

الجمهورية الجزائرية الديمقراطية الشعبية

République Algérienne Démocratique et Populaire

وزارة التعليم العالي والبحث العلمي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



جامعة الإخوة منتوري قسنطينة I
Frères Mentouri Constantine I University
Université Frères Mentouri Constantine I

Faculté des Sciences de la Nature et de la Vie
Département de Biochimie et Biologie Cellulaire Moléculaire

كلية علوم الطبيعة والحياة
قسم

Mémoire présenté en vue de l'obtention du diplôme de Master

Domaine : Sciences de la Nature et de la Vie

Filière : Sciences Biologiques

Spécialité : Biochimie Appliquée

N° d'ordre :

N° de série :

Intitulé :

*Modélisation de l'activité inhibitrice MRP1 par
l'approche QSAR*

Présenté par : MEZAACHE Rayene
KHENIFAR Sara

Le 22/06/2022

Jury d'évaluation :

- **Encadrant:** Mr. Khalid BOUHEDJAR. MRB CRBt Constantine.
- **Co-encadrant:** Mr. DEMS MOHAMED ABDESSELEM. MRA CRBt Constantine.
- **Examineur 1:** Mr. BENSEGUENI ABDERRAHMANE Pr. UFM Constantine-1
- **Examineur 2:** Mr. MOKRANI ELHASSEN. MCB. UFM Constantine-1.

Année universitaire
2021 - 2022

الجمهورية الجزائرية الديمقراطية الشعبية

République Algérienne Démocratique et Populaire

وزارة التعليم العالي والبحث العلمي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



جامعة الإخوة منتوري قسنطينة I
Frères Mentouri Constantine I University
Université Frères Mentouri Constantine I

Faculté des Sciences de la Nature et de la Vie
Département de Biochimie et Biologie Cellulaire Moléculaire

كلية علوم الطبيعة والحياة
قسم

Mémoire présenté en vue de l'obtention du diplôme de Master

Domaine : Sciences de la Nature et de la Vie

Filière : Sciences Biologiques

Spécialité : Biochimie Appliquée

N° d'ordre :

N° de série :

Intitulé :

*Modélisation de l'activité inhibitrice MRP1 par
l'approche QSAR*

Présenté par : MEZAACHE Rayene
KHENIFAR Sara

Le 22/06/2022

Jury d'évaluation :

- **Encadrant:** Mr. Khalid BOUHEDJAR. MRB CRBt Constantine.
- **Co-encadrant:** Mr. DEMS MOHAMED ABDESSELEM. MRA CRBt Constantine.
- **Examineur 1:** Mr BENSEGUENI ABDERRAHMANE Pr. UFM Constantine-1
- **Examineur 2:** Mr. MOKRANI ELHASSEN. MCB. UFM Constantine-1.

Année universitaire
2021 - 2022



The page is framed by a decorative border. At the top corners, there are black silhouettes of graduates wearing caps, facing each other. Below these are intricate floral designs in shades of pink, red, and green, featuring roses and other flowers. The border is symmetrical and extends down the sides and across the bottom of the page.

Remerciements

On remercie Dieu le tout puissant de nous avoir donné la santé, la volonté et la patience d'entamer et de terminer ce mémoire.

A notre encadrant

Nous souhaitons remercier vivement notre encadrant Mr Khalid BOUHEDJAR chercheur au centre de recherche en biotechnologie (CRBt), pour sa patience, sa disponibilité, sa gentillesse et surtout ses bons conseils afin d'élaborer ce mémoire dans de bonnes conditions.

A notre Co-encadrant

Nous adressons nos sincères remerciements à **Mr. DEMS MOHAMED ABDESSELEM**, maître de recherche classe A au centre de recherche en biotechnologie (CRBt) qui a dirigé notre travail. Ses conseils et ses commentaires précieux nous ont permis de surmonter nos difficultés et de progresser dans notre étude.

A l'équipe du (CRBt)

Nous tenons à montrer notre gratitude à toutes les personnes qu'on a croisées au (CRBt), ce qui a contribué à créer une si bonne atmosphère au cours de notre séjour dans le laboratoire 18.

Aux examinateurs :

Nous exprimons nos vifs remerciements et reconnaissance au **Mr. BENSEGUENI A.** Professeur à l'université des frères Mentouri- Constantine 1 pour l'intérêt qu'il a porté à notre recherche en acceptant d'examiner notre travail .

Nous tenons également à remercier **Mr. MOKRANI E.H.** Maitre-assistant B à l'université des frères Mentouri- Constantine1, D'avoir pris le temps d'examiner notre travail.

Nous adressons nos sincères remerciements à tous les enseignants qui ont assuré notre formation durant les cinq ans d'étude et toutes les personnes qui ont contribué de près ou de loin durant tous notre cycle d'étude.



Dédicace :

*C'est un moment de plaisir et de bonheur
de dédier ce mémoire ainsi qu'exprimer nos
sincères sentiments à nos chers parents
pour leurs encouragements continue, leur
aide, leurs sacrifices, leurs patiences
illimitées et leurs profond et sincère
amours.*



ملخص

هدفت هذه الدراسة الى إنشاء نماذج QSAR موثوقة، مستقرة وتنبئية.

استعملت منهجية QSAR في نمذجة النشاط المثبط MRP1 الذي يمثلته (pIC_{50}) ل 105 من مشتقات بيرولوبيريميدين وذلك من خلال مقارنة نتائج البرنامجين (CORAL و DRAGON-QSARINS). تم التحقق من صحة النماذج التي أنشأها هذان البرنامجان من خلال التحقق الداخلي والخارجي للحكم على استقرارهما، قوتهما وقدراتهما التنبئية.

نتائج QSARINS : ($Q^2 = 81,5\%$ / $R^2 = 83,02\%$) حيث ان النشاط المثبط كان مرتبطاً بثلاث واصفات (GATS2v VEA2 ، T (N-F)). اذ تم استخدام شروط Tropsha و Golbraikh والتحقق منها.

نتائج CORAL : ($Q^2 = 85\%$ ، $R^2 = 87\%$) حيث وجدنا أن هذا النموذج قوي للغاية، وله معاملات تنبئية داخلية وخارجية ممتازة. لذا فهي قادرة على التنبؤ بالنشاط المدروس للمركبات الجديدة.

اذن يمكن استخدام هذا النموذج بشكل فعال لتقدير الجزيئات التي لم يتم اختبارها تجريبياً.

Mots clés : Pyrrolopyrimidine, MRP1, QSAR, méthode Monte Carlo, chimio bioinformatique.

Résumé

Ce mémoire a pour objectifs d'établir des modèles QSAR fiables, stables et prédictifs.

La méthodologie QSAR dont deux approches (DRAGON-QSARINS et CORAL) sont utilisées pour modéliser l'activité inhibitrice MRP1 représenté par (pIC_{50}) de 105 dérivés du pyrrolopyrimidine, en comparant leurs résultats, ainsi choisir le model le plus performant.

Les modèles établis de ces deux approches ont prouvé leurs validations internes et externes confirmant leurs stabilités, leurs robustesses et leurs capacités prédictives.

Pour le modèle DRAGON-QSARINS ($R^2 = 83.02 \%$, $Q^2 = 81.5 \%$) dont l'activité inhibitrice a été corrélés avec 3 descripteurs (T(N..F), GATS2v, VEA2). Les résultats sont conformes aux principales recommandations du Golbraikh et Tropsha.

Les résultats obtenus par CORAL ($R^2 = 87 \%$, $Q^2 = 85 \%$) nous ont permis de conclure que ce dernier possède d'excellents paramètres prédictifs internes et externes, et un caractère très robuste et stable pouvant prédire l'activité pour de nouveaux composés.

Enfin, le model CORAL peut être utilisé efficacement pour estimer des molécules qui n'ont pas été testées expérimentalement.

Mots clés : Pyrrolopyrimidine, MRP1, QSAR, méthode Monte Carlo, chimio bioinformatique.

Abstract

This study aims to develop stable, predictive and reliable QSAR models. The QSAR methodology uses both QSARINS and CORAL softwares to generate the MRP1 inhibitory activity represented by pIC_{50} of 105 pyrrolopyrimidine derivatives through comparison. The established models have proven the internal and external validation, results throughout their stability, predictiveness and reliability.

The DRAGON-QSARINS model ($R^2 = 83.02\%$, $Q^2 = 81.5\%$) have been correlated the inhibitory activity with 3 descriptive parameters; (T(N..F), GATS2v, VEA2). These results were in concordance with the Golbraikh and Tropsha conditions.

The CORAL model ($R^2 = 87\%$, $Q^2 = 85\%$) has excellent internal and external predictive parameters and reliable and solid character that can help predict the activity of novel compounds.

In conclusion, the CORAL model can be reliably used to estimate molecules and compounds that have not been experimentally tested.

Keywords: Pyrrolopyrimidine, MRP1, QSAR, méthode Monte Carlo, chimio bioinformatique.

Liste des tableaux

Tableau	Titre	Page
Tableau 1	Classification des transporteurs ABC humains.	13
Tableau 2	Substrats de MRP1.	16
Tableau 3	Exemples de molécules avec leur code SMILES.	21
Tableau 4	Les descripteurs intervenant dans les modèles.	41
Tableau 5	Statistiques du modèle 1.	42
Tableau 6	Statistiques du modèle 2.	42
Tableau 7	Statistiques du modèle sélectionné.	43
Tableau 8	Les paramètres potentiels prédictifs d'un modèle pour les 3 choix.	47
Tableau 9	Variation des paramètres statistiques selon l'approche utilisée.	49

Liste des Annexes

Annexes	Titre	Page
Annexes 01	Valeurs expérimentales prédit et calculé par l'approche DRAGON-QSARINS pour les 105 dérivés de pyrrolopyrimidines.	A
Annexes 02	Activité inhibitrice envers MRP1 de 105 dérivés de pyrroloririmidines le choix de split pour l'approche CORAL.	E

Liste des figures

Figure	Titre	Page
Figure 1	les facteurs externes du cancer(NCI).	4
Figure 2	Facteurs de risque modifiables et non modifiables.	5
Figure 3	Vue d'ensemble de la recherche et développement pour un nouveau médicament.	8
Figure 4	Les quatre grands mécanismes de résistance.	12
Figure 5	L'orientation et la topologie de MRP1.	14
Figure 6	Modèle hypothétique du cycle du transport par MRP1.	18
Figure 7	Matrice d'adjacence du 2-méthylbutane représenté après élimination de l'hydrogène.	20
Figure 8	Exemple de courbe dose-réponse pour la DL ₅₀ .	23
Figure 9	Les méthodes disponibles pour tester les molécules.	24
Figure 10	Stratégie globale d'une étude QSAR.	32
Figure 11	Diverses voies de synthèse pour les dérivés de la pyrrolopyrimidine.	34
Figure 12	Activités biologiques des pyrrolopyrimidine.	35
Figure 13	Droites d'ajustement pour le modèle (03) des valeurs expérimentales et prédits de la notation pIC_{50} pour le modèle de QSAR.	44
Figure 14	Principe du test de randomisation.	45
Figure 15	Tests de randomisation.	46
Figure 16	Diagramme de Williams.	46

Liste des Symboles et abréviations

Abreviation	Signification
R_0^2	Coefficient of determination: observed versus predicted activities.
R_0^2	Coefficient of determination (determined the predicted versus observed activities).
ABC	Transporteurs ATP-binding cassette.
ADME	absorption, distribution, métabolisme et excrétion.
AEM	L'Agence européenne des médicaments.
AG-SSV	Sélection de sous-ensembles de variables par algorithme génétique.
AM1	Austin Mode 1.
ATP	Adénosine triphosphate.
C	Concentration de l'effet.
CBNPC	Cancers bronchiques non à petites cellules.
CBPC	Cancers bronchiques à petites cellules.
CCC	Coefficient de corrélation de concordance.
CHC	Carcinome hépatocellulaire.
CIRC	Centre international de recherche sur le cancer.
CORAL	CORrelation And Logic
CTD	Common Technical Document.
DCW	Correlation weights / degree cooling weeks
DL ₅₀	La dose qui cause la mort de 50% de la population.
EC ₅₀	Concentration d'effet.
EGFR	Récepteur du facteur de croissance épidermique.
Es	Constante de substituant de Taft.
F	Indice de Fisher.
HUGO	Human Genome Organisation.
IC ₅₀	Concentration inhibitrice.
IIC	Index of ideality of correlation.
INCa	Institut nationale du cancer.

InChI	International Chemical Identifier.
K	Slope: predicted versus observed activities regression lines through the origin.
k'	Slope: observed versus predicted activities regression lines through the origin.
LD ₅₀	La dose nécessaire pour la moitié des membres d'une population testée après une durée de test spécifiée.
LOO	Validation croisée par omission d'une observation : Cross-validation by leaveone-out.
MAE	Mean absolute error.
MDR	Multi drug résistance.
MLR	Régression linéaire multiple.
MM+	Mécanique Moléculaire (+).
MRP1	Multidrug resistance-associated protein 1.
N	Dimension de la population (échantillon).
N	The number of compounds in the se.
NBD	Nucleodide-binding domain.
OCDE	Organisation de Coopération et de Développement Economiques.
MCO	Moindres carrés ordinaires.
OMS	Organisation mondiale de la santé.
P	Coefficient de partage octanol-eau.
PDK1	3-Phosphoinositide-dependent kinase 1.
PGP	La glycoprotéine P.
<i>p</i> IC ₅₀	log 1/ (IC ₅₀).
PM3	Parametrization Method 3.
PRESS	Somme des carrés des erreurs de prédiction.
Q	Cross-validated correlation coefficient.
Q ²	Coefficient de validation externe.
Q ² _{LMO}	Coefficient de prédiction (Leave-Many-Out).
Q ² _{LOO}	Coefficient de prédiction (Leave-One-Out).
QSAR	Relation quantitative-structure-activité.
QSARINS	QSAR-INSUBRIA.

QSPR	Relation quantitative-propriété-activité.
R	Coefficient de corrélation.
R ²	Coefficient de détermination.
R ² _{adj}	Coefficient de détermination ajusté.
RMN	Résonance Magnétique Nucléaire.
S	Standard error of estimation.
SAR	Relation -structure-activité.
SCC	Carcinome spinocellulaire.
SFR	Relapse-free survival.
SMILES	Simplified Molecular Input Line Entry System.
TMD	Trans-membrane domaine.
Tr	Trining.
UICPA	L'Union internationale de chimie pure et appliquée.
$\bar{y}(i)$	Valeur prédite.
\hat{y}_i	Valeur estimée.
y_i	Valeur observée.



Sommaire

Table des matières

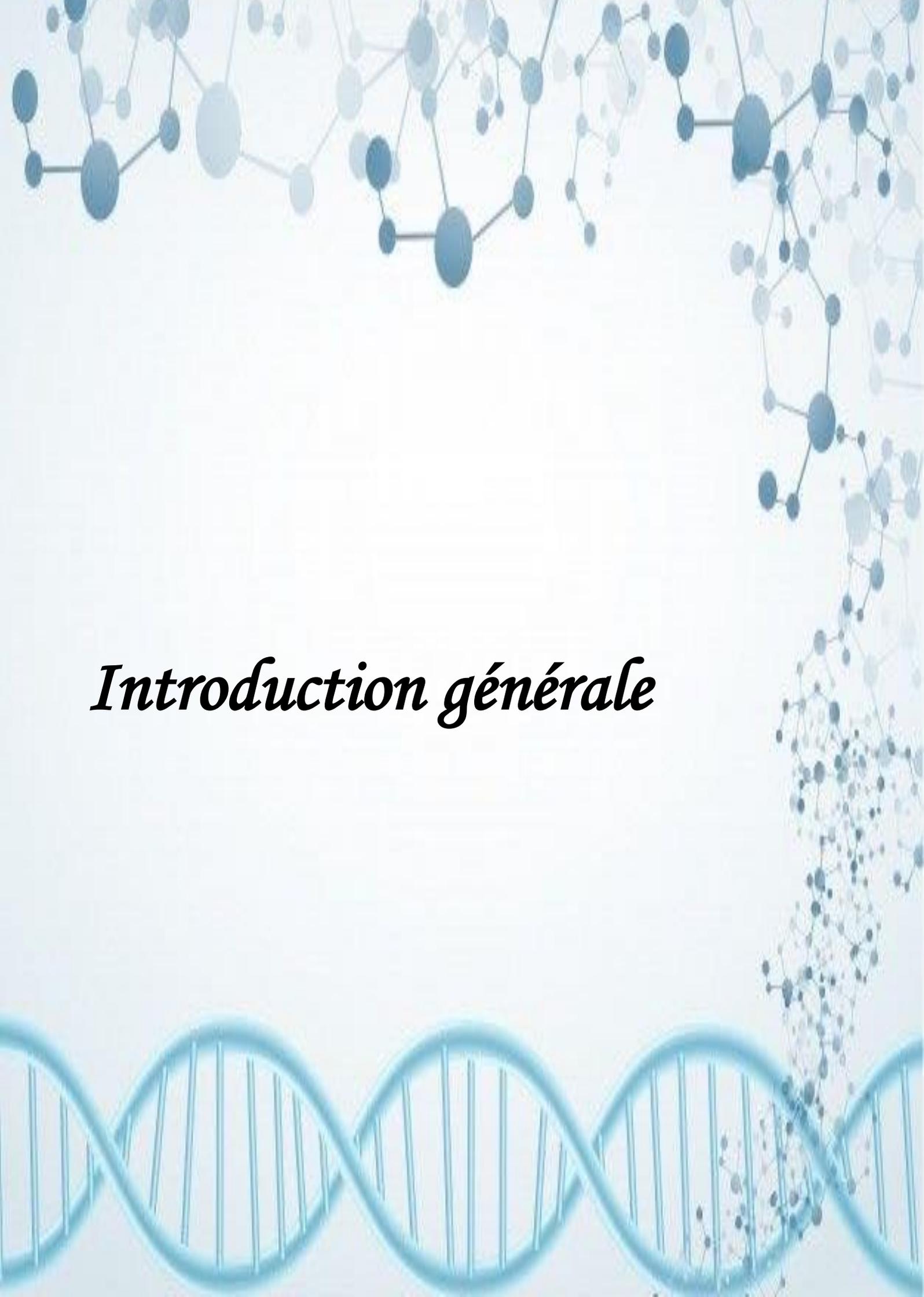
<i>Remerciements</i>	II
<i>Dédicace</i> :	III
<i>Résumé</i>	V
<i>Liste des tableaux</i>	VI
<i>Liste des Annexes</i>	VII
<i>Liste des figures</i>	VIII
<i>Liste des Symboles et abréviations</i>	IX
<i>Introduction générale</i>	1
<i>Partie I : Synthèse bibliographique</i>	3
<i>Chapitre 1 : La résistance des protéines aux médicaments anticancéreux</i>	4
1. Le cancer :	3
1.1. Généralités :	3
1.2. Les facteurs de risque des cancers :	3
1.3. Les cancers les plus fréquents :	4
1.3.1. Cancer du sein :	4
1.3.2. Cancer des poumons :	6
1.1.3. Cancer du côlon :	6
1.1.4. Cancer de la prostate :	7
2. La conception de médicaments ou Drug Design :	7
2.1. Les étapes de développement du médicament :	7
2.1.1. La recherche :	8
2.1.2. Les études précliniques :	8
2.1.3. Les études cliniques :	9
2.1.4. Enregistrement et Autorisation de Mise sur le Marché :	10
2.2. Les approches de la conception de médicaments :	10
2.2.1. Approche « Ligand-based » :	10
2.2.2. Approche « Structure-based » :	10
3. Multi-Drug résistance dans le cancer :	11
3.1. Les transporteurs ABC :	12
3.1.1. La protéine de résistance à de multiples drogues : MRP1	13
3.1.2. Topologie du transporteur MRP1 :	14
3.1.3. Localisation de MRP1 :	14
3.1.4. Fonction physiologique de MRP1 :	15
3.1.5. Substrats transportés par MRP1 et spécificité :	16

3.1.6.	<i>Cycle du transport de MRP1 :</i>	17
3.1.7.	<i>Implication clinique de MRP1 dans le cancer :</i>	18
<i>Chapitre02 : La méthodologie QSAR</i>		20
1.	<i>Chimio-Bio-informatique :</i>	21
2.	<i>La représentation moléculaire :</i>	21
2.1.	<i>Théorie des graphes :</i>	21
2.2.	<i>La notation SMILES :</i>	22
3.	<i>La biologiques des produits chimiques et toxicologie :</i>	22
4.	<i>Essais biologiques de toxicité :</i>	23
5.	<i>Méthodes in silico méthodes alternatives :</i>	25
5.1.	<i>QSAR :</i>	26
5.1.1.	<i>Descripteurs Moléculaire :</i>	26
5.1.2.	<i>Les types de descripteurs moléculaires :</i>	27
5.1.2.1.	<i>Les descripteurs 1-D :</i>	27
5.1.2.2.	<i>Les descripteurs 2-D :</i>	27
5.1.2.3.	<i>Les descripteurs 3-D :</i>	27
5.1.3.	<i>Construction de modèles de QSAR :</i>	28
5.1.4.	<i>Paramètres d'évaluation des modèles :</i>	29
5.1.5.	<i>Interprétation du Modèle :</i>	31
5.1.6.	<i>Stratégie globale d'une étude QSAR :</i>	32
6.	<i>Conclusions :</i>	33
<i>Partie II : Etude Expérimentale</i>		34
1.	<i>Matériels et méthodes :</i>	34
1.1.	<i>Les pyrrolopyrimidines :</i>	34
1.1.1.	<i>Stratégies de synthèse des pyrrolopyrimidines :</i>	34
1.1.2.	<i>Profil pharmacologique des pyrrolopyrimidines :</i>	35
1.1.3.	<i>Activité anticancéreuse :</i>	35
1.2.	<i>Origine des données :</i>	37
1.3.	<i>Traitements des données :</i>	37
1.4.	<i>Répartition des échantillons :</i>	38
1.5.	<i>Méthode de sélection de descripteurs :</i>	38
1.6.	<i>La Régression Linéaire multiple MLR :</i>	38
1.7.	<i>Régression par la méthode de Monte Carlo :</i>	39
2.	<i>Résultat et discussions :</i>	41
2.1.	<i>Construction model dragon descripteur :</i>	41
2.1.1.	<i>Analyse de la régression :</i>	41

2.1.1.1.	<i>Model pour 1 Descripteur T(N..F) :</i>	42
2.1.1.2.	<i>Model pour 2 Descripteurs T (N..F) BELe2 :</i>	42
2.1.1.3.	<i>Model pour 3 Descripteur T(N..F) GATS2v VEA2:</i>	42
2.1.2.	<i>Qualité de l'ajustement :</i>	43
2.1.3.	<i>Validation :</i>	44
2.1.3.1.	<i>Les Principes de l'OCDE :</i>	44
2.1.3.2.	<i>Validation interne :</i>	45
2.1.3.3.	<i>Validation externe :</i>	45
2.1.3.4.	<i>test de randomisation (Y scrambling):</i>	45
2.1.4.	<i>Le domaine d'application :</i>	46
2.1.5.	<i>Conclusion :</i>	47
2.2.	<i>Construction model CORAL :</i>	47
2.2.1.	<i>Descripteurs optimal SMILES-based :</i>	47
2.2.2.	<i>Conclusion :</i>	48
2.3.	<i>Comparaison entre les modèles DRAGON-Based et Optimal SMILE-Based :</i>	48
<i>Conclusion générale</i>		49

Annexes

Liste des références



Introduction générale

Introduction générale

Aujourd'hui, le cancer subit une croissance qui va de pair avec celle de la population mondiale. Il était donc une maladie parmi tant d'autres, devant laquelle les médecins étaient contraints de constater leur impuissance. Ce n'est que dans les années 1950 que le cancer est devenu un problème pour la société dans son ensemble. [1]

L'innovation thérapeutique qui lutte contre cette maladie consiste à la mise en œuvre de méthodes de traitement basées sur les caractéristiques génétiques des tumeurs. Mais malheureusement, l'émergence de la résistance aux drogues qui cause des échecs thérapeutiques a actuellement invalidé l'action de nombreux agents chimiothérapeutiques qui sont des principaux moyens de contrôle des cancers.

L'échec des chimiothérapies contre les microbes pathogènes et les cancers est dû aux exportateurs ABC responsable du transport d'un large éventail de médicaments, dont la protéine MRP1 exprimée dans une large gamme de tissus, normaux et tumoraux. De nombreuses études relient sa présence à un mauvais pronostic de traitement et sa responsabilité dans la résistance à un large spectre de médicaments. Des inhibiteurs de cette protéine ont été évalués dans des essais cliniques contre le cancer et les maladies liées au cancer, mais avec un succès limité et avec des effets secondaires graves.

Le développement de nouveaux composés inhibant cette protéine (MRP1), avec une puissance, une sélectivité élevées et sans effets secondaires graves est souhaitable. Tel que Les pyrrolopyrimidines qui représente une classes de ligands intéressantes, car elle été largement explorées par les chimistes médicaux en raison de leur vaste profil pharmacologique anticancéreux. [2]

En effet, le processus de recherche et de développement, dans l'industrie pharmaceutique dure en moyenne dix à quinze ans pour environ une molécule sur 10 000 testées, et il utilise plusieurs méthodes coûteuses, difficiles et inefficaces [3], Ce qui a obligé les chercheurs à s'orienter vers de nouvelles méthodes qui permettent de modéliser et de prédire rapidement l'activité molécules avant même qu'elles soient synthétisées. Parmi ces méthodes, la recherche de similarité, le criblage à l'aide de pharmacophore et les méthodes QSAR. [4]

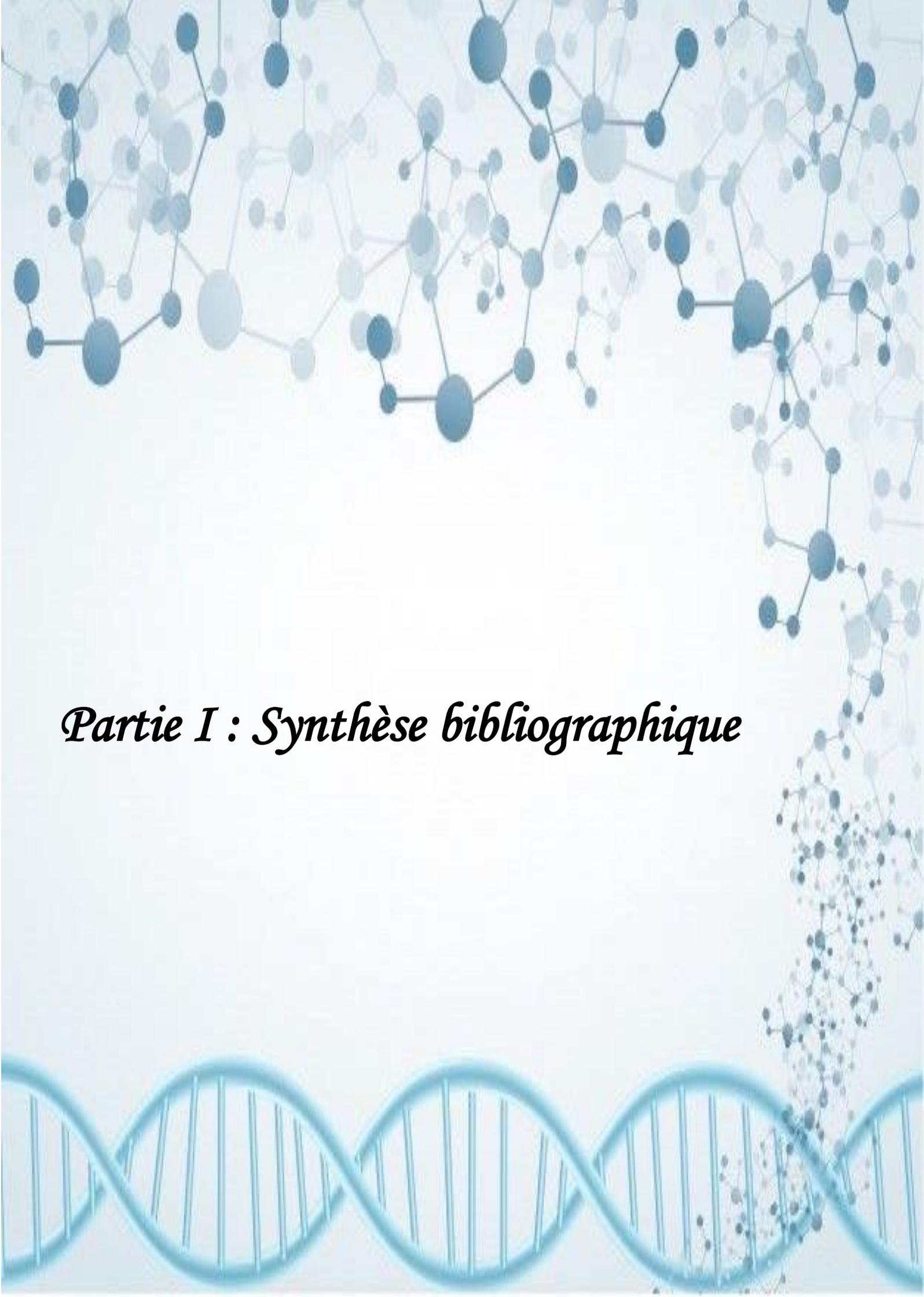
Ces dernières comprennent l'utilisation de méthodes d'analyse multidimensionnelle afin de modéliser la relation entre les structures des composées et leurs activités biologiques grâce à l'apprentissage statistique.

Notre étude vise à développer un modèle robuste et fiable en se basant sur l'utilisation de l'approche QSAR dans le but d'expliquer et prédire l'activité anticancéreuse d'une série de 105 dérivés de pyrrolopyrimidine inhibiteurs de MRP1.

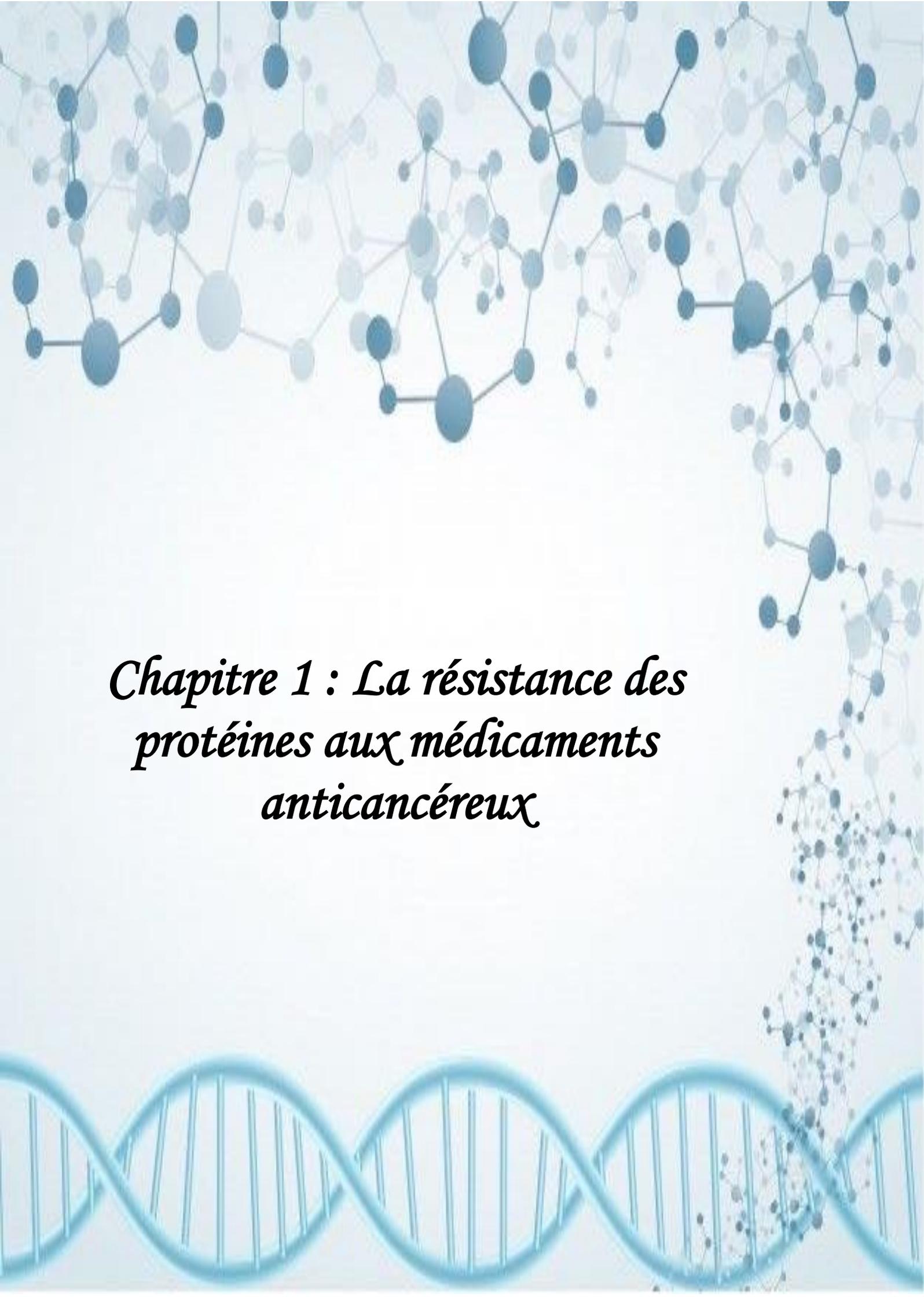
Le manuscrit de ce mémoire est articulé sur deux grandes parties précédées d'une introduction générale :

- une partie bibliographique qui contient deux chapitres :
 - le 1^{er} chapitre est intitulé « La résistance aux médicaments dans le cancer » où nous avons abordés un aperçu général sur le cancer, La conception de médicaments (Drug Design) et la multi-drug résistance d'ont la protéine MRP1.
 - le 2^{ème} chapitre nommé « Les méthodes QSAR » qui représente un Rappel sur la méthodologie QSAR.
- une deuxième partie expérimentale où nous avons présenté, discuté et comparé les résultats obtenus par l'approche QSAR en utilisant 2 logiciels QSARINS et CORAL SEA que nous avons traités tout au long de notre séjour au centre de recherche en biotechnologie de Constantine (CRBt).

Enfin, nous clôturons par une conclusion générale suivie d'une annexe.



Partie I : Synthèse bibliographique

The background features a light blue gradient with several molecular models. At the top, there are complex, interconnected networks of blue and grey spheres representing atoms and molecules. At the bottom, a prominent blue DNA double helix structure is visible, winding across the width of the page. The overall aesthetic is scientific and clean.

*Chapitre 1 : La résistance des
protéines aux médicaments
anticancéreux*

1. Le cancer :

1.1. Généralités :

Le cancer, appelé également tumeur maligne ou néoplasme est défini par l'Organisation Mondiale de la Santé (OMS) comme : « un grand groupe de maladies pouvant toucher n'importe quelle partie de l'organisme ». Selon le Centre international de recherche sur le cancer (CIRC) de l'OMS, une personne sur cinq développe un cancer au cours de sa vie. Un homme sur huit et une femme sur onze meurent de cette maladie. Son incidence a considérablement augmenté.

Il s'agit d'un processus multi-étapes et progressif, qui s'étend sur plusieurs années, voire décennies avant d'être détectable cliniquement. Ces cellules commencent à se diviser de façon désordonnée au lieu de mourir. Bien qu'il existe de nombreux types différents de cancer, ils se forment tous à la suite du développement anormal et incontrôlé de cellules. Qui à la fin forme une masse appelée tumeur, celle-ci peut être cancéreuse, tumeur maligne, tumeur bénigne.

Des cellules peuvent se détacher de la tumeur, migrer par l'intermédiaire des vaisseaux sanguins et lymphatiques, puis envahir les tissus voisins et essaimer dans d'autres organes. On parlera alors de métastases. Chaque catégorie de cancer a ses propres caractéristiques. [5]

1.2. Les facteurs de risque des cancers :

L'étiologie des cancers est multifactorielle et complexe. Malgré tout, à ce jour, de nombreux facteurs de risque de cancer ont été identifiés [6].selon (CIRC), On cite quatre facteurs principaux qui sont : l'essor démographique, le vieillissement de la population, l'environnement et le mode de vie.

Il existe tout d'abord des facteurs de risque dits « internes », liés à l'âge ou l'histoire familiale. En vieillissant, le risque de développer un cancer devient plus important. Différents facteurs génétiques liés au fait de porter une mutation sur un ou plusieurs gènes peuvent également jouer un rôle dans le risque de développer un cancer.

Il existe ensuite des facteurs dits « externes » liés au mode de vie ou à l'environnement. Nous pouvons citer, par exemple : le stress, les habitudes alimentaires, le tabagisme, le surpoids, l'obésité, le manque d'activité physique, l'exposition excessive aux rayonnements UV et l'exposition à certains virus ou bactéries.

L'Institut national du cancer (INCa) a classé Les facteurs de risque liés à l'environnement et qui sent : le contact avec certains produits chimiques, notamment dans le cadre professionnel (amiante, benzène, poussière de bois, etc.) ; l'exposition à certaines substances polluantes présentes dans l'air, l'eau ou la terre ; les rayonnements d'origine naturelle, comme le radon, ou d'origine artificielle comme les rayons X .la Figure1 montre les facteurs externes du cancer.

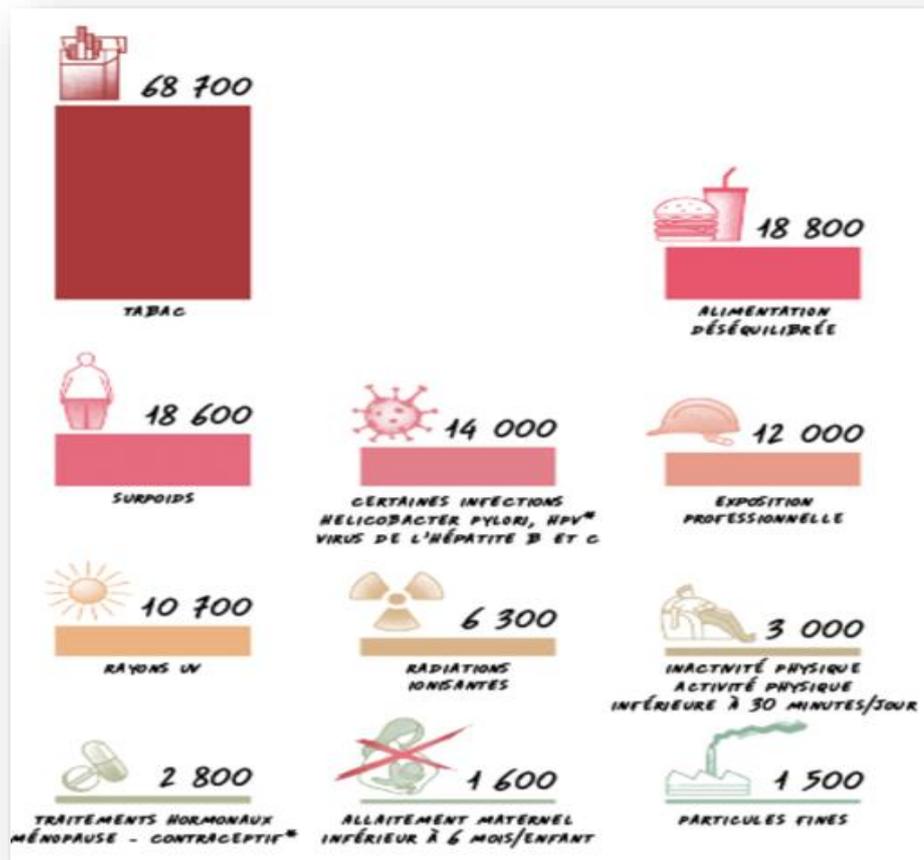


Figure 1 : les facteurs externes du cancer (INCa)

1.3. Les cancers les plus fréquents :

en termes de nombre de cas recensés par (OMS) en 2020 les cancers les plus courants étaient les suivants [7] :

1.3.1. Cancer du sein :

Le cancer du sein a été signalé comme étant le cancer le plus répandu à l'échelle mondiale avec 2,26 millions de cas par ans. Les taux d'incidence sont élevés dans les pays les plus développés, tandis que les taux dans les pays moins développés sont faibles mais en augmentation. Son incidence augmente rapidement avec l'âge pendant les années de

reproduction, puis augmente à un rythme plus lent à l'âge moyen de la ménopause; il avait quasiment doublé entre 1980 et 2005, puis stagner et diminuer depuis. [8], [9]

C'est une forme de tumeur maligne qui se développe au niveau du sein. Il existe différents types de cancer du sein. Les plus fréquents (95 %) sont des adénocarcinomes : ils se développent à partir des cellules épithéliales (= carcinome) de la glande mammaire. Il existe également d'autres formes, beaucoup plus rares, comme des lymphomes et des sarcomes. [10]

La figure 2 présente les Facteurs de risque modifiables et non modifiables, Soulignant les facteurs particulièrement importants ou qui ont été établis récemment, dont l'effet d'œstradiol plus celui de la saturation des dépistages lié aux mammographies ainsi que la diminution du nombre de prescriptions de traitements hormonaux substitutifs pour les femmes ménopausées.

Les antécédents personnels du cancer du sein sont un facteur de risque majeur ainsi que L'âge, Les facteurs familiaux ou génétiques. [11]–[14]



Figure 2 : Facteurs de risque modifiables et non modifiables

Malgré tout cela, Il est considéré comme un cancer de bon pronostic ; La diminution de la mortalité relative est liée à deux éléments principaux : les progrès des traitements d'une part, et d'autre part, le dépistage précoce des cancers. La majorité des cancers du sein sont diagnostiqués alors qu'ils sont encore à un stade peu avancé. [15]

1.3.2. Cancer des poumons :

Cancer du poumon, appelé aussi cancer bronchique ou cancer broncho-pulmonaire ; avec 2,21 millions de cas en 2020 d'après (OMS) il prend la 2^{ème} place des cancers les plus fréquent mondialement, reflétant ainsi l'efficacité limitée des chimiothérapies cytotoxiques conventionnelles chez les patients atteints de cette maladie.

Différentes cibles moléculaires ont été découvertes à ce jour. Mais toutes ne disposent pas d'un inhibiteur conçu avec succès. L'évolution rapide de la maladie après le diagnostic, reflète ensuite la faible survie à 5 ans ; elle est fortement influencée par l'histologie et le stade au moment du diagnostic. [16], [17]

Les cellules qui se multiplie de façon anarchique tapissent les alvéoles pulmonaires jusqu'à former une masse appelée tumeur maligne. Les principaux types histologiques de cancer du poumon englobent les cancers bronchiques à petites cellules (CBPC, 15 % des cas) et les cancers bronchiques non à petites cellules (CBNPC), 85 % des cas). Le NSCLC comprend l'adénocarcinome (AC) (environ 40% à 50 % de tous les cancers du poumon), le carcinome épidermoïde (SQCC) et le carcinome à grandes cellules (LCLC) (environ 25 % du nombre total de cas). [18], [19].

Le tabac est le premier facteur de risque du cancer du poumon. Il y a aussi D'autres facteurs environnementaux ou professionnels (l'amiante, des gaz d'échappement des moteurs, du radon, d'hydrocarbures polycycliques aromatiques, de certains rayonnements ionisants, de la silice, du cadmium). [20]

1.1.3. Cancer du côlon :

Le cancer du côlon est considéré comme le troisième cancer le plus fréquent avec 1,93 million cas en 2020 et la deuxième cause de décès par cancer chez les hommes et les femmes avec 916 000 décès dans le monde entier d'après l'OMS. Il a été mis en cause dans le comportement humain en ce qui concerne un régime alimentaire hyper protéique, riche en graisses et pauvre en fibres, le tabagisme, le manque d'activité physique, l'obésité et le stress, tous cela constituent des facteurs de risque majeurs de ce cancer. [21]–[23]

Dans leur grande majorité (70 %), les tumeurs du côlon se développent sur la dernière partie du côlon : le sigmoïde. L'adénocarcinome est la forme la plus fréquente des cancers touchant le côlon (95 %) (d'adénocarcinomes lieberkühniens (85 %), ou d'adénocarcinomes mucineux). [24], [25]

Comme traitement actuel pour le cancer du côlon, une combinaison entre la chimiothérapie, la chirurgie et la radiothérapie. Cette dernière ne fournit qu'un succès modéré, en particulier pendant les dernières étapes du cancer. Il est donc nécessaire d'adopter de nouvelles approches en matière d'agents prophylactiques et thérapeutiques contre ce cancer (bactéries probiotiques qui présentent de nombreuses fonctions bénéfiques dans le tractus gastro-intestinal et ce pour leur potentiel biothérapeutique alternatif contre le cancer du côlon). [26]–[29]

1.1.4. Cancer de la prostate :

Le cancer de la prostate est le quatrième cancer le plus fréquent avec 1,41 million de cas et le deuxième cancer le plus fréquent chez les hommes. [30], [31]

Il faut savoir que le cancer de la prostate évolue le plus souvent lentement, parfois pendant plus de 10 ans donc il peut être asymptomatique à un stade précoce. Sous forme de tumeur qui évolue généralement lentement, à partir de cellules de la glande prostatique devenues anormales. Selon l'avancée de la maladie, on parle de :

- ❖ tumeur localisée, lorsque le cancer est confiné à la prostate ;
- ❖ cancer localement avancé, lorsque la tumeur a commencé à toucher la capsule prostatique et les organes adjacents ;
- ❖ tumeur avec atteinte ganglionnaire, lorsque les ganglions lymphatiques de la région de la prostate sont atteints ;
- ❖ cancer métastatique, lorsqu'il s'est disséminé à d'autres organes distants.

L'âge constitue le premier facteur de risque du cancer de la prostate mais l'hérédité, l'alimentation riche en graisses et l'exposition à certains produits, notamment les polluants chimiques joueraient aussi un rôle dans la survenue de la maladie. [32], [33]

2. La conception de médicaments ou Drug Design :

2.1. Les étapes de développement du médicament :

Le chemin parcouru pour la recherche d'une molécule médicamenteuse est souvent long, fastidieux, incertain et réglementé. En effet, le processus de recherche et de développement dans l'industrie pharmaceutique, ne réussit que pour environ une molécule sur 10 000 testées et dure en moyenne dix voire quinze ans. [34], [35]

Intéressons-nous donc au parcours du médicament depuis la potentielle molécule thérapeutique jusqu'à sa mise sur le marché (voir Figure 3)



Figure 3 : Vue d'ensemble sur la recherche et le développement d'un nouveau médicament

Celui-ci comprend quatre étapes majeures :

- 1ère étape : La recherche ;
- 2ème étape : Les études précliniques ;
- 3ème étape : Les études cliniques ;
- 4ème étape : Enregistrement et autorisation de mise sur le marché.

2.1.1. La recherche :

L'axe de recherche de l'industrie pharmaceutique est déterminé en fonction de l'avancement de la recherche fondamentale, des besoins médicaux exprimés et des stratégies internes de l'entreprise selon Les Entreprises du médicament (Leem) [36], [37]. Cette étape comprend :

- ✓ L'identification et la validation de la cible ;
- ✓ L'Identification de principes actifs ou identification du « hit » ;
- ✓ Optimisation moléculaire ou identification du « lead ».

2.1.2. Les études précliniques :

L'objectif de ce stade est d'affiner encore la sélection des molécules à tester sur l'Homme. Ces études consistent à évaluer en laboratoire la molécule qui doit répondre à un besoin thérapeutique pour une ou plusieurs pathologies et qui doit avoir des effets secondaires minimales. Pour trouver ces molécules, les chercheurs peuvent mieux utiliser des outils de modélisation informatique (modélisation moléculaire, criblage virtuel) que des tests expérimentaux (criblage dans le monde réel, tests in vivo et in vitro) pour garantir la sécurité,

la sélectivité et l'efficacité des molécules synthétiques et dont le but est d'identifier le profil pharmacologique, pharmacocinétique et toxicologique du candidat médicament.

Les tests *in vivo* sont effectués sur trois espèces animales distinctes avant de passer au stade clinique. Cependant, des méthodes alternatives ont vu le jour (tests *in vitro*) permettant désormais de sauver des vies animales. [38]

2.1.3. Les études cliniques :

Ils sont strictement encadrés par la loi et ne peuvent débuter qu'après autorisation par les autorités de santé telles que L'Agence européenne des médicaments (AEM).

Après avoir passé les tests précliniques, une série d'évaluations suivra dans le but de prouver le caractère inoffensif et l'efficacité des futurs médicaments chez l'homme. [39]

a. Phase I «phase d'innocuité» :

Les tests sont effectués sur des volontaires sains, sauf pour des molécules anticancéreuses et antibiotiques. Son intérêt se porte sur les propriétés pharmacocinétiques ADME de la molécule.

b. Phase II « tests de biodisponibilité ou d'efficacité » :

Réalisée sur des groupes des volontaires malades, cette phase permet de déterminer la posologie du traitement et également évaluer les premiers effets secondaires. Si ces derniers sont trop importants, en comparaison de l'effet thérapeutique du médicament, le développement du traitement est définitivement suspendu.

c. Phase III « double aveugle » :

Le but de cette étape est d'établir le rapport bénéfices/risques. Au cours de cette phase, un grand nombre de patients est soumis, soit à l'effet de la molécule contenant le principe actif, soit à un placebo. Les résultats sont ensuite comparés avec ceux de traitements déjà existants, ce qui devrait confirmer l'efficacité du médicament.

d. Phase IV « post marketing » ou de « pharmacovigilance » :

Il s'agit d'essais cliniques après commercialisation de la substance thérapeutique. En résumé, cette dernière phase consiste à relever les éventuels effets secondaires apparaissant plus tard dans la prise du médicament.

2.1.4. Enregistrement et Autorisation de Mise sur le Marché :

A l'issue de l'essai clinique, l'entreprise pharmaceutique responsable du développement dépose un dossier CTD ou "Common Technical Document" contenant l'ensemble des données obtenues au cours du processus de développement (rapports d'études in vitro et in vivo, propriétés chimiques de la molécule, formulation galénique et processus de production, pharmacocinétique et toxicité in vitro et in vivo, résultats cliniques, sécurité, etc.). Qui sera examiné par l'autorité sanitaire du pays dans lequel il a été conçu. Si l'avis est favorable, le médicament recevra une AMM.

2.2. Les approches de la conception de médicaments :

Deux types d'approches de drug design sont connus. La première fait référence au ligand et se nomme « ligand- based drug design » et la deuxième se réfère à la cible moléculaire et se nomme « structure- based drug design ». [40]

2.2.1. Approche « Ligand-based » :

Cette approche s'appuie sur les connaissances acquises sur d'autres molécules qui interagissent avec la cible d'intérêt biologique et qui définit les caractéristiques minimales qu'une molécule doit posséder pour interagir avec le système biologique d'intérêt. Ici, un bras robotique détermine des relations quantitatives structure-activité (QSAR), ces relations peuvent être utilisées pour prédire de nouveaux analogues [41]. Pour générer un pharmacophore, plusieurs étapes sont nécessaires : [42]–[45]

- ✓ Recherche et identification des ligands de référence.
- ✓ La recherche conformationnelle.
- ✓ Identification des points pharmacophoriques de chaque ligand de référence.
- ✓ Détermination et Evaluation des modèles pharmacophoriques.

2.2.2. Approche « Structure-based » :

Elle repose sur les connaissances de la structure tridimensionnelle obtenues par des méthodes expérimentales telles que la cristallographie par rayons X et la spectroscopie par Résonance Magnétique Nucléaire (RMN) de la cible biologique si sa structure est connue. Par contre, pour toutes les structures non déterminées expérimentalement il est possible d'obtenir un modèle, construit avec la modélisation par homologie, basé sur la structure expérimentale d'une protéine proche. Tous ces développements ont eu un impact positif sur l'évolution de la méthode "structure-based" [42], [46], [47]. Par conséquent, cette méthode se divise en deux types :

a. Approche basée sur le récepteur :

La première catégorie consiste à « trouver » un ligand pour une cible ou un récepteur donné dans une base de données utilisées pour la recherche. La méthode implique le criblage d'un grand nombre de ligands pour extraire les ligands qui s'insèrent dans la poche de liaison au récepteur. L'avantage de cette méthode consiste en l'économie d'efforts de synthèse dans la recherche de nouveaux « composés de tête ». [42]

b. Approche basée sur le complexe ligand-récepteur :

La deuxième catégorie concerne « la construction » du ligand, généralement appelée «receptor-based ». En assemblant de petites fractions de molécules de manière progressive, les ligands sont construits sans tenir compte des contraintes structurelles dues aux poches de liaison. Les différentes parties en question peuvent être des fragments moléculaires ou des atomes. L'avantage de cette méthode (qui connaît de plus en plus d'engouement) est la construction de molécules originales qui ne sont pas dans les bases de données. [48], [49]

3. Multi-Drug résistance dans le cancer :

La chimiothérapie est l'un des outils les plus efficaces de la médecine moderne permettant le traitement de nombreuses infections microbiennes ainsi qu'un des principaux moyens de contrôle des cancers. [50], [51]

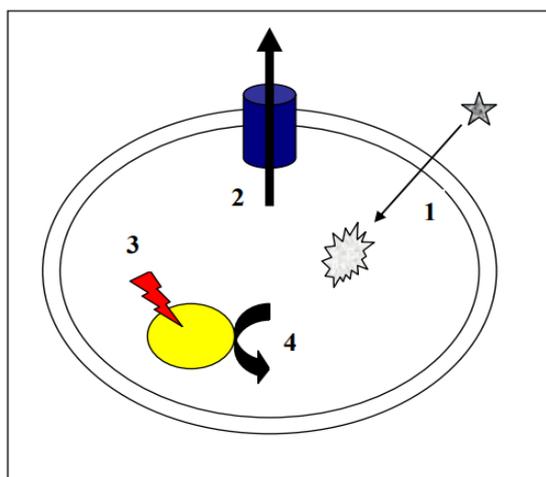
La résistance aux médicaments est l'obstacle majeur au traitement de nombreux cancers. Les causes d'échecs des chimiothérapies sont multiples, la principale étant MDR (Multi Drug Resistance) intrinsèque de la cellule tumorale.

La résistance des cellules tumorales à plusieurs classes de produits dont les anthracyclines, taxanes et les épipodophyllotoxines est associée aux transporteurs ATP-binding cassette (ABC) ((la glycoprotéine P (PGP) et multidrug resistance-associated protein1 (MRP1)), qui fonctionnent en tant que pompes d'efflux de médicaments. [52], [53]

Trois changements majeurs sont visibles dans les cellules présentant le phénotype MDR:

- La réduction de l'accumulation de médicaments cytotoxiques ;
- Modifications de l'activité ou de l'expression de certaines protéines cellulaires ;
- Modifications de la physiologie cellulaire affectant le pH cytoplasmique ou le trafic membranaire intracellulaire.

Ces changements majeurs aboutissent à 4 grands mécanismes de résistance [54], représenté par la Figure 4.



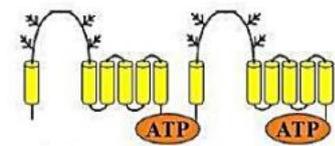
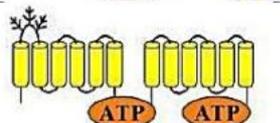
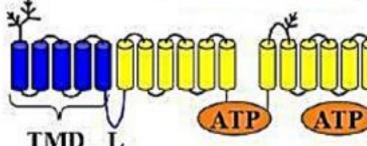
- 1- Inactivation moléculaire.
- 2- Efflux actif en dehors de la cellule (transporteurs ABC).
- 3- Altération de la cible moléculaire (ADN topoisomérases).
- 4- Anomalies de réponses (apoptose, augmentation de la réparation de l'ADN).

Figure 4 : Les quatre grands mécanismes de résistance

3.1. Les transporteurs ABC :

Découverts il y a environ 50 ans, les protéines ABC constituent la plus grande famille de protéines, dont l'identification est basée sur la présence de la signature ABC. Elles sont exprimées de manière ubiquitaire et sont impliquées dans des activités essentielles pour la survie cellulaire. Ces transporteurs actifs primaires utilisent l'ATP pour assurer l'échange et/ou la communication de la cellule avec l'environnement. Ils participent à de nombreux processus biologiques tels que l'homéostasie des lipides, la présentation des antigènes, la transduction des signaux, la détoxification et l'absorption des nutriments. Leurs substrats appartiennent à un large spectre de molécules inorganiques et organiques de tailles multiples. Ils transportent des sucres, des acides aminés, peptides, nucléosides, vitamines, métaux et médicaments. D'un point de vue négatif, plusieurs de ces protéines membranaires biologiquement pertinentes contribuent à la résistance aux médicaments chez l'homme, les champignons et les bactéries. Des exemples de ces transporteurs est ABCB1 [55], ABCG2 [56] et MRP1 [57] qui sont responsables de la chimiorésistance chez l'homme. [50], [58] (voir tableau 1)

Tableau 1 : Classification des transporteurs ABC humains [59]

<p>Subfamily ABCA</p>		<p>12 transporteurs complets appartiennent à cette sous-famille. Les exemples sont ABCA4 (transporteur de rétinol ABCR) et ABCA1 (transporteur du cholestérol).</p>
<p>ABCB</p>		<p>Elle est la seule sous-famille contenant des transporteurs complets (4) et des demi-transporteurs (7). Un exemple est ABCB1 (transporteur de médicaments P-glycoprotéine).</p>
<p>ABCC</p>		<p>13 transporteurs complets appartiennent à cette sous-famille. Les exemples sont ABCC1 (transporteur de médicaments MPRP1) et ABCC7 (canal Cl- CFTR)</p>
<p>ABCD</p>		<p>La moitié des transporteurs font partie de cette sous-famille. L'un d'entre eux est ABCD3 (PMP70 peroxisomal transporteur)</p>
<p>ABCE/F</p>		<p>Ces sous-familles sont composées uniquement par le NBD. La sous-famille E est composée par 1 transporteur et la sous-famille F compte 3 protéines. Un exemple est ABCF1</p>
<p>ABCG</p>		<p>5 demi-transporteurs appartiennent à cette sous-famille. La particularité de ce groupe est qu'ils sont appelés inversés puisque le côté N-terminal correspond au la NBD et le C-terminal à la TMD. Un exemple est ABCG2</p>

3.1.1. La protéine de résistance à de multiples drogues : MRP1

Le transporteur MRP1 a été découvert en 1992 à la suite de recherche de nouvelles protéines ayant la capacité de conférer le phénotype MDR. Ainsi, à la fin des années 80, de nombreuses lignées cellulaires sélectionnées par différentes drogues avaient les caractéristiques classiques du phénotype MDR, mais ceci sans augmenter l'expression de la PGP. Une de ces lignées, H69AR, dérivée de la lignée cellulaire de cancer à petites cellules H69, par sélection à la doxorubicine, était alors résistante à une grande variété de drogues naturelles telles que les anthracyclines, les Vinca alcaloïdes et les épipodophyllotoxines. Une analyse différentielle des ARNm de la lignée H69AR en comparaison avec la lignée H69 a montré l'expression d'un nouveau transporteur ABC. Cette protéine, initialement appelée multidrug resistance-associated protein ou MRP puis MRP1, est alors devenu le membre fondateur d'une grande sous-famille de la superfamille des transporteurs ABC, les MRPs (ABCC). [57], [60], [61]

3.1.2. Topologie du transporteur MRP1 :

Un certain nombre de topologies a été proposé pour MRP1, dont la plupart prédisent trois domaines transmembranaires (TMD) (TMD₀, TMD₁, TMD₂) contenant 17 hélices transmembranaires (TM) répartis entre ces trois TMD : TM 1–5 (TMD₀), TM 6–11 (TMD₁) et TM 12–17 (TMD₂) et deux domaines de fixation des nucléotides (NBD₁, NBD₂) ; Ce modèle est cohérent avec presque toutes les données expérimentales disponibles (Figure 5). [62]–[65]

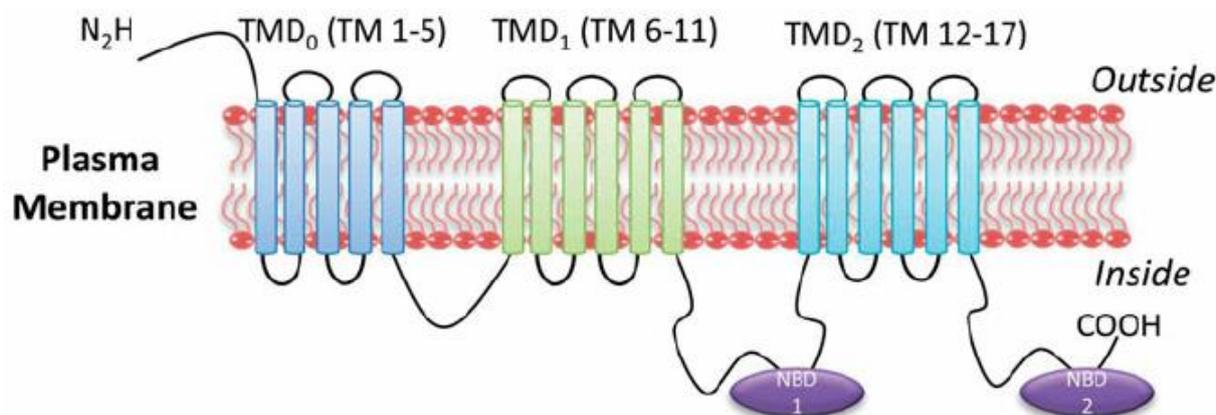


Figure 5 : L'orientation et la topologie de MRP1

3.1.3. Localisation de MRP1 :

La distribution tissulaire et subcellulaire des différents membres de la sous-famille des MRP est très variable. Dans le cas de MRP1, ce transporteur est largement exprimé par les poumons, les testicules, les reins, les muscles squelettiques et cardiaques, mais il est faiblement détectable dans le foie humain adulte, alors que dans les hépatocytes en prolifération et dans les lignées cancéreuses hépatiques, telles que HepG2, l'expression de ce transporteur est considérablement augmentée. [66], [67]

Bien que la localisation de MRP1 soit distribuée dans tout l'organisme, on a retrouvé ce transporteur dans certaines lignées cellulaires bien spécifiques tel que les cellules épithéliales bronchiques, les pneumocytes hyperplasiques de type II dans le poumon, les mastocytes et les érythrocytes dans le système circulatoire. MRP1 est aussi exprimée dans le cerveau dans un type cellulaire spécifique, comme dans la barrière hématoencéphalique et le plexus choroïde de la barrière cérébrospinale. [68]–[72]

3.1.4. Fonction physiologique de MRP1 :

La diversité des substrats et la distribution étendue de MRP1 suggèrent la possibilité de nombreuses fonctions physiologiques dont beaucoup restent à confirmer. Son substrat physiologique préférentiel est le LTC₄. Sa capacité à transporter le GSH et le GSSG indique la possibilité que le transporteur pourrait contribuer à maintenir le potentiel redox de la cellule. [73]

➤ **Transport de molécules organiques par MRP1 :**

Le MRP1 Transporte un spectre de substrats similaire à celui attribué à un ou plusieurs transporteurs connus sous le nom de transporteur d'anions organiques multi-spécifique (MOAT). La plupart de ces substrats d'anions organiques sont des molécules conjuguées au glutathion (GSH), au glucuronide ou au sulfate conjuguées. [74]

➤ **Phénotype de résistance aux médicaments médiée par MRP1 :**

Depuis sa découverte dans le H69AR très résistante lignée cellulaire du cancer du poumon, MRP1 s'est avéré être surexprimé dans des lignées cellulaires multirésistantes dérivées de nombreux types de tissus et de tumeurs, y compris les cancers du poumon, les carcinomes du côlon, du sein, de la prostate et diverses formes de leucémie. [62], [75], [76]

La surexpression de MRP1 dans certaines lignées cellulaires est également associée à une modification de la distribution intracellulaire altérée des médicaments intrinsèquement actifs souvent avec l'accumulation de drogue dans des vésicules cytoplasmiques. [62], [77]–[80]

➤ **Rôle du GSH dans la résistance aux médicaments médiée par MRP1 et le stress oxydatif :**

Étant donné que MRP1 ne transporte apparemment pas directement les produits naturels chimiothérapeutiques non modifiés mais peut transporter les produits anioniques de biotransformation de phase II conjugués au GSH ou à d'autres petites molécules in vitro, on a émis l'hypothèse que MRP1 confère une résistance en éliminant les conjugués de médicaments des cellules. Bien qu'il semble peu probable que l'élimination accrue des métabolites de médicaments (qui sont généralement moins toxiques) par MRP1 ait des effets chimioprotecteurs importants, l'élimination efficace de ces produits de biotransformation peut empêcher l'inhibition des enzymes de phase II nécessaires à la conjugaison des xénobiotiques.[81], [82]

➤ **Transport de drogues non modifiées et GSH :**

L'incapacité de MRP1 à transporter directement les agents chimiothérapeutiques à base de produits naturels est en accord avec le fait que ces xénobiotiques sont de mauvais Inhibiteurs du transport concernons d'autres substrats médié par MRP1. Même à des concentrations extrêmement élevées de médicaments une faible inhibition du transport de LTC₄ a été observée.

Cependant, la puissance inhibitrice de certains de ces médicaments était nettement accrue en présence d'un taux réduit de GSH. [83], [84]

- Analyses structure-fonction de MRP1 ;
- Activité ATPase du MRP1.

3.1.5. Substrats transportés par MRP1 et spécificité :

La plupart des études se sont focalisés sur les substrats xénobiotiques en raison du rôle potentiel de MRP1 dans le phénomène de chimiorésistance et dans la protection de l'organisme contre les agents toxiques environnementaux. D'autres études cherchent à identifier des substrats endogènes dans l'objectif de mieux comprendre les fonctions physiologiques de la protéine (tableau 2). [85]

Tableau 2 : Substrats de MRP1

Composés	Exemples
<i>Médicaments /xénobiotiques</i>	
Antinéoplasiques	- Antimétabolites (methotrexate, edatrexate, ZD1694) - Anthracyclines (doxorubicine, épirubicine) - Alkaloïdes (étoposide, vincristine, vinblastine, paclitaxel, irinotecan)
Antiviraux	Saquinavir, ritonavir
Antibiotiques	Difloxacine, grepafloxacine
Metalloïdes	Arsénite, arséniate, antimoine
Sondes fluorescentes	Calcéïne, Fluo-3, BCECF, SNARF
Toxines	Aflatoxine B ₁ , chloropropham, fenitrothion, methoxychlor
<i>Médicaments/xénobiotiques conjugués</i>	
Conjugués au GSH	2,4-Dinitrophényl-SG, bimine-SG, N-éthylmaléimide-SG, doxorubicine-SG, thiopta-SG, cyclophosphamide-SG, melphalan-SG, chlorambucil-SG, acide ethacrynique-SG, metolachlor-SG, atrazine-SG, sulfuraphane-SG, aflatoxine B ₁ -epoxyde-SG, 4-nitroquinoline 1-oxyle-SG, As(SG) ₃
Conjugués à des glucuronides	Etoposide-Gluc, 4-(méthylnitrosamino)-1-(3-pyridyl)-1-butanol (NNAL)-3β-O-Gluc, SN-38-Gluc, 4-méthylumbelliferyl-β-D-Gluc, 6-hydroxy-5,7-diméthyl-2-méthylamino-4-(3-pyridylméthyl)benzothiazole sulphate (E3040S)-Gluc
<i>Composés naturels</i>	
Conjugués au GSH	Leucotriène C ₄ , prostaglandine A ₂ -SG, 15-deoxy-Δ ^{12,14} prostaglandine J ₂ -SG, hydroxynonenal-SG
Conjugués à des glucuronides	17β-Estradiol-17-β-D-Gluc, glucuronosylbilirubine, bis-glucuronosylbilirubine, hyodesoxycholate-6-α-Gluc
Conjugués à des sulfates	Estrone-3-sulfate, deshydroepiandrosterone sulfate, sulfatolithocholate
Folates	Acide folique, L-leucovorine
Peptides	GSH, GSSG
Autres	Bilirubine

3.1.6. Cycle du transport de MRP1 :

Le cycle de transport le plus accepté de MRP1 peut être découpé en six étapes différentes [86]–[91] (Figure 6):

1→ D'abord la Fixation du substrat (aussi du GSH si le transport du substrat est dépendant de ce dernier) avec un site de haute affinité sur MRP1.

2→ Cette interaction, à son tour, induit la fixation de l'ATP sur NBD1 stabilise l'interaction entre les NBD en établissant des contacts avec le motif Walker C de NBD2, facilitant ainsi la prochaine étape.

3→ La fixation de la seconde molécule d'ATP complète la formation du dimère de NBD et cause des changements conformationnels dans NBD2. Cette "fermeture" des deux domaines NBD est la clé du mouvement unidirectionnel des substrats à travers les membranes.

4→ La protéine maintient un état de faible affinité après hydrolyse de l'ATP sur NBD2, tant que NBD1 est occupé par l'ATP et que l'ADP n'a pas été relégué du NBD2. Cependant, bien que les étapes dans le modèle menant à la formation d'un état de faible affinité pour le substrat soient très bien décrites, le retour à l'état initial de la protéine reste à déterminer.

5→ Si le NBD1 n'a pas d'activité ATPasique, le retour de la protéine à l'état initial se fait après relargage de l'ADP, NBD2 et de l'ATP de NBD1.

6→ C'est possible que l'hydrolyse de l'ATP soit requise par NBD1.

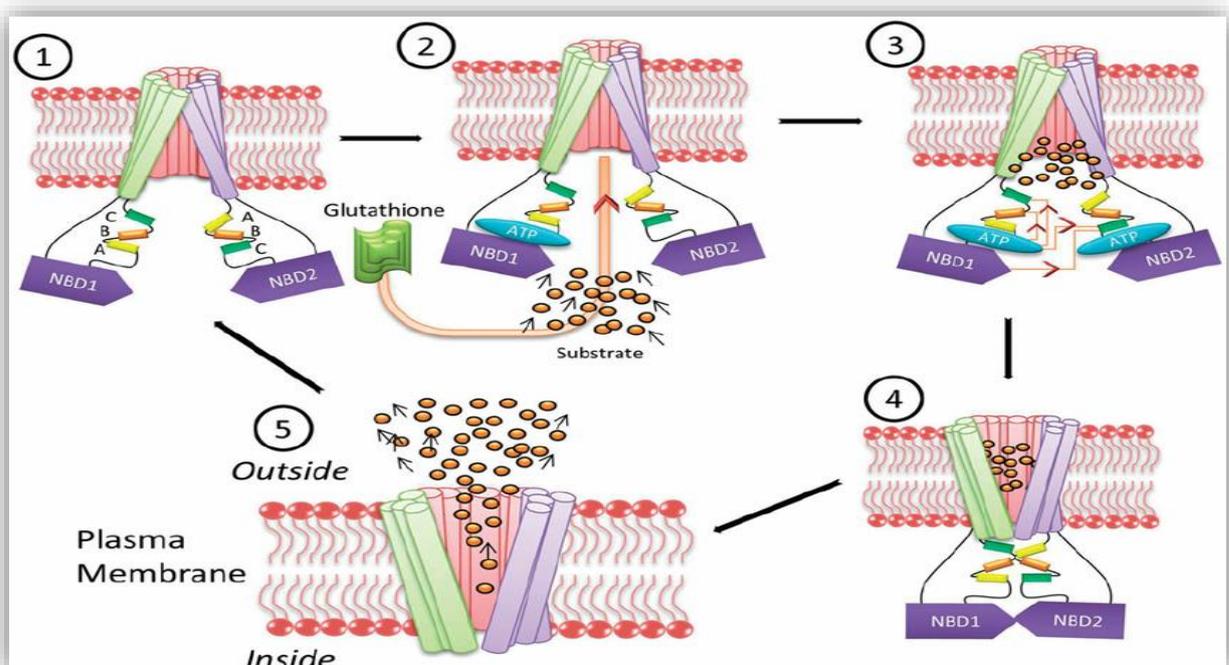


Figure 6 : Modèle hypothétique du cycle du transport par MRP1 [92]

3.1.7. Implication clinique de MRP1 dans le cancer :

De considérables essais sur des animaux et des multiples études in vitro ont montré la capacité des transporteurs ABC dans l'augmentation de la résistance à de nombreux agents chimiothérapeutiques. L'étude de la contribution de MRP1 à la résistance clinique des cancers n'en est aujourd'hui qu'à ses premiers stades. [93]–[96]

a. Cancer du poumon :

Dans le NSCLC, MRP1 s'est avéré être systématiquement surexprimé dans une large population de patients avant le traitement, particulièrement au niveau des adénocarcinomes. Bien que la surexpression de MRP1 ait rendu les tumeurs résistantes à la chimiothérapie, on a observé que ces tumeurs étaient moins agressives et plus différenciées, en plus il a été constaté que l'expression de MRP1 semble être prévalente dans les stades précoces de la maladie, mais qu'elle diminue avec sa progression. D'autres études ont conclu que des niveaux élevés de MRP1 sont un indicateur négatif, impliquant une faible réponse aux traitements chimiothérapeutiques. [97], [98]

b. Cancer de la prostate :

La corrélation clinique de hauts niveaux de MRP1 et la mutation P53 sont prouvées par des études in vitro de la régulation du gène MRP1 qui montrent que la protéine P53 est un suppresseur très puissant de la transcription de ce dernier. Le niveau de MRP1 est augmenté avec l'avancée du cancer et durant le processus invasif, et il est positivement associé avec le statut de la mutation P53 des tumeurs, une association également observée dans les cancers NSCLC. [99], [100]

c. Neuroblastome :

Dans le neuroblastome, il y a eu beaucoup de preuves convaincantes indiquant une forte association de MRP1 avec l'amplification de NMYC. Une étude a fourni la preuve que MRP1 est un indicateur pronostique fort dans le neuroblastome ; Il est intéressant de noter que les mêmes résultats n'ont pas été trouvés avec P-gp/ABCB1. En effet, l'induction de MYCN dans une lignée cellulaire de neuroblastome a également provoqué une augmentation de l'ARNm ABCC1 ainsi que de la protéine MRP1. Cela renforce la coordination de la régulation entre l'expression de MYCN et de MRP1, fournissant un lien mécaniste entre l'agressivité tumorale

et la protéine MRP1. Dans l'ensemble, MRP1 pourrait être le principal mécanisme MDR dans le neuroblastome, et sa surexpression à n'importe quel stade conduirait à un mauvais résultat clinique. [101]–[104]

d. Cancer du sein :

Beaucoup d'études indépendantes montrent que l'expression de MRP1 est un pronostic négatif pour plusieurs types de cancers du sein. Trois d'entre elles ont conclu que la présence de MRP1 à un stade avancé du cancer du sein était associée à une rechute rapide après chimiothérapie. dans ce cancer, le MRP1 peut être un fort marqueur de pronostic pour l'OS et la SFR (relapse-free survival) lorsqu'il est exprimé de manière inhérente à des niveaux élevés avant la chimiothérapie. L'analyse de l'expression de MRP1 dans des nodules lymphatiques du cancers du sein indique que le niveau de ce dernier était plus élevé dans les métastases que dans la tumeur primaire, alors que l'inverse se produit pour la Glycoprotéine-P. [105]–[107]



*Chapitre02 : La méthodologie
QSAR*

1. Chimio-Bio-informatique :

Une discipline scientifique qui a beaucoup évolué durant ces 40 dernières années concernant le développement, la création, l'organisation, le stockage, la diffusion, l'analyse, la visualisation et l'utilisation de l'information chimique et biologique, représente un domaine en évolution rapide.

Actuellement, les méthodes spécifiques de la Chimio-Bio-informatique sont devenues indispensables pour le développement de nouveaux composés, matériaux et processus. D'un autre point de vue elle représente un lien entre la chimie, biologie et l'informatique. [108]–[110]

2. La représentation moléculaire :

2.1. Théorie des graphes :

La théorie des graphes, défini et établie au XVIIIe par Deux scientifiques, Alexander C. Brown et James J. Sylvester, qui ont développé la représentation moléculaire sous forme de nœuds (atomes, indiqués par leur nom) et de liaisons Où les électrons sont délocalisés. Aujourd'hui, les atomes d'hydrogène sont implicitement représentés dans le graphe puisqu'ils sont supposés remplir les valences inutilisées. [111], [112]

Une représentation commune du graphe est la matrice d'adjacence, un carrée avec une dimension N égale au nombre d'atomes. Chaque position (i, j) dans la matrice spécifie l'absence (0 valeur) ou la présence d'une liaison reliant les atomes i et j, remplie de 1, 2, ou 3 pour indiquer une liaison simple, double ou triple, 4 pour une liaison amide, et 5 pour la liaison aromatique. Les éléments en diagonale sont toujours zéro [113]. Un exemple de représentation d'une matrice est dans la figure 7 :

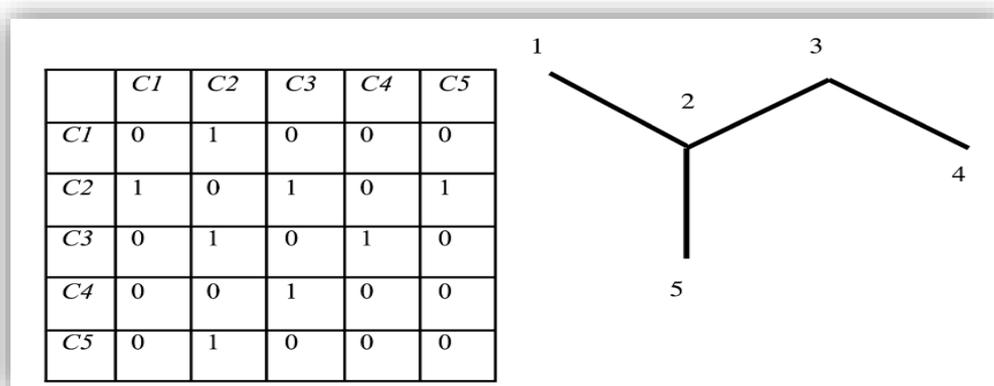


Figure 7: Matrice d'adjacence du 2-méthylbutane représenté après élimination de l'hydrogène

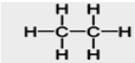
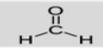
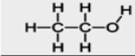
Le système SMILES (Simplified Molecular Input Line Entry System) est un système de notation chimique conçu pour le traitement moderne de l'information chimique. Il est basé sur les principes de la théorie des graphes moléculaires. SMILES est une notation linéaire qui représente les structures chimiques dans une définition basée sur un graphique, où les atomes, les liaisons et les cycles sont codés dans un graphique et représentés dans des séquences de texte. [114], [115]

2.2. La notation SMILES :

Une molécule peut avoir plusieurs représentations SMILE correspondantes, puisqu'une molécule peut être codée en commençant n'importe où ; dans le tableau 3, quatre chaînes SMILES, toutes correctes, représentent l'éthanol. Pour résoudre ce problème et fournir une correspondance unique entre les SMILES et les molécules, de multiples algorithmes de canonisation sont inventés pour garantir l'unicité de la représentation des molécules. [116], [117]

Les développements récents en matière de notations chimiques sont les codes InChI (International Chemical Identifier), soutenus par l'Union internationale de chimie pure et appliquée (IUPAC). peut décrire de manière unique une molécule, à différents niveaux de détail, mais n'est pas destiné à être lu par l'homme. [118]

Tableau 3 : Exemples de molécules avec leur code SMILES

SMILES	Nom	Formule	Graphe
CC	Ethane	CH_3CH_3	
C=O O=C	Forme aldéhyde	CH_2O	
CCO OCC C(C)O C(O)C	Ethanol	CH_3CH_2OH	

3. La biologie des produits chimiques et toxicologie :

Depuis le XIXe siècle, la pratique de l'expérimentation animale s'est imposée en physiologie, en microbiologie et en chirurgie. L'explosion de la biologie moléculaire dans la seconde moitié du vingtième siècle a accru l'importance des modèles in vivo. [119]

Tous les modèles ont leurs limites, leur prédiction peut être mauvaise, et leur transférabilité aux phénomènes réels qu'ils modélisent peut être insatisfaisante. Ainsi l'extrapolation de données issues des modèles animaux à l'environnement ou à la santé humaine dépend de la mesure dans laquelle ce modèle est un reflet approprié de la condition étudiée. [120]

Ces limites sont toutefois intrinsèques à toute modélisation approches. La plupart des questions sur les modèles animaux sont éthiques plus que scientifique. En santé publique, l'utilisation de modèles animaux est imposée par la réglementation et peu probable que toute autorité sanitaire autorise les nouveaux médicaments sans tester sur les animaux. [121]

4. Essais biologiques de toxicité :

La toxicité est le degré auquel une substance peut endommager un organisme; une propriété préoccupante pour toute substance chimique.

Paracelse Theophrastus Phillipus von Hohenheim (1493-1541) a écrit : " Tout est poison, rien n'est poison : c'est la dose qui fait le poison". [122]

La relation entre la dose et ses effets sur l'organisme exposé est d'une grande importance en matière de toxicologie [123], [124]. Le processus d'utilisation des animaux pour évaluer la toxicité des produits chimiques a été défini comme suit :

- La toxicité peut être mesurée par ses effets sur la cible ;
- Comme les individus réagissent différemment à une même dose de toxine, une mesure de la toxicité au niveau de la population est souvent utilisé qui relie les probabilités d'un résultat pour une donnée individu dans une population. L'exemple est DL_{50} : la dose qui cause la mort de 50% de la population ;
- Lorsque la dose est distincte, définir les « facteurs de sûreté ».par exemple, si une dose est sans danger pour un rat, on peut supposer qu'un faible pourcentage de dose serait sans danger pour un être humain ce qui permet d'utiliser un facteur de sécurité 100, par exemple.

La mesure continue de l'activité biologique la plus courante est le logarithme (IC_{50}) (concentration inhibitrice), qui mesure la concentration d'un composé particulier nécessaire pour induire une inhibition de 50 % de l'activité biologique étudiée. De même, la dose létale médiane, DL_{50} , est la dose nécessaire pour la mort de la moitié des membres d'une population testée après une durée de test spécifiée.c'est pas la dose létale pour tous les sujets, mais seulement pour la moitié d'entre eux. [125], [126]

La relation dose-réponse décrit le changement de l'effet sur un organisme causé par des niveaux différents de doses d'un produit chimique après un certain temps d'exposition . Une courbe dose-réponse est un graphique x-y reliant la dose à la réponse de l'organisme. [127], [128]

- La dose mesurée est représentée sur l'axe X et la réponse sur l'axe Y. Dans ce cas, la courbe est typiquement sigmoïdale, avec la partie la plus raide au milieu (la figure 8) montre un exemple de dose –courbe de réponse pour DL₅₀.
- La réponse peut être une réponse physiologique ou biochimique; La DL₅₀ est utilisée en toxicologie humaine .la concentration d'inhibition IC₅₀ et sa double concentration d'effet EC₅₀ sont utilisées en pharmacologie.

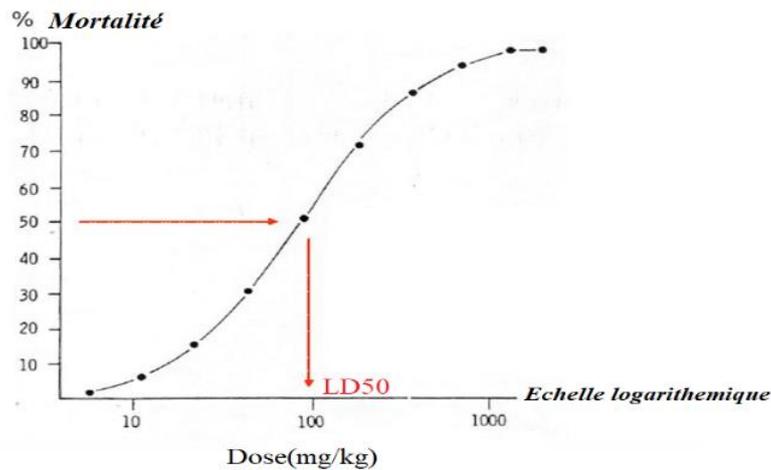


Figure 8 : Exemple de courbe dose-réponse pour la DL₅₀

Aujourd'hui, des tests in vitro sont également disponibles. Il s'agit de l'analyse scientifique des effets d'un produit chimique sur des bactéries ou des cellules de mammifères en culture. Les expériences utilisant des systèmes in vitro sont utiles dans les premières phases d'études médicales où il est nécessaire de sélectionner un grand nombre de candidats thérapeutiques potentiels, ou pour effectuer des tests rapides. Les systèmes in vitro sont non physiologiques et présentent d'importantes limitations, car leurs résultats ne sont pas en corrélation avec ceux obtenus in vivo. Toutefois, l'utilisation de systèmes in vitro présente des avantages substantiels pour améliorer la compréhension mécaniste des activités des substances toxiques et l'utilisation de cellules humaines pour définir les effets toxiques propres à l'humain. [129]–[131]

Les méthodes in vivo et in vitro utilisent toutes deux la substance chimique ; Dans le cas des étapes initiales de la conception de nouveaux médicaments ou de nouveaux matériaux, la substance chimique est seulement conçue, mais pas encore produite. D'autres méthodes qui ne nécessitent que les structures chimiques sont nécessaires, et elles sont collectivement nommées in silico. [132]

5. Méthodes *in silico* méthodes alternatives :

L'expérimentation animale se réfère à l'utilisation d'animaux dans les expériences. Dans le monde entier, on estime que le nombre d'animaux utilisés annuellement pour des expériences est de l'ordre de dizaines de millions. Dans la toxicité, les tests animaux sont appelés modèles *in vivo* ; ils donnent des doses pour certaines espèces pour extrapoler les données à la santé humaine ou à l'environnement. Des fois l'extrapolation des données d'une espèce à l'autre n'est pas évidente. Par exemple, les doses létales pour les rats et les souris sont parfois très différentes. [133]

Comment construire un modèle qui relie une structure chimique à un effet ? Ce sujet a été étudié avant même l'existence des ordinateurs. Le terme "*in silico*" recouvre aujourd'hui les méthodes consacrées à cette fin et fait référence au fait que l'on utilise des ordinateurs. [134]

Les méthodes *in silico* les plus connues sont les méthodes de QSAR basées sur le principe que la structure moléculaire est responsable de toutes les activités et visent à trouver la dose ou la sous-structure responsable de l'activité. Elle est donc basée sur le concept de similarité (des molécules similaires devraient avoir des effets similaires).

La lecture transversale est une autre méthode *in silico*, qui consiste à trouver la molécule la plus similaire et à adapter sa valeur expérimentale à la molécule étudiée [135]–[138]. (La figure 9) illustre les méthodes *in vivo*, *in vitro* et *in silico* dans son ensemble.

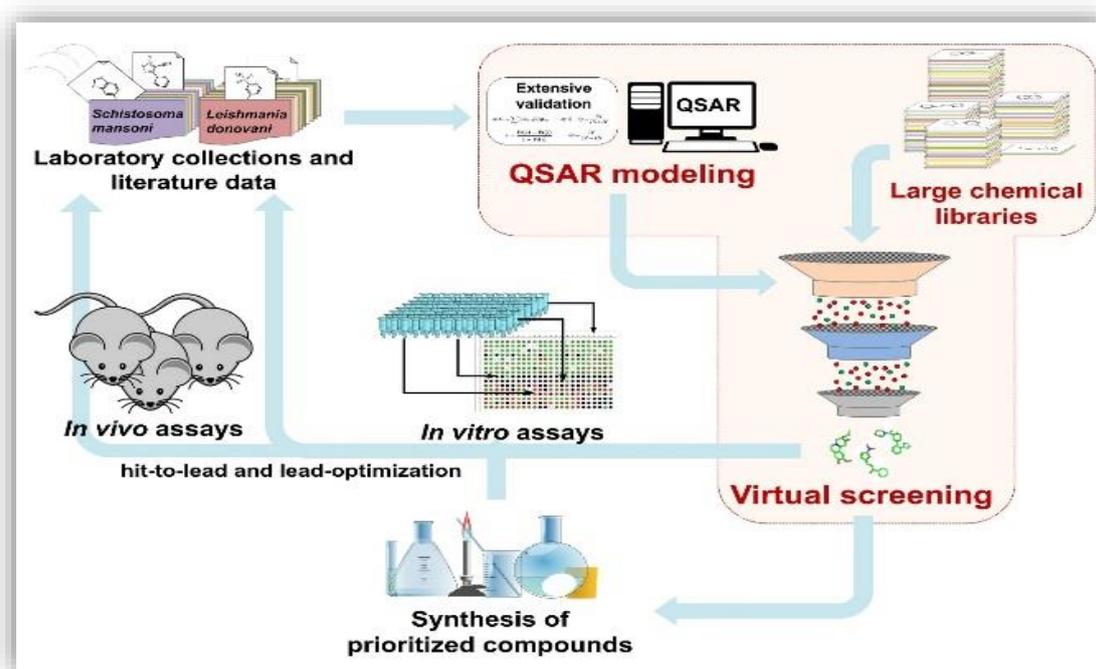


Figure 9 : Les méthodes disponibles pour tester les molécules

5.1. QSAR :

QSAR vise à Corréler une variable de réponse particulière avec des descripteurs moléculaires qui ont été calculés ou même mesurés à partir des molécules elles-mêmes. Ces méthodes ont d'abord été mises au point par Corwin Hansch dans les années 40, en analysant des séries de composés et formulant l'équation QSAR [139], [140] :

$$\text{Log } 1/C = a p + b s + c E_s + \text{const} \quad (\text{eq 1})$$

- [C] : concentration de l'effet.
- [p] : coefficient de partage octanol-eau.
- [s] : constante de substitution de Hammett (électronique).
- [Es] : constante de substituant de Taft.

Log P s'agit d'une mesure de la différence de solubilité du composé dans ces deux solvants. Normalement, l'un des solvants est l'eau tandis que le second est hydrophobique comme l'octanol. Avec un coefficient élevé de partage octanol/eau, la substance chimique est hydrophobe et préférentiellement répartis dans des compartiments hydrophobes tels que la membrane cellulaire, tandis que l'hydrophilie se trouve dans des compartiments hydrophiles tels que sérum sanguin. Les valeurs LogP aujourd'hui sont souvent prédites. [141][142]

Les définitions de la structure chimique et de la fonction restent un défi aujourd'hui, mais la relation entre la structure et la propriété est largement adoptée dans la découverte de médicaments et dans l'évaluation des risques. [143]

Parfois, les méthodes QSAR prennent un nom plus spécifique comme QSPR (relation quantitative structure-propriété) lorsqu'elles sont utilisées pour des propriétés physicochimiques, (le point d'ébullition, la solubilité, le logP...). Ils sont tous corrélés à une variable dépendante (l'effet ou la réponse) avec un ensemble de variables indépendantes (généralement des propriétés calculées, ou descripteurs) Ils sont des modèles statistiques et peuvent être appliqués à prédire les réponses pour les points de données invisibles sans avoir besoin d'utiliser le produit chimique réel, même sans qu'il soit synthétisé. [144][145]

5.1.1. Descripteurs Moléculaire :

Afin d'exploiter au maximum les informations contenues dans les structures moléculaires, celles-ci sont converties en une série de grandeurs qui représentent des

caractéristiques saillantes physico-chimiques, structurelles, locales ou globales de la molécule. Ces grandeurs sont appelés descripteurs. [146]–[148]

5.1.2. Les types de descripteurs moléculaires :

Il existe aujourd'hui plus de 10 000 descripteurs moléculaires. Ils permettent en effet d'effectuer des prédictions sans avoir à synthétiser les molécules, ce qui est un des objectifs de la modélisation. Dans ce qui suit, nous allons présenter les descripteurs moléculaires les plus utilisés, en commençant par les descripteurs les plus simples.

Historiquement, deux grands schémas pour la classification des descripteurs moléculaires ont été établis: [147]

- l'un en fonction de leur origine (constitutionnel, topologique, géométrique, quantique, thermodynamique...);
- un autre sur leur dimensionnalité (1D, 2D, 3D ou 4D).

5.1.2.1. Les descripteurs 1-D :

Tous les descripteurs moléculaires qui peuvent être calculés à partir de la formule brute de la molécule à l'aide de la composition moléculaire appartiennent aux descripteurs 1D. c'est-à-dire les atomes qui la constituent, et ils représentent des propriétés générales telles que : les pourcentages massiques des atomes, la masse molaire, le poids moléculaire.... Ils sont souvent présentés comme des empreintes digitales (un vecteur binaire où 1 indique la présence d'une sous-structure et 0 son absence). [149]

5.1.2.2. Les descripteurs 2-D :

Ils sont obtenus à partir de la structure plane de la molécule. Elles contiennent des informations relatives à la connectivité ou à certains fragments moléculaires, mais aussi des estimations des propriétés physico-chimiques [150], [151]. Dans cette catégorie on trouve principalement :

- Les indices topologiques ;
- Les indices constitutionnels.

5.1.2.3. Les descripteurs 3-D :

Ils sont évalués à partir des positions relatives des atomes dans l'espace, et décrivent des caractéristiques plus complexes; leurs calculs nécessitent donc de connaître, le plus souvent par modélisation moléculaire empirique ou ab-initio, la géométrie 3D de la molécule. Ils

s'avèrent relativement coûteux en temps de calcul mais apportent davantage d'informations [152]–[156]. On distingue plusieurs familles importantes de descripteurs 3D :

- Descripteurs géométriques ;
- Les descripteurs électrostatiques,
- Les descripteurs physico-chimiques.

5.1.3. Construction de modèles de QSAR :

- ✓ La première étape est la préparation d'un ensemble contenant les données expérimentales disponibles sur le point final sélectionné. Celle-ci est très exigeante, car peu d'ensembles de données trouvés dans la littérature sont complets et validés. Souvent, (QSAR) présente des modèles élaborés à partir des quelques données disponibles de bonne qualité. Mais dans la plupart des applications réelles, les données doivent être trouvées à partir des sources principales (études de laboratoire) et vérifiées pour être cohérentes et vraiment comparable. [108], [157]
- ✓ La deuxième étape consiste à calculer et à sélectionner les descripteurs. Aujourd'hui, des centaines ou des milliers de descripteurs 2D sont facilement calculés ; il est souvent préférable d'en calculer un grand nombre et de sélectionner ensuite les quelques descripteurs les plus pertinents en utilisant des méthodes simples telles que les algorithmes génétique, méthode d'accumulation (ajout d'un descripteur à la fois), méthode de suppression d'une variable à la fois, ou des méthodes d'optimisation pour la sélection des variables.
- ✓ Quel que soit l'algorithme choisi pour développer des modèles prédictifs, il est important de tenir compte des statistiques sur la qualité des modèles et de s'assurer qu'une méthodologie de modélisation correcte est utilisée pour éviter les débordements. De nombreuses statistiques sur les modèles sont disponibles qui peuvent indiquer si de nouveaux points de données, à partir desquels les réponses doivent être prédites, peuvent être appliqués au modèle. [158]

Deux types de modèles d'apprentissage supervisé sont intéressants :

- Les méthodes de classification attribuant la molécule cible à deux ou plusieurs classes, le plus souvent biologiquement actives ou inactives ;
- Les méthodes de régression tentent d'utiliser des données continues, comme une variable de réponse biologique mesurée, pour corréler les molécules avec ces données afin de prédire une valeur numérique continue pour de nouvelles molécules et inconnues en utilisant la génération modélisée ;

Une autre distinction est faite entre :

- les méthodes basées sur la similarité, qui tendent à regrouper les molécules similaires ;
- les méthodes basées sur les caractéristiques, qui utilisent un ensemble de caractéristiques moléculaires (comme descripteurs) pour construire la fonction de classification.

Une pléthore d'algorithmes et de méthodes sont disponibles pour la modélisation, notamment la régression statistique (analyse discriminante linéaire, moindres carrés partiels et régression linéaire multiple) et les méthodes d'apprentissage (classificateur bayésien naïf, arbres de décision, partitionnement récursif, forêt aléatoire, réseaux neuronaux artificiels et machines à vecteurs de support). [159]–[161]

Au cours des deux dernières décennies, de nombreuses méthodes dérivées de l'intelligence artificielle sont devenues des méthodes d'élection dans le domaine du QSAR [162], car elles ne sont pas paramétriques et permettent de trouver des fonctions de modélisation sans décisions a priori.

5.1.4. Paramètres d'évaluation des modèles :

Afin d'évaluer l'importance des modèles QSAR, la validation de ces modèles reste une étape très sensible dans les études statistiques. Un modèle étant le résultat d'une analyse statistique, son interprétation et son application doivent se faire dans le cadre très précis du domaine couvert par l'analyse. [163]

Différents paramètres statistiques sont employés, pour déterminer la qualité d'un modèle tels que :

- Le coefficient de détermination multiple qui reflète le degré de corrélation entre les données de la propriété observées et prévues de l'ensemble de test.

$$R^2 = 1 - \frac{SCE}{SCT} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (\text{eq 2})$$

- Où Y_i et \hat{Y}_i sont les données de la propriété observée et prédite pour les composés de test,
- \bar{y} : Est la valeur moyenne des valeurs observées pour l'ensemble de calibration.

Les modèles avec des valeurs de R^2 supérieures à la valeur stipulée de 0,5 sont considérés comme bien prédictifs.

La vérification de la stabilité du modèle QSAR publiés est généralement explorée en utilisant :

- "La validation croisée par omission d'une observation" (LOO : cross-validation by leave-one-out) : qui consiste à recalculer le modèle sur $(n - 1)$ composés de calibrage, le modèle obtenu servira alors à estimer la valeur de la propriété du composé éliminé noté $\hat{y}(i)$. On répète le procédé pour chacun des n composés de l'ensemble de calibrage. [163], [164]

On l'utilise pour définir le coefficient de prédiction:

$$Q_{LOO}^2 = \frac{SCT - PRESS}{SCT} \quad (\text{eq 3})$$

Contrairement à R^2 qui augmente avec le nombre de paramètres du modèle, le facteur Q_{LOO}^2 affiche une courbe obtenu pour un certain nombre de descripteurs, puis décroît de façon monotone. Ce fait confère une grande importance au coefficient Q_{LOO}^2 .

Une valeur de $Q_{LOO}^2 > 0,5$ est considérée comme satisfaisante, et une valeur supérieure à 0,9 est excellente. [164]

Dans le cas où on a suffisamment de données qui n'ont pas servi dans la création du modèle ou après collecte de nouvelles données, on peut ou on doit procéder à la validation de ce dernier, c'est la validation externe.

Notée que Q_{ext}^2 , est calculée comme suit :

$$Q_{ext}^2 = 1 - \frac{\sum_{i=1}^{n_{ext}} (y_i - \hat{y}_{(i)})^2 / n_{ext}}{\sum_{i=1}^{n_{tr}} (y_i - \bar{y}_{tr})^2 / n_{tr}} \quad (\text{eq 4})$$

Pour une grande valeur de Q_{LOO}^2 , une valeur élevée de Q_{ext}^2 permet de présager d'une bonne capacité prédictive du modèle.

Depuis la dernière décennie, un grand débat a eu lieu sur l'application de la validation interne ou externe pour assurer la prédiction des modèles [165]. L'objectif moyen de la validation interne est de vérifier la stabilité des modèles, cependant, ce type de validation ne peut pas atteindre une véritable validation externe. [166]

✓ **Les conditions de Golbraikh et Tropsha** : ont montré que le pouvoir prédictif des modèles QSAR ne peut être revendiqué que si le modèle a été appliqué avec succès pour la prédiction des composés de l'ensemble de test externe (les composés qui n'ont pas été utilisés dans le développement du modèle). Pour refléter la qualité des prévisions, diverses mesures fondées sur les critères de Golbraikh et de Tropsha [167], [168] ont été proposées [163], [167], [169], comme Q_{F1}^2 , Q_{F2}^2 , Q_{F3}^2 et r_m^2 qui sont détaillées par les équations suivantes :

$$Q_{F1}^2 = 1 - \frac{\sum(Y_{\text{obs}(\text{test})} - Y_{\text{pred}(\text{test})})^2}{\sum(Y_{\text{obs}(\text{test})} - \bar{Y}_{\text{training}})^2} \quad (\text{eq 5})$$

Q_{F2}^2 Proposé par Schuurmann *et al.* [170] , donnée par:

$$Q_{F2}^2 = 1 - \frac{\sum(Y_{\text{obs}(\text{test})} - Y_{\text{pred}(\text{test})})^2}{\sum(Y_{\text{obs}(\text{test})} - \bar{Y}_{\text{test}})^2} \quad (\text{eq 6})$$

$$Q_{F3}^2 = 1 - \frac{\sum(Y_{\text{obs}(\text{test})} - Y_{\text{pred}(\text{test})})^2 / n_{\text{test}}}{\sum(Y_{\text{obs}(\text{train})} - \bar{Y}_{\text{train}})^2 / n_{\text{test}}} \quad (\text{eq 7})$$

$$r_m^2 = r^2 \cdot (1 - \sqrt{r^2 + r_0^2}) \quad (\text{eq 8})$$

Selon ces auteurs, les modèles sont considérés comme satisfaisants, si l'ensemble des conditions suivantes sont simultanément réalisées : $R^2 > 0.6$, $Q^2 > 0.5$

5.1.5. Interprétation du Modèle :

L'interprétation des modèles est un point important recommandé dans la démarche de validation des modèles pour la prédiction dans un cadre réglementaire. [171], [172]

Les chercheurs essaient de comprendre les modèles à partir des principes de base connus. Le petit nombre de descripteurs utilisés et leur rôle dans des équations linéaires simples sont généralement considérés comme nécessaires pour accepter les résultats QSAR.

Cependant, si l'objectif principal de QSAR est la prédiction, l'accent doit être mis sur la qualité du modèle, et non sur son interprétation. Il est courant de faire la distinction entre le QSAR prédictif, qui se concentre sur la précision prédictive, et le QSAR descriptif (SAR), qui se concentre sur l'interprétation. [173], [174]

Cette vision du «QSAR » et de «SAR » est en quelque sorte obsolète, comme le souligne Polishchuk [174], qui explique que le concept d'interprétation des modèles QSAR a évolué.

Livingstone [136] résume ainsi l'intelligibilité des modèles QSAR : « Le besoin de compréhensibilité dépend de l'application, puisqu'un modèle mathématique validé reliant les propriétés cibles aux propriétés chimiques peut dans certains cas constituer une estimation précise de l'activité chimique » .

SAR et QSAR sont des modèles statistiques prédictifs, ils souffrent donc de problèmes dans la théorie d'apprentissage statistique. Parmi les nombreuses fonctionnalités qui peuvent accomplir la tâche d'induire des modèles, la performance et la simplicité sont les caractéristiques des modèles les plus populaires. [175], [176]

5.1.6. Stratégie globale d'une étude QSAR :

La stratégie de développement d'un QSAR et plus générale qu'un QSPR (voir figure 10), elle s'articule autour de 6 points :

- Constituer une base de données Structure - Propriété à partir de mesures quantitatives, fiables et normalisées de la propriété cible, pour chaque composé et Sélectionner des descripteurs chimiques en relation avec la propriété cible ;
- Diviser cet ensemble de données en un ensemble jeu d'apprentissage et un ensemble de test ;
- Construire des modèles à partir d'un ensemble d'apprentissage avec les outils statistiques souhaités. Caractériser ces modèles par leurs indices de validation internes et vérifier leur robustesse par un test de hasardisation ;
- Valider les modèles avec l'ensemble de test et calculer leur indice de corrélation externe.
- Répéter l'opération de division pour obtenir d'autres ensembles d'apprentissage et de test. La division optimale donne le plus petit ensemble d'apprentissage capable de bonnes prévisions pour le plus grand ensemble de test ;
- Explorer et exploiter les modèles validés pour comprendre les mécanismes possibles et faire des prévisions.

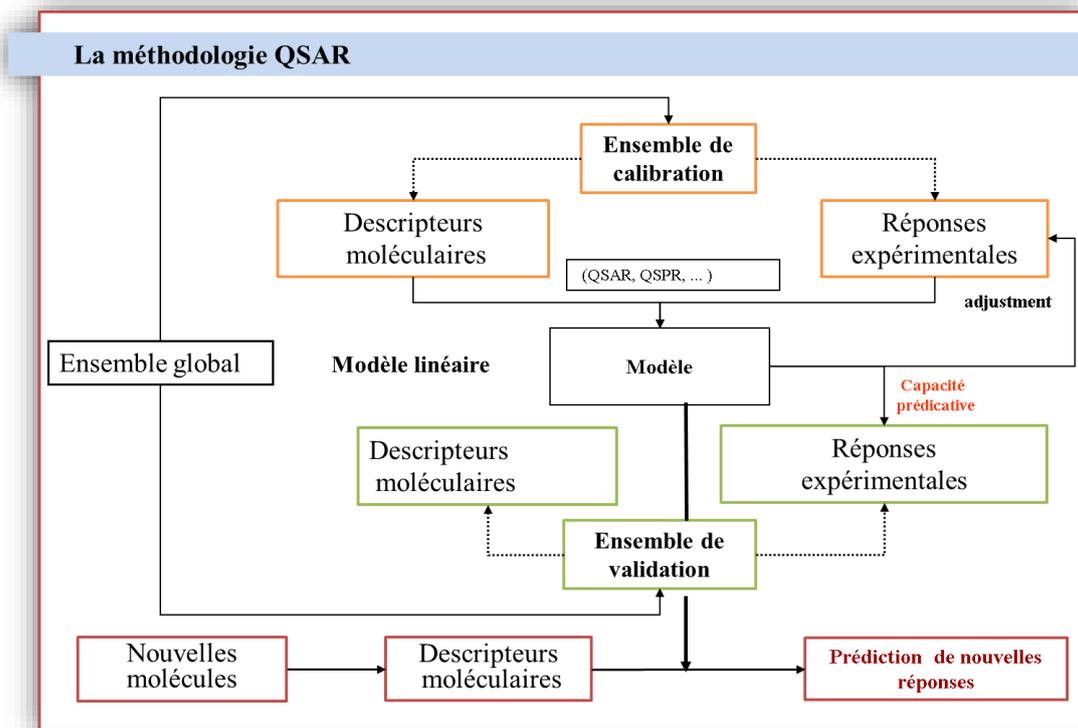


Figure 10 : Stratégie globale d'une étude QSAR

6. Conclusions :

Outre les méthodes classiques telles que les expériences *in vivo* et *in vitro*, l'utilisation d'outils informatiques suscite de plus en plus d'intérêt au sein de la communauté scientifique. Les outils informatiques fournissent des simulations, des moyens d'accélération de la conception moléculaire, offrant des prédictions, et expliquant en quelque sorte le mécanisme. Les QSAR tirent leurs méthodes à partir de la chimiométrie et des simulations physiques, et en les combinant avec les statistiques pour construire des systèmes pouvant être prédictifs.

Les QSAR peuvent s'étendre à partir de méthodes très prédictives, qui habituellement utilisent des corrélations complexes et non linéaires explicatifs et qui ont accès à une interprétation simple.

L'utilisation des modèles QSAR est en croissance, car ils fournissent la seule méthode abordable pour filtrer de grandes quantités de données.

The background of the page features a light blue gradient. At the top, there are several complex molecular structures composed of interconnected spheres and lines, representing various chemical compounds. At the bottom, a prominent blue DNA double helix structure is visible, winding across the width of the page. The text is centered in the middle of the page.

Partie II : Etude Expérimentale



Matériels et méthodes

1. Matériels et méthodes :

1.1. Les pyrrolopyrimidines :

Les pyrimidines fusionnées sont l'une des classes intéressantes d'hétérocycles, qui ont été largement explorées par les chimistes médicaux en raison de leur vaste profil pharmacologique. Dont les pyrrolopyrimidines qui ont gagné l'attention de nombreux chercheurs comme un échafaudage utile avec des activités biologiques variées (agents anticancéreux, antibactériens, anti-inflammatoires et antiviraux). [177], [178]

Actuellement, certains dérivés de la pyrrolopyrimidine sont en cours d'essais cliniques contre diverses cibles pathologiques. Leur développement a commencé au début des années 1970 et le premier Schéma de synthèse est apparu en 1974, impliquant des pyrimidines monocycliques comme substrat clé [179]. Plus tard, différentes stratégies de synthèse ont été développées sur la base des dérivés du pyrrole, et les modifications du squelette des pyrimidines bicycliques. [180], [181]

1.1.1. Stratégies de synthèse des pyrrolopyrimidines :

En raison de leur immense importance pharmacologique, de nombreux efforts ont été déployés pour générer des séries de pyrrolopyrimidines et leurs dérivés. [182]–[190]

La figure 11 résume quelques schémas de synthèse de pyrrolopyrimidines et leurs dérivés (Voies synthétiques 1 à 17).

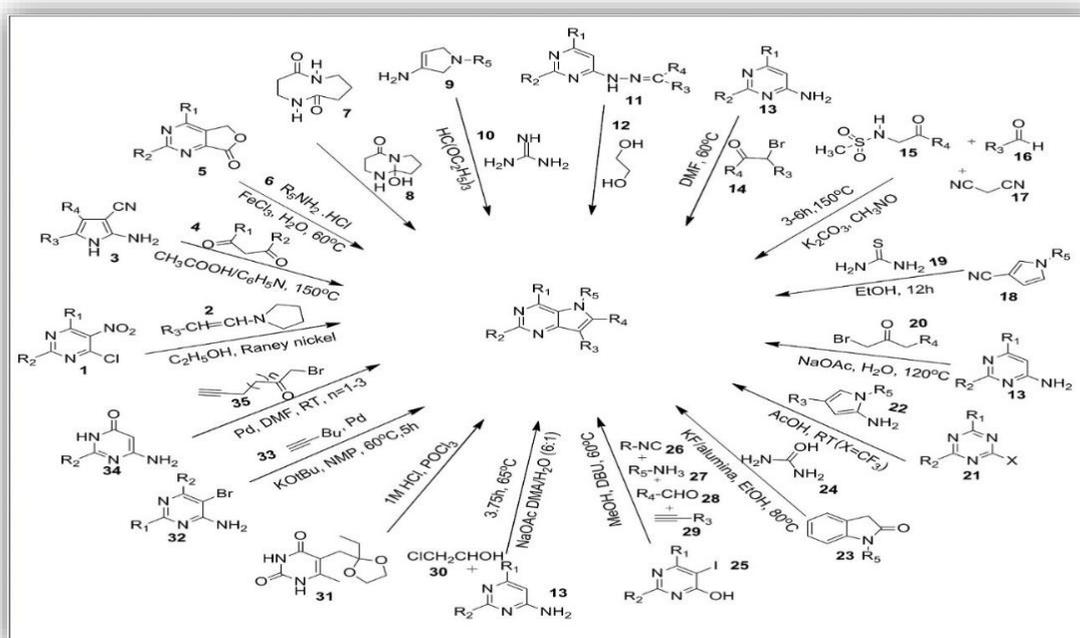


Figure 11 : Diverses voies de synthèse pour les dérivés de la pyrrolopyrimidine

1.1.2. Profil pharmacologique des pyrrolopyrimidines :

L'étude biologique des pyrrolopyrimidines suggère l'implication de nombreuses voies dans leur mécanisme d'action pour leur potentiel pharmacologique varié. Pour l'activité anticancéreuse, ils y'a plusieurs mécanismes, notamment l'inhibition de membres cruciaux des voies de signalisation, la déstabilisation des microtubules, l'inhibition de la Src-tyrosine kinase, l'inhibition de MRP1 et l'inhibition de la carbonhydrase résumés dans figure 12. [191]

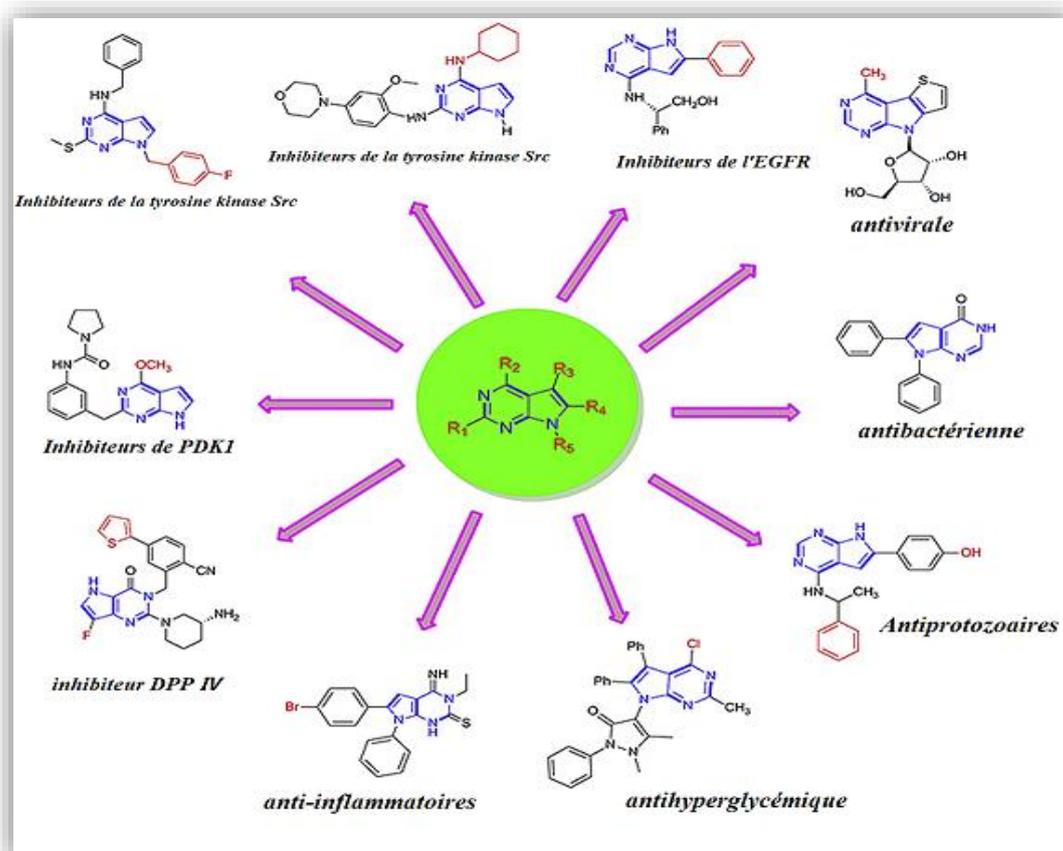


Figure 12 : Activités biologiques des pyrrolopyrimidine

1.1.3. Activité anticancéreuse :

a) Inhibiteurs multi-résistants de la protéine 1 (MRP1) :

En 2016, Schmit et al [192]. Ont synthétisé une série de dérivés de la pyrrolopyrimidine et ont évalué la puissance et la sélectivité de ces composés contre les maladies associées à la résistance aux médicaments protéine 1 (MRP1, ABCC1). Récemment, des dérivés de pyrrolo et d'indolopyrimidines ont été observés en tant qu'inhibiteurs puissants de MRP1 [192]. Ci-après d'autres activités anticancéreuses :

- i)** Inhibiteur de HCK.
- j)** Inhibiteurs du récepteur du facteur de croissance de type 1 de l'insuline (IGF-1R).
- k)** Inhibiteurs de la tyrosine kinase Src.
- l)** Inhibiteurs de l'EGFR (récepteur du facteur de croissance épidermique).
- m)** Inhibiteurs de l'anhydrase carbonique.

- b)** gents antiprolifératifs.
- c)** Inhibiteurs de la kinase Mps1.
- d)** Inhibiteur de HCK et de FLT3-ITD.
- e)** Inhibiteurs d'Akt.
- f)** Inhibiteurs doubles de l'EGFR et du VEGFR2
- g)** Agent de déstabilisation des microtubules.
- h)** Inhibiteurs de PDK1 (3-Phosphoinositide-dependent kinase1).

Les autres activités de la pyrrolopyrimidine :

- Activité antivirale.
- Agents antidiabétiques.
- Activité antihypertensive.
- Activité anti-inflammatoire.
- Activité anti-protozoaire.
- Activité antibactérienne.

1.2. Origine des données :

Dans notre travail l'homogénéité des données est essentielle, l'activité doit être mesurée par un seul et même protocole, avec des conditions expérimentales identiques. L'activité de tous les composés utilisés dans l'étude QSAR a été mesurée par un essai d'accumulation du médicament en utilisant la lignée cellulaire COR.L23/R exprimant MRP1, les résultats sont rapportées en tant que IC₅₀ pour l'accumulation de daunomycine, où l'accumulation de 100% est celle observée en présence d'un modulateur MRP1 connu : vérapamil à 100 µM. Voir annexe 01. [193][194]

1.3. Traitements des données :

Le développement du modèle QSAR nécessite le passage de notre base de composés chimiques par trois étapes :

- Etape 1 Préparation de l'ensemble de données ;
- Etape 2 : Calcule des descripteurs ;
- Etape 3 : La modélisation.

En utilisant plusieurs logiciels de traitement :

✓ **ChemDraw** : Est un outil complet destiné aux chimistes et biologistes, intégrant toute une gamme d'outils intelligents. Il est imposé comme une référence pour le dessin de structures des composés chimiques. Simple à utiliser, puissant et qui permet de dessiner de manière intuitive et efficace en deux et trois dimensions. [195], [196]

✓ **HyperChem** : est un environnement de modélisation moléculaire sophistiqué, connu pour sa qualité, sa flexibilité et sa facilité d'utilisation. Il propose différentes méthodes d'optimisation (PM3, MM+, AM1, etc...) En optimisant la molécule, les descripteurs moléculaires peuvent être calculés telle que le volume moléculaire, les énergies HUMO et LUMO...etc. Les descripteurs moléculaires théoriques ont été calculés par le processus suivant : les structures moléculaires ont été pré-optimisées par le champ de force de la mécanique moléculaire MM+ issu de ce logiciel de modélisation moléculaire. [197], [198]

✓ **Dragon** : est une application de calcul de descripteurs moléculaires, Sachant qu'il n'a pas été conçu comme un logiciel QSAR, il ne fournit que des descripteurs moléculaires et ne peut pas effectuer une analyse QSAR (au contraire le MobyDigs). Cependant, DRAGON permet la fusion des descripteurs moléculaires calculés (1664 descripteurs appartenant à différentes classes) et des propriétés définies par l'utilisateur pour un ensemble de molécules, fournissant un fichier de sortie complet qui est facilement chargé. [199]

✓ **QSARINS (QSAR-INSUBRIA)** : est un nouveau logiciel pour le développement et la validation des modèles de régression linéaire multiple (MLR) par les moindres carrés ordinaires

(MCO) et algorithme génétique pour la sélection de variables. Ce programme se concentre principalement sur la validation externe des modèles QSAR. la réduction des descripteurs moléculaires d'entrée, la division de l'ensemble de données en ensembles d'apprentissage et de prédiction, détection de valeurs aberrantes et prédictions interpolées ou extrapolées, validation interne et externe par différents paramètres, modélisation consensuelle et diverses graphs pour les visualisations sont implémentées. QSARINS est une plate-forme conviviale pour la modélisation QSAR en accord avec les Principes de l'OCDE et pour l'analyse de la fiabilité des données prédites obtenues. [200], [201]

✓ **CORAL SEA** : est un logiciel libre qui génère des descripteurs optimaux spéciaux utilisé dans la construction des modèles QSAR pour les substances représentées par (SMILES) à l'aide de la méthode Monte Carlo (www.insilico.eu/Coral). [202]

1.4 Répartition des échantillons :

Pour la sélection des composés de calibrage et de validation, nous avons utilisé :

- **Le choix aléatoire** : est la technique la plus simple. elle suppose qu'un ensemble d'échantillons tirés aléatoirement d'un lot suffisamment grand suit la distribution statistique du lot entier. [203]

1.5 Méthode de sélection de descripteurs :

Les meilleurs descripteurs sont sélectionnés en explorant la qualité statistique de toutes les combinaisons possibles des descripteurs disponibles, en utilisant la régression par les moindres carrés ordinaires (OLS) et la sélection de sous-ensembles de variables par algorithme génétique (AG-SSV) du logiciel QSARINS. Cette procédure de sélection de variables génère une «population» de modèles, classés selon les valeurs décroissantes de R^2 . Les meilleurs modèles ont été choisis en utilisant le Q_{LOO}^2 comme valeur d'optimisation et en tenant compte de la dimension des modèles qui devrait être aussi petite que possible. En outre, la corrélation entre les descripteurs et la réponse modélisée est vérifiée par la règle QUIK pour exclure les modèles à forte colinéarité entre les descripteurs. [200]

1.6. La Régression Linéaire multiple MLR :

Est l'une des méthodes de modélisation les plus populaires grâce à sa simplicité d'utilisation et de sa facilité d'interprétation. L'avantage important de la régression linéaire multiple est qu'elle est très transparente, puisque l'algorithme est disponible, et que les prédictions peuvent être réalisées facilement. [204]

1.7. Régression par la méthode de Monte Carlo :

L'approche des descripteurs hybrides optimaux est basée sur les structures SMILES et sur la représentation graphique moléculaire qui sont basées sur l'optimisation des poids de corrélation. Le programme CORAL (www.insilico.eu/Coral) a été utilisé pour construire le modèle de régression qui est implémenté dans la méthode de Monte Carlo. [205]–[207]

Le calcul par CORAL passe par les étapes suivantes :

- La structure moléculaire peut être représentée par SMILES et/ou un graphique moléculaire ;
- Les descripteurs hybrides optimaux sont sensibles aux deux représentations susmentionnées de la structure moléculaire ;
- Les descripteurs optimaux hybrides sont calculés par optimisation des poids de corrélation de la Attributs SMILES avec les poids de corrélation des invariants du graphique ;
- Le descripteur hybride optimal DCW (T, N) est appliqué à un modèle prédictif de paramètre :

$$\text{Activité biologique} = C_0 + C_1 \times \text{DCW}(T, N) \quad (\text{eq 9})$$

$$\text{DCW}(T, N) = \text{DCWSMILES}(T, N) + \text{DCWgraph}(T, N) \quad (\text{eq 10})$$

Où l'activité est le critère d'évaluation, C_0 et C_1 sont les coefficients de régression.

DCW (T^*N^*) est le descripteur optimal, de la fonction des caractéristiques moléculaires extraites de SMILES. T^* , N^* sont le meilleur seuil et le meilleur nombre d'époques calculées pendant la construction, respectivement. [208], [209]

Le T est le seuil pour définir deux classes de caractéristiques moléculaires (attributs structurels) : rare et fréquent. Si un attribut structurel est fréquent dans les composés d'entraînement, elle peut être associée à l'activité, tandis que les attributs rares ont moins de signification et sont traités comme du bruit. Le descripteur optimal est calculé avec les poids de corrélation (CW) des attributs structurels actifs (ASs) : pour les attributs rares, les CW sont mis à zéro, c'est-à-dire qu'ils ne sont pas impliqués dans la modélisation. [207], [210]

Le N est le nombre d'époque est une séquence de variation d'un CW de la caractéristique moléculaire. Les valeurs numériques CW sont obtenues avec l'optimisation Monte Carlo pendant la construction. L'objectif de la technique de Monte Carlo est de maximiser le coefficient de corrélation entre le descripteur optimal et le paramètre. L'optimisation est faite

avec la formation et les ensembles de formation invisibles afin d'éviter la surformation. Dans ce travail, l'optimisation a été effectuée au moyen de l'équilibre des corrélations. [205], [211] Les descripteurs optimaux utilisés ici sont calculés via le logiciel CORAL comme suit :

$$\mathbf{DCW}(T^*, N^*) = _CW(Sk) + _CW(SSk) + _CW(SSSk) \quad (\text{eq 11})$$

Les Sk sont des atomes SMILES, c'est-à-dire un symbole de la notation SMILES ou deux symboles qui ne peuvent pas être examinés séparément (par exemple, 'Cl', 'Br', etc.). Les SSk et SSSk sont des combinaisons de deux et trois atomes SMILES, respectivement. Les CW(Sk), CW(SSk), et CW(SSSk) sont ce qu'on appelle la corrélation poids des attributs SMILES ci-dessus. La corrélation poids CW(X), où X = Sk, SSk, SSSk, est un coefficient utilisé pour calculer le DCW(T*, N*) [212], [213]. L'optimisation Monte Carlo vise à fournir CW(X) qui donnent la fonction cible maximale (TFa) calculée comme suit :

$$\mathbf{TFa} = R + R' / R - R' \times 0.1 + a \cdot \text{IICCLB} \quad (\text{eq 12})$$

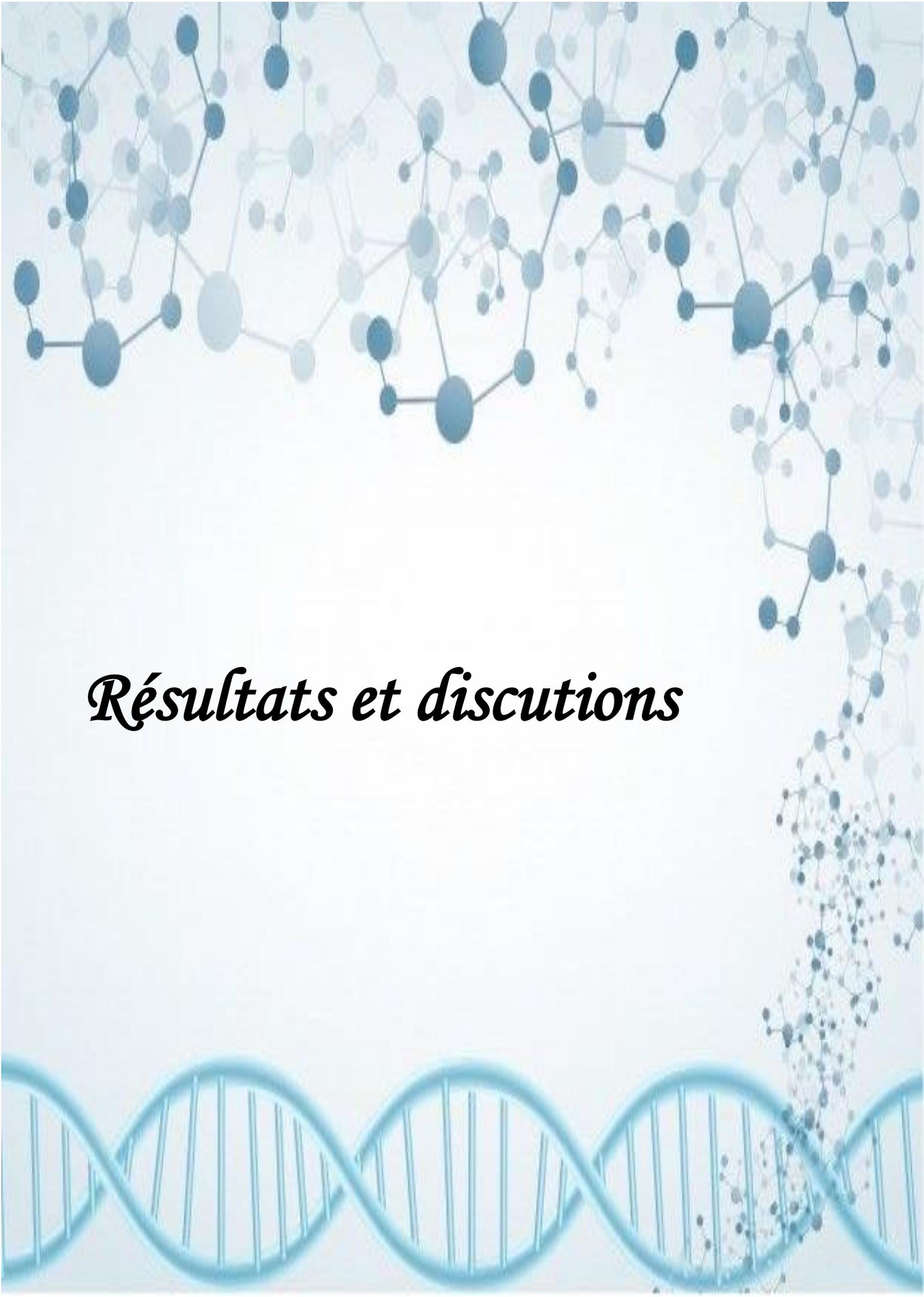
Les coefficients R et R' sont des coefficients de corrélation entre DCW(T*,N*) et expérimental pLC₅₀ pour la formation et les ensembles de formation invisibles, respectivement; le IICCLB est l'indice d'idéalité de corrélation pour l'ensemble d'étalonnage suggéré dans la littérature. [214]–[216]

- a = 0.0 donne la fonction cible de différents modèles CORAL ;
- a = 0,1 donne modèle le CORAL de nouvelle qualité.

Où la qualité statistique de l'ensemble de calibrage est améliorée au détriment de l'ensemble de formation. Le modèle de pLC₅₀ est le suivant :

$$p\mathbf{LC}_{50} = C_0 + C_1 \times \mathbf{DCW}(T^*, N^*) \quad (\text{eq 13})$$

Les valeurs T = T* et N = N* sont des valeurs des paramètres qui fournissent les meilleures caractéristiques statistiques du modèle pour l'ensemble d'étalonnage. [208], [210], [211]

The background features a light blue gradient with various molecular models. At the top, there are several ball-and-stick molecular structures in shades of blue and grey. At the bottom, a prominent DNA double helix is rendered in a glowing blue color. The text is centered in the middle of the page.

Résultats et discussions

2. Résultat et discussions :

2.1. Construction model dragon descripteur :

La construction du modèle QSAR a été déterminée en prenant pour variable dépendante le logarithme de l'inverse de IC₅₀ (en mmol / litre) pour les inhibiteurs MRP1, et pour les variables explicatives on a choisis entre 3 modèles :

- Model un descripteur (T(N..F) ;
- Model deux descripteur (T (N..F) , BELe2) ;
- Model trois descripteur (T(N..F), GATS2v, VEA2).

Tableau 4 regroupe les définitions de différents descripteurs.

Tableau 4 : Les descripteurs intervenant dans les modèles

Descripteurs	Classe	Signification
T (N..F)	2D Paires d'atomes	La Somme de la distance topologique entre N..F
BELe2	BCUT descripteur	valeur propre n. 2 la plus basse de la matrice de Burden / pondérée par les électronégativités atomiques de Sanderson
GATS2v	2D autocorrélation	Autocorrélation de Geary du retard 2 pondérée par le volume de van der Waals
VEA2	VEA2	descripteur topologique

Ces 105 composés ont été divisés en un ensemble d'apprentissage (84 composés) et un ensemble de test (21 composés) en utilisant la randomisation ainsi que la diversité chimique et biologique.

2.1.1. Analyse de la régression :

L'activité inhibitrice MRP1 représenté par le cologarithme décimal d'IC₅₀ (*p*IC₅₀) de 105 dérivés de pyrrolopyrimidines. Parmi les modèles sélectionnés par l'algorithme génétique nous avons retenu les modèles dont les équations sont :

2.1.1.1. Model pour 1 Descripteur T(N..F) :

Équation du model :

$$\text{Log (1/IC}_{50}) = -0.233 (\pm 0.072) + 0.0096 (\pm 0.0007) \text{ T (N..F)} \quad (\text{eq 14})$$

Le meilleur model pour 1 dimension a été obtenu avec T(N..F) qui représente la somme de la distance topologique entre N et F. les paramètres statistiques sont regroupé dans le tableau 5.

Tableau 5 : Paramètres statistiques du modèle 1

ID	MODEL	R ²	Q ² _{LOO}	Q ² _{LMO} 10%	Q ² _{LMO} 20%	Q ² _{LMO} 30%	Q ² _{ext}	Q ² _{F1}	Q ² _{F2}	Q ² _{F3}	KΔ	K	F	S
1	T (N..F)	0,706	0,692	0,69	0,69	0,69	0,681	0,678	0,674	0,718	0,84	0,93	199,8	0,49

2.1.1.2. Model pour 2 Descripteurs T (N..F) BELe2 :

Équation du model :

$$\text{Log (1/IC}_{50}) = -22.641 (\pm 4.327) + 0.0083 (\pm 0.0006) \text{ T (N..F)} + 12.194 (\pm 2.354) \text{ BELe2} \quad (\text{eq 15})$$

Le meilleur model pour 2 dimension a été obtenu grâce à l'utilisation de l'algorithme génétique avec T(N..F) qui représente la somme de la distance topologique entre N et F et BELe2 qui montre la valeur propre inferieur n.2 de la matrice BURDEN. Les paramètres statistiques sont regroupés dans le tableau 6 :

Tableau 6 : Paramètre statistiques du modèle 2

ID	MODEL	R ²	Q ² _{LOO}	Q ² _{LMO} 10%	Q ² _{LMO} 20%	Q ² _{LMO} 30%	Q ² _{ext}	Q ² _{F1}	Q ² _{F2}	Q ² _{F3}	KΔ	K	F	S
2	T (N..F) BELe2	0,778	0,763	0,76	0,764	0,759	0,745	0,742	0,739	0,774	0,23	0,94	144,4	0,42

2.1.1.3. Model pour 3 Descripteur T(N..F) GATS2v VEA2:

Équation du model :

$$\text{Log (1/IC}_{50}) = 6.712 (\pm 0.922) + 0.0055 (\pm 0.0007) \text{ T (N..F)} - 4.321 (\pm 0.716) \text{ GATS2v} - 21.194 (\pm 6.516) \text{ VEA2} \quad (\text{eq 16})$$

Le meilleur model pour 3 dimension a été obtenu grâce à utilisation de l'algorithme génétique avec T(N..F) qui représente la somme de la distance topologique entre N et F et

GATS2v qui représente l'autocorrélation de Geary du retard 2 pondérée par le volume de Van Der Waals et VEA2. Les paramètres statistiques sont regroupés dans le tableau 7.

Tableau 7 : Paramètres statistiques du modèle sélectionné

ID	MODEL	R ²	Q ² _{LOO}	Q ² _{LMO} 10%	Q ² _{LMO} 20%	Q ² _{LMO} 30%	Q ² _{ext}	Q ² _{F1}	Q ² _{F2}	Q ² _{F3}	KΔ	K	F	S
3	T (N..F) GATS2v VEA2	0,8302	0,815	0,81	0,81	0,81	0,811	0,80	0,798	0,825	0,18	0,998	132	0,37

D'après les résultats sélectionner (les 3 tableaux) on a choisie celui de 3 descripteurs (R² =83.02 %) (Q²_{LOO} =81.5 %) (Q²_{ext}=81.1 %)

2.1.2. Qualité de l'ajustement :

Les valeurs des paramètres statistiques (Tableau 7) prennent en considération la corrélation entre les trois descripteurs (T(N..F) GATS2v VEA2) des 105 dérivés de pyrrolopyrimidines.

La valeur élevée de F = 132 plus les valeurs de (R² =83.02 %) et de (Q² =81.5%) montrent la bonne qualité de l'ajustement, et la valeur du coefficient de détermination R² signifie que 83.02% de la variabilité de pIC₅₀ est expliquée par ces descripteurs.

le tableau 7 et la figure 13, représentent la bonne qualité d'ajustement, la robustesse et le pouvoir prédictif de notre modèle (R², Q²_{LOO}, Q²_{ext}), les valeurs des erreurs relativement faibles, et celles de la dispersion des points autour de la droite, Montrent que les valeurs prédites sont en adéquation avec les valeurs expérimentales, caractéristique d'un bon ajustement.

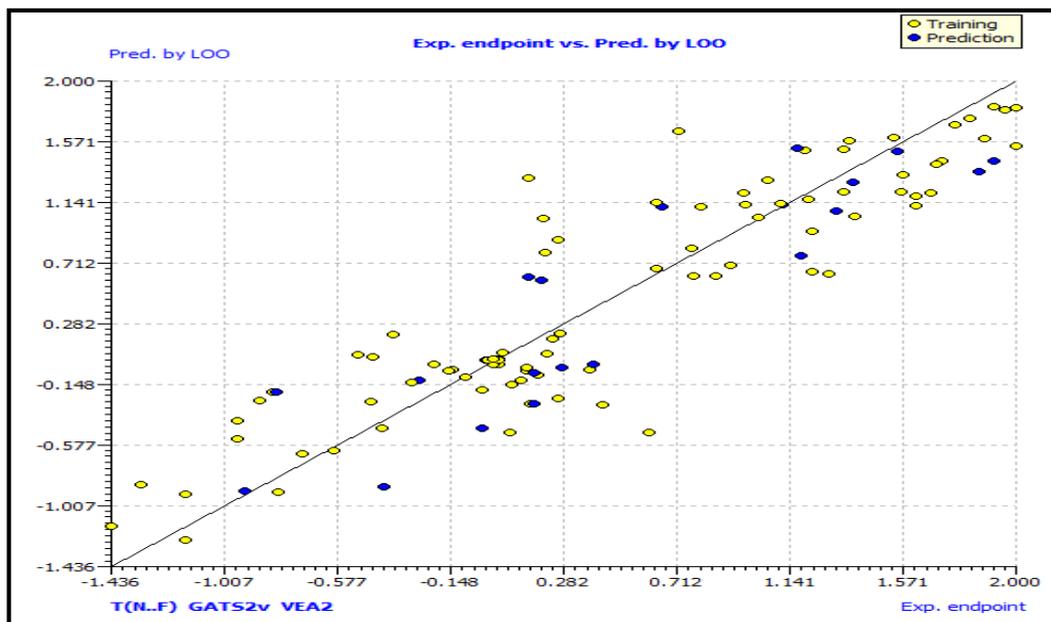


Figure 13 : Droites d'ajustement pour le modèle (03) des valeurs expérimentales et prédits de la notation pIC_{50} pour le modèle de QSAR

2.1.3. Validation :

Dans cette étude l'ensemble des données a été divisé en deux sous-ensemble : sous-ensemble d'entraînement (80 % = 84 molécules) pour la construction du modèle et la validation interne, et un sous-ensemble de Test (20 % = 21 molécules) pour la validation externe.

2.1.3.1. Les Principes de l'OCDE :

L'OCDE sont les meilleurs éléments possibles des points essentiels à résoudre tout en développant des modèles QSAR fiables et reproductibles [217]. Les principes ont été formulés par des experts du QSAR lors d'une réunion tenue à Setúbal (Portugal) en mars 2002 comme lignes directrices pour la validation des modèles QSAR. Les 5 directives adoptées par l'OCDE pour la validité du modèle QSAR sont :

- ❖ Principe 1— Définition précise de la propriété prédite,
- ❖ Principe 2— Equation mathématique (ou algorithme) sans équivoque (reproductible),
- ❖ Principe 3— Domaine d'applicabilité défini,
- ❖ Principe 4— Mesures appropriées des performances,
- ❖ Principe 5— Interprétation des mécanismes sous-jacents (si possible).

Le défi actuel de QSAR/QSPR n'est plus au développement d'un modèle statistiquement solide, c'est l'élaboration d'un modèle capable de prédire avec précision l'activité ou la propriété de nouveaux produits chimiques.

2.1.3.2. Validation interne :

Le coefficient de détermination de la validation croisée du modèle $Q_{LOO}^2 = 81.5 \%$, témoigne d'une bonne corrélation entre l'activité prédite et l'activité réelle, et reflète une précision du modèle.

2.1.3.3. Validation externe :

La validation interne, pour les nouveaux composés ne suffit pas pour juger la capacité prédictive du modèle développé. La validation externe doit être réalisée pour une bonne prévision des composés qui n'ont pas été utilisés lors de l'élaboration du modèle.

Les paramètres de la validation externe $Q_{ext}^2 = 81.1 \%$ confirme la bonne capacité prédictive du modèle pour les composés non impliqués dans les calculs.

2.1.3.4. test de randomisation (Y scrambling):

Pour éviter les corrélations dus au hasard et valider le modèle MLR calculé, le test de randomisation a été appliqué. Il consiste à mélanger les valeurs de la réponse Y entre elles sans changer la valeur des descripteurs Xi (figure 14). [218], [219]

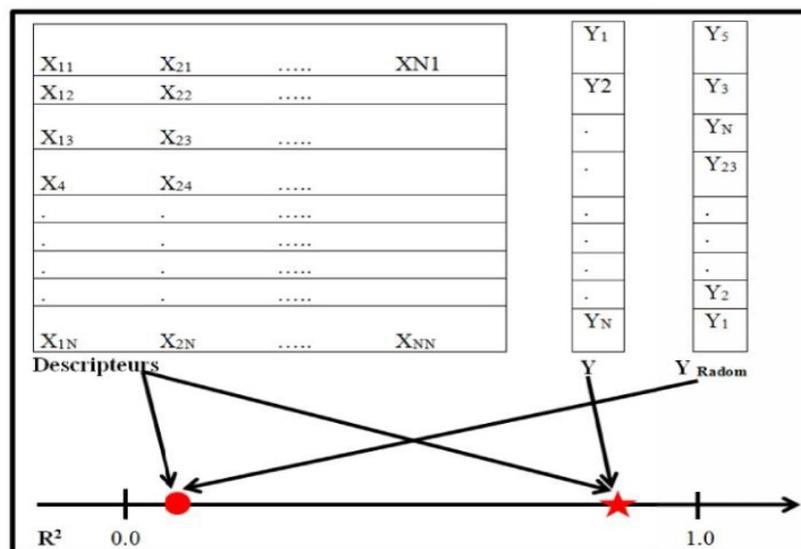


Figure 14: principe du test de randomisation [220]

Les 105 modèles ont été développés, et la figure 15 montre les faibles valeurs de Q^2 et R^2 obtenues après chaque mélange, faisant ressortir que les résultats du modèle original ne sont pas aléatoires.

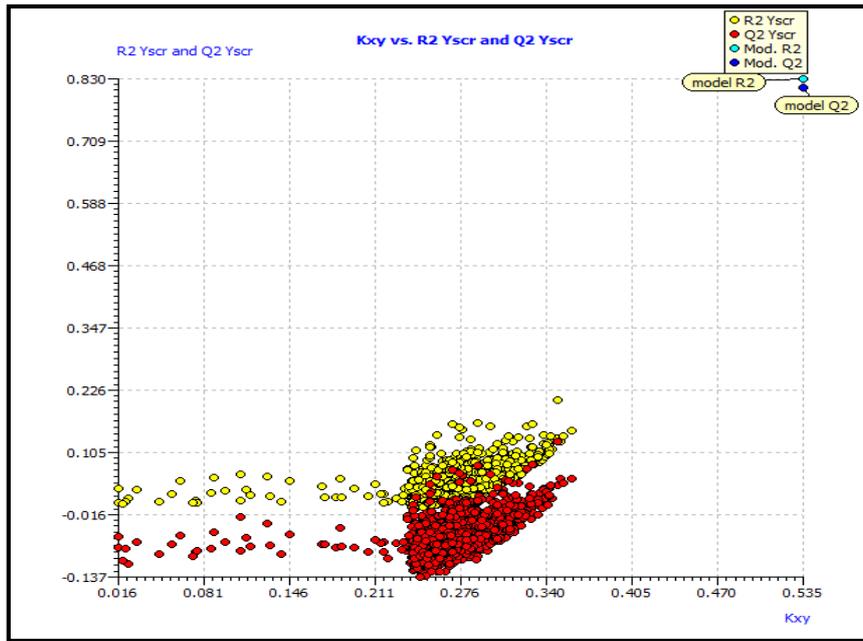


Figure 15 : Tests de randomisation

2.1.4. Le domaine d'application :

Le diagramme de Williams désigne les composés qui se trouvent dans le domaine d'application de notre model (Figure 16), la plupart des composés de l'ensemble de données sont dans l'intervalle ± 3 du domaine d'application, à l'exception des 3 composés (82/83/103) qui dépassent le seuil $h^* = 0,14$ (points levier). Ainsi, le modèle proposé pourrait être utilisé pour examiner les bases de données existantes.

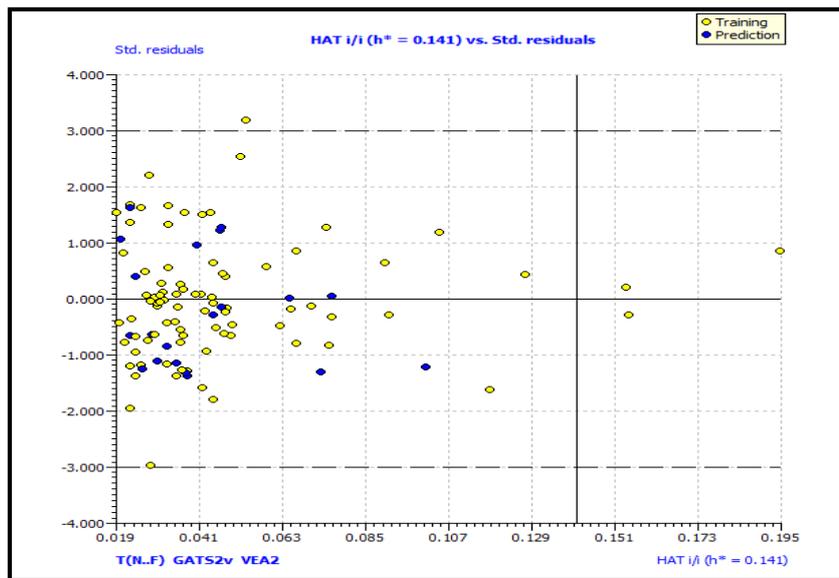


Figure 16 : Diagramme de Williams

2.1.5. Conclusion :

Nous avons développé une équation utile de QSAR qui relie des descripteurs théoriques à l'IC₅₀ expérimental de 105 molécules. Les valeurs très élevées des paramètres statistiques (Q_{LOO}^2 , Q_{LMO}^2 , Q_{F1}^2 , Q_{F2}^2 , Q_{F3}^2), montrent que le modèle proposé est robuste et possède une bonne capacité prédictive. On peut alors conclure que pour cette famille de composés, le modèle proposé pourrait prédire et donner des informations sur l'inhibition du MRP1.

2.2. Construction model CORAL :

2.2.1. Descripteurs optimal SMILES-based :

Afin de vérifier la reproductibilité des modèles CORAL [31], Il est convenable de tester plusieurs fractionnements dans le sous-système d'entraînement (c'est-à-dire les ensembles d'entraînement (73), calibration (16) et de validation (16)). Le tableau 8 des trois fractionnements aléatoires donne les modèles suivants :

Tableau 8 : Les paramètres potentiels prédictifs d'un modèle pour les 3 choix

Choix 1, (T=2/N=4)	n	R ²	CCC	IIC	Q ²	Q _{F1} ²	Q _{F2} ²	Q _{F3} ²	\overline{R}_m^2	S	MAE	F
Training	73	0.85	0.92	0.76	0.84					0.32	0.25	409
Calibration	16	0.85	0.80	0.92	0.81	0.72	0.72	0.59	0.82	0.55	0.45	82
Validation	16	0.85	0.92	0.85	0.80					0.36	0.27	82
pIC₅₀ = -4.3416741 (± 0.0283116) + 0.0523884 (± 0.0002997) * DCW(3,7)												
Choix 2, (T=1/N=6)	n	R ²	CCC	IIC	Q ²	Q _{F1} ²	Q _{F2} ²	Q _{F3} ²	\overline{R}_m^2	S	MAE	F
Training	73	0.81	0.89	0.74	0.79					0.38	0.30	294
Calibration	16	0.84	0.90	0.92	0.78	0.82	0.81	0.79	0.73	0.40	0.31	76
Validation	16	0.85	0.91	0.51	0.81						0.27	82
pIC₅₀ = -3.6454358 (± 0.0282435) + 0.1072802 (± 0.0006958) * DCW(3,7)												
Choix 3, (T=1/N=5)	n	R ²	CCC	IIC	Q ²	Q _{F1} ²	Q _{F2} ²	Q _{F3} ²	\overline{R}_m^2	S	MAE	F
Training	73	0.86	0.93	0.90	0.86					0.30	0.22	453
Calibration	16	0.88	0.93	0.94	0.84	0.88	0.85	0.85	0.78	0.33	0.26	98
Validation	16	0.86	0.93	0.82	0.89					0.37	0.29	153
pIC₅₀ = -4.4198667 (± 0.0337899) + 0.0584289 (± 0.0003850) * DCW(3,7)												

T : le seuil

N : époques

Différentes équations sont examinées pour déterminer si des corrélations quantitatives étaient possibles entre les caractéristiques moléculaires fondamentales basées sur SMILES et les activistes biologiques MRP1 observées sur 105 molécules.

Les approches que nous avons appliquées utilisent différents niveaux d'information associés à la technique de Monte carlo [221]–[223]. L'approche prend en compte l'erreur observée sur les prédictions dans le processus de raffinement du modèle. Cela sert à identifier les faiblesses des modèles, permettant de sélectionner les plus robustes. L'optimisation donne rapidement une valeur élevée pour le coefficient de détermination pour l'ensemble de calibration et de validation.

Cependant, avec plus d'époques, les coefficients de détermination diminuent. L'optimisation utilise plus d'époques pour optimiser, mais les résultats sont plus fiables.

2.2.2. Conclusion :

Le modèle CORAL est basé uniquement sur les valeurs de SMILES et de l'activiste observés, sans qu'il soit nécessaire de calculer les descripteurs, généralement fournis par un logiciel supplémentaire. Cela présente deux avantages :

- la simplicité de l'approche ;
- la stabilité et reproductibilité plus élevées.

Ainsi, l'approche est plus simple et ne nécessite pas une série d'outils, comme dans les autres approches, qui procèdent d'abord au calcul des descripteurs/caractéristiques moléculaires, et puis au calcul de l'algorithme.

2.3. Comparaison entre les modèles DRAGON-Based et Optimal SMILE-Based :

D'après les principes de base pour l'établissement des modèles QSAR, Il convient d'estimer le potentiel prédictif d'une approche avec un groupe de split en ensembles d'apprentissage de validation. Le tableau 6 montre la qualité statistique des modèles optimisés par la méthode Monte Carlo implémentée dans le logiciel CORAL à partir de la base SMILES et celui de QSARINS avec DRAGON descripteurs. D'après les résultats obtenus pour les deux approches on trouve qu'il y a une différence notable, le meilleur modèle a été obtenu avec CORAL en utilisant l'optimisation Monte Carlo.

Tableau 9 : Variation des paramètres statistiques selon l'approche utilisée

Approche	Choix 1, (T=1, N=5)	N	R ²	Q ²	Q _{F1} ²	Q _{F2} ²	Q _{F3} ²	S	F
SMILE- Based	Training	73	0.86	0.86				0.30	453
	Calibration	16	0.88	0.84	0.88	0.85	0.85	0.33	98
	Validation	16	0.86	0.89				0.37	153
DAGON- Based	T (N..F) GATS2v VEA2		0.83	0.82	0.80	0.80	0.83	0.37	132

The background features a light blue gradient with several molecular models. At the top, there are complex, interconnected networks of blue and grey spheres connected by thin lines, resembling a protein structure or a molecular network. At the bottom, a prominent blue DNA double helix structure is visible, winding across the width of the page. The overall aesthetic is clean and scientific.

Conclusion générale

Conclusion générale

La relation entre la structure moléculaire et l'activité biologique est l'un des défis majeurs de la conception des médicaments, dont le développement de nouveaux composés anticancéreux.

Notre but consiste à développer des modèles QSAR fiables sur la base d'une série de 105 dérivés pyrrolopyrimidine pour prédire l'activité inhibitrice de la protéine MRP1.

Dans la première partie, l'utilisation des descripteurs du logiciel SQARINS aboutissant à la construction des modèles QSAR par la sélection d'une approche algorithmique génétique, et dont les résultats obtenus ainsi que les paramètres statistiques des modèles doivent être acceptables.

Pour ce faire, le logiciel QSARINS a été utilisé ainsi qu'une analyse de régression linéaire multiple (MLR) nous permettant d'aboutir à une corrélation entre pIC_{50} et trois descripteurs sélectionnés : (T(N..F), BELe2, GATS2v VEA2). Dont l'objectif est de former l'équation finale de QSAR.

Les principales techniques de validation (interne, externe, la randomisation des Y...). Ont été utilisées pour évaluer le modèle QSAR obtenu en accord avec les cinq principes de l'OCDE ; nous avons également vérifié les critères d'acceptabilité du modèle de Golbarikh et Tropsha.

Le modèle QSAR a été déterminé en prenant pour variable :

- Dépendante : le logarithme de l'inverse de IC_{50} (en mmol / litre).
- Explicatives : trois descripteurs (T(N..F), BELe2, GATS2v VEA2).

Selon les conditions de Golbreich et Tropsha :

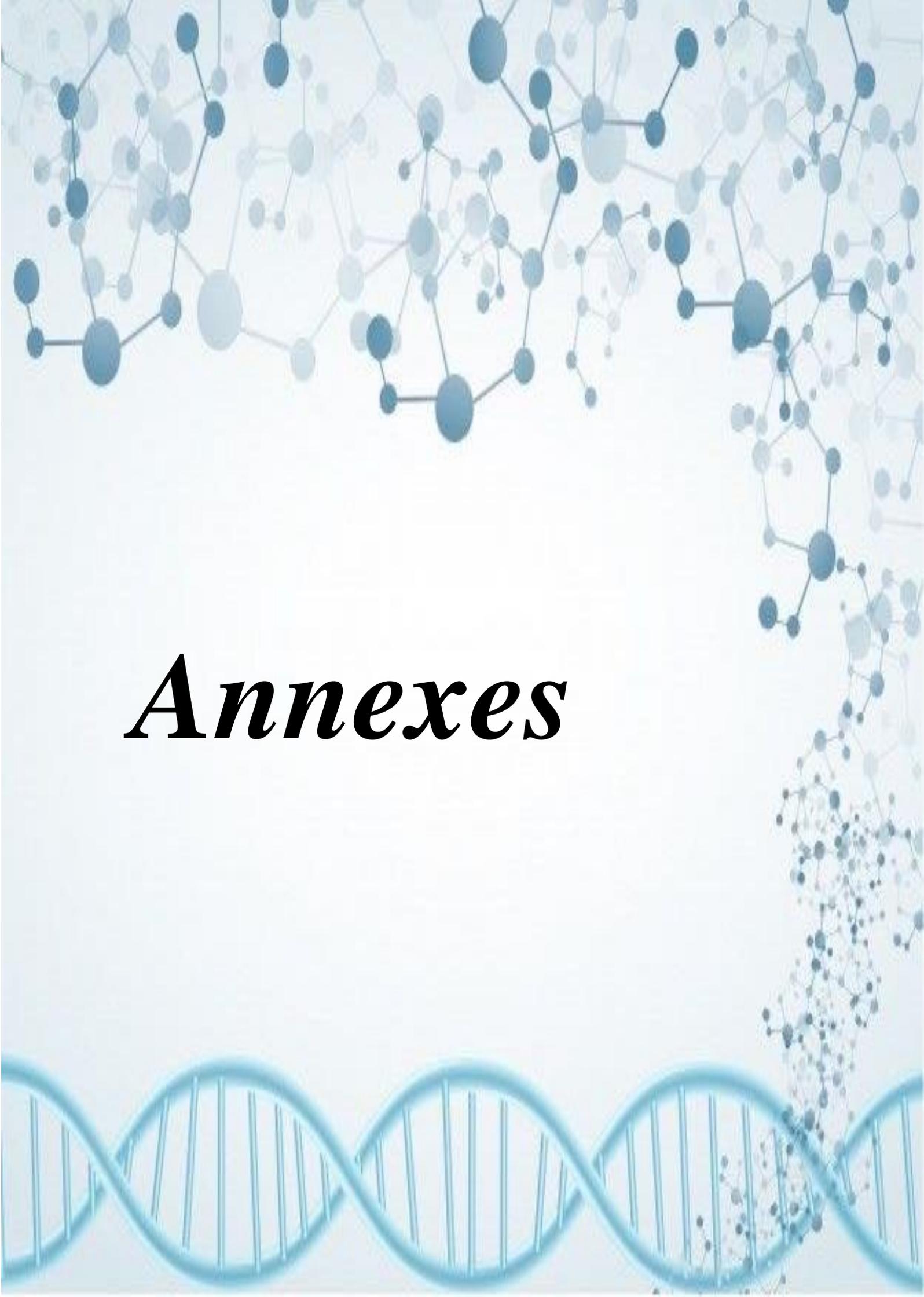
Le modèle trouvé : $R^2 = 83.02 \%$ $Q^2 = 81.51 \%$

Dans la deuxième partie, le logiciel CORAL, basé uniquement sur les valeurs de SMILES et l'activité observée donc d'une approche simple, nous a permis de construire des modèles QSAR.

Les résultats : $R^2 = 87 \%$ $Q^2 = 85 \%$

Enfin la troisième partie représente une étude comparative entre les descripteurs calculés par le logiciel QSARINS et ceux calculés par CORAL concluant ainsi que le modèle CORAL possède un caractère très robuste et d'excellents paramètres prédictifs internes et externes.

Notre objectif principal nous a permis d'entrevoir des perspectives assez prometteuses dans cette discipline ou nous prévoyons de poursuivre l'élaboration des modèles en utilisant d'autres méthodes telles que (Machine Learning, Deep Learning, les graphs machines) afin de poursuivre notre chemin de recherche.



Annexes

Annexes 01 : Les valeurs expérimentales prédit et calculé par l'approche
DRAGON-QSARINS pour les 105 dérivés de pyrrolopyrimidines

	Code	Ensemble	γ Exp	γ Pred
1	1069464	Entrainement	-0,591	-0,6094
2	9828452	Entrainement	1,215	1,174
3	9979854	Entrainement	0,222	0,0793
4	9980860	Entrainement	0,187	-0,0718
5	10004696	Entrainement	-0,137	-0,0421
6	10029563	Test	1,119	1,1269
7	10046544	Test	-0,267	-0,1144
8	10047863	Entrainement	0,64	0,6754
9	10051663	Test	0,662	1,1118
10	10071405	Entrainement	0,387	-0,0246
11	10092403	Entrainement	-0,866	-0,2744
12	10094280	Test	0,281	-0,0246
13	10322116	Entrainement	0,146	-0,0421
14	10344755	Entrainement	0,273	0,2219
15	10346312	Entrainement	0,052	0,0845
16	10388318	Entrainement	-0,823	-0,2094
17	10407830	Entrainement	0,161	-0,2696
18	10432300	Entrainement	0,036	-0,0022
19	10454772	Entrainement	0,041	0,0365
20	10455921	Entrainement	0,86	0,6362
21	10459316	Entrainement	0,244	0,1858
22	11164648	Test	-0,803	-0,1965
23	11177641	Test	0,399	-0,0022
24	11186582	Entrainement	-0,954	-0,4236
25	11189272	Entrainement	-0,447	-0,2659
26	11258566	Entrainement	-0,362	0,1831
27	11337594	Entrainement	-1,32	-0,8903
28	11351150	Test	0,171	-0,0589
29	11359019	Entrainement	-0,799	-0,8987
30	11362459	Entrainement	-0,151	-0,0461
31	11408751	Entrainement	0,921	0,7092
32	11418752	Entrainement	0,02	-0,0022
33	11432257	Entrainement	0,208	1,0067
34	11444700	Entrainement	1,367	1,5668
35	11454819	Entrainement	0,979	1,1227
36	10031184	Entrainement	1,721	1,4578
37	10320818	Test	1,187	0,7684
38	11156535	Entrainement	0,15	1,2549
39	11168500	Entrainement	1,699	1,4293
40	11177283	Entrainement	1,027	1,0405
41	11179994	Entrainement	1,108	1,1384

42	11200621	Test	0,199	0,5968
43	11201011	Test	1,678	1,2284
44	11212921	Test	1,387	1,2889
45	11213871	Test	1,319	1,0846
46	11214464	Entrainement	1,824	1,7458
47	11226516	Entrainement	1,77	1,6966
48	11246811	Test	0,15	0,6196
49	11247947	Entrainement	1,229	0,9504
50	11249529	Test	1,863	1,3622
51	11261262	Entrainement	1,388	1,0621
52	11270541	Test	1,553	1,5059
53	11270563	Entrainement	1,229	0,6796
54	11293906	Test	1,921	1,4409
55	11293909	Entrainement	1,921	1,8341
56	11293927	Entrainement	0,636	1,1329
57	11295806	Entrainement	2	1,8229
58	11304370	Entrainement	1,62	1,1379
59	11305058	Entrainement	1,347	1,5103
60	11316308	Entrainement	0,777	0,627
61	11318194	Entrainement	1,959	1,8068
62	11339085	Entrainement	0,772	0,8161
63	11350733	Entrainement	0,216	0,7824
64	11376266	Entrainement	1,538	1,6036
65	11385818	Entrainement	0,807	1,1069
66	11396606	Entrainement	1,06	1,2889
67	11419244	Entrainement	1,201	1,4917
68	11443156	Entrainement	1,886	1,6165
69	11443942	Entrainement	1,577	1,3512
70	11453657	Entrainement	0,721	1,6008
71	11454703	Entrainement	2	1,557
72	11455462	Entrainement	0,971	1,2021
73	11488138	Entrainement	1,569	1,2277
74	11751643	Test	1,174	1,5315
75	11752839	Entrainement	0,266	0,8608
76	11774980	Entrainement	1,292	0,6691
77	11775287	Entrainement	1,62	1,2
78	44276937	Entrainement	1,347	1,2208
79	917533	Entrainement	-1,152	-0,9411
80	9975676	Entrainement	-0,408	-0,4542
81	10048669	Entrainement	-0,009	0,0271
82	10333981	Entrainement	-1,436	-1,2025
83	10379829	Entrainement	-0,704	-0,6414
84	10383921	Entrainement	-0,025	-0,1739
85	10407830	Entrainement	0,432	-0,2696
86	10472817	Entrainement	0,146	-0,0093
87	11359019	Test	-0,923	-0,8987

88	127048170	Entrainement	-0,29	-0,152
89	127048171	Entrainement	-0,953	-0,5737
90	127049502	Test	0,173	-0,282
91	127049508	Entrainement	-0,21	-0,0074
92	127049509	Test	-0,398	-0,8648
93	127050480	Entrainement	0	0,0276
94	127050481	Entrainement	-0,442	0,0367
95	127050796	Entrainement	0,041	0,0276
96	127050797	Entrainement	0,02	0,0353
97	127050798	Entrainement	-0,498	0,0486
98	127051981	Entrainement	0,265	-0,2261
99	127051982	Entrainement	0,086	-0,1396
100	127051983	Entrainement	0,607	-0,4507
101	127051984	Entrainement	0,125	-0,1102
102	127051985	Entrainement	-0,086	-0,0924
103	127052887	Entrainement	-1,152	-1,2296
104	127052888	Entrainement	0,079	-0,419
105	127052889	Test	-0,025	-0,4532

Annexes 02 : Activité inhibitrice envers MRP1 de 105 dérivés de pyrrolorimidines le choix de split pour l'approche CORAL.

Code	SMILES	Choix 1	Choix 2	Choix3	pIC ₅₀	
1	1069464	<chem>C1CCC2=C(C3=C(N2CC1)C(=NC=N3)SCC4=CC=C(C=C4)Cl)C#N</chem>	Entraînement	Entraînement	Calibration	-0.591
2	9828452	<chem>C1CCN2C(=C(C3=C2C(=NC=N3)N4CCN(CC4)CCC5=CC(=C(C=C5)F)F)C(=O)NCC6=CC=CC=N6)C1</chem>	Entraînement	Entraînement	Entraînement	1.215
3	9979854	<chem>C1CCN2C(=C(C3=C2C(=NC=N3)N4CCN(CC4)CCC5=CN(C6=CC=CC=C65)C#N)C1</chem>	Entraînement	Calibration	Test	0.222
4	9980860	<chem>CC1=NC(=NO1)C2=C3CCCCN3C4=C2N=CN=C4N5CCN(CC5)CCC6=CC=CC=C6</chem>	Test	Entraînement	Entraînement	0.187
5	10004696	<chem>C1CCN2C(=C(C3=C2C(=NC=N3)N4CCN(CC4)CCC5=CC=CC=C5)C(=O)C6=CC=CC=C6)C1</chem>	Calibration	Test	Test	-0.137
6	10029563	<chem>C1CCN2C(=C(C3=C2C(=NC=N3)N4CCN(CC4)CCC5=CC(=C(C=C5)F)F)C(=O)NC6=CN=CC=C6)C1</chem>	Entraînement	Calibration	Entraînement	1.119
7	10046544	<chem>C1CCN2C(=C(C3=C2C(=NC=N3)N4CCN(CC4)CC(=O)C5=CC=CC=C5)C#N)C1</chem>	Entraînement	Entraînement	Entraînement	-0.267
8	10047863	<chem>C1CCN2C(=C(C3=C2C(=NC=N3)N4CCN(CC4)CCC5=CC(=C(C=C5)F)F)C#N)C1</chem>	Entraînement	Entraînement	Entraînement	0.64
9	10051663	<chem>C1CCN2C(=C(C3=C2C(=NC(=N3)C4=CN=CC=C4)N5CCN(CC5)CCC6=CC(=C(C=C6)F)F)C#N)C1</chem>	Entraînement	Entraînement	Entraînement	0.662
10	10071405	<chem>C1CCN2C(=C(C3=C2C(=NC(=N3)N4CCN(CC4)CCC5=CC(=CC=C5)[N+](=O)[O-])C#N)C1</chem>	Entraînement	Entraînement	Entraînement	0.387

11	10092403	C1CCN2C(=C(C3=C2C(=NC=N3)N4CCN(CC4)CCCC5=C C=CC=C5)C#N)C1	Calibration	Calibration	Entrainement	-0.866
12	10094280	C1CCN2C(=C(C3=C2C(=NC=N3)N4CCN(CC4)CCC5=CC =C(C=C5)[N+](=O)[O-])C#N)C1	Entrainement	Test	Entrainement	0.281
13	10322116	C1CCN2C(=C(C3=C2C(=NC(=N3)C4=CC=CC=C4)N5CCN (CC5)CCC6=CC=CC=C6)C# N)C1	Entrainement	Entrainement	Entrainement	0.146
14	10344755	CC1=C(SC(=N1)C2=C3CCCC N3C4=C2N=CN=C4N5CCN(CC5)CCC6=CC=CC=C6)C	Entrainement	Entrainement	Entrainement	0.273
15	10346312	C1CCN2C(=C(C3=C2C(=NC(=N3)I)N4CCN(CC4)CCC5=C C=CC=C5)C#N)C1	Entrainement	Test	Calibration	0.052
16	10388318	C1CCN2C(=C(C3=C2C(=NC=N3)N4CCN(CC4)S(=O)(=O)CC5=CC=CC=C5)C#N)C1	Calibration	Entrainement	Test	-0.823
17	10407830	C1CCN2C(=C(C3=C2C(=NC=N3)N4CCN(CC4)CCC5=CC =CC=C5)C#N)C1	Entrainement	Entrainement	Entrainement	0.161
18	10432300	C1CCN2C(=C(C3=C2C(=NC=N3)N4CCN(CC4)CCC5=CC (=CC=C5)Cl)C#N)C1	Entrainement	Entrainement	Entrainement	0.036
19	10454772	COC(=O)C1=C2CCCCN2C3= C1N=CN=C3N4CCN(CC4)C CC5=CC=CC=C5	Test	Test	Entrainement	0.041
20	10455921	C1CCN2C(=C(C3=C2C(=NC=N3)N4CCN(CC4)CCC5=CC (=C(C=C5)F)F)C(=O)N)C1	Test	Test	Test	0.86
21	10459316	CC(=O)NC1=CC=C(C=C1)C2 =NC3=C(C(=N2)N4CCN(CC 4)CCC5=CC=CC=C5)N6CCC CC6=C3C#N	Entrainement	Entrainement	Test	0.244
22	11164648	C1CCN2C(=C(C3=C2C(=NC=N3)NCCN4CCC5=CC=CC= C5C4)C#N)C1	Calibration	Test	Entrainement	-0.803

23	11177641	C1CCN2C(=C(C3=C2C(=NC =N3)N4CCN(CC4)CCC5=CC =C(C=C5)Cl)C#N)C1	Calibration	Entrainement	Entrainement	0.399
24	11186582	CN(CCC1=CC=CC=C1)C2=N C=NC3=C2N4CCCC4=C3C #N	Entrainement	Test	Entrainement	-0.954
25	11189272	CCC(=O)C1=C2CCCCN2C3= C1N=CN=C3N4CCN(CC4)C CC5=CC=CC=C5	Entrainement	Calibration	Entrainement	-0.447
26	11258566	COC1=CC=CC(=C1)CCN2CC N(CC2)C3=NC=NC4=C3N5 CCCC5=C4C#N	Entrainement	Entrainement	Calibration	-0.362
27	11337594	C1CCN2C(=C(C3=C2C(=NC =N3)NS(=O)(=O)CC4=CC=C C=C4)C#N)C1	Entrainement	Test	Entrainement	-1.32
28	11351150	C1CCN2C(=C(C3=C2C(=NC =N3)N4CCN(CC4)CCC5=CC =CC=C5)C6=NC=CC=N6)C1	Test	Calibration	Calibration	0.171
29	11359019	C1CCN2C(=C(C3=C2C(=NC =N3)NCCC4=CC=CC=C4)C# N)C1	Test	Calibration	Entrainement	-0.799
30	11362459	C1CCN2C(=C(C3=C2C(=NC =N3)N4CCN(CC4)CCC5=CC =CC=C5[N+])(=O)[O-])C#N)C1	Entrainement	Test	Entrainement	-0.151
31	11408751	CCNC(=O)C1=C2CCCCN2C 3=C1N=CN=C3N4CCN(CC4)CCC5=CC(=C(C=C5)F)F	Entrainement	Entrainement	Entrainement	0.921
32	11418752	C1CCN2C(=C(C3=C2C(=NC =N3)N4CCN(CC4)CCC5=CC =CC=C5Cl)C#N)C1	Entrainement	Calibration	Entrainement	0.02
33	11432257	CC1=CC=CC=C1C2=NC3=C(C(=N2)N4CCN(CC4)CCC5= CC(=C(C=C5)F)F)N6CCCC 6=C3C#N	Entrainement	Test	Calibration	0.208
34	11444700	COC1=CC=CC(=C1OC)CNC(=O)C2=C3CCCCN3C4=C2N =CN=C4N5CCN(CC5)CCC6= CC(=C(C=C6)F)F	Entrainement	Entrainement	Entrainement	1.367
35	11454819	C1CCN2C(=C(C3=C2C(=NC(=N3)C4=CC=NC=C4)N5CC	Test	Test	Entrainement	0.979

		<chem>N(CC5)CCC6=CC(=C(C=C6)F)C#N)C1</chem>				
36	10031184	<chem>C1CN(CCN1CCC2=CC(=C(C=C2)F)F)C3=NC(=NC4=C3NC5=C(C=C(C=C54)[N+](=O)[O-])[N+](=O)[O-])CN6CCOCC6</chem>	Calibration	Entrainement	Entrainement	1.721
37	10320818	<chem>C1CN(CCN1CCC2=CC(=C(C=C2)F)F)C3=NC=NC4=C3NC5=CC=CC(=C54)C(=O)N</chem>	Entrainement	Entrainement	Entrainement	1.187
38	11156535	<chem>C1CN(CCN1CCC2=CC(=C(C=C2)F)F)C3=NC=NC4=C3NC5=CC=CC(=C54)C(=O)NC6=CC=NC=C6</chem>	Calibration	Entrainement	Entrainement	0.15
39	11168500	<chem>C1CN(CCN1CCC2=CC(=C(C=C2)F)F)C3=NC=NC4=C3N(C5=CC=CC(=C54)C(=O)N)C6=CC=NC=C6</chem>	Calibration	Entrainement	Entrainement	1.699
40	11177283	<chem>CN1C2=C(C3=C1C(=NC=N3)N4CCN(CC4)CCC5=CC(=C(C=C5)F)F)N=CC=C2</chem>	Entrainement	Entrainement	Test	1.027
41	11179994	<chem>CCOC(=O)CN1C2=C(C=C(C=C2)[N+](=O)[O-])C3=C1C(=NC=N3)N4CCN(CC4)CCC5=CC(=C(C=C5)F)F</chem>	Test	Calibration	Entrainement	1.108
42	11200621	<chem>CC1=C2C(=CC=C1)NC3=C2N=CN=C3N4CCN(CC4)CCC5=CC(=C(C=C5)F)F</chem>	Calibration	Entrainement	Entrainement	0.199
43	11201011	<chem>CC1=NC=CC2=C1C3=C(N2)C(=NC=N3)N4CCN(CC4)CCC5=CC(=C(C=C5)F)F</chem>	Entrainement	Entrainement	Entrainement	1.678
44	11212921	<chem>CN1C2=C(C3=C1C(=NC=N3)N4CCN(CC4)CCC5=CC(=C(C=C5)F)F)C(=NC=C2)Cl</chem>	Entrainement	Entrainement	Test	1.387
45	11213871	<chem>C1CN(CCN1CCC2=CC(=C(C=C2)F)F)C3=NC=NC4=C3N(C5=C4C=NC=C5)CC6=CC=C(C=C6)</chem>	Entrainement	Test	Entrainement	1.319
46	11214464	<chem>CN1C2=C(C3=C1C(=NC=N3)N4CCN(CC4)CCC5=CC(=C(C=C5)F)F)N=CC=C2</chem>	Entrainement	Calibration	Entrainement	1.824

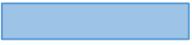
		C=C5)F)F)C(=NC=C2)NCCC6 =CC=CC=N6				
47	11226516	CN1C2=C(C3=C1C(=NC(=N3)CN4CCOCC4)N5CCN(CC5)CCC6=CC(=C(C=C6)F)F)C(=NC=C2)Cl	Entrainement	Entrainement	Entrainement	1.77
48	11246811	C1CN(CCN1CCC2=CC(=C(C=C2)F)F)C3=NC=NC4=C3NC5=C4C(=CC=C5)N	Entrainement	Entrainement	Calibration	0.15
49	11247947	COC(=O)C1=CC2=C(C=C1)NC3=C2N=CN=C3N4CCN(C4)CCC5=CC(=C(C=C5)F)F	Entrainement	Entrainement	Entrainement	1.229
50	11249529	C1CN(CCN1CCC2=CC(=C(C=C2)F)F)C3=NC=NC4=C3N(C5=C4C(=CC=C5)[N+](=O)[O-])CC6=CC=NC=C6	Calibration	Entrainement	Entrainement	1.863
51	11261262	C1CN(CCN1CCC2=CC(=C(C=C2)F)F)C3=NC=NC4=C3N(C5=C4C(=CC=C5)[N+](=O)[O-])CCC6=CC=CC=C6	Test	Entrainement	Test	1.388
52	11270541	CNC1=NC=CC2=C1C3=C(N2C)C(=NC=N3)N4CCN(CC4)CCC5=CC(=C(C=C5)F)F	Entrainement	Entrainement	Test	1.553
53	11270563	C1CN(CCN1CCC2=CC(=C(C=C2)F)F)C3=NC=NC4=C3NC5=C4C(=CC=C5)[N+](=O)[O-]	Entrainement	Entrainement	Entrainement	1.229
54	11293906	CN1C2=C(C3=C(C=CN=C31)C(=O)N)N=CN=C2N4CCN(CC4)CCC5=CC(=C(C=C5)F)F	Test	Entrainement	Entrainement	1.921
55	11293909	CN1C2=C(C3=C1C(=NC=N3)N4CCN(CC4)CCC5=CC(=C(C=C5)F)F)C(=NC=C2)N(C)C	Entrainement	Entrainement	Test	1.921
56	11293927	CN1C2=C(C(=CC=C2)[N+](=O)[O-])C3=C1C(=NC=N3)N4CCN(CC4)CCC5=CC(=C(C=C5)F)F	Entrainement	Calibration	Entrainement	0.636
57	11295806	CN1C2=C(C3=C(C=CN=C31)C(=O)N)N=C(N=C2N4CCN(CC4)CCC5=CC(=C(C=C5)F)F)CN6CCOCC6	Entrainement	Entrainement	Entrainement	2

58	11304370	CN1C2=C(C=NC=C2)C3=C1 C(=NC=N3)N4CCN(CC4)CC C5=CC(=C(C=C5)F)F	Entrainement	Entrainement	Calibration	1.62
59	11305058	CN1C2=C(C3=C(C=CN=C31)C#N)N=CN=C2N4CCN(CC4)CCC5=CC(=C(C=C5)F)F	Entrainement	Entrainement	Entrainement	1.347
60	11316308	C1CN(CCN1CCC2=CC(=C(C =C2)F)F)C3=NC=NC4=C3N C5=C4C(=CC=C5)Cl	Entrainement	Entrainement	Entrainement	0.777
61	11318194	CC1=C(C(=C2C3=C(C(=NC= N3)N4CCN(CC4)CCC5=CC(=C(C=C5)F)F)N(C2=N1)C)C(=O)N)[N+](=O)[O-]	Entrainement	Calibration	Test	1.959
62	11339085	C1CN(CCN1CCC2=CC(=C(C =C2)F)F)C3=NC=NC4=C3N C5=CC=CC(=C54)C#N	Entrainement	Test	Entrainement	0.772
63	11350733	COC1=CC2=C(C=C1)NC3=C 2N=CN=C3N4CCN(CC4)CC C5=CC(=C(C=C5)F)F	Entrainement	Entrainement	Calibration	0.216
64	11376266	CN1C2=CC=CC(=C2C3=C1C (=NC(=N3)CN4CCOCC4)N5 CCN(CC5)CCC6=CC(=C(C=C 6)F)F)C(=O)N	Calibration	Calibration	Entrainement	1.538
65	11385818	COC1=C(C2=C(C=C1)NC3= C2N=CN=C3N4CCN(CC4)C CC5=CC(=C(C=C5)F)F)[N+](=O)[O-]	Entrainement	Entrainement	Entrainement	0.807
66	11396606	CN1C2=CC(=NC=C2C3=C1 C(=NC=N3)N4CCN(CC4)CC C5=CC(=C(C=C5)F)F)Cl	Entrainement	Entrainement	Entrainement	1.06
67	11419244	CN1C2=C(C3=C1C(=NC=N3)N4CCN(CC4)CCC5=CC(=C(C=C5)F)F)C(=NC=C2)OC	Test	Entrainement	Entrainement	1.201
68	11443156	CN1C2=C(C3=C1C(=NC(=N 3)COC)N4CCN(CC4)CCC5= CC(=C(C=C5)F)F)C(=NC=C2)Cl	Entrainement	Entrainement	Entrainement	1.886
69	11443942	C1CN(CCN1CCC2=CC(=C(C =C2)F)F)C3=NC=NC4=C3N(C5=C4C(=CC=C5)[N+](=O)[O-])CC6=CN=CC=C6	Calibration	Entrainement	Calibration	1.577

70	11453657	CC1=NC2=C(C(=C1)C#N)C3=C(N2C)C(=NC=N3)N4CCN(CC4)CCC5=CC(=C(C=C5)F)F	Test	Calibration	Calibration	0.721
71	11454703	CN1C2=C(C3=C1C(=NC=N3)N4CCN(CC4)CCC5=CC(=C(C=C5)F)F)C(=NC=C2)N6CCOCC6	Test	Entrainement	Test	2
72	11455462	C1CN(CCN1CCC2=CC(=C(C=C2)F)F)C3=NC(=NC4=C3NC5=C4C=C(C=C5)[N+](=O)[O-])CN6CCOCC6	Entrainement	Entrainement	Test	0.971
73	11488138	CC(C)OC1=NC=CC2=C1C3=C(N2C)C(=NC=N3)N4CCN(CC4)CCC5=CC(=C(C=C5)F)F	Test	Test	Test	1.569
74	11751643	CC1=NC2=C(C(=C1)C(=O)N)C3=C(N2C)C(=NC=N3)N4CCN(CC4)CCC5=CC(=C(C=C5)F)F	Entrainement	Entrainement	Entrainement	1.174
75	11752839	C1CCN(C1)S(=O)(=O)C2=C3=C(C=C2)NC4=C3N=CN=C4N5CCN(CC5)CCC6=CC(=C(C=C6)F)F	Calibration	Calibration	Entrainement	0.266
76	11774980	C1CN(CCN1CCC2=CC(=C(C=C2)F)F)C3=NC=NC4=C3NC5=C4C=C(C=C5)[N+](=O)[O-]	Entrainement	Entrainement	Entrainement	1.292
77	11775287	CN1C2=CC=CC(=C2C3=C1C(=NC=N3)N4CCN(CC4)CCC5=CC(=C(C=C5)F)F)C(=O)N	Calibration	Entrainement	Entrainement	1.62
78	44276937	CCOC(=O)N1C2=CC=CC(=C2C3=C1C(=NC=N3)N4CCN(CC4)CCC5=CC(=C(C=C5)F)F)C(=O)N	Entrainement	Entrainement	Calibration	1.347
79	917533	C1CCN2C(=C(C3=C2C(=NC=N3)NCC4=CC=CC=C4)C#N)C1	Test	Entrainement	Test	-1.152
80	9975676	C1CCN2C(=C(C3=C2C(=NC=N3)N4CCN(CC4)C5=CC=C(C=C5)C#N)C1	Entrainement	Entrainement	Entrainement	-0.408

81	10048669	C1CN(CCN1CCC2=CC=CC=C2)C3=NC=NC4=C3N(C=C4C#N)CCC5=CC=CC=C5	Entraînement	Entraînement	Entraînement	-0.009
82	10333981	C1CCN2C(=C(C3=C2C(=NC=N3)N4CCNCC4)C#N)C1	Entraînement	Calibration	Test	-1.436
83	10379829	CN1CCN(CC1)C2=NC=NC3=C2N4CCCC4=C3C#N	Entraînement	Test	Calibration	-0.704
84	10383921	CCN1C=C(C2=C1C(=NC=N2)N3CCN(CC3)CCC4=CC=CC=C4)C#N	Entraînement	Calibration	Entraînement	-0.025
85	10407830	C1CCN2C(=C(C3=C2C(=NC=N3)N4CCN(CC4)CCC5=CC=CC=C5)C#N)C1	Entraînement	Entraînement	Entraînement	0.432
86	10472817	CN1C=C(C2=C1C(=NC=N2)N3CCN(CC3)CCC4=CC=CC=C4)C#N	Entraînement	Entraînement	Entraînement	0.146
87	11359019	C1CCN2C(=C(C3=C2C(=NC=N3)NCCC4=CC=CC=C4)C#N)C1	Entraînement	Test	Calibration	-0.923
88	127048170	C1CCCC(CC1)C2=C(C3=C(N2C4CCCCC4)C(=NC=N3)N5CCN(CC5)C(C6=CC=CC=C6)C7=CC=CC=C7)C#N	Entraînement	Test	Entraînement	-0.29
89	127048171	C1CCCC(CC1)C2=C(C3=C(N2C4CCCCC4)C(=NC=N3)N5CCN(CC5)C6=CC=CC=C6)C#N	Calibration	Entraînement	Entraînement	-0.953
90	127049502	C1CCN2C(=C(C3=C2C(=NC=N3)N4CCN(CC4)CC5=CC=CC=C5)C#N)C1	Entraînement	Test	Entraînement	0.173
91	127049508	C1CCN2C(=C(C3=C2C(=NC=N3)N4CCN(CC4)C(C5=CC=CC=C5)C6=CC=CC=C6)C#N)C1	Entraînement	Entraînement	Entraînement	-0.21
92	127049509	C1CCN2C(=C(C3=C2C(=NC=N3)NCCCC4=CC=CC=C4)C#N)C1	Entraînement	Entraînement	Calibration	-0.398
93	127050480	C1CN(CCN1CCC2=CC=CC=C2)C3=NC=NC4=C3N(C=C4C#N)CCC5=CC=CC=C5	Entraînement	Entraînement	Entraînement	0

94	127050481	C1CN(CCN1CC2=CC=CC=C2)C3=NC=NC4=C3N(C=C4C#N)CC5=CC=CC=C5	Entrainement	Entrainement	Entrainement	-0.442
95	127050796	C1CN(CCN1CC2=CC=CC=C2)C3=NC=NC4=C3N(C=C4C#N)CCC5=CC=CC=C5	Entrainement	Entrainement	Entrainement	0.041
96	127050797	C1CN(CCN1CCC2=CC=CC=C2)C3=NC=NC4=C3N(C=C4C#N)CCCC5=CC=CC=C5	Entrainement	Entrainement	Calibration	0.02
97	127050798	C1CN(CCN1CC2=CC=CC=C2)C3=NC=NC4=C3N(C=C4C#N)CCCC5=CC=CC=C5	Entrainement	Entrainement	Entrainement	-0.498
98	127051981	CCCN1C=C(C2=C1C(=NC=N2)N3CCN(CC3)CCC4=CC=C(C=C4)C#N	Entrainement	Entrainement	Entrainement	0.265
99	127051982	CC(C)N1C=C(C2=C1C(=NC=N2)N3CCN(CC3)CCC4=CC=CC=C4)C#N	Entrainement	Entrainement	Entrainement	0.086
100	127051983	C1CC1N2C=C(C3=C2C(=NC=N3)N4CCN(CC4)CCC5=CC=CC=C5)C#N	Entrainement	Entrainement	Entrainement	0.607
101	127051984	C1CN(CCN1CCC2=CC=CC=C2)C3=NC=NC4=C3N(C=C4C#N)C5=CC=CC=C5	Test	Calibration	Entrainement	0.125
102	127051985	C1CN(CCN1CC2=CC=CC=C2)C3=NC=NC4=C3N(C=C4C#N)C5=CC=CC=C5	Test	Entrainement	Entrainement	-0.086
103	127052887	C1CCN2C(=C(C3=C2C(=NC=N3)NC4=CC=CC=C4)C#N)C1	Entrainement	Entrainement	Calibration	-1.152
104	127052888	C1CCCC(CC1)C2=C(C3=C(N2C4CCCCC4)C(=NC=N3)N5CCN(CC5)CCC6=CC=CC=C6)C#N	Entrainement	Entrainement	Entrainement	0.079
105	127052889	C1CCCC(CC1)C2=C(C3=C(N2C4CCCCC4)C(=NC=N3)N5CCN(CC5)CC6=CC=CC=C6)C#N	Calibration	Entrainement	Entrainement	-0.025



The background features a light blue gradient with various molecular models. At the top, there are several ball-and-stick molecular structures in shades of blue and grey. At the bottom, a prominent blue DNA double helix is visible, extending across the width of the page. The overall aesthetic is scientific and clean.

***Liste des références
bibliographique***

Références bibliographiques:

- [1] W. IB, “No Title,” “*The Orig. Hum. cancer Mol. Mech. Carcinog. their Implic. cancer Prev. Treat. G.H.A. Clowes Meml. Award Lect.*,” no. Cancer research, 1988.
- [2] L. M. Break, M. A. M. Mohamed, O. A. A. Al-Thubaiti, and F. E. M. Eibaih, “Nucleosides 10: Synthesis of New Derivatives of Pyrimidine and Fused Pyrimidine Nucleosides of Expected Biological Activity,” *Int. J. Org. Chem.*, vol. 09, no. 03, pp. 107–120, 2019, doi: 10.4236/ijoc.2019.93010.
- [3] F. Malerba, L. Orsenigo, and S. Selosse, “Biotechnologies et industrie pharmaceutique,” *Rev. l’OFCE*, vol. no96, no. 1, pp. 157–209, Feb. 2006, doi: 10.3917/reof.096.0157.
- [4] D. Rognan and P. Bonnet, “Les chimiothèques et le criblage virtuel,” *médecine/sciences*, vol. 30, no. 12, pp. 1152–1160, Dec. 2014, doi: 10.1051/medsci/20143012019.
- [5] R. A. Weinberg, *The Biology of Cancer*. W.W. Norton & Company, 2013. doi: 10.1201/9780429258794.
- [6] S. McGuire, “World Cancer Report 2014. Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, WHO Press, 2015,” *Adv. Nutr.*, vol. 7, no. 2, pp. 418–419, Mar. 2016, doi: 10.3945/an.116.012211.
- [7] J. Ferlay *et al.*, “Cancer statistics for the year 2020: An overview,” *Int. J. Cancer*, vol. 149, no. 4, pp. 778–789, Aug. 2021, doi: 10.1002/ijc.33588.
- [8] F. Binder-Foucard, N. Bossard, P. Delafosse, A. Belot, A.-S. Woronoff, and L. Remontet, “Cancer incidence and mortality in France over the 1980–2012 period: Solid tumors,” *Rev. Epidemiol. Sante Publique*, vol. 62, no. 2, pp. 95–108, Apr. 2014, doi: 10.1016/j.respe.2013.11.073.
- [9] S. Le Guyader-Peyrou *et al.*, “Cancer incidence in France over the 1980–2012 period: Hematological malignancies,” *Rev. Epidemiol. Sante Publique*, vol. 64, no. 2, pp. 103–112, Apr. 2016, doi: 10.1016/j.respe.2015.12.017.
- [10] É. Terrat, “Traitements des cancers pulmonaires,” *L’Aide-Soignante*, vol. 27, no. 146, pp. 15–17, 2013, doi: 10.1016/j.aidsoi.2013.02.007.
- [11] D. Z. J. Chu, “Re: Breast Cancer Incidence, 1980-2006: Combined Roles of Menopausal Hormone Therapy, Screening Mammography, and Estrogen Receptor Status,” *JNCI J. Natl. Cancer Inst.*, vol. 100, no. 8, pp. 596–597, Apr. 2008, doi: 10.1093/jnci/djn077.
- [12] A. Antoniou *et al.*, “Average Risks of Breast and Ovarian Cancer Associated with BRCA1 or BRCA2 Mutations Detected in Case Series Unselected for Family History: A Combined Analysis of 22 Studies,” *Am. J. Hum. Genet.*, vol. 72, no. 5, pp. 1117–1130, May 2003, doi: 10.1086/375033.
- [13] N. Mavaddat, A. C. Antoniou, D. F. Easton, and M. Garcia-Closas, “Genetic susceptibility to breast cancer,” *Mol. Oncol.*, vol. 4, no. 3, pp. 174–191, Jun. 2010, doi: 10.1016/j.molonc.2010.04.011.
- [14] R. J. Santen and R. Mansel, “Benign Breast Disorders,” *N. Engl. J. Med.*, vol. 353, no.

- 3, pp. 275–285, Jul. 2005, doi: 10.1056/NEJMra035692.
- [15] A. Cowppli-Bony *et al.*, “Survie des personnes atteintes de cancers solides en France métropolitaine, 1989–2013,” *Rev. Epidemiol. Sante Publique*, vol. 64, pp. S201–S202, Sep. 2016, doi: 10.1016/j.respe.2016.06.084.
- [16] U. Claude and B. Lyon, “One-carbon metabolism and smoking related cancers Anouar Fanidi To cite this version : HAL Id : tel-01383446,” 2016.
- [17] P. M. Bedano and N. H. Hanna, “Salvage Therapy in Patients with Advanced Non-small Cell Lung Cancer,” *J. Thorac. Oncol.*, vol. 1, no. 6, pp. 582–587, Jul. 2006, doi: 10.1016/S1556-0864(15)30364-6.
- [18] E. Shtivelman *et al.*, “Molecular pathways and therapeutic targets in lung cancer,” *Oncotarget*, vol. 5, no. 6, pp. 1392–1433, Mar. 2014, doi: 10.18632/oncotarget.1891.
- [19] Institut National du Cancer, “Les traitements des cancers du poumon - Guide patients,” 2017.
- [20] D. E. S. Réponses and À. V. O. S. Questions, “Et Comment Me Faire Aider ?”.
- [21] R. Wu *et al.*, “Epigenetics/epigenomics and prevention by curcumin of early stages of inflammatory-driven colon cancer,” *Mol. Carcinog.*, vol. 59, no. 2, pp. 227–236, Feb. 2020, doi: 10.1002/mc.23146.
- [22] M. M. Center, A. Jemal, and E. Ward, “International Trends in Colorectal Cancer Incidence Rates,” *Cancer Epidemiol. Biomarkers Prev.*, vol. 18, no. 6, pp. 1688–1694, Jun. 2009, doi: 10.1158/1055-9965.EPI-09-0090.
- [23] M. Song, A. T. Chan, and J. Sun, “Influence of the Gut Microbiome, Diet, and Environment on Risk of Colorectal Cancer,” *Gastroenterology*, vol. 158, no. 2, pp. 322–340, Jan. 2020, doi: 10.1053/j.gastro.2019.06.048.
- [24] H. Ueno *et al.*, “Risk factors for an adverse outcome in early invasive colorectal carcinoma,” *Gastroenterology*, vol. 127, no. 2, pp. 385–394, Aug. 2004, doi: 10.1053/j.gastro.2004.04.022.
- [25] J. Viguier, P. J. Bousquet, D. Bessette, F. de Bels, C. Ferrari, and C. Morin, “52. Cancer et politique de lutte contre le cancer,” in *Traité de santé publique*, Lavoisier, 2016, p. 490. doi: 10.3917/lav.bourd.2016.01.0515.
- [26] F. Hajiaghaalipour, M. S. Kanthimathi, J. Sanusi, and J. Rajarajeswaran, “White tea (*Camellia sinensis*) inhibits proliferation of the colon cancer cell line, HT-29, activates caspases and protects DNA of normal cells against oxidative damage,” *Food Chem.*, vol. 169, pp. 401–410, Feb. 2015, doi: 10.1016/j.foodchem.2014.07.005.
- [27] M. E. Sanders, “Probiotics: Considerations for Human Health,” *Nutr. Rev.*, vol. 61, no. 3, pp. 91–99, Mar. 2003, doi: 10.1301/nr.2003.marr.91-99.
- [28] A. Górška, D. Przystupski, M. J. Niemczura, and J. Kulbacka, “Probiotic Bacteria: A Promising Tool in Cancer Prevention and Therapy,” *Curr. Microbiol.*, vol. 76, no. 8, pp. 939–949, Aug. 2019, doi: 10.1007/s00284-019-01679-8.
- [29] B. C. S. Cruz, M. M. Sarandy, A. C. Messias, R. V. Gonçalves, C. L. L. F. Ferreira, and M. C. G. Peluzio, “Preclinical and clinical relevance of probiotics and synbiotics in colorectal carcinogenesis: a systematic review,” *Nutr. Rev.*, vol. 78, no. 8, pp. 667–687, Aug. 2020, doi: 10.1093/nutrit/nuz087.

- [30] P.-J. Lamy, "Cancer de la prostate : les niveaux de preuve des biomarqueurs de la détection précoce," *Médecine Nucléaire*, vol. 38, no. 1, pp. 14–17, Feb. 2014, doi: 10.1016/j.mednuc.2013.12.007.
- [31] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA. Cancer J. Clin.*, vol. 68, no. 6, pp. 394–424, Nov. 2018, doi: 10.3322/caac.21492.
- [32] I. M. Thompson *et al.*, "Prevalence of Prostate Cancer among Men with a Prostate-Specific Antigen Level ≤ 4.0 ng per Milliliter," *N. Engl. J. Med.*, vol. 350, no. 22, pp. 2239–2246, May 2004, doi: 10.1056/NEJMoa031918.
- [33] D. G. Bostwick, J. Qian, and C. Schlesinger, "Contemporary pathology of prostate cancer," *Urol. Clin. North Am.*, vol. 30, no. 2, pp. 181–207, May 2003, doi: 10.1016/S0094-0143(02)00189-1.
- [34] D. E. Salazar and G. Gormley, "Modern Drug Discovery and Development," in *Clinical and Translational Science*, Elsevier, 2017, pp. 719–743. doi: 10.1016/B978-0-12-802101-9.00041-7.
- [35] J. Hughes, S. Rees, S. Kalindjian, and K. Philpott, "Principles of early drug discovery," *Br. J. Pharmacol.*, vol. 162, no. 6, pp. 1239–1249, Mar. 2011, doi: 10.1111/j.1476-5381.2010.01127.x.
- [36] "No Title," "*La recherche cognitive, aux origines de tout | LEEM - Les entreprises du médicament.*," 2016.
- [37] "No Title," "*Pharma: Biopharmaceutical Research&Development: The process behind medicines.*," 2016.
- [38] Y. Asses, "Conception par modélisation et criblage in silico d' inhibiteurs du récepteur c-Met.," p. 142, 2011.
- [39] S. K. Hanks and T. Hunter, "The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification 1," *FASEB J.*, vol. 9, no. 8, pp. 576–596, May 1995, doi: 10.1096/fasebj.9.8.7768349.
- [40] S.-Y. Yang, "Pharmacophore modeling and applications in drug discovery: challenges and recent advances," *Drug Discov. Today*, vol. 15, no. 11–12, pp. 444–450, Jun. 2010, doi: 10.1016/j.drudis.2010.03.013.
- [41] O. F. Güner, "Pharmacophore perception, development, and use in drug design," 2000.
- [42] O. Dror, A. Shulman-Peleg, R. Nussinov, and H. Wolfson, "Predicting Molecular Interactions in silico: I. A Guide to Pharmacophore Identification and its Applications to Drug Design," *Curr. Med. Chem.*, vol. 11, no. 1, pp. 71–90, Jan. 2004, doi: 10.2174/0929867043456287.
- [43] S. Handschuh, M. Wagener, and J. Gasteiger, "Superposition of Three-Dimensional Chemical Structures Allowing for Conformational Flexibility by a Hybrid Method," *J. Chem. Inf. Comput. Sci.*, vol. 38, no. 2, pp. 220–232, Mar. 1998, doi: 10.1021/ci970438r.
- [44] A. Vuorinen and D. Schuster, "Methods for generating and applying pharmacophore models as virtual screening filters and for bioactivity profiling," *Methods*, vol. 71, pp.

- 113–134, Jan. 2015, doi: 10.1016/j.ymeth.2014.10.013.
- [45] S. P. Leelananda and S. Lindert, “Computational methods in drug discovery,” *Beilstein J. Org. Chem.*, vol. 12, pp. 2694–2718, Dec. 2016, doi: 10.3762/bjoc.12.267.
- [46] H.-J. Böhm, “The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure,” *J. Comput. Aided. Mol. Des.*, vol. 8, no. 3, pp. 243–256, Jun. 1994, doi: 10.1007/BF00126743.
- [47] H. A. Carlson *et al.*, “Developing a Dynamic Pharmacophore Model for HIV-1 Integrase,” *J. Med. Chem.*, vol. 43, no. 11, pp. 2100–2114, Jun. 2000, doi: 10.1021/jm990322h.
- [48] G. Schneider and U. Fechner, “Computer-based de novo design of drug-like molecules,” *Nat. Rev. Drug Discov.*, vol. 4, no. 8, pp. 649–663, Aug. 2005, doi: 10.1038/nrd1799.
- [49] M. P. A. Sanders *et al.*, “From the protein’s perspective: the benefits and challenges of protein structure-based pharmacophore modeling,” *Medchemcomm*, vol. 3, no. 1, pp. 28–38, 2012, doi: 10.1039/C1MD00210D.
- [50] R. L. Juliano and V. Ling, “A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants,” *Biochim. Biophys. Acta - Biomembr.*, vol. 455, no. 1, pp. 152–162, Nov. 1976, doi: 10.1016/0005-2736(76)90160-7.
- [51] N. M. Kuderer, D. C. Dale, J. Crawford, and G. H. Lyman, “Impact of Primary Prophylaxis With Granulocyte Colony-Stimulating Factor on Febrile Neutropenia and Mortality in Adult Cancer Patients Receiving Chemotherapy: A Systematic Review,” *J. Clin. Oncol.*, vol. 25, no. 21, pp. 3158–3167, Jul. 2007, doi: 10.1200/JCO.2006.08.8823.
- [52] H. Hamada and T. Tsuruo, “Functional role for the 170- to 180-kDa glycoprotein specific to drug-resistant tumor cells as revealed by monoclonal antibodies,” *Proc. Natl. Acad. Sci.*, vol. 83, no. 20, pp. 7785–7789, Oct. 1986, doi: 10.1073/pnas.83.20.7785.
- [53] D. R. Gill, S. C. Hyde, C. F. Higgins, M. A. Valverde, G. M. Mintenig, and F. V. Sepúlveda, “Separation of drug transport and chloride channel functions of the human multidrug resistance P-glycoprotein,” *Cell*, vol. 71, no. 1, pp. 23–32, Oct. 1992, doi: 10.1016/0092-8674(92)90263-C.
- [54] J. M. Walker, *MMB vol.596 Zhou J. (ed.) Multi-Drug Resistance in Cancer (Humana Press, 2009)(ISBN 1607614154)(492s)*. 2009.
- [55] M. M. Gottesman, T. Fojo, and S. E. Bates, “Multidrug resistance in cancer: role of ATP-dependent transporters,” *Nat. Rev. Cancer*, vol. 2, no. 1, pp. 48–58, Jan. 2002, doi: 10.1038/nrc706.
- [56] L. A. Doyle *et al.*, “A multidrug resistance transporter from human MCF-7 breast cancer cells,” *Proc. Natl. Acad. Sci.*, vol. 95, no. 26, pp. 15665–15670, Dec. 1998, doi: 10.1073/pnas.95.26.15665.
- [57] S. Cole *et al.*, “Overexpression of a transporter gene in a multidrug-resistant human lung cancer cell line,” *Science (80-)*, vol. 258, no. 5088, pp. 1650–1654, Dec. 1992, doi: 10.1126/science.1360704.

- [58] O. Lewinson and N. Livnat-Levanon, “Mechanism of Action of ABC Importers: Conservation, Divergence, and Physiological Adaptations,” *J. Mol. Biol.*, vol. 429, no. 5, pp. 606–619, Mar. 2017, doi: 10.1016/j.jmb.2017.01.010.
- [59] F. Kobayashi, C. A. Ross, and J. R. P. Ross, “Thailandina and Neothailandina, and their subfamily Thailandininae: an example of an invalid taxonomic group of Permian fusuline foraminifera,” *J. Paleontol.*, vol. 84, no. 2, pp. 360–361, Mar. 2010, doi: 10.1666/09-065R1.1.
- [60] S. E. L. Mirski, J. H. Gerlach, and S. P. C. Cole, “Multidrug Resistance in a Human Small Cell Lung Cancer Cell Line Selected in Adriamycin1,” *Cancer Res.*, vol. 47, no. 10, pp. 2594–2598, May 1987.
- [61] S. P. C. Cole, “Patterns of cross-resistance in a multidrug-resistant small-cell lung carcinoma cell line,” *Cancer Chemother. Pharmacol.*, vol. 26, no. 4, pp. 250–256, Jul. 1990, doi: 10.1007/BF02897225.
- [62] D. R. Hipfner, R. G. Deeley, and S. P. . Cole, “Structural, mechanistic and clinical aspects of MRP1,” *Biochim. Biophys. Acta - Biomembr.*, vol. 1461, no. 2, pp. 359–376, Dec. 1999, doi: 10.1016/S0005-2736(99)00168-6.
- [63] C. Kast and P. Gros, “Epitope Insertion Favors a Six Transmembrane Domain Model for the Carboxy-Terminal Portion of the Multidrug Resistance-Associated Protein,” *Biochemistry*, vol. 37, no. 8, pp. 2305–2313, Feb. 1998, doi: 10.1021/bi972332v.
- [64] S. P. C. Cole, “Targeting Multidrug Resistance Protein 1 (MRP1, ABCC1): Past, Present, and Future,” *Annu. Rev. Pharmacol. Toxicol.*, vol. 54, no. 1, pp. 95–117, Jan. 2014, doi: 10.1146/annurev-pharmtox-011613-135959.
- [65] M. K. DeGorter, G. Conseil, R. G. Deeley, R. L. Campbell, and S. P. C. Cole, “Molecular modeling of the human multidrug resistance protein 1 (MRP1/ABCC1),” *Biochem. Biophys. Res. Commun.*, vol. 365, no. 1, pp. 29–34, Jan. 2008, doi: 10.1016/j.bbrc.2007.10.141.
- [66] E. M. Leslie, R. G. Deeley, and S. P. C. Cole, “Multidrug resistance proteins: role of P-glycoprotein, MRP1, MRP2, and BCRP (ABCG2) in tissue defense,” *Toxicol. Appl. Pharmacol.*, vol. 204, no. 3, pp. 216–237, May 2005, doi: 10.1016/j.taap.2004.10.012.
- [67] C. Mercier, C. Masseguin, F. Roux, J. Gabrion, and J.-M. Scherrmann, “Expression of P-glycoprotein (ABCB1) and Mrp1 (ABCC1) in adult rat brain: focus on astrocytes,” *Brain Res.*, vol. 1021, no. 1, pp. 32–40, Sep. 2004, doi: 10.1016/j.brainres.2004.06.034.
- [68] H. Roelofsen *et al.*, “Increased levels of the multidrug resistance protein in lateral membranes of proliferating hepatocyte-derived cells,” *Gastroenterology*, vol. 112, no. 2, pp. 511–521, Feb. 1997, doi: 10.1053/gast.1997.v112.pm9024305.
- [69] K.-C. Peng *et al.*, “Tissue and Cell Distribution of the Multidrug Resistance-Associated Protein (MRP) in Mouse Intestine and Kidney,” *J. Histochem. Cytochem.*, vol. 47, no. 6, pp. 757–767, Jun. 1999, doi: 10.1177/002215549904700605.
- [70] D. E. Atkinson, S. L. Greenwood, C. P. Sibley, J. D. Glazier, and L. J. Fairbairn, “Role of MDR1 and MRP1 in trophoblast cells, elucidated using retroviral gene transfer,” *Am. J. Physiol. Physiol.*, vol. 285, no. 3, pp. C584–C591, Sep. 2003, doi: 10.1152/ajpcell.00418.2002.
- [71] M. Nagashige *et al.*, “Basal Membrane Localization of MRP1 in Human Placental

- Trophoblast,” *Placenta*, vol. 24, no. 10, pp. 951–958, Nov. 2003, doi: 10.1016/S0143-4004(03)00170-X.
- [72] E. DELANGE, “Potential role of ABC transporters as a detoxification system at the blood?CSF barrier,” *Adv. Drug Deliv. Rev.*, vol. 56, no. 12, pp. 1793–1809, Oct. 2004, doi: 10.1016/j.addr.2004.07.009.
- [73] J. Hirrlinger, J. König, D. Keppler, J. Lindenau, J. B. Schulz, and R. Dringen, “The multidrug resistance protein MRP1 mediates the release of glutathione disulfide from rat astrocytes during oxidative stress,” *J. Neurochem.*, vol. 76, no. 2, pp. 627–636, Jan. 2001, doi: 10.1046/j.1471-4159.2001.00101.x.
- [74] R. P. J. Oude Elferink, D. K. F. Meijer, F. Kuipers, P. L. M. Jansen, A. K. Groen, and G. M. M. Groothuis, “Hepatobiliary secretion of organic compounds; molecular mechanisms of membrane transport,” *Biochim. Biophys. Acta - Rev. Biomembr.*, vol. 1241, no. 2, pp. 215–268, Jul. 1995, doi: 10.1016/0304-4157(95)00006-D.
- [75] I. Hellström, K. E. Hellström, and P. D. Senter, “Development and Activities of the BR96-Doxorubicin Immunoconjugate,” in *Immunotoxin Methods and Protocols*, New Jersey: Humana Press, pp. 3–16. doi: 10.1385/1-59259-114-0:3.
- [76] W. Berger, L. Elbling, E. Hauptmann, and M. Micksche, “Expression of the multidrug resistance-associated protein (MRP) and chemoresistance of human non-small-cell lung cancer cells,” *Int. J. Cancer*, vol. 73, no. 1, pp. 84–93, Sep. 1997, doi: 10.1002/(SICI)1097-0215(19970926)73:1<84::AID-IJC14>3.0.CO;2-5.
- [77] D. W. Loe, K. C. Almquist, S. P. C. Cole, and R. G. Deeley, “ATP-dependent 17 β -Estradiol 17-(β -D-Glucuronide) Transport by Multidrug Resistance Protein (MRP),” *J. Biol. Chem.*, vol. 271, no. 16, pp. 9683–9689, Apr. 1996, doi: 10.1074/jbc.271.16.9683.
- [78] J. Renes, E. G. E. De Vries, E. F. Nienhuis, P. L. M. Jansen, and M. Müller, “ATP- and glutathione-dependent transport of chemotherapeutic drugs by the multidrug resistance protein MRP1,” *Br. J. Pharmacol.*, vol. 126, no. 3, pp. 681–688, Feb. 1999, doi: 10.1038/sj.bjp.0702360.
- [79] H. Coley, W. Amos, P. Twentyman, and P. Workman, “Examination by laser scanning confocal fluorescence imaging microscopy of the subcellular localisation of anthracyclines in parent and multidrug resistant cell lines,” *Br. J. Cancer*, vol. 67, no. 6, pp. 1316–1323, Jun. 1993, doi: 10.1038/bjc.1993.244.
- [80] M. Binaschi *et al.*, “MRP gene overexpression in a human doxorubicin-resistant SCLC cell line: Alterations in cellular pharmacokinetics and in pattern of cross-resistance,” *Int. J. Cancer*, vol. 62, no. 1, pp. 84–89, Jul. 1995, doi: 10.1002/ijc.2910620116.
- [81] H. W. van Veen, C. F. Higgins, and W. N. Konings, “Multidrug transport by ATP binding cassette transporters: a proposed two-cylinder engine mechanism,” *Res. Microbiol.*, vol. 152, no. 3–4, pp. 365–374, Apr. 2001, doi: 10.1016/S0923-2508(01)01208-6.
- [82] R. J. Raggars, A. van Helvoort, R. Evers, and G. van Meer, “The human multidrug resistance protein MRP1 translocates sphingolipid analogs across the plasma membrane,” *J. Cell Sci.*, vol. 112, no. 3, pp. 415–422, Feb. 1999, doi: 10.1242/jcs.112.3.415.

- [83] J. H. Hooijberg *et al.*, “Potent interaction of flavopiridol with MRP1,” *Br. J. Cancer*, vol. 81, no. 2, pp. 269–276, Sep. 1999, doi: 10.1038/sj.bjc.6690687.
- [84] M. Gao, D. W. Loe, C. E. Grant, S. P. C. Cole, and R. G. Deeley, “Reconstitution of ATP-dependent Leukotriene C4 Transport by Co-expression of Both Half-molecules of Human Multidrug Resistance Protein in Insect Cells,” *J. Biol. Chem.*, vol. 271, no. 44, pp. 27782–27787, Nov. 1996, doi: 10.1074/jbc.271.44.27782.
- [85] E. Leslie, “Toxicological relevance of the multidrug resistance protein 1, MRP1 (ABCC1) and related transporters,” *Toxicology*, vol. 167, no. 1, pp. 3–23, Oct. 2001, doi: 10.1016/S0300-483X(01)00454-1.
- [86] J. Zaitseva, S. Jenewein, T. Jumpertz, I. B. Holland, and L. Schmitt, “H662 is the linchpin of ATP hydrolysis in the nucleotide-binding domain of the ABC transporter HlyB,” *EMBO J.*, vol. 24, no. 11, pp. 1901–1910, Jun. 2005, doi: 10.1038/sj.emboj.7600657.
- [87] K. Hollenstein, R. J. Dawson, and K. P. Locher, “Structure and mechanism of ABC transporter proteins,” *Curr. Opin. Struct. Biol.*, vol. 17, no. 4, pp. 412–418, Aug. 2007, doi: 10.1016/j.sbi.2007.07.003.
- [88] L. Manciu *et al.*, “Intermediate Structural States Involved in MRP1-mediated Drug Transport,” *J. Biol. Chem.*, vol. 278, no. 5, pp. 3347–3356, Jan. 2003, doi: 10.1074/jbc.M207963200.
- [89] P. C. Smith *et al.*, “ATP Binding to the Motor Domain from an ABC Transporter Drives Formation of a Nucleotide Sandwich Dimer,” *Mol. Cell*, vol. 10, no. 1, pp. 139–149, Jul. 2002, doi: 10.1016/S1097-2765(02)00576-2.
- [90] R. G. Deeley and S. P. C. Cole, “Substrate recognition and transport by multidrug resistance protein 1 (ABCC1),” *FEBS Lett.*, vol. 580, no. 4, pp. 1103–1111, Feb. 2006, doi: 10.1016/j.febslet.2005.12.036.
- [91] A. Rothnie, R. Callaghan, R. G. Deeley, and S. P. C. Cole, “Role of GSH in Estrone Sulfate Binding and Translocation by the Multidrug Resistance Protein 1 (MRP1/ABCC1),” *J. Biol. Chem.*, vol. 281, no. 20, pp. 13906–13914, May 2006, doi: 10.1074/jbc.M600869200.
- [92] J. F. Lu, D. Pokharel, and M. Bebawy, “MRP1 and its role in anticancer drug resistance,” *Drug Metab. Rev.*, vol. 47, no. 4, pp. 406–419, Oct. 2015, doi: 10.3109/03602532.2015.1105253.
- [93] A. J. Gerrard, “Gravity waves and mesospheric clouds in the summer middle atmosphere: A comparison of lidar measurements and ray modeling of gravity waves over Sondrestrom, Greenland,” *J. Geophys. Res.*, vol. 109, no. D10, p. D10103, 2004, doi: 10.1029/2002JD002783.
- [94] É. Bakos and L. Homolya, “Portrait of multifaceted transporter, the multidrug resistance-associated protein 1 (MRP1/ABCC1),” *Pflügers Arch. - Eur. J. Physiol.*, vol. 453, no. 5, pp. 621–641, Jan. 2007, doi: 10.1007/s00424-006-0160-8.
- [95] J. F. Lu, F. Luk, J. Gong, R. Jaiswal, G. E. R. Grau, and M. Bebawy, “Microparticles mediate MRP1 intercellular transfer and the re-templating of intrinsic resistance pathways,” *Pharmacol. Res.*, vol. 76, pp. 77–83, Oct. 2013, doi: 10.1016/j.phrs.2013.07.009.

- [96] M. Munoz, M. Henderson, M. Haber, and M. Norris, "Role of the MRP1/ABCC1 Multidrug Transporter Protein in Cancer," *IUBMB Life*, vol. 59, no. 12, pp. 752–757, 2007, doi: 10.1080/15216540701736285.
- [97] W. Berger *et al.*, "Multidrug resistance markers P-glycoprotein, multidrug resistance protein 1, and lung resistance protein in non-small cell lung cancer: prognostic implications," *J. Cancer Res. Clin. Oncol.*, vol. 131, no. 6, pp. 355–363, Jun. 2005, doi: 10.1007/s00432-004-0653-9.
- [98] I. Sugawara *et al.*, "Preferential expression of the multidrug-resistance-associated protein (MRP) in adenocarcinoma of the lung," *Int. J. Cancer*, vol. 64, no. 5, pp. 322–325, Oct. 1995, doi: 10.1002/ijc.2910640507.
- [99] G. F. Sullivan, P. S. Amenta, J. D. Villanueva, C. J. Alvarez, J. M. Yang, and W. N. Hait, "The expression of drug resistance gene products during the progression of human prostate cancer," *Clin. Cancer Res.*, vol. 4, no. 6, p. 1393–1403, 1998, [Online]. Available: <http://europepmc.org/abstract/MED/9626455>
- [100] T.-H. Kuo, F.-Y. Liu, C.-Y. Chuang, H.-S. Wu, J.-J. Wang, and A. Kao, "To predict response chemotherapy using technetium-99m tetrofosmin chest images in patients with untreated small cell lung cancer and compare with p-glycoprotein, multidrug resistance related protein-1, and lung resistance-related protein expression," *Nucl. Med. Biol.*, vol. 30, no. 6, pp. 627–632, Aug. 2003, doi: 10.1016/S0969-8051(03)00058-1.
- [101] M. Haber *et al.*, "Altered expression of the MYCN oncogene modulates MRP gene expression and response to cytotoxic drugs in neuroblastoma cells," *Oncogene*, vol. 18, no. 17, pp. 2777–2782, Apr. 1999, doi: 10.1038/sj.onc.1202859.
- [102] R. Domingo-Fernandez, K. Watters, O. Piskareva, R. L. Stallings, and I. Bray, "The role of genetic and epigenetic alterations in neuroblastoma disease pathogenesis," *Pediatr. Surg. Int.*, vol. 29, no. 2, pp. 101–119, Feb. 2013, doi: 10.1007/s00383-012-3239-7.
- [103] S. . Cotterill *et al.*, "Clinical prognostic factors in 1277 patients with neuroblastoma: results of The European Neuroblastoma Study Group 'Survey' 1982–1992," *Eur. J. Cancer*, vol. 36, no. 7, pp. 901–908, May 2000, doi: 10.1016/S0959-8049(00)00058-7.
- [104] C. F. Manohar *et al.*, "MYCN-mediated regulation of the MRP1 promoter in human neuroblastoma," *Oncogene*, vol. 23, no. 3, pp. 753–762, Jan. 2004, doi: 10.1038/sj.onc.1207151.
- [105] K. Nooter *et al.*, "The prognostic significance of expression of the multidrug resistance-associated protein (MRP) in primary breast cancer," *Br. J. Cancer*, vol. 76, no. 4, pp. 486–493, Aug. 1997, doi: 10.1038/bjc.1997.414.
- [106] K. Nooter, G. B. de la Riviere, J. Klijn, G. Stoter, and J. Foekens, "Multidrug resistance protein in recurrent breast cancer," *Lancet*, vol. 349, no. 9069, pp. 1885–1886, Jun. 1997, doi: 10.1016/S0140-6736(05)63876-7.
- [107] M. Filipits *et al.*, "92 P - MRP and MDR1 Gene expression in primary breast carcinomas," *Eur. J. Cancer*, vol. 32, p. S19, Jan. 1996, doi: 10.1016/0959-8049(96)84850-7.
- [108] N. Brown, "Chemoinformatics—an introduction for computer scientists," *ACM Comput. Surv.*, vol. 41, no. 2, pp. 1–38, Feb. 2009, doi: 10.1145/1459352.1459353.

- [109] N. A. Woody, "Chemoinformatics: a textbook, Johann Gasteiger and Thomas Engel (eds), Wiley-VCH, Weinheim, 2003, ISBN 3-527-30681-1," *J. Chemom.*, vol. 18, no. 6, pp. 314–315, Jun. 2004, doi: 10.1002/cem.871.
- [110] Y. C. Martin, J. L. Kofron, and L. M. Traphagen, "Do Structurally Similar Molecules Have Similar Biological Activity?," *J. Med. Chem.*, vol. 45, no. 19, pp. 4350–4358, Sep. 2002, doi: 10.1021/jm020155c.
- [111] A. T. Balaban, "Applications of graph theory in chemistry," *J. Chem. Inf. Comput. Sci.*, vol. 25, no. 3, pp. 334–343, Aug. 1985, doi: 10.1021/ci00047a033.
- [112] M. Minoux, "Plus court chemin avec contraintes : Algorithmes et applications," *Ann. Des Télécommunications*, vol. 30, no. 11–12, pp. 383–394, Nov. 1975, doi: 10.1007/BF03004231.
- [113] C. Flament, *Théorie des graphes et structures sociales*. De Gruyter, 1968. doi: 10.1515/9783111670577.
- [114] S. Wang, Y. Guo, Y. Wang, H. Sun, and J. Huang, "SMILES-BERT," in *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, Sep. 2019, pp. 429–436. doi: 10.1145/3307339.3342186.
- [115] D. Weininger, "SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules," *J. Chem. Inf. Model.*, vol. 28, no. 1, pp. 31–36, Feb. 1988, doi: 10.1021/ci00057a005.
- [116] G. Neglur, R. L. Grossman, and B. Liu, "Assigning Unique Keys to Chemical Compounds for Data Integration: Some Interesting Counter Examples," 2005, pp. 145–157. doi: 10.1007/11530084_13.
- [117] D. Weininger, A. Weininger, and J. L. Weininger, "SMILES. 2. Algorithm for generation of unique SMILES notation," *J. Chem. Inf. Comput. Sci.*, vol. 29, no. 2, pp. 97–101, May 1989, doi: 10.1021/ci00062a008.
- [118] D. Adam, "Chemists synthesize a single naming system," *Nature*, vol. 417, no. 6887, pp. 369–369, May 2002, doi: 10.1038/417369a.
- [119] P. K. H. Chow, R. T. H. Ng, and B. E. Ogden, *Using Animal Models in Biomedical Research*. WORLD SCIENTIFIC, 2008. doi: 10.1142/6454.
- [120] F. Clarac, "Les modèles animaux et leur limites," *Hist. la Rech. Contemp.*, no. Tome IV-N°1, pp. 15–22, Jun. 2015, doi: 10.4000/hrc.906.
- [121] Y. Christen, "L'animal est-il une personne?," *Biol. Aujourd'hui.*, vol. 204, no. 3, pp. 235–242, Oct. 2010, doi: 10.1051/jbio/2010012.
- [122] P. Grandjean, "Paracelsus Revisited: The Dose Concept in a Complex World," *Basic Clin. Pharmacol. Toxicol.*, vol. 119, no. 2, pp. 126–132, Aug. 2016, doi: 10.1111/bcpt.12622.
- [123] A. M. Tsatsakis *et al.*, "The dose response principle from philosophy to modern toxicology: The impact of ancient philosophy and medicine in modern toxicology science," *Toxicol. Reports*, vol. 5, pp. 1107–1113, 2018, doi: 10.1016/j.toxrep.2018.10.001.
- [124] W. W. Piegorsch, "Quantification of Toxic Response and the Development of the Median Effective Dose (Ed 50)—a Historical Perspective," *Toxicol. Ind. Health*, vol. 5,

- no. 1, pp. 55–62, Jan. 1989, doi: 10.1177/074823378900500105.
- [125] G. Zbinden, “Scope of Acute Toxicity Testing — Current Methodology,” in *The Contribution of Acute Toxicity Testing to the Evaluation of Pharmaceuticals*, Berlin, Heidelberg: Springer Berlin Heidelberg, 1986, pp. 5–9. doi: 10.1007/978-3-642-70390-4_2.
- [126] J. C. Dearden and M. Hewitt, “Prediction of Human Lethal Doses and Concentrations of MEIC Chemicals from Rodent LD 50 Values: An Attempt to Make Some Reparation,” *Altern. to Lab. Anim.*, vol. 49, no. 1–2, pp. 10–21, Jan. 2021, doi: 10.1177/0261192921994754.
- [127] A. Wilson, D. M. Reif, and B. J. Reich, “Hierarchical dose-response modeling for high-throughput toxicity screening of environmental chemicals,” *Biometrics*, vol. 70, no. 1, pp. 237–246, Mar. 2014, doi: 10.1111/biom.12114.
- [128] P. Brunel, G. Flesch, and G. Meno-Tetang, *Voyage au coeur de la relation dose-réponse du médicament*. EDP Sciences, 2020. doi: 10.1051/978-2-7598-1999-7.
- [129] D. J. Hill, C. S. Hosking, and L. V. Reyes-Benito, “Reducing the need for food allergen challenges in young children: a comparison of in vitro with in vivo tests,” *Clin. Exp. Allergy*, vol. 31, no. 7, pp. 1031–1035, Jul. 2001, doi: 10.1046/j.1365-2222.2001.01113.x.
- [130] A. Guillouzo, “Liver cell models in in vitro toxicology,” *Environ. Health Perspect.*, vol. 106, no. suppl 2, pp. 511–532, Apr. 1998, doi: 10.1289/ehp.98106511.
- [131] H. Sawada, “A Toxicogenomic Approach to Drug-Induced Phospholipidosis: Analysis of Its Induction Mechanism and Establishment of a Novel in Vitro Screening System,” *Toxicol. Sci.*, vol. 83, no. 2, pp. 282–292, Nov. 2004, doi: 10.1093/toxsci/kfh264.
- [132] M. Hayashi, E. Kamata, A. Hirose, M. Takahashi, T. Morita, and M. Ema, “In silico assessment of chemical mutagenesis in comparison with results of Salmonella microsome assay on 909 chemicals,” *Mutat. Res. Toxicol. Environ. Mutagen.*, vol. 588, no. 2, pp. 129–135, Dec. 2005, doi: 10.1016/j.mrgentox.2005.09.009.
- [133] H. Ellinger-Ziegelbauer, H. Gmuender, A. Bandenburg, and H. J. Ahr, “Prediction of a carcinogenic potential of rat hepatocarcinogens using toxicogenomics analysis of short-term in vivo studies,” *Mutat. Res. Mol. Mech. Mutagen.*, vol. 637, no. 1–2, pp. 23–39, Jan. 2008, doi: 10.1016/j.mrfmmm.2007.06.010.
- [134] J. Kirchmair *et al.*, “Fast and Efficient in Silico 3D Screening: Toward Maximum Computational Efficiency of Pharmacophore-Based and Shape-Based Approaches,” *J. Chem. Inf. Model.*, vol. 47, no. 6, pp. 2182–2196, Nov. 2007, doi: 10.1021/ci700024q.
- [135] E. Benfenati and G. Gini, “Computational predictive programs (expert systems) in toxicology,” *Toxicology*, vol. 119, no. 3, pp. 213–225, May 1997, doi: 10.1016/S0300-483X(97)03631-7.
- [136] D. J. Livingstone, “The Characterization of Chemical Structures Using Molecular Properties. A Survey,” *J. Chem. Inf. Comput. Sci.*, vol. 40, no. 2, pp. 195–209, Mar. 2000, doi: 10.1021/ci990162i.
- [137] N. Claude, F. Goldfain-Blanc, and A. Guillouzo, “La place des méthodes in silico, in vitro, in omic dans l’évaluation de la sécurité des médicaments,” *médecine/sciences*, vol. 25, no. 1, pp. 105–110, Jan. 2009, doi: 10.1051/medsci/2009251105.

- [138] J. D. McKinney, "The Practice of Structure Activity Relationships (SAR) in Toxicology," *Toxicol. Sci.*, vol. 56, no. 1, pp. 8–17, Jul. 2000, doi: 10.1093/toxsci/56.1.8.
- [139] N. Nikolova-Jeliazkova and J. Jaworska, "An Approach to Determining Applicability Domains for QSAR Group Contribution Models: An Analysis of SRC KOWWIN," *Altern. to Lab. Anim.*, vol. 33, no. 5, pp. 461–470, Oct. 2005, doi: 10.1177/026119290503300510.
- [140] C. HANSCH, P. P. MALONEY, T. FUJITA, and R. M. MUIR, "Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients," *Nature*, vol. 194, no. 4824, pp. 178–180, Apr. 1962, doi: 10.1038/194178b0.
- [141] A. K. Ghose and G. M. Crippen, "Atomic physicochemical parameters for three-dimensional-structure-directed quantitative structure-activity relationships. 2. Modeling dispersive and hydrophobic interactions," *J. Chem. Inf. Comput. Sci.*, vol. 27, no. 1, pp. 21–35, Feb. 1987, doi: 10.1021/ci00053a005.
- [142] H. Kubinyi, "From Narcosis to Hyperspace: The History of QSAR," *Quant. Struct. Relationships*, vol. 21, no. 4, pp. 348–356, Oct. 2002, doi: 10.1002/1521-3838(200210)21:4<348::AID-QSAR348>3.0.CO;2-D.
- [143] C. Selassie and R. P. Verma, "History of Quantitative Structure–Activity Relationships," in *Burger's Medicinal Chemistry and Drug Discovery*, Wiley, 2010, pp. 1–96. doi: 10.1002/0471266949.bmc001.pub2.
- [144] S. Yousefinejad and B. Hemmateenejad, "Chemometrics tools in QSAR/QSPR studies: A historical perspective," *Chemom. Intell. Lab. Syst.*, vol. 149, pp. 177–204, Dec. 2015, doi: 10.1016/j.chemolab.2015.06.016.
- [145] A. R. Katritzky, U. Maran, V. S. Lobanov, and M. Karelson, "Structurally Diverse Quantitative Structure–Property Relationship Correlations of Technologically Relevant Physical Properties," *J. Chem. Inf. Comput. Sci.*, vol. 40, no. 1, pp. 1–18, Jan. 2000, doi: 10.1021/ci9903206.
- [146] G. Aranoff, "Thought Experiment: Marginal Cost versus John M. Clark's Workable Competition Pricing," *Mod. Econ.*, vol. 11, no. 11, pp. 1742–1751, 2020, doi: 10.4236/me.2020.1111119.
- [147] A. Dudek, T. Arodz, and J. Galvez, "Computational Methods in Developing Quantitative Structure-Activity Relationships (QSAR): A Review," *Comb. Chem. High Throughput Screen.*, vol. 9, no. 3, pp. 213–228, Mar. 2006, doi: 10.2174/138620706776055539.
- [148] C. Nieto-Draghi *et al.*, "A General Guidebook for the Theoretical Prediction of Physicochemical Properties of Chemicals for Regulatory Purposes," *Chem. Rev.*, vol. 115, no. 24, pp. 13093–13164, Dec. 2015, doi: 10.1021/acs.chemrev.5b00215.
- [149] A. Regina Todeschini and S. Hakomori, "Functional role of glycosphingolipids and gangliosides in control of cell adhesion, motility, and growth, through glycosynaptic microdomains," *Biochim. Biophys. Acta - Gen. Subj.*, vol. 1780, no. 3, pp. 421–433, Mar. 2008, doi: 10.1016/j.bbagen.2007.10.008.
- [150] H. P. Schultz, "Topological organic chemistry. 1. Graph theory and topological indices

- of alkanes,” *J. Chem. Inf. Model.*, vol. 29, no. 3, pp. 227–228, Aug. 1989, doi: 10.1021/ci00063a012.
- [151] H. Wiener, “Structural Determination of Paraffin Boiling Points,” *J. Am. Chem. Soc.*, vol. 69, no. 1, pp. 17–20, Jan. 1947, doi: 10.1021/ja01193a005.
- [152] P. Labute, “A widely applicable set of descriptors,” *J. Mol. Graph. Model.*, vol. 18, no. 4–5, pp. 464–477, 2000, doi: 10.1016/S1093-3263(00)00068-1.
- [153] M. Ghamali, S. Chtita, A. Aouidate, A. Ghaleb, M. Bouachrine, and T. Lakhlifi, “Combining DFT and QSAR computation to predict the interaction of flavonoids with the GABA (A) receptor using electronic and topological descriptors,” *J. Taibah Univ. Sci.*, vol. 11, no. 3, pp. 422–433, May 2017, doi: 10.1016/j.jtusci.2016.06.005.
- [154] “QSRR Study of Linear Retention Indices for Volatile Compounds using Statistical Methods,” *Chem. Sci. Trans.*, vol. 7, no. 4, Oct. 2018, doi: 10.7598/cst2018.1501.
- [155] R. Bosque, J. Sales, E. Bosch, M. Rosés, M. C. García-Alvarez-Coque, and J. R. Torres-Lapasió, “A QSPR Study of the p Solute Polarity Parameter to Estimate Retention in HPLC,” *J. Chem. Inf. Comput. Sci.*, vol. 43, no. 4, pp. 1240–1247, Jul. 2003, doi: 10.1021/ci0340660.
- [156] V. N. Viswanadhan, A. K. Ghose, G. R. Revankar, and R. K. Robins, “Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain,” *J. Chem. Inf. Model.*, vol. 29, no. 3, pp. 163–172, Aug. 1989, doi: 10.1021/ci00063a006.
- [157] D. Gadaleta, A. Lombardo, C. Toma, and E. Benfenati, “A new semi-automated workflow for chemical data retrieval and quality checking for modeling applications,” *J. Cheminform.*, vol. 10, no. 1, p. 60, Dec. 2018, doi: 10.1186/s13321-018-0315-6.
- [158] T. Hastie, J. Friedman, and R. Tibshirani, *The Elements of Statistical Learning*. New York, NY: Springer New York, 2001. doi: 10.1007/978-0-387-21606-5.
- [159] M. T. C. Tomasz Puzyn, Jerzy Leszczynski, *Recent Advances in QSAR Studies: Methods and Applications Volume 8 de Challenges and Advances in Computational Chemistry and Physics*, Illustrée.
- [160] G. Khalaf, “Improving the Ordinary Least Squares Estimator by Ridge Regression,” *OALib*, vol. 09, no. 05, pp. 1–8, 2022, doi: 10.4236/oalib.1108738.
- [161] L. Xu and K. Yu, “A Note on Dynamic Roy’s Identity,” *Theor. Econ. Lett.*, vol. 04, no. 07, pp. 513–516, 2014, doi: 10.4236/tel.2014.47064.
- [162] G. C. Gini, “Predictive Toxicology of Chemicals : Experiences and Impact of AI Tools Papers Predictive Toxicology of Chemicals : Experiences and Impact of AI Tools,” no. March 1999, 2019.
- [163] A. Tropsha, P. Gramatica, and V. Gombar, “The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models,” *QSAR Comb. Sci.*, vol. 22, no. 1, pp. 69–77, Apr. 2003, doi: 10.1002/qsar.200390007.
- [164] M. Shen, C. Béguin, A. Golbraikh, J. P. Stables, H. Kohn, and A. Tropsha,

- “Application of Predictive QSAR Models to Database Mining: Identification and Experimental Validation of Novel Anticonvulsant Compounds,” *J. Med. Chem.*, vol. 47, no. 9, pp. 2356–2364, Apr. 2004, doi: 10.1021/jm030584q.
- [165] P. Gramatica, “Principles of QSAR models validation: internal and external,” *QSAR Comb. Sci.*, vol. 26, no. 5, pp. 694–701, May 2007, doi: 10.1002/qsar.200610151.
- [166] A. Tropsha and A. Golbraikh, “Predictive QSAR Modeling Workflow, Model Applicability Domains, and Virtual Screening,” *Curr. Pharm. Des.*, vol. 13, no. 34, pp. 3494–3504, Dec. 2007, doi: 10.2174/138161207782794257.
- [167] A. Golbraikh and A. Tropsha, “Beware of q^2 ,” *J. Mol. Graph. Model.*, vol. 20, no. 4, pp. 269–276, Jan. 2002, doi: 10.1016/S1093-3263(01)00123-1.
- [168] A. Tropsha, “Best Practices for QSAR Model Development, Validation, and Exploitation,” *Mol. Inform.*, vol. 29, no. 6–7, pp. 476–488, Jul. 2010, doi: 10.1002/minf.201000061.
- [169] G. Schüürmann, R.-U. Ebert, J. Chen, B. Wang, and R. Kühne, “External Validation and Prediction Employing the Predictive Squared Correlation Coefficient — Test Set Activity Mean vs Training Set Activity Mean,” *J. Chem. Inf. Model.*, vol. 48, no. 11, pp. 2140–2145, Nov. 2008, doi: 10.1021/ci800253u.
- [170] K. Roy, S. Kar, and R. N. Das, *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*. Elsevier Science, 2015.
- [171] D. T. Stanton, “On the Physical Interpretation of QSAR Models,” *J. Chem. Inf. Comput. Sci.*, vol. 43, no. 5, pp. 1423–1433, Sep. 2003, doi: 10.1021/ci0340658.
- [172] T. I. Netzeva *et al.*, “Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure–Activity Relationships,” *Altern. to Lab. Anim.*, vol. 33, no. 2, pp. 155–173, Apr. 2005, doi: 10.1177/026119290503300209.
- [173] N. S. Zefirov and V. A. Palyulin, “QSAR for Boiling Points of ‘Small’ Sulfides. Are the ‘High-Quality Structure–Property–Activity Regressions’ the Real High Quality QSAR Models?,” *J. Chem. Inf. Comput. Sci.*, vol. 41, no. 4, pp. 1022–1027, Jul. 2001, doi: 10.1021/ci0001637.
- [174] P. Polishchuk, “Interpretation of Quantitative Structure–Activity Relationship Models: Past, Present, and Future,” *J. Chem. Inf. Model.*, vol. 57, no. 11, pp. 2618–2639, Nov. 2017, doi: 10.1021/acs.jcim.7b00274.
- [175] D. H. Wolpert and W. G. Macready, “No free lunch theorems for optimization,” *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 67–82, Apr. 1997, doi: 10.1109/4235.585893.
- [176] J. G. Topliss and R. P. Edwards, “Chance factors in studies of quantitative structure–activity relationships,” *J. Med. Chem.*, vol. 22, no. 10, pp. 1238–1244, Oct. 1979, doi: 10.1021/jm00196a017.
- [177] X. Cao *et al.*, “Synthesis and Biological Evaluation of Fused Tricyclic Heterocycle Piperazine (Piperidine) Derivatives As Potential Multireceptor Atypical Antipsychotics,” *J. Med. Chem.*, vol. 61, no. 22, pp. 10017–10039, Nov. 2018, doi: 10.1021/acs.jmedchem.8b01096.
- [178] O. M. Khalil, A. M. Kamal, S. Bua, H. El Sayed Teba, Y. M. Nissan, and C. T. Supuran, “Pyrrolo and pyrrolopyrimidine sulfonamides act as cytotoxic agents in

- hypoxia via inhibition of transmembrane carbonic anhydrases,” *Eur. J. Med. Chem.*, vol. 188, p. 112021, Feb. 2020, doi: 10.1016/j.ejmech.2019.112021.
- [179] V. AMARNATH and R. MADHAV, “A Survey of Methods for the Preparation of Pyrrolopyrimidines,” *Synthesis (Stuttg.)*, vol. 1974, no. 12, pp. 837–859, 1974, doi: 10.1055/s-1974-23448.
- [180] A. V. Kadushkin, A. S. Sokolova, N. P. Solov’eva, and V. G. Granik, “Lactam acetals and acid amides. 74. Studies of the synthesis and antitumor activity of 5,6-polymethylenepyrrolo [3,2-d]pyrimidines,” *Pharm. Chem. J.*, vol. 28, no. 11, pp. 792–798, Nov. 1994, doi: 10.1007/BF02218707.
- [181] N. Azizi, A. Khajeh-Amiri, H. Ghafari, M. Bolourtchian, and M. Saidi, “Iron-Catalyzed Inexpensive and Practical Synthesis of N-Substituted Pyrroles in Water,” *Synlett*, vol. 2009, no. 14, pp. 2245–2248, Sep. 2009, doi: 10.1055/s-0029-1217799.
- [182] H. V. Chavan *et al.*, “Design, synthesis, characterization and anti-inflammatory evaluation of novel pyrazole amalgamated flavones,” *Bioorg. Med. Chem. Lett.*, vol. 23, no. 5, pp. 1315–1321, Mar. 2013, doi: 10.1016/j.bmcl.2012.12.094.
- [183] A. Gangjee, S. Kurup, M. A. Inhat, J. E. Thorpe, and S. S. Shenoy, “Synthesis and biological activity of N4-phenylsubstituted-6-(2,4-dichloro phenylmethyl)-7H-pyrrolo[2,3-d]pyrimidine-2,4-diamines as vascular endothelial growth factor receptor-2 inhibitors and antiangiogenic and antitumor agents,” *Bioorg. Med. Chem.*, vol. 18, no. 10, pp. 3575–3587, May 2010, doi: 10.1016/j.bmc.2010.03.052.
- [184] M. De Rosa and D. Arnold, “Electronic and Steric Effects on the Mechanism of the Inverse Electron Demand Diels–Alder Reaction of 2-Aminopyrroles with 1,3,5-Triazines: Identification of Five Intermediates by 1 H, 13 C, 15 N, and 19 F NMR Spectroscopy,” *J. Org. Chem.*, vol. 74, no. 1, pp. 319–328, Jan. 2009, doi: 10.1021/jo8019626.
- [185] P. Mizar and B. Myrboh, “Three-component synthesis of 5:6 and 6:6 fused pyrimidines using KF–alumina as a catalyst,” *Tetrahedron Lett.*, vol. 49, no. 36, pp. 5283–5285, Sep. 2008, doi: 10.1016/j.tetlet.2008.06.087.
- [186] L. El Kaim, L. Grimaud, and S. Wagschal, “ChemInform Abstract: Toward Pyrrolo[2,3-d]pyrimidine Scaffolds,” *ChemInform*, vol. 41, no. 50, p. no-no, Dec. 2010, doi: 10.1002/chin.201050163.
- [187] J. Shi *et al.*, “EC144 Is a Potent Inhibitor of the Heat Shock Protein 90,” *J. Med. Chem.*, vol. 55, no. 17, pp. 7786–7795, Sep. 2012, doi: 10.1021/jm300810x.
- [188] V. Prieur, N. Heindler, J. Rubio-Martínez, G. Guillaumet, and M. D. Pujol, “One-pot synthesis of 4-aminated pyrrolo[2,3-d]pyrimidines from alkynylpyrimidines under metal-catalyst-free conditions,” *Tetrahedron*, vol. 71, no. 8, pp. 1207–1214, Feb. 2015, doi: 10.1016/j.tet.2015.01.012.
- [189] D. E. Ehmann and S. D. Lahiri, “Novel compounds targeting bacterial DNA topoisomerase/DNA gyrase,” *Curr. Opin. Pharmacol.*, vol. 18, pp. 76–83, Oct. 2014, doi: 10.1016/j.coph.2014.09.007.
- [190] S. K. Desmoulin *et al.*, “Functional Loss of the Reduced Folate Carrier Enhances the Antitumor Activities of Novel Antifolates with Selective Uptake by the Proton-Coupled Folate Transporter,” *Mol. Pharmacol.*, vol. 82, no. 4, pp. 591–600, Oct. 2012,

- doi: 10.1124/mol.112.079004.
- [191] S. Pathania and R. K. Rawal, "Pyrrolopyrimidines: An update on recent advancements in their medicinal attributes," *Eur. J. Med. Chem.*, vol. 157, pp. 503–526, Sep. 2018, doi: 10.1016/j.ejmech.2018.08.023.
- [192] S. M. Schmitt, K. Stefan, and M. Wiese, "Pyrrolopyrimidine Derivatives as Novel Inhibitors of Multidrug Resistance-Associated Protein 1 (MRP1, ABCC1)," *J. Med. Chem.*, vol. 59, no. 7, pp. 3018–3033, Apr. 2016, doi: 10.1021/acs.jmedchem.5b01644.
- [193] V. Lather and A. K. Madan, "Topological model for the prediction of MRP1 inhibitory activity of pyrrolopyrimidines and templates derived from pyrrolopyrimidine," *Bioorg. Med. Chem. Lett.*, vol. 15, no. 22, pp. 4967–4972, Nov. 2005, doi: 10.1016/j.bmcl.2005.08.011.
- [194] S. Wang *et al.*, "Design and Synthesis of New Templates Derived from Pyrrolopyrimidine as Selective Multidrug-Resistance-Associated Protein Inhibitors in Multidrug Resistance," *J. Med. Chem.*, vol. 47, no. 6, pp. 1339–1350, Mar. 2004, doi: 10.1021/jm0310129.
- [195] AMANDA YARNELL, "Happy Birthday, ChemDraw," *Chem. Eng. News Arch.*, vol. 93, no. 27, p. 3, Jul. 2015, doi: 10.1021/cen-09327-editorial.
- [196] "Chemdraw Ultra," « *Molecular modelling, structure drawing, Semi-empirical calculations, structure display, MOPAC, solvation energy, MM2* ». Copyright Cambridge Soft Coperation, 2002.
- [197] A. Hinchliffe, "HyperChem Release 4.5 for Windows," *Electron. J. Theor. Chem.*, vol. 1, no. 1, pp. 62–64, Mar. 2001, doi: 10.1002/ejtc.9.
- [198] U. Hypercube, Inc., Gainesville, Florida, "HyperChem." 1999. [Online]. Available: <http://www.hyper.com>
- [199] T. Srl, "Dragon for windows," (*Software for Molecular Descriptor Calculation*). Milano, Italy, 2007.
- [200] P. Gramatica, N. Chirico, E. Papa, S. Cassani, and S. Kovarich, "QSARINS: A new software for the development, analysis, and validation of QSAR MLR models," *J. Comput. Chem.*, vol. 34, no. 24, pp. 2121–2132, Sep. 2013, doi: 10.1002/jcc.23361.
- [201] P. Gramatica, S. Cassani, and N. Chirico, "QSARINS-chem: Insubria datasets and new QSAR/QSPR models for environmental pollutants in QSARINS," *J. Comput. Chem.*, vol. 35, no. 13, pp. 1036–1044, May 2014, doi: 10.1002/jcc.23576.
- [202] D. M. Ceccarelli, "The Coral Sea," in *World Seas: an Environmental Evaluation*, Elsevier, 2019, pp. 679–698. doi: 10.1016/B978-0-08-100853-9.00040-3.
- [203] W. Wu and D. L. Massart, "Artificial neural networks in classification of NIR spectral data: Selection of the input," *Chemom. Intell. Lab. Syst.*, vol. 35, no. 1, pp. 127–135, Nov. 1996, doi: 10.1016/S0169-7439(96)00034-2.
- [204] P. Pratim Roy, S. Paul, I. Mitra, and K. Roy, "On Two Novel Parameters for Validation of Predictive QSAR Models," *Molecules*, vol. 14, no. 5, pp. 1660–1701, Apr. 2009, doi: 10.3390/molecules14051660.
- [205] A. P. Toropova, A. A. Toropov, E. Benfenati, R. Rallo, D. Leszczynska, and J. Leszczynski, "Development of Monte Carlo Approaches in Support of Environmental

- Research,” 2017, pp. 453–469. doi: 10.1007/978-3-319-56850-8_12.
- [206] A. P. Toropova and A. A. Toropov, “The index of ideality of correlation: improvement of models for toxicity to algae,” *Nat. Prod. Res.*, vol. 33, no. 15, pp. 2200–2207, Aug. 2019, doi: 10.1080/14786419.2018.1493591.
- [207] A. P. Toropova and A. A. Toropov, “The index of ideality of correlation: A criterion of predictability of QSAR models for skin permeability?,” *Sci. Total Environ.*, vol. 586, pp. 466–472, May 2017, doi: 10.1016/j.scitotenv.2017.01.198.
- [208] A. A. Toropov, A. P. Toropova, M. Marzo, J. Lou Dorne, N. Georgiadis, and E. Benfenati, “QSAR models for predicting acute toxicity of pesticides in rainbow trout using the CORAL software and EFSA’s OpenFoodTox database,” *Environ. Toxicol. Pharmacol.*, vol. 53, pp. 158–163, Jul. 2017, doi: 10.1016/j.etap.2017.05.011.
- [209] A. A. Toropov and A. P. Toropova, “Use of the index of ideality of correlation to improve predictive potential for biochemical endpoints,” *Toxicol. Mech. Methods*, vol. 29, no. 1, pp. 43–52, Jan. 2019, doi: 10.1080/15376516.2018.1506851.
- [210] A. A. Toropov, R. Carbó-Dorca, and A. P. Toropova, “Index of Ideality of Correlation: new possibilities to validate QSAR: a case study,” *Struct. Chem.*, vol. 29, no. 1, pp. 33–38, Feb. 2018, doi: 10.1007/s11224-017-0997-9.
- [211] P. Kumar, A. Kumar, and J. Sindhu, “Design and development of novel focal adhesion kinase (FAK) inhibitors using Monte Carlo method with index of ideality of correlation to validate QSAR,” *SAR QSAR Environ. Res.*, vol. 30, no. 2, pp. 63–80, Feb. 2019, doi: 10.1080/1062936X.2018.1564067.
- [212] A. P. Toropova, A. A. Toropov, and E. Benfenati, “CORAL: Prediction of binding affinity and efficacy of thyroid hormone receptor ligands,” *Eur. J. Med. Chem.*, vol. 101, pp. 452–461, Aug. 2015, doi: 10.1016/j.ejmech.2015.07.012.
- [213] K. Roy and S. Kar, Eds., “How to Judge Predictive Quality of Classification and Regression Based QSAR Models?,” in *Frontiers in Computational Chemistry*, BENTHAM SCIENCE PUBLISHERS, 2015, pp. 71–120. doi: 10.2174/9781608059782115020005.
- [214] P. G. R. Achary, “Simplified molecular input line entry system-based optimal descriptors: QSAR modelling for voltage-gated potassium channel subunit Kv7.2,” *SAR QSAR Environ. Res.*, vol. 25, no. 1, pp. 73–90, Jan. 2014, doi: 10.1080/1062936X.2013.842930.
- [215] M. A. Islam and T. S. Pillay, “Simplified molecular input line entry system-based descriptors in QSAR modeling for HIV-protease inhibitors,” *Chemom. Intell. Lab. Syst.*, vol. 153, pp. 67–74, Apr. 2016, doi: 10.1016/j.chemolab.2016.02.008.
- [216] A. Rescifina *et al.*, “Development of a Sigma-2 Receptor affinity filter through a Monte Carlo based QSAR analysis,” *Eur. J. Pharm. Sci.*, vol. 106, pp. 94–101, Aug. 2017, doi: 10.1016/j.ejps.2017.05.061.
- [217] J. S. Jaworska, M. Comber, C. Auer, and C. J. Van Leeuwen, “Summary of a workshop on regulatory acceptance of (Q)SARs for human health and environmental endpoints.,” *Environ. Health Perspect.*, vol. 111, no. 10, pp. 1358–1360, Aug. 2003, doi: 10.1289/ehp.5757.
- [218] A. Linusson, M. Elofsson, I. E. Andersson, and M. K. Dahlgren, “Statistical Molecular

- Design of Balanced Compound Libraries for QSAR Modeling,” *Curr. Med. Chem.*, vol. 17, no. 19, pp. 2001–2016, Jul. 2010, doi: 10.2174/092986710791233661.
- [219] K. Roy, I. Mitra, S. Kar, P. K. Ojha, R. N. Das, and H. Kabir, “Comparative Studies on Some Metrics for External Validation of QSPR Models,” *J. Chem. Inf. Model.*, vol. 52, no. 2, pp. 396–408, Feb. 2012, doi: 10.1021/ci200520g.
- [220] R. ZAKARIA and S. M. AMINE, “Analyse de la relation quantitative Structure / Activité des inhibiteurs de la tyrosine kinase,” Université Frères Mentouri Constantine 1 Université des Frères Mentouri Constant, 2021.
- [221] P. Kumar and A. Kumar, “In silico enhancement of azo dye adsorption affinity for cellulose fibre through mechanistic interpretation under guidance of QSPR models using Monte Carlo method with index of ideality correlation,” *SAR QSAR Environ. Res.*, vol. 31, no. 9, pp. 697–715, Sep. 2020, doi: 10.1080/1062936X.2020.1806105.
- [222] A. A. Toropov, A. P. Toropova, G. Selvestrel, D. Baderna, and E. Benfenati, “Prediction of No Observed Adverse Effect Concentration for inhalation toxicity using Monte Carlo approach,” *SAR QSAR Environ. Res.*, vol. 31, no. 12, pp. 1–12, Dec. 2020, doi: 10.1080/1062936X.2020.1841827.
- [223] A. A. Toropov, A. P. Toropova, A. Lombardo, A. Roncaglioni, G. J. Lavado, and E. Benfenati, “The Monte Carlo method to build up models of the hydrolysis half-lives of organic compounds,” *SAR QSAR Environ. Res.*, vol. 32, no. 6, pp. 463–471, Jun. 2021, doi: 10.1080/1062936X.2021.1914156.

Année universitaire :

2021-2022

Présenté par : MEZAACHE RAYENE

KHENIFAR SARA

Modélisation de l'activité inhibitrice MRP1 par l'approche QSAR

Mémoire pour l'obtention du diplôme de Master en Biochimie Appliquée

Résumé :

Ce mémoire a pour objectifs d'établir des modèles QSAR fiables, stables et prédictifs. La méthodologie QSAR dont deux approches (DRAGON-QSARINS et CORAL) sont utilisées pour modéliser l'activité inhibitrice MRP1 représenté par (pIC_{50}) de 105 dérivés du pyrrolopyrimidine, en comparant leurs résultats, ainsi choisir le model le plus performant.

Les modèles établis de ces deux approches ont prouvé leur validation interne et externe confirmant leurs stabilités, leurs robustesses et leurs capacités prédictives.

Pour le modèle DRAGON-QSARINS ($R^2=83.02\%$, $Q^2= 81.5\%$) dont l'activité inhibitrice a été corrélés avec 3 descripteurs (T(N..F), GATS2v, VEA2). Les résultats sont conformes aux principales recommandations du Golbraikh et Tropsha.

Les résultats obtenus par CORAL ($R^2 = 87\%$, $Q^2= 85\%$) nous ont permis de conclure que ce dernier possède d'excellents paramètres prédictifs internes et externes, et un caractère très robuste et stable pouvant prédire l'activité pour de nouveaux composés.

Enfin, le model CORAL peut être utilisé efficacement pour estimer des molécules qui n'ont pas été testées expérimentalement.

Mots clés : Pyrrolopyrimidine, MRP1, QSAR, méthode Monte Carlo, Chimio-métrie.

Laboratoire de recherche : laboratoire de Bio-informatique centre de recherche de la biotechnologie (CRBt), Constantine.

Jury devaluation:

- **Encadrant:** Mr.BOUHEDJAR KHALID. MRA CRBt Constantine.
- **Co-encadrant:** Mr.DEMS MOHAMED ABDESSELEM.MRA CRBt Constantine.
- **Examineur 1:** Mr.BENSEGUENI ABDERRAHMANE. Pr .UFM Constantine-1.
- **Examineur 2:** Mr.MOKRANI ELHASSEN MCB. UFM Constantine-1.

Date de dépôt : 22-06-2022