules cancéreuses ou dans leur microenvironnement pour freiner la croissance tumorale.

B. L'immunothérapie active :

vise à stimuler le système immunitaire du patient pour qu'il réagisse contre les cellules tumorales. [38]

I.6. CONCLUSION

Le cancer est une maladie complexe et multifactorielle ; Bien qu'il reste l'un des défis les plus importants en matière de santé publique, des avancées significatives ont été réalisées en matière de prévention, de détection précoce et de traitement du cancer. Il est important de continuer à investir dans la recherche sur le cancer pour améliorer notre compréhension de la maladie et développer de nouveaux traitements et stratégies de prévention ; surtout qu'il existe nombreux types de cancer, Le cancer de la peau est l'une des formes les plus courantes de cancer dans le monde, et il peut être prévenu en prenant des mesures simples pour réduire l'exposition au soleil et en surveillant régulièrement la peau pour détecter tout changement suspect. [39]

LE CANCER DE LA PEAU

II.1. DEFINITION :

Un cancer de la peau se manifeste par la prolifération incontrôlée et anormale de cellule de la peau et plus précisément de l'épiderme, la couche externe de la peau. Cette multiplication pathologique est causée par les lésions non réparés de l'ADN, qui entraine des mutations de ces cellules, venant à former une tumeur maligne, Il peut être traité avec succès s'il est diagnostiqué tôt. [40]

II.2. LES TYPES :

II.2.1. Le carcinome basocellulaire :

C'est le type le plus commun de cancer de la peau. Il se développe dans les cellules basales de l'épiderme, la couche supérieure de la peau. Le carcinome basocellulaire est généralement causé par une exposition au soleil excessif et est plus fréquent chez les personnes de plus de 50 ans. Il a un taux de guérison élevé et est rarement fatal.

II.2.2. Le carcinome épidermoïde :

Également appelé cancer de la peau à cellules squameuses, se développe dans les cellules squameuses de l'épiderme. Il est également souvent causé par une exposition excessive au soleil et peut se produire sur n'importe quelle partie de la peau, mais il est plus fréquent sur les zones exposées au soleil. Comme le carcinome basocellulaire, le carcinome épidermoïde a un taux de guérison élevé s'il est diagnostiqué et traité tôt.

II.2.3. Le mélanome :

C'est un type plus rare de cancer de la peau, mais il est également le plus dangereux. Le mélanome se développe dans les cellules qui produisent le pigment de la peau (mélanocytes) et peut se propager rapidement à d'autres parties du corps. Le mélanome est souvent causé par une exposition au soleil, mais il peut également être causé par d'autres facteurs tels que des antécédents familiaux de mélanome. [41].

La protéine kinase BRAF est une enzyme qui joue un rôle crucial dans la voie de signalisation RAS/MAPK (mitogen-activated protein kinase), impliquée dans la régulation de la croissance cellulaire, la différenciation et la survie cellulaire. BRAF est une kinase, ce qui signifie qu'elle est responsable de la phosphorylation d'autres protéines, modifiant ainsi leur activité et leur fonction. La protéine BRAF fait partie d'une famille de protéines comprenant

également les kinases ARAF et CRAF (ou RAF-1). Lorsque des facteurs de croissance ou d'autres signaux stimulent les cellules, la voie RAS/MAPK est activée pour transmettre ces signaux à l'intérieur de la cellule. Dans cette voie, BRAF est activé en aval de la protéine RAS et active ensuite une cascade de phosphorylations, aboutissant à l'activation des protéines MAP kinase (ERK1/2), qui régulent divers processus cellulaires.

Les mutations de BRAF sont fréquemment observées dans de nombreux cancers, y compris le mélanome cutané, le cancer colorectal et le cancer papillaire de la thyroïde. La mutation la plus courante est la mutation V600E, qui entraîne une activation constante de la voie RAS/MAPK, indépendamment des signaux de croissance. En conséquence, les inhibiteurs de BRAF ont été développés pour cibler spécifiquement les cellules cancéreuses avec des mutations BRAF, offrant ainsi des options thérapeutiques dans le traitement de certains types de cancers.

L'enzyme V600E_BRAF est une mutation spécifique de la protéine kinase BRAF, qui est impliquée dans la voie de signalisation RAS/MAPK. Cette mutation est souvent présente dans les mélanomes cutanés, y compris la lignée cellulaire

Le rôle de l'enzyme V600E_BRAF dans le mélanome SK-MEL-2 est de stimuler de manière constitutive la voie de signalisation RAS/MAPK. Normalement, cette voie de signalisation est activée de manière temporaire en réponse à des signaux de croissance cellulaire. Cependant, la mutation V600E_BRAF entraîne une activation constante de cette voie, indépendamment des signaux environnants. Cette activation constitutive de la voie RAS/MAPK par l'enzyme V600E_BRAF conduit à une prolifération cellulaire excessive, une survie cellulaire accrue et une capacité invasive des cellules du mélanome.

En conséquence, la mutation V600E_BRAF est considérée comme un facteur clé dans le développement et la progression du mélanome SK-MEL-2, et elle représente une cible thérapeutique importante pour le traitement de cette maladie. [42]

II.2.3.1. les causes et les facteurs :

Les causes exactes du cancer de la peau ne sont pas entièrement comprises, mais plusieurs facteurs de risque ont été identifiés. Les principales causes sont l4 exposition aux rayons ultraviolets (UV) du soleil et des lits de bronzage, ainsi que des facteurs génétiques Et environnementaux. Les facteurs de risque pour le cancer de la peau comprennent :

A. L'exposition aux rayons UV : Une exposition excessive aux rayons UV du soleil ou des lits de bronzage est le principal facteur de risque de tous les types de cancer de la peau.

Les rayons UV endommagent l'ADN des cellules cutanées, ce qui peut entraîner une croissance anormale et un cancer de la peau.

B. Antécédents familiaux de cancer de la peau : Les personnes ayant des antécédents familiaux de cancer de la peau ont un risque plus élevé de développer la maladie. Certains types de cancer de la peau, tels que le mélanome, peuvent être hérités.

C. Teint de peau clair : Les personnes ayant une peau claire sont plus susceptibles de Développer un cancer de la peau, car leur peau contient moins de mélanine, qui protège contre Les rayons UV.

D. Exposition professionnelle : Certaines professions, telles que les agriculteurs et les Travailleurs de la construction, peuvent être exposées à des niveaux élevés de rayons UV ou à Des produits chimiques carcinogènes, augmentant ainsi leur risque de cancer de la peau.

E. Maladies cutanées préexistantes : Les personnes atteintes de certaines maladies Cutanées, telles que le xeroderma pigmentosum, ont un risque accru de développer un cancer de la peau

F. Médicaments immunosuppresseurs: Les personnes qui prennent des médicaments immunosuppresseurs, tels que les transplantés d'organes, sont plus susceptibles de Développer un cancer de la peau en raison de leur système immunitaire affaibli [10].

II.2.3.2. Les stades :

Le stade du cancer de la peau fait référence à la gravité et à la progression de la maladie. Il est Déterminé par la taille de la lésion, la profondeur de l'invasion tumorale et la présence de métastases (Propagation de la tumeur à d'autres parties du corps). Les stades d'évolution du cancer de la peau Les plus couramment utilisés sont :

Stade 0 : Le cancer de la peau est limité à la couche la plus superficielle de la peau (épiderme) et ne S'est pas propagé à d'autres parties du corps. Ce stade est souvent appelé carcinome in situ.

Stade I : Le cancer de la peau mesure moins de 2 centimètres de diamètre et n'a pas encore atteint les ganglions lymphatiques ou d'autres parties du corps.

Stade II : Le cancer de la peau mesure plus de 2 centimètres de diamètre et/ou s'est propagé aux ganglions lymphatiques adjacents, mais n'a pas encore atteint d'autres parties du corps.

Stade III : Le cancer de la peau s'est propagé aux ganglions lymphatiques régionaux ou à des tissus voisins, mais n'a pas encore atteint d'autres parties du corps.

Stade IV : Le cancer de la peau s'est propagé à des organes distants du corps, tels que le cerveau, les poumons ou le foie. Ce stade est souvent appelé cancer de la peau métastatique

ou avancé Il est important de noter que les stades d'évolution peuvent varier en fonction du type de cancer de la peau [43]

II.2.3.3 Les symptômes :

Les symptômes de mélanome peuvent varier d'une personne a l'autre. Voici quelques signes et symptômes courants :

- Changement dans la taille, la forme ou la couleur d'une tache de naissance existante.
- Apparition d'une nouvelle tache de couleur brun foncé, noire ou multicolore sur la peau.
- Asymétrie : une moitié de la tache ou de la lésion est différente de l'autre.
- Bords irréguliers ou flous autour de la tache ou de la lésion.
- Évolution rapide de la taille ou de l'apparence de la tache ou de la lésion.
- Présence de plusieurs couleurs dans la tache ou la lésion, telles que du noir, du brun, du rouge, du bleu ou du blanc.
- Sensation de démangeaisons, de picotements ou de douleur au niveau de la tache ou de la lésion.
- Saignement ou formation de croûtes sur la tache ou la lésion.
- Ulcération ou formation d'une bosse au niveau de la tache ou de la lésion. [44]

II.2.3.4. Le dépistage du mélanome

Le dépistage du mélanome est une méthode de détection précoce de cette maladie avant l'apparition de symptômes ou de signes visibles. Cette méthode peut impliquer un auto-examen régulier de la peau ou un examen réalisé par un professionnel de santé, comme un dermatologue.[45]

❖ Selon l'OMS :

L'Organisation mondiale de la santé (OMS) définit le dépistage du mélanome comme un ensemble de mesures qui visent à identifier les signes précoces du cancer de la peau avant que les symptômes ne se manifestent. Ces mesures comprennent notamment :[46]

-L'auto-examen régulier de la peau pour détecter tout changement dans l'apparence de grains de beauté ou de lésions cutanées.

- L'examen clinique de la peau effectué par un dermatologue ou un autre professionnel de la santé pour détecter tout signe de cancer de la peau.

- La photographie de la peau pour surveiller les changements dans l'apparence de la peau au fil du temps.

- La biopsie pour confirmer un diagnostic de mélanome.[47]

II.2.3.5. Le diagnostic du mélanome

Le diagnostic du mélanome est souvent effectué par un dermatologue ou un médecin de soins primaires. Voici les étapes courantes du diagnostic :

A. Examen visuel : le médecin examine la peau à la recherche de lésions cutanées qui pourraient être un signe de mélanome

B. Biopsie : si le médecin suspecte un mélanome, il peut prélever un échantillon de tissu de la lésion cutanée pour une biopsie. Le tissu prélevé est examiné au microscope pour déterminer s'il s'agit d'un mélanome et, si oui, quel type.

C. Imagerie : dans certains cas, une imagerie comme une radiographie, une échographie ou une tomodensitométrie (TDM) peut être utilisée pour aider à déterminer si le mélanome s'est propagé à d'autres parties du corps.

Il est important de noter que tous les cancers de la peau ne nécessitent pas une biopsie, et que les examens d'imagerie ne sont généralement pas nécessaires pour les stades précoces de la maladie. Le diagnostic précoce est crucial pour un traitement efficace du mélanome. [48]

II.2.3.6. Le traitement du mélanome

Les traitements pour le mélanome varient selon le type et le stade du cancer. Voici les principales options de traitement :

- ✓ **Chirurgie :** la chirurgie est le traitement le plus courant pour le mélanome. Elle consiste à retirer la tumeur et une marge de tissu sain autour de la tumeur. Selon la taille et la profondeur de la tumeur, la chirurgie peut être réalisée sous anesthésie locale ou générale.
- ✓ **Radiothérapie :** la radiothérapie utilise des rayonnements pour détruire les cellules cancéreuses. Elle est généralement utilisée pour traiter les mélanomes qui sont difficiles à retirer chirurgicalement ou pour les patients qui ne peuvent pas subir une intervention chirurgicale.
- ✓ **Chimiothérapie :** la chimiothérapie consiste à utiliser des médicaments pour tuer les cellules cancéreuses. Elle est généralement utilisée pour traiter les cancers de la peau avancés qui ont commencé à se propager à d'autres parties du corps.
- ✓ **Immunothérapie :** l'immunothérapie utilise le système immunitaire du corps pour combattre le cancer. Elle est souvent utilisée pour traiter les cancers de la peau avancés.

- ✓ **Thérapies ciblées :** les thérapies ciblées sont des médicaments qui sont conçus pour cibler des protéines spécifiques dans les cellules cancéreuses. Elles sont souvent utilisées pour traiter les cancers de la peau avancés.[49]

Il est important de noter que le traitement du mélanome est souvent très efficace, surtout lorsqu'il est détecté et traité tôt. Si vous avez un mélanome , il est recommandé de discuter avec votre médecin des options de traitement les plus appropriées pour votre situation.[50]

II.3. CONCLUSION

Le mélanome est une maladie qui se développe lorsque les cellules de la peau commencent à croître de manière anormale et incontrôlable. Il existe trois principaux types de cancer de la peau : le carcinome basocellulaire, le carcinome épidermoïde et le mélanome. Le principal facteur de risque est l'exposition aux rayons UV, notamment ceux du soleil, ainsi que les antécédents familiaux et la couleur de la peau. Le traitement dépend du type et de la gravité du cancer, allant de la chirurgie à la radiothérapie ou à la chimiothérapie. Une prévention efficace consiste à protéger la peau du soleil avec des crèmes solaires et à éviter les expositions prolongées aux UV. [51]

CONCEPTION DES MEDICAMENTS ASSISTES PAR ORDINATEUR

III .1.Approche Basé sur la structure

III .1.1. GENERALITES

Le développement d'un nouveau médicament pour des maladies nécessite le plus souvent des investissements considérables en plus leurs processus est long. Même si les dépenses consacrées à la découverte des médicaments ont été multipliées par plusieurs fois, le nombre de nouveaux médicaments approuvés restent considérablement réduit. Suite au développement de l'informatique et de l'existence sur le marché de logiciels professionnels adapté moins couteux et rapide il y a une croissance exceptionnelle de l'intérêt pour les relations quantitative structure- activité (QSAR), qui utilisent des méthodes d'analyses multidimensionnelles afin de modéliser des activité en fonction des paramètres structuraux moléculaire appelés descripteurs[17]

III.1.2. DEFINITION

Une relation quantitative structure à activité (en anglais : quantitative structure activity relationship ou QSAR , parfois désignée sous le nom de relation quantitative structure à propriété- en anglais : quantitative structure property relationship ou QSPR) est le procédé par lequel une structure chimique est corrélée avec un effet bien déterminé comme l'activité biologique ou la réactivité chimique. Ainsi, l'activité biologique peut être exprimée de manière quantitative, comme pour la concentration de substance nécessaire pour obtenir une certaine réponse biologique.

De plus lorsque les propriétés ou structure physico-chimiques sont exprimées par des chiffres, on peut proposer une relation mathématique, ou relation quantitative structure à activité entre les deux. L'expression mathématique obtenue peut alors être utilisée comme moyen prédictif de la réponse biologique pour des structures similaires. La QSAR la plus commune est de la forme : Activité= f (propriétés physico-chimiques et /ou structurales) [52]

III.1.3. Historique

Il y a plus d'un siècle et demi, des scientifiques ont observé que des propriétés physicochimiques des alcanes augmentent quand le nombre d'atomes de carbone augmentent. Ils ont observé également une diminution de la solubilité dans l'eau des alcools avec l'augmentation du nombre d'atomes de carbone, cela est considéré depuis, comme laPremière

formulation générale en QSAR. Cinq ans après, en 1868, Crum-Brown et fraser [53] postulèrent que l'activité biologique d'une molécule est une fonction de sa constitution chimique selon la forme suivant Ø=f(c)Quelques décennies plus tard, en 1893 Richet a découvert que la toxicité des composés organique a un rapport avec leur solubilité dans l'eau est définie par : αØ=f (αc).

A la fin de 19^{éme} siècle, Meyer en 1899 et Overton en 1901, ont observé une relation linéaire entre l'activité des narcotiques et leurs coefficients de partage huile-eau.Six ans après, en 1907 Fühner et Neubauer [54] ont montré pour une série de narcotique homologue, que l'activité augmente en fonction de la progression géométrique de la série de composés, cela montrait l'importance de la contribution d'additivité de groupement fonctionnel pour l'activité biologique.

En 1962, Hansen [55] a montré l'existence d'une corrélation entre la toxicité des acides benzoïques substitués et les constantes électronique (σ) des substituant.

L'année 1964 est considérée comme le début des méthodes QSAR modernes. Hansh et Fujita [8] ont établi les premières corrélations entre les propriétés physico-chimiques (log p, pkA, Approches de base, développement, validation et application des méthodes QSAR/QSPR 21 paramètres stérique et électronique) et l'activité Biologique (activité enzymatique, pharmacologique).

L'approche QSAR résulte de la combinaison de différents paramètres physico-chimique de façon linéaire additive (équation 1) ; log (1/c) est le logarithme de l'inverse de la dose molaire qui produit ou empêche une certaine réponse biologique, log p est le logarithme du coefficient départage de n-octanol/eau). D'autres méthodes utilisent un paramètre connu sous le nom de paramètre de lipophilecalculée (equation 2), il est employé à la place des valeurs mesurées de log p, et la formulation d'une équation parabolique pour la description quantitative non-linéaire des rapports lipophile-activité (Equation.3)

$$\text{Log}\frac{1}{c} = a \log p + b_0 + \ldots + \text{const} \qquad \text{(Eq.01)}$$

$$\pi\chi = \log p_r x - \log p_r h \qquad \text{(Eq.02)}$$

$$\text{Log}\frac{1}{c} = a \log p + b \log p + c\sigma + \ldots + \text{const} \qquad \text{(Eq.03)}$$

Sept ans plus tard, Hansh et Lien ont réalisé une étude QSAR sur différentes familles biologiques. Ils ont observé que quel que soit la famille utilisée, l'activité antifongique

Dépend du coefficient de partage Eau-Octanol, expérimental ou calculé [56].Ces études ont été extrapolées aux techniques séparatives en corrélant les propriétés physico-chimiques des analyses avec les temps de rétention notés QSRR [57]..

Récemment, des méthodes basées sur des structures 3D comme l'étude CoMFA (Comparative Molecular Field Analysis) et CoMSIA (Comparative MolecularSimilarity Indices Analysis) [58] permettent de traiter les relations structure-activité en trois dimensions, 3D-QSAR/QSPR.

III.1.4. OBJECTIF

La modélisation QSAR (Quantitative Structure Activity Relationship) en chémoinformatique, vise à construire des modèles statistiques prédictifs reliant la structure chimique de composé à leurs propriétés physico-chimiques ou biologique. Ces modèles recherchent et exploitent des régularités dans les données stockées dans les bases de données en chimie. [59] Il importe donc de maîtriser la façon dont on stock et recherche de l'information sur des molécules.

Les approches chimio-informatique sont très utilisées dans l'industrie chimique pour modéliser les propriétés physico-chimique des molécules et des matériaux, ainsi que dans l'industrie pharmaceutique pour effectuer des criblages virtuels ou prédire des propriétés pharmacodynamiques et pharmacocinétiques [60].

III.1.5. PRINCIPE

L'étude QSAR/QSPR consiste à définir une relation mathématique entre une activité biologique, ou une propriété mesurée pour une série de composé similaire dans les mêmes conditions expérimentales, avec des descripteurs moléculaires à l'aide des méthodes statistique. L'objectif de ces études est d'analyser les données structurales afin de détecter les facteurs déterminants pour l'activité ou la propriété étudiée. Pour ce faire, différents types de méthodes statistiques peuvent être employées. L'expression mathématique obtenue est utilisé comme un moyen prédictif de l'activité/propriété étudiée pour de nouvelles molécules ou bien des molécules pour lesquelles les données expérimentales ne sont pas disponibles. [61]

III.1.6. STRATEGIE GLOBALE

Pour développer un modèle, il faut commencer par la recherche du maximum possible des données expérimentales fiables. Ensuite, le développement d'une série de descripteurs qui caractérisent les structures moléculaires des composés de la base de données en vue de les relier à l'activité/propriété expérimentale étudiée. Une fois développé, le modèle doit être

validé en termes de corrélation. L'influence des composés du jeu d'entrainement sur le modèle est estimée par des méthodes de validation interne. Pour estimer le pouvoir prédictif du modèle, il est nécessaire de disposer de données expérimentales supplémentaires, c'est la validation externe [57]

III.1.7. MODELES QSAR/QSPR

Au cours des années passées, les relations quantitatives structure activité, quantitative structure propriété (QSAR/QSPR) sont devenu un puissant outil théorique, à la place de la mécanique quantique, pour la description et la prédiction des propriétés des systèmes moléculaires complexes dans différentes circonstances. L'approche QSAR procède à l'hypothèse d'une relation dépendante entre n'importe quelle propriété physique, affinité chimique, ou activité biologique d'un composé chimique et sa structure moléculaire. Cette correspondance univoque peut être représentée par la composition chimique, la connectivité des atomes, la surface d'énergie potentielle, et la fonction d'onde électronique d'un composé.

Différents descripteurs moléculaires physico-chimiques représentant la structure peuvent être déterminés empiriquement où en utilisant des méthodes théoriques et computationnelles de différents niveaux de complexité. Il a fallu mettre en évidence que la connaissance de la constitution chimique exacte et/ou de la structure moléculaire tridimensionnelle des composés chimiques étudies est une condition essentielle à l'application de l'approche QSAR. La réalisation de l'approche QSAR, dépend de l'utilisation appropriée des descripteurs moléculaires [57]

III.1.8. LES DESCRIPTEURS MOLECULAIRES

Les descripteurs moléculaires sont les produits finaux de procédures mathématiques transformant les informations chimiques codées dans une structure moléculaire en un représentant numérique. Un descripteur moléculaire est un paramètre (une valeur numérique) propre à une structure chimique donnée. Ces valeurs peuvent être obtenues expérimentalement ou calculées à partir de la structure de la molécule.

Les descripteurs calculés, permettent d'effectuer des prédictions sans avoir à synthétiser les molécules, ce qui est l'un des objectifs de la modélisation moléculaire. Les descripteurs moléculaires jouent un rôle fondamental dans les études de la relation quantitative structure activité/propriété. Ils sont utilisés en tant que variables indépendantes pour prédire une variable dépendante (activité ou propriété). L'utilisation des descripteurs moléculaires dans le développement de modèles QSAR/QSPR n'est pas une tâche aisée. Tout

d'abord, un très grand nombre de descripteurs moléculaires, de différentes complexités et de conceptions diverses a été introduit au cours des dernières années.

Ensuite, pendant ce temps, aucune règle stricte n'a été établie pour la sélection de descripteurs adaptés parmi le grand nombre de descripteurs disponibles. Ce choix a souvent été basé sur l'intuition chimique des chercheurs, ou en se pliant à la tradition [62].

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

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 [63].

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 électro-topologique et certains autres invariants [64]

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 [65]. 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 [66].

III.1.9. METHODOLOGIE GENERALE

III.1.9.1. Base de données

Les études QSAR/QSPR étant avant tout des analyses statistiques, l'une des étapes absolument capitales est celle de la sélection des données initiales. En effet, un modèle QSAR/QSPR est très dépendant de données expérimentales de référence. Le choix de la base de données est décisif dans le développement de tel modèle. Dans la plupart des cas, les données expérimentales sont issues de la littérature. Pour être de qualité, une base de données doit être composée de données expérimentales aussi fiables que possible obtenues en suivant un protocole unique puisque les erreurs sur celles-ci se propageront dans le modèle final.

Il y'a plusieurs éléments à vérifier dans les étapes de nettoyage d'une base de données. Il faut tout d'abord vérifier que les structures sont correctes d'un point de vue chimique (règle de valence, …), des structures erronées entrainent la génération de mauvais descripteurs et donc de mauvais modèles.

III.1.9.2. Descripteurs moléculaires

Avant toute modélisation, il est nécessaire de calculer ou de mesurer un grand nombre de descripteurs différents, car les mécanismes qui déterminent l'activité d'une molécule ou une de ses propriétés sont fréquemment mal connus. Il faut ensuite sélectionner parmi ces descripteurs ceux qui sont les plus pertinentes pour la modélisation. En fait, chaque

modélisation repose sur le nombre de descripteurs pertinents utilisés par ce dernier. Une règle empirique inspirait de la littérature, selon laquelle le nombre maximal de descripteurs utilisés devrait idéalement être de l'ordre du cinquième du nombre de composés dans le jeu d'apprentissage [67]

Les descripteurs moléculaires sont généralement classés en trois catégories : les descripteurs physicochimiques, topologiques et électroniques. Ces descripteurs sont caractéristiques de la structure bidimensionnelle ou tridimensionnelle de la molécule. Parmi les descripteurs les plus utilisées dans notre étude QSAR/QSPR]

III.1.9.3. Sélection des descripteurs

Lorsqu'un grand nombre de descripteurs est introduit, certains d'entre eux peuvent contenir des informations redondantes, entraînant un problème de colinéarité. De plus, les descripteurs calculés n'ont pas nécessairement une influence sur l'activité à modéliser. Les descripteurs employés doivent être, autant que possible, porteurs de sens et facilement interprétables d'un point de vue chimique, aussi les modèles QSAR/QSPR devraient être simples, transparents et compréhensibles d'un point de vue phénoménologique. Il est nécessaire donc d'éliminer les descripteurs dont l'influence est inférieure à celle de l'erreur, et de sélectionner uniquement les plus pertinents d'entre eux.

De manière générale, pour qu'un descripteur soit retenu, il faut que son retrait entraîne une décroissance significative de performance du modèle. Il faut donc être attentif à ne pas perdre de l'information essentielle. [68] Finalement, le sens chimique des descripteurs utilisés doit être pris en considération puisque, plus les descripteurs sont reliés chimiquement au phénomène, plus la probabilité de faire intervenir des descripteurs par hasard est réduite. La procédure de sélection et de réduction des descripteurs peut être effectuée en deux étapes : Sélection objective¬ Sélection subjective¬

III.9.3.1 Sélection objective

Elle consiste en la sélection des variables en réduisant le nombre de descripteurs sans faire participer la variable dépendante (l'activité biologique). La première étape de cette procédure consiste à exclure tous les descripteurs ayant un pourcentage élevé de valeurs identiques pour l'ensemble des composés (variance non significative). Cela permet de s'assurer que de tels descripteurs ne sont pas inclus par chance dans le modèle final. De même, lorsque deux descripteurs sont fortement corrélés et leur combinaison possède un

coefficient de détermination supérieur au seuil requit (R2>0,95), seul celui présentant la plus grande variance est retenu [69].

Non seulement ces procédures évitent l'introduction, dans le modèle, de descripteurs inappropriés mais elles rendent la suite de l'analyse moins coûteuse en terme de temps de calcul, puisqu'elles réduisent le nombre de descripteurs restant à traiter

III.9.3.2 Sélection subjective

A .Introduction progressive

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

B. Elimination progressive

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

C. Sélection pas à pas

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

III.1.9.4. Division de la base de données

Il existe plusieurs approches possibles pour la sélection de la série d'apprentissage sur laquelle on établit un modèle QSAR/QSPR et la série de test pour tester le pouvoir prédictif de ce modèle d'une famille des composés [70]. L'approche la plus utilisée par les praticiens de QSAR est la division aléatoire qui consiste à diviser la base de données par simple sélection aléatoire.

III.1.9.5. Méthodes d'analyse de données

L'élaboration d'un modèle QSAR/QSPR a besoin d'une méthode d'analyse de données qui permet de quantifier la relation qui existe entre la propriété/activité et la structure (descripteurs). Il existe plusieurs méthodes pour construire un modèle et analyser les données statistiques de ce dernier, certaines sont linéaires telles que la régression linéaire multiple (MLR), la régression aux moindres carrés partiels (PLS), d'autres sont non linéaires comme la régression non-linéaire multiple (MNLR), les réseaux de neurones artificiels (RNA)… ces

méthodes sont disponibles dans des logiciels tels que, Excel, Systat, Exstat, Minitab, Statistica, SPSS, etc.….

III.1.9.6. Interprétation et validation d'un modèle

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. Sa robustesse, c'est-à-dire l'influence des composés de la série d'apprentissage sur le modèle, est estimée par des méthodes de validation interne. Afin d'estimer son pouvoir prédictif, des données expérimentales supplémentaires sont nécessaires afin de déterminer la capacité du modèle à prédire ces valeurs c'est ce que l'on appelle validation externe. Enfin, il est important de savoir quel type de molécules utilisées avec quel modèle. On parle alors de domaine d'applicabilité [71] .

III.9.6.1. Validation interne

La validation interne d'un modèle QSAR/QSPR a été réalisée en utilisant la validation croisée LOO (Leave-One-Out) ou LMO (Leave-Many-Out) qui est quantifiée par le coefficient R2. Ce processus consiste à extraire un certain nombre k de molécules du jeu initial à N molécules et à construire un nouveau modèle avec les (N-k) molécules restantes à l'aide des descripteurs choisis (seules les constantes de la régression changent). Ce processus est ensuite réitéré pour retirer et prédire les valeurs de toutes les molécules de la série d'apprentissage. En fonction du nombre de molécules retirées à chaque itération, on parlera de Leave-One-Out (LOO) ou de Leave-Many-Out (LMO) selon qu'une ou plusieurs molécules est (sont) retirée(s) [67] Cependant, la validation interne est insuffisante pour étudier le pouvoir prédictif d'un modèle. Pour cette raison la validation externe du modèle est devenue une norme et une partie obligatoire dans la modélisation QSAR/QSPR [55]

III.9.6.2. Validation externe

Cette méthode consiste à prédire la propriété/activité d'une série de molécules appelée généralement série de test qui ne sont pas dans la série de développement du modèle, cette validation est caractérisée par le paramètre R2. Récemment plusieurs études ont montré l'insuffisance du paramètre R 2 pour vérifier le pouvoir prédictif des modèles QSAR/QSPR. 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» [71]

III.1.9.7. Domaine d'applicabilité

Un modèle QSAR/QSPR 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 [72]. 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 [73].

III.2.Approche Basé sur le ligand

III.2.1.DEFINITION :

III.2.1.1. ADMET

ADME est un acronyme qui représente les étapes clés de la pharmacocinétique, c'est-à-dire l'étude de l'absorption, de la distribution, du métabolisme et de l'excrétion des médicaments dans l'organisme. Voici la signification de chaque lettre :

Absorption : Il s'agit du processus par lequel un médicament pénètre dans l'organisme, en passant généralement par les voies orales, intraveineuse, intramusculaire, transdermique ou inhalée. L'absorption détermine la quantité de médicament disponible pour être distribuée à travers l'organisme.

Distribution : Cela fait référence à la répartition du médicament dans les différents tissus de l'organisme, après son absorption. La distribution est influencée par plusieurs facteurs tels que la solubilité du médicament, le taux de liaison aux protéines plasmatiques, la taille moléculaire, etc.

Métabolisme : Il s'agit du processus de transformation du médicament par des enzymes hépatiques ou d'autres tissus en métabolites qui peuvent être éliminés par les reins ou la bile. Le métabolisme peut avoir un effet important sur l'efficacité et la sécurité du médicament.

Elimination : C'est le processus par lequel le médicament ou ses métabolites quittent l'organisme, principalement par les reins ou la bile. L'excrétion est importante pour maintenir des niveaux thérapeutiques optimaux du médicament et éviter les effets toxiques.

Toxicité : La toxicité se réfère à la capacité d'une substance à causer des effets néfastes sur les Organismes vivants. Une substance toxique peut entraîner des dommages ou des

altérations fonctionnelles dans les cellules, les tissus, les organes ou les systèmes biologiques. [74]

III.2.1.2.Docking

Le Docking, ou "Docking moléculaire", est une méthode de modélisation informatique qui permet de prédire comment deux molécules se lient ou interagissent entre elles, généralement une petite molécule (comme un médicament) et une grande molécule (comme une protéine). Le but du Docking est de prédire la géométrie et l'énergie de liaison des complexes moléculaires formés par ces molécules.

Le Docking est un outil important dans la découverte de médicaments, car il permet de prédire comment un composé chimique va interagir avec une cible biologique, comme une protéine impliquée dans une maladie. Cette information est ensuite utilisée pour concevoir de nouveaux composés chimiques qui ont une meilleure affinité et une meilleure spécificité pour leur cible, ce qui peut conduire à de nouveaux médicaments plus efficaces et plus sûr [75].

III.2.2. Le principe du Docking :

Le principe du Docking est de trouver la conformation la plus favorable ou la plus stable pour la liaison entre le ligand et le récepteur, en utilisant des informations sur leurs structures moléculaires tridimensionnelles et leurs énergies d'interaction. Le Docking est utilisé en chimie médicinale pour identifier de nouveaux composés médicamenteux ou pour optimiser des composés existants en les faisant interagir avec leur cible moléculaire.[65]

III.2.3. Les étapes de Docking :

Les étapes générales du processus de Docking :

1) *Préparation des structures moléculaires :* Les structures moléculaires de la petite molécule (appelée ligand) et de la grande molécule (appelée récepteur) sont préparées en utilisant des techniques telles que la cristallographie aux rayons X, la RMN ou la modélisation informatique.
2) *Recherche de l'espace de recherche :* L'espace de recherche est défini en sélectionnant les zones de la surface du récepteur où le ligand peut se lier. Cela peut être basé sur des informations expérimentales ou des simulations de dynamique moléculaire.
3) *Recherche des poses :* Les poses possibles du ligand sont explorées en utilisant des algorithmes de recherche de conformation, tels que Monte Carlo ou algorithmes génétiques.

4) *Évaluation des poses :* Chaque pose est évaluée pour sa faisabilité géométrique et pour son énergie, qui est calculée en utilisant des fonctions d'énergie telles que la force de van der Waals, les liaisons hydrogène et les interactions électrostatiques.
5) *Raffinement des poses :* Les poses les plus prometteuses sont sélectionnées pour un raffinement plus détaillé de leur conformation, en utilisant des techniques telles que la minimisation d'énergie ou la dynamique moléculaire.
6) *Analyse des résultats :* Les poses finales sont classées en fonction de leur affinité ou de leur énergie de liaison prévue. Ces informations peuvent être utilisées pour prédire l'activité biologique ou pour guider la conception de nouveaux composés [6].

La fonction de score :

La fonction de score est un élément essentiel du processus de Docking qui permet d'évaluer la qualité des différentes poses proposées pour la liaison entre un ligand et un récepteur. Le score est une mesure de l'affinité entre le ligand et le récepteur, qui est généralement calculé en utilisant une fonction d'énergie.

La fonction de score prend en compte différents termes énergétiques qui décrivent les interactions entre les atomes du ligand et du récepteur, comme les forces de van der Waals, les liaisons hydrogène, les interactions électrostatiques et les termes entropiques. En utilisant ces termes, la fonction de score calcule une énergie globale pour chaque pose du ligand dans le récepteur.

L'objectif est de trouver la pose avec la plus faible énergie, ce qui indique une affinité plus forte et donc une probabilité plus élevée que cette pose représente la conformation réelle de la liaison. Les fonctions de score sont souvent développées et optimisées en utilisant des données expérimentales sur les liaisons connues, afin d'améliorer leur capacité à prédire les liaisons potentielles.

Il existe de nombreuses fonctions de score différentes utilisées dans le Docking, chacune ayant ses avantages et ses limites en termes de précision et de temps de calcul. Les choix de la fonction de score dépendent souvent de la nature des ligands et des récepteurs, ainsi que des objectifs de la recherche [76].

III.2.4. Les types :

Les trois principaux types de Docking moléculaire sont les suivants :

Docking rigide (rigid Docking) : Dans le Docking rigide, à la fois la cible et le ligand sont considérés comme rigides, ce qui signifie qu'aucun changement conformationnel n'est autorisé

pendant la simulation. Les positions et orientations optimales du ligand sont recherchées pour maximiser l'ajustement dans le site actif de la cible. C'est le type de Docking le plus simple et le plus rapide, mais il ne tient pas compte des changements conformationnels possibles de la cible lors de l'interaction.

Docking semi-flexible (semi-flexible Docking) : Dans le Docking semi-flexible, la cible est considérée comme rigide tandis que le ligand est autorisé à subir des changements conformationnels restreints. Cela permet au ligand de s'ajuster à la cible en adoptant différentes conformations, mais la cible reste fixe. Cette approche est plus réaliste que le Docking rigide, car elle permet une certaine flexibilité du ligand, mais elle ne tient pas compte des éventuels changements conformationnels de la cible.

Docking flexible (flexible Docking) : Le Docking flexible permet à la fois à la cible et au ligand de subir des changements conformationnels lors de la simulation. Cela permet de prendre en compte les ajustements conformationnels possibles des deux molécules pendant l'interaction. Le Docking flexible est plus complexe et nécessite davantage de ressources de calcul, mais il offre une représentation plus réaliste de l'interaction ligand-cible. Cette approche est souvent utilisée lorsque des changements conformationnels importants de la cible sont attendus lors de la liaison.

Ces trois types de Docking offrent différents 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 présumée des molécules impliquées et des ressources disponibles en termes de temps de calcul et de puissance informatique [77].

III.11. CONCLUSION

La modélisation QSAR/QSPR de l'activité biologique ou des propriétés physicochimiques de molécules constitue ces dernières années, un champ de recherche très important et une phase innovante qui peut guider la synthèse de médicaments. La mise en place de modèles QSAR/QSPR n'est pas une chose aisée. Un des problèmes importants réside également dans le traitement de données en quantités importantes. Un grand nombre de descripteurs et de molécules peuvent être analysés, mais aucune règle stricte n'existe quant au choix des paramètres structuraux les plus importants parmi le jeu complet de ceux disponibles. Une bonne pratique de la modélisation QSAR/QSPR, si l'utilisation des lignes directrices recommandées de l'organisation de coopération et de développement économiques (OCDE) peut développer des bons modèles prédictifs aves des applications pratiques

démontrées dans divers domaines biologiques et chimiques qui peuvent encore renforcer son acceptabilité par la communauté scientifique [61].

SUPPORT VICTOR MACHINE

IV.1. INTRODUCTION

Les machines à vecteurs de support ou séparateurs à vaste marge (en anglais support-Victor machine, SVM) sont un ensemble de techniques d'apprentissage supervisées, destinées à résoudre des problèmes de discrimination et de régression. Les SVM sont une généralisation des classifieurs linéaires. Les SVM ont été appliqués dans de nombreux domaines (bioinformatique, recherche d'information, vision par ordinateur, finance…). [78] Selon les données, la performance des machines à vecteurs de support est de même ordre, ou même supérieure, à celle d'un réseau de neurones.

IV.2. PRINCIPE GENERALE

Les SVM peuvent être utilisés pour résoudre des problèmes de discrimination afin de pouvoir décider à quelle classe appartient un échantillon, ou de régression, c'est-à-dire permettent à prédire la valeur numérique d'une variable. La résolution de ces deux problèmes passe par la construction d'une fonction H qui à un vecteur d'entrée X ait correspondre une sortie Y. On se limite pour l'instant à un problème de discrimination à deux classes (discrimination binaire), c'est-à-dire $Y\epsilon\ \{-1,1\}$, le vecteur d'entrée x étant dans un espace X muni d'un produit scalaire. On peut prendre par exemple $X=R^N$ [79]

IV.3. BUT

Le but de ce travail est de présenter les méthodes dites SVM (Support Vector Machine), issues de la théorie de l'apprentissage statistique, pour l'identification des systèmes. La théorie de l'apprentissage sur laquelle s'appuient les SVMs permet de définir des bornes pour l'erreur du modèle sur des données futures, i.e. non disponibles lors de l'apprentissage (erreur de généralisation). La minimisation de ces bornes conduit à l'algorithme d'apprentissage SVM qui ne minimise donc pas simplement l'erreur sur les données disponibles comme la plupart des algorithmes d'apprentissage de réseaux de neurones. Les SVMs, ont été développées pour la reconnaissance de formes ou la classification. Dans ces domaines, elles se sont rapidement imposées comme l'état de l'art sur diverses applications [80] Une formulation des SVMs pour la régression, appelée SVR (Support VectorRegression), existe aussi. Cette méthode a récemment été appliquée avec succès sur un certain nombre de problèmes d'identification. Certains auteurs ont aussi proposé

des améliorations de la méthode pour son application spécifique à l'identification. Ce travail présente la méthode SVM pour la régression tout d'abord pour le cas linéaire, puis pour le cas non-linéaire dans sa forme originelle qui conduit à un problème de programmation quadratique.

IV.4. REGRESSION PAR MACHINES A VECTEUR DE SUPPORT

La régression par Machines à Vecteurs de Support (SVR) consiste à trouver la fonction f(x) qui a au plus une déviation ε par rapport aux exemples d'apprentissage (xi, yi), pour i = 1, . . ., N, et qui est le plus petit possible. Cela revient à ne pas considérer les erreurs inférieures à ε et à interdire celles supérieures à ε [81]. Maximiser la platitude de la fonction permet de minimiser la complexité du modèle qui influe sur ses performances en généralisation. En effet, la théorie de l'apprentissage permet de borner l'erreur de généralisation par une somme de deux termes : l'un dépendant de la complexité du modèle et l'autre dépendant de l'erreur sur les données d'apprentissage. Les méthodes SVMs sont basées sur le contrôle de la complexité du modèle lors de l'apprentissage. Les Sections suivantes présentent l'algorithme SVR pour un modèle linéaire, puis pour le cas non-linéaire.

IV.4.1. Problème linéaire

La méthode est d'abord décrite pour une fonction linéaire de la forme :

$$F(x) = (w, x) + b \qquad \text{(IV.01)}$$

Avec $x \in Rp$ le vecteur d'entrée, $w \in Rp$ le vecteur des paramètres (ou poids) et b une constante à déterminer. Pour assurer la platitude de la fonction f, la norme des poids $||w||$ est minimisée (contrainte sur la dérivée de la fonction f). Le problème revient donc à minimiser cette norme en garantissant que les erreurs sont inférieures à ε et peut s'écrire

$$\text{Min}\frac{1}{2}\left|\left|w\right|\right|^2 \qquad \text{(IV.02)}$$

$$\text{S.C.}|y_i - (w, x_i) - b| \leq \varepsilon,\ i = 1,\ldots N \qquad \text{(IV.03)}$$

On remarque que dans cette formulation du problème, l'objectif n'est pas de minimiser l'erreur d'apprentissage comme dans les réseaux de neurones ou la plupart des algorithmes de régression, mais d'assurer la platitude de la fonction. La minimisation de l'erreur n'apparaît que sous forme de contraintes ou aucune erreur n'est autorisée. Cette description du problème considère donc qu'une fonction linéaire qui approxime avec une précision tous les exemples. En pratique, en présence du bruit trop important ou de valeurs aberrantes, il est aussi plus important d'autoriser certaines erreurs. Dans ce cas, le concept de marge souple (soft margin)

est utilisé. Il consiste à introduire des variables de relâchement (slack variables) ξ_i, ξ_i^* pour rendre faisables les contraintes du problème d'optimisation qui devient

$$\text{Min}\frac{1}{2}\|w\|^2+C\sum_{i=1}^{N=1}(\xi i\ \xi i\ *) \qquad \text{(IV. 04)}$$

Sous les contraintes

$$yi - (w,xi) - b \leq \varepsilon + \xi_i \qquad \text{(IV.05)}$$

$$(w,xi) + b - y_i \leq \varepsilon + \xi_i^* \qquad \text{(IV.06)}$$

$$\xi_i,\xi_i^* \geq 0\ ,\ i = 1,...,N \qquad \text{(IV.07)}$$

Où ξi et $\xi i\ *$représentent respectivement les erreurs positives et négatives. La constante $C > 0$ est un hyper paramètre permettant de régler le compromis entre la quantité d'erreur autorisée et la platitude de la fonction. Cette formulation du problème revient à utiliser une fonction d'erreur $|\xi|\varepsilon$ appelée ε-insensible (ε-insensitive) de la forme:

$$|Y\text{-}f(x)|\varepsilon = \begin{cases} 0 & \text{pour} |y - f(x)| < \varepsilon \\ y - f(x) & \text{pour} |y - f(x)| \geq \varepsilon \end{cases} \qquad \text{(IV.08)}$$

Représentée sur la figure IV.1. Cette fonction peut être interprétée comme en créant un tube d'insensibilité de rayon ε autour de la fonction f(x) (voir la Figure. IV.2). Les variables ξ_i et ξ_i^*ou représentent alors la distance selon l'axe y entre le point (xi, yi) et le bord ce tube.

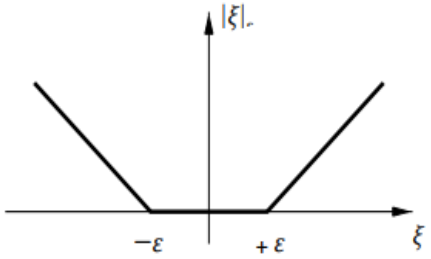


Figure IV.1: fonction d'erreur ε-insensible

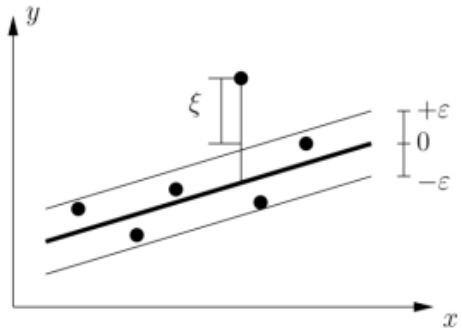


Figure IV.2 : Tube d'insensibilité de rayon ε.

Le problème (3) se résout en minimisant le Lagrangien L donné par

$L=\frac{1}{2}\|w\|^2+c\sum_{i=1}^{N}(\xi i + \xi i*) - \sum_{i=1}^{N}(ni\xi i + n\xi i*)$ (IV.09)

$\sum_{i=1}^{N} = \alpha i(\varepsilon + \xi i - yi + (w, xi) + b)$

$\sum_{i=1}^{N} = \alpha i * (\varepsilon + \xi i * - yi + (w, xi) + b)$

Où les ηi, ηi*, αi, αi* représentent les multiplicateurs de Lagrange. Le calcul des dérivées de L par rapport aux variables primales (w,b,ξi,ξi *) donne

$\frac{\partial l}{\partial b}=\sum_{i=1}^{N} i = 1 - \alpha i)0$ (IV.10)

$\frac{\partial l}{\partial b}\sum_{i=1}^{N}(\alpha * - \alpha i)\chi i = 0$ (IV. 11)

$\frac{\partial l}{\partial} = C- \alpha_i - n_i =0$ (IV.12)

$\frac{\partial l}{\partial} = C-\alpha_i^*-n_i^*=0$ (IV.13)

Les équations (IV.12) et (IV.13) permettent de supprimer les variables $\eta_i= c- \alpha_i$et $\eta_i^*=c-\alpha_i^*$En réinjectant ces résultats dans (IV.05), on obtient le Lagrangien dual qui doit être maximisé.

$l_D= -\frac{1}{2}\sum_{ij=1}^{N}(-\alpha i *)(\alpha j - \alpha j *)(xi, xj)$ (IV.14)

Sous les contraintes

$\sum_{i=1}^{N} \alpha i - \alpha i = 0$Et $\alpha_i, \alpha_i^* \in [0, C]$ (IV.15)

A partir de l'équation(9), les poids du modèle sont déterminés par :

$W=\sum_{i=1}^{N}(\alpha i - \alpha i)\, xi$ (IV.16)

Le modèle s'écrit donc :

$F(x) =\sum_{i=1}^{N=1}(\alpha i - \alpha i *)\,(x, xi) + b$ (IV. 17)

Le paramètre de biais b peut se calculer grâce aux conditions de Karush-Kuhn-Tucker (KKT) selon lesquelles les produits entre les variables duales et les contraintes sont nuls à la solution :

$\alpha_i= (\varepsilon+\xi_i -y_i+ (w,x_i) +b) =0$ (IV.18)

$\alpha i^*= (\varepsilon+\xi_i^* -yi+ (w,x_i) +b) =0$ (IV.19)

$ni\xi i=(C-\alpha_i)\xi_i=0$ (IV.20)

$n_i^*\xi_i*=(C-\alpha_i^*)\xi_i^*=0$ (IV.21)

b peut par exemple être déterminé par (Eq.18) sur un point particulier pour lequel xi, yi et ξi sont connus et αi a été précédemment évaluée par l'algorithme de maximisation du Lagrangien dual L_D.

On peut remarquer que, d'après (Eq.20) et (Eq.21), les seuls points pouvant être à l'extérieur du tube d'insensibilité ($\xi i, \xi i^*$=60) sont ceux pour lesquels les multiplicateurs de Lagrange correspondants respectent $\alpha_i = C$ ou $\alpha_i^* = C$. D'autre part, pour les exemples à l'intérieur du tube

($|yi - f(xi)| < \varepsilon$ et $\xi i, \xi i^* = 0$), on a $\alpha i, \alpha i^* = 0$ pour pouvoir vérifier les conditions (IV.18) et (IV.19). Cela entraîne la propriété de parcimonie (sparsity) des SVMs : seuls les exemples avec des αi, αi*correspondants non nuls sont nécessaires pour implémenter le modèle (IV.20) .Ces exemples sont appelés Vecteurs de Support ou Support Vectors (SVs) et ne constituent dans la pratique qu'une petite fraction des données d'apprentissage [82].

IV.4.2. Problème non-linéaire

L'extension des SVMs au cas non-linéaire est assez simple et repose sur la projection des données dans un espace de dimension supérieure dans lequel le problème devient linéaire. L'exemple suivant [83] montre l'implémentation d'une fonction quadratique par un SVM linéaire. Soit la projection Φ définie par :

$\phi:(x_1,x_2) = (x_1\sqrt{2}\, x1, x_2, x_2)$ (IV.22)

L'algorithme SVM précédemment décrit appliqué aux données projetées dans R^3 par Φ donnerait une fonction linéaire dans R3 correspondant à une fonction quadratique dans R^2 . Cette approche semble raisonnable pour cet exemple en deux dimensions, mais peut vite devenir coûteuse en temps de calculs pour des non-linéarités polynomiales d'ordre supérieur ou des problèmes à grande dimension d'entrée [84] Pour résoudre ces problèmes, une projection implicite par fonctions noyaux est utilisée.

Le fait que les données xi n'apparaissent que sous forme de produit scalaire aussi bien dans la formulation du problème d'optimisation (IV.16) que dans la fonction f constitue un des principaux avantages des SVMs. Il n'est pas nécessaire de connaître la projection Φ : R p→ F explicitement, mais uniquement le résultat $k(x_i,x) = h\Phi(x_i),\Phi(x)_i$ du produit scalaire entre les images des points dans l'espace étendu F, aussi appelé featurespace. Le problème non linéaire peut donc être formulé ainsi

$\text{Max } l_d = -\frac{1}{2}\sum_{ij=1}^{N}(\alpha i - \alpha i*)(\alpha j - \alpha j*)k(xi, xj)$ (IV.23)

$-\varepsilon\sum_{i=1}^{N}(\alpha i + \alpha i*) + \sum_{i=1}^{N} yi(\alpha i - \alpha i*)$ (IV.24)

Sous les contraintes

$\sum_{i=1}^{N} \alpha i - \alpha i* = 0 \quad et\ \alpha, \alpha i* \in [0, C]$ (IV.25)

De même, les poids sont donnés par :

$\sum_{i=1}^{N}(\alpha i - \alpha i*) = 0\Phi(xi)$ (IV.26)

Et le modèle par :

$F(x)= \sum_{i=1}^{N}(\alpha i - \alpha i *) = k(xi, x) + b$ (IV.27)

La principale différence avec le cas linéaire est que le vecteur des poids w n'est plus exprimé explicitement dans Rp mais implicitement dans F. De même, l'algorithme cherchera la fonction la plus plate dans l'espace étendu F et non plus dans l'espace d'entrée R p . La fonction k est appelée fonction noyau ou noyau (kernel function ou kernel). Dans la pratique, seule cette fonction est connue, la projection Φ ne l'est pas.[82] Les noyaux les plus couramment utilisés pour les SVMs sont les noyaux polynomiaux, sigmoïdaux et à fonction de base radiale (Radial Basis Function, RBF) définis ainsi :

– linéaire : k(x,x') = (x,x'');

– polynomial : $k(x,x') = (\gamma(x,x')+ c)^{d}$;

– sigmoïdal : k(x,x') = tanh(γ(x,x') + c);

IV.5. FONCTION DE BASE RADIALE RBF

Pour être admissible, une fonction noyau doit vérifier les conditions de Mercer [83] qui se résument à vérifier que la fonction k : R p×p→ R corresponde bien à un produit scalaire définissant une norme dans un espace étendu F.

Programmation linéaire et SVR

Une autre formulation basée sur la minimisation de la norme l_1 des paramètres conduit à un problème de programmation linéaire et non plus quadratique. L'idée n'est plus de maximiser la platitude de la fonction, mais plutôt de minimiser le nombre de SVs qui influe aussi sur les bornes de l'erreur de généralisation. Une estimée du nombre de SVs et donnée par (αi + αi*). La minimisation de ce terme, aussi appelée "rétrécissement" des paramètres (shrinkage)

Conduit en effet à maximiser le nombre de paramètres nuls dans le modèle (IV.28) et donc à minimiser le nombre de vecteurs de support nécessaires pour le calcul de f. Par rapport à l'apprentissage SVM par programmation quadratique, l'apprentissage par programmation linéaire conduit généralement à des modèles utilisant moins de vecteurs de support. Pour la fonction d'erreur ε-insensible, le problème devient [85]

$\text{Min} \sum_{i=1}^{N}(\alpha i + \alpha i *) + C \sum_{i=1}^{N}(\xi i + \xi i *)$ (IV.28)

Sous les contraintes

$yi - \sum_{j=1}^{N}(\alpha j + \alpha j *)k(xi , xj) - b \leq \varepsilon + \xi i$ (IV.29)

$\sum_{j=1}^{N}(\alpha j + \alpha j *)k(xi , xj) + b - yi \leq \varepsilon + \xi i$ (IV.30)

Pour i = 1,..., N. Contrairement au SVR classique, le passage à la formulation duale n'apporte aucune simplification. Cependant ce problème peut se résoudre par des techniques de programmation linéaire habituelles. Une autre formulation donnée par Mangasarian et qui implique moins de variables dans le problème[83].

$\sum_{i=1}^{N} \alpha i C \sum_{i=1}^{N} \varepsilon i$ (IV.31)

Sous les contraintes

$-\xi i \leq \sum_{j=1}^{N} \alpha jk(xj, xi) + b - Yi \leq \xi i$ (IV.32)

$0 \leq \varepsilon \leq \xi i$

$-ai \leq \alpha \leq ai$, i = 1,..., N

Soit, sous forme matricielle

Min 1 T a + C1 T ξ (IV.33)

S.c. $-\xi \leq Ka+b1-y \leq \xi$

$-0 \leq \xi \leq \xi$

$-a \leq \alpha \leq a$

Dans ce cas, le modèle est donné par :

$K(x,x')=\exp\left(\frac{-|x-x'|2}{2\sigma 2}\right)$ (IV.34)

$F(x) = \sum \alpha_i k(x, x_i) + b = k(x, X^T)\alpha + b$ (IV.35)

Sous cette forme, le problème ne requiert pas que le noyau k respecte les conditions de Mercer. La seule contrainte est qu'il soit symétrique [83]. En effet, contrairement au problème quadratique qui minimise une norme des poids induite par le noyau dans le feature space F, la minimisation de la norme des paramètres est ici réalisée dans RN. Il n'est donc pas nécessaire de disposer d'un noyau qui définit un produit scalaire et induit une norme dans F.

IV.6. PRISE EN CHARGE DE LA REGRESSION VECTORIELLE

IV.6.1. SVM Le SVMr

Est l'une des méthodes d'apprentissage automatique, basée sur la théorie de l'apprentissage statistique (SLT), qui implémente la théorie de la minimisation des risques de structure (SRM).Considérons la base de données d'apprentissage de N_d points (xi, yi), avec i= 1,.., N_d , X représente les entrées du modèle et Y sa sortie. L'objectif de SVMr est de trouver une fonction de régression non linéaire pour produire la sortie en se basant sur les données d'apprentissageN_d. Un SVM modèle prend la forme Eq. (IV.36) [86]

$Y(x)=\omega T.Ø(x)+ b$ (IV.36)

Où Ø (.) représente la fonction non linéaire, ω représente la Vecteur des poids, et b représente le biais de la régression fonction .La perte ε-SVM insensible fonction de Vapnik, et basée sur

le principe du risque structurel Minimisation (SRM), w et b sont estimés en minimisant l'équation de la fonction de risque structurel (SRM) suivante Eq.37 [87].

Min R w,c, ξ, ξ')=$\frac{1}{2}$ ||w||²+$\sum_{i=1}^{Nd}$(ξi + ξ′i)

$$sc\begin{cases} yi - yi = yi(w^T\acute{O}(xi) + b)\xi + \xi i \\ yi - yi = (w^T\acute{O}(xi) + b) - yi \leq \xi + \xi' i \\ \xi i, \xi' i \geq 0 \; i = 1,2 \ldots, Nd \end{cases} \quad (IV.37)$$

Le compromis entre la complexité du modèle et l'erreur empirique est spécifiée par un paramètre régularisé C et est appelé paramètre de pénalité. ξ et ξ ' sont variables lentes représentant les erreurs d'apprentissage supérieures et inférieures, respectivement. Selon la méthode des multiplicateurs de Lagrange, le problème d'optimisation mentionné ci-dessus peut être résolu dans sa forme double Eq. (IV.38)et différenciation aux variables primaires (w ; C ; ξ ; ξ')

Min $\sum_{i=1}^{Nd}$ yi(αi − αi ∗)

$$\xi\sum_{i=1}^{Nd}(\alpha i + \alpha i) * -\frac{1}{2}\sum_{i=1}^{Nd}\sum_{j=1}^{Nd} \quad (IV.38)$$

(αi−αi')(αj−αj') Ó(xi)^T . Ó(xj)

$$sc\begin{cases} \sum_{i=1}^{Nd}(a1 - a'i) = 0 \\ 0 \leq 'ai - a'i \leq c \text{ for: } i = 1,2 \ldots, Nd \end{cases} \quad (IV.39)$$

Où α et α′ sont les doubles multiplicateurs de Lagrange. Les multiplicateurs de Lagrange optimaux α* sont résolus par algorithme de programmation quadratique standard ; puis la régression fonction peut être réécrite comme suit : (IV.40)

$$F(x)\sum_{i=1}^{Nd}(ai\ a'i)k(xi,xj)b \quad (IV.40)$$

Avec K(xi,xj)=. Ó(xi)^T . Ó (xj) (IV.41)

K (xi, xj) est la fonction noyau. Les paramètres du noyau fonction sont importants pour la précision de la prédiction. Différentes fonctions du noyau ont une influence importante sur l'efficacité de la Support Vector Machine [86] Certains noyaux commun sont été testés dans ce modélisation et parmi ces fonctions, on a observé que la fonction noyau de la gaussienne Radial Basic (GRBF) Éq. 6 donnes des résultats prometteurs :

K ($x - i, x - j$)=$e^{(-1/2\sigma 2*|xi\text{-}xj|^2)}$ (IV.42)

σ² désigne la largeur du paramètre du noyau gaussien. Le taux de ε et C peut affecter le nombre de vecteurs de support utilisés pour construire la fonction de régression. Par

conséquent, pour optimiser et sélectionner la fonctionnalité appropriée du modèle SVMr, la valeur C et ε doivent être considérées exhaustivement. Lors de l'application de SVMr, la conception les valeurs de C, ε et de largeur σ de la fonction de base radiale ont été optimisé par la méthode de la libellule (DA). (Dragonfly)

IV.6.2 Dragonfly algorithme

Dragonfly (DA) est une méthode récente pour optimiser l'intelligence des essaims. Les essaims de libellules ont besoin attirer les sources de nourriture et distraire les ennemis externes selon cinq opérations principales que sont l'alignement, la cohésion, séparation, attirance pour les sources de nourriture et distraction par ennemis. Ces opérations sont représentées par les éléments suivants les modèles mathématiques en utilisant les équations. (IV.43) à (IV.47) [86].La séparation (S) : le but de cette étape est d'éviter la statique collision entre des individus et d'autres individus dans le quartier. Cette séparation s'exprime par ce qui suit équation :

$$Si=\sum_{j=1}^{N} X - X \qquad \text{(IV.43)}$$

Où X est la localisation de l'individu courant, Xj montre l'emplacement du jème individu proche, et N est le nombre d'individus proches personnes. Alignement (A) : qui indique la correspondance des vitesses entre les membres d'un même groupe est décrite comme suit :

$$Ai=\sum_{j=1}^{N} \frac{Vj}{N} \qquad \text{(IV.44)}$$

Où Vj représente la vitesse du jième individu voisin La cohésion (C) : qui représente l'attraction des individus vers le centre de la masse du quartier et est décrit comme suit : k

$$Ci=(\sum_{j=1}^{N} \frac{Xj}{N})-X \qquad \text{(IV.45)}$$

L'attirance pour une source de nourriture est déterminée comme suit

$$Fi=(X^{+}) - X \qquad \text{(IV.46)}$$

X^{+} Représente l'emplacement de l'aliment.La déviation vers l'extérieur d'un ennemi est décrite comme suit :

$$Ei=X^{-}+X \qquad \text{(IV.47)}$$

X^{-} Représente l'emplacement de l'ennemi. Le comportement des libellules a permis l'accord deces cinq modèles correctifs. Pour mettre à jour l'emplacement de l'artificielle libellule dans un environnement de recherche et simuler leur mouvements, deux vecteurs sont étudiés : le vecteur pas (ΔX) Eq.(Eq.48) et le vecteur position (X) IV.49 :

$$\Delta X_{t+1} = (sSi + aA_i + cCi + f F_i + eE_i) + W\delta x_t \qquad \text{(IV.48)}$$

Où s représente le poids de séparation, Si indique la séparation du ième individu, a est le poids d'alignement, Ai est le alignement du i ème individu, c indique la cohésion. Poids, Ci est la cohésion du ième individu, f est la nourriture facteur, Fi est la source de nourriture du ième individu, e est l'ennemi acteur, et Ei est l'emplacement de l'ennemi du ième individu. W est le poids de l'inertie, et t est le compteur d'itérations. Après calcul du vecteur pas, la vectrice position est calculé comme suit :

$$X_{t+1} = X_t + \Delta X_{t+1} \quad \text{(IV.49)}$$

Dans le cas où il n'y a pas d'individus voisins, une méthode de marche aléatoire (vol de Lévy) est appliquée pour améliorer le caractère aléatoire de la recherche [88]. Dans ce cas, l'équation du vecteur de position est montrée dans l'équation(IV.50).

$$X_{t+1} = X_t + \text{Lévy (d)} * X_t \quad \text{(IV.50)}$$

D est la dimension des vecteurs de position. Le Lévy se calcule comme suit :

$$\text{Lévy(x)} = 0.01 * \frac{r1 * \sigma}{r2\ 1/} \quad \text{(IV.51)}$$

Où r_1 et r_2 sont deux nombres aléatoires dans [0,1], β est une constante, et σ est calculée comme suit

$$\sigma = \Gamma(1+\beta) * \sin(\pi\beta\ 2\ )\ \Gamma(\ (1+\beta)\ 2\ ) * \beta * 2\ (\ \beta - 1\ 2) \quad \text{(IV.52)}$$

$$\Gamma(x) = (x - 1) \quad \text{(IV.53)}$$

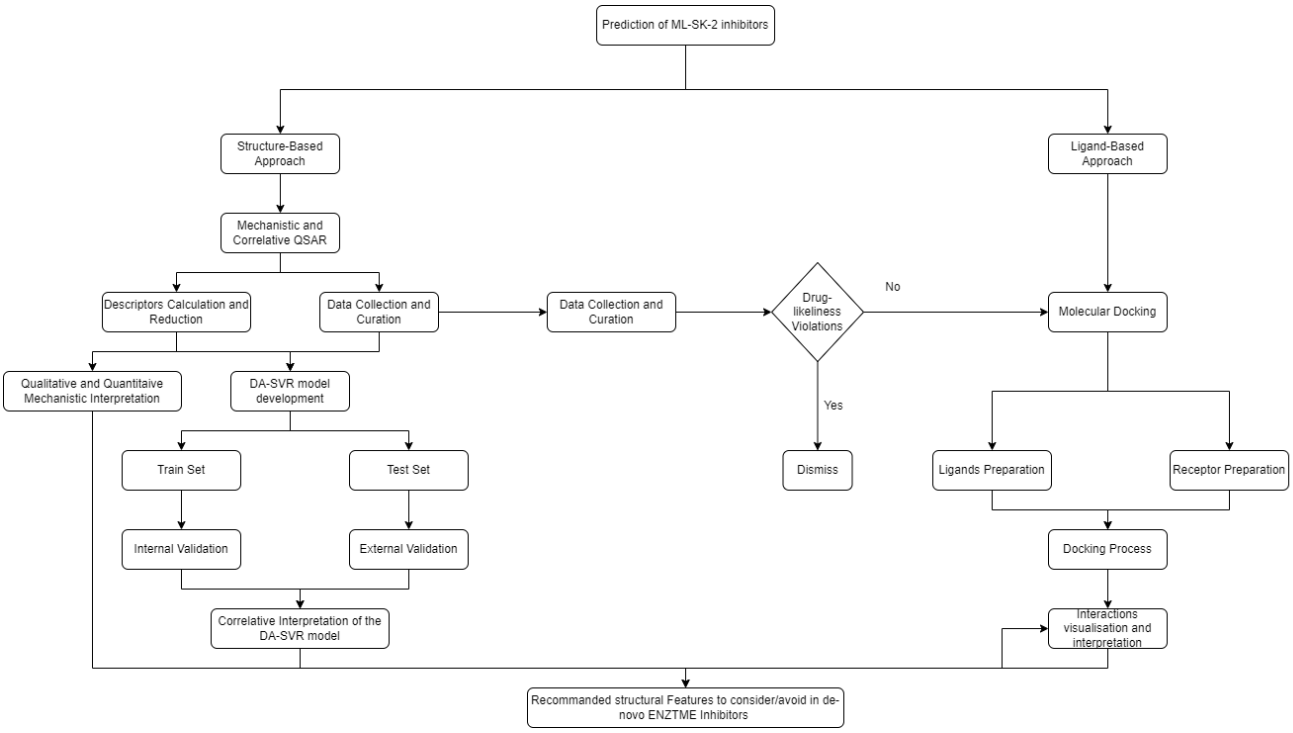
IV.7.CONCLUSION :

Une introduction sur les méthodes SVM pour l'identification des systèmes a été présentée. Certains points importants pour l'application pratique des SVM tels que le réglage des hyper paramètres, l'implémentation algorithmique ou l'application concrète à l'identification ont ensuite été traités. , la méthode SVM donne des résultats équivalents avec plus de vecteurs de support quand les données sont en nombre suffisant. Quand les données sont peu disponibles, la méthode SVM donne des résultats avec moins de centres. L'avantage de ce modèle mathématique est sa capacité de résoudre les problèmes non linéaires pour des phénomènes complexes. L'utilisation de L'algorithme Dragonfly consiste à généraliser le noyau paramètres dans le SVM et à améliorer la robustesse de la maquette [83]

EXPLOITATION DES RESULTATS

V.1.METHODOLOGIE DE CALCUL :

La méthodologie adoptée pour le calcul de la concentration inhibitrice (PIG50) par la technique SVM-DA peut être schématisée suivant l'organigramme ci-dessus :



V.1.1. Collecte de la base de données

Les données ont été recueillies au moyen du Pubchem et de la littérature (QSAR_Cancer. pdf) sur la base de leur activité prouvée contre SK-MEL-2 et V600E-BRAF 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 moitié inhibitrice ont été conservées. À des fins de linéarisation. Les valeurs expérimentales des IG50 (nM) ont été converties en échelle logarithmique décimale négative (pGI50 = -Log10 (pGI 50)

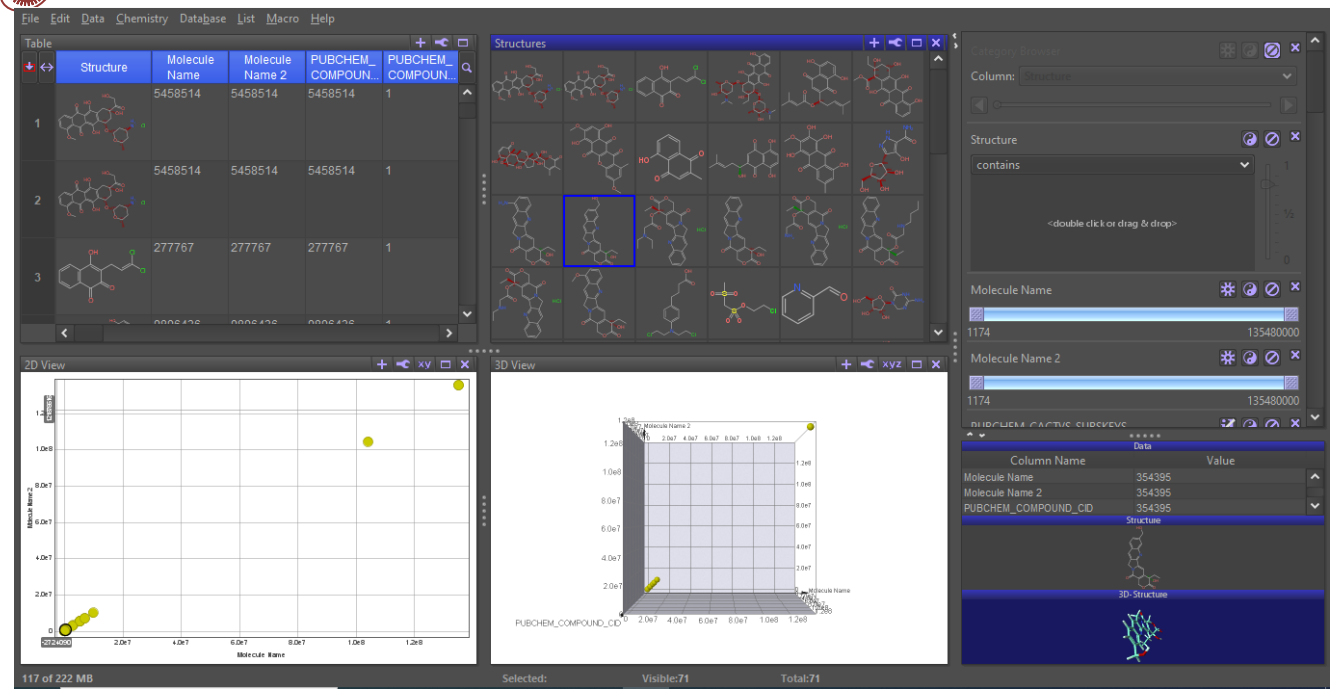
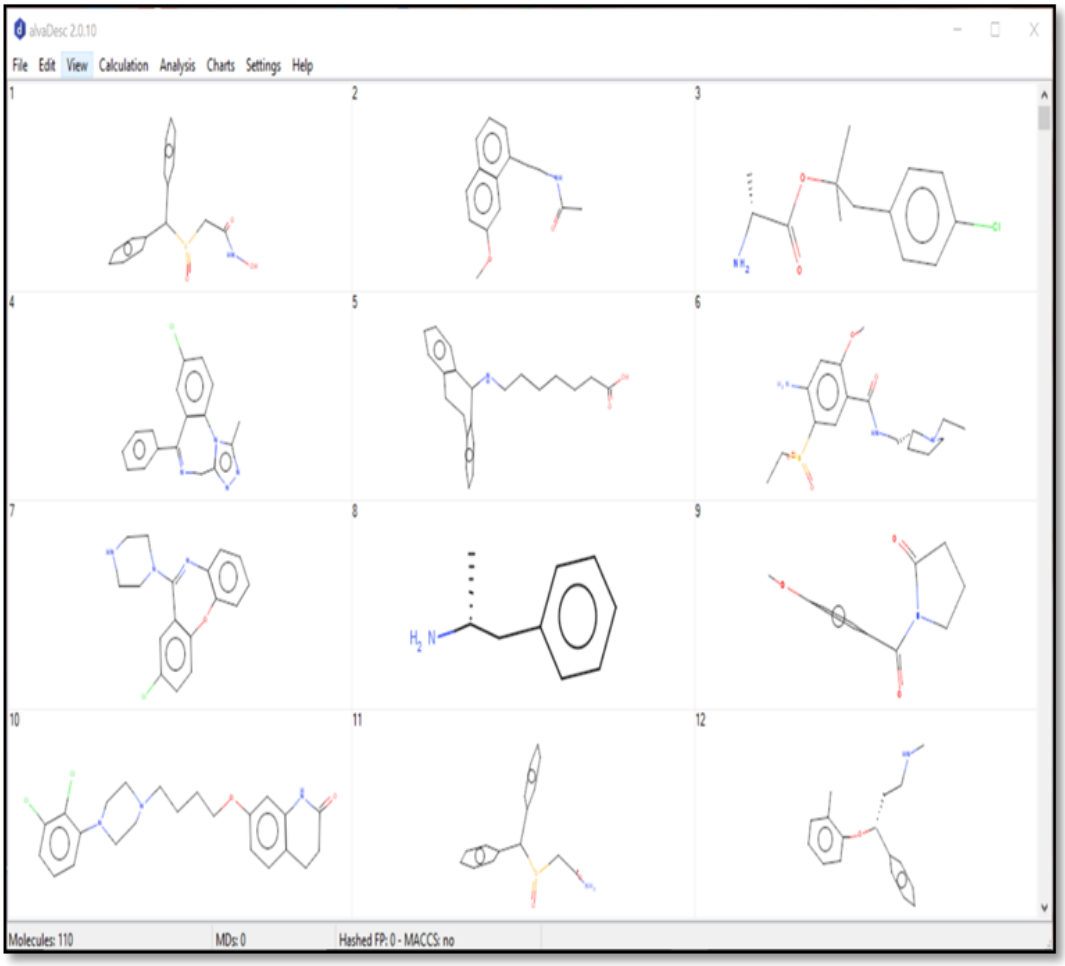


Figure V.1 : capture d'écran Data Warriorr

V.1. 2. 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'outil Stepwise-MLR a été utilisé deux fois pour sélectionner uniquement les descripteurs représentatifs réduisant le nombre de descripteurs à 10 et formant ainsi une matrice de 71x10.



FigureV.2 : Capture d'écran AlvaDesc

V.1. 3. Division de la base de données et développement du modèle

L'ensemble de données a été distribué de manière optimale en un ensemble d'apprentissage contenant 76% de l'ensemble de données.et un ensemble de test. Contenant 24% de l'ensemble de données. Les modèles QSAR ont été développés en utilisant l'approche : DA_SVR, réalisés au moyen MATLAB

V.1.3.1. Support Vecteur Machine (SVM)

Les machines à vecteurs de support ou séparateurs à vaste marge (en anglais support-vector machine. SVM) sont un ensemble de techniques d'apprentissage supervisées. Destinées à résoudre des problèmes de discrimination et de régression. Les SVM sont une généralisation des classifieurs linéaires. Les SVM ont été appliqués dans de nombreux domaines (chimio-informatique. recherche d'information. vision par ordinateur et finance).

V.1.3.2. 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.

A. Validation interne

La validation interne d'un modèle QSAR/QSPR a été réalisée en utilisant la validation croisée LOO (Leave-One-Out) ou LMO (Leave-Many-Out) qui est quantifiée par le coefficient R^2. Ce processus consiste à extraire un certain nombre k de molécules du jeu initial à N molécules et à construire un nouveau modèle avec les (N-k) molécules restantes à l'aide des descripteurs choisis (seules les constantes de la régression changent). Ce processus est ensuite réitéré pour prédire les valeurs de toutes les molécules de la série d'apprentissage. En fonction du nombre de molécules retirées à chaque itération.

On parlera de Leave-One-Out (LOO) ou de Leave-ManyOut (LMO) selon qu'une ou plusieurs molécules est (sont) retirée(s). Cependant.la validation interne est insuffisante pour étudier le pouvoir prédictif d'un modèle. Pour cette raison la validation interne du modèle est devenue une norme et une partie obligatoire dans la modélisation QSAR/QSPR

***B.* Validation externe**

Cette méthode consiste à prédire la propriété/activité d'une série de molécules appelée généralement série de test qui ne sont pas dans la série de développement du modèle. Cette validation est caractérisée par le paramètre R^2. Récemment plusieurs études ont montré l'insuffisance du paramètre R^2 pour vérifier le pouvoir prédictif des modèles QSAR/QSPR. 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 ».

R^2 : coefficient de détermination

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

$R0^2$: coefficient de détermination sans intercept

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

c

$Q2F^2$

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

Scaled Avg.Rm^2 : Rm² moyen normalisé

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

Scaled DeltaRm^2 : moyenne de rm²

$$\left|r_m^2 - r_m^{'2}\right|$$

RMSE : Erreur quadratique moyenne

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

V.1.4. Domaine d'applicabilité

Un modèle QSAR/QSPR 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.

V.2. RESULTATS ET DISCUSSION

V.2.1. Sélection des descripteurs représentatifs

La sélection a été effectuée à l'aide du logiciel Stepwise révélant les dix descripteurs les plus pertinents et représentatifs : MATS3m, CATS3D_08_DN, MATS1p, CAT2D_04_DD, CATS2D_08_DD, F10 [N_N], Mor30p, Mor26i, Mor10i, s2_pathlength.

- **MATS3m**

La mesure d'auto corrélation de Moran de lag 3 pondérée par la masse permet de quantifier la similarité spatiale des valeurs d'une variable en prenant en compte l'influence relative des différentes régions environnantes.

- **CATS3D_08_DN**

Est un terme technique qui peut être utilisé dans le domaine de la chimie ou de la biochimie. Ce terme peut faire référence à une plage de valeurs de distance (8,000 - 9,000 Å) utilisée pour classer un certain type d'interaction ou de propriété moléculaire, peut-être lié à la présence ou à l'absence d'un groupe donneur ou accepteur dans une molécule. "CATS3D" est un logiciel utilisé pour l'analyse des structures et des propriétés moléculaires.

- **MATS1p**

Est une mesure qui évalue la similarité spatiale des propriétés moléculaires en prenant en compte leur polarisabilité. Elle peut aider à identifier les groupes de molécules ayant des propriétés similaires et liées à la polarisabilité

- **CATS2D_04_DD**

Est une mesure utilisée pour évaluer les interactions donneur-donneur entre les atomes d'une molécule à une distance de 4 liaisons chimiques. Cette mesure est basée sur la technique CATS (Chemical Atom Typing Scheme) qui attribue des types d'atomes chimiques à chaque atome de la molécule. Les interactions donneur-donneur peuvent influencer les propriétés chimiques et physiques de la molécule, telles que la stabilité, la solubilité ou la réactivité.

- **CATS2D_08_DD**

Est une méthode pour quantifier la connectivité moléculaire dans un plan bidimensionnel et peut être utilisée pour étudier la structure moléculaire et les propriétés chimiques des composés.

- **F10 [N-N]**

Elle est importante pour comprendre les propriétés physiques et chimiques de la molécule et peut être utilisée pour la conception de nouveaux composés

- **Mor30p**

Une mesure de la similarité spatiale entre les valeurs d'une variable et de ses voisins jusqu'à une distance de 30 unités spatiale, en prenant en compte la polarisabilité de chaque point

- **Mor26i**

Une mesure de la similarité spatiale entre les valeurs d'une variable et de ses voisins jusqu'à une distance de 26 unités spatiale, en prenant en compte le potentiel d'ionisation de chaque point

- **Mor10i**

Une mesure de la similarité spatiale entre les valeurs d'une variable et de ses voisins jusqu'à une distance de 10 unités spatiale, en prenant en compte le potentiel d'ionisation de chaque point

- **s2_pathLength**

La distance maximale, mesurée en nombre de liaisons chimiques, entre l'atome central d'une molécule et l'atome le plus éloigné du substituant 2 ; Cette mesure est souvent utilisée en chimie organique pour caractériser la structure spatiale des molécules



Figure V.3: Word Cloud de l'importance des Descripteurs

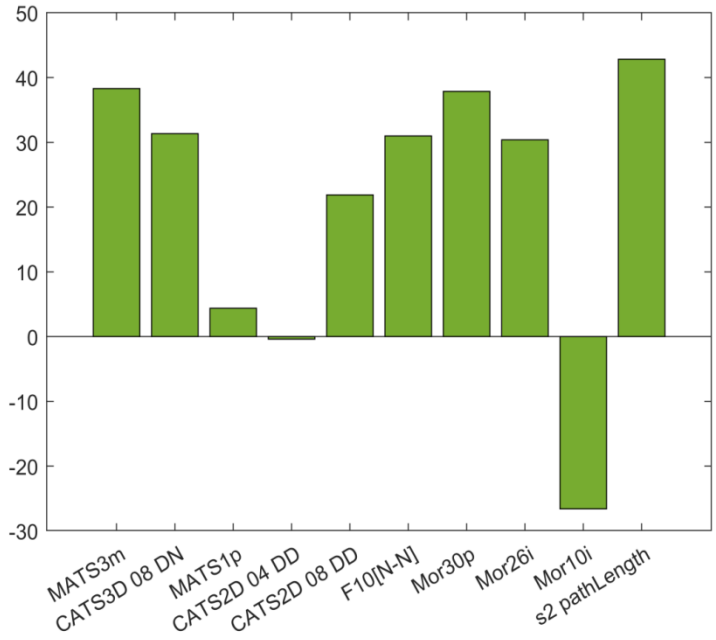


Figure V.4: Bar plot de l'importance relative des descripteurs

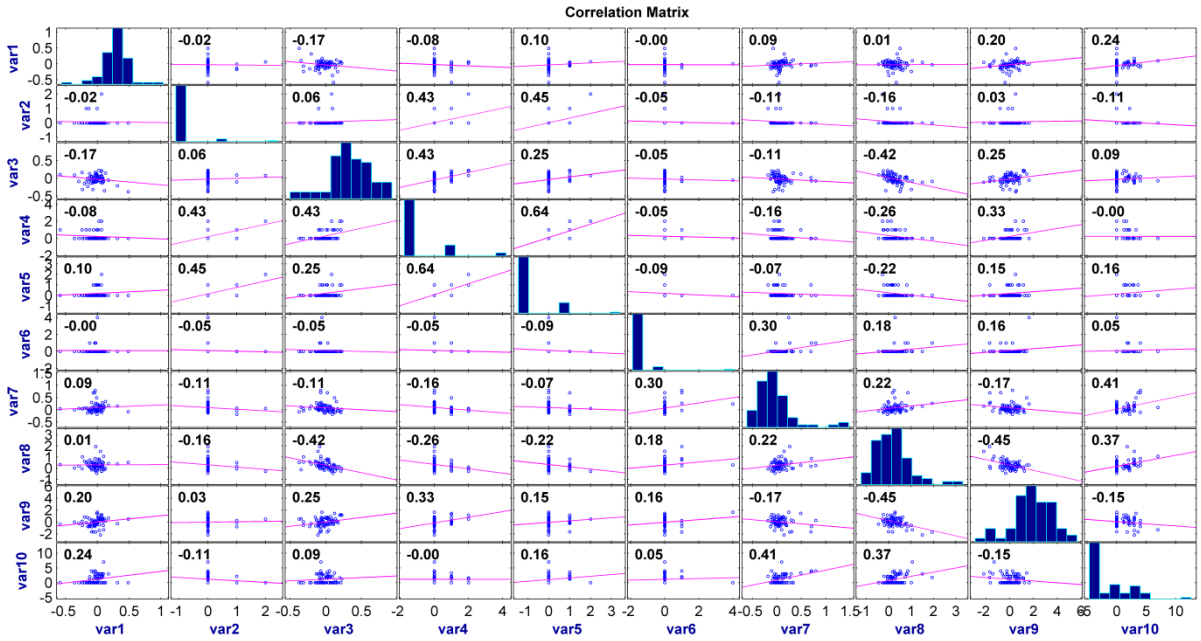


Figure V.5: Diagramme de corrélation

Des valeurs positives de R^2 indiquent une corrélation positive, tandis que des valeurs négatives de R_2 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 V.1 : les descripteurs pertinents

Nom de descripteurs	Dimension	Nom complets
MATS3m	2D	auto corrélation de Moran de lag 3 pondérée par la masse
CATS3D_08_DN	2D	Topologie d'atomes chimiques 3D Donneur-Négative à un décalage de 08
MATS1p	2D	Autocorrélation de Moran de décalage 1 pondérée par la polarisabilité
CAT2D_04_DD	2D	Topologie d'atomes chimiques 2D Donneur-Donneur à un décalage de 04
CATS2D_08_DD	2D	Topologie d'atomes chimiques 2D Donneur-Donneur à un décalage de 08
F10 [N_N]	2D	Fréquence des liaisons N-N à une distance topologique de 10
Mor30p	2D	Autocorrélation de Moran de décalage 30 pondérée par la polarisabilité
Mor26i	2D	Autocorrélation de Moran de décalage 26 pondérée par le potentiel d'ionisation
Mor10i	2D	Autocorrélation de Moran de décalage 10 pondérée par le potentiel d'ionisation
s2_pathlength	2D	longueur de chemin maximale du substituant 2

Tableau V.2 : descripteurs retenus

Numéro	MATS3m	CATS3D_08_DN	MATS1p	CATS2D_04_DD	CATS2D_08_DD	F10[N-N]	Mor30p	Mor26i	Mor10i	s2_pathLength	pGI50
1	-0,0043	1	0,0926	1	1	0	-0,1824	-0,0609	0,2140	2,2	6,904
2	-0,0043	0	0,0926	1	1	0	-0,0676	-0,2135	0,0326	2,2	5,614

3	-0,010	0	0,0967	0	0	0	-0,0369	-0,4549	-0,3465	0	4,756
4	6,90E-04	0	0,0130	0	1	0	0,1665	0,8108	-0,2315	2,9090	7,644
5	-0,0956	0	0,0513	0	0	0	-0,1052	0,5149	-0,2452	4	5,697
6	0,0618	0	0,1946	2	1	0	-0,0562	0,1096	1,1467	1,6666	5,506
7	-0,0102	0	0,1204	1	0	0	0,4895	0,2712	1,4291	3,2727	6,346
8	0,0216	0	0,0297	0	0	0	0,0987	0,4962	0,6743	0	6,063
9	-0,0274	0	0,1271	0	0	0	-0,1008	-0,2449	0,0799	0	5,732
10	-0,0312	0	0,1680	0	0	0	-0,0359	0,1527	-0,1922	0	5,71
11	0,0651	0	0,2099	1	0	0	-0,1039	-0,2471	1,2387	0	4
12	0,0249	0	0,2215	2	1	0	0,0219	-0,0552	1,15196	1,25	4,072
13	0,0675	0	0,0642	0	0	0	0,1288	0,1348	0,8877	2	6,54
14	0,0673	0	0,0461	0	0	0	0,0373	0,3123	0,4900	2	4,744
15	0,0768	0	-0,1335	0	0	0	0,2915	0,5692	0,3141	3	6,418
16	0,0843	0	0,0084	0	0	0	0,0735	0,0997	0,5704	2	6,506
17	0,0721	0	-0,0651	0	0	0	0,1191	0,2903	0,7646	3	6,308
18	0,1747	0	-0,0551	0	0	0	0,1753	0,4256	0,6576	3	5,503
19	0,0939	0	-0,1029	0	0	0	0,1925	0,49611	-0,6050	3	5,557
20	0,0845	0	-0,0310	0	0	0	0,1814	0,26908	9,02E-04	2	7
21	-0,0885	0	-0,0251	0	0	0	-0,1294	0,0185	-0,8045	0	4,498
22	-0,1530	0	-0,2280	0	0	0	0,0398	0,5801	-0,7553	0	3,463
23	-0,0474	0	-0,1039	0	0	0	0,0288	0,1486	-0,4229	0	4,671

24	-0,0610	0	-0,0550	1	1	0	0,1203	0,2816	0,0406	1,25	5,941
25	-0,1235	0	0,1320	0	0	0	-0,0861	0,1327	0,5423	1	4,567
26	0,0143	0	0,0067	1	1	0	-0,0487	0,1390	0,6912	1,6666	3,746
27	-0,2414	0	0,1646	1	0	0	0,0407	-0,0415	0,1154	0	4
28	-0,0390	0	-0,0189	0	0	0	0,1988	-0,1958	-0,6029	0	5,566
29	0,0659	2	0,0791	2	2	0	0,0543	-0,2788	0,4770	0	9,924
30	-0,0077	1	-0,0867	1	1	0	0,0022	-0,0362	-0,1081	1,25	5
31	0,3164	0	-0,2074	0	0	0	0,1521	0,3780	-0,2836	0	5,2
32	-0,1752	1	-0,1014	1	0	0	0,0708	0,5060	-0,1270	0	4,832
33	-0,0406	0	-0,1014	0	0	0	0,0256	0,2940	-0,1376	0	4,037
34	0,0014	0	-0,0554	0	0	4	0,2406	0,2872	1,6279	0	8
35	-0,2369	0	-0,3158	0	0	0	0,0385	1,0107	-0,5728	0	3,606
36	0,0332	0	-0,0675	0	0	0	0,1114	0,3918	0,4782	3	4,097
37	0,0270	0	0,0498	0	0	0	0,1617	-0,0212	0,4226	2	6,381
38	0,0338	0	-0,0533	0	0	0	0,0448	0,1388	0,1564	0	4,644
39	0,0892	0	-0,0393	0	0	0	-0,0077	-0,0872	0,3894	0	4,072
40	-0,028	0	-0,0870	0	0	1	0,2643	1,9391	-1,7011	1,7777	9,904
41	-0,1065	0	-0,1426	0	0	0	0,0779	0,3808	-0,6586	0	3,612
42	-0,2925	0	-0,0879	0	0	0	0,0549	0,1957	-0,2571	0	3,101
43	-0,5870	1	0,1050	1	0	0	0,0440	0,3371	-0,3908	0	2,027
44	-0,0709	0	-0,3154	0	0	0	-0,0417	0,5410	0,0169	0	3,419
45	0,0312	0	-0,0056	0	0	0	0,1854	0,7070	-0,4810	0	5,104
46	-0,3617	0	0,0912	0	0	0	-9,74E-04	0,0082	-0,0148	0	2,767

47	-0,0927	0	0,0402	0	0	0	-0,0513	-0,1405	0,2573	0	4,013
48	-0,0911	0	-0,0695	0	0	0	0,3044	-0,0092	-1,6456	0	5,265
49	-0,1219	1	0,0266	1	0	1	-0,0235	0,7396	-0,3089	2	4,042
50	-0,015	0	-0,0502	0	0	0	-0,0208	0,4986	-0,4379	0	4,691
51	0,0071	0	0,0184	0	1	0	-0,0170	0,1894	0,7948	1,5	5,743
52	-0,0692	0	-0,1301	0	0	0	0,0399	0,1934	0,3427	0	4
53	-0,0399	0	0,0256	0	0	0	0,0472	0,0790	-0,5142	2,1666	5,695
54	-0,1227	0	-0,0209	0	0	0	0,0250	0,0984	-0,6302	0	3,558
55	-0,0633	0	-0,1007	0	0	0	0,0349	0,4227	-0,1937	0	4,023
56	-0,0317	0	-0,0450	0	1	0	-0,0134	0,4207	-0,2050	2,75	4,85
57	-0,1340	0	-0,1018	0	0	0	-0,0188	0,1758	-0,8069	0	6,139
58	-0,2041	0	-0,1484	0	0	0	0,0680	0,6270	0,8213	0	2,1
59	-0,0245	0	-0,1068	0	0	0	-0,0041	0,3817	-0,1345	0	3,611
60	-0,0657	0	0,0180	0	1	0	0,6992	0,1711	-1,6755	3,6363	7,295
61	0,1113	0	-0,3496	0	0	0	0,1143	-0,1331	0,5464	0	4,211
62	-0,0506	0	0,0647	0	0	0	0,0853	-0,0704	-0,3522	0	5,242
63	0,0306	0	0,0383	0	0	0	-0,0472	0,7904	-1,4394	3	7,656
64	-0,0230	0	0,0961	0	0	0	0,0111	0,2180	0,0771	0	5,77
65	-0,1562	0	-0,0103	0	0	0	-0,1246	0,2845	0,7392	0	4,526
66	0,4885	0	-0,3757	0	0	0	-0,0768	0,5123	0,4921	0	4,746
67	0,1284	0	0,1463	0	0	0	0,3352	-0,0193	0,1620	0	6,265
68	-0,0393	0	-0,0723	0	0	1	0,7075	0,8561	-0,7719	4	9,015
69	-0,0359	0	-0,0728	0	0	1	0,7977	0,2447	0,2362	4	6,87

70	0,0491	0	-0,1130	0	0	0	0,1475	1,5020	-2,1932	3	8,402
71	0,0990	0	-0,0717	0	0	0	0,0976	1,0570	-1,3210	7	7,921

L'analyse des descripteurs obtenus indique que :

- Une valeur élevée de "longueur de chemin maximale du substituant 2" est bénéfique pour les inhibiteurs de SK-MEL-2, indiquant une plus grande distance spatiale entre le substituant et le reste de la molécule, ce qui peut favoriser la flexibilité moléculaire et des interactions de liaison améliorées.
- Une faible valeur d'auto corrélation de Moran de décalage 10 pondérée par le potentiel d'ionisation est bénéfique pour les inhibiteurs de SK-MEL-2, suggérant une distribution diversifiée des potentiels d'ionisation et une adaptabilité moléculaire au site de liaison.
- Une valeur élevée d'auto corrélation de Moran de décalage 26 pondérée par le potentiel d'ionisation est considérée comme bénéfique pour les inhibiteurs de SK-MEL-2 en raison d'une meilleure affinité de liaison, de caractéristiques pharmacophoriques spécifiques et d'interactions électrostatiques favorables.
- Une valeur élevée d'auto corrélation de Moran de décalage 30 pondérée par la polarisabilité est bénéfique pour les inhibiteurs de SK-MEL-2 en raison des interactions renforcées, de compatibilité structurale avec le site de liaison et d'une puissance inhibitrice accrue.
- Une valeur élevée de Fréquence des liaisons N-N à une distance topologique de 10 peut indiquer des motifs structuraux favorables et une diversité structurale parmi les inhibiteurs.
- Une haute valeur de Topologie d'atomes chimiques 2D donneur-donneur à un décalage de 08 est bénéfique pour les inhibiteurs de SK-MEL-2 en raison d'interactions donneur-donneur fortes, favorisant une meilleure affinité de liaison et une complémentarité structurale.
- Une faible valeur de Topologie d'atomes chimiques 2D donneur-donneur à un décalage de 04 est bénéfique pour les inhibiteurs de SK-MEL-2 en raison d'une meilleure affinité de liaison, d'une flexibilité moléculaire et d'un moindre encombrement stérique.
- Une valeur élevée d'auto corrélation de Moran de décalage 1 pondérée par la polarisabilité est bénéfique pour les inhibiteurs de SK-MEL-2 en raison d'interactions moléculaires renforcées, d'une reconnaissance spécifique du site de liaison et d'une puissance accrue.
- Une valeur élevée de Topologie d'atomes chimiques 3D donneur-négative à un décalage de 08 est avantageuse pour les inhibiteurs de SK-MEL-2 en raison d'interactions de liaison

renforcées, de complémentarité structurale avec le site de liaison et d'une reconnaissance moléculaire améliorée.

- Une valeur élevée de MATS3m peut être bénéfique pour les inhibiteurs de SK-MEL-2 en raison de la connectivité spatiale accrue et des interactions potentiellement renforcées avec le site de liaison. Cependant, d'autres facteurs doivent également être pris en compte pour évaluer l'efficacité globale d'un inhibiteur, tels que les interactions spécifiques, les propriétés pharmacocinétiques et la spécificité de la cible.

2. 2. Développement des modèles QSAR

. Modèle SVR-Dragonfly

Dans cette étude, un modèle DA_SVR a été développé pour prédire le pouvoir inhibiteur de croissance des diverses molécules contre le mélanome et l'enzyme tumoral V600E-BRAF L'objectif de la combinaison des algorithmes DA et SVR était de trouver la meilleure combinaison d'hyper paramètres en testant différentes valeurs. La meilleure combinaison obtenue était C=150, γ=1, et ε=0.05, avec l'utilisation d'une fonction gaussienne comme fonction Kernel. L'hybridation de SVM et DA permet d'éviter le sur-ajustement du modèle, en obtenant des coefficients R et R^2 élevés respectivement (0.9816 0.9206 0.9731) (0.9613 0.8230 0.9428) et une valeur de RMSE faible (0.3770). Le tableau V.3 présente les paramètres statistiques du modèle DA_SVR, tandis que la figure V.4 illustre la corrélation entre les valeurs prédites du pGI50 et les valeurs observées pour l'apprentissage, le test et l'ensemble des données.

Tableau V.3 : Critères statistiques pour le modèle DA_SVR

Critères de validation	Apprentissage	Test	Global
R^2	0,9635	0,8474	0,9468
R_0^2 test	0,9635	0,8474	0,9467
$R_0^{2'}$ test	0,9616	0,8230	0,9429
Q2F1	0,9852	0,9373	0,9781
Q2F2	0,9634	0,8434	0,9467

Scaled Avg.Rm^2	0,9410	0,7742	0,9132
Scaled DeltaRm^2	0,0323	0,1178	0,0470
CCC	0,9812	0,91 !!!69	0,9724

R^2également 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.

R_0^2 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 R_0^2 test élevé indique une bonne capacité de prédiction du modèle, tandis qu'un R_0^2 test faible suggère une moins bonne performance prédictive.

$R_0^{2'}$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).

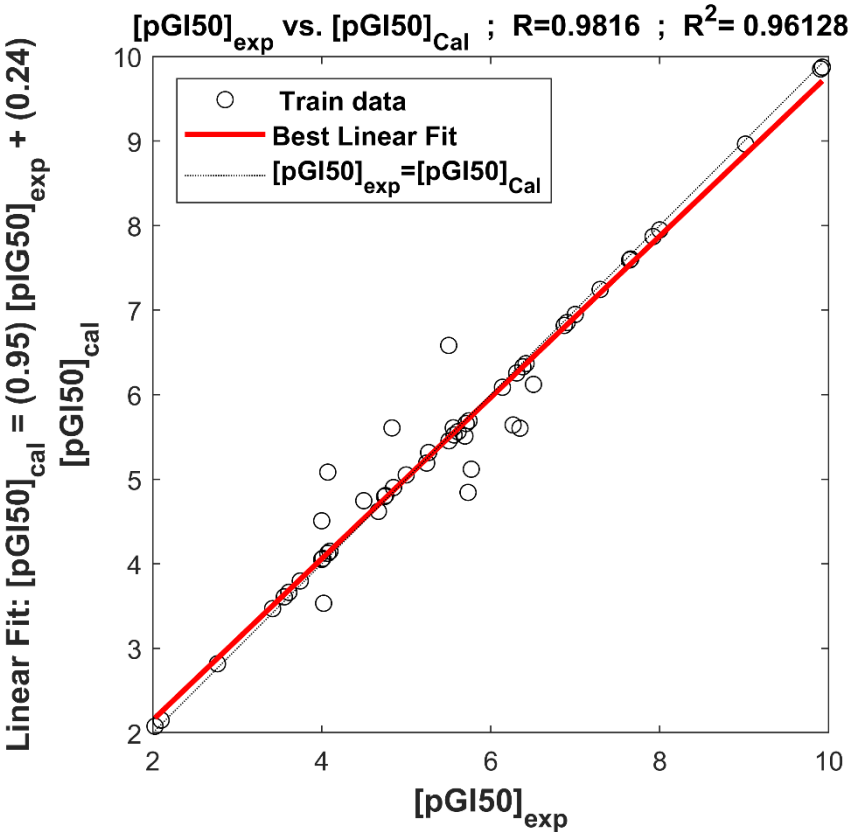
Q2F1: est basé sur la comparaison entre les valeurs expérimentales et les valeurs prédites par le modèle QSAR. Il mesure la précision du modèle en termes de prédiction de l'activité biologique des composés chimiques. Un Q2F1 élevé indique une bonne capacité du modèle à prédire avec précision les activités des composés.

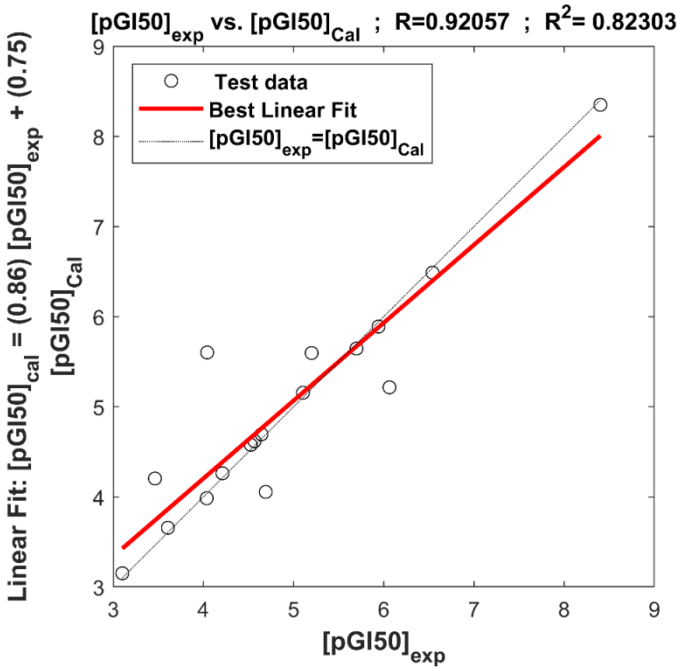
Q2F2: est basé sur la prédiction des composés de l'ensemble de test (Q2) en utilisant une procédure de validation croisée (F2) proposée par Schüürmann et al.

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.

CCC: le critère CCC est utilisé pour évaluer la qualité d'un modèle QSAR en termes de précision et de concordance entre les prédictions et les observations. Une valeur élevée de CCC indique une bonne concordance entre le modèle et les données réelles.


[pGI50]exp vs. [pGI50]Cal ; R=0.9816 ; R2= 0.96128
Train data
Best Linear Fit
[pGI50]exp=[pGI50]Cal
Linear Fit: [pGI50]cal = (0.95) [pGI50]exp + (0.24)
[pGI50]cal
[pGI50]exp


[pGI50]exp vs. [pGI50]Cal ; R=0.92057 ; R2= 0.82303
Test data
Best Linear Fit
[pGI50]exp=[pGI50]Cal
Linear Fit: [pGI50]cal = (0.86) [pGI50]exp + (0.75)
[pGI50]Cal
[pGI50]exp

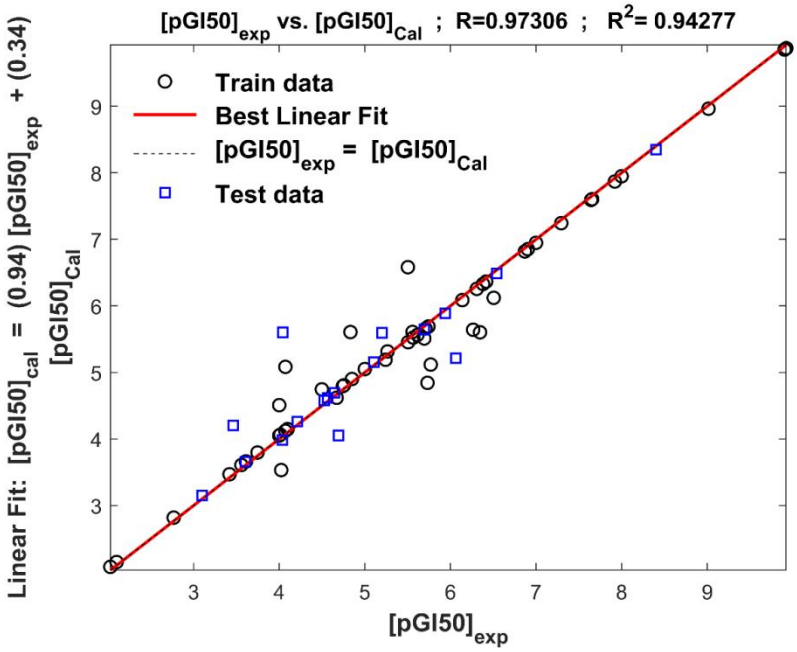


Figure V.6 : Courbes de régression SVM-DA

(a) apprentissage. (b) : test. (c) global

V.3.VALIDATION EXTERNE

Le calcul des critères de validation externe sur le jeu de test de modèle montre une bonne performance pour le DA_SVR (R^2 : 0.921) et des valeurs faibles des RMSEP (0.240). 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 DA_SVR a une meilleure performance en termes de mesure de validation externe telles que le CCC et les valeurs des RMSE indiquant que le modèle DA _SVR est plus précis. En conséquence le modèle DA_SVR a montré une bonne signification statistique et une bonne robustesse.

V.4. DOMAINE D'APPLICABILITE

V.4.1. Algorithme SVR-Dragonfly

Le graphe de William pour le modèle SVR-PSO (figure V.7) a démontré que près de 3% des molécules (soit 2 molécules) se trouvent en dehors de la plage des résidus standards (-3, 3), ce qui suggère que ces molécules présentent des valeurs aberrantes de réponse. L'une de ces molécules est intégrée au groupe d'apprentissage, tandis que l'autre fait partie du groupe de test.

Deux des molécules d'apprentissage, représentant environ 3% du total, ont dépassé la valeur seuil de 0,6 en termes de levier. Ces deux molécules présentent des structures aberrantes. La plupart des points étaient alignés avec les résidus standards zéro, tandis qu'un

petit nombre de points du groupe de test et du groupe d'apprentissage étaient dispersés dans le domaine spatial.

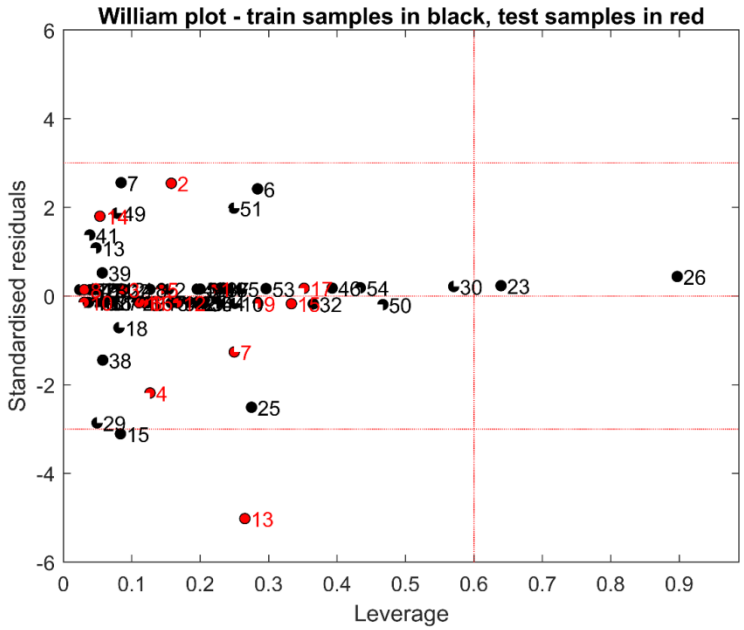


Figure V.7 :. Domaine d'applicabilité

V.5.EXPLOITATION DES RESULTATS ' DOCKING

V.5.1.ADMET

Pour évaluer la compatibilité avec un médicament des propriétés ADMET des 71 composés, la règle de Lipinski des Cinq a été prise en compte, La filtration a donné lieu à 58 molécules présentant des propriétés semblables à celles des médicaments et ont donc été conservées qui selon la règle de Pfizer indique une possible toxicité. Ainsi que les paramètres de toxicité montrent que les molécules ne sont pas toxiques, à l'exception des molécules numéros (13) et (48) Cela pourrait être dû à la fonction acide sulfonique qui est présentée en excès dans ces deux molécules et l'absence de la fonction d'oxygène, par contre dans les autres on ne trouve pas la fonction sulfonique en excès, et ça explique pourquoi la molécule (36) est parmi les bons résultats, c'est la seule molécule qui ne contient pas la fonction sulfonique.

V.5.2. Prédiction des propriétés pharmacocinétiques (ADMET) et de la Similitude avec les médicaments :

L'utilisation d'outils informatiques pour l'identification de nouveaux candidats médicaments permet de réduire les expérimentations et d'améliorer les taux de réussite. Dans ce contexte, la règle de Lipinski a été appliquée comme première étape de criblage pour évaluer la similarité avec les médicaments. Par la suite, un criblage secondaire a été effectué

en utilisant l'outil en ligne ADMET Lab. 2.0 pour calculer les propriétés ADMET en tant que mesure de la pharmacocinétique.

V.5.3. Préparation des structures :

Les 58 composés qui ont été filtrés par Swiss ADMET ont été soumis à une optimisation géométrique et à une minimisation d'énergie à l'aide de Data Warrior et Open Babel, en utilisant l'approche MFF94S.

V.5.4. Préparation du ligand et de la protéine et étude du Docking:

La structure cristallographique du récepteur V600E-BRAF en complexe avec le ligand vemurafenib (PDB CODE : 3OG7) a été obtenue depuis la banque de données PDB. Après avoir éliminé l'excès de molécules d'eau et la chaîne A de la protéine à l'aide de Chimera, les coordonnées tridimensionnelles du site actif ont été déterminées (x= 18.5167, y=7.79361 et z=10.3298). Ensuite, le docking a été réalisé en utilisant AutoDock Vina avec un script dans l'environnement Ubuntu/Linux, configuré pour un maximum de 1000 itérations et une taille d'évolution simplex de 50. La simulation du docking a été exécutée au minimum 40 fois pour les 10 poses, et les meilleures poses ont été sélectionnées en fonction des scores Mol Dock, Rerank, de l'interaction E et de la liaison E-H. De plus, Discovery Studio (DS) Visualizer v. 3.5 a été utilisé pour visualiser diverses interactions intermoléculaires telles que les liaisons H, les interactions hydrophobes et les liaisons halogènes

V.5.5. Interactions entre le récepteur et le ligand:

Les inhibiteurs V600E-BRAF optimaux ont été triés en fonction du score du Docking le plus faible, et la plupart des composés testés sont efficaces pour cibler V600E-BRAF. De plus, pour mieux comprendre l'excellent score du Docking observé, les inhibiteurs les plus puissants ont été étudiés avec V600E-BRAF. La conformation liée, des six composes est présentée dans les figures V.7-V.12 ainsi que les types d'interactions impliqués dans les modèles sélectionnées et les composés conçus.

La molécule 48 (Amsacrine) se lie au domaine V600E-BRAF avec une énergie de liaison libre de -10. 22 kcal/mol, Trois liaisons hydrogène ont été observées entre le récepteur et la molécule (ligand), dont deux étaient des liaisons hydrogène conventionnelles avec PHE595 et GLY596 de la protéine. De plus, une liaison C-H a été observée avec CLY593. De plus, VAL471, ALA481, LEU514, ILE463, PHE583 et CYS532 formaient des liaisons π-alkyle avec le ligand, tandis que TRP531 et PHE583 formaient une interaction empilée π - π avec le ligand. En outre, LYS483 formait une liaison π -cation

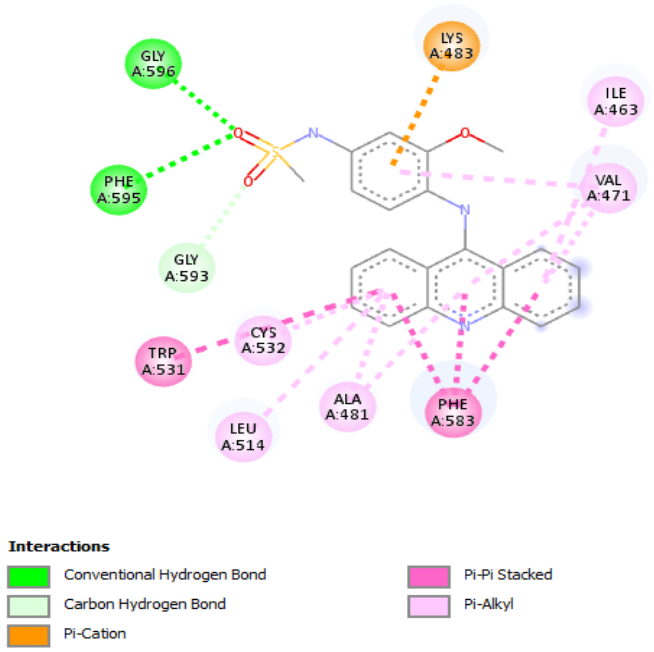


Figure V.8 : Diagramme des interactions 2D d'Amsacrine avecV600E-BRAF

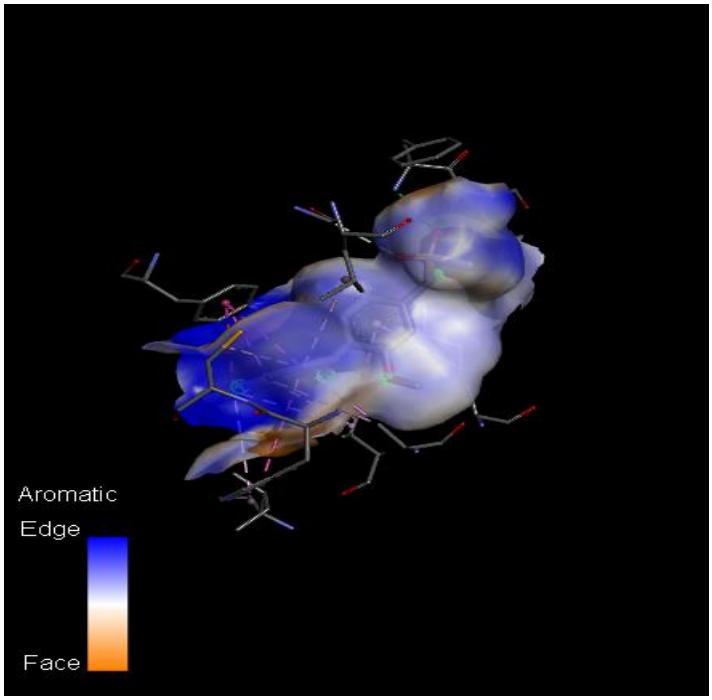


Figure V.9 : Vue en 3D du compose d'Amsacrine dans le site de liaison de V600-BRAF

La molécule 6 (Arnebin 1) présente une affinité de liaison élevée envers le domaine V600E-BRAF, avec une énergie de liaison libre mesurée à -9.865 kcal/mol, tel qu'il illustré sur la Figure V.10 Cette interaction est caractérisée par la présence de deux liaisons hydrogène conventionnelles entre le récepteur et le ligand. Plus spécifiquement, les résidus

ASP594 et ILE527 sont impliqués dans ces liaisons. Toutefois, il est important de noter que le résidu LYS483 forme des liaisons désavantageuses de type donneur-donneur et π -cation avec le ligand, le résidu THR529 établit une liaison π -sigma, contribuant ainsi à une interaction favorable entre le ligand et le domaine V600E-BRAF.

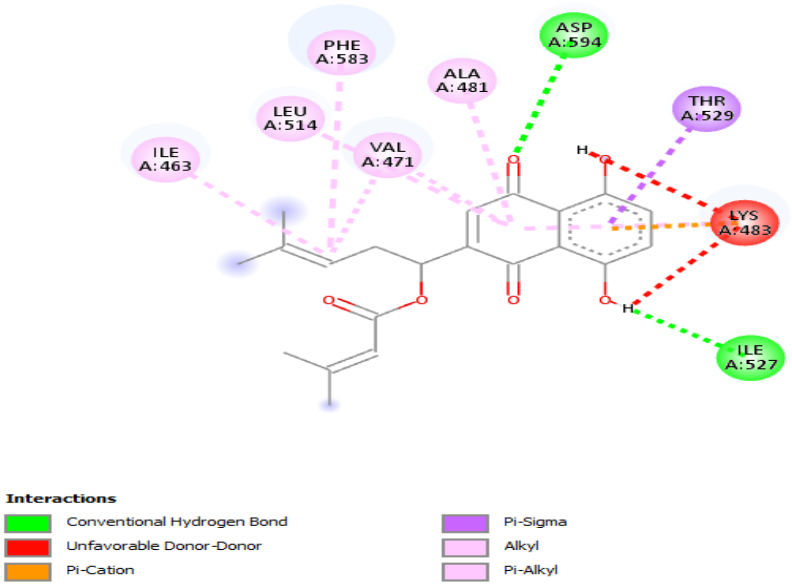


Figure V.10 : Diagramme des interactions 2D d'Arnebin 1 avecV600E-BRAF

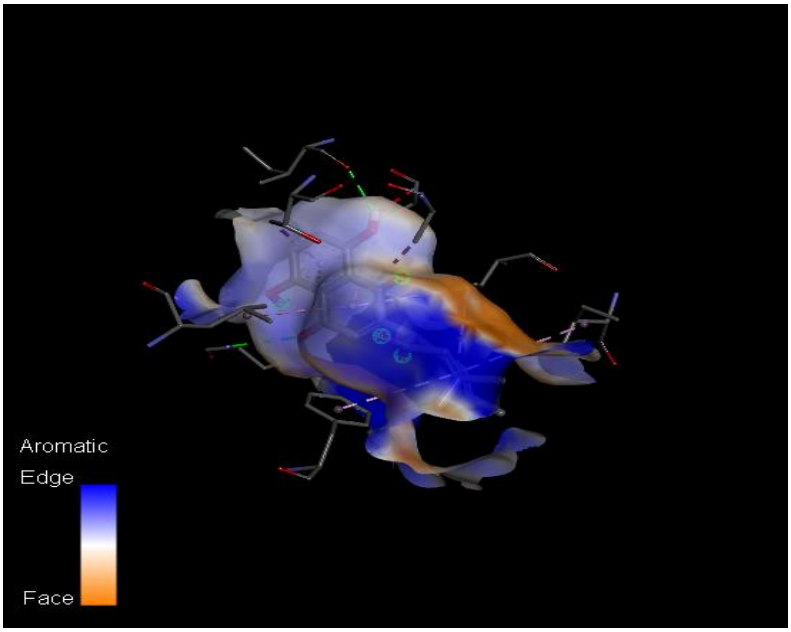


Figure V.11: Vue en 3D du composé d'Arnebin 1 dans le site de liaison de V600-BRAF.

La molécule 12 (Camptothecin Butylglycinate Ester Hydrochloride) présente une affinité de liaison avec le domaine V600E-BRAF, avec une énergie de liaison libre mesurée à -10.06 kcal/mol, comme illustré dans la Figure V.10. Lors de cette interaction, deux liaisons hydrogène conventionnelles ont été observées entre le récepteur et le ligand. Ces liaisons

impliquent les résidus THR529 et GLN530 de la protéine. De plus, il a été observé que la CYS532 forme des liaisons défavorables de type donneur-donneur, tandis que le résidu VAL471 établit une liaison π -sigma.

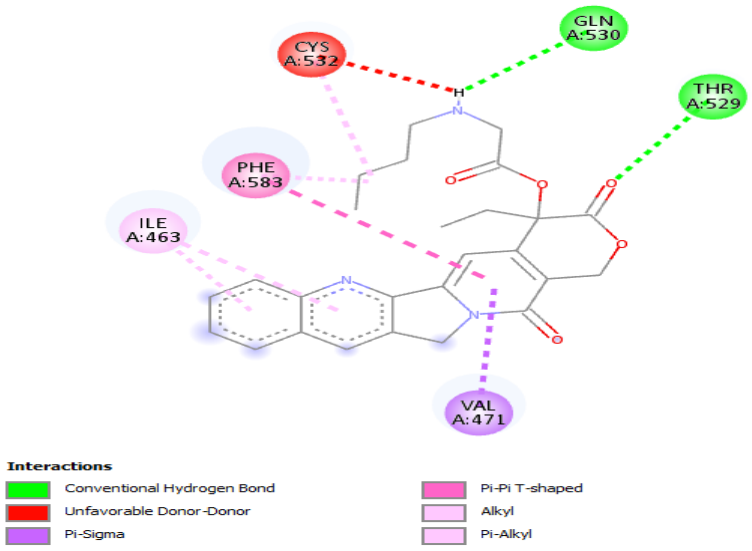


Figure V.12 Diagramme des interactions 2D de Camptothecin Butylglycinate Ester Hydrochloride avecV600E-BRAF

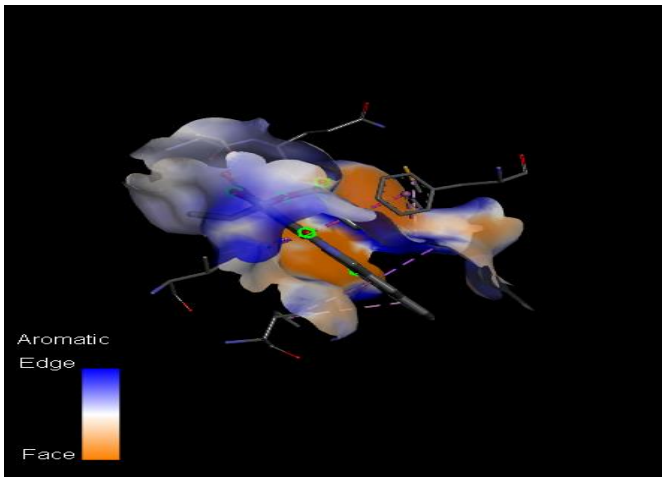


Figure V.13: Vue en 3D du compose de Camptothecin Butylglycinate Ester Hydrochloride dans le site de liaison de V600-BRAF..

La molécule 46 (Hydroxyurea) présente une affinité de liaison avec le domaine V600E-BRAF, avec une énergie de liaison libre mesurée à -3.356 kcal/mol, comme illustré sur la Figure V.14. Lors de cette interaction, trois liaisons hydrogène conventionnelles ont été observées entre le récepteur et le ligand. Ces liaisons impliquent les résidus ALA481, ILE527 et THR529 de la protéine.

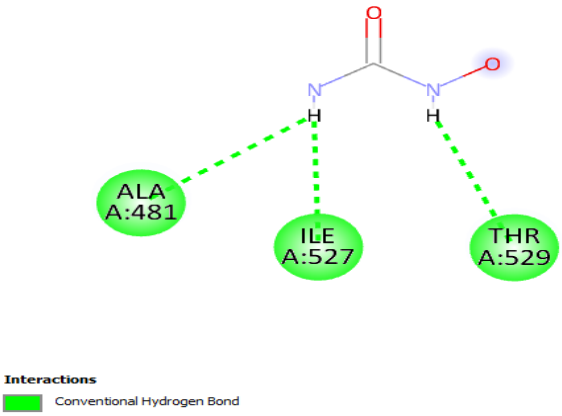


Figure V.14: Diagramme des interactions 2D de Hydroxyurea avecV600E-BRAF

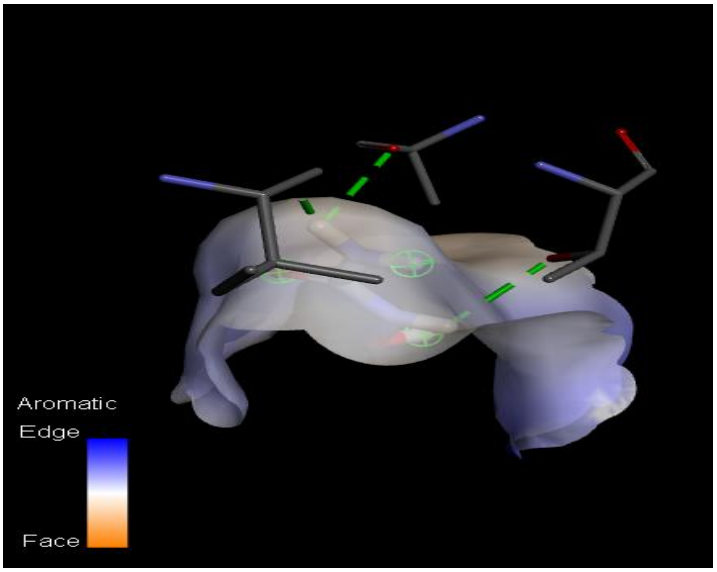


Figure V.15 : Vue en 3D du compose Hydroxyurea dans le site de liaison de V600-BRAF.

La molécule 45 (Hycanthone Mesylate) démontre une forte affinité de liaison envers le domaine V600E-BRAF, avec une énergie de liaison libre mesurée à -3.889 kcal/mol, comme indiqué sur la Figure V.16. Cette interaction se caractérise par la présence de trois liaisons hydrogène conventionnelles entre le récepteur et le ligand. Plus spécifiquement, les résidus ARG509 et ARG506 sont impliqués dans ces liaisons. De plus, il convient de souligner que les résidus LYS570 et LYS507 ont formé trois liaisons de type π-cation avec le ligand

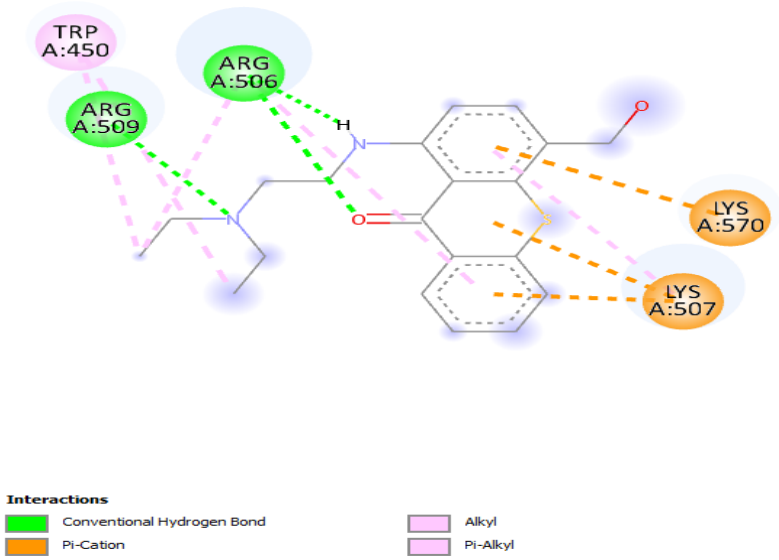


Figure V.16 : Diagramme des interactions 2D de Hycanthone MesylateavecV600E-BRAF

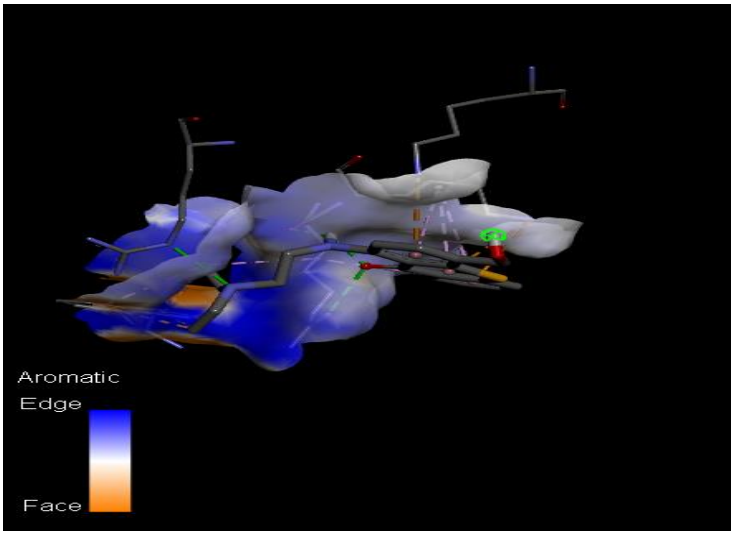


Figure V.17 : Vue en 3D du compose hycantone Mesylate dans le site de liaison de V600-BRAF.

La molécule 65 (thiotepa) présente une forte affinité de liaison envers le domaine V600E-BRAF, avec une énergie de liaison libre mesurée à -3.963 kcal/mol, tel qu'il est illustré sur la Figure V.18. Cette interaction est caractérisée par la formation de trois liaisons Hydrogène conventionnelles entre le récepteur et le ligand. Les résidus THR529 et GLN530 sont spécifiquement impliqués dans ces liaisons. De plus, il est important de souligner que le résidu PHE583 établit une liaison de type π-cation avec le ligand.

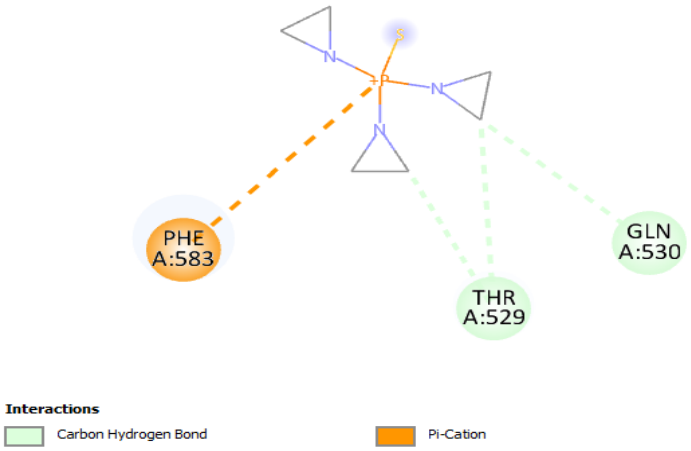


Figure V.18 : Diagramme des interactions 2D de thiotepa avecV600E-BRAF

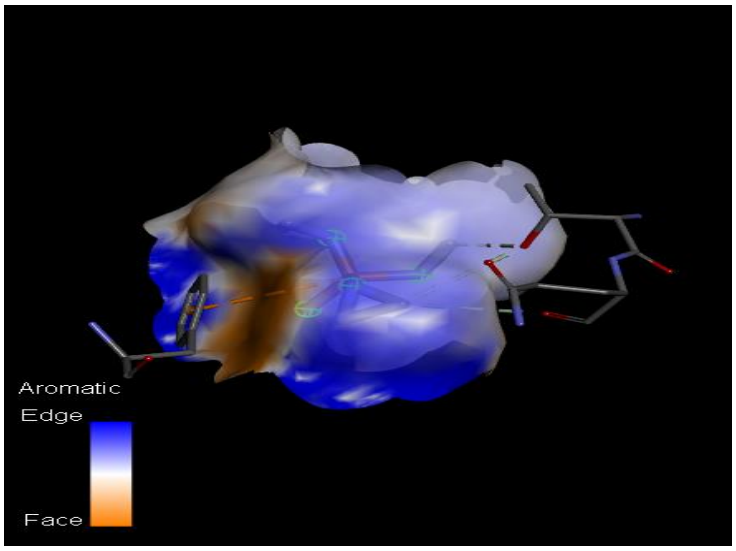


Figure V.19: Vue en 3D du composé Thiotepa dans le site de liaison de V600-BRAF.

CONCLUSION GENERALE

Cette étude vise à établir un modèle QSAR en utilisant des techniques d'intelligence artificielle pour prédire l'efficacité des nouvelles molécules. L'objectif principal est de développer un outil capable de prédire l'activité inhibitrice de ces composés. Pour ce faire, des méthodes d'intelligence artificielle ont été utilisées dans le processus de modélisation. L'accent est mis sur la prédiction précise de la concentration QSAR occupe une place essentielle dans la recherche de médicaments, offrant une approche largement adoptée pour prédire l'activité de molécules potentielles.

Cette méthode constitue une étape préliminaire cruciale dans le processus de découverte de médicaments. Grâce à QSAR, il devient possible d'évaluer et de prédire l'efficacité des composés avant les essais cliniques, accélérant ainsi le développement de nouveaux traitements. Ce travail présente, la modélisation de l'activité inhibitrice de nouvelles molécules synthétisées contre l'enzyme tumorale SK-MEL en utilisant une machine à vecteur de support (SVM), basée sur des descripteurs moléculaires.

La source de données utilisée était Pubchem, à partir de laquelle la base de données a été obtenue. Des descripteurs moléculaires en 2D ont été générés, et une sélection minutieuse de descripteurs pertinents a été effectuée, Cette sélection a abouti à l'identification de 10 descripteurs pertinents pour les 71 molécules considérées.

Le modèle QSAR développé dans cette étude est le SVM-DA, en utilisant l'environnement MATLAB. Les critères de validation internes et externes ont démontré une excellente performance du modèle SVM-DA en termes de productivité et de robustesse, avec des valeurs des R^2= [0,9613 0,8230, 0,9428] et des RMSE= [0.3275, 0.5027 ,0.3769]. Ces résultats confirment la fiabilité du modèle SVM-DA dans la prédiction de l'activité inhibitrice des molécules étudiées contre l'enzyme V600E-BRAF

Dans le cadre de ce mémoire, l'utilisation de SwissADME et d'AdmetLab 2.0, ainsi que la technique du docking, a permis d'approfondir la compréhension des interactions moléculaires. SwissADME s'est avéré être une ressource précieuse pour prédire les propriétés d'absorption, de distribution, de métabolisme et d'excrétion des composés. De plus, AdmetLab 2.0 a fourni des informations essentielles sur l'admettabilité des molécules, évaluant leur potentiel d'absorption, de toxicité et de biodisponibilité. L'intégration de ces outils avec le docking a permis d'étudier en détail les interactions spécifiques entre les ligands

et les cibles, ouvrant ainsi de nouvelles perspectives pour la conception de médicaments efficaces et sûr le modèle développé dans ce travail présente des applications prometteuses dans le domaine du docking moléculaire et du criblage virtuel, offrant ainsi des avantages pour la conception de médicaments innovants.

À l'avenir, il est envisageable d'explorer d'autres techniques de modélisation et d'optimisation émergentes, telles que les algorithmes génétiques et la logique floue, afin de prédire de nouveaux paramètres pertinents. Cette approche ouvrira de nouvelles perspectives de recherche et permettrait d'améliorer les performances du modèle, conduisant ainsi à la conception de médicaments plus efficaces.

REFERENCES BIBLIOGRAPHIQUES

[1] Leo Ghemtio. Simulation numérique et approche orientée connaissance pour la découverte de nouvelles molécules thérapeutiques. Autre. Université Henri Poincaré - Nancy 1, 2010.

[2] Chemsa I. Contribution à l'étude de l'évaluation biologique de co-traitement par la chimiothérapie et par l'Ephedra Alata chez des femmes cancéreuses de la région d'El Oued. 2020.

[3] Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, et al. Observatoire mondial du cancer : « Cancer Today ». Lyon : Centre international de recherche sur le cancer, 2020

[4] Cuenot A. Les Succédanés du Tabac (Note Préliminaire). In*:* Journal d'agriculture tropicale et de botanique appliquée, vol. 14, n°6-7, Juin-juillet1967. pp. 191-252.

[5] Cuenot A. Les Succédanés du Tabac (Note Préliminaire). In*:* Journald'agriculture tropicale et de botanique appliquée, vol. 14, n°6-7, Juin-juillet 1967. pp. 191-252.

[6] Ikram D. Le tabagisme et les maladies inflammatoires. Jun-2022:university center of abdalhafidboussouf- MILAhttp://dspace.centre-univ mila.dz/jspui/handle/123456789/2122

[7] Loubelo E. Impact des produits forestiers non ligneux (PFNL) sur l'économie des ménages et la sécurité alimentaire: cas de la République du Congo: Université Rennes 2; 2012

[8] Selponi Y. Apprendre les drogues en classe: La construction sociale des psychotropes par leurs effets lors des actions de prévention des conduites addictives en milieu scolaire. Agora. 2019(1):109-21.

[9] Delavigne V, editor. Les mots des patients atteints de cancer. Socioterminologie et pratique professionnelle en santé. Les langues de spécialité en question : perspectives d'étude et applications 12ème journée scientifique de la CRL; 2007 2007-11-17; Paris, France;

[10] Perry MC. Prise en charge d'un patient atteint de cancer. Goldman's Cecil Medicine Cancérologie. 2015:1.

[11] François Lucia. Développement d'un modèle prédictif de réponse thérapeutique et de toxicité dans les cancers du col de l'utérus : de l'imagerie à la clinique. Médecine humaine et pathologie. Université de Bretagne occidentale - Brest, 2021.

[12] Marion Torres. Statut nutritionnel de la personne âgée vivant à domicile : prévalence, facteurs associés et conséquences. Santé publique et épidémiologie. Université de Bordeaux, 2014.

[13] Kaya M, Bladhane B, Aggoune A, Ayat A. Dépistage du cancer du col de l'utérus. 2022.

[14] Johanna Galmès. Isolement et caractérisation de nouvelles espèces de Torque Teno Mini Virus (TTMV) : implication potentielle dans la pathogenèse de la pneumonie. Biologie cellulaire. Ecole normale supérieure de lyon - ENS LYON, 2013.

[15] Service de neurochirurgie, CHU Ibn-Rochd, BP 20100, Casablanca, Maroc
Received 5 December 2007, Accepted 22 October 2008, Available online 17 December 2008.

[16] Sommer-Bühler J, Dozier C, Stalder H. Les céphalées. PrimaryCare. 2004;4(12).

[17] DelAmo Sanchez. P..*et al.*, (The BABAR Collaboration) Phys. Rev. D 82, 011101(R) – Published 9 July 2010

[18] Binet Alfred. C.-A. Strong, La psychologie de la douleur ; Herbert Nichols, Nerfs de la douleur ; L. Oppenheimer, Douleur et sensationsthermiques.In: L'année psychologique. vol. 2. pp. 706-710, 1895.

[19] Li, Y., Chiocca, A., Numerical modelling and in-situ radiographic study of the grain nucleation and growth of inoculated aluminum alloys IOP Conf. Ser.: Mater. Sci.Eng. 84, 2015.

[20] Laudy D. traitements expérimentaux dans le cas des enfants atteints de cancer. Laval théologique et philosophique. 1998;54:277-305.

[21] E. Stoeckle,Chirurgie des sarcomes des tissus mous des membres et de la paroi du tronc,Cancer/Radiothérapie,Volume 10, Issues 1–2,2006,

[22] Moulinié V. La chirurgie des âges: corps, sexualité et représentations du sang: Les Editions de la MSH; 1998.

[23] Santos, J.S., Conceição, M., Costa Carlito, O., Nascimento Sobrinho, L., da Conceição, M., Da SILVA, M MELO, Anthropometric profile and food intake of

adolescents in Teixeira de Freitas - Bahia, Brazil Rev. Nutr., Campinas, 18(5):623-632, set./out., 2005

[24] El Bousaadani, A., Eljahd, L., Abada, R., Rouadi, Service d'ORL et de chirurgie cervicofaciale, hôpital du 20-Août-1953, CHU Ibn-Rochd, 20240 Casablanca, Maroc
[25] Pérol M, Arpin D. Traitement de maintenance dans les cancers bronchiques non à petites cellules métastatiques: une révolution conceptuelle? La Presse Médicale. 2011;40(4):404-14. Daniel YC Heng

[26] Déchelette M, Krakowski I, Arbiol É, Chvetzoff G, Cullet D, Delorme T, et al. Douleur et cancer: informations à l'usage des personnes malades et de leurs proches. Douleur et analgésie. 2007;20(4):272-81.

[27] Marieb E, Hoehn K. Anatomie et physiologie humaines: Livre+ eText+ plateforme numérique MonLab-Licence étudiant 60 mois: Pearson Education France; 2014.

[28] Beylot-Barry, M., Dereure, O., Vergier, B., Barete, S., Prise en charge des lymphomes T cutanés : recommandations du Groupe français d'étude des lymphomes cutanés, Annales de Dermatologie et de Vénéréologie, Volume 137, Issue 10,2010,

[29] Bonvalot S, Vanel D, Terrier P, Le Pechoux C, Le Cesne A. Principes du traitement des sarcomes des tissus mous de l'adulte. EMC-Rhumatologie-Orthopédie. 2004;1(6):521-41.

[30] Corde, S. Développement méthodologique de l'application d'agents pharmacologiques renforçateurs de l'effet photoélectrique pour l'utilisation du rayonnement synchrotron en radiothérapie anticancéreuse. Ingénierie biomédicale. Université Joseph-Fourier - Grenoble I, 2002.

[31] Pignot, P. Sargos, Place de la radiothérapie adjuvante après cystectomie pour les tumeurs de vessie infiltrant le muscle, Progrès en Urologie, Volume 31, Issue 3, 2021,

[32] (2009). La radiothérapie. In: Les cancers. Comprendre, prévenir, traiter. Springer, Paris.
[33] Patrick Vassal. Fusion d'images multi-modales pour la radiothérapie conformationnelle : application au repositionnement du patient. Ingénierie biomédicale. Université Joseph-Fourier - Grenoble I, 1998.

[34] Pauline Gilson. Caractérisation in vitro et in vivo de nouveaux agents pyrrolo-pyrimidine ciblant les microtubules pour le traitement des cancers. Médecine humaine et pathologie. Université Grenoble Alpes, 2017.

[35] Aline Mongis. Une nouvelle stratégie d'immunothérapie : cibler directement des immunostimulants à la surface des cellules tumorales par ligation bio-orthogonale. Biologie cellulaire. Université d'Orléans, 2017.

[36] Despiau, F., Zagala, Y., Delord, J.P., Montastruc, M., Étude observationnelle du temps passé en hôpital de jour selon la voie d'administration (intraveineuse vs sous-cutanée) d'une thérapie ciblée, Bulletin du Cancer, Vol,104, Issue 10, 2017,

[37] Burstein H.J, Temin S, Anderson H, Buchholz T.A, Davidson N.E, Adjuvant endocrine therapy for womenwith hormone receptor-positive breast cancer: Journal of Clinical Oncology. 2014.

[38] Dossier réalisé en collaboration avec Eric Vivier, équipe Cellules innées lymphoïdes, Centre d'immunologie de Marseille-Luminy (unité 1104 Inserm/CNRS/Aix-Marseille Université), PU-PH Aix Marseille Université/Assistance publique des hôpitaux de Marseille, directeur scientifique d'Innate-Pharma
.

[39] Publique, T.S. Contribution à l'étude de l'efficacité des interventions Domotiques et Téléassistances dans la prévention des chutes à domicile des personnes âgées en perte d'autonomie et Alzheimer au stade léger amodéré: Université de limoges; 2013.

[40] Elodie W. Le grenadier (PUNICA GRANATUM): Plante historique et évolutions. 2009.

[41] Pottier As. Étude de l'implication de la protéine matricellulaire SPARC et des fibroblastes du microenvironnement lymphatique dans la résistance thérapeutique des mélanomes: Nice; 2015.

[42] Lièvre, A., Puig, P.L. Bronchiolite : qu'est-ce que c'est ? Hôpital Ambroise Paré, 9, avenue Charles de Gaulle, F-92100 Boulogne-Billancourt, 2023

[43] Sangaré M.F. Etude des aspects épidémiologiques, cliniques et thérapeutiques des cancers gynécologiques et mammaires au centre de santé de référence de la Commune V du district de Bamako. 2003.

[44] Mcconville B.J, Steichen-Asch P, Harris R, Neudorf S, Sambrano J, Lampkin B, et al. Pediatric bone marrow transplants :psychological aspects. The Canadian Journal of Psychiatry. 1990;35(9):769-75.

[45] Alzeeb, G., Développement d'un modèle cellulaire flexible tridimensionnel : un focus sur le cancer gastrique., École doctorale Biologie-Santé (Nantes) , en partenariat avec Génétique, génomique fonctionnelle et biotechnologies, thèse de doctorat Soutenue le 28-01-2022 à Brest

[46] Duchemann, B. Intelligence artificielle et santé publique: normes, savoirs et appropriations. Médecine humaine et pathologie. Université Paris Cité, 2021.

[47] Bleunven A. Contribution à la conception d'un système d'imagerie polarimétrique en vue d'applications pour la détection précoce du mélanome: École doctorale Santé, en partenariat avec Ingénierie de la vision, thèse de doctorat Soutenue le 01-12-201628-01-2022 à Brest.

[48] Yable, K.D., Caractérisation optique du Mycobacterium Ulcerans dans le cadre du diagnostic précoce de l'ulcère de Buruli, Thèse du doctorat de l'institut national polytechnique félix houphouët-boigny de yamoussoukro, soutenue le 03/11/ 2022

[49] Kassab, T. Stratégies thérapeutiques actuelles dans la prise en charge des mélanomes et éducation thérapeutique du patient, de l'hôpital à l'officine. Sciences pharmaceutiques. 2015.

[50] Le Gay, D. Expériences, opinions et attentes des médecins généralistes en matière de prévention solaire et de dépistage du mélanome dans les Alpes-Maritimes : Médecine humaine et pathologie. 2015.

[51] Berthé B. Activités de cytologie à l'unité d'oncologie du département de pédiatrie du CHU Gabriel Touré de Bamako: USTTB; 2022.

[52] Bouhedjar, K., Messadi, D., Nasr-Eddine, C., Une étude QSAR basée sur la méthode préconisée par l'OCDE: Une comparaison entre les deux modèles Free-logP et logP-Based. 2012.

[53] Patani, G.A., LaVoie, E.J., Bioisosterism: A Rational Approach in Drug Design.," Chem. Rev., vol. 96 8, pp. 3147–3176, 1996.

[54] Brown, A.C., Fraser, T.R, On the Connection between Chemical Constitution and Physiological Action. Part II.—On the Physiological Action of the Ammonium Bases derived from A tropia and Conia," Trans. R. Soc. Edinburgh, vol. 25, no. 2, pp. 693–739, 1869,

[55] Ghasemi, F., Mehridehnavi, A., Pérez-Garrido, A., Neural network and deep-learning algorithms used in QSAR studies :merits and drawbacks.," Drug Discov. Today, vol. 23, no. 10, pp. 1784–1790, Oct. 2018,

[56] Roy, K., Kar, S., Das, R., Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment. 2015.

[57] Nantasenamat, C. Isarankura, C. , Ayudhya, N., Naenna, T., A practical overview of quantitative structure-activity relationship," 2009.

[58] Klebe, V., Abraham, U., Mietzner, T., Molecularsimilarity indices in a comparative analysis (CoMSIA) of drugmolecules to correlate and predicttheirbiologicalactivity.," J. Med. Chem., vol. 37 24, pp. 4130–4146, 1994.

[59] Chtita, S., Mohammed,G., Lakhlifi, T., Modélisation de molécules organiques heterocycliques biologiquement actives par des méthodes QSAR / QSPR . Recherche de nouveaux médicaments HAL Id August, 2017.

[60] Jaworska J, Nikolova-Jeliazkova N, Aldenberg T. QSAR applicabilty domain estimation by projection of the training set descriptor space: a review. Altern Lab Anim. Oct;33(5):445-59, 2005

[61] Lévêque, L., Tahiri, N., Goldsmith, M.R., Verner, M.A., Quantitative Structure-Activity Relationship (QSAR) modeling to predict the transfer of environmental chemicals across the placenta,Computational Toxicology, Volume 21, 100211, 2022.

[62] Sushko, Y., Novotarskyi, S., Körner, R., Vogt, J., Abdelaziz, A., Tetko, I. Prediction-driven matched molecular pairs to interpret QSARs and aid the molecular optimization process.," J. Cheminform., vol. 6, no. 1, p. 48, 2014,

[63] Hamadache, M., Benkortbi, O., Hanini, S., Amrane, A., Khaouane, L., A Quantitative Structure Activity Relationship for acute oral toxicity of pesticides on rats: Validation, domain of application and prediction," J. Hazard. Mater., vol. 303, pp. 28–40, 2016,.

[64] Frisch, M.J. Gaussian 03, Revision D.01," Gaussian, Inc., Wallingford, CT, Jan. 2004.

[65] Roy, P.P., Paul, S., Mitra, I., Roy, K., On two novel parameters for validation of predictive QSAR models," Molecules, vol. 14, no. 5, pp. 1660–1701, 2009,

[66] Zhang, L., Zhu, H., Oprea, T.I., Golbraikh, A., Tropsha, A., QSAR modeling of the blood-brainbarrierpermeability for diverse organic compounds.," Pharm. Res., vol. 25, no. 8, pp. 1902–1914, Aug. 2008.

[67] He, L., Jurs, P.C., Assessing the reliability of a QSAR model'spredictions," J. Mol. Graph. Model., vol. 23, no. 6, pp. 503–523, 2005,

[68] Verma, J., Khedkar, V.M., Coutinho, E.C., 3D-QSAR in drug design--a review." Curr. Top. Med. Chem., vol. 10, no. 1, pp. 95–115, 2010,

[69] Golbraikh A, Tropsha A. Beware of q2! J Mol Graph Model. Jan; 20(4):269-76, 2002.

[70] Tetko, I.V., Critical assessment of QSAR models of environmental toxicity against Tetra hymenapyriformis :focusing on applicability domain and over fitting by variable selection.," J. Chem. Inf. Model., vol. 48, no. 9, pp. 1733–1746, Sep. 2008,.

[71] Tham, S.Y., Agatonovic-Kustrin,S., Application of the artificial neural network in quantitative structure-gradient elution retention relationship of phenyl thio carbamyl amino acids derivatives.," J. Pharm. Biomed. Anal., vol. 28 3–4, pp. 581–590, 2002.

[72] Netzeva, T.I., CurrentStatus of Methods for Defining the Applicability Domain of (Quantitative) Structure-Activity Relationships," Altern. to Lab. Anim., vol. 33, pp. 155–173, 2005.

[73] Buatois S, Le Merdy M, Labat L, Scherrmann J-M, Decleves X. Principales modifications pharmacocinétiques chez l'enfant. Toxicologie analytique et clinique.;26 (3):156-64. 2014.

[74] Renvez G. Application des modes statiques à l'étude de la flexibilité des protéines: vers un processus de docking: Université Paul Sabatier-Toulouse III; 2010.

[75] Rouaissia, H., Etude par Docking Moléculaire de l'activité inhibitrice des Furano coumarines Contenues dans le Jus de Pamplemousse sur les Cytochromes CYP3A4. 2020.

[76] Ikram, A.R. Etude 3D-QSAR et docking moléculaire de dérivés de pyrazoline l'activité antiamibienne contre la souche HM1: IMSS d'E. histolytica 2022.

[77] Arrault A. Stratégies de docking-scoring assistées par analyse de données. Application au criblage virtuel des cibles thérapeutiques COX-2 et PPAR gamma: Université d'Orléans; 2007.

[78] Brut M. Nouvelle approche méthodologique pour la prise en compte de la flexibilité dans les interactions entre moleculesbiologiques: les Modes Statiques: Université Paul Sabatier-Toulouse III; 2009.

[79] Chtita S. Modélisation de molécules organiques hétérocycliques biologiquement actives par des méthodes QSAR/QSPR. Recherche de nouveaux médicaments: Université Moulay Ismaïl, Meknès; 2017.

[80] Mbai, J.S., U Migration during Weathering of Urani ferrous Syenite on Tropical Climate: Case of Ngombas in the Southern Cameroon," J. Geosci. Geomatics, vol. 10, no. 2, pp. 99–111, May 2022

[81] Schölkopf, B., Smola, A.J., Learning withKernels: Support Vector Machines, Regularization, Optimization, and Beyond.” The MIT Press, Jun. 05, 2018.

[82] Vapnik, V., The Nature of Statistical Learning Theory,” in Statistics for Engineering and Information Science, vol. 8, , pp. 1–15, 2000.

[83] Smola, A.J., Schölkopf, B., A tutorial on support vector regression,” Stat. Comput., vol. 14, no. 3, pp. 199–222, 2004,

[84] Tharwat, A., Classification assessmentmethods,” Appl. Comput. Informatics, vol. 17, no. 1, pp. 168–192, 2018,

[85] Liansheng, H., Din, Z.M., Lai, Y.M., Spatial distribution and temporal dynamics of Frankliniella occidentalis Pergande, 1895 and Thrips palmi Karny, Life Excit. Biol., vol. 1, no. 4, pp. 176–196, 2013.

[86] Zhang, N., Shen, S.L., Zhou, A.N., Chen, J., A brief report on the explosions at a chemical factory in Xiang shui, China, Process Saf. Prog., vol. 38, no. 2, 2019.

[87] Ibrir, A., Kerchich, Y., Hadidi, N., Merabet, H., Hentabli, M., Prediction of the concentrations of PM1, PM2.5, PM4, and PM10 by using the hybrid dragon fly-SVM algorithm,” Air Qual. Atmos. Heal., vol. 14, no. 3, pp. 313–323, 2021,

[88] Lauer F, Bloch G, Méthodes SVM pour l'identification. Journées Identification et Modélisation Expérimentale; Nov. 2006, Poitiers, France. pp. 0110344.