



Université Mohamed Khider de Biskra
Faculté des sciences exactes et des sciences naturelles et de la nature
et de la vie
Département des sciences de la matière

MÉMOIRE DE MASTER

Domaine sciences de la matière
Filière Chimie
Spécialité chimie pharmaceutique

Réf. :

Présenté et soutenu par :
Fatma BOUDAUCHE

Le :

Techniques de modélisation moléculaires appliquées à l'étude et l'optimisation des inhibiteurs de la protéine Bcl-2

Jury :

M.	Salah BELAIDI	Prof	Université Med Khider de Biskra	Président
M ^{me} .	Imane Almi	MCB	Université Med Khider de Biskra	Rapporteur
M ^{me} .	Saida KHAMOULI	MCA	Université Med Khider de Biskra	Examineur

Année universitaire : 2023/2024.

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

REMERCIEMENTS

Tout d'abord, je tiens à remercier ALLAH le Tout-Puissant de m'avoir donné la santé, la volonté, le courage et la patience pour mener à bien ma formation et réaliser ce travail de recherche.

*Tout d'abord, ce travail ne serait pas aussi riche et n'aurait pu voir le jour sans l'aide et l'encadrement de **Dr. Almi Imane**. Je la remercie pour la qualité exceptionnelle de son encadrement, pour sa patience, sa disponibilité, sa gentillesse et surtout ses bons conseils afin d'élaborer ce mémoire dans de bonnes conditions.*

Je tiens également à remercier les membres du jury de thèse pour avoir pris le temps de lire ce manuscrit ainsi que pour les remarques qu'ils m'adresseront lors de cette soutenance afin d'améliorer.

*J'exprime ma profonde gratitude et mon respect à **M. Melkmi Nadjib** pour son soutien indéfectible et son dévouement constants à notre égard, ainsi que pour son traitement bienveillant, ses conseils précieux et sa grande patience. Les leçons que j'ai acquises grâce à lui au fil des années ont été d'une valeur inestimable.*

Je ne saurai jamais assez remercier la lumière de ma vie, mes parents, qui, par leurs prières et leurs encouragements, m'ont permis de surmonter tous les obstacles.

Enfin, je tiens à remercier chaleureusement les enseignants et le personnel du Département de Chimie pour leur dévouement et leur abnégation. Je n'oublie pas non plus mes collègues de notre promotion, dont le soutien et la camaraderie ont été essentiels tout au long de cette aventure académique.

DÉDICACE

Du profond de mon cœur, je dédie ce travail à tous ceux qui me sont chers.

À la personne qui s'est toujours sacrifiée pour me voir réussir, à toi, mon père,

Je tiens à vous exprimer ma gratitude d'avoir cru en mes capacités, malgré mes limites et mes faiblesses. Votre confiance en moi a été une lumière dans les temps sombres et votre présence constante a été mon rocher. Je vous offre ce certificat avec un cœur plein de gratitude.

Je remercie aussi Dieu de t'avoir choisi, mon père, parmi toute l'humanité. Vous êtes pour moi une source inestimable de soutien et de force.

À ma chère mère,

Merci infiniment pour le soutien constant et l'amour inconditionnel que vous m'entourez depuis mon enfance. Que votre bénédiction accompagne toujours mes pas. J'espère que cette humble œuvre sera la réalisation de vos vœux les plus chers et le fruit de vos innombrables sacrifices. Que Dieu vous accorde santé, bonheur et longue vie.

À mes frères, Omar, Abd al Rahman, Bissan, Kawthar.

À mes tantes, A.B et H.B

À mon seul oncle, qui était comme un père pour moi ; merci de votre soutien et de vos encouragements constants.

À ma cousine, Iatidale, qui a toujours été à mes côtés tout au long de cette recherche.

*À mes amies proches qui étaient comme des sœurs pour moi : Manar,
Warda et Soumaia.*

*À tous mes collègues de deuxième année de Master du Département de
Chimie Pharmaceutique.*

Résumé

Le présent travail propose une technique fiable, stable et prédictive de la relation quantitative structure-activité (QSAR) pour des séries de molécules anticancéreuses, élaborées à partir d'une base de données de 108 molécules inhibitrices de la protéine Bcl-2. Nous avons employé une approche in silico pour prédire les propriétés physicochimiques des molécules à l'aide de l'outil web SwissADME, afin de sélectionner les molécules en fonction de leur propriétés pharmacocinétique.

Différents descripteurs moléculaires ont été calculés à l'aide de la plateforme web gratuite ChemDes. Le logiciel RapidMiner a été utilisé pour sélectionner les descriptions appropriées parmi 1135 d'entre elles. L'ensemble des données a été soumis à des études statistiques (MLR) via le logiciel SPSS. Les modèles obtenus ont été validés par une méthode de validation interne.

Une forte corrélation entre les valeurs des activités expérimentales et prédites a été observée, indiquant la fiabilité, la robustesse et la qualité de modèle obtenue. Ce modèle pourra être appliqué avec succès pour prédire les affinités des nouveaux composés, dont les données expérimentales sont indisponibles.

Mots-clés : QSAR, MLR, Bcl-2, Cancer, ADME.

Abstract

The present work proposes a reliable, stable and predictive technique of the quantitative structure-activity relationship (QSAR) for series of anticancer molecules, developed from a database of 108 inhibitor molecules of the Bcl-2 protein. We employed an in silico approach to predict the physicochemical properties of molecules using the SwissADME web tool, in order to select molecules based on their pharmacological properties.

Different molecular descriptors were calculated using the free web platform ChemDes. RapidMiner software was used to select the appropriate descriptions from 1135 of them. All data was subjected to statistical studies (MLR) via SPSS software. The models obtained were validated by an internal validation method.

A strong correlation between the experimental and predicted activity values was observed, indicating the reliability, robustness and model quality obtained. This model can be successfully applied to predict the affinities of new compounds, for which experimental data are unavailable.

Keywords: QSAR, MLR, Bcl-2, Cancer, ADME.

Table des matières -----	I
Table des figures -----	V
Liste des tableaux -----	VII
Introduction générale -----	01

CHAPITRE 1: L'APOPTOSE ET LE MORT CELLULAIRE

I. 1. Partie 01 : Maladie du cancer -----	05
I. 1.1. Introduction -----	05
I. 1.2. Le cancer -----	05
I. 1.2.1. La cellule cancéreuse -----	06
I. 1.3. Type de cancer -----	07
I. 1.3.1. Cancer du sein -----	07
I. 1.3.2. Cancer de l'estomac -----	08
I. 1.3.3. Carcinome hépatocellulaire (le cancer du foie) -----	09
I. 1.3.4. Cancer de la prostate -----	10
I. 1.4. Pathogenèses du cancer -----	11
I. 1.4.1. Les risques liés aux comportements -----	11
I. 1.4.1.1. Tabagisme -----	11
I. 1.4.1.2. Alimentation -----	11
I. 1.4.1.3. L'alcool -----	12
I. 1.4.2. Les risques liés à l'environnement -----	12
I. 1.4.2.1. La pollution de l'air -----	12
I. 1.4.2.2. Les infections -----	13
I. 1.4.3. Les risques liés à l'individu et son histoire -----	14
I. 1.4.3.1. L'âge avancé -----	14
I. 1.4.3.2. Les hormones -----	14
I. 1.4.3.3. Origine héréditaire -----	14

I. 1.4.3.4. Maladies inflammatoires ou auto-immunes -----	14
I. 1.4.3.5. Traitements -----	15
I. 1.5. Les principales options thérapeutiques -----	15
I. 1.5.1. La chirurgie -----	15
I. 1.5.2. La radiothérapie -----	16
I. 1.5.3. Les traitements médicamenteux du cancer -----	16
I. 1.5.3.1. La chimiothérapie -----	16
I. 1.5.3.2. L'hormonothérapie -----	16
I. 1.5.4. Médecine de précision -----	17
I. 1.5.4.1. Thérapies ciblées et immunothérapie spécifique -----	17
I. 2. Partie 02 : La famille des protéines BCL-2 -----	18
I .2.1. Introduction -----	18
I .2.2. Le cycle cellulaire -----	18
I .2.2.1. Les phases du cycle cellulaire -----	19
I .2.3. La famille des protéines BCL-2 -----	24
I .2.3.1. Classification des protéines de la famille Bcl-2 -----	25
I .2.3.1.1. Les protéines anti-apoptotiques -----	25
I .2.3.1.2. Les protéines pro-apoptotiques multi-domaines -----	26
I .2.3.1.3. Les protéines pro-apoptotiques à BH3 -----	26
I .2.4. Structure et Localisation -----	27
I .2.5. Implication de la famille Bcl-2 dans la cancer -----	28

CHAPITRE 02 : LES TECHNIQUES DE MODELISATION MOLECULAIRE

II .1. Introduction -----	36
II .2. Le criblage virtuel -----	37
II .2.1. Criblage virtuel « ligand-based » -----	39
II .2.1.1. Recherche de similarité -----	39
II .2.1.2. Modèles pharmacophoriques « ligand-based » -----	40
II .2.1.3. Quantitative Structure-Activity Relationship (QSAR) « ligand-based » --	40
II .2.1.3.1. Les étapes principales pour la construction du modèle QSAR -----	42
II .2.1.3.1.1. Collecte et traitement de l'ensemble des données -----	42

TABLE DES MATIERES

II .2.1.3.1.2.	Génération sur les Descripteurs moléculaires -----	43
II .2.1.3.1.3.	Sélection des descripteurs -----	44
II .2.1.3.1.4.	Construction du modèle QSAR -----	45
A.	Méthodes Statistiques -----	45
A .1.	La régression linéaire multiple (RLM) -----	45
A .2.	Régression des moindres carrés partiels (PLS) -----	46
A .3.	Régression non linéaire multiple (RNLM) -----	46
B.	Test de la signification globale des methodes statistiques ---	47
B .1.	Coefficient de détermination R^2 -----	47
B .2.	Coefficient de détermination ajusté R^2_{adj} -----	47
B .3.	L'erreur quadratique moyenne « MSE » et l'erreur type résiduel « s » -----	47
B .4.	Le test de Fisher F -----	48
II .2.1.4.	Validation du modèle QSAR -----	49
II .3.1.4.1.	Validation interne -----	49
A.	Validation croisée -----	49
B.	Y-randomisation -----	50
II .3.1.4.2.	Validation externe -----	51
II .2.1.5.	Domaine d'applicabilité -----	51
II .2.2.	Criblage virtuel « structure-based » -----	52
II .2.2.1.	Sélection d'un site de liaison -----	52
II .2.2.2.	Prédiction du site de liaison -----	52
II .2.2.3.	Modèles pharmacophoriques « structure-based » -----	52
II .2.2.4.	La relation quantitative structure activité (3D-QSAR) -----	53
II .2.2.5.	La conception de novo -----	53
II .2.2.6.	Méthodes de docking -----	54
II 3.2.6.1.	Types du Docking moléculaire -----	55
A.	Docking avec ligand rigide -----	55
B.	Docking avec ligand flexible -----	56
C.	Docking semi-flexible -----	56
II .3	Les propriétés ADME -----	56

CHAPITRE 03 : RESULTATS ET DISCUSSION

III .1.	Introduction -----	64
III .2.	Matériels et méthodes -----	65
III .2.1.	Ensemble de données -----	65
III .2.2.	Calculs des descripteurs -----	76
III .2.3.	Sélection des descripteurs -----	77
III .2.4.	Développement de modèle QSAR -----	79
III .2.5.	Validation de modèle QSAR -----	81
	Conclusion générale -----	87

CHAPITRE 1: L'APOPTOSE ET LE MORT CELLULAIRE

I. 1. Différences dans le mécanisme de division cellulaire entre une cellule normale et cancéreuse -----	07
I. 2. Cancer du sein -----	08
I. 3. Cancer gastrique (cancer de l'estomac) -----	08
I. 4. Carcinome hépatocellulaire (le cancer du foie) -----	10
I. 5. Cancer de la prostate -----	10
I. 6. Représentation schématique du cycle cellulaire -----	19
I. 7. Représentation schématique de L'interphase -----	21
I. 8. Représentation schématique de la prophase -----	22
I. 9. Représentation schématique de la métaphase -----	23
I. 10. Représentation schématique de l'anaphase -----	23
I. 11. Représentation schématique de la télophase -----	24
I. 12. Structure des protéines de la famille Bcl-2. -----	25

CHAPITRE 02 : LES TECHNIQUES DE MODELISATION MOLECULAIRE

II .1. Classification des méthodes de criblage virtuel « ligand-based » et « structure-based » --	38
II .2. Principales étapes de l'élucidation d'un pharmacophore -----	40
II .3. Schéma général des différentes étapes d'une étude QSAR -----	41
II .4. Etapes effectuées pour le développement d'un modèle QSAR -----	42
II .5. Génération de descripteurs moléculaires indispensables pour établir la relation QSAR --	44
II .6. Méthodes de validation du modèle QSAR -----	49
II .7. Représentation schématique des étapes principales d'une conception de novo -----	54
II .8. Représentation schématique du Docking moléculaire -----	55
II .9. Model de docking moléculaire (docking flexible) -----	56

CHAPITRE 03 : RESULTATS ET DISCUSSION

III .1. Méthodologie générale utilisée dans l'étude QSAR -----	65
III .2. Graphique « BOILED-Egg » -----	71

III .3. Illustration de l'interface de RapidMiner. Dans la zone centrale, la chaîne de traitement contient 4 blocs. Les panneaux gauche et droite sont respectivement dédiés à l'ajout de bloc et au paramétrage du bloc sélectionné ----- 77

III .4. Corrélation entre les affinités biologiques expérimentales et prédites pour le modèle ---- 81

III .5. Représentations graphique des valeurs résiduelles en fonction des valeurs expérimentales pour le modèle ----- 82

CHAPITRE 02 : LES TECHNIQUES DE MODELISATION MOLECULAIRE

II .1. les descripteurs classés en fonction de la dimension chimique ----- 43

CHAPITRE 03 : RESULTATS ET DISCUSSION

III .1. Les propriétés physicochimiques de 108 molécules à l'aide de l'outil web SwissADME - 66
III .2. Les structures des composés de la base de données et leurs valeurs de Pki rapportées --- 72
III .3. Les descripteurs moléculaires calculés ----- 78
III .4. Matrice de corrélation pour le modèle QSAR avec les facteurs d'inflation de la variance
(VIF) des descripteurs selectionne ----- 80

INTRODUCTION GENERALE

Introduction générale :

Le cancer est l'une des principales causes de décès dans le monde, et le fardeau qu'il représente ne cesse de s'alourdir. En 2021, un nouveau seuil préoccupant a été franchi : selon les estimations, 20 millions de personnes ont été diagnostiquées avec un cancer et 10 millions sont décédées des suites de cette maladie. Ces chiffres continueront de croître rapidement dans les décennies à venir. La charge mondiale du cancer devrait atteindre 30 millions de nouveaux cas et 16,3 millions de décès par an d'ici 2040 [1].

Il est souvent complexe de déterminer les raisons précises pour lesquelles une personne développe un cancer. Cependant, l'étude des mécanismes biologiques à l'origine des cancers, appelée cancérogenèse, a permis d'identifier plusieurs facteurs augmentant le risque de cancer [2].

Les progrès scientifiques de ces dernières années permettent aujourd'hui de déchiffrer les codes génétiques du cancer et de comprendre à quel point cette maladie est liée aux mécanismes fondamentaux de la vie [3]. La recherche de nouvelles molécules thérapeutiques capables de stopper la prolifération tumorale constitue l'un des principaux axes de recherche en cancérologie [4].

Le développement de médicaments est un processus lent et laborieux, nécessitant un investissement considérable. Une entreprise pharmaceutique mettra de huit à douze ans pour produire un médicament. La phase de découverte, à elle seule, consomme une part significative des ressources de l'entreprise pour la synthèse et les tests biochimiques, et dure entre trois et cinq ans [5]. Les dépenses en recherche et développement étaient estimées à dépasser aujourd'hui les 900 millions de dollars, dont les deux tiers sont consacrés à la découverte de composés précurseurs potentiels (lead compound) qui, malheureusement, ne passent pas l'étape des tests précliniques [6].

La conception de médicaments in silico représente un nouveau paradigme ayant un impact significatif sur le processus global de découverte de médicaments. Elle permet de réduire le temps et les coûts, tout en produisant de nouveaux produits avec des propriétés thérapeutiques optimales et sans effets secondaires indésirables, de la manière la plus efficace et la moins coûteuse possible. Cette approche intègre diverses méthodes de découverte, telles que les approches basées sur les réseaux, les ligands et les relations structure-activité [7].

Avec le développement de l'informatique et la disponibilité sur le marché des logiciels professionnels adaptés, peu coûteux et rapides, l'intérêt pour les relations quantitatives structure-activité (QSAR) a connu une croissance exceptionnelle. Ces méthodes utilisent des analyses multidimensionnelles pour modéliser l'activité en fonction des paramètres structuraux moléculaires, appelés descripteurs [8].

La découverte de telles relations permet de prédire les activités et les propriétés des nouveaux composés et, par conséquent, de guider les synthèses de nouvelles molécules sans avoir à les réaliser [9].

le but de ce travail de recherche est de développer des inhibiteurs de la protéine Bcl-2 qui jouent un rôle crucial en tant que régulateurs de la voie mitochondriale de l'apoptose. Leur fonction principale, historiquement définie, est de contrôler la perméabilisation de la membrane externe mitochondriale (MOMP), ce qui active la voie intrinsèque de l'apoptose [10].

Pour atteindre cet objectif, nous avons envisagé de produire un modèle QSAR pour un ensemble de molécules.

Pour cela, le contenu de ce mémoire est composé de trois chapitres :

- **Introduction générale** : Dans laquelle nous rappelons quelques points sur le cancer et identifions l'objectif principal de ce travail.
- **Le premier chapitre**: Il est divisé en deux parties:
 - **Partie (A)**: La première section se concentrera sur le cancer, examinant les divers aspects de cette maladie, notamment ses facteurs de risque et ses traitements.
 - **Partie (B)**: se concentrera sur la famille des protéines Bcl-2.
- **Le deuxième chapitre** : Il couvre les principales approches ainsi que les étapes de développement d'un modèle QSAR.
- **Le troisième chapitre** : Nous discutons les résultats obtenus.
- **Une conclusion générale** : Conclure l'ensemble du travail réalisé.

Les Références :

- [1]. IAEA BULLETIN agence internationale de l'énergie atomique , la publication phare de l'AIEA (Février 2022) .
- [2]. Fondation arc pour la recherche sur le cancer " Le cancer collection comprendre et agir "
- [3]. Morgans, A. K., Van Bommel, A. C., Stowell, C., Abrahm, J. L., Basch, E., Bekelman, J. E., ... & Penson, D. F. (2015). Development of a standardized set of patient-centered outcomes for advanced prostate cancer: an international effort for a unified approach. *European urology*, 68(5), 891-898.
- [4]. El Ghalia, H. (2018). Corrélation-(structure-activité anticancéreuse) par les méthodes QSAR des molécules hétérocycliques précurseurs de médicaments.
- [5]. Kraim, k. (2009). *etudes de qsar sur des activites biologiques utilisant des produits d'origines naturels* (doctoral dissertation).
- [6]. Tropsha, A., & Oprea, T. (2005). Chemoinformatics in Drug Discovery. *Methods and Principles in Medicinal Chemistry*, 23, 440-441.
- [7]. KHAINNAR, M., & AIMEUR, O. Etude des relations Quantitatives Structure-Activité (QSAR) de plusieurs familles de composés sur des données biologiques.
- [8]. Bouhedjar, k. (2011). Toxicité des phénols: Une étude QSAR basée sur la méthode préconisée par l'OCDE.
- [9]. Chtita, S. (2017). *Modélisation de molécules organiques hétérocycliques biologiquement actives par des méthodes QSAR/QSPR. Recherche de nouveaux médicaments* (Doctoral dissertation, Université Moulay Ismaïl, Meknès).
- [10]. Brayer, S. (2013). *Les fonctions non-apoptotiques et pro-fibrosantes de la protéine pro-apoptotique BAX dans la fibrose pulmonaire idiopathique* (Doctoral dissertation, Paris Est).

**CHAPITRE 1:
L`APOPTOSE ET LE MORT
CELLULAIRE**

Partie 01 : Maladie du cancer

I .1.1. Introduction:

Le cancer est responsable d'un décès sur huit dans le monde [1]. Il regroupe plus de 100 maladies distinctes, avec des facteurs de risque et des caractéristiques épidémiologiques variés, provenant de la plupart des types de cellules et des organes du corps humain. Ces maladies se caractérisent par une prolifération relativement incontrôlée des cellules, capables d'envahir les tissus normaux et de métastaser vers des organes distants [2] .

Le cancer constitue un défi de santé mondial majeur. D'après l'Organisation mondiale de la santé (OMS), les données de 2012 révèlent environ 14,1 millions de nouveaux cas de cancer et 8,2 millions de décès attribués à cette maladie. En 2008, on estimait que 70% des décès liés au cancer dans le monde se produisaient dans les pays en développement.

Bien que l'incidence du cancer soit en augmentation dans la plupart des régions, les taux les plus élevés sont généralement observés dans les régions les plus développées. Toutefois, la mortalité demeure significativement plus élevée dans les pays en développement, principalement en raison du déficit de détection précoce et d'accès aux traitements [3].

La pandémie de COVID-19 a provoqué des retards dans le diagnostic et le traitement du cancer en raison de la fermeture des établissements de santé, des perturbations dans l'emploi et l'assurance maladie, ainsi que de la crainte de l'exposition au virus. Ces effets ont été particulièrement ressentis lors du pic de la pandémie à la mi-2020 [4]. Ce nombre est estimé atteindre 11 millions d'ici à 2030 [5].

I .1.2. Le cancer :

Le terme « cancer » désigne un large éventail de maladies pouvant affecter n'importe quelle partie du corps. On utilise également les termes « tumeurs malignes » et « néoplasmes ». Une caractéristique distinctive du cancer est la croissance rapide de cellules anormales issues d'un même clone, qui peuvent envahir les tissus environnants et migrer vers d'autres organes pour former des métastases. La présence de métastases étendues est la principale cause de décès associés au cancer [6 - 7].

Un cancer se caractérise par une prolifération anarchique de cellules, perturbant le fonctionnement normal des organes et pouvant conduire au décès [8 – 9]. Chaque organe et tissu présente un arrangement spécifique de cellules, habituellement bien ordonné [9]. Les échanges de signaux entre ces cellules maintiennent l'homéostasie tissulaire, garantissant que chaque cellule occupe sa place et exerce sa fonction. En revanche, les cellules cancéreuses ignorent ces signaux régulateurs [8], se multipliant de manière incontrôlée et envahissant les tissus environnants, sans remplir leur fonction physiologique. Ce phénomène compromet progressivement le fonctionnement des organes touchés, pouvant ultimement conduire au décès, selon la localisation et l'évolution de la tumeur [10]. Un cancer peut émerger à partir d'une seule cellule qui subit des mutations favorisant sa survie et sa multiplication [8]. Cette cellule donne naissance à une masse de tissu anormal, appelée tumeur, qui envahit progressivement les tissus sains à mesure que le cancer progresse. Le processus de métastase survient lorsque des cellules se détachent de la tumeur d'origine, appelée tumeur primaire, pour se propager dans la circulation sanguine et former des tumeurs dans d'autres parties du corps [8]. Cette dissémination rend le traitement du cancer considérablement plus complexe, car les approches thérapeutiques ciblées sur une zone précise sont moins efficaces [10].

I.1.2.1. La cellule cancéreuse :

Bien que les cellules tumorales conservent souvent des caractéristiques biologiques résiduelles des cellules normales (**figure I.1**), elles acquièrent de nouvelles propriétés en raison de la transformation maligne. Souvent, les cellules tumorales sont plus grandes que leurs cellules d'origine, pouvant même devenir des cellules géantes dans certains cas. Cette augmentation de taille n'est pas uniforme au sein d'une même population, phénomène désigné par le terme d'anisocytose, bien que cette caractéristique ne soit pas spécifique aux lignées cancéreuses [11].

Le plasmalemme est altéré, avec une augmentation de sa teneur en acides sialiques et en gangliosides, ce qui augmente souvent la mobilité latérale des protéines intrinsèques [11]. Le cytoplasme présente généralement une basophilie accrue due à une augmentation de la quantité d'ARNm. Il est peu abondant, souvent vacuolisé, avec un réseau de filaments d'actine désorganisé et un nombre accru de ribosomes. La membrane nucléaire est épaissie et forme souvent des invaginations intranucléaires [11]. Le nombre de mitochondries est réduit, parfois fragmentées,

avec un nombre de crêtes inférieur à celui des cellules non cancéreuses. De plus, les enzymes mitochondriales peuvent être altérées ou réduites en quantité [12].

La division des cellules cancéreuses devient incontrôlée, provoquant une prolifération anarchique et la génération de lignées dites "immortelles". Ce processus est crucial dans la transformation maligne car il permet l'accumulation d'un grand nombre d'anomalies génomiques. Des anomalies peuvent survenir lors des mitoses, telles qu'une dispersion anarchique avec une réparation chromosomique inégale [11].

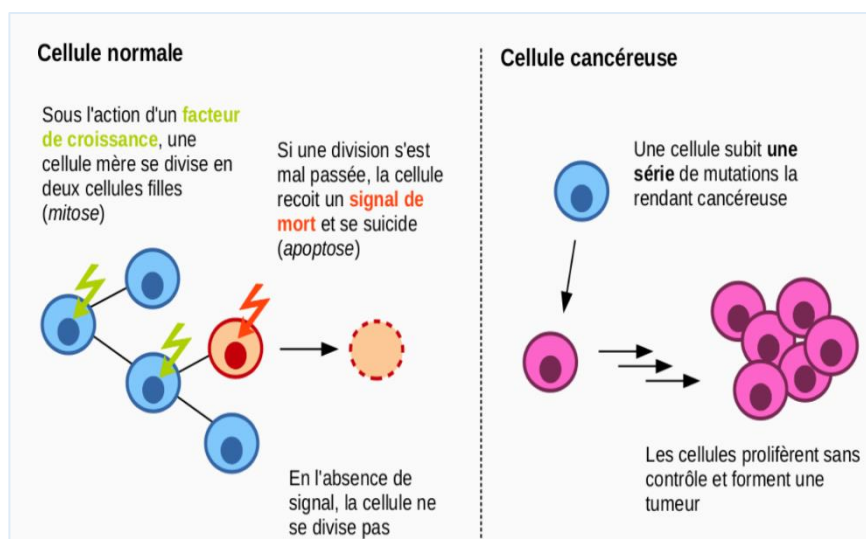


Figure I.1 : Différences dans le mécanisme de division cellulaire entre une cellule normale et cancéreuse.

I .1.3. Type de cancer:

I .1.3.1. Cancer du sein :

Le cancer du sein désigne un ensemble de maladies où les cellules du tissu mammaire se transforment et se multiplient de manière incontrôlée, souvent aboutissant à la formation d'une masse ou d'une tumeur. Généralement, ces cancers prennent leur origine dans les glandes mammaires (lobules) ou les canaux qui relient ces glandes au mamelon (**figure I.2**) [13].

Au début, le cancer du sein peut être asymptomatique lorsque la tumeur est de petite taille et plus facile à traiter. C'est pourquoi le dépistage précoce est essentiel pour détecter la maladie. Le signe physique le plus courant est la présence d'une masse indolore. Parfois, le cancer du sein peut se

propager aux ganglions lymphatiques de l'aisselle, provoquant une masse ou un gonflement, même avant que la tumeur initiale du sein ne soit suffisamment grande pour être palpable [13].

Les symptômes moins fréquents incluent des douleurs ou une sensation de lourdeur dans le sein, des plis, un gonflement, un épaississement ou une rougeur de la peau mammaire, ainsi que des changements au niveau du mamelon tels qu'un écoulement spontané (surtout s'il est sanglant), des croûtes ou une rétraction [13].

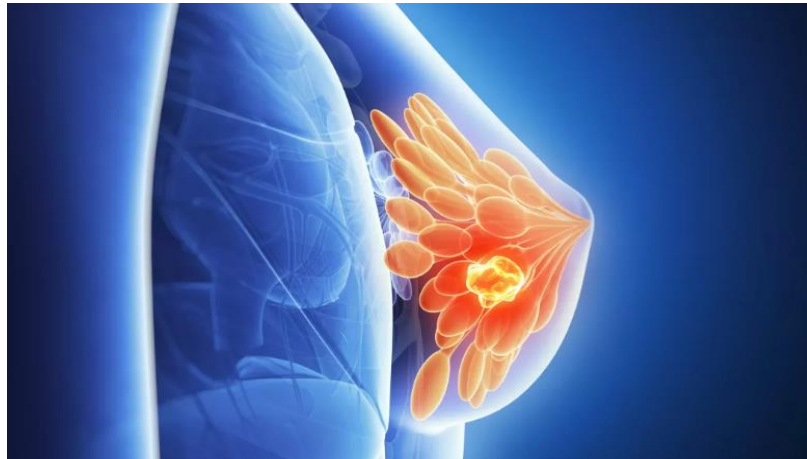


Figure I.2 : Cancer du sein.

I.1.3.2. Cancer de l'estomac :

Le cancer de l'estomac, connu également sous le nom de cancer gastrique, émerge à partir d'une cellule pariétale de l'estomac, initialement saine, qui subit une multiplication anarchique, conduisant à la formation d'une masse désignée comme une tumeur maligne (**figure I.3**) [14].

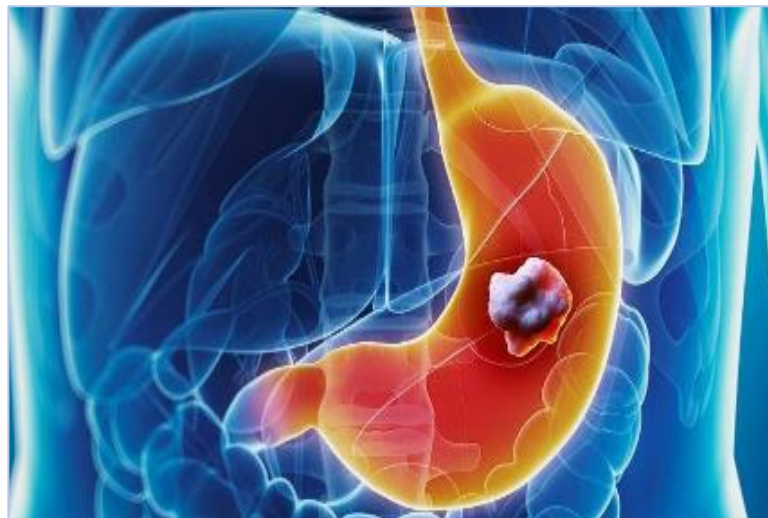


Figure I.3 : Cancer gastrique (cancer de l'estomac).

En l'absence de traitement, cette affection peut progresser et donner lieu à des métastases vers les ganglions lymphatiques ou d'autres organes [15].

Généralement, le cancer de l'estomac prend la forme d'un adénocarcinome, dont les symptômes demeurent souvent atypiques. Plusieurs facteurs peuvent favoriser son développement, notamment une gastrite chronique, principalement induite par la bactérie *Helicobacter pylori*, le tabagisme, une alimentation riche en sel et pauvre en fruits et légumes, des antécédents familiaux de cancer de l'estomac, ou encore plus rarement, une prédisposition génétique [15].

I.1.3.3. Carcinome hépatocellulaire : (le cancer du foie)

Selon l'Organisation mondiale de la Santé (OMS), le cancer du foie est défini comme un cancer qui se développe dans les cellules hépatiques. Le type de cancer le plus fréquent au niveau du foie est le carcinome hépatocellulaire (CHC) (**figure I.4**).

Le foie, qui fait partie de l'appareil digestif, est situé dans la partie supérieure droite de l'abdomen et est le plus grand organe plein du corps. Il produit des enzymes et de la bile, qui facilitent la digestion des aliments, ainsi que des protéines essentielles à la coagulation du sang. De plus, le foie purifie le sang en éliminant les substances nocives. Dans certains cas, des modifications des cellules hépatiques peuvent entraîner un cancer du foie [16].

La plupart du temps, ce cancer se développe dans les cellules qui constituent la majeure partie du foie, appelées hépatocytes. Ce type de cancer est connu sous le nom de carcinome hépatocellulaire et est le type de cancer primitif du foie le plus courant.

Le cancer du foie peut également apparaître dans les cellules qui tapissent les canaux biliaires à l'intérieur du foie. Ce type de cancer est appelé cholangiocarcinome ou cancer des canaux biliaires intrahépatiques.

D'autres types de cancers peuvent se propager au foie, mais ils sont différents du cancer primitif du foie. Lorsque le cancer prend naissance dans une autre partie du corps et se propage au foie, on parle de métastases hépatiques. Les métastases hépatiques sont traitées différemment du cancer primitif du foie [16].

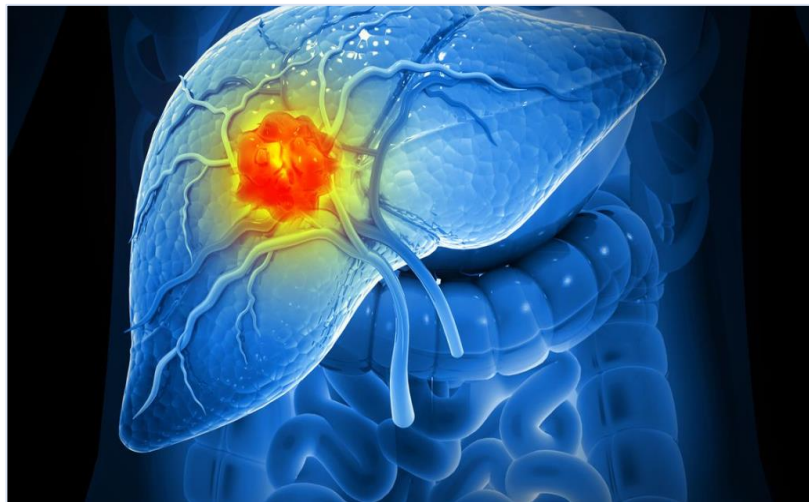


Figure I.4 : Carcinome hépatocellulaire (le cancer du foie).

I .1.3.4. Cancer de la prostate :

Le cancer de la prostate est défini par l'Organisation mondiale de la santé (OMS) comme une malignité caractérisée par la croissance anormale de cellules dans la prostate, une glande du système reproducteur masculin. Cette croissance incontrôlée peut former une masse tumorale pouvant envahir les tissus environnants et se propager à d'autres parties du corps, entraînant éventuellement des métastases. Cette définition est largement reconnue et est également soutenue par des recherches et des études publiées dans des journaux médicaux et scientifiques (**figure I.5**).



Figure I.5 : Cancer de la prostate.

I .1.4. Pathogénéses du cancer :

Il demeure souvent ardu de déterminer les causes précises du cancer chez un individu. Les cancers, étant des maladies multifactorielles, résultent d'une combinaison complexe de divers éléments, dont certains demeurent encore inconnus. Néanmoins, l'exploration des mécanismes biologiques sous-jacents, appelée cancérogenèse, a permis d'identifier plusieurs facteurs de risque associés à cette maladie.

Ces facteurs de risque sont généralement classés en deux catégories distinctes : les facteurs évitables et les facteurs non évitables. Les premiers incluent des aspects liés à notre mode de vie et à nos comportements, tels que le tabagisme, le régime alimentaire, l'exposition au soleil, ainsi que les infections virales ou bactériennes spécifiques. Ces facteurs évitables offrent des opportunités de prévention. Quant aux facteurs non évitables, ils sont intrinsèquement liés à des aspects tels que l'âge, le sexe et la génétique, sur lesquels il est difficile d'agir directement [17].

I .1.4.1. Les risques liés aux comportements :

I .1.4.1.1. Tabagisme :

Lorsqu'une cigarette se consume, elle libère environ 250 substances dangereuses, dont 70 sont des agents cancérogènes avérés. Parmi ces agents, on retrouve de nombreux composants des goudrons ainsi que certains métaux lourds tels que l'arsenic et le plomb. Leur effet nocif ne se limite pas aux voies respiratoires, car ces composés traversent les poumons et pénètrent dans la circulation sanguine, contaminant ainsi l'ensemble de l'organisme. Une fois à l'intérieur d'une cellule, ces produits altèrent son ADN, augmentant ainsi le risque de mutations génétiques, dont certaines peuvent déclencher un cancer[17].

I .1.4.1.2. Alimentation :

La consommation de certains aliments ou composés de notre alimentation a été étroitement liée, selon des preuves solides, à un risque accru ou réduit de développer certains cancers chez les individus en bonne santé.

Les éléments nutritionnels liés à une diminution du risque de cancer comprennent :

- Les fibres, bénéfiques pour prévenir le cancer du côlon-rectum.

- Les fruits et légumes, associés à une réduction du risque de cancers aérodigestifs.
- Les céréales complètes, favorables à la prévention du cancer du côlon-rectum.
- Les produits laitiers, liés à une diminution du risque de cancer du côlon-rectum.
- Le café, associé à une réduction du risque de cancers du foie et de l'endomètre.

Les éléments nutritionnels associés à une augmentation du risque de cancer incluent :

- La consommation d'alcool, liée à un risque accru de cancers aérodigestifs supérieurs, de sein avant et après la ménopause, de foie, d'estomac et de côlon-rectum.
- La consommation de charcuterie et de viande rouge, associée à un risque accru de cancer du côlon-rectum.
- La consommation de sel et d'aliments conservés par le sel, liée à un risque accru de cancer de l'estomac [18].

I .1.4.1.3. L'alcool :

L'alcool est impliqué dans la formation de lésions qui favorisent le développement du cancer. Deux molécules sont principalement responsables de ces effets : l'éthanol et son produit de dégradation principal, l'acétaldéhyde. L'acétaldéhyde est classé comme cancérigène avéré pour l'Homme par le Centre international de recherche sur le cancer. Cette molécule est produite dès la consommation de boissons alcoolisées, puis en grande quantité dans le foie où elle est libérée dans la circulation sanguine. Elle exerce son effet cancérigène en altérant l'ADN des cellules, affectant ainsi les muqueuses de la cavité buccale, du larynx, du pharynx et de l'œsophage, ainsi que d'autres organes de l'organisme [17].

I .1.4.2. Les risques liés à l'environnement :

I .1.4.2.1. La pollution de l'air :

La pollution de l'air constitue un facteur de risque de cancer, pouvant varier considérablement en fonction du lieu, de la saison, de la température, de la proximité des sources de pollution et de l'humidité.

I.1.4.2.2. Les infections :

Au niveau mondial, environ 20 % des cas de cancer sont associés à des infections virales.

a. Les virus :

- **Les infections par les virus de l'hépatite B et C (VHB et VHC)** peuvent évoluer vers une cirrhose hépatique, un état favorisant le développement du cancer du foie. À eux seuls, ces virus sont responsables de 70 à 80 % des cas de cancer du foie.
- **Le papillomavirus** peut induire des lésions du col de l'utérus qui ont le potentiel de se transformer en cancer. Environ 70 % des cas de cancer du col de l'utérus sont attribuables à deux types de papillomavirus humains (HPV16, HPV18). De plus, les infections à papillomavirus peuvent également conduire à des cancers de l'anus, de la vulve, du vagin, du pénis et de l'oropharynx.
- **Le virus Epstein-Barr** est associé à des lymphomes non hodgkiniens et des cancers du nasopharynx.
- **Le virus de l'herpès HHV8** est responsable du sarcome de Kaposi.
- **Le virus HTLV** est lié à des leucémies et des lymphomes.
- **Le virus VIH** (responsable du sida) favorise l'apparition de complications cancéreuses en raison de l'immunodépression qu'il induit, entraînant une diminution des défenses naturelles de l'organisme.

b. Les bactéries :

Helicobacter pylori est la première bactérie identifiée comme étant à l'origine d'un type de cancer, en l'occurrence le cancer de l'estomac. L'infection par cette bactérie peut multiplier par 5 ou 6 le risque de développer un cancer gastrique. On estime qu'elle est responsable de 63 % des cas de cancer de l'estomac dans le monde, ce qui équivaut à 5,5 % de tous les cancers diagnostiqués à l'échelle mondiale.

Cette bactérie infecte plus de la moitié de la population mondiale, avec des taux variant de 20 à 90 % selon les pays, souvent contractée durant l'enfance, par transmission d'individu à individu, principalement par voie orale et fréquemment au sein des familles (transmission mère/enfant, entre frères et sœurs).

I .1.4.3. Les risques liés à l'individu et son histoire :

L'avancement en âge, les prédispositions génétiques, les traitements antérieurs contre le cancer, entre autres, sont autant de facteurs de la vie personnelle qui peuvent influencer le risque de développer la maladie.

I .1.4.3.1. L'âge avancé :

L'âge avancé, un facteur inévitable par excellence, accroît la probabilité de cancer. Avec le temps, le nombre de lésions cellulaires susceptibles de s'accumuler augmente. Ainsi, à mesure que nous vieillissons, le risque que l'une de nos cellules devienne anormale et engendre une tumeur augmente également.

I .1.4.3.2. Les hormones :

Les fluctuations hormonales tout au long de la vie peuvent altérer le risque de cancer. Des éléments tels que l'âge de la puberté et de la ménopause, le nombre et l'âge des grossesses, ainsi que la prise de contraceptifs oraux ou de traitements hormonaux substitutifs à la ménopause (THS) peuvent avoir un impact significatif.

I .1.4.3.3. Origine héréditaire:

Entre 5 et 10 % des cas de cancer ont une origine héréditaire. Il est estimé que 5 à 10 % des cancers de l'ovaire et du sein, ainsi que 3 % des cancers du côlon, sont associés à une prédisposition génétique héritée. D'autres cancers, plus rares, ont souvent une origine héréditaire, comme les rétinoblastomes (tumeurs de l'œil chez les enfants) ou certains types de tumeurs rénales (comme le néphroblastome, également connu sous le nom de tumeur de Wilms).

I .1.4.3.4. Maladies inflammatoires ou auto-immunes:

Dans le cas de certaines maladies chroniques telles que les maladies inflammatoires ou auto-immunes, les déficits immunitaires sont liés à un risque accru de cancer ou de lymphome.

I .1.4.3.5. Traitements:

Certains traitements comme la radiothérapie, la chimiothérapie ou l'utilisation d'immunosuppresseurs peuvent également augmenter le risque de cancer, en particulier celui des cancers hématologiques (affectant le sang) et cutanés [17].

I .1.5. Les principales options thérapeutiques :

Il est crucial de poser un diagnostic précis du cancer afin de pouvoir le traiter de manière adaptée et efficace. En effet, chaque forme de cancer requiert un protocole thérapeutique spécifique. Généralement, le traitement du cancer s'appuie sur la chirurgie, la radiothérapie et/ou des traitements systémiques tels que la chimiothérapie, les traitements hormonaux ou les thérapies biologiques ciblées. Le choix du traitement approprié dépend à la fois du type de cancer et du profil du patient. Il est primordial que le protocole soit suivi dans un délai déterminé afin d'obtenir les résultats thérapeutiques escomptés [6].

I .1.5.1. La chirurgie :

La chirurgie demeure la méthode la plus traditionnelle de traitement du cancer. Encore aujourd'hui, elle est fréquemment recommandée en premier lieu aux patients ayant reçu un diagnostic de cancer [19].

Les cancers qui nécessitent souvent un traitement chirurgical impliquent généralement l'exérèse non seulement de la tumeur ou de l'organe affecté, mais également d'une marge de tissu sain environnant ainsi que des ganglions lymphatiques adjacents. Cette excision étendue, visant à éliminer localement toutes les cellules cancéreuses, constitue un élément essentiel du processus de guérison.

Les progrès significatifs réalisés dans le domaine de l'imagerie ont permis d'améliorer considérablement les techniques chirurgicales. Dans les interventions abdominales, il est souvent possible d'éviter l'ouverture de la paroi abdominale en optant pour la coelioscopie, une approche chirurgicale moins invasive utilisant un équipement de plus en plus sophistiqué [20].

I.1.5.2. La radiothérapie :

La radiothérapie est largement utilisée dans le traitement des cancers, étant administrée à plus de la moitié des patients atteints de cette maladie.

Ce traitement implique l'exposition de la zone affectée par les cellules cancéreuses à des rayonnements. Cette exposition entraîne une altération des cellules, les privant de leur capacité à se reproduire. Pour limiter les dommages aux tissus sains avoisinants, seuls les secteurs abritant la tumeur sont visés par les rayons, minimisant ainsi les risques pour les cellules saines [19].

I.1.5.3. Les traitements médicamenteux du cancer :

I.1.5.3.1. La chimiothérapie :

Dans le domaine de l'oncologie, la chimiothérapie désigne l'utilisation de cytostatiques, qui sont plus d'une centaine. L'oncologue les prescrit en fonction du type de cancer et de son stade.

Les cellules humaines suivent un ordre précis de division en plusieurs phases. Les cytostatiques perturbent cet ordre de division des cellules cancéreuses, ce qui ralentit la croissance de la tumeur. Cependant, ces médicaments affectent également les cellules saines, ce qui constitue la principale cause des effets secondaires de la chimiothérapie. Heureusement, les organes ont la capacité de se régénérer, ce qui réduit les effets secondaires après le traitement.

Tous les cytostatiques peuvent entraîner des effets secondaires qui varient d'une personne à l'autre. La nature, la fréquence et l'intensité de ces effets dépendent de plusieurs facteurs, notamment les caractéristiques individuelles du patient, le médicament utilisé et la combinaison de médicaments administrés dans le cadre de la chimiothérapie [21].

I.1.5.3.2. L'hormonothérapie :

Ce type de traitement repose, comme pour la chimiothérapie, sur la prise de médicaments [19].

L'hormonothérapie est un traitement visant à bloquer l'action des hormones qui pourraient favoriser la croissance des cellules cancéreuses. Principalement utilisée dans le traitement des cancers gynécologiques et prostatiques, qui sont dits hormono-dépendants, cette thérapie cible les tumeurs sensibles aux hormones.

Les cancers du sein ou de la prostate présentent souvent une sensibilité aux hormones féminines telles que les œstrogènes et la progestérone, ou aux hormones masculines comme la testostérone, qui stimulent leur développement.

L'hormonothérapie intervient en bloquant l'action de ces hormones sur les cellules cancéreuses, afin d'entraver la progression du cancer [14].

I .1.5.4. Médecine de précision :

I .1.5.4.1 Thérapies ciblées et immunothérapie spécifique :

Les thérapies ciblées représentent une autre catégorie de traitements contre le cancer, souvent disponibles sous forme de médicaments oraux. En visant des molécules spécifiques de l'organisme, elles entravent des mécanismes essentiels à la multiplication des cellules cancéreuses ou, de manière plus générale, à la croissance de la tumeur. Certains de ces traitements agissent directement sur les cellules cancéreuses, tandis que d'autres ciblent les cellules du microenvironnement tumoral [17].

les traitements d'immunothérapie spécifique réactivent le système immunitaire en ciblant des éléments spécifiques de la tumeur ou de son environnement, favorisant ainsi la destruction des cellules cancéreuses.

La médecine de précision constitue désormais une option thérapeutique dans la lutte contre le cancer. Elle ne se substitue pas aux traitements existants qui ont déjà permis de guérir un cancer sur deux, mais elle vient compléter l'arsenal thérapeutique disponible. En offrant de nouvelles avenues, la médecine de précision représente une lueur d'espoir pour les patients atteints de cancers pour lesquels les traitements conventionnels se révèlent insuffisants [14].

Partie 02 : La famille des protéines BCL-2

I .2.1. Introduction:

L'apoptose a été formellement définie et baptisée en 1972 en tant que réponse morphologique spécifique à divers types de stress cellulaire, Cependant, malgré l'originalité et l'importance du concept, peu de recherches expérimentales ont été menées au cours des 20 années suivantes en raison de l'absence d'outils pour manipuler le processus. Au début des années 1990, deux découvertes fondamentales ont changé la donne. Tout d'abord, lors de l'élucidation méticuleuse de la séquence complète de développement du nématode *Caenorhabditis elegans* au niveau cellulaire individuel, il a été observé qu'un nombre fixe et prévisible de "cellules intermédiaires" étaient destinées à mourir, et que ce processus était régulé positivement et négativement par des gènes spécifiques. Deuxièmement, la découverte d'un nouveau gène appelé B-cell CLL/lymphoma 2 (Bcl-2; codé par BCL2) en tant que partenaire dans une translocation chromosomique réciproque dans une tumeur humaine a révélé qu'il fonctionnait non pas comme un oncogène classique en stimulant la division cellulaire, mais plutôt en prévenant l'apoptose. Lorsqu'il a été constaté que le BCL2 des mammifères pouvait remplacer CED-9, le gène de *C. elegans* qui inhibe la mort cellulaire, la généralité du processus a été reconnue, et la littérature scientifique a explosé avec désormais plus de 105 publications sur l'apoptose. Cependant, il est ironique de noter qu'après 20 années supplémentaires d'investigations intensives, il est devenu clair que le mécanisme d'action de Bcl-2 est tout à fait distinct de celui de Ced-9, qui séquestre l'activateur de la caspase protéase, l'effector final de l'apoptose. En revanche, Bcl-2 agit principalement en se liant à d'autres protéines apparentées qui régulent perméabilisation de la membrane externe mitochondriale (MOM) [22].

I .2.2. Le cycle cellulaire :

L'organisme humain, constitué d'environ 100 000 milliards de cellules, maintient son fonctionnement grâce à un équilibre dynamique entre la production de nouvelles cellules et l'élimination des cellules existantes. Ce contrôle est exercé par deux mécanismes principaux : le cycle cellulaire, qui régule la division cellulaire et la création de nouvelles cellules, et l'apoptose, un processus programmé de mort cellulaire qui élimine les cellules âgées ou endommagées. À l'exception des cellules reproductrices telles que les spermatozoïdes et les ovules [23].

Le cycle cellulaire représente le pilier central de tous les organismes, défini comme une série ordonnée de processus permettant à une cellule de se diviser en deux cellules filles présentant les mêmes caractéristiques morphologiques et physiologiques que la cellule mère [24]. C'est un processus fondamental et commun à toutes les cellules eucaryotes, composé de deux grandes étapes : l'interphase, durant laquelle la cellule augmente sa masse et duplique son matériel génétique, et la phase mitotique (la mitose ou la phase M), pendant laquelle deux cellules filles sont formées, chacune avec un ensemble complet de chromosomes (**figure I.6**) .

Lorsque la cellule ne se divise pas, elle est dans un état de quiescence, appelé la phase G0 [25]. Ce processus de division cellulaire passe par plusieurs étapes regroupées dans le cycle cellulaire, impliquant divers acteurs tels que des protéines et de l'acide ribonucléique (ARN).

Le cycle cellulaire se compose de cinq phases distinctes : une phase de repos, une première phase de croissance, une phase de synthèse, une seconde phase de croissance et enfin la phase de mitose proprement dite, où la division cellulaire se produit [23].

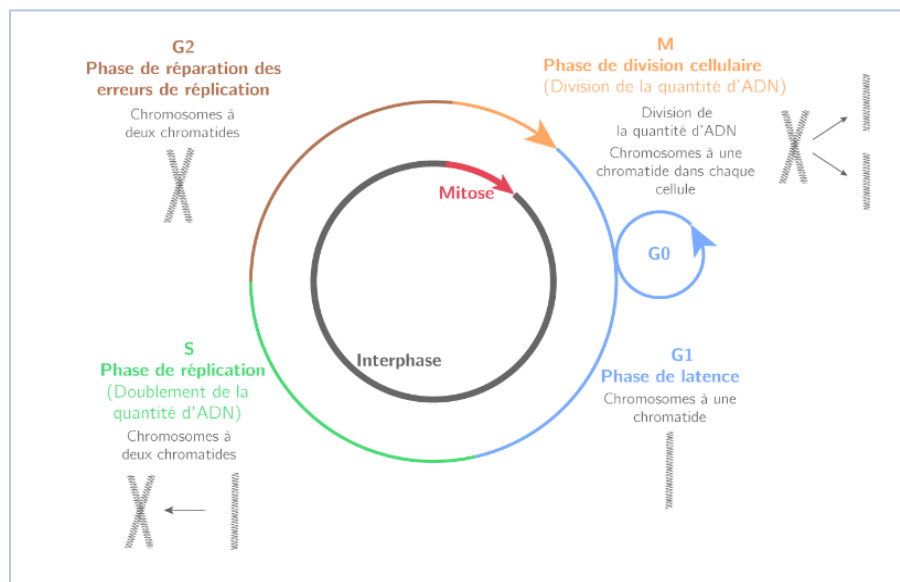


Figure I.6 : Représentation schématique du cycle cellulaire.

I .2.2.1. Les phases du cycle cellulaire :

Le cycle cellulaire se compose de deux principales étapes l'interphase et la mitose :

a. L'interphase :

La phase interphasique est la période la plus longue du cycle cellulaire s'étendant de la fin d'une division cellulaire au début de la suivante. Sa durée dépend de la nature de la cellule et des conditions physiologiques qui lui sont propres (**figure I.7**).

L'interphase est divisée en trois phases successives : la phase G1 (Gap 1), la phase S (synthèse) et la phase G2 (Gap 2) [26].

La phase G1 : généralement la plus longue du cycle cellulaire, est principalement une période de croissance initiale. Elle se caractérise par une activité métabolique intense, comprenant la transcription des ARN et la synthèse des protéines nécessaires à l'accroissement de la cellule. Pendant cette phase, les cellules franchissent le point de restriction, un seuil au-delà duquel le cycle devient irrévocablement engagé et l'entrée en division n'est plus conditionnée par la présence de signaux mitogènes [27].

Au cours de cette phase, la réplication de l'ADN ne se produit pas, mais le volume cytoplasmique de la cellule augmente par rapport au volume du noyau, ce qui permet de repérer le moment où la cellule entre dans la phase S [28].

La phase S : Elle commence lorsque le rapport de volume entre le cytoplasme et le noyau atteint un seuil spécifique. C'est alors que la transcription des gènes impliqués dans la synthèse des enzymes de synthèse de l'ADN est activée. Sa durée, généralement peu variable, dépend principalement de la quantité d'ADN à répliquer et du nombre de sites de réplication actifs simultanément [28].

Pendant cette phase, le matériel chromosomique est doublé par duplication, chaque chromosome dupliqué étant constitué de deux chromatides sœurs. La durée de la phase S varie en fonction de la quantité d'ADN à répliquer et du nombre de sites de réplication actifs. La synthèse de l'ADN est coordonnée avec celle des histones. Les ARNm des histones sont produits pendant la phase S, puis dégradés immédiatement après la réplication. Cette phase se caractérise également par la fin de la synthèse d'autres ARNm [29].

La phase G2 : Les cellules en phase G2 présentent une quantité d'ADN doublée due à la synthèse d'ADN pendant la phase S. Cette phase est généralement très courte, voire parfois inexistante. Pendant la phase G2, la cellule vérifie si la réplication de l'ADN a été correctement effectuée (réparation post-réplivative) et se prépare à la division cellulaire [29]. Il s'agit d'une phase de sécurité au cours de laquelle la cellule a l'occasion de réparer d'éventuelles erreurs survenues lors de la réplication de son ADN. Le déclenchement de la mitose à la fin de la phase G2 est marqué par la phosphorylation de nombreuses protéines. La phase G2 a une durée de 2 à 3 heures [29].

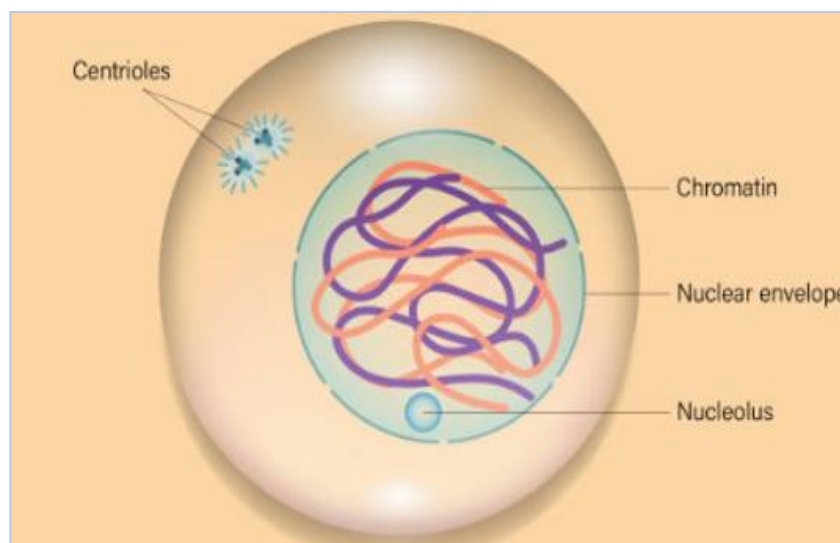


Figure I.7 : Représentation schématique de L'interphase.

b. La phase mitotique (la phase M ou mitose) :

La mitose, une phase cruciale du cycle cellulaire, est hautement conservée chez toutes les cellules eucaryotes [30]. Elle comprend à la fois une division nucléaire, appelée caryocinèse, et une division cytoplasmique, nommée cytokinèse. La durée de la mitose est généralement de 1 à 3 heures. Elle se déroule en cinq étapes caractéristiques, de la prophase à la télophase [31].

La prophase : La prophase se caractérise par une réorganisation de la chromatine, qui se condense en chromosomes compacts visibles sous microscope photonique [32]. C'est la première phase de la division cellulaire lors de la mitose (**figure I.8**). Pendant cette phase, les centrosomes se séparent et migrent vers les pôles opposés de la cellule [31]. À la fin de la prophase, le noyau disparaît lorsque l'enveloppe nucléaire se rompt [33].

Une autre particularité de la prophase est la maturation des centrosomes, riches en tubuline, qui agissent comme des centres organisateurs de microtubules (MTOC). Ils se séparent et migrent progressivement de manière diamétralement opposée [34].

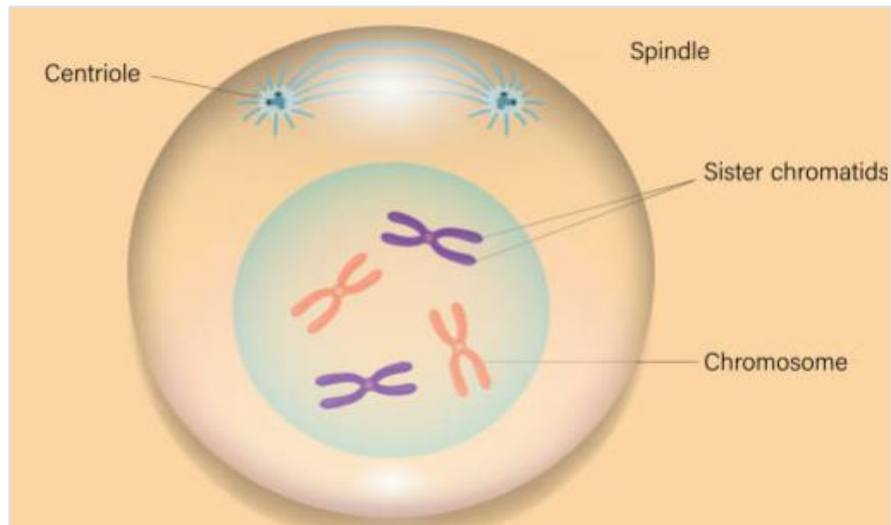


Figure I.8 : Représentation schématique de la prophase.

La prométaphase : Certains auteurs considèrent la prométaphase comme une extension de la prophase. Pendant cette phase, la condensation de l'ADN des chromosomes se poursuit, tandis que des complexes protéiques spécialisés, appelés kinétochores, se forment au niveau des centromères des chromosomes. Les microtubules s'allongent à partir de deux centrosomes et se fixent aux kinétochores [31].

La métaphase : C'est la phase où les chromosomes se positionnent de part et d'autre du plan équatorial, engendrant ainsi des forces opposées qui tirent chaque chromatide sœur vers un pôle du fuseau [35].

À la métaphase de la mitose, les chromosomes atteignent leur degré de condensation maximal, devenant ainsi le plus visibles (**figure I.9**). C'est à ce stade qu'ils illustrent le mieux l'organisation discontinue du génome eucaryote. De plus, étant donné que l'enveloppe nucléaire a disparu, ils sont les plus faciles à analyser par simple écrasement de la cellule, car ils s'étalent sans problème dans les préparations cytologiques [36].

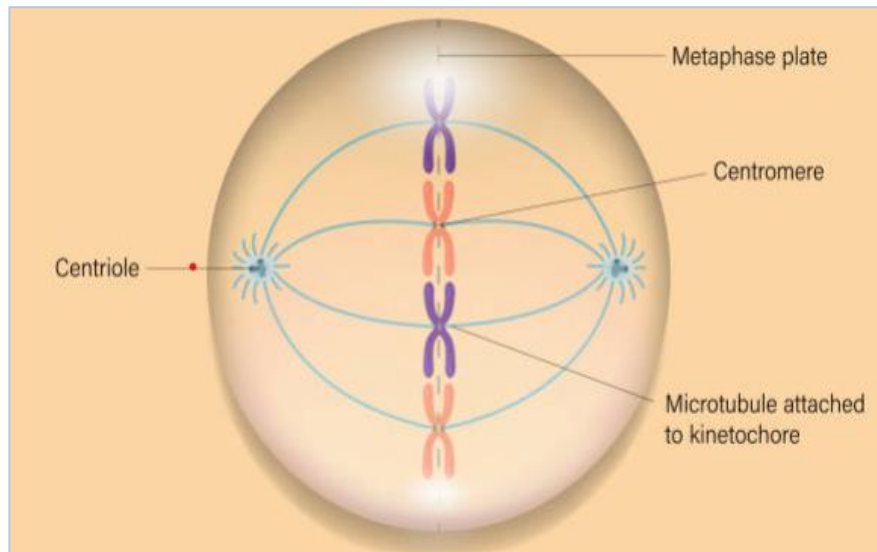


Figure I.9 : Représentation schématique de la métaphase.

L'anaphase : est marquée par la ségrégation des chromatides sœurs de chaque chromosome en deux lots égaux de chromosomes à une chromatide, qui migrent vers les pôles opposés du fuseau (**figure I.10**). Dans un premier temps, appelé anaphase A, les microtubules kinétochoriens se dépolymérisent et tirent les chromatides vers chacun des pôles. Puis, dans un second temps, appelé anaphase B, les microtubules polaires s'allongent par polymérisation, éloignant ainsi les deux pôles du fuseau l'un de l'autre [37].

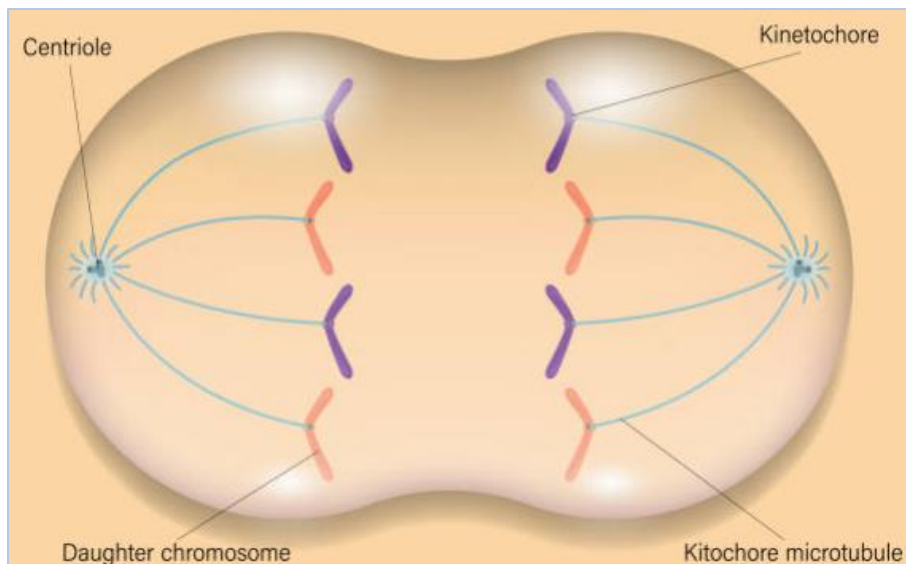


Figure I.10 : Représentation schématique de l'anaphase.

La télophase : Durant cette phase, les microtubules kinétochoriens se détachent, les chromosomes se décondensent et redeviennent des brins d'ADN [30]. Deux enveloppes nucléaires se forment autour des chromosomes filles. Les autres microtubules continuent de s'allonger jusqu'à ce que deux cellules filles distinctes se forment [38] (**figure I.11**).

Les microtubules jouent un rôle crucial dans la cellule. En perturbant leur instabilité dynamique, cela entraîne un blocage de la division cellulaire en métaphase et peut conduire à l'apoptose. De ce fait, il est médicalement intéressant de développer des molécules capables d'agir sur cet équilibre [39].

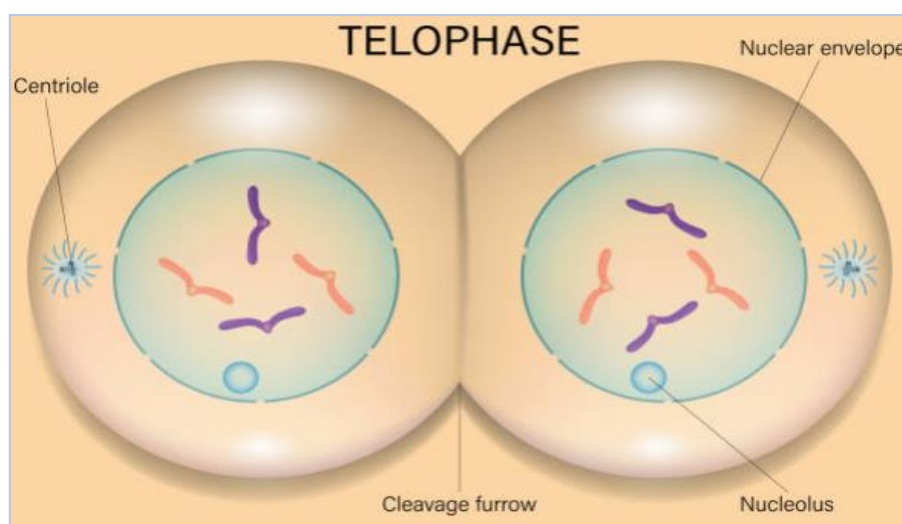


Figure I.11 : Représentation schématique de la télophase.

I .2.3. La famille des protéines BCL-2 :

La famille Bcl-2 représente un ensemble de protéines intracellulaires qui jouent un rôle crucial dans la régulation de l'apoptose. De plus, il est désormais évident qu'elles interviennent également dans divers autres processus cellulaires. Leur dysfonctionnement est souvent lié à la progression tumorale et aux troubles neurodégénératifs. Depuis la découverte du membre fondateur de cette famille, Bcl-2, il y a environ trente ans, des homologues de cette protéine ont été identifiés dans de nombreuses espèces, mais les mécanismes moléculaires régissant leur activité, qu'elle soit apoptotique ou non, demeurent encore mal définis [40].

I .2.3.1. Classification des protéines de la famille Bcl-2 :

Les protéines de la famille Bcl-2 se caractérisent par une structure comprenant 4 domaines d'homologie appelés domaines BH (Bcl-2 homology domain), qui sont des séquences courtes de moins de 20 résidus d'acides aminés [41]. Ces protéines jouent un rôle dans l'apoptose, notamment lors de la libération du cytochrome c dans le cytosol : certaines protéines de cette famille sont pro-apoptotiques car elles favorisent la libération du cytochrome c dans le cytosol, tandis que d'autres sont anti-apoptotiques [42] (**figure I.12**).

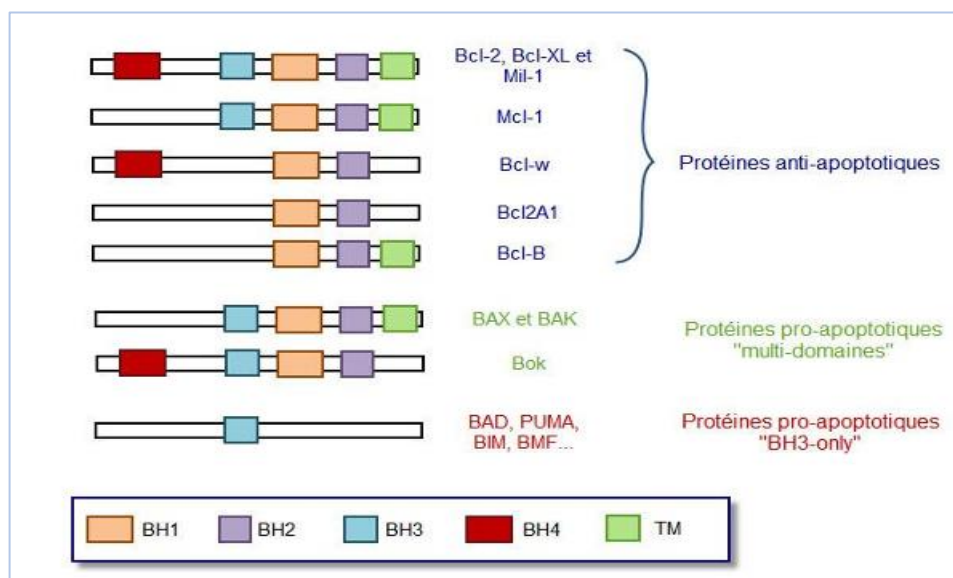


Figure I.12 : Structure des protéines de la famille Bcl-2.

Ces deux caractéristiques permettent de classer ces protéines en 3 sous-familles (**figure 12**) :

- Les protéines **anti-apoptotiques**, pouvant présenter des homologies pour les 4 domaines BH .
- Les protéines **pro-apoptotiques multi-domaines** possédant plusieurs domaines BH .
- Les protéines **pro-apoptotiques** dites « BH3 only », ne possédant que le domaine BH3.

I .2.3.1.1. Les protéines anti-apoptotiques :

Les protéines anti-apoptotiques, qui peuvent comporter de 2 à 4 domaines BH, sont principalement localisées dans la membrane externe mitochondriale, bien qu'elles puissent parfois se trouver dans le cytosol ou la membrane du réticulum endoplasmique. Cette sous-famille comprend les protéines Bcl-2, Bcl-XL, Mcl-1, Bcl-W, Bcl2A1 et Bcl-B [42].

Elles se distinguent principalement par leur capacité à inhiber l'apoptose et présentent une région N-terminale variable en longueur, contenant une ou plusieurs hélices alpha ainsi que le domaine BH4. Ce domaine, caractérisé par une hélice α amphipatique, a été identifié grâce à une comparaison de séquences entre Bcl-2, Bcl-xL et Bcl-w [43].

I .2.3.1.2. Les protéines pro-apoptiques multi-domaines :

Ces protéines ont d'abord été identifiées par la présence de trois domaines BH (BH1 à BH3) et l'absence de domaine BH4. Cependant, l'alignement des protéines Bcl-2 basé sur leur structure a révélé la présence d'un motif BH4 conservé également présent dans les protéines pro-apoptiques multi-domaines [44].

BAX et BAK sont les principaux membres de ce sous-groupe des protéines de la famille BCL-2. Pendant l'apoptose, ces protéines pro-apoptiques subissent un changement de conformation conduisant à leur oligomérisation et à leur insertion dans la membrane externe de la mitochondrie, formant des pores. Cela entraîne la perméabilisation de la membrane externe de la mitochondrie et l'activation de la cascade des caspases [45]. Chez les mammifères, une surexpression de ces effecteurs pro-apoptiques conduit à la mort cellulaire [46].

I .2.3.1.3. Les protéines pro-apoptiques à BH3 :

Les protéines à BH3-seulement, comme leur nom l'indique, ne comportent qu'un seul domaine d'homologie BH3. Cette famille comprend huit membres : Bid, Bad, Bim, Bmf, Noxa, Puma, Bik et Hrk. Ces protéines agissent comme des "senseurs de la mort", déclenchant l'apoptose en réponse à un stress cellulaire [40].

Bad et Bid se trouvent initialement dans le cytosol sous forme inactive [47]. Lorsqu'elles sont activées par un stimulus apoptotique, elles se dirigent vers les mitochondries [40].

Bim et Bmf sont également localisées dans le cytosol, mais elles sont associées respectivement aux microtubules et aux microfilaments d'actine [48]. , ce qui en fait des sentinelles de la dynamique du cytosquelette [38].

Puma et Noxa sont activées transcriptionnellement par la protéine p53 [49]. se localisent directement dans les mitochondries pour promouvoir l'apoptose [40].

Enfin, Bik et Hrk possèdent un domaine hydrophobe C-terminal qui leur permet de se localiser dans les membranes intracellulaires [50]. Avec Bik étant principalement présente dans le réticulum endoplasmique [51].

I .2.4. Structure et Localization :

La protéine anti-apoptotique Bcl-2, constituée de 239 résidus d'acides aminés et ayant une masse moléculaire de 26 266 Da, est fréquemment surexprimée dans de nombreux cancers solides, notamment les cancers des voies aérodigestives supérieures (VADS) [52] , les leucémies [53], et les lymphomes [54].

Elle possède quatre domaines BH et un domaine transmembranaire situé du côté C-terminal, facilitant sa localisation dans les membranes externes mitochondriales (MEM), l'enveloppe nucléaire (EN) et la membrane du réticulum endoplasmique (RE) [55].

La structure secondaire de la protéine est conservée parmi les différentes protéines anti-apoptotiques et se compose de huit hélices- α reliées par des boucles. Les hélices 5 et 6 forment le cœur de la protéine et sont principalement hydrophobes, entourées par d'autres hélices- α , avec les hélices 1 et 2 d'un côté et les hélices 3 et 4 de l'autre côté [56]. De plus, les hélices 3, 4, 5 et 6 constituent une poche hydrophobe significative qui facilite ses interactions avec les membres pro-apoptotiques de la famille Bcl-2 [56]. Bcl-2 comporte également une boucle non structurée qui joue un rôle dans ses modifications post-traductionnelles [42].

Effectivement, Bcl-2 peut subir des phosphorylations à différents résidus (thréonine 69, sérines 70 et 87) situés dans cette région non structurée. Les conséquences de ces phosphorylations varient en fonction du type cellulaire et des stimuli cellulaires : elles peuvent entraîner une inactivation de la protéine et déclencher l'apoptose [57]. Ces phosphorylations peuvent également stabiliser la protéine et renforcer son interaction avec Bax, renforçant ainsi sa propriété anti-apoptotique [58]. En fin , La phosphorylation de Bcl-2 joue également un rôle dans la régulation de l'homéostasie calcique [59] et l'autophagie [60].

L'expression de la protéine Bcl-2 est également régulée au niveau post-transcriptionnel par l'intermédiaire de différents micro-ARN (miARN), tels que les miARN-15 et miARN-16 [61], miARN-129 [62], miRNA 30b [63].

I .2.5. Implication de la famille Bcl-2 dans la cancer :

Le constat que plusieurs médicaments anticancéreux induisent l'apoptose suggère que les tumeurs résistantes à la chimiothérapie pourraient être incapables d'activer ce processus de mort cellulaire [64]. En effet, des études ont montré que le développement de la résistance dans différentes cellules cancéreuses est lié, au moins dans certains cas, à une surexpression de protéines anti-apoptotiques telles que Bcl-2 et Bcl-XL [65]. La sensibilité au cisplatine in vitro et in vivo a été associée à la surexpression de Bax, une protéine pro-apoptotique [66].

Ces protéines font partie d'une vaste famille de protéines impliquées dans la régulation de l'apoptose, connue sous le nom de famille Bcl-2. Certains de ces membres agissent comme des suppresseurs de la mort cellulaire, tels que Bcl-2, Bcl-XL et Mcl-1, tandis que d'autres agissent comme des promoteurs de la mort cellulaire, comme Bax, Bcl-Xs, Bak, Bad et Bik [64]. La sensibilité des cellules à un signal de mort cellulaire semble être en partie déterminée par l'équilibre entre les protéines pro-apoptotiques et anti-apoptotiques [64]. Dans ce cadre, il a été observé que la surexpression artificielle de Bcl-2 et Bcl-XL confère une résistance à la chimiothérapie dans des tests de survie à court terme in vitro dans diverses lignées cellulaires tumorales [65].

Leur implication dans la résistance aux médicaments est également renforcée par le constat que plusieurs lignées cellulaires cancéreuses résistantes aux médicaments surexpriment souvent Bcl-2 ou Bcl-XL [64]. D'autre part, des études ont révélé que la surexpression de Bax et de Bad pouvait augmenter la sensibilité des cellules cancéreuses humaines in vitro à certaines drogues chimiothérapeutiques [67]. Il a été signalé que l'inactivation de protéines pro-apoptotiques telles que Bax et Bad est observée dans des cellules de cancer ovarien, ce qui semble réduire leur sensibilité à différentes drogues [68]. L'inactivation de Bcl-2 et Bcl-XL dans des cellules exprimant différents niveaux de ces deux protéines conduit à l'induction de l'apoptose [69].

Les altérations de la voie apoptotique mitochondriale sont fréquentes dans les cellules cancéreuses. Étant donné l'importance des membres de la famille Bcl-2 dans la régulation de cette voie, il n'est pas étonnant que ces gènes soient modifiés dans les tumeurs. En effet, comme mentionné précédemment, Bcl-2 a été initialement identifié en raison de sa translocation dans des lymphomes folliculaires et est surexprimé dans divers types de cancers [65].

Outre les lymphomes folliculaires, les taux de Bcl-2 sont élevés dans plusieurs autres cancers humains, notamment les carcinomes du sein, de la prostate, gastrique, du foie .

Un grand nombre d'études ont examiné les effets des antisens (AS) ciblant la protéine Bcl-2 dans divers types de cellules cancéreuses. Il a été démontré qu'un antisens ciblant la protéine Bcl-2 inhibe la croissance des cellules de carcinome de Merkel [70].

Un antisens (AS) ciblant la protéine Bcl-XL, un autre membre de la famille des protéines Bcl-2, a été utilisé dans des cellules cancéreuses de la prostate, ce qui les rend plus sensibles à divers agents chimiothérapeutiques [71]. . De plus, un autre antisens dirigé contre Bcl-XL a été capable d'induire l'anoïkose, une forme d'apoptose déclenchée par la perturbation ou la perte d'adhérence cellulaire, et a supprimé la croissance des cellules tumorales ovariennes [72]. Par ailleurs, le groupe de Piché et al. a réussi à moduler les niveaux de protéines Bcl-2 en utilisant un anticorps monovalent modifié dans les cellules cancéreuses du sein MCF-7, augmentant ainsi leur sensibilité à divers médicaments [73]. Ces résultats suggèrent fortement le rôle de la protéine Bcl-2 dans le phénomène de résistance aux traitements [64].

En fin, la plupart des recherches visant à élucider le rôle d'une protéine de la famille Bcl-2 dans le phénomène de résistance s'appuient sur la surexpression de cette protéine dans une lignée cellulaire spécifique au type de cancer étudié. Toutefois, cette surexpression peut entraîner une surcharge cellulaire, rendant difficile la détermination précise de son action [64].

Les Références :

- [1]. Garcia, M., Jemal, A. W. E. C. M. H. Y. S. R. T. M. J., Ward, E. M., Center, M. M., Hao, Y., Siegel, R. L., & Thun, M. J. (2007). Global cancer facts & figures 2007. *Atlanta, GA: American cancer society*, 1(3), 52.
- [2]. Stratton, M. R., Campbell, P. J., & Futreal, P. A. (2009). The cancer genome. *Nature*, 458(7239), 719-724.
- [3]. République Algérienne Démocratique et Populaire PLAN NATIONAL CANCER 2019 Nouvelle vision stratégique centrée sur le malade Octobre 2014
- [4]. Ghoshal, S., Rigney, G., Cheng, D., Brumit, R., Gee, M. S., Hodin, R. A., ... & Succi, M. D. (2022). Institutional surgical response and associated volume trends throughout the COVID-19 pandemic and postvaccination recovery period. *JAMA network Open*, 5(8), e2227443-e2227443.
- [5]. Benson, J. R., & Jatoi, I. (2012). The global breast cancer burden. *Future oncology*, 8(6), 697-702.
- [6]. World Health Organization. (2022, February 2). Cancer.
- [7]. Goret, L. (2020). Prise en charge du cancer du sein: thérapeutiques actuelles et rôle du pharmacien d'officine (Doctoral dissertation).
- [8]. Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *cell*, 100(1), 57-70.
- [9]. Feitelson, M. A., Arzumanyan, A., Kulathinal, R. J., Blain, S. W., Holcombe, R. F., Mahajna, J., ... & Nowsheen, S. (2015, December). *Sustained proliferation in cancer: Mechanisms and novel therapeutic targets*. In *Seminars in cancer biology* (Vol. 35, pp. S25-S54). Academic Press.
- [10]. Joncas, F. H. (2017). *Localisation intracellulaire des isoformes d'Akt dans la chimiorésistance des cancers gynécologiques* (Doctoral dissertation, Université du Québec à Trois-Rivières).
- [11]. Leclers, D. (2005). *Conception de récepteurs circulants anti-lymphangiogéniques par clonage de la séquence codante du gène humain VEGFR-3:: application en cancérologie* (Doctoral dissertation, Limoges).
- [12]. Hamahand W., Wenberg R., 1996. Article de référence sur .la biologie moléculaire du cancer. The hallmarks of cancer cell., 100: 57-70.
- [13]. American Cancer Society. (2022). *Breast Cancer Facts & Figures 2022-2024*.
- [14]. République française institut national du cancer.
- [15]. Cliniques universitaires saint- luc, ucl bruxelles, service de chirurgie et transplantation abdominale, brochure à l'attention des patients et de leur famille

- [16]. Société canadienne du cancer
- [17]. Fondation ARC pour la recherche sur le cancer.
- [18]. Centre Léon Bérard de lutte contre le cancer.
- [19]. Centre de cancérologie de la porte de saint-cloud, American Hospital of Paris.
- [20]. Ligue contre le cancer pour la vie, Recherche Information - prévention - dépistage Actions pour les malades et leurs proches, BRK002 Juin _ 2011.
- [21]. Un guide de la Ligue contre le cancer pour les personnes touchées et leurs proches, Les traitements médicamenteux du cancer Chimiothérapies et autres traitements
- [22]. Shamas-Din, A., Kale, J., Leber, B., & Andrews, D. W. (2013). Mechanisms of action of Bcl-2 family proteins. *Cold Spring Harbor perspectives in biology*, 5(4), a008714.
- [23]. République française institut national du cancer 04/01/2021.
- [24]. Coisy-Quivy, M. (2004). *Rôle de Brm dans le contrôle du cycle cellulaire et Étude de l'équilibre prolifération/différenciation des kératinocytes* (Doctoral dissertation, Université Montpellier II-Sciences et Techniques du Languedoc).
- [25]. Meijer, L. (2003). Le cycle de division cellulaire et sa régulation. *ONCOLOGIE PARIS-*, 5(7/8), 311-326.
- [26]. Université Oran 1 Ahmed Benbella , Faculté de Médecine, Département de Médecine, Service d'Histologie-Embryologie Pr Ag Belarbi-Amar N Cours de Cytologie des étudiants de 1ère année de Médecine.
- [27]. Kachaner, D. (2012). *Optineurine, un nouveau régulateur de la mitose* (Doctoral dissertation, Université Pierre et Marie Curie-Paris VI).
- [28]. Bekkal Brikci, F. (2005). Modélisation du cycle cellulaire et couplage avec la dynamique de population cellulaire (Doctoral dissertation, Paris 6).
- [29]. Ricci, E. P., Kucukural, A., Cenik, C., Mercier, B. C., Singh, G., Heyer, E. E., ... & Moore, M. J. (2014). Staufen1 senses overall transcript secondary structure to regulate translation. *Nature structural & molecular biology*, 21(1), 26-35.
- [30]. Mary, H. (2015). *Analyse et modélisation de la dynamique des chromosomes durant la mitose chez la levure à fission* (Doctoral dissertation, Université de Toulouse, Université Toulouse III-Paul Sabatier).
- [31]. Kim, Y. K., Furic, L., Parisien, M., Major, F., DesGroseillers, L., & Maquat, L. E. (2007). *Staufen1 regulates diverse classes of mammalian transcripts*. *The EMBO journal*, 26(11), 2670-2681.

- [32]. ROBINET C. (2011). *Rôles et régulation du PI(4,5) P2 dans le remodelage cortical et la morphogénèse cellulaire en mitose*. Thèse de doctorat en cancérologie, Université Toulouse III – Paul Sabatier, France.
- [33]. Herbert, M., Levasseur, M., Homer, H., Yallop, K., Murdoch, A., & McDougall, A. (2003). *Homologue disjunction in mouse oocytes requires proteolysis of securin and cyclin B1*. *Nature cell biology*, 5(11), 1023-1025.
- [34]. Doxsey, S., Zimmerman, W., & Mikule, K. (2005). *Centrosome control of the cell cycle*. *Trends in cell biology*, 15(6), 303-311.
- [35]. Mayer, T. U., Kapoor, T. M., Haggarty, S. J., King, R. W., Schreiber, S. L., & Mitchison, T. J. (1999). *Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen*. *Science*, 286(5441), 971-974.
- [36]. Callen, J. C., & Perasso, R. (2005). *Biologie cellulaire-2e éd.: Des molécules aux organismes*. Dunod.
- [37]. Cheeseman, I. M., & Desai, A. (2008). Molecular architecture of the kinetochore–microtubule interface. *Nature reviews Molecular cell biology*, 9(1), 33-46.
- [38]. Colombel, V. (2009). *Synthèse et étude de l'activité biologique de nouveaux analogues du N-acétylcolchinol* (Doctoral dissertation, Université Claude Bernard-Lyon I).
- [39]. Luch, A. (2002). *Cell cycle control and cell division: implications for chemically induced carcinogenesis*. *ChemBiochem*, 3(6), 506-516.
- [40]. Bonneau, B. (2013). *Implication des protéines de la famille Bcl-2 dans la régulation des flux calciques au cours du développement embryonnaire précoce du poisson zèbre* (Doctoral dissertation, Université Claude Bernard-Lyon I).
- [41]. Chittenden, T., Flemington, C., Houghton, A. B., Ebb, R. G., Gallo, G. J., Elangovan, B., ... & Lutz, R. J. (1995). *A conserved domain in Bak, distinct from BH1 and BH2, mediates cell death and protein binding functions*. *The EMBO journal*, 14(22), 5589-5596.
- [42]. Gilormini, M. (2015). *Inhibition des protéines anti-apoptotiques de la famille Bcl-2 par l'ABT-737: intérêt pour le traitement des cancers des voies aérodigestives supérieures* (Doctoral dissertation, Université Claude Bernard-Lyon I).
- [43]. Reed, J. C., Zha, H., Aime-Sempe, C., Takayama, S., & Wang, H. G. (1996). Structure—Function Analysis of Bcl-2 Family Proteins: Regulators of Programmed Cell Death. *Mechanisms of Lymphocyte Activation and Immune Regulation VI: Cell Cycle and Programmed Cell Death in the Immune System*, 99-112.
- [44]. Kvensakul, M., Yang, H., Fairlie, W. D., Czabotar, P. E., Fischer, S. F., Perugini, M. A., ... & Colman, P. M. (2008). *Vaccinia virus anti-apoptotic FIL is a novel Bcl-2-like domain-swapped*

dimer that binds a highly selective subset of BH3-containing death ligands. Cell Death & Differentiation, 15(10), 1564-1571.

[45]. Chipuk, J. E., & Green, D. R. (2008). How do BCL-2 proteins induce mitochondrial outer membrane permeabilization?. *Trends in cell biology, 18(4), 157-164.*

[46]. Brayer, S. (2013). *Les fonctions non-apoptotiques et pro-fibrosantes de la protéine pro-apoptotique BAX dans la fibrose pulmonaire idiopathique* (Doctoral dissertation, Paris Est).

[47]. Li, H., Zhu, H., Xu, C. J., & Yuan, J. (1998). Cleavage of BID by caspase 8 mediates the mitochondrial damage in the Fas pathway of apoptosis. *Cell, 94(4), 491-501.*

[48]. Puthalakath, H., Huang, D. C., O'Reilly, L. A., King, S. M., & Strasser, A. (1999). The proapoptotic activity of the Bcl-2 family member Bim is regulated by interaction with the dynein motor complex. *Molecular cell, 3(3), 287-296.*

[49]. Nakano, K., & Vousden, K. H. (2001). PUMA, a novel proapoptotic gene, is induced by p53. *Molecular cell, 7(3), 683-694.*

[50]. Inohara, N., Ding, L., Chen, S., & Núñez, G. (1997). harakiri, a novel regulator of cell death, encodes a protein that activates apoptosis and interacts selectively with survival-promoting proteins Bcl-2 and Bcl-XL. *The EMBO journal.*

[51]. Germain, M., Mathai, J. P., & Shore, G. C. (2002). BH-3-only BIK functions at the endoplasmic reticulum to stimulate cytochrome c release from mitochondria. *Journal of Biological Chemistry, 277(20), 18053-18060.*

[52]. Sharma, H., Sen, S., Mathur, M., Bahadur, S., & Singh, N. (2004). Combined evaluation of expression of telomerase, survivin, and anti-apoptotic Bcl-2 family members in relation to loss of differentiation and apoptosis in human head and neck cancers. *Head & Neck: Journal for the Sciences and Specialties of the Head and Neck, 26(8), 733-740.*

[53]. Campos, L., Rouault, J. P., Sabido, O., Oriol, P., Roubi, N., Vasselon, C., ... & Guyotat, D. (1993). High expression of bcl-2 protein in acute myeloid leukemia cells is associated with poor response to chemotherapy.

[54]. Gaulard, P., d'Agay, M. F., Peuchmaur, M., Brousse, N., Gisselbrecht, C., Solal-Celigny, P., ... & Mason, D. Y. (1992). Expression of the bcl-2 gene product in follicular lymphoma. *The American journal of pathology, 140(5), 1089.*

[55]. Krajewski, S., Tanaka, S., Takayama, S., Schibler, M. J., Fenton, W., & Reed, J. C. (1993). Investigation of the subcellular distribution of the bcl-2 oncoprotein: residence in the nuclear envelope, endoplasmic reticulum, and outer mitochondrial membranes. *Cancer research, 53(19), 4701-4714.*

- [56]. Petros, A. M., Medek, A., Nettesheim, D. G., Kim, D. H., Yoon, H. S., Swift, K., ... & Fesik, S. W. (2001). Solution structure of the antiapoptotic protein bcl-2. *Proceedings of the National Academy of Sciences*, 98(6), 3012-3017.
- [57]. Haldar, S., Basu, A., & Croce, C. M. (1998). Serine-70 is one of the critical sites for drug-induced Bcl2 phosphorylation in cancer cells. *Cancer research*, 58(8), 1609-1615.
- [58]. Ruvolo, P. P., Clark, W., Mumby, M., Gao, F., & May, W. S. (2002). A functional role for the B56 α -subunit of protein phosphatase 2A in ceramide-mediated regulation of Bcl2 phosphorylation status and function. *Journal of Biological Chemistry*, 277(25), 22847-22852.
- [59]. Bassik, M. C., Scorrano, L., Oakes, S. A., Pozzan, T., & Korsmeyer, S. J. (2004). Phosphorylation of BCL-2 regulates ER Ca²⁺ homeostasis and apoptosis. *The EMBO journal*, 23(5), 1207-1216.
- [60]. Wei, Y., Patingre, S., Sinha, S., Bassik, M., & Levine, B. (2008). JNK1-mediated phosphorylation of Bcl-2 regulates starvation-induced autophagy. *Molecular cell*, 30(6), 678-688.
- [61]. Cimmino, A., Calin, G. A., Fabbri, M., Iorio, M. V., Ferracin, M., Shimizu, M., ... & Croce, C. M. (2005). miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proceedings of the National Academy of Sciences*, 102(39), 13944-13949.
- [62]. Karaayvaz, M., Zhai, H., & Ju, J. (2013). miR-129 promotes apoptosis and enhances chemosensitivity to 5-fluorouracil in colorectal cancer. *Cell death & disease*, 4(6), e659-e659.
- [63]. Liao, W. T., Ye, Y. P., Zhang, N. J., Li, T. T., Wang, S. Y., Cui, Y. M., ... & Ding, Y. Q. (2014). MicroRNA-30b functions as a tumour suppressor in human colorectal cancer by targeting KRAS, PIK3CD and BCL2. *The Journal of pathology*, 232(4), 415-427.
- [64]. Bélanger, S. (2003). *Résistance au cisplatine dans le cancer ovarien: rôle de la protéine anti-apoptotique Bcl-2?*. éditeur non identifié.
- [65]. Reed, J. C. (1995). Regulation of apoptosis by bcl-2 family proteins and its role in cancer and chemoresistance. *Current opinion in oncology*, 7(6), 541-546.
- [66]. Perego, P., Giarola, M., Righetti, S. C., Supino, R., Caserini, C., Delia, D., ... & Zunino, F. (1996). Association between cisplatin resistance and mutation of p53 gene and reduced bax expression in ovarian carcinoma cell systems. *Cancer research*, 56(3), 556-562.
- [67]. Strobel, T., Swanson, L., Korsmeyer, S., & Cannistra, S. A. (1996). BAX enhances paclitaxel-induced apoptosis through a p53-independent pathway. *Proceedings of the National Academy of Sciences*, 93(24), 14094-14099.
- [68]. Meijerink, J. P., Mensink, E. J., Wang, K., Sedlak, T. W., Slöetjes, A. W., De Witte, T., ... & Korsmeyer, S. J. (1998). Hematopoietic malignancies demonstrate loss-of-function mutations of BAX. *Blood, The Journal of the American Society of Hematology*, 91(8), 2991-2997.

- [69]. Zangemeister-Wittke, U., Leech, S. H., Olie, R. A., Simões-Wüst, A. P., Gautschi, O., Luedke, G. H., ... & Stahel, R. A. (2000). A novel bispecific antisense oligonucleotide inhibiting both bcl-2 and bcl-xL expression efficiently induces apoptosis in tumor cells. *Clinical cancer research*, 6(6), 2547-2555.
- [70]. Schlagbauer-Wadl, H., Klosner, G., Heere-Ress, E., Waltering, S., Moll, I., Wolff, K., ... & Jansen, B. (2000). Bcl-2 antisense oligonucleotides (G3139) inhibit Merkel cell carcinoma growth in SCID mice. *Journal of investigative dermatology*, 114(4), 725-730.
- [71]. Lebedeva, I., Rando, R., Ojwang, J., Cossum, P., & Stein, C. A. (2000). Bcl-xL in prostate cancer cells: effects of overexpression and down-regulation on chemosensitivity. *Cancer Research*, 60(21), 6052-6060.
- [72]. Frankel, A., Rosen, K., Filmus, J., & Kerbel, R. S. (2001). Induction of anoikis and suppression of human ovarian tumor growth in vivo by down-regulation of Bcl-XL. *Cancer research*, 61(12), 4837-4841.
- [73]. Piché, A., Grim, J., Rancourt, C., Gómez-Navarro, J., Reed, J. C., & Curiel, D. T. (1998). Modulation of Bcl-2 protein levels by an intracellular anti-Bcl-2 single-chain antibody increases drug-induced cytotoxicity in the breast cancer cell line MCF-7. *Cancer research*, 58(10), 2134-2140.

CHAPITRE 02 :
LES TECHNIQUES DE
MODELISATION MOLECULAIRE

II .1. Introduction :

Les ordinateurs jouent un rôle indispensable dans la chimie pharmaceutique moderne, tant pour la découverte que pour le développement de nouveaux médicaments. Les avancées rapides dans les logiciels et le matériel informatique ont permis aux pharmaco-chimistes d'utiliser couramment les ordinateurs dans leurs laboratoires, à condition qu'ils maîtrisent les bases de la mécanique quantique et d'autres équations relatives aux molécules [1].

Pour répondre à ce besoin croissant, deux disciplines de la "chimie computationnelle" ont émergé : les relations structure-activité (QSAR - Quantitative Structure-Activity Relationships) et les relations structure-propriété (QSPR - Quantitative Structure-Property Relationships) [2].

Le terme QSAR, ou Relation Quantitative entre la Structure et l'Activité, désigne un domaine de recherche visant à établir des corrélations entre les structures moléculaires, représentées par des descripteurs numériques, et leurs propriétés ou activités biologiques. Cette discipline utilise des méthodes de modélisation informatique et mathématique pour parvenir à ses objectifs [3].

L'analyse des relations structure-activité quantitative (QSAR) repose sur le principe général de la chimie médicinale selon lequel l'activité biologique d'un ligand ou d'un composé est liée à sa structure moléculaire ou à ses propriétés, et des molécules structurellement similaires peuvent avoir des activités biologiques similaires [4]. De telles informations structurales moléculaires sont encodées dans des descripteurs moléculaires, et un modèle QSAR définit des relations mathématiques entre les descripteurs et les activités biologiques des ligands connus afin de prédire les activités des ligands inconnus. Les méthodes QSAR ont été appliquées dans plusieurs domaines scientifiques, notamment la chimie, la biologie, la toxicologie et la découverte de médicaments, pour prédire et classer les activités biologiques de composés virtuels ou nouvellement synthétisés [5,6].

QSAR s'est révélé être un outil précieux et avantageux, en particulier dans le domaine de la recherche pharmaceutique industrielle, où la disponibilité d'échantillons est souvent limitée et les mesures expérimentales peuvent être dangereuses, longues et coûteuses. Les prédictions fournies par QSAR peuvent guider efficacement le développement d'un pharmacophore optimal, offrant ainsi un soutien précieux pour valider ou rejeter une hypothèse [7,8].

Même sans recourir à des instruments analytiques sophistiqués, QSAR peut fournir des informations cruciales sur les caractéristiques structurales d'un composé impliqué dans une activité biologique. Bien que la dérivation d'une relation directe avec la structure moléculaire puisse être complexe, il est néanmoins possible d'identifier plusieurs facteurs structurels connus sous le nom de descripteurs moléculaires, qui influent sur la propriété moléculaire sélectionnée [3].

II .2. Le criblage virtuel :

Le terme "criblage virtuel" désigne un ensemble de techniques computationnelles visant à explorer des bases de données de composés à la recherche de molécules d'intérêt. Une analogie fréquemment utilisée compare ces techniques à des filtres, permettant de constituer des ensembles de molécules partageant certaines propriétés, de sélectionner les plus susceptibles d'interagir avec une cible donnée et d'éliminer les composés supposés inactifs ou les molécules indésirables [9].

Le criblage virtuel est un élément essentiel du processus de recherche et développement de médicaments. Il permet d'explorer plusieurs chimiothèques de manière virtuelle à l'aide d'algorithmes et de formules de calcul pour découvrir de nouvelles substances actives. Ces méthodes se sont avérées efficaces dans de nombreuses études, atteignant des taux de réussite concurrentiels par rapport aux méthodes de criblage classiques telles que le criblage à haut débit (HTS) [10]

Selon la nature des données expérimentales disponibles, la méthode de criblage in silico se divise en deux approches distinctes. La première repose sur la structure de la cible pour identifier les interactions potentielles au sein du site de liaison, permettant ainsi de sélectionner des molécules capables de s'y lier. Cette approche est connue sous le nom de "structure-based virtual screening» [11,12]. La seconde utilise des ligands connus comme points de référence pour définir les propriétés responsables de leur affinité pour la cible, et ainsi identifier des molécules similaires. Cette méthode est appelée "ligand-based virtual screening" (Figure II.1) [13,14].

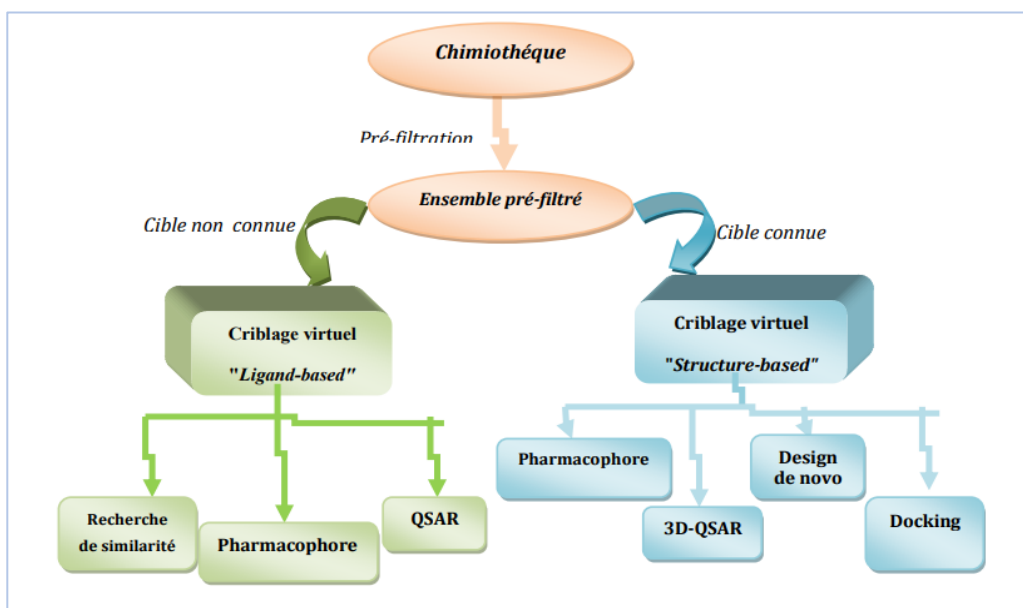


Figure II.1 : Classification des méthodes de criblage virtuel « ligand-based » et « structure-based ».

L'adoption de ces méthodes est déterminée par les données expérimentales disponibles sur le système en question [15]. Par conséquent, lorsque la configuration tridimensionnelle de la cible biologique est élucidée (par cristallographie, diffraction des rayons X, résonance magnétique nucléaire (RMN), microscopie électronique ou modélisation de structure), le choix de l'approche est guidé en conséquence" [16].

Ces techniques peuvent être employées de manière indépendante ou combinée. L'intégration de ces deux approches au sein d'une même étude offre une opportunité d'accroître les probabilités de renforcer ou de perfectionner des molécules déjà identifiées[17].

Chimiothèques :

Avant de débiter les criblages, il est nécessaire de choisir une chimiothèque de composés. Le succès des criblages dépend en grande partie de cette sélection, tenant compte de la taille, de la composition, de la diversité et des caractéristiques des molécules incluses dans la chimiothèque [18].

II .2.1. Criblage virtuel « ligand-based » :

Quand au moins un ligand de la cible étudiée est identifié, on peut mettre en œuvre un criblage virtuel basé sur les ligands, également connu sous le nom de "ligand-based". Le principe fondamental commun à toutes les méthodes basées sur les ligands est que des molécules similaires ont tendance à présenter des profils d'activité similaires [19]. La similitude entre les molécules peut être évaluée en recherchant des propriétés communes, qui serviront de descripteurs de similitude. Selon le nombre de ligands de référence disponibles pour la cible et le type de descripteurs utilisés, différentes méthodes peuvent être employées : la recherche de similitude, le criblage à l'aide de pharmacophores et les méthodes QSAR [20].

II .2.1.1. Recherche de similarité :

La recherche de similarité est la stratégie à adopter lorsque peu de ligands ont été répertoriés pour la cible biologique sélectionnée. Dès qu'un ligand actif est identifié, il est possible d'entreprendre une recherche de similarité [21]. Cette approche repose sur l'utilisation de descripteurs et de mesures de similitude pour comparer les molécules à cribler avec un ou plusieurs ligands de référence, permettant ainsi de prédire leur profil d'activité[20].

II .2.1.2. Modèles pharmacophoriques « ligand-based » :

Le pharmacophore est défini comme l'agencement tridimensionnel des caractéristiques nécessaires à une molécule pour avoir une activité biologique spécifique. Il englobe les propriétés stériques et électrostatiques, telles que les donneurs et accepteurs de liaisons hydrogènes, les interactions électrostatiques, ainsi que les groupes aromatiques et hydrophobes, essentiels pour établir des interactions optimales avec une cible biologique spécifique, permettant ainsi de déclencher ou de bloquer une réponse biologique. Les pharmacophores servent de référence pour le criblage de chimiothèques, en identifiant les molécules qui correspondent à ces caractéristiques. Un pharmacophore est dit "ligand-based" lorsqu'il est établi à partir de composés actifs utilisés comme référence, sans tenir compte de la structure de leur récepteur. La plupart des méthodes de construction de pharmacophore fonctionnent selon le même principe (**Figure II.2**) [22].

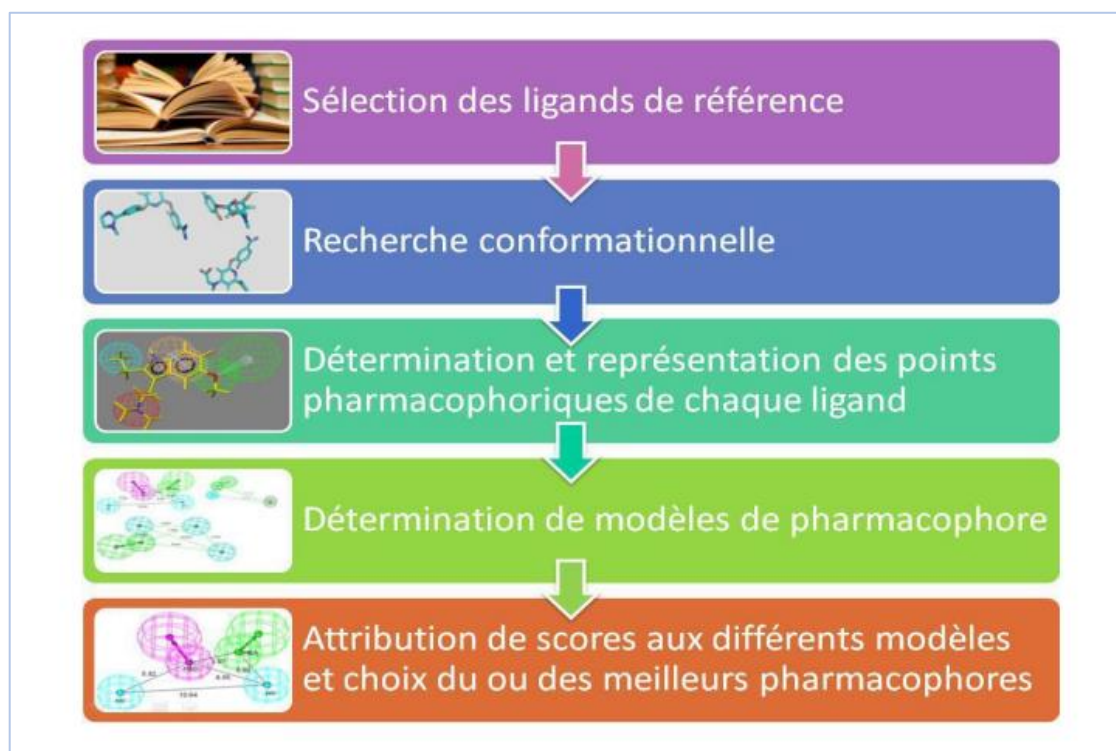


Figure II.2 : Principales étapes de l'élucidation d'un pharmacophore.

II .2.1.3. Quantitative Structure-Activity Relationship (QSAR) « ligand-based »:

L'analyse quantitative des relations entre les structures d'un groupe de composés et leurs activités permet d'identifier et d'évaluer l'impact des propriétés influençant l'activité biologique. Cette corrélation peut être exprimée par une équation mathématique qui lie les influences réciproques des paramètres concernés. L'extrapolation de ces résultats peut être utilisée pour prédire l'activité de nouveaux composés [23].

Dans le cadre d'une étude QSAR (**Figure II.3**), il est nécessaire de calculer tous les descripteurs moléculaires possibles à partir de la structure moléculaire de plusieurs composés actifs. Après avoir éliminé les descripteurs dont la variation est négligeable sur l'ensemble des molécules, une analyse multivariée suivie d'une évaluation statistique est réalisée afin d'obtenir le modèle QSAR. Une dernière étape de validation du modèle est indispensable pour garantir sa fiabilité [20].

Il existe plus de 3000 descripteurs moléculaires disponibles pour les études QSAR [24], provenant de diverses théories et méthodes, et pouvant être dérivés soit directement à partir de la structure chimique, soit grâce à l'utilisation de logiciels spécialisés[20]. Les descripteurs les plus couramment utilisés comprennent les descripteurs topologiques (2D-QSAR), les descripteurs géométriques ou basés sur des grilles (3D-QSAR). De nouvelles approches plus avancées, telles que QSAR-4D, -5D et -6D, ont été développées pour améliorer les approches QSAR-3D. Ces approches prennent en compte différentes conformations des ligands (4D) [25], l'adaptation structurale de la cible au ligand (5D) [26], ainsi que les effets des solvants (6D) [27].

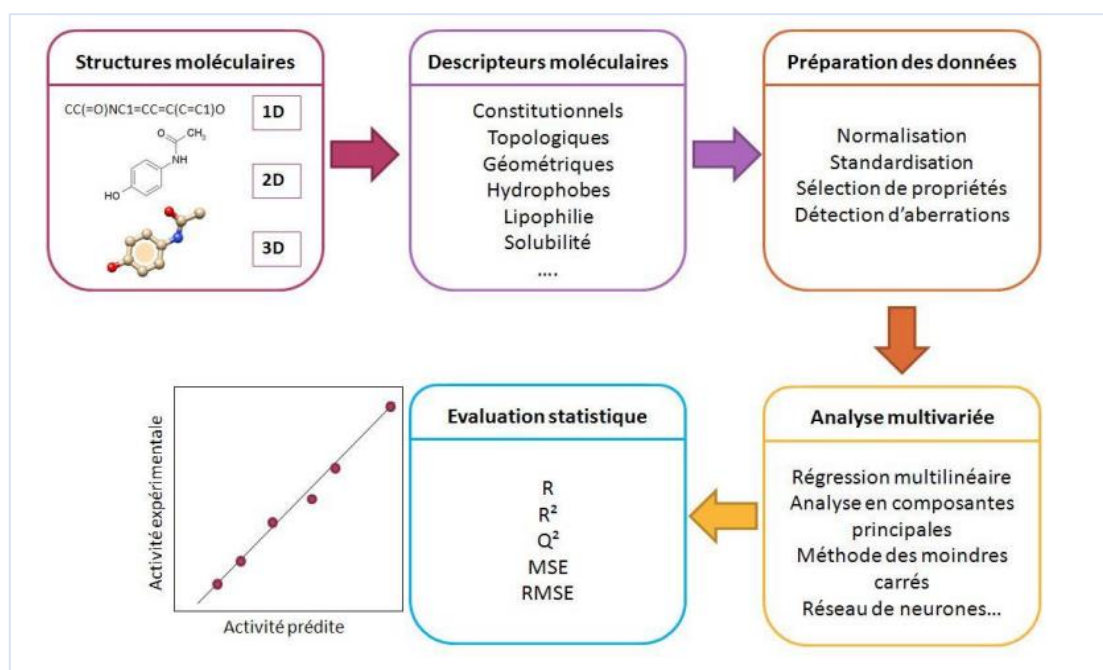


Figure II.3: Schéma général des différentes étapes d'une étude QSAR.

II .2.1.3.1. Les étapes principales pour la construction du modèle QSAR :

En général, pour prédire une activité biologique ou une propriété par la méthode QSAR, la construction du modèle mathématique se déroule en plusieurs étapes principales (**Figure II.4**) [28].

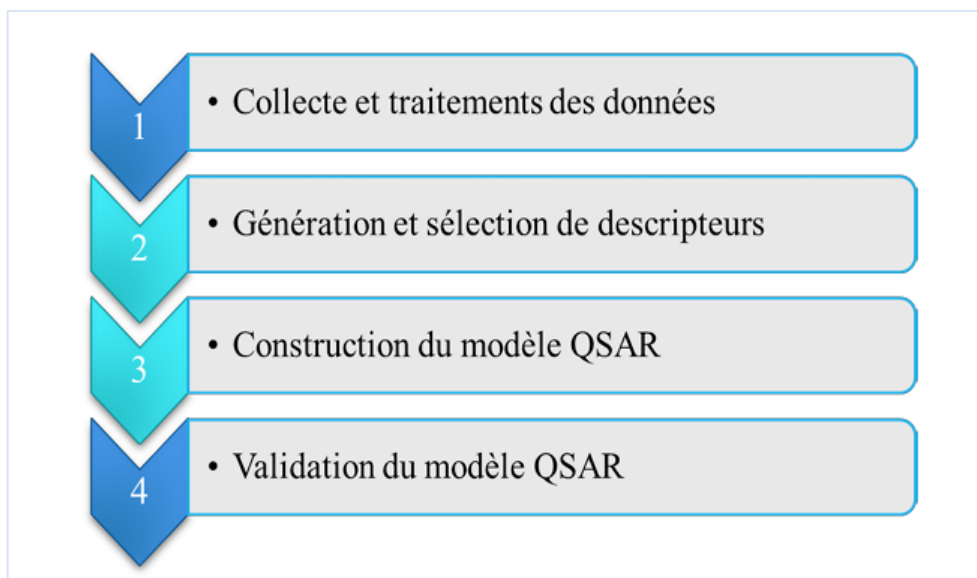


Figure II.4: Etapes effectuées pour le développement d'un modèle QSAR.

II .2.1.3.1.1. Collecte et traitement de l'ensemble des données :

Les modèles QSAR sont fortement tributaires des données expérimentales utilisées pour leur élaboration. Ainsi, il est essentiel de sélectionner soigneusement l'ensemble de données à modéliser, car cela représente une étape cruciale dans la construction des modèles QSAR [28].

Les données utilisées idéalement devraient être de haute qualité, fiables et homogènes, avec des incertitudes minimales pour réduire les erreurs expérimentales [28].

L'efficacité d'un modèle QSAR dépend également du type de molécules incluses. Plus le modèle contient de composés présentant des structures proches et similaires, plus il a de chances d'être performant. La présence de molécules dont le comportement est atypique peut influencer considérablement le modèle final [28].

II .2.1.3.1.2. Génération sur les Descripteurs moléculaires :

Les descripteurs sont des valeurs numériques qui codent les caractéristiques physicochimiques des molécules à partir de leurs structures. La génération de ces descripteurs est une étape cruciale dans l'analyse de QSAR, et la qualité d'un modèle dépend largement de la nature des descripteurs produits [29].

Le (**Tableau II.1**) présente une méthode de catégorisation des descripteurs en fonction de la dimension chimique utilisée pour leur définition :

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

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

- **Les descripteurs 1D :**

Se déduisent de la formule brute de la molécule et caractérisent les propriétés générales du composé. Ils incluent notamment sa composition atomique ainsi que sa masse molaire. Il convient de noter que ces descripteurs ne permettent pas de différencier les isomères de constitution [30].

- **Les descripteurs 2D :**

sont dérivés de la formule développée de la molécule et peuvent revêtir diverses formes.

Les indices constitutionnels détaillent les constituants moléculaires, comprenant le nombre de liaisons simples ou multiples, ainsi que le nombre de cycles, par exemple.

Les indices topologiques sont obtenus à partir de la structure 2D de la molécule, fournissant des données sur sa taille, sa forme globale et ses ramifications [30]. Parmi les plus couramment utilisés figurent l'indice de Wiener [31], l'indice de Randić [32], l'indice de connectivité de valence de Kier-

Hall [33] et l'indice de Balaban[34]. L'indice de Wiener, par exemple, quantifie le volume moléculaire et la ramification en additionnant les distances topologiques entre les atomes. L'indice de Randić est largement utilisé comme indicateur de l'accessibilité moléculaire au solvant [30].

Bien que ces descripteurs 2D capturent souvent les propriétés physiques, ils peuvent s'avérer insuffisants pour expliquer certaines caractéristiques ou activités, notamment les activités biologiques. Pour pallier cela, des descripteurs basés sur la structure 3D des molécules ont été développés grâce aux avancées technologiques et méthodologiques[30]. (Figure II.5)

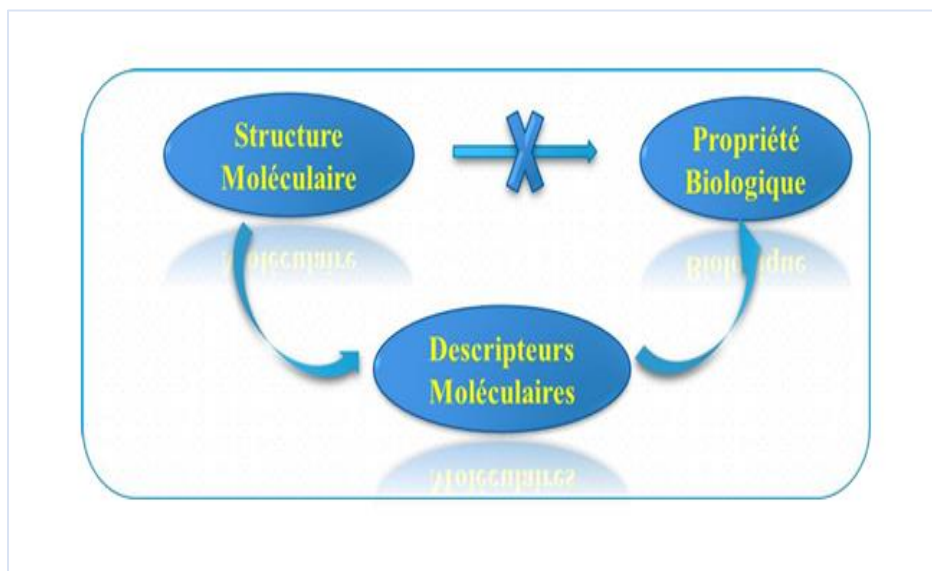


Figure II.5 : Génération de descripteurs moléculaires indispensables pour établir la relation QSAR.

II .2.1.3.1.3. Sélection des descripteurs :

Lorsque de nombreux descripteurs sont collectés, certains d'entre eux peuvent contenir des informations redondantes, créant ainsi un problème de colinéarité. De plus, tous les descripteurs calculés ne contribuent pas forcément à modéliser l'activité[35].

Il est donc nécessaire d'éliminer les descripteurs dont l'influence est inférieure à celle de l'erreur, en ne sélectionnant que les plus pertinents parmi eux. En général, pour qu'un descripteur soit conservé, son retrait doit entraîner une baisse significative des performances du modèle[35].

II .2.1.3.1.4. Construction du modèle QSAR :

Les modèles QSAR sont construits en établissant des relations empiriques, utilisant des stratégies basées sur la régression et la classification, qu'elles soient linéaires ou non linéaires, entre les valeurs des descripteurs chimiques calculés à partir des structures moléculaires et les propriétés ou activités biologiques mesurées expérimentalement de ces molécules. Ces modèles sont ensuite appliqués pour prédire ou concevoir de nouveaux produits chimiques caractéristiques[36].

En plus des méthodes classiques, une variété d'outils d'apprentissage automatique peuvent être employés pour analyser la modélisation QSAR, en particulier pour traiter des données complexes et de grande taille provenant de la chimie informatique. Ces données présentent souvent des relations non linéaires avec les réponses[37].

Les méthodes mathématiques utilisées pour construire des modèles QSAR permettent de résoudre les problèmes de régression et de classification dans la structuration des données. Parmi ces méthodes, l'une des plus répandues pour la classification ou la reconnaissance de formes est la régression linéaire multiple (MLR) [38].

A. Méthodes Statistiques :

Les techniques statistiques ou chimio-métriques fournissent le fondement mathématique de la construction d'un modèle QSAR.

A .1. La régression linéaire multiple (RLM) :

La régression linéaire multiple (RLM) représente la méthode statistique de modélisation la plus simple et la plus couramment utilisée dans les études de la relation structure-activité. Popularisée par Hansch, elle consiste à relier l'activité biologique à des propriétés expérimentales telles que la lipophilie, l'électronique et la stéricité pour des séries de composés [39].

La méthode RLM est fondée sur l'hypothèse d'une relation linéaire entre une variable dépendante Y (ici, l'activité biologique) et un ensemble de n variables indépendantes Xi (les descripteurs moléculaires) [35]. L'objectif est d'obtenir une équation mathématique de la forme (éq II.1) :

$$Y = a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n \quad (\text{éq II.1})$$

La régression linéaire est simple à mettre en œuvre, et les coefficients a_n obtenus peuvent être interprétés : ils mesurent l'influence de chaque variable sur la grandeur étudiée. Cependant, dans de nombreux cas, il est nécessaire d'utiliser des modèles qui prennent en compte la corrélation non linéaire[40].

A .2. Régression des moindres carrés partiels (PLS) :

La régression PLS représente une méthode rapide, efficace et optimale pour minimiser les covariances de manière précise. Elle est particulièrement utile lorsqu'il y a un grand nombre de variables prédictives ou lorsque ces variables présentent une forte colinéarité[41].

Tout d'abord, la PLS réduit les variables prédictives à un ensemble plus restreint de composantes non corrélées. Ensuite, elle effectue la régression par les moindres carrés sur ces composantes plutôt que sur les données initiales[42].

Contrairement à la régression linéaire multiple (MLR), la PLS ne suppose pas que les variables prédictives sont fixes. Ainsi, elle tolère les erreurs dans les mesures de ces variables en prenant en compte les incertitudes des mesures[43].

A .3. Régression non linéaire multiple (RNLM) :

La régression non linéaire multiple (RNLM) est une méthode non linéaire, utilisant des fonctions telles que exponentielle, logarithmique, polynomiale, etc., pour déterminer un modèle mathématique capable d'expliquer de manière optimale la variabilité d'une propriété ou d'une activité Y en fonction des descripteurs moléculaires. La formule générale de la RNLM est donnée par (éq II.2) [44] :

$$Y = a_0 + \sum_{i=1}^n a_i x_i + \dots + b_i x_i^n \quad (\text{éq II.2})$$

Avec :

Y : est la variable dépendante (à expliquer ou à prédire)

X_i : sont les variables indépendantes (explicatives)

n : est le nombre de variables explicatives

a₀ : est la constante de l'équation du modèle

a_i et b_i : sont les coefficients de descripteurs dans l'équation du modèle [45].

B .Test de la signification globale des methodes statistiques :

Pour évaluer la performance d'un modèle, divers paramètres statistiques sont utilisés, notamment les erreurs quadratiques moyennes (Mean Square Errors) et les coefficients de corrélation, fréquemment employés dans les études QSAR. Cette section détaille ces paramètres.

B .1. Coefficient de détermination R^2 :

Le coefficient de corrélation de Bravais-Pearson entre y_i et \hat{y}_i , représentant respectivement les valeurs observées et prédites par le modèle de régression, est noté R. Sa valeur oscille entre 0 et 1 [46]. Mathématiquement, il est défini comme suit (éq II.3) :

$$R^2 = \sqrt{1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2}} \quad (\text{éq II.3})$$

B .2. Coefficient de détermination ajusté R^2_{adj} :

Le coefficient de détermination ajusté R^2_{adj} prend en considération le nombre de variables et leur importance.

Au contraire du R^2 , il n'augmente pas nécessairement lorsque de nouvelles variables sont ajoutées au modèle[47]. On l'explique ainsi (éq II.4) :

$$R^2_{\text{adj}} = 1 - \left[\frac{n-1}{n-p-1} (1 - R^2) \right] \quad (\text{éq II.4})$$

Avec :

n est le nombre des observations (les molécules) ; p est le nombre de variables indépendantes (les descripteurs) ; est R^2 le coefficient de détermination du modèle [48].

B .3. L'erreur quadratique moyenne « MSE » et l'erreur type résiduel « s » :

L'erreur quadratique moyenne mathématiquement, il est défini comme suit (éq II.5) :

$$MSE = \frac{\sum |(\hat{y}_1 - y_1)^2|}{n} \quad (\text{éq II.5})$$

Ou encore, l'erreur type résiduel « s » (éq II.6) :

$$s = \sqrt{\frac{\sum(\hat{y}_i - y_i)^2}{n - p - 1}} \quad (\text{éq II.6})$$

Avec :

y_i et \hat{y}_i sont, respectivement, les valeurs observées et calculées de la variable dépendante ; n est le nombre des observations ; p est le nombre de variables indépendantes.

La variation de l'activité cible non expliquée par le modèle QSAR est mesurée à l'aide de ces paramètres.

Plus la déviation standard est faible, plus la corrélation est élevée. La valeur dépend toujours de l'unité de mesure de l'activité cible et prend également en considération les erreurs expérimentales, ce qui explique qu'une valeur trop faible n'a aucun sens[48].

B .4. Le test de Fisher F :

L'indice de test F de Fisher est utilisé pour évaluer le niveau de signifiante statistique du modèle à "x%" (le niveau habituel étant 95%), ce qui équivaut à évaluer la qualité du choix des paramètres. La conclusion obtenue ne doit pas nous amener à croire que la corrélation a "x%" de chances d'être vraie, mais simplement que la corrélation est vérifiée pour "x%" des composés pris comme référence, avec une abstraction pour les autres.

Hypothèses :

H_0 : les variances des échantillons sont homogènes.

H_1 : les variances des échantillons ne sont pas homogènes.

On calcule le F (observé) à partir de la formule (éq II.7) :

$$F(\text{observé}) = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \hat{y}_i)^2} \frac{n - p - 1}{p} \quad (\text{éq II.7})$$

Avec :

F est l'indice de Fisher ; y_i et \hat{y}_i sont, respectivement, les valeurs observées et calculées de la variable dépendante ; \bar{y} est la valeur moyenne des valeurs prédites ; n est le nombre des observations (les molécules) ; p est le nombre de variables indépendantes (les descripteurs).

Si le F observé est plus grand que le F théorique, cela entraîne le rejet de l'hypothèse nulle H_0 , ce qui indique que les variances des échantillons sont trop différentes pour être considérées comme homogènes.

Si le F observé est plus petit que le F théorique, cela conduit à l'acceptation de l'hypothèse nulle H_1 , ce qui signifie que les deux variances ont des valeurs suffisamment proches pour que l'on accepte l'idée qu'elles soient homogènes[48].

II .2.1.4. Validation du modèle QSAR :

La validation des modèles est essentielle pour évaluer leur fiabilité et déterminer la reproductibilité des résultats ainsi que la pertinence d'un modèle développé pour une application spécifique [49].

Trois méthodes sont couramment utilisées pour la validation d'un modèle QSAR : la validation interne, la validation externe et le test de randomisation[3]. (Figure II.6)

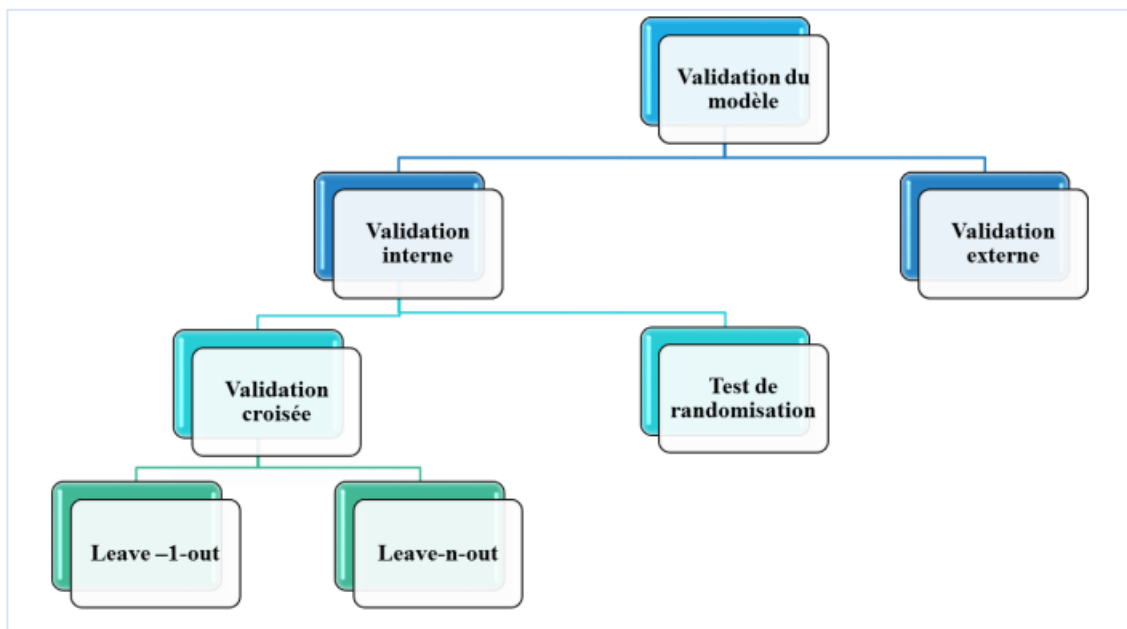


Figure II.6 : Méthodes de validation du modèle QSAR.

II .2.1.4.1. Validation interne :

A .Validation croisée :

La méthode de validation croisée, une technique de validation interne largement utilisée en QSAR, comprend principalement deux procédures : Leave-One-Out (LOO) et Leave-n-Out, où "n" représente le nombre de composés retirés à chaque itération[3].

Dans la procédure Leave-One-Out, une molécule est retirée séquentiellement de l'ensemble d'apprentissage (TSET). Un modèle QSAR est alors construit sur un ensemble de m-1 composés, et la molécule retirée est prédite par le modèle ainsi formé, où "m" représente le nombre total de molécules utilisées dans la construction du modèle (TSET). Cette procédure est répétée m fois afin de prédire les propriétés de toutes les molécules[3].

La procédure Leave-n-Out implique de diviser l'ensemble de données en plusieurs parties. À tour de rôle, chaque partie est utilisée comme ensemble de test interne, tandis que les autres parties constituent l'ensemble d'entraînement. Les molécules de chaque groupe exclu sont prédites par le modèle construit sur l'ensemble d'entraînement. Cette procédure est répétée p fois pour prédire les propriétés de toutes les molécules, où p représente le nombre de groupes de molécules exclus[49].

La performance des modèles de régression est évaluée à l'aide de paramètres statistiques de la validation croisée, tels que l'écart-type de la validation croisée (S_{cv-100}) et le coefficient de détermination de validation (Q^2_{cv-100}), qui mesure la capacité du modèle QSAR à prédire les propriétés des molécules [3].

Q^2_{cv-100} est calculé à partir de l'équation(éq II.8) suivante :

$$Q^2_{cv-100} = \frac{\sum_{i=1}^N (y_i - \bar{y}_i)^2 - \sum_{i=1}^N (y_i - y_{i_{cv}})^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} \quad (\text{éq II.8})$$

Avec :

$y_{i_{cv}}$ est la valeur prévue de la propriété y par la méthode de la validation croisée. \bar{y}_i est la moyenne arithmétique des y.

Q^2_{cv-100} est employé en tant qu'outil de diagnostic pour évaluer le pouvoir prédictif ou la qualité de la prévision d'un modèle.

B. Y-randomisation :

Le pouvoir de prévision interne par la procédure de la validation croisée peut être surestimé. Une valeur élevée du coefficient de corrélation q^2 (ou r^2_{cv}) peut résulter d'une corrélation due au hasard [42] ou de la redondance des structures lorsque les différences entre les composés du jeu d'apprentissage sont minimales (l'autocorrélation) [48] . Pour contrôler la robustesse d'un modèle, on utilise souvent le test d'hasardisation[50]. Dans cette approche de validation, les valeurs de la

variable cible sont redistribuées aléatoirement sur l'ensemble du jeu d'apprentissage et un nouveau modèle est dérivé. L'opération est répétée plusieurs fois et si la moyenne des coefficients de corrélation résultants reste élevée, on peut conclure qu'aucun modèle acceptable ne peut être obtenu par cette méthode statistique sur ce jeu de données[48].

II .2.1.4.2. Validation externe :

Cette méthode implique la prédiction de la propriété ou de l'activité d'une série de molécules, généralement appelée série de test, qui ne fait pas partie de la série utilisée pour développer le modèle. La validation externe est caractérisée par des paramètres tels que R^2 (test) et R^2_{cv} (test) [51]. Récemment, plusieurs études ont souligné les limites des paramètres R^2 et R^2_{CV} pour évaluer la capacité prédictive des modèles QSAR [52]. Par conséquent, d'autres paramètres doivent être pris en compte à cette fin. Ces paramètres sont souvent appelés "critères de validation externe" ou "critères de Trophsa" [52].

II .2.1.5. Domaine d'applicabilité :

Le domaine d'application se réfère à la région de l'espace chimique délimitée par les molécules présentes dans l'ensemble d'apprentissage du modèle. Les modèles QSAR ne peuvent pas prédire de manière fiable les propriétés pour l'ensemble des composés chimiques existants[53].

Un modèle QSAR ne peut être considéré comme universel car il est développé sur un nombre restreint de composés qui ne couvrent pas la totalité de l'espace chimique. Ainsi, l'activité ou la propriété prédite pour un composé chimiquement différent de ceux présents dans l'ensemble d'apprentissage ne peut pas être considérée comme fiable[54]. Un modèle idéal serait celui capable de prédire l'activité ou la propriété de n'importe quelle molécule imaginable, mais cela est souvent difficile à réaliser. La taille limitée de l'ensemble d'entraînement restreint l'espace chimique des modèles construits, et par conséquent, lorsque des molécules se situent en dehors de cet espace, les prédictions ne sont plus fiables[55].

En effet, un modèle QSAR n'est pas conçu pour être utilisé en dehors de son domaine d'application, c'est-à-dire en dehors de l'espace chimique couvert par son ensemble d'apprentissage[53].

II .2.2. Criblage virtuel « structure-based » :

Le criblage virtuel "structure-based" tire parti de la connaissance de la structure de la cible thérapeutique pour repérer des composés actifs. La détermination de la structure tridimensionnelle (3D) d'une protéine est principalement réalisée par la cristallographie aux rayons X ou la résonance magnétique nucléaire (RMN) [56].

En l'absence de la structure de la cible thérapeutique, la modélisation par homologie peut être utilisée [57]. Cette approche consiste à construire un modèle basé sur une ou plusieurs protéines dont la structure est résolue et qui présentent une forte homologie de séquence avec la cible. Cette méthode permet d'obtenir rapidement et à moindre coût les données essentielles pour mettre en œuvre les méthodes de criblage "structure-based" de divers types. Ainsi, environ 25% des études de docking publiées entre 2000 et 2009 auraient utilisé un modèle construit par homologie de séquence [58].

II .2.2.1. Sélection d'un site de liaison :

Si la structure de la cible a été établie expérimentalement et qu'elle est cocrystallisée avec un ligand, le site de liaison du ligand est identifié. Dans ce cas, un espace de recherche peut être défini autour de ce site pour appliquer les méthodes requises. En revanche, en l'absence de cette information, des outils de prédiction des sites de liaison peuvent être utilisés pour guider la recherche [59].

II .2.2.2. Prédiction du site de liaison :

Les outils de prédiction des poches "druggables" sont conçus pour repérer les sites potentiels de liaison d'un ligand sur la cible. On peut les classer en trois catégories : les méthodes fondées sur la connaissance, celles basées sur la géométrie, et celles reposant sur l'énergie d'interaction [60].

II .2.2.3. Modèles pharmacophoriques « structure-based » :

Les pharmacophores 3D décrivent l'agencement spatial des propriétés chimiques nécessaires à l'activité biologique, en se basant sur un ensemble de ligands actifs de référence[61]. Cependant, intégrer la structure 3D du site de liaison, lorsque disponible, permet de créer de nouveaux modèles pharmacophoriques qui intègrent des informations cruciales, rendant ainsi les méthodes pharmacophoriques plus puissantes et exploitant des aspects jusqu'ici sous-utilisés dans les modèles basés sur les ligands[62]. Les pharmacophores 3D peuvent être dérivés de la structure 3D

de la protéine seule, selon une approche dite "receptor-based", ou à partir d'une structure co-cristallisée avec un ligand complexe, appelée approche "complex-based"[63].

II .2.2.4. La relation quantitative structure activité (3D-QSAR) :

La méthode la plus répandue dans la conception de médicaments est la relation quantitative structure-activité tridimensionnelle (3D-QSAR), qui englobe l'analyse comparative du champ moléculaire (CoMFA) et l'analyse comparative des indices de similarité moléculaire (CoMSIA) [64].

Les modèles 3D-QSAR sont utilisés lorsque la structure tridimensionnelle de la cible biologique est connue [65]. Ces approches visent à établir des corrélations entre l'activité biologique d'une série de composés actifs de référence et l'arrangement spatial de nombreuses propriétés moléculaires telles que les propriétés stériques, lipophiles et électroniques. Elles offrent ainsi des indications pour l'optimisation par pharmacomodulation et la conception de nouveaux composés présentant des profils d'activité améliorés[66].

II .2.2.5. La conception de novo :

Cette méthode implique la conception progressive de composés bioactifs en construisant des ligands dans le site actif du récepteur cible, en utilisant la structure du site d'interaction comme point de départ pour la recherche de nouveaux composés actifs [67].

Initialement, toutes les interactions ligand-récepteur, telles que les liaisons hydrogène, électrostatiques et hydrophobes, sont extraites afin d'identifier les sites d'interaction nécessaires pour le placement de divers blocs dans le site d'interaction ligand-récepteur[68]. Ces blocs peuvent être des atomes individuels ou des fragments moléculaires[68]. L'utilisation d'atomes comme blocs permet d'introduire une diversité structurale plus grande, mais génère également un nombre considérable de solutions possibles, rendant la sélection des composés les plus prometteurs complexe[68]. En revanche, l'approche basée sur les fragments permet de restreindre l'espace chimique à étudier, surtout lorsqu'ils sont choisis de manière judicieuse. Par conséquent, les méthodes basées sur les fragments sont les plus couramment utilisées, où les fragments identifiés sont ensuite assemblés pour former de nouveaux composés [69]. (Figure II.7)

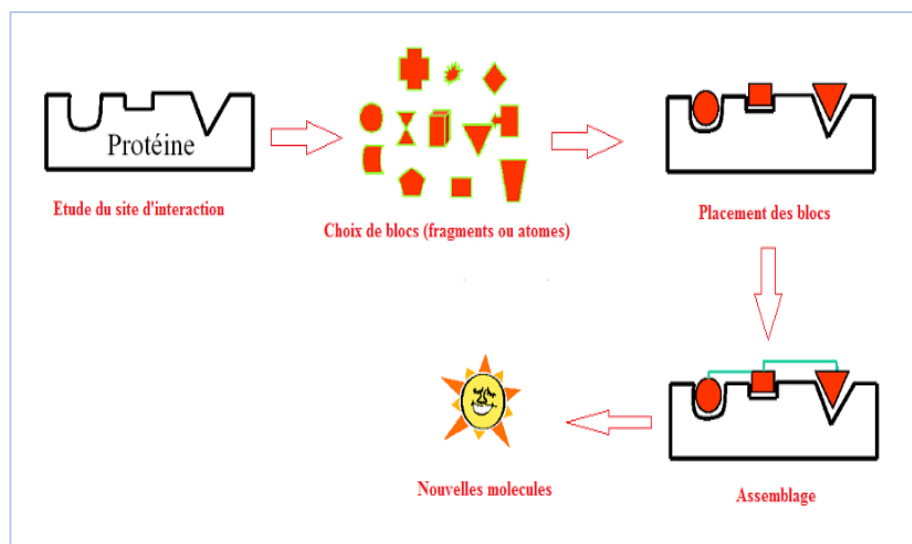


Figure II.7 : Représentation schématique des étapes principales d'une conception de novo.

II .2.2.6. Méthodes de docking :

Le Docking in silico, est une technique informatique visant à prédire la structure d'un complexe formé par deux molécules (une cible et un ligand) ou plus. La formation de ces complexes repose sur la reconnaissance de la structure tridimensionnelle d'un ligand par un site récepteur et sur le contrôle de l'activité de nombreuses molécules. Le récepteur est le plus souvent une protéine dotée d'un ou de plusieurs sites actifs spécifiques, dont l'accessibilité peut varier selon le contexte. Quant au ligand, il s'agit généralement d'une petite molécule flexible et étrangère [70].

Les méthodes de docking sont utilisées depuis plus de 30 ans dans la conception de médicaments, avec deux objectifs principaux :

- l'identification des ligands pour une cible donnée à partir de vastes bibliothèques chimiques, en prédisant les degrés d'affinité ou en déterminant les valeurs réelles d'affinité (K_i , IC_{50} , etc.)
- la prédiction des modes de liaison adoptés par ces ligands dans le complexe formé.

Le processus de docking se divise en deux phases distinctes :

- la recherche algorithmique, qui génère un grand nombre de structures potentielles et détermine les méthodes de communication.

- la notation, utilisant des méthodes mathématiques pour estimer la force de l'interaction et l'affinité de liaison entre les molécules après la phase de docking[71]. (Figure II.8)

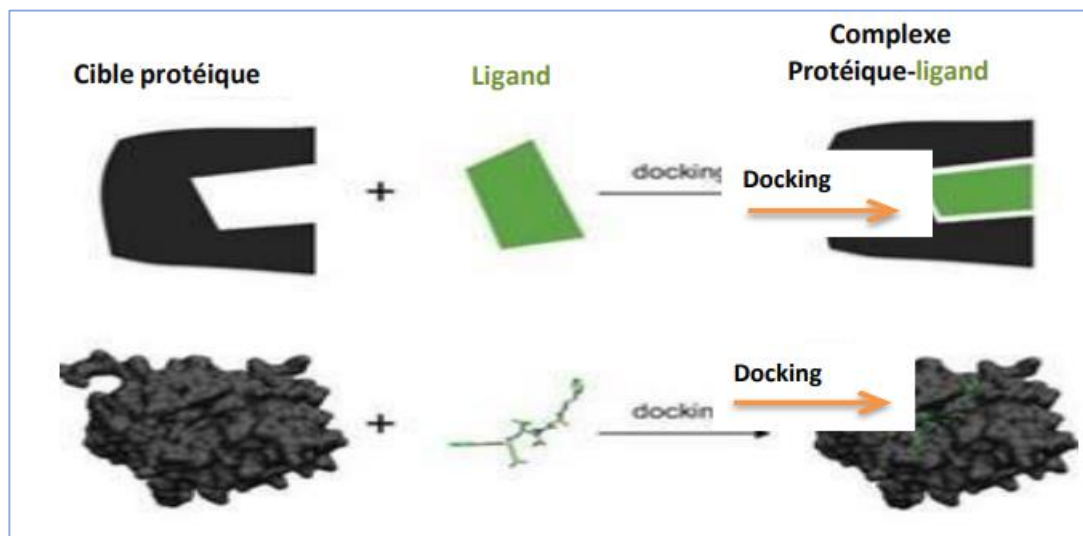


Figure II.8 : Représentation schématique du Docking moléculaire.

II .2.2.6.1. Types du Docking moléculaire :

A .Docking avec ligand rigide :

Pendant une période prolongée, le processus de liaison entre un ligand et son récepteur était conceptualisé comme un mécanisme statique. Selon cette vision, le ligand agissait comme une clé ayant une forme complémentaire à celle de la serrure qu'il était destiné à "ouvrir", représentée par le récepteur. Ce modèle était connu sous le nom de modèle "clé-serrure" ou "lock-and-key"[72].

Ensuite, une représentation négative du site de liaison était utilisée pour éliminer toutes les poses incompatibles avec le site actif, en tenant compte des "clashes" ou des distances.

Enfin, les poses sélectionnées étaient évaluées et les meilleures étaient optimisées[20].

Les méthodes de docking ligand rigide sont appréciées pour leur rapidité et sont souvent utilisées lors de criblages virtuels pour filtrer rapidement les molécules inadéquates, comme celles qui sont trop grandes ou qui présentent une mauvaise complémentarité avec le site de liaison[73].

B .Docking avec ligand flexible :

Considérer le système biologique comme un système dynamique révèle que la liaison du ligand au site d'interaction se produit principalement par un processus de sélection de conformation, qui nécessite une adaptation mutuelle entre le ligand et le récepteur et conduit à des changements conformationnels importants pour les deux parties[74]. Cette vision simpliste du problème a été améliorée par l'introduction de la flexibilité du ligand, facilitée par l'avènement de capacités informatiques avancées, permettant l'exploration des différentes conformations qu'il peut adopter lors de son ancrage au site actif[75]. Cette approche est actuellement largement préférée dans la plupart des programmes d'amarrage car elle permet une évaluation rapide d'un large ensemble de molécules, s'élevant à des millions[76]. (Figure II.9)

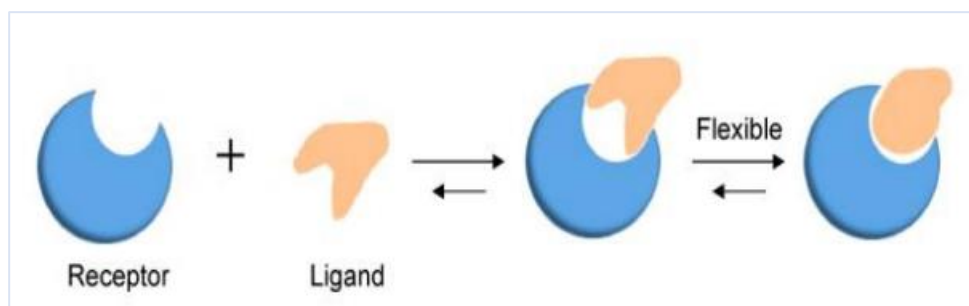


Figure II.9 : Model de docking moléculaire (docking flexible) .

C .Docking semi-flexible :

Le docking moléculaire semi-flexible est une approche qui a permis d'obtenir des résultats plus précis en tenant compte de la flexibilité du ligand tout en considérant le récepteur comme un corps rigide [77].

II .3. Les propriétés ADME :

Le parcours d'un médicament à travers l'organisme peut être subdivisé en quatre étapes distinctes. Tout d'abord, la phase de résorption ou d'absorption, où le médicament quitte son site d'administration pour entrer dans la circulation systémique. Ensuite, la phase de distribution, caractérisée par la dispersion du médicament dans les divers tissus de l'organisme. Puis vient la phase de métabolisme, au cours de laquelle le médicament est modifié par le corps. Enfin, la phase d'élimination, qui concerne les processus par lesquels le médicament est éliminé du corps.

Ces quatre phases sont souvent désignées par l'acronyme ADME. Certains chercheurs élargissent ce concept en y ajoutant une phase de libération du principe actif à partir de la forme galénique administrée, ainsi qu'une phase de réponse au principe actif. Dans cette perspective, le parcours est décrit comme suit : Libération - Absorption - Distribution - Métabolisme - Élimination - Réponse. Cette approche est alors représentée par l'acronyme L'ADME pour englober l'ensemble de ces étapes [78].

Les caractéristiques d'absorption, de distribution, de métabolisme, d'excrétion et de toxicité (ADMET) des médicaments candidats ou des produits chimiques environnementaux jouent un rôle crucial dans la recherche de nouveaux médicaments et dans l'évaluation des risques environnementaux. Pour qu'une molécule puisse être efficace en tant que médicament, elle doit pouvoir atteindre sa cible dans le corps à une concentration adéquate et y demeurer sous une forme bioactive suffisamment longtemps pour permettre le déroulement des processus biologiques attendus. C'est pourquoi le développement de médicaments implique désormais une évaluation précoce des processus d'absorption, de distribution, de métabolisme et d'excrétion (ADME), dès les premières phases de la découverte, alors que le nombre de composés à examiner est élevé mais que l'accès aux échantillons physiques est souvent limité [79].

Les Références :

- [1]. Grover, M., Singh, B., Bakshi, M., & Singh, S. (2000). Quantitative structure–property relationships in pharmaceutical research–Part 2. *Pharmaceutical science & technology today*, 3(2), 50-57.
- [2]. Cronin, M. T. (2010). Quantitative structure–activity relationships (QSARs)—applications and methodology. *Recent advances in QSAR studies: methods and applications*, 3-11.
- [3]. Kraim, k. (2009). Etudes de qsar sur des activites biologiques utilisant des produits d'origines naturels (Doctoral dissertation).
- [4]. Esposito, E. X., Hopfinger, A. J., & Madura, J. D. (2004). Methods for applying the quantitative structure-activity relationship paradigm. *Chemoinformatics: Concepts, methods, and tools for drug discovery*, 131-213.
- [5]. Bradbury, S. P. (1995). Quantitative structure-activity relationships and ecological risk assessment: an overview of predictive aquatic toxicology research. *Toxicology letters*, 79(1-3), 229-237.
- [6]. Chen, J. Z., Han, X. W., Liu, Q., Makriyannis, A., Wang, J., & Xie, X. Q. (2006). 3D-QSAR studies of arylpyrazole antagonists of cannabinoid receptor subtypes CB1 and CB2. A combined NMR and CoMFA approach. *Journal of medicinal chemistry*, 49(2), 625-636.
- [7]. Gasteiger, J. (2003). Handbook of Chemoinformatics: From data to knowledge in 4 volumes. (*No Title*).
- [8]. Mannhold, R., Krogsgaard-Larsen, P., & Timmerman, H. (2008). *QSAR: Hansch analysis and related approaches*. John Wiley & Sons.
- [9]. Khadidja, M., & Ikram, L. (2019). Conception in silico de nouveaux composés anticancéreux [Mémoire]. *Biskra: Université Med Khider*.
- [10]. Daoudi, s., chernai, m., ikerrouyene, m., & haddadou, a. (2022). conception d'un antimigraineux par criblage virtuel.
- [11]. Ghosh, S., Nie, A., An, J., & Huang, Z. (2006). Structure-based virtual screening of chemical libraries for drug discovery. *Current opinion in chemical biology*, 10(3), 194-202.
- [12]. Lyne, P. D. (2002). Structure-based virtual screening: an overview. *Drug discovery today*, 7(20), 1047-1055.
- [13]. Bajorath, J. (2002). Integration of virtual and high-throughput screening. *Nature Reviews Drug Discovery*, 1(11), 882-894.
- [14]. Lengauer, T., Lemmen, C., Rarey, M., & Zimmermann, M. (2004). Novel technologies for virtual screening. *Drug discovery today*, 9(1), 27-34.

- [15]. Oprea, T. I., & Matter, H. (2004). Integrating virtual screening in lead discovery. *Current opinion in chemical biology*, 8(4), 349-358.
- [16]. Ekins, S., Mestres, J., & Testa, B. (2007). In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *British journal of pharmacology*, 152(1), 9-20.
- [17]. Muegge, I., & Oloff, S. (2006). Advances in virtual screening. *Drug discovery today: technologies*, 3(4), 405-411.
- [18]. Lavecchia, A., & Di Giovanni, C. (2013). Virtual screening strategies in drug discovery: a critical review. *Current medicinal chemistry*, 20(23), 2839-2860.
- [19]. Johnson, M. A., & Maggiora, G. M. (1990). *Concepts and applications of molecular similarity*.
- [20]. Lagarde, N. (2014). Méthodes de criblage virtuel in silico: importance de l'évaluation et application à la recherche de nouveaux inhibiteurs de l'interleukine 6 (Doctoral dissertation, Paris, CNAM).
- [21]. Koeppen, H., Kriegl, J., Lessel, U., Tautermann, C. S., & Wellenzohn, B. (2011). Ligand-Based Virtual Screening. *Virtual screening: principles, challenges, and practical guidelines*, 61-85.
- [22]. BENCHEIKH, B. (2023). *C riblage in silico des inhibiteurs du SARS CoV 2 par Docking moléculaire et QSAR* (Doctoral dissertation).
- [23]. Patrick, G.L.(2002). *Aspects quantitatifs des relations structure-activité (RSA)*, in *Chimie Pharmaceutique*. p. 648.
- [24]. Todeschini, R., & Consonni, V. (2008). *Handbook of molecular descriptors*. John Wiley & Sons.
- [25]. Andrade, C. H., Pasqualoto, K. F., Ferreira, E. I., & Hopfinger, A. J. (2010). 4D-QSAR: perspectives in drug design. *Molecules*, 15(5), 3281-3294.
- [26]. Vedani, A., & Dobler, M. (2002). 5D-QSAR: the key for simulating induced fit?. *Journal of medicinal chemistry*, 45(11), 2139-2149.
- [27]. Vedani, A., Descloux, A. V., Spreafico, M., & Ernst, B. (2007). Predicting the toxic potential of drugs and chemicals in silico: A model for the peroxisome proliferator-activated receptor γ (PPAR γ). *Toxicology letters*, 173(1), 17-23.
- [28]. Benouis, S .(2022).*Etude théorique de l'activité inhibitrice de quelques enzymes par le criblage virtuel et le docking moléculaire*.Université Badji mokhtar - annaba. Chimie organique appliquée. P43 .
- [29]. KRAIM, k.(2009). *Etude QSAR sur activites biologiques en utilisant des produits d'origines naturels*, universite de Guelma.

- [30]. Goulon-Sigwalt-Abram, A. (2008). *Une nouvelle méthode d'apprentissage de données structurées: applications à l'aide à la découverte de médicaments* (Doctoral dissertation, Université Pierre et Marie Curie-Paris VI).
- [31]. Wiener, H. (1947). *Structural determination of paraffin boiling points. Journal of the American chemical society*, 69(1), 17-20.
- [32]. Randic, M. (1975). *Characterization of molecular branching. Journal of the American Chemical Society*, 97(23), 6609-6615.
- [33]. Kier, L.B., et Hall, L.H.(1976). *Molecular connectivity in chemistry and drug research*. New York : Academic Press.
- [34]. Balaban, A. T. (1982). Highly discriminating distance-based topological index. *Chemical physics letters*, 89(5), 399-404.
- [35]. El Ghalia, H. (2018). *Corrélation-(structure-activité anticancéreuse) par les méthodes QSAR des molécules hétérocycliques précurseurs de médicaments*.
- [36]. Gupta, M., Lee, H. J., Barden, C. J., & Weaver, D. F. (2019). The blood–brainbarrier (BBB) score. *Journal of medicinal chemistry*, 62(21), 9824-9836.
- [37]. Muratov, E. N., Bajorath, J., Sheridan, R. P., Tetko, I. V., Filimonov, D., Poroikov, V., Oprea, T. I., Baskin, I. I., Varnek, A., Roitberg, A., Isayev, O., Curtalolo, S., Fourches, D., Cohen, Y., Aspuru-Guzik, A., Winkler, D. A., Agrafiotis, D., Cherkasov, A., & Tropsha, A. (2020). QSAR without borders. *Chemical Society Reviews*, 49(11), 3525-3564
- [38]. Roy, K., Kar, S., & Das, R. N. (2015). *Statistical methods in QSAR/QSPR*. Springerbriefs in molecular.
- [39]. Besalú, E., de Julián-Ortiz, J. V., & Pogliani, L. (2007). Trends and plot methods in MLR studies. *Journal of chemical information and modeling*, 47(3), 751-760.
- [40]. Consonni, V., Ballabio, D., & Todeschini, R. (2010). Evaluation of model predictive ability by external validation techniques. *Journal of chemometrics*, 24(3-4), 194-201.
- [41]. Wold, S. (1994). PLS for multivariate linear modeling. *QSAR: chemometric methods in molecular design Methods and principles in medicinal chemistry*.
- [42]. Clark, M., & Cramer III, R. D. (1993). The probability of chance correlation using partial least squares (PLS). *Quantitative Structure-Activity Relationships*, 12(2), 137-145.
- [43]. Sjöström, M., Eriksson, L.(1955). *han van deWaterbeemd*, p 63- 90.
- [44]. Niculescu, S. P., Lewis, M. A., & Tigner, J. (2008). Probabilistic neural networks modeling of the 48-h LC50 acute toxicity endpoint to *Daphnia magna*. *SAR and QSAR in Environmental Research*, 19(7-8), 735-750.
- [45]. Mitra, I., Saha, A., Roy, K.(2010) *Mol Simult*, p 1067–1079.

- [46]. Hubbard, R. E. (2005). 3D structure and the drug-discovery process. *Molecular BioSystems*, 1(5-6), 391-406.
- [47]. Djelloul.k.(2021). *Etude QSRR de dérivés benzéniques.universite Badji mokhtar -annaba*. Chimie analytique et environnement p 50.
- [48]. Chtita, S. (2017). *Modélisation de molécules organiques hétérocycliques biologiquement actives par des méthodes QSAR/QSPR. Recherche de nouveaux médicaments* (Doctoral dissertation, Université Moulay Ismaïl, Meknès).
- [49]. Bultinck, P., Winter, H. D., Langenaeker, W., J. P. Tollenaere, J. P.(2004) "Computational Chemistry for Drug Discovery" Ed. Marcel Dekker. Etats Unies.
- [50]. Wold, S., Eriksson, L., & Clementi, S. (1995). Statistical validation of QSAR results. *Chemometric methods in molecular design*, 309-338.
- [51]. Ghamali, M., Chtita, S., Bouachrine, M., & Lakhlifi, T. (2016). Méthodologie générale d'une étude RQSA/RQSP. *Revue Interdisciplinaire*, 1(1).
- [52]. Golbraikh, A., & Tropsha, A. (2002). Beware of q²!. *Journal of molecular graphics and modelling*, 20(4), 269-276.
- [53]. El Ghalia, H. (2018). Corrélation-(structure-activité anticancéreuse) par les méthodes QSAR des molécules hétérocycliques précurseurs de médicaments.
- [54]. Tetko, I. V., Sushko, I., Pandey, A. K., Zhu, H., Tropsha, A., Papa, E., ... & Varnek, A. (2008). Critical assessment of QSAR models of environmental toxicity against *Tetrahymena pyriformis*: focusing on applicability domain and overfitting by variable selection. *Journal of chemical information and modeling*, 48(9), 1733-1746.
- [55]. Principles for the Validation, for Regulatory Purposes, of Structure-Activity Relationship Models, Organisation de Coopération et de Développement Economique, Paris, 2009.
- [56]. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ... & Bourne, P. E. (2000). The protein data bank. *Nucleic acids research*, 28(1), 235-242.
- [57]. Lavecchia, A., & Di Giovanni, C. (2013). Virtual screening strategies in drug discovery: a critical review. *Current medicinal chemistry*, 20(23), 2839-2860.
- [58]. Tuccinardi, T. (2009). Docking-based virtual screening: recent developments. *Combinatorial chemistry & high throughput screening*, 12(3), 303-314.
- [59]. Empereur-Mot, C. (2017). *Développement d'outils statistiques d'évaluation de méthodes de criblage virtuel: courbes de prédictivité & Screening Explorer* (Doctoral dissertation, Conservatoire national des arts et métiers-CNAM).

- [60]. R Laurie, A. T., & Jackson, R. M. (2006). Methods for the prediction of protein-ligand binding sites for structure-based drug design and virtual ligand screening. *Current Protein and Peptide Science*, 7(5), 395-406.
- [61]. Hessler, G. et Baringhaus, K.H. (2010). Le potentiel de saut d'échafaudage des pharmacophores. *La découverte de médicaments aujourd'hui : technologies*, 7(4), e263-e269.
- [62]. Zerroug, E. (2020). *Analyse in silico de plusieurs séries de pharmacomolécules hétérocycliques appliquée à la conception de médicaments* (Doctoral dissertation, Université Mohamed Khider–BISKRA).
- [63]. Dror, O., Shulman-Peleg, A., Nussinov, R. et Wolfson, H.J. (2004). Prédire les interactions moléculaires in silico : I. Un guide pour l'identification des pharmacophores et ses applications à la conception de médicaments. *Chimie médicinale actuelle*, 11(1), 71-90.
- [64]. Huang, L. L., Han, J., Ran, J. X., Chen, X. P., Wang, Z. H., & Wu, F. H. (2018). 3D-QSAR, molecular docking and molecular dynamics simulations of oxazepane amidoacetonitrile derivatives as novel DPPI inhibitors. *Journal of Molecular Structure*, 1168, 223-233.
- [65]. Faculté des Sciences de la Nature et de la Vie Département de Biochimie et Biologie cellulaire et moléculaire, Mémoire du diplôme de Master sciences Biologiques 2023, Etude comparative de l'efficacité prédictive des modèles 2D-QSAR et 3D-QSAR pour la conception de molécules anti-malaria P 24.
- [66]. BELKAIBECH, I. (2021). Mémoire de master, Etude 3D-QSAR et docking moléculaire de dérivés de pyrazoline l'activité anti-amibienne contre la souche HK1 :IMSS d'E.histolytica, UNIVERSITÉ MOHAMED BOUDIAF - M'SILA .
- [67]. Bristol, J. A. (1993). *Annual reports in medicinal chemistry* (Vol. 28). Academic Press.
- [78]. Saoussene, M. L. (2016). *Etude docking et synthèse de dérivés de xanthone: voie d'accès à de nouveaux inhibiteurs de l'α-glucosidase* (Doctoral dissertation, Université de Annaba).
- [69]. Schneider, G., & Fechner, U. (2005). Computer-based de novo design of drug-like molecules. *Nature Reviews Drug Discovery*, 4(00), 649-663.)
- [70]. van Gunsteren, Wilfred F. ; Berendsen, Herman J. C *Angewandte Chemie*, (September 1990), vol. 102, issue 9, pp. 1020-1055.
- [71]. Bencheikh, B. (2023). *C riblage in silico des inhibiteurs du SARS CoV 2 par Docking moléculaire et QSAR* (Doctoral dissertation).
- [72]. Fischer, E. (1894). Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte der deutschen chemischen Gesellschaft*, 27(3), 2985-2993.
- [73]. Miteva, M. A., Lee, W. H., Montes, M. O., & Villoutreix, B. O. (2005). Fast structure-based virtual ligand screening combining FRED, DOCK, and Surflex. *Journal of medicinal chemistry*, 48(19), 6012-6022.

- [74]. Vogt, A. D., & Di Cera, E. (2013). Conformational selection is a dominant mechanism of ligand binding. *Biochemistry*, 52(34), 5723-5729.
- [75]. Yuriev, E., Agostino, M., & Ramsland, P. A. (2011). Challenges and advances in computational docking: 2009 in review. *Journal of Molecular Recognition*, 24(2), 149-164.
- [76]. Mohan, V., Gibbs, A. C., Cummings, M. D., Jaeger, E. P., & DesJarlais, R. L. (2005). Docking: successes and challenges. *Current pharmaceutical design*, 11(3), 323-333.
- [77]. Mokrani, E. H., & Bensegueni, A. (2012). Contribution à l'amélioration de l'activité biologique des inhibiteurs de la dipeptidyl-peptidase 4 dans le diabète de type 2 par simulation informatique (Doctoral dissertation, Université Frères Mentouri Constantine 1).
- [78]. Ritschel, W. A. (1976). *Handbook of basic pharmacokinetics*.
- [79]. Ben sahal Sirine, S. B. La prédiction in silico des propriétés ADME des molécules des huiles essentielles de *Myrtus communis* L.

CHAPITRE 03 :

RESULTATS ET DISCUSSION

III .1. Introduction:

La découverte de nouveaux médicaments est un processus complexe et chronophage, nécessitant en moyenne plus de 12 années et un investissement financier considérable avant qu'un médicament puisse être disponible pour les patients. Ce processus est également associé à un risque élevé [1,2]. Grâce à un système dynamique de découverte de médicaments, les molécules candidates acquièrent un contenu informationnel élevé, traité ensuite par des méthodes sophistiquées d'analyse de données, faisant partie intégrante du domaine de la chimie informatique [3].

Dans ce contexte, plusieurs techniques *in silico* sont utilisées dans le processus de découverte et de conception de médicaments. La principale parmi celles-ci est l'utilisation des méthodes QSAR, qui étudient les relations qualitatives et quantitatives entre la structure et l'activité (SAR/QSAR).

Le principe fondamental des méthodes QSAR (Quantitative Structure Activity Relationship) réside dans l'établissement d'une relation mathématique à partir d'analyses de données. Ces méthodes permettent de corréler des propriétés moléculaires microscopiques, désignées comme descripteurs, à un effet expérimental tel que l'activité biologique ou la toxicité, pour une série de composés chimiques similaires. La construction de ces modèles QSAR repose sur la définition de descripteurs moléculaires, qu'ils soient empiriques ou théoriques, intégrant des informations sur la structure et les caractéristiques physico-chimiques des molécules [4].

Le choix d'une base de données expérimentale de référence est une étape cruciale dans toute étude QSAR, car la fiabilité du modèle dépend en grande partie de cette base. Idéalement, cette base devrait être composée de données expérimentales obtenues de manière fiable et selon un protocole expérimental cohérent [4].

Enfin, l'association entre les descripteurs et la base de données se réalise à l'aide de divers outils d'analyse, tels que les régressions multi-linéaires (MLR), les régressions aux moindres carrés partiels (PLS), les arbres de décision, les réseaux de neurones et les algorithmes génétiques [4].

Les informations déduites des études QSAR peuvent être utilisées pour approfondir notre compréhension des structures moléculaires ainsi que des mécanismes d'action potentiels au niveau moléculaire. Ces données sont également précieuses pour prédire les propriétés physico-chimiques et les activités biologiques de nouveaux composés, et pour guider la conception de structures chimiques innovantes [5].

À titre d'application, nous avons proposé l'élaboration d'un modèle statistique QSAR pour coder l'information chimique sous forme d'équation mathématique pour une série des inhibiteurs de la protéine Bcl-2 , en vue d'évaluer leur activité anticancéreuse.

Ce projet repose sur l'application de méthodes avancées de modélisation moléculaire ainsi que sur l'évaluation des propriétés ADME. Dans cette optique, nous avons conçu un schéma synthétique qui illustre de manière concise les diverses phases de nos recherches (**Figure III.1**) :

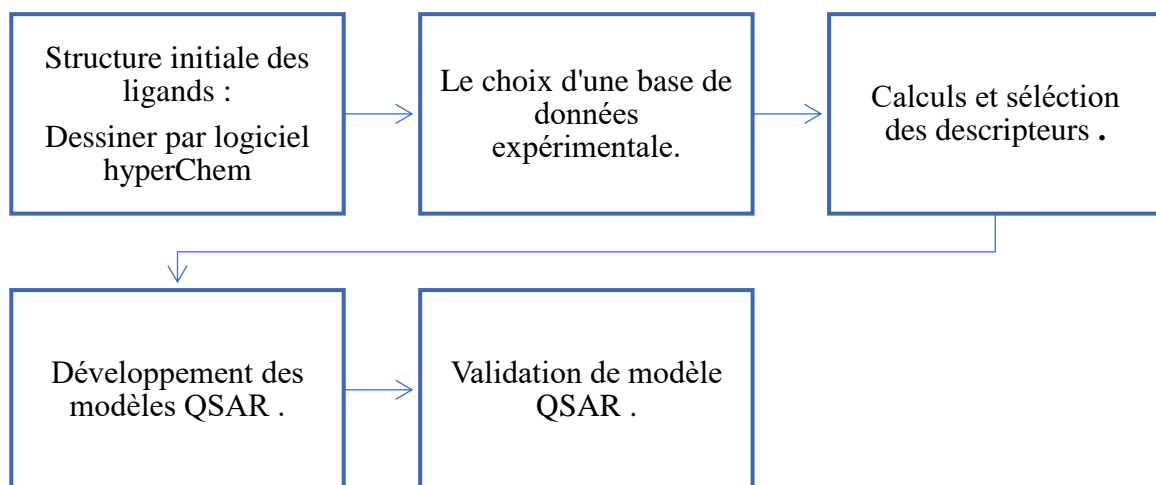


Figure III.1 : Méthodologie générale utilisée dans l'étude QSAR.

III .2. Matériels et méthodes :

III .2.1. Ensemble de données :

Un candidat médicament de qualité se caractérise par de bonnes propriétés pharmacocinétiques, notamment une biodisponibilité adéquate et une faible toxicité ainsi que des effets secondaires minimes [6].

Comme point de départ de cette étude, nous avons utilisé HyperChem pour construire et optimiser les molécules. Pour parvenir à cette amélioration, nous avons utilisé la méthode de mécanique moléculaire MM⁺, puis la méthode semi empirique AM1.

Nous avons utilisé une approche in silico pour prédire les propriétés physicochimiques de 108 molécules inhibiteurs de la protéine Bcl-2 [7 , 8 , 9] à l'aide du site web SwissADME (**Tableau**

III.1) . Cette analyse vise à évaluer la biodisponibilité potentielle des médicaments, laquelle dépend en grande partie de la solubilité et de la capacité de la molécule à traverser la membrane intestinale. Des caractéristiques telles que la solubilité dans l'eau, le coefficient de partage octanol/eau (LogP), le nombre de liaisons rotatives, la surface non polaire, etc. Les composés qui ne respectent pas la très connue règle de Lipinski, Verber et Egan ont généralement une mauvaise pharmacocinétique Propriétés [10].

En utilisant cette approche intégrée, nous pouvons évaluer et sélectionner les « leads » candidats avec des propriétés ADME optimales, en nous concentrant sur ceux qui sont les plus susceptibles d'être bien absorbés par le tractus gastro-intestinal et donc d'avoir une efficacité clinique accrue.

Tableau III.1 : Les propriétés physico-chimiques de 108 molécules à l'aide de l'outil web SwissADME .

Les Molécules	Propriétés physico-chimique						Drug Likeness		
	TPSA (Å ²)	n-ROT	MW (g/mol)	MLog P	n-HA	n-HD	Lipinski	Veber	Egan
	(0~140)	(0~11)	(100~500)	WLogP (0~5)	(0~12)	(0~7)			
Molécule 01	194.54	19	879.49	3.10	10	2	Refusé	Refusé	Refusé
				8.99					
Molécule 02	194.54	19	934.25	3.58	10	2	Refusé	Refusé	Refusé
				8.79					
Molécule 03	194.54	19	889.80	3.50	10	2	Refusé	Refusé	Refusé
				8.69					
Molécule 04	194.54	20	903.82	3.66	10	2	Refusé	Refusé	Refusé
				8.88					
Molécule 05	240.36	20	914.38	2.53	12	2	Refusé	Refusé	Refusé
				8.25					
Molécule 06	158.01	11	596.76	3.12	5	2	Acceptée	Refusé	Refusé
				6.83					
Molécule 07	190.79	16	786.96	2.70	8	2	Refusé	Refusé	Refusé
				5.96					
Molécule 08	170.48	19	711.93	2.96	7	2	Refusé	Refusé	Refusé
				7.31					
Molécule 09	161.25	16	767.88	3.80	9	2	Acceptée	Refusé	Refusé
				9.49					
Molécule 10	161.25	15	734.33	3.81	6	2	Acceptée	Refusé	Refusé
				7.98					

CHAPITRE 03 : RESULTATS ET DISCUSSION

Molécule 11	161.25	15	717.87	3.72	7	2	Acceptée	Refusé	Refusé
				7.88					
Molécule 12	185.04	15	724.89	2.71	7	2	Refusé	Refusé	Refusé
				7.19					
Molécule 13	161.25	16	767.88	3.80	9	2	Acceptée	Refusé	Refusé
				9.49					
Molécule 14	161.25	15	734.33	3.81	6	2	Acceptée	Refusé	Refusé
				7.98					
Molécule 15	161.25	15	717.87	3.72	7	2	Acceptée	Refusé	Refusé
				7.88					
Molécule 16	174.14	15	700.87	2.42	7	2	Refusé	Refusé	Refusé
				6.72					
Molécule 17	174.14	15	700.87	2.42	7	2	Refusé	Refusé	Refusé
				6.72					
Molécule 18	161.25	14	639.83	2.73	6	2	Acceptée	Refusé	Refusé
				6.37					
Molécule 19	181.56	16	716.87	2.07	7	2	Refusé	Refusé	Refusé
				5.33					
Molécule 20	193.59	17	731.88	2.41	7	3	Refusé	Refusé	Refusé
				5.53					
Molécule 21	207.01	16	766.95	1.93	9	2	Refusé	Refusé	Refusé
				6.27					
Molécule 22	164.49	16	702.89	2.49	7	2	Refusé	Refusé	Refusé
				5.54					
Molécule 23	164.49	16	716.91	2.67	7	2	Refusé	Refusé	Refusé
				5.85					
Molécule 24	164.49	17	813.43	3.76	7	2	Refusé	Refusé	Refusé
				7.86					
Molécule 25	177.38	17	779.97	2.41	8	2	Refusé	Refusé	Refusé
				6.60					
Molécule 26	176.52	18	800.04	3.04	7	3	Refusé	Refusé	Refusé
				7.09					
Molécule 27	207.01	17	780.98	2.11	9	2	Refusé	Refusé	Refusé
				6.02					
Molécule 28	170.48	17	711.93	2.96	7	2	Refusé	Refusé	Refusé
				7.17					
Molécule 29	170.48	17	697.91	2.78	7	2	Refusé	Refusé	Refusé
				6.78					
Molécule 30	170.48	17	766.37	3.55	7	2	Refusé	Refusé	Refusé
				7.63					
Molécule 31	170.48	17	766.37	3.55	7	2	Refusé	Refusé	Refusé
				7.63					
Molécule 32	170.48	17	737.97	3.31	7	2	Refusé	Refusé	Refusé
				7.70					
Molécule 33	170.48	17	731.92	3.11	7	2	Refusé	Refusé	Refusé
				6.98					
Molécule 34	170.48	17	810.82	3.64	7	2	Refusé	Refusé	Refusé
				7.74					
Molécule 35	170.48	17	745.95	3.28	7	2	Refusé	Refusé	Refusé
				7.28					

CHAPITRE 03 : RESULTATS ET DISCUSSION

Molécule 36	170.48	17	749.91	3.47	8	2	Refusé	Refusé	Refusé
				7.53					
Molécule 37	170.48	16	729.91	3.31	7	2	Refusé	Refusé	Refusé
				7.33					
Molécule 38	179.71	18	761.95	2.80	8	2	Refusé	Refusé	Refusé
				6.98					
Molécule 39	173.72	17	732.91	2.45	8	2	Refusé	Refusé	Refusé
				5.55					
Molécule 40	115.81	4	418.44	2.10	6	4	Acceptée	Acceptée	Acceptée
				4.06					
Molécule 41	115.81	3	390.39	1.67	6	4	Acceptée	Acceptée	Acceptée
				3.24					
Molécule 42	118.89	8	344.36	1.22	5	5	Acceptée	Acceptée	Acceptée
				2.39					
Molécule 43	118.89	7	378.38	1.95	5	5	Acceptée	Acceptée	Acceptée
				2.64					
Molécule 44	118.89	9	406.43	2.38	5	5	Acceptée	Acceptée	Acceptée
				3.23					
Molécule 45	94.83	0	256.21	0.11	5	3	Acceptée	Acceptée	Acceptée
				1.58					
Molécule 46	95.58	3	360.36	2.39	5	3	Acceptée	Acceptée	Acceptée
				3.23					
Molécule 47	106.44	0	194.14	-0.56	5	4	Acceptée	Acceptée	Acceptée
				0.04					
Molécule 48	187.55	14	641.76	2.25	7	2	Refusé	Refusé	Refusé
				5.10					
Molécule 49	170.48	14	681.87	2.60	7	2	Refusé	Refusé	Refusé
				5.76					
Molécule 50	184.03	15	639.83	3.00	6	3	Acceptée	Refusé	Refusé
				6.55					
Molécule 51	170.48	14	681.87	2.60	7	2	Refusé	Refusé	Refusé
				5.76					
Molécule 52	158.01	11	568.71	3.00	5	2	Acceptée	Refusé	Refusé
				6.05					
Molécule 53	178.32	14	653.81	2.30	6	2	Refusé	Refusé	Refusé
				5.90					
Molécule 54	181.56	13	612.72	1.86	6	2	Refusé	Refusé	Refusé
				3.85					
Molécule 55	161.25	14	639.83	2.73	6	2	Acceptée	Refusé	Refusé
				6.37					
Molécule 56	161.25	14	639.83	2.73	6	2	Acceptée	Refusé	Refusé
				6.37					
Molécule 57	161.25	15	653.86	2.91	6	2	Acceptée	Refusé	Refusé
				6.76					
Molécule 58	161.25	16	667.88	3.10	6	2	Acceptée	Refusé	Refusé
				7.15					
Molécule 59	184.31	16	662.78	2.95	7	2	Refusé	Refusé	Refusé
				6.50					
Molécule 60	193.54	16	704.81	2.16	8	2	Refusé	Refusé	Refusé
				5.89					

CHAPITRE 03 : RESULTATS ET DISCUSSION

Molécule 61	184.31	14	634.72	2.84	7	2	Refusé	Refusé	Refusé
				6.19					
Molécule 62	167.24	16	648.79	3.37	7	2	Refusé	Refusé	Refusé
				6.98					
Molécule 63	179.71	19	761.95	2.53	9	2	Refusé	Refusé	Refusé
				6.69					
Molécule 64	196.78	19	775.93	2.12	9	2	Refusé	Refusé	Refusé
				6.21					
Molécule 65	179.71	21	790.00	2.87	9	2	Refusé	Refusé	Refusé
				7.47					
Molécule 66	202.49	20	761.95	2.53	9	3	Refusé	Refusé	Refusé
				6.86					
Molécule 67	179.71	20	775.98	2.70	9	2	Refusé	Refusé	Refusé
				7.08					
Molécule 68	179.71	19	761.95	2.53	9	2	Refusé	Refusé	Refusé
				6.69					
Molécule 69	188.94	19	803.99	2.12	10	2	Refusé	Refusé	Refusé
				6.08					
Molécule 70	176.47	16	690.83	2.58	8	2	Refusé	Refusé	Refusé
				6.37					
Molécule 71	184.23	15	621.72	3.00	7	3	Refusé	Refusé	Refusé
				6.41					
Molécule 72	167.24	15	634.77	3.18	7	2	Refusé	Refusé	Refusé
				6.59					
Molécule 73	167.24	14	620.74	3.00	7	2	Refusé	Refusé	Refusé
				6.39					
Molécule 74	213.72	9	632.66	-4.60	10	8	Refusé	Refusé	Refusé
				4.43					
Molécule 75	213.72	13	688.76	0.54	10	8	Refusé	Refusé	Refusé
				5.99					
Molécule 76	213.72	15	716.82	0.88	10	8	Refusé	Refusé	Refusé
				6.77					
Molécule 77	191.72	19	772.92	1.55	10	6	Refusé	Refusé	Refusé
				7.73					
Molécule 78	191.72	17	744.87	1.22	10	6	Refusé	Refusé	Refusé
				6.94					
Molécule 79	191.72	13	688.76	0.54	10	6	Refusé	Refusé	Refusé
				5.38					
Molécule 80	191.72	10	646.68	0.41	10	6	Refusé	Refusé	Refusé
				5.20					
Molécule 81	135.55	8	510.60	2.09	7	4	Acceptée	Acceptée	Refusé
				4.19					
Molécule 82	81.00	3	327.37	2.08	6	4	Acceptée	Acceptée	Acceptée
				2.59					
Molécule 83	89.79	4	287.31	2.02	4	4	Acceptée	Acceptée	Acceptée
				2.99					
Molécule 84	181.31	13	679.80	1.86	10	4	Refusé	Refusé	Refusé
				5.41					
Molécule 85	64.35	7	412.48	4.13	4	1	Acceptée	Acceptée	Acceptée
				5.79					

CHAPITRE 03 : RESULTATS ET DISCUSSION

Molécule 86	104.28	7	612.71	5.03	6	3	Refusé	Acceptée	Refusé
				9.05					
Molécule 87	111.91	13	605.68	4.32	7	2	Refusé	Refusé	Refusé
				7.91					
Molécule 88	111.91	12	591.65	4.14	7	2	Refusé	Refusé	Refusé
				7.67					
Molécule 89	64.35	6	364.44	3.51	4	1	Acceptée	Acceptée	Acceptée
				5.15					
Molécule 90	73.58	7	414.45	3.65	5	1	Acceptée	Acceptée	Acceptée
				5.82					
Molécule 91	64.35	5	322.36	2.84	4	1	Acceptée	Acceptée	Acceptée
				4.02					
Molécule 92	95.05	4	506.59	4.50	5	3	Refusé	Acceptée	Refusé
				6.75					
Molécule 93	102.68	11	575.65	4.65	6	2	Refusé	Refusé	Refusé
				7.79					
Molécule 94	111.91	12	577.63	3.69	7	2	Refusé	Refusé	Refusé
				7.28					
Molécule 95	111.91	13	591.65	3.87	7	2	Acceptée	Refusé	Refusé
				7.52					
Molécule 96	102.68	10	519.55	3.91	6	2	Acceptée	Acceptée	Refusé
				6.49					
Molécule 97	73.58	7	400.43	3.44	5	1	Acceptée	Acceptée	Acceptée
				5.43					
Molécule 98	64.35	5	308.33	2.60	4	1	Acceptée	Acceptée	Acceptée
				3.64					
Molécule 99	100.91	13	605.68	4.05	7	1	Acceptée	Refusé	Refusé
				7.75					
Molécule 100	89.79	4	287.31	2.02	4	4	Acceptée	Acceptée	Acceptée
				2.99					
Molécule 101	132.31	9	573.70	4.75	6	4	Refusé	Acceptée	Refusé
				7.79					
Molécule 102	99.31	12	615.78	5.30	6	1	Refusé	Refusé	Refusé
				8.70					
Molécule 103	132.31	7	475.51	3.40	6	4	Acceptée	Acceptée	Refusé
				5.37					
Molécule 104	89.79	5	335.35	2.74	4	4	Acceptée	Acceptée	Acceptée
				3.46					
Molécule 105	185.31	16	862.57	3.53	9	2	Refusé	Refusé	Refusé
				8.02					
Molécule 106	185.31	16	828.13	3.09	9	2	Refusé	Refusé	Refusé
				7.37					
Molécule 107	185.31	16	842.15	3.26	9	2	Refusé	Refusé	Refusé
				7.68					
Molécule 108	185.31	16	763.26	2.74	9	2	Refusé	Refusé	Refusé
				6.91					

Pour qu'une molécule puissante, dotée d'une affinité conférant une activité biologique, puisse atteindre sa cible dans l'organisme, elle doit surmonter divers obstacles afin d'y parvenir avec une concentration suffisante pour déclencher les effets biologiques escomptés. Les études ADMET jouent un rôle crucial en tant qu'élément essentiel de la découverte et du développement des médicaments.

Une fois que nous avons les prédictions ADME pour nos composés, nous utilisons l'outil œuf dur « BOILED-Egg », ce qui permet une évaluation intuitive de l'absorption gastro-intestinale passive (HIA) et de la pénétration cérébrale (BBB) en fonction de la position des molécules dans le cadre WLOGP versus TPSA. **Figure III.2** visualise les résultats et les « leads » candidats sélectionnés en fonction de leur absorption gastro-intestinale (HIA).

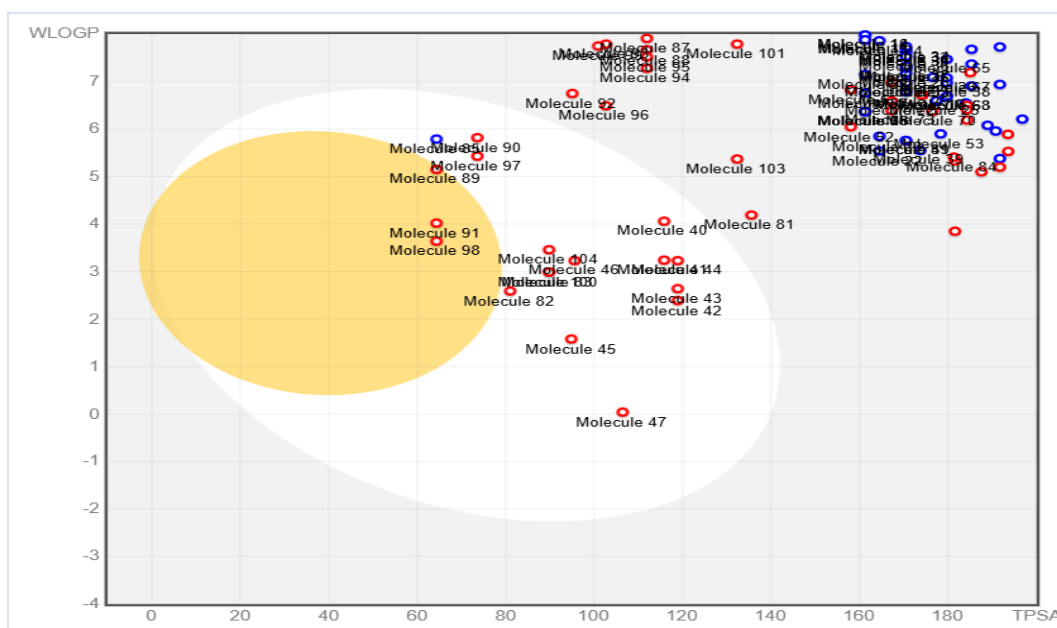


Figure III.2 : Graphique « BOILED-Egg ».

Ce graphique, appelé l'œuf dur, prédit simultanément deux paramètres ADME clés : l'absorption gastro-intestinale passive (HIA) et l'accès au cerveau (BBB). Malgré sa simplicité conceptuelle reposant sur deux descripteurs physico-chimiques (WLOGP et TPSA), ce modèle a été construit avec rigueur statistique et robustesse. L'œuf dur présente deux zones distinctes, jaune pour la perméation BBB et blanche pour l'absorption HIA, avec une région grise pour les molécules peu absorbées et à pénétration cérébrale limitée. Intégré à SwissADME, il permet une évaluation complète de l'absorption passive, de l'accès cérébral et de l'efflux actif, avec une coloration des

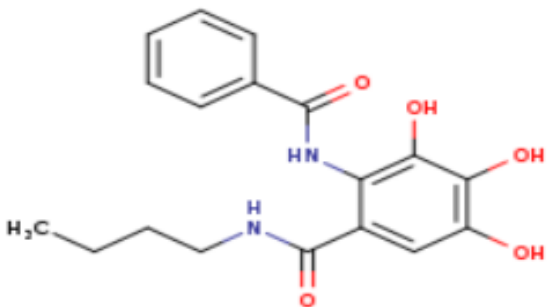
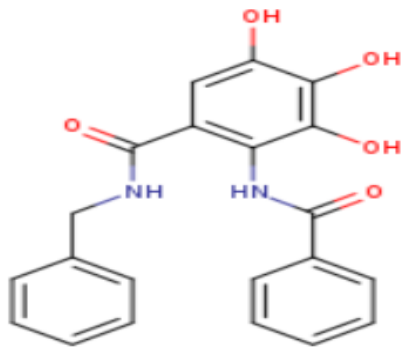
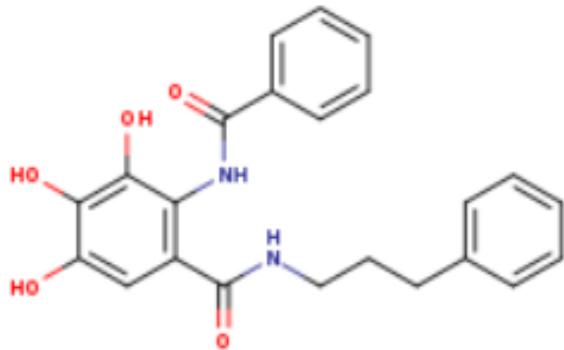
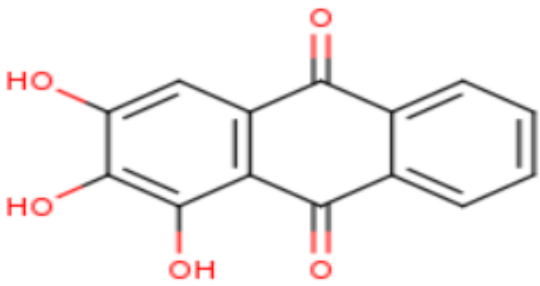
points en bleu pour les substrats de P-gp (PGP+) et en rouge pour les non-substrats de P-gp (PGP-) [11].

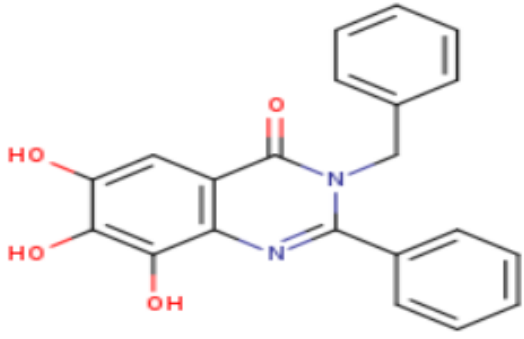
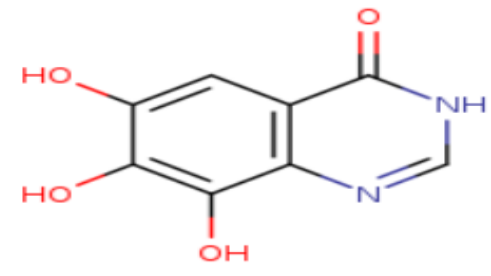
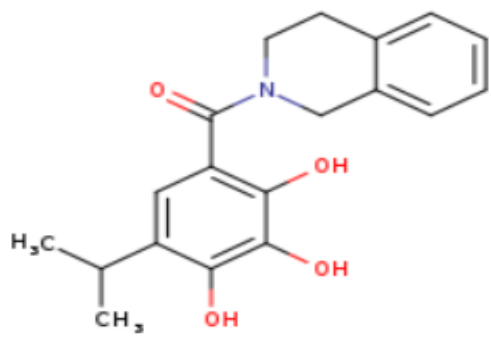
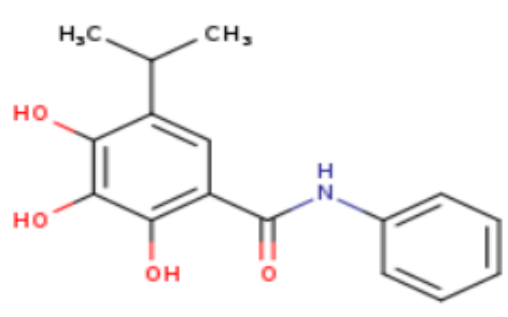
En se référant aux deux documents précédents, nous avons retenu les 14 composés (**Tableau III.2**) se trouvant dans la zone blanche, car ils présentent une forte absorption dans le tractus gastro-intestinal tout en étant conformes aux règles de Lipinski, Veber et Egan. Ceci témoigne de leurs excellentes propriétés pharmacocinétiques.

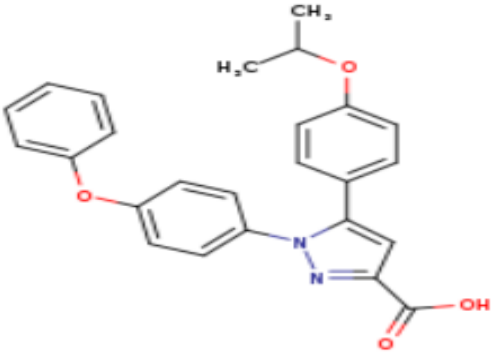
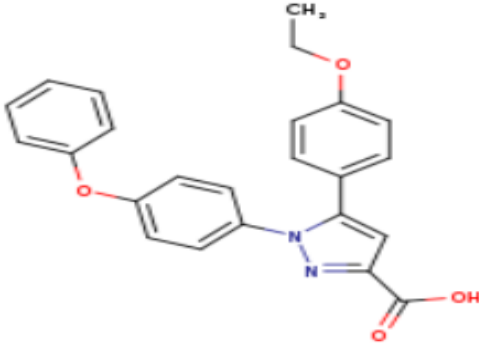
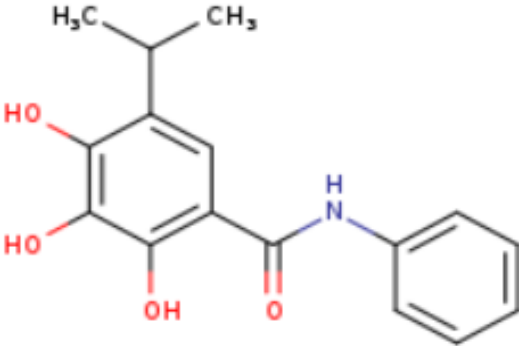
Tableau III.2 : Les structures des composés de la base de données et leurs valeurs de Pki rapportées.

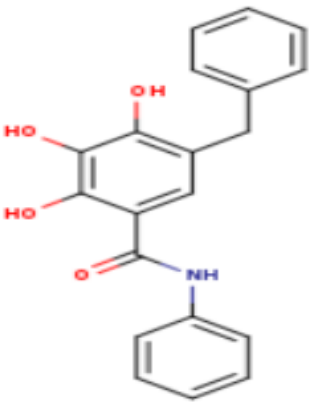
N° de molécule	Molécule	Ki	Pki
1		1113	-5.954
2		179	-6.747

CHAPITRE 03 : RESULTATS ET DISCUSSION

3		1925	-5.716
4		1323	-5.878
5		8040	-5.095
6		1325	-5.878

7		959	-6.018
8		9282	-5.032
9		3100	-5.509
10		31900	-4.496

<p>11</p>		<p>932</p>	<p>-6.031</p>
<p>12</p>		<p>694</p>	<p>-6.159</p>
<p>13</p>		<p>24100</p>	<p>-4.618</p>

14		8300	-5.081
----	---	------	--------

III .2.2. Calculs des descripteurs :

Les descripteurs moléculaires se réfèrent à des propriétés de molécules mesurées expérimentalement ou dérivées théoriquement [12].

Plus précisément, ils offrent des représentations quantitatives des caractéristiques physiques, chimiques ou topologiques des molécules, résumant ainsi notre compréhension de la structure moléculaire et de son activité sous différents aspects.

À la fois, les descripteurs moléculaires et les empreintes moléculaires jouent un rôle central dans l'analyse QSAR/SAR, le criblage virtuel de molécules, la recherche de composés basée sur la similarité, le classement des molécules cibles, la prédiction de l'ADME/Tox des médicaments, ainsi que d'autres processus de découverte de médicaments [13–14].

Dans cette étude, nous avons utilisé une plateforme Web gratuite appelée ChemDes, qui fournit un service en ligne permettant de calculer facilement et instantanément divers descripteurs moléculaires et empreintes digitales. Plus précisément, ChemDes peut calculer 3679 descripteurs et 59 empreintes moléculaires. De plus, ChemDes fournit trois utilitaires utiles, appelés ChemCONV, ChemMOP et ChemFPS, pour une conversion de format pratique, l'optimisation MOPAC et le calcul de similarité d'empreintes digitales [15].

La plateforme Web mentionnée précédemment a permis de calculer et d'identifier un vaste éventail de descripteurs moléculaires, totalisant 1 313 en tout. Dans le cadre du processus de modélisation, divers descripteurs sont rassemblés pour représenter une propriété spécifique, étant donné que les facteurs influençant le processus étudié ne sont pas tous initialement connus. Toutefois, tous les

descripteurs envisagés ne contribuent pas nécessairement de manière significative à la propriété modélisée, et certaines variables peuvent ne pas être indépendantes.

De plus, le nombre de descripteurs, c'est-à-dire la dimension de la base de données d'entrée, détermine la dimension du vecteur de paramètres à ajuster. Si cette dimension est disproportionnée par rapport au nombre d'observations (molécules) dans la base d'apprentissage, le modèle risque d'être surajusté aux exemples donnés, incapable de généraliser pour prédire la propriété modélisée pour de nouvelles molécules, et il peut contenir des informations redondantes. Il est donc essentiel que les descripteurs moléculaires utilisés soient pertinents et interprétables d'un point de vue chimique. En sélectionnant des descripteurs pertinents, ils offrent des indications sur les mécanismes sous-jacents, rendant ainsi les modèles QSAR simples, transparents et compréhensibles [16].

III .2.3. Sélection des descripteurs :

Un nombre minimum de descripteurs est nécessaire pour expliquer la propriété ciblée sans perdre d'informations. Selon Einstein, il est important de rendre tout aussi simple que possible, mais pas plus simple. Pour une modélisation QSAR, il est essentiel que le modèle soit le plus simple possible en termes de paramètres tout en exprimant au mieux les informations contenues dans la propriété [17].

Nous avons d'abord supprimé manuellement les descripteurs avec une valeur nulle, ceux avec des valeurs répétées dans la plupart, voire toutes les molécules, et les descripteurs à valeurs constantes (écarts types inférieurs à 0,001) qui ne fournissent aucune information.

Pour compléter cette procédure, nous avons utilisé le logiciel RapidMiner, anciennement connu sous le nom de YALE (Yet Another Learning Environment) [18]. C'est un outil puissant permettant de mettre en place rapidement une chaîne complète de traitement de données, depuis l'entrée des données jusqu'à la visualisation des résultats d'analyse [19].

Dans RapidMiner, une chaîne de traitement est constituée d'une série de "blocs" connectés les uns aux autres pour représenter la séquence des opérations. Chaque bloc correspond à une opération spécifique : chargement des données depuis un fichier, transformation des données ou l'application d'algorithmes de fouille de données. Une fois la chaîne de traitement terminée, les données peuvent être visualisées ou exportées [19]. L'interface du logiciel, illustrée dans la (**FigureIII.3**), présente une chaîne de traitement simple au centre.

Il est essentiel d'effectuer une analyse bivariée pour un ensemble de descripteurs différents, ce qui implique le calcul du coefficient de corrélation entre chaque paire de descripteurs. Si ce coefficient présente une signification statistique ($R > 0,7$), ces deux indicateurs sont considérés comme étant fortement liés et ne peuvent pas être utilisés en même temps dans l'analyse QSAR. Dans la réalité, ils seront ensuite supprimés lors de la sélection. En pratique, ils seront alors retirés lors de la sélection [20].

Cette approche d'analyse est appelée l'analyse objective. Son objectif est de réduire le nombre de critères sans prendre en compte la variable dépendante (l'affinité dans notre cas) [21].

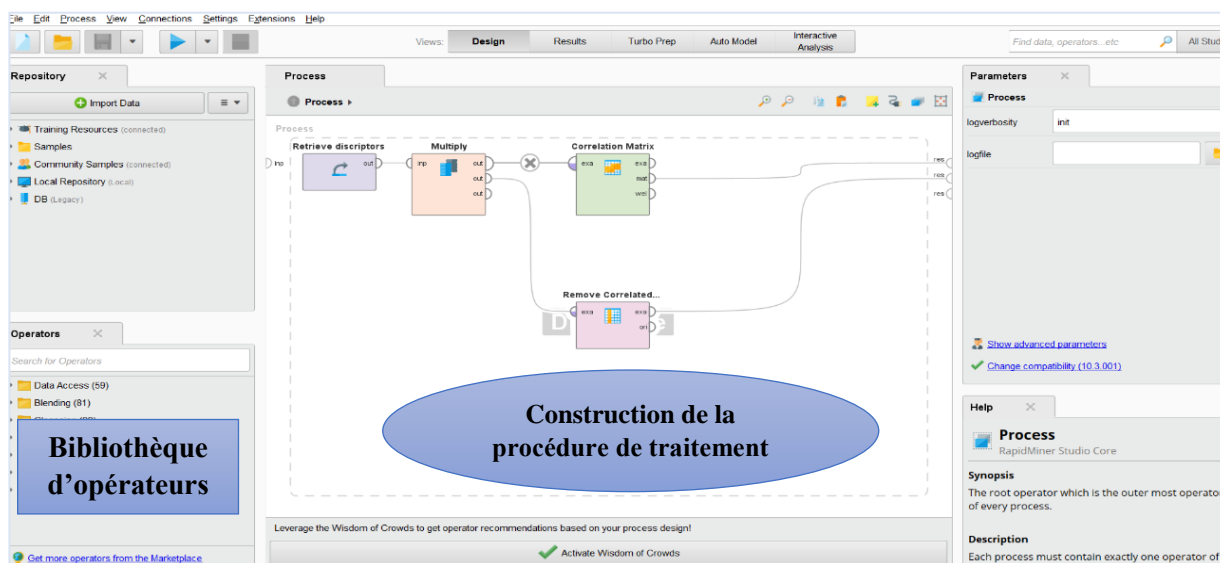


Figure III.3 : Illustration de l'interface de RapidMiner. Dans la zone centrale, la chaîne de traitement contient 4 blocs. Les panneaux gauche et droite sont respectivement dédiés à l'ajout de bloc et au paramétrage du bloc sélectionné.

À la fin de ce processus, nous avons obtenu 14 descripteurs moléculaires, comme le montre **Tableau III.3:**

Tableau III.3 : les descripteurs moléculaires calculés.

N°	Platt	Smax34	Tac	ndonr	S35	kappa3	MRVSA4	naccr
1	100	13.41	6.79	4	13.41	4.94	26.31	7
2	94	13.276	6.605	4	13.276	4.48	13.468	7
3	70	12.308	6.315	5	24.607	5.736	19.765	5
4	82	12.606	6.654	5	25.025	6.016	6.545	5
5	86	12.644	6.824	5	25.101	6.998	12.842	5

6	64	12.233	4.628	3	24.446	2.086	0	5
7	86	13.162	5.815	3	13.162	4.16	6.545	6
8	44	11.195	4.65	4	11.195	1.76	0	5
9	76	12.842	5.202	3	12.842	3.759	32.731	4
10	62	12.209	5.066	4	12.209	4.05	19.765	4
11	94	11.546	5.145	1	11.546	6.533	19.951	5
12	90	11.51	5.1	1	11.51	6.081	6.924	5
13	62	12.209	5.066	4	12.209	4.05	19.765	4
14	74	12.439	5.489	4	12.439	5.042	6.421	4
N°	EstateVSA7	EstateVSA2	EstateVSA3	PEOEVSA7	PEOEVSA6	slogPVSA1	Ki	Log Ki
1	4.984	11.127	11.63	11.127	23.614	29.977	1113	-5.954
2	4.984	11.127	11.63	11.127	19.056	29.977	179	-6.747
3	10.634	12.108	18.908	12.108	24.62	33.679	1925	-5.716
4	10.634	5.563	11.63	12.108	23.763	27.134	1323	-5.878
5	10.634	18.529	18.051	12.108	36.604	33.679	8040	-5.095
6	0	0	6.066	16.69	6.066	26.886	1325	-5.878
7	4.984	11.388	11.63	5.563	11.63	24.871	959	-6.018
8	9.968	0	12.394	0	6.066	25.288	9282	-5.032
9	6.066	18.653	11.984	18.653	29.532	32.671	3100	-5.509
10	5.317	11.251	0	11.251	24.117	21.227	31900	-4.496
11	5.099	11.443	22.75	5.563	80.578	26.96	932	-6.031
12	5.099	18.05	22.75	5.563	73.654	27.463	694	-6.159
13	5.317	11.251	0	11.251	24.117	21.227	24100	-4.618
14	5.317	17.672	5.563	17.672	23.763	21.227	8300	-5.081

III .2.4. Développement de modèle QSAR :

Le modèle 2D-QSAR a été élaboré à l'aide de la méthode de régression linéaire multiple (MLR) via le logiciel SPSS. Ce modèle met en évidence la relation entre l'affinité biologique (variable dépendante) et les descripteurs moléculaires (variables indépendantes) en utilisant des équations linéaires. La méthode de régression linéaire multiple estime les coefficients de régression en appliquant la technique de l'ajustement des moindres carrés [22].

L'élaboration d'un modèle QSAR requiert un ensemble de données varié, nécessitant ainsi la prise en compte de nombreux descripteurs [23]. Pour le développement du modèle MLR, 14 descripteurs ont été utilisés. Les valeurs des descripteurs sélectionnés pour ce modèle sont présentées dans le **Tableau III.3**.

Le meilleur modèle a été sélectionné sur la base de paramètres statistiques, notamment le coefficient de corrélation au carré ($R^2 = 0.918 > 0,6$), qui mesure la qualité relative de l'ajustement [24]. L'erreur standard d'estimation (SEE=0.2252) représente une mesure absolue de la qualité de

l'ajustement. La valeur de Fischer (F=29.784), ou ratio de Fisher, reflète le rapport entre la variance expliquée par le modèle et la variance due à l'erreur dans la régression. Des valeurs élevées du test F indiquent que le modèle est statistiquement significatif [25].

Matrice de corrélation :

Tableau III.4 : matrice de corrélation pour le modèle QSAR avec les facteurs d'inflation de la variance (VIF) des descripteurs selectionne.

	platt	naccr	slogPVSA1	LogKi	vif
Platt	1	0.495	0.317	-0.768	1.363
naccr		1	0.364	-0.814	1.414
slogPVSA1			1	-0.620	1.187
LogKi				1	

Équation et analyse des données:

La corrélation entre l'affinité biologique (Pki) et les descripteurs est exprimée par la relation suivante :

$$Pki = -1,090 - 0,366 Nacc - 0,049 S \log(PVSA1) - 1,090 Platt$$

Les caractéristiques statistiques de l'équation obtenue sont :

$$R^2= 0.918 \quad ; \quad F = 29.784 \quad ; \quad SEE = 0.2252$$

$$N= 14 \quad ; \quad p\text{-value} < 0.05 \quad ; \quad RMSE = 0.225$$

Où N est le nombre de composés (ensemble d'apprentissage). Le modèle présente un bon coefficient de corrélation (R²) entre les descripteurs (nacc, SlogPVSA1, Platt).

Dans ce modèle, les valeurs nacc, SlogPVSA1 et Platt indiquent que l'affinité diminue à mesure que les valeurs des descripteurs augmentent, en raison de leur coefficient négatif dans l'équation précédente.

(1) La p-value est inférieure à 0.05 ,ce qui signifie que nous prendrons un risque inférieur à 0,01% en supposant que l'hypothèse nulle (pas d'effet des variables explicatives) est fausse et que les équations de régression sont statistiquement significatives [21]. Nous pouvons donc conclure avec certitude que les variables choisies fournissent beaucoup d'informations.

(2) La valeur plus élevée de R^2 , associée à une celle des erreurs relativement faibles (**RMSE**) plus faible, indique que le modèle proposé est prédictif et fiable.

(3) Le VIF a été défini comme $VIF = \frac{1}{1-R^2}$ est le coefficient de corrélation pour une variable indépendante par rapport à tous les autres descripteurs du modèle. Les variables ayant un VIF supérieur à 5 sont instables et doivent être éliminées, tandis que les modèles avec des valeurs de VIF comprises entre 1 et 4 peuvent être acceptés [21]. Comme le montre le **tableau III.4**, les valeurs de VIF des trois descripteurs sont toutes inférieures à 5,0, indiquant qu'il n'y a pas de colinéarité parmi les descripteurs sélectionnés et que le modèle résultant présente une bonne stabilité.

(4) La statistique de Fisher F nous permet de déterminer la présence ou l'absence de corrélation entre l'ensemble des descripteurs et l'activité biologique [26]. La valeur ($F = 29.784$) est jugée statistiquement significative au niveau de 95 %, étant donné que la valeur F calculée est supérieure aux valeurs critiques du tableau (voir l'annexe 1) [27].

(5) La valeur SSE la plus petite et la plus proche de 0 est toujours requise pour un modèle QSAR prédictif. Une faible valeur de **SEE = 0.2252** indique que le modèle proposé est fiable et bien ajusté. L'équation de régression a une signification statistique, comme le confirme la valeur F.

III .2.5. Validation de modèle QSAR :

Afin de tester la validité prédictive du modèle QSAR sélectionné, la technique de validation croisée LOO a été utilisée. Le modèle développé a été validé par le calcul des paramètres statistiques suivants :

$$R^2_{adj} = 0.887 \quad ; \quad PRESS = 0.785 \quad ; \quad Q^2_{loo} = 0.841$$

- (1) La valeur élevée du coefficient de détermination ($Q^2_{loo} = 0,887$) pour la validation du modèle prouve le pouvoir prédictif de cette approche et la stabilité du modèle. Elle montre une bonne corrélation entre l'affinité prédite et l'affinité réelle, reflétant ainsi la précision du modèle .
- (2) PRESS est un paramètre important dans la validation croisée, car c'est une bonne approximation de la véritable erreur prédictive des modèles. Sa faible valeur peut être considérée comme statistiquement significative [24]. Plus la valeur PRESS est faible, plus

le modèle est prévisible. Par conséquent, une valeur faible ($PRESS = 0,785$) indique une meilleure prévisibilité du modèle.

- (3) La valeur plus élevée de coefficient du détermination $R^2_{adj} = 0,887$ sont des critères de base pour un meilleur modèle QSAR. La valeur de R^2 indique que **0,918** de variation totale est expliquée par le modèle (bon ajustement), alors que la valeur très élevée du rapport de la variance expliquée par le modèle à la variance résiduelle ($F = 29,784$; $P < 0,05$) permet une très bonne prédiction des ($n = 14$) valeurs de PK_i de l'ensemble de calibrage, (erreur standard $SE = 0,21$), la valeur très élevée de Q^2_{LOO} , qui diffère très peu de celle de R^2 , renseigne sur la robustesse du modèle.

La **figure III.4** ci-dessous illustre la courbe de corrélation obtenue par régression linéaire multiple (MLR) entre les valeurs prédites et les valeurs expérimentales de l'affinité biologique pK_i des composés . La courbe montre un coefficient de corrélation R^2 de **0.918**, ce qui est significatif. Cela indique que ce modèle peut être appliqué avec succès pour prédire l'affinité biologique de cette série de molécules.

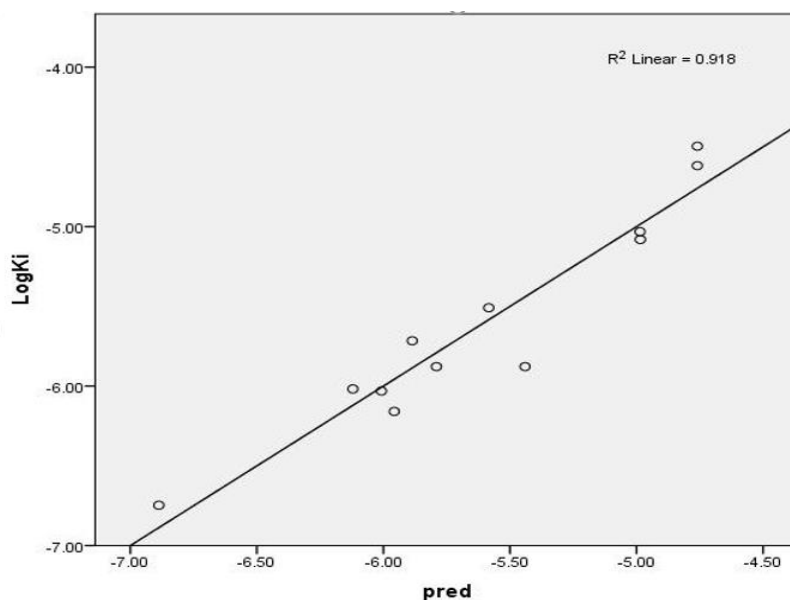


Figure III.4 : corrélation entre les affinités biologiques expérimentales et prédites pour le modèle.

Dans l'élaboration du modèle QSAR, les valeurs résiduelles de l'affinité biologique ont été tracées en fonction des valeurs expérimentales afin de détecter une éventuelle erreur systématique, comme le montre la **figure III.5**. La répartition des résidus de part et d'autre de l'axe des abscisses indique l'absence d'erreur systématique [28]. Cela suggère que ce modèle peut être appliqué avec succès pour prédire l'affinité biologique d'une série de molécules susceptibles d'inhiber l'enzyme BCL-2.

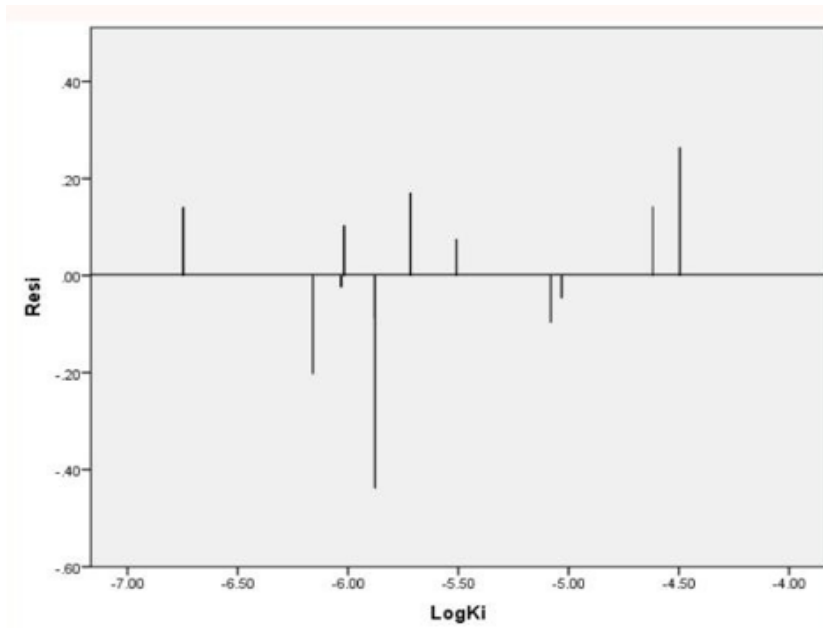


Figure III.5 : Représentations graphique des valeurs résiduelles en fonction des valeurs expérimentales pour le modèle.

Les références :

- [1]. Supuran, C. T., Casini, A., & Scozzafava, A. (2003). Protease inhibitors of the sulfonamide type: anticancer, antiinflammatory, and antiviral agents. *Medicinal Research Reviews*, 23(5), 535-558.
- [2]. Al-Mohammed, N. N., Alias, Y., Abdullah, Z., Shakir, R. M., Taha, E. M., & Hamid, A. A. (2013). Synthesis and antibacterial evaluation of some novel imidazole and benzimidazole sulfonamides. *Molecules*, 18(10), 11978-11995.
- [3]. El Ghalia, H. (2018). Corrélation-(structure-activité anticancéreuse) par les méthodes QSAR des molécules hétérocycliques précurseurs de médicaments.
- [4]. MEDJAHED, S. *Etude par des méthodes QSAR de plusieurs séries de molécules hétérocycliques à intérêt pharmaceutique* (Doctoral dissertation, Université Mohamed Boudiaf des Sciences et de la Technologie-Mohamed Boudiaf d'Oran).
- [5]. Huynh, T. N. P. (2007). *Synthèse et études des relations structure/activité quantitatives (QSAR/2D) d'analyse benzo [c] phénanthridiniques* (Doctoral dissertation, Université d'Angers).
- [6]. Jian-Bo, T. O. N. G., Zhang, X., Ding, L. U. O., & Shuai, B. I. A. N. (2021). Molecular design, molecular docking and ADMET study of cyclic sulfonamide derivatives as SARS-CoV2 inhibitors. *Chinese Journal of Analytical Chemistry*, 49(12), 63-73.
- [7]. Wang, G., Nikolovska-Coleska, Z., Yang, C. Y., Wang, R., Tang, G., Guo, J., ... & Wang, S. (2006). Structure-based design of potent small-molecule inhibitors of anti-apoptotic Bcl-2 proteins. *Journal of medicinal chemistry*, 49(21), 6139-6142.
- [8]. Zhang, Z., Liang, X., Li, X., Song, T., Chen, Q., & Sheng, H. (2013). Design and application of a rigid quinazolone scaffold based on two-face Bim α -helix mimicking. *European journal of medicinal chemistry*, 69, 711-718.
- [9]. Bruncko, M., Oost, T. K., Belli, B. A., Ding, H., Joseph, M. K., Kunzer, A., ... & Elmore, S. W. (2007). Studies leading to potent, dual inhibitors of Bcl-2 and Bcl-xL. *Journal of medicinal chemistry*, 50(4), 641-662.
- [10]. Mandlik, V., Bejugam, P. R., & Singh, S. (2016). Application of artificial neural networks in modern drug discovery. In *Artificial neural network for drug design, delivery and disposition* (pp. 123-139). Academic Press.
- [11]. Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific reports*, 7(1), 42717.
- [12]. Todeschini, R., & Consonni, V. (2008). *Handbook of molecular descriptors*. John Wiley & Sons.
- [13]. Geppert, H., Vogt, M., & Bajorath, J. (2010). Current trends in ligand-based virtual screening: molecular representations, data mining methods, new application areas, and performance evaluation. *Journal of chemical information and modeling*, 2(50), 205-216.

- [14]. Cao, D. S., Dong, J., Wang, N. N., Wen, M., Deng, B. C., Zeng, W. B., ... & Chen, A. F. (2015). In silico toxicity prediction of chemicals from EPA toxicity database by kernel fusion-based support vector machines. *Chemometrics and Intelligent Laboratory Systems*, 146, 494-502.
- [15]. Dong, J., Cao, D. S., Miao, H. Y., Liu, S., Deng, B. C., Yun, Y. H., ... & Chen, A. F. (2015). ChemDes: an integrated web-based platform for molecular descriptor and fingerprint computation. *Journal of cheminformatics*, 7, 1-10.
- [16]. Less, F. P. (1980). Loss prevention in the process industries: hazard identification, assessment and control. Butterworth-Heinemann.
- [17]. Crawley, M. J. (2005). Statistics: an introduction using R. Wiley. Chichester, West Sussex, England.
- [18]. Mierswa I., Wurst M., Klinkenberg R., Scholz M., Euler T. (2006). YALE: Rapid Prototyping for Complex Data Mining Tasks, Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
- [19]. Guyet, T., Nicolas, H., Ghedamsi, B., & Athané, É. (2013, September). Fouille d'images géoréférencées avec RapidMiner. In *Conférence Internationale de Géomatique et d'Analyse Spatiale (SAGEO)*.
- [20]. Trinajstić, N., Nikolić, S., Basak, S. C., & Lukovits, I. (2001). Distance indices and their hyper-counterparts: Intercorrelation and use in the structure-property modeling. *SAR and QSAR in Environmental Research*, 12(1-2), 31-54.
- [21]. Chtita, S. (2017). *Modélisation de molécules organiques hétérocycliques biologiquement actives par des méthodes QSAR/QSPR. Recherche de nouveaux médicaments* (Doctoral dissertation, Université Moulay Ismaïl, Meknès).
- [22]. Sadeghi, F., Afkhami, A., Madrakian, T., & Ghavami, R. (2022). QSAR analysis on a large and diverse set of potent phosphoinositide 3-kinase gamma (PI3K γ) inhibitors using MLR and ANN methods. *Scientific reports*, 12(1), 6090. <https://doi.org/10.1038/s41598-022-09843-0>
- [23]. Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ. Computer science*, 7, e623. <https://doi.org/10.7717/peerj-cs.623>.
- [24]. Khamouli, S., Belaidi, S., & Belkhiri, L. (2018). QSAR Studies of amino-pyrimidine derivatives as Mycobacterium tuberculosis Protein Kinase B inhibitors. *Turkish Computational and Theoretical Chemistry*, 2(2), 16-27.
- [25]. Verma, R. P., & Hansch, C. (2010). QSAR modeling of taxane analogues against colon cancer. *European journal of medicinal chemistry*, 45(4), 1470-1477.
- [26]. Saidani sarra, a. r. (2022). inhibiteurs du sars-cov-2: relation quantitative structure-activité (qsar).

[27]. Dermeche, K., Tchouar, N., Belaidi, S., & Salah, T. (2015). Qualitative Structure-Activity Relationships and 2D-QSAR Modeling of TNF- α Inhibition by Thalidomide Derivatives. *Journal of Bionanoscience*, 9(5), 395-400.

[28]. Tabet, R. Etude des propriétés QSAR d'une série des isiondoles à activité thérapeutique (Doctoral dissertation).

CONCLUSION GENERALE

Conclusion générale :

Le cancer demeure la principale cause de décès à travers le monde, ce qui en fait un problème majeur de santé publique. Le développement de la résistance dans diverses cellules cancéreuses est, dans certains cas, lié à une surexpression de protéines anti-apoptotiques. La famille Bcl-2 représente un ensemble de protéines intracellulaires qui jouent un rôle crucial dans la régulation de l'apoptose.

Pour atteindre notre objectif nous avons appliqué les techniques de modélisation moléculaire à l'étude et à l'optimisation des inhibiteurs de la protéine Bcl-2, une cible cruciale dans le traitement du cancer. Notre travail s'est consacré à la modélisation de la relation quantitative structure-activité (QSAR). Nous avons développé des modèles fiables, robustes, stables et précis, capables de prédire efficacement l'activité anticancéreuse de divers composés appartenant aux inhibiteurs de Bcl-2.

Pour notre étude, nous avons commencé par sélectionner une base de données expérimentale de référence, ce qui contient un ensemble de 108 inhibiteurs différents. Nous nous concentrons sur ceux qui ont le plus de chances d'être bien absorbés par le tractus gastro-intestinal, augmentant ainsi leur efficacité clinique, en même temps présentant des propriétés ADME optimales. Ensuite, nous avons calculé un grand nombre de descripteurs moléculaires, incluant des descripteurs de charge, topologiques, constitutionnels et physicochimiques.

Ce modèle a été élaboré en utilisant une méthode statistique MLR. Les analyses effectuées montrent que des descripteurs spécifiques, tels que nacc, SlogPVSA1 et Platt, sont particulièrement pertinents pour prédire l'activité anticancéreuse des inhibiteurs de Bcl-2. Une fois le modèle QSAR établi, des techniques de validation essentielles, telles que des tests statistiques standard et une validation interne, ont été appliquées pour garantir sa fiabilité et son exactitude. Les résultats obtenus démontrent l'efficacité des modèles QSAR développés pour prédire l'activité des inhibiteurs de Bcl-2.

À partir des différents résultats obtenus au cours de ce travail, nous pouvons affirmer que notre objectif initial a été largement atteint. Les ensembles des molécules, les traitements statistiques et les techniques informatiques utilisés pour le développement et l'analyse des modèles QSAR ont donné de bons résultats. En conclure que les modèles QSAR développés peuvent être utilisés comme outils prédictifs précieux dans la conception de nouveaux composés, potentiellement augmentant l'efficacité des traitements anticancéreux.

LISTE DES ABREVIATIONS

LISTE DES ABREVIATIONS

OMS : L'Organisation mondiale de la santé.

ARNm : L'acide ribonucléique messenger.

ADN : Acide désoxyribonucléique.

CHC : Le carcinome hépatocellulaire.

VHB : Le virus de l'hépatite B.

VHC : Le virus de l'hépatite C.

HPV : Le virus du papillome humain.

HHV8 : Herpès virus humain type 8.

HTLV : Human T-lymphotropic virus.

VIH : Le virus de l'immunodéficience humaine.

THS : Traitement hormonal substitutif.

ARN : L'acide ribonucléique.

G1 : Phase de latence.

S: Phase de réplication.

G2: Phase de réparation.

MTOC: Centres organisateurs de microtubules.

miARN: Les micro-ARN.

VADS : Les cancers des voies aérodigestives supérieures.

MCF7 : Est le nom de la lignée de cellules tumorales mammaires la plus utilisée dans les laboratoires de recherche sur le cancer du sein, (Michigan Cancer Fondation).

AS : un antisens.

EN : Enveloppe nucléaire.

RE : Réticulum endoplasmique.

MEM : Membrane externe mitochondriale.

LISTE DES ABREVIATIONS

QSAR : Quantitative Structure-Activity Relationships.

QSPR : Quantitative Structure-Property Relationships.

ADME : Absorption, de distribution, de métabolisme et d'excrétion.

ADMET : Absorption, de distribution, de métabolisme, d'excrétion et de toxicité.

HTS : High throughput screening (en français, criblage à haut débit).

RMN : Résonance magnétique nucléaire.

2D : 2- Dimensional.

3D : 3-Dimensional.

4D : 4- Dimensional.

5D : 5- Dimensional.

6D : 6- Dimensional.

MLR : La régression linéaire multiple.

RLM : La régression linéaire multiple.

PLS : Régression des moindres carrés partiels.

RNLM : Régression non linéaire multiple.

LOO: Leave-One-Out.

CoMFA: Comparative Molecular Field Analysis.

CoMSIA: Comparative Molecular Similarity Indices Analysis.

ADME: Absorption, Distribution, Métabolisme, et Élimination.

YALE: Yet Another Learning Environment

SEE: L'erreur standard d'estimation.

F: La valeur de Fischer.

PRESS : Somme des carrés des erreurs de prédiction

R² : Coefficient de corrélation au carré

LISTE DES ABREVIATIONS

VIF : Facteurs d'inflation de la variance

RMSE : Erreur Quadratique Moyenne (Root Mean Squared Error)

Q^2_{100} : coefficient de détermination pour la validation croisée

LES ANNEXES

Les annexes:

Tableau 01 : Les significations et les classes des différents descripteurs.

N°	Symbole	Signification	Classe
1	ndonr	Number of H-bond donors	Constitutional descriptors
2	naccr	Number of H-bond acceptors	
3	Platt	Platt number: The Platt number is also known as the total edge adjacency (AE) index.	Topology descriptors
4	Tac	total of absolute charges	Charge descriptors
5	Smax34	Maximum of E-State value of specified atom type(Smax1~Smax79)	Electrotopological State Indices descriptors
6	S35	Sum of E-State of atom type: dO	
7	Kappa3	Molecular shape Kappa index for 3 bonded fragment	Kappa descriptors
8	MRVSA4	Molar Refractivity and Surface area	Surface and molar refractivity descriptors
9	EstateVSA7	MOE-type descriptors using Estate indices and surface area contributions	MOE-type descriptors
10	EstateVSA2		
11	EstateVSA3		
12	PEOEVSA7	MOE-type descriptors using partial charges and surface area contributions	
13	PEOEVSA6		
14	SlogPVSA1	MOE-type descriptors using SLogP contributions and surface area contributions	

LES ANNEXES

Tableau 02 : Table de la loi de Fisher.

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	8	12	24	>25
1	161.4	199.5	215.7	224.6	230.2	234.0	238.9	243.9	249.0	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.41	19.45	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.74	8.64	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.91	5.77	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.53	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.74	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.79	2.61	2.40
12	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.69	2.50	2.30
13	4.67	3.80	3.41	3.18	3.02	2.92	2.77	2.60	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.05	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.23	2.03	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.38	2.20	2.00	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.18	1.98	1.73
25	4.24	3.38	2.99	2.76	2.60	2.49	2.34	2.16	1.96	1.71
26	4.22	3.37	2.98	2.74	2.59	2.47	2.32	2.15	1.95	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.30	2.13	1.93	1.67
28	4.20	3.34	2.95	2.71	2.56	2.44	2.29	2.12	1.91	1.65
29	4.18	3.33	2.93	2.70	2.54	2.43	2.28	2.10	1.90	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.51
60	4.00	3.15	2.76	2.52	2.37	2.25	2.10	1.92	1.70	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.02	1.83	1.61	1.25
>120	3.84	2.99	2.60	2.37	2.21	2.10	1.94	1.75	1.52	1.00



تصريح شرفي

خاص بالالتزام بقواعد النزاهة العلمية لإنجاز بحث

(ملحق القرار 1082 المؤرخ في 2021/12/27)

أنا الممضي أسفله،

السيدة: السيد:
.....

الصفة: طالب سنة ثانية ماستر كيمياء

الحامل (ة) لبطاقة التعريف الوطنية رقم:
الصادرة بتاريخ: 2021/12/22

المسجل بكلية: علوم الطبيعة وعلوم المادة... قسم: علوم المادة.....
والحيات

والمكلف بإنجاز أعمال بحث: مذكرة ماستر في الكيمياء

عنوان البحث: Techniques de modélisation moléculaire
appliquées à l'étude et l'optimisation des inhibiteurs
de la protéine Bcl-2.

أصرح بشرفي أنني ألتزم بمراعاة المعايير العلمية والمنهجية ومعايير الأخلاقيات المهنية والنزاهة
الأكاديمية المطلوبة في إنجاز البحث المذكور أعلاه وفق ما ينص عليه القرار رقم 1082 المؤرخ في
2021/12/27 المحدد للقواعد المتعلقة بالوقاية من السرقة العلمية ومكافحتها.

التاريخ: 23 - 05 - 2024

إمضاء المعني بالأمر