



الجمهورية الجزائرية الديمقراطية الشعبية

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE

وزارة التعليم العالي والبحث العلمي

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE  
SCIENTIFIQUE

Université des Frères Mentouri Constantine 1

جامعة الاخوة منتوري قسنطينة 1

Faculté des Sciences de la Nature et de la Vie

كلية علوم الطبيعة و الحياة

قسم الكيمياء الحيوية والبيولوجيا الخلوية والجزيئية

**Département de Biochimie et Biologie Cellulaire et Moléculaire**

**Mémoire présenté en vue de l'obtention du Diplôme de Master**

**Filière : Sciences Biologiques**

**Spécialité : Biochimie appliquée**

Intitulé

**Etude chimio métrique des dérivés de quinazoline**

**Pour le traitement de cancer du sein**

Présenté et soutenu par :

Le : 30/09/2020

**BOUFENNARA FERIAL**

**BRIK MERIEM**

Devant le jury :

**Président du jury : Mr. BENSEGUENI ABDERRAHMANE Pr. UFM, Constantine 1**

**Encadreur: Mr. DEMS MOHAMED ABDESSELEM. MRA CRBt Constantine**

**Examineur : Mr. MOKRANI ELHASSEN MAA UFM, Constantine 1**

Année universitaire

2019- 2020



الجمهورية الجزائرية الديمقراطية الشعبية

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE

وزارة التعليم العالي والبحث العلمي

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE  
SCIENTIFIQUE

Université des Frères Mentouri Constantine 1

جامعة الاخوة منتوري قسنطينة 1

Faculté des Sciences de la Nature et de la Vie

كلية علوم الطبيعة و الحياة

قسم الكيمياء الحيوية والبيولوجيا الخلوية والجزيئية

**Département de Biochimie et Biologie Cellulaire et Moléculaire**

**Mémoire présenté en vue de l'obtention du Diplôme de Master**

**Filière : Sciences Biologiques**

**Spécialité : Biochimie appliquée**

Intitulé

**Etude chimio métrique des dérivés de quinazoline**

**Pour le traitement de cancer du sein**

Présenté et soutenu par :

Le : 30/09/2020

**BOUFENNARA FERIAL**

**BRIK MERIEM**

Devant le jury :

**Président du jury : Mr. BENSEGUENI ABDERRAHMANE Pr. UFM, Constantine 1**

**Encadreur: Mr. DEMS MOHAMED ABDESSELEM. MRA CRBt Constantine**

**Examineur : Mr. MOKRANI ELHASSEN MAA UFM, Constantine 1**

Année universitaire

2019- 2020

# Remerciements

Tout d'abord, nous tenons à remercier Dieu, de nous avoir donné la santé, la volonté et la patience pour mener à terminer notre formation et pouvoir réaliser ce travail.

## **A notre encadreur :**

***Monsieur MOHAMED ABDESSELEM DEMS maitre de recherche classe A au centre de recherche en biotechnologie (CRBt).***

Nous vous remercions d'avoir été à l'initiative de ce travail. Vous nous avez guidés dans l'élaboration de ce mémoire par votre grande disponibilité, vos conseils et votre précieuse aide. Merci de nous avoir fait partager vos connaissances, nous avons été heureux de travailler avec vous et nous tenons à vous en remercier sincèrement.

Nous tenons à remercier ***Monsieur KHALID BOUHEDJAR*** pour sa disponibilité et nous avoir fait partager ces connaissances.

## **A notre Président du jury :**

***Pr. BENSEGUENI A. Professeur à l'université des frères Mentouri- Constantine 1.***

Nous vous remercions de nous faire l'honneur de présider le jury de ce travail, Soyez assuré de notre profonde reconnaissance.

## **A notre examinateur :**

***Mr. MOKRANI E.H. Maitre-assistant A à l'université des frères Mentouri- Constantine 1.***

Nous vous remercions d'avoir accepté d'examiner ce travail et pour l'honneur que vous nous avez fait en participant au jury.

Nous remercions tous les enseignants de notre spécialité qui ont contribué à l'acquisition de certaines de nos connaissances. Un grand merci également à toute l'équipe du centre de recherche en biotechnologie(C,R,Bt). Enfin, nous remercions tous ceux et celles qui ont contribué de près ou de loin à la réalisation de ce travail.

# Dédicace

Je dédie ce modeste travail :

♥ **A ma très chère Maman** *Nora*

A la lumière de mes jours, la source de mes efforts, la flamme de mon cœur, de ma vie et de mon bonheur MAMAN. Votre amour et votre patience m'ont accompagné chacun de mes pas. J'ai conscience de tous les sacrifices que vous avez dû faire pour me permettre de mener mes études dans les meilleures conditions possibles.

♥ **A mon cher Papa** *Hichem*

Mon homme de vie, mon éternel exemple, mon appui moral et source de joie et de bonheur, celui qui s'est toujours battu pour me voir réussir

♥ **A mon cher frère** *Mohamed Cherif*

♥ **A ma petite sœur** *Rania*

♥ **A mon grand-père** *Mohamed*

♥ **A la mémoire de ma grand-mère** *Mimiya*

♥ **A mes amis** *Somia, Mahmoud, Rayane et Amani*

A tous ceux qui me sont chers.

**BOUFENNARA FERAL**

# Dédicace



*Pour l'âme de mon très cher père que la mort ne lui a laissé le temps pour lire le rapport de sa fille*

*Pour la fleur de ma vie ma mère dont la patience de son cœur manque à plusieurs hommes*

*À vous mes frères zineeddine, nadjib, mes belles sœurs, Asma et sana, qui m'avez toujours soutenu et encouragé durant ces années d'études*

*À toutes les personnes de ma grande famille, à ma cousine Besma pour l'encouragement et l'aide*

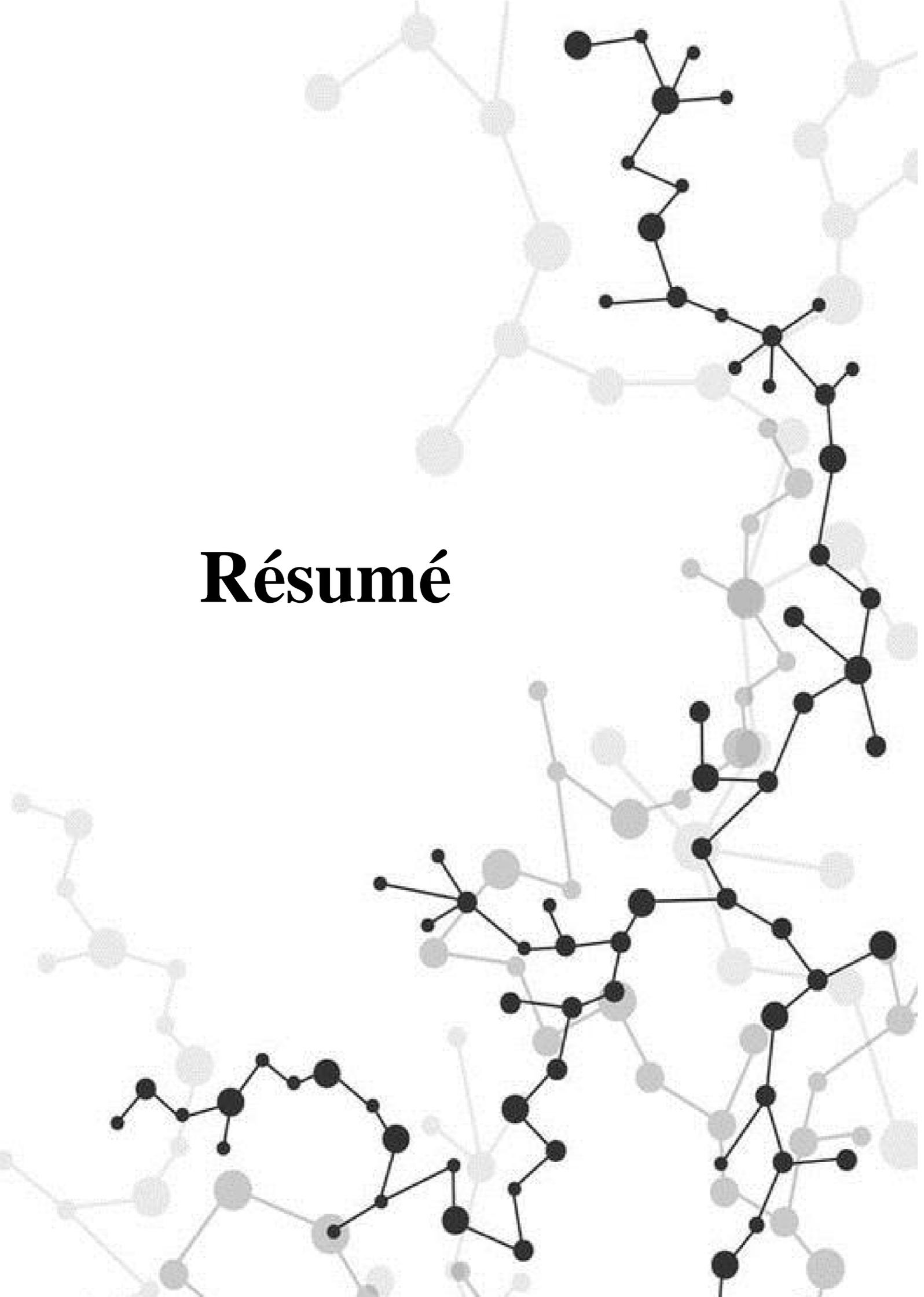
*À mes meilleures amies ferial, rayan, amani, hadjer, à tous les amis de la promotion*

*Et à tous ceux qui ont contribué de près ou de loin pour que ce projet soit possible*

*Je vous dis*

*MERCI*

**Brik Meriem**

The background features a complex network of interconnected nodes and lines, resembling a molecular or network structure. The nodes are represented by circles of varying sizes and shades of gray, while the lines are thin and black. The overall appearance is that of a dense, interconnected web of points and connections, typical of a graph or a molecular model.

# Résumé

في عملنا، تم استخدام أداة القياس الكيميائي QSAR لنموذجة النشاط المثبط لـ ABCG2 لـ 87 مشتقاً من الكينازولين الممثل باللوغاريثم العشري العكسي لـ  $IC_{50}$ ،  $pIC_{50}$ . تم بناء نموذج QSAR يتكون من 5 واصفات MATS7v، MATS2m، MATS5p، RDF010m، EEig12xc مرتبطة بـ  $pIC_{50}$ .

أعطى نموذج QSAR ناتج ذات خصائص إحصائية جيدة ( $R^2 = 75.89$ ،  $Q^2_{LOO} = 70.99$ ،  $Q^2_{ext} = 74.37$ ،  $40.2982$ ) للنبؤ بالنشاط المثبط لـ ABCG2. لقد أثبت النموذج المقترح متانته وقدرته التنبؤية الجيدة (داخلياً وخارجياً) بالإضافة إلى ثباته الجيد. يمكن استخدامه بشكل فعال لتقدير الجزيئات التي لم يتم اختبارها تجريبياً.

#### الكلمات الدالة:

الكينازولين، QSAR، ABCG2،  $pIC_{50}$ ، القياسات الكيميائية

## Summary

In our work the QSAR chemometrics tool was used to model the inhibitory activity of ABCG2 represented by negative decimal logarithm of the  $IC_{50}$  ( $pIC_{50}$ ) of 87 quinazoline derivatives. A QSAR model was built from 5 descriptors MATS2m, MATS7v, MATS5p, RDF010m, EEig12xc correlated with  $pIC_{50}$ .

The resulting QSAR model gave better statistical parameters ( $R^2=75.89$ ,  $Q^2_{LOO} =70.99$ ,  $Q^2_{ext}=74.37$ ,  $F=40.2982$ ) to predict inhibitory activity of ABCG2. The proposed model has proven its robustness, its good predictive capacity (internal and external) as well as good stability. It could be used effectively to estimate molecules not tested experimentally.

## Keywords:

Quinazolines, QSAR, ABCG2,  $pIC_{50}$ , Chemometrics

## Résumé

Dans notre travail l'outil QSAR de la chimio-métrie a été utilisé pour modéliser l'activité inhibitrice d'ABCG2 représenté par le cologarithme décimal d'IC<sub>50</sub> (pIC<sub>50</sub>) de 87 dérivés de quinazoline. Un model QSAR a été construit à partir de 5 descripteurs MATS2m, MATS7v, MATS5p, RDF010m, EEig12xc corrélés avec pIC<sub>50</sub>.

Le modèle QSAR obtenu a donné des meilleurs paramètres statistiques ( $R^2=75.89$ ,  $Q^2_{LOO}=70.99$ ,  $Q^2_{ext}=74.37$ ,  $F=40.2982$ ) pour prédire l'activité inhibitrice. Le modèle proposé a prouvé sa robustesse, sa bonne capacité prédictive (Interne et externe) ainsi qu'une bonne stabilité. Il pourrait être utilisé efficacement pour estimer les molécules non testées expérimentalement

## Mots clés :

Quinazoline, QSAR, ABCG2, pIC<sub>50</sub>, Chimio-métrie

### Liste des tableaux

Tableaux	Titres	Pages
<b>Tableau 1</b>	L'activité inhibitrice envers ABCG2 de 87 dérivés de quinazoline.	23
<b>Tableau 2</b>	Matrice de corrélation entre $pIC_{50}$ et les descripteurs impliqués dans le modèle QSAR (valeurs de R).	29
<b>Tableau 3</b>	Statistiques du modèle sélectionné.	30
<b>Tableau 4</b>	Les valeurs expérimentales, prédit et calculé pour $pIC_{50}$ pour les 87 dérivés de quinazoline.	31
<b>Tableau 5</b>	Signification des descripteurs moléculaires.	Annexe
<b>Tableau 6</b>	Les valeurs de l' $IC_{50}$ et les cinq descripteurs.	Annexe

## Liste des figures

<b>Figures</b>	<b>Titres</b>	<b>Pages</b>
<b>Figure 1</b>	Topologie des transporteurs ABC.	4
<b>Figure 2</b>	Représentation schématique du transfert des substrats à travers la membrane par les TMD.	5
<b>Figure 3</b>	Modèle de transport des drogues "ATP switch model".	6
<b>Figure 4</b>	Les transporteurs ABC et leur implication dans la résistance aux médicaments.	8
<b>Figure 5</b>	Vue schématique de l'expression d'ABCG2 dans l'organisme.	9
<b>Figure 6</b>	Schéma d'ABCG2 et liste (non exhaustive) de ses substrats.	11
<b>Figure 7</b>	Structure de Gefitinib.	12
<b>Figure 8</b>	Structure d'Erlotinib.	13
<b>Figure 9</b>	Structure de Lapatinib.	13
<b>Figure 10</b>	Processus de la construction d'un modèle QSAR.	15
<b>Figure 11</b>	principe de la validation croisée pour 5 plis.	20
<b>Figure 12</b>	Structure de dérivés de quinazoline.	23
<b>Figure 13</b>	La droite de régression des valeurs expérimentales et prédits de la notation $pIC_{50}$ pour le modèle de QSAR.	31
<b>Figure 14</b>	Test de randomisation associé au modèle QSAR 1000 itérations.	37
<b>Figure 15</b>	Diagramme de williams des résidus standardisés.	38

## Symboles et abréviation

### **A**

**ABCG<sub>2</sub>** : ATP-Binding Cassette famille 'G'

**ABC** : ATP-Binding Cassette

**AG** : algorithmes génétiques

**ATP** : adénosine triphosphate

### **B**

**BCRP**: Breast Cancer Resistance Protein

### **C**

**CIRC** : Centre international de recherche sur le cancer

### **E**

**EGFR** : récepteur du facteur de croissance épidermique

**EQMP** : erreur quadratique moyenne prédite

**EQMC** : erreur quadratique moyenne calculé

### **F**

**F**: test de Fisher

### **H**

**HUGO** : Human Genome Organisation

**HER2** : facteur de croissance épidermique humain 2

### **I**

**IC 50** : La concentration inhibitrice 50 (CI 50, ou IC 50 en anglais)

### **K**

**K** : indice de corrélation multivariée

### **M**

**MDR**: Multi Drug Résistance

**MXR**: MitoXantrone Resistance-associated gene

**MLR** : Régression linéaire multiple

### **N**

**NBD** : Nucleotide-Binding Domain

### **O**

**OMS** : Organisation Mondiale de la Santé

### **P**

**PRESS:** Somme des carrés des erreurs de prédiction.

**pIC<sub>50</sub>:**  $\log 1/(\text{IC}_{50})$ .

## Q

**QSAR:** Quantitative Structure Activity Relationships (Relations Quantitatives Structure Activité).

**Q<sup>2</sup>:** coefficient de validation

**Q<sup>2</sup><sub>LoO</sub>:** Coefficient de validation (leave-one-out)

**Q<sup>2</sup><sub>Boot</sub>:** Coefficient de prédiction par la technique du bootstrap.

**Q<sup>2</sup><sub>ext</sub> :** Coefficient de la validation externe

## R

**R<sup>2</sup> :** Coefficient de détermination.

**RMSE:** Root mean square error (l'écart type)

**R<sup>2</sup><sub>adj</sub> :** Coefficient de détermination ajusté.

## S

**SDEC :** standard de calcul

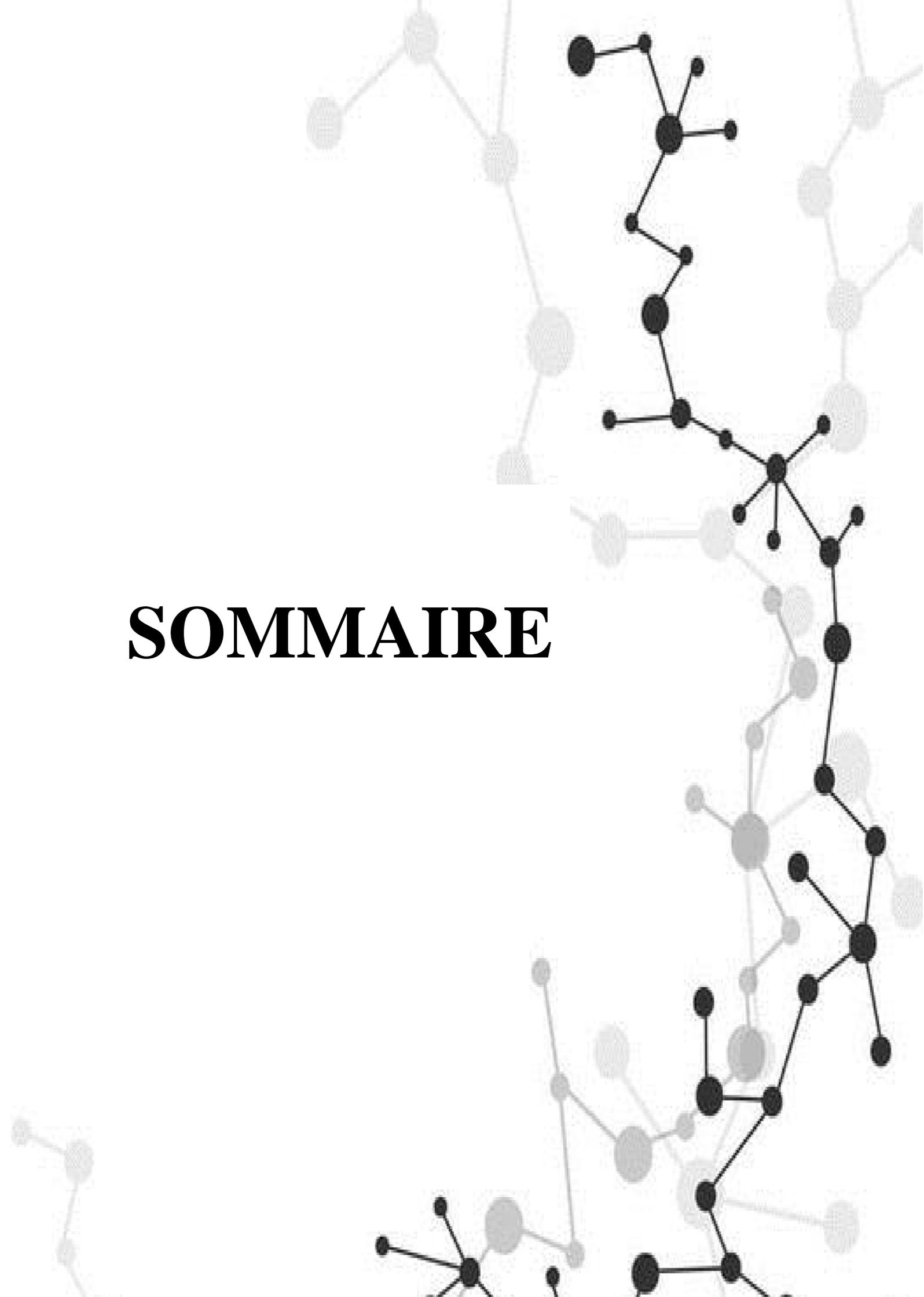
**SDEP :** standard de prediction

**SDEP<sub>ext</sub> :** standard de prediction externe

**SE:** Erreur standard

## T

**TMD:** Trans-Membrane Domain

The background of the page is filled with abstract, interconnected molecular or network-like structures. These structures are composed of small circular nodes connected by thin lines. Some nodes are solid black, while others are semi-transparent grey. The overall effect is a complex, organic pattern that resembles a molecular model or a data network, set against a plain white background.

# **SOMMAIRE**

## Sommaire

Remerciements.....	I
Dédicace.....	II
Résumé .....	IV
Liste des tableaux .....	VIII
Liste des figures .....	13
Liste des abréviations.....	X
Introduction générale.....	1

### Partie1 : synthèse bibliographique

#### Chapitre1 : cancer du sein et les transporteurs ABCG

1. introduction : .....	3
1.1. Statistiques du cancer du sein : .....	3
2. Les transporteurs des médicaments ABC : .....	3
2.1. Topologie et organisation structurale des transporteurs ABC : .....	4
2.1.1. Les domaines transmembranaires (TMDs) : .....	4
2.1.2. Les domaines de fixation des nucléotides (NBDs) : .....	5
2.2. Classification : .....	5
2.3. Mécanisme d'action des transporteurs ABC : .....	6
2.4. Les transporteurs ABC humains : .....	6
2.5. Les transporteurs « Multi Drug Résistance » : .....	7
2.6. Les transporteurs ABCG2 : .....	8
2.6.1. Localisation : .....	8
2.6.2. Rôles physiologiques : .....	9
2.6.3. Substrats d'ABCG2 : .....	10
2.6.4. Inhibiteurs d'ABCG2 : .....	11
2.7. Les quinazolines : .....	11
2.7.1. Synthèse : .....	11

2.8. Intérêt biologique et pharmacologique de quelques inhibiteurs d'ABC2 :.....	12
--	----

## **CHAPITRE2 : la méthodologie relation quantitative structure-activité (QSAR)**

1. introduction : .....	14
1.1. Historique: .....	14
1.2. Principe : .....	14
1.3. Processus du model QSAR : .....	15
1.4. Les descripteurs : .....	16
1.4.1. Types de descripteurs : .....	16
1.4.1.1. Les descripteurs 1-D : .....	16
1.4.1.2. Les descripteurs 2-D : .....	16
1.4.1.3. Les descripteurs 3-D : .....	16
1.5. Technique Relation Structure Activité Quantitative : .....	17
1.5.1. Régression linéaire multiple : .....	17
1.6. Paramètres statistiques : .....	18
1.6.1. Coefficient de détermination : .....	18
1.6.2. Le Coefficient de détermination Ajuste noté <b>Rajusté2</b> : .....	18
1.6.3. Le test de Fisher ou F : .....	19
1.7. Validation des modèles : .....	19
1.7.1. Validation interne : .....	19
1.7.2 Validation externe : .....	20
1.7.3. Test de randomisation : .....	21
1.8. Domaine D'application (DA) : .....	21

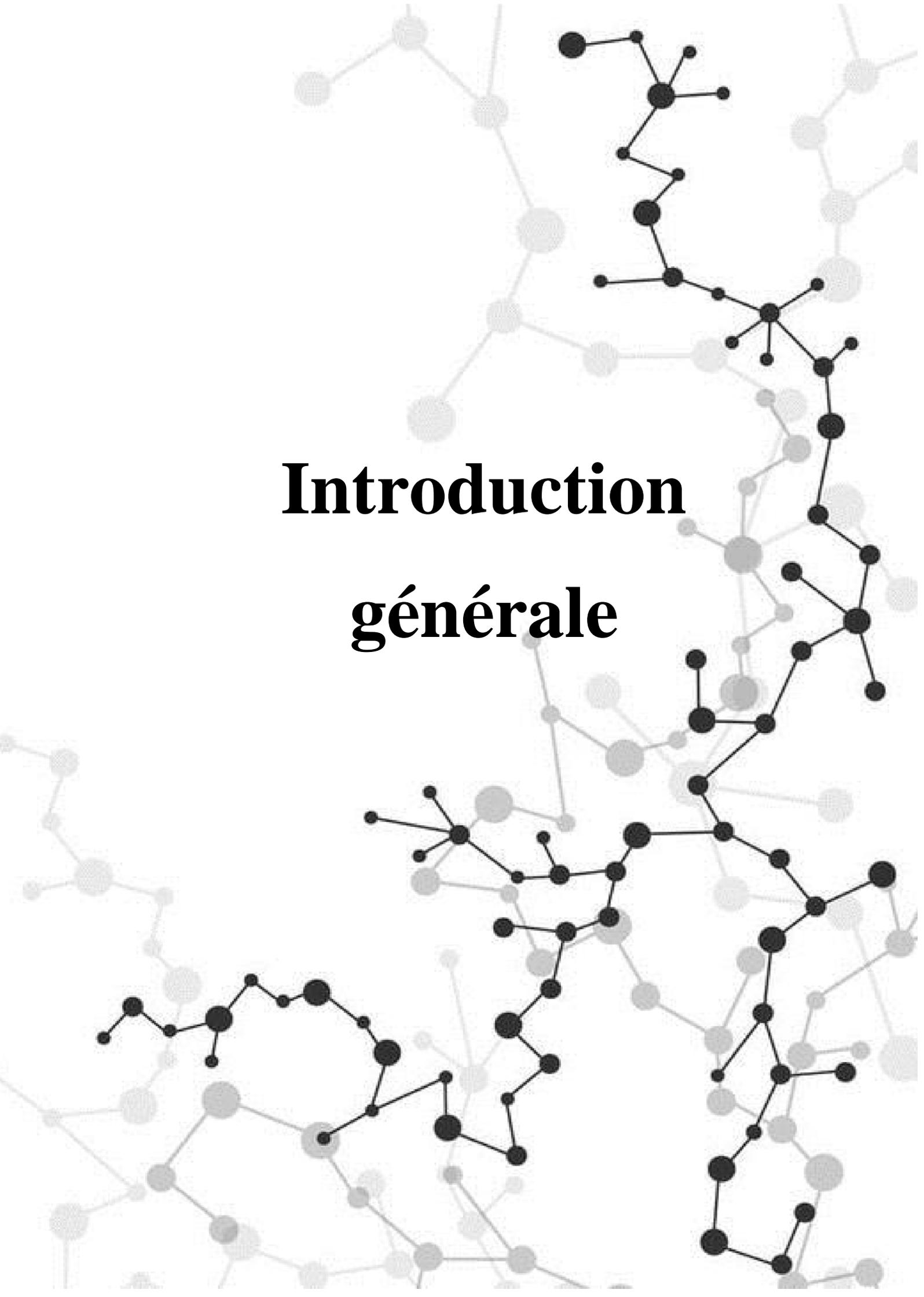
## **Partie2 : Etude expérimentale**

1. Méthodologie : .....	23
1.1. Collecte des données : .....	23
1.2. Calcul des descripteurs moléculaires : .....	27
1.2.1. Préparation de l'ensemble de données : .....	27
1.2.2. Développement et outils de validation du model : .....	27

1.2.3. Evaluation du modèle :.....	27
2.Résultats et discussion : .....	28
2.1. La colinéarité :.....	28
2.2. Analyse de la régression :.....	29
2.3. Qualité de l'ajustement : .....	30
2.4. Validation :.....	36
2.4.1 Validation interne :.....	36
2.4.2 Validation externe : .....	36
2.4.3 Test de randomisation : .....	36
2.5. Domaine chimique : .....	37
Conclusion générale.....	39

## **Références**

## **Annexes**



# **Introduction générale**

## Introduction générale

Le cancer du sein est un problème majeur de santé publique. C'est le premier cancer le plus fréquemment diagnostiqué dans le monde et le premier cancer chez la femme. C'est une maladie qui désigne les tumeurs malignes agressives où les cellules deviennent de plus en plus grosses [1].

Par conséquent, l'un des enjeux majeurs dans le traitement du cancer du sein est la découverte de nouveaux médicaments d'une manière plus efficace, économique et rapide [2]. Il s'agit en effet de trouver des molécules qui présentent des propriétés thérapeutiques particulières, ce processus demande généralement entre 10 et 15 ans de recherches de plus, les dépenses pour la recherche et développement dépassent les 900 millions de Dollars [3,4].

Il est donc crucial d'utiliser des méthodes pour compenser les inconvénients de l'expérience, pour cette raison, l'industrie pharmaceutique s'oriente vers de nouvelles méthodes spécifiques de la chimio-informatique qui aide à modéliser et prédire de façon rapide l'activité avant même que celles-ci ne soient synthétisées [5]. Parmi ces méthodes, les relations structure-activité ou **QSAR** (en anglais **Quantitative Structure-Activity Relationships**). Elles consistent essentiellement la recherche d'une relation entre les structures des molécules et leurs activités en utilisant des méthodes d'analyses multidimensionnelles par apprentissage statistique afin de modéliser l'activité biologique [6-9].

De nombreux chercheurs ont étudié les quinazolines et ces dérivés (**gefitinib**) comme nouveaux médicaments anticancéreux qui présentent un large spectre d'activités pharmacologiques [10,11], ces composées agissent comme des inhibiteurs d'ABCG2, qui est une protéine de transport d'efflux polymorphe [12-14].

L'objectif de ce travail vise à utiliser la méthodologie QSAR afin d'élaborer un modèle robuste et fiable qui explique et prédit l'activité anti-cancéreuse d'une série de 87 dérivés de quinazoline inhibiteurs d'ABCG2.

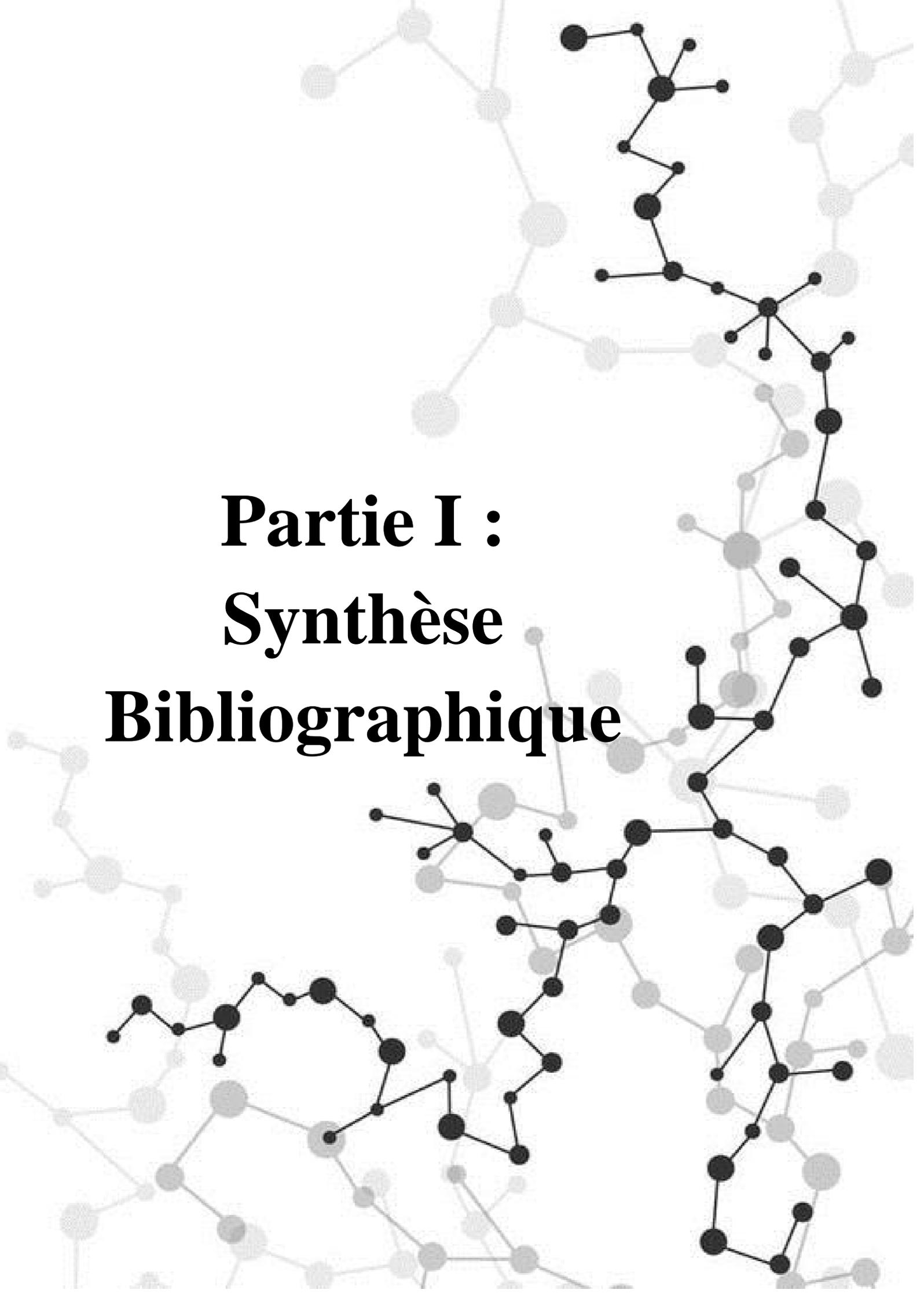
Notre mémoire est divisée en deux parties précédées par une introduction générale :

Une partie bibliographique contient deux chapitres :

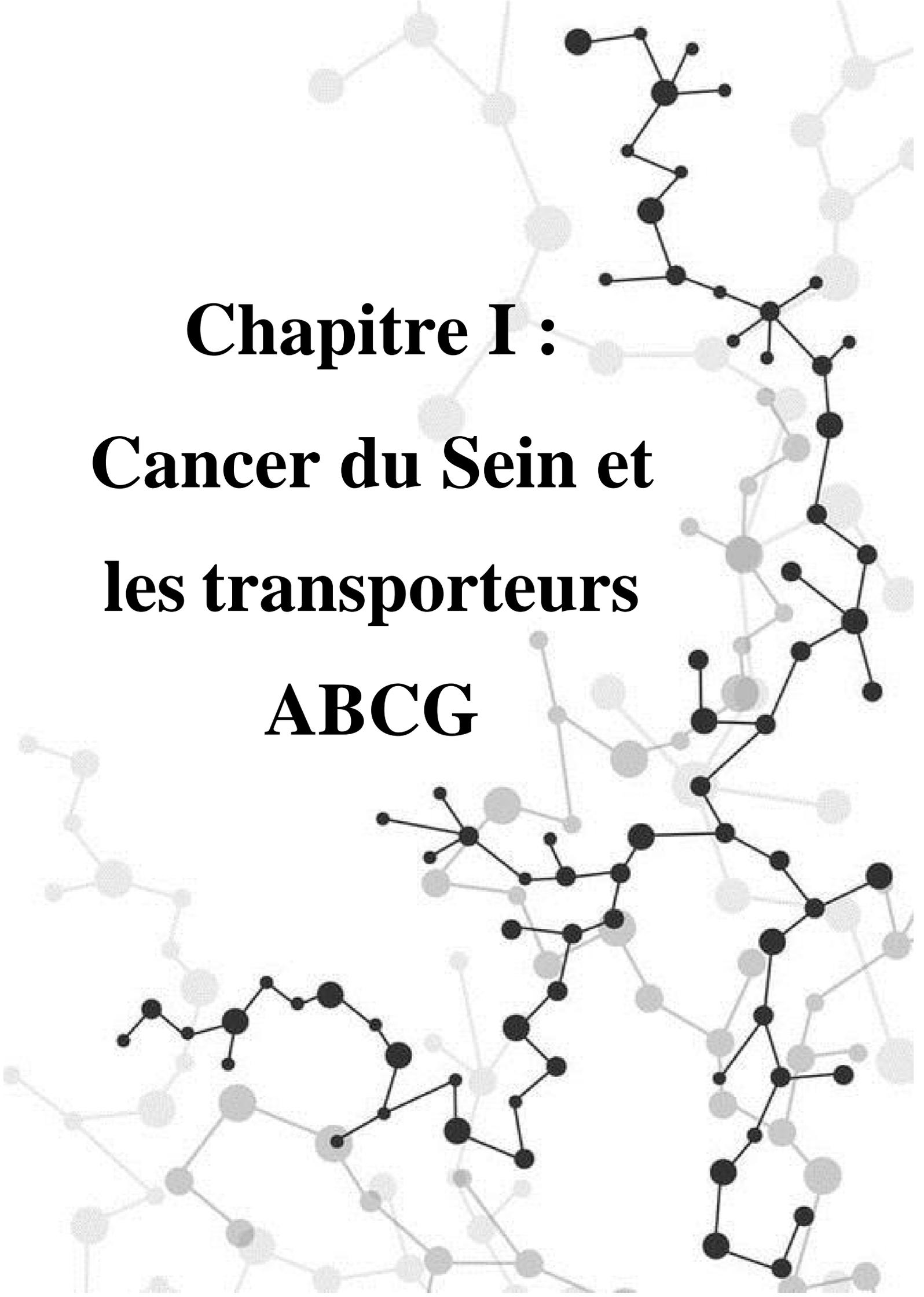
\* Dans le premier chapitre nous présentons un aperçu général sur le cancer du sein et les transporteurs ABCG.

\* L'objectif du deuxième chapitre est de rappeler la méthodologie QSAR (de l'anglais quantitative structure activity relationship).

Et une partie expérimentale où nous présenterons et nous discuterons les résultats obtenus pour la méthodologie QSAR que nous avons effectuée au long de notre séjour au centre de recherche en biotechnologie de Constantine. Et nous achèverons le travail par une conclusion générale.

The background features a complex network of molecular structures. It consists of numerous interconnected nodes (represented by circles) and edges (represented by lines). The nodes and lines are rendered in two colors: a light gray and a dark black. The structures are scattered across the page, with some appearing more prominent than others. The overall effect is a dense, abstract representation of a chemical or biological network.

**Partie I :**  
**Synthèse**  
**Bibliographique**

The background features a complex network of molecular structures. It consists of numerous interconnected nodes (represented by circles of varying sizes and shades of gray) and edges (represented by thin lines). Some structures are rendered in a darker, more prominent black color, while others are lighter and more faded, creating a sense of depth and complexity. The overall appearance is that of a molecular graph or a network diagram, typical of scientific or medical publications.

**Chapitre I :**  
**Cancer du Sein et**  
**les transporteurs**  
**ABCG**

### 1. introduction :

Le cancer du sein désigne les tumeurs malignes agressives c'est-à-dire des cellules qui se multiplient de manière anarchique et qui deviennent de plus en plus grosses [1]. C'est une maladie très courante surtout chez les femmes et qui est l'une des principales causes de décès [15]. Seule une partie des cancers du sein, 5 à 10 %, sont héréditaires dû à une anomalie au niveau d'un gène (anomalie génétique) qui se transmet d'une génération à une autre [16]. Chez les femmes ménopausées, l'obésité favorise le développement du cancer du sein en raison de l'augmentation des concentrations sériques d'œstrogènes libres. Par contre, avant la ménopause l'obésité a été associée à une diminution du risque de cancer du sein, en raison de l'augmentation des cycles anovulatoires menant à une diminution des taux d'œstrogènes circulants [17].

#### 1.1. Statistiques du cancer du sein :

Au monde, selon des chiffres récents en 2018, le cancer du sein est de loin le cancer le plus fréquent chez les femmes (24 %) avec 2,1 millions de personnes diagnostiquées et de 627.000 décès (72 décès/heure) dans le monde.

Les taux d'incidence (voir l'annexe) du cancer du sein sont les plus élevés dans les régions développées. (OMS ,2018).

En ALGERIE selon l'OMS En 2018, il y a eu 53076 nouvelles personnes touchées par le cancer et plus de 29 000 décès en Algérie et selon les derniers chiffres du Centre international de recherche sur le cancer (CIRC), au cours des cinq dernières années, le nombre de cas prévalent égale à 127 306 malades. Il est également indiqué que le nombre de femmes malades environ de 29112, avec en première position le cancer du sein.

### 2. Les transporteurs des médicaments ABC :

Les transporteurs ABC « ATP-Binding Cassette » est une large famille des transporteurs, regroupe un nombre élevé de protéines chez toutes les espèces, des micro-organismes jusqu'à l'homme. Il existe 70 transporteurs ABC chez la bactérie [18] impliqués dans l'import et l'export des molécules et de la résistance aux antibiotiques. Chez l'homme, on a 48 gènes codant pour ces transporteurs [19].

## 2.1. Topologie et organisation structurale des transporteurs ABC :

Les transporteurs ABC sont constitués de deux domaines transmembranaires (**TMD** en anglais Trans-Membrane Domain) assurent la reconnaissance des substrats et deux domaines de fixation des nucléotides (**NBD** en anglais Nucleotide-Binding Domain) qui hydrolysent l'ATP et fournissent l'énergie nécessaire au transport de ces substrats, d'où le nom « ATP-Binding Cassette ». Cette structure est relativement conservée. On peut trouver ces domaines sur une seule chaîne polypeptidiques ou sur deux chaînes polypeptidiques (demi-transporteurs) comporte chacune un TMD et un NBD et afin d'être fonctionnels ils doivent s'associer par deux. La figure 1 montre la topologie des transporteurs ABC [20-25].

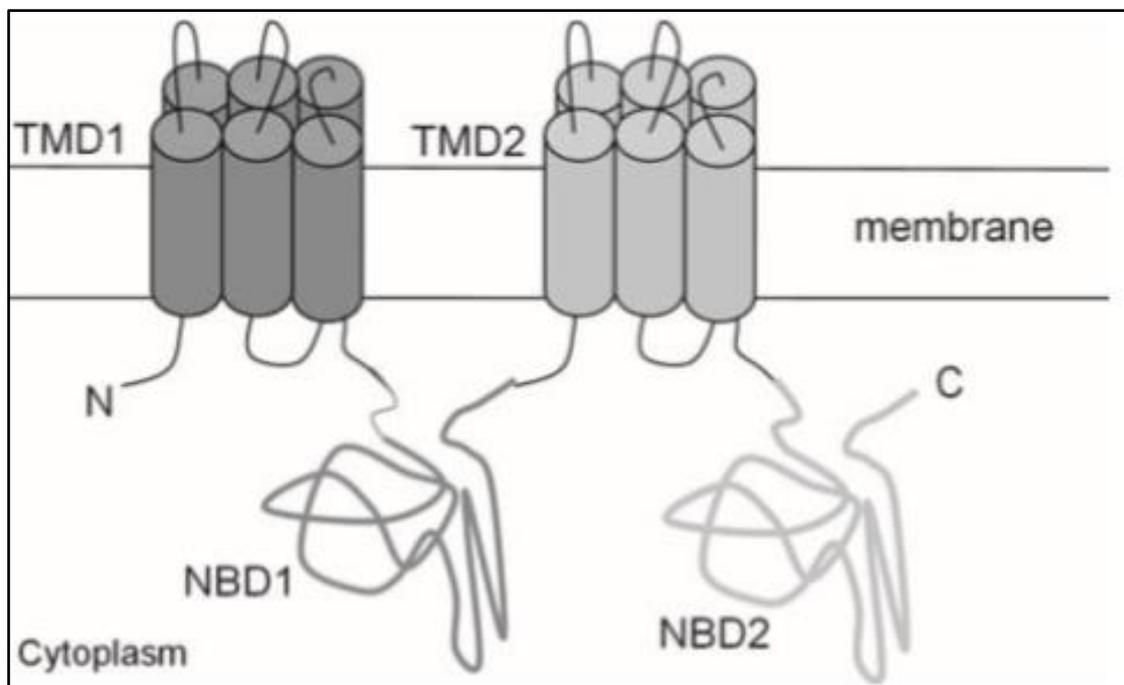
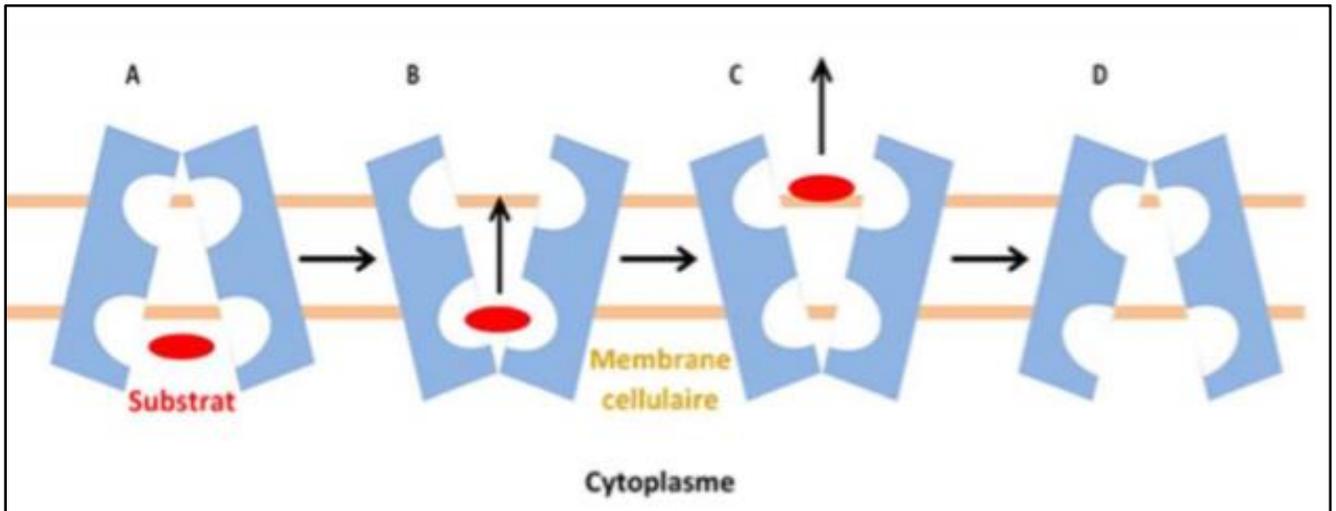


Figure 1 : Topologie des transporteurs ABC [26]

### 2.1.1. Les domaines transmembranaires (TMDs) :

Les TMDs sont très peu conservés entre les différents transporteurs ABC, ce qui est en accord avec la grande variété de substrats que l'on peut retrouver au sein de cette famille de protéines. Les deux TMD s'associent pour former un dimère traversant la membrane plasmique, une extrémité en contact du cytoplasme et l'autre avec le milieu extracellulaire (Voir figure 2). Elles sont donc capables de prendre en charge de nombreuses molécules différentes allant de xénobiotiques aux produits de métabolisation, en passant par des peptides, des vitamines, le cholestérol ou les ions [21,24,25].



**Figure 2 :** Représentation schématique du transfert des substrats à travers la membrane par les TMD

### 2.1.2. Les domaines de fixation des nucléotides (NBDs) :

Les NBD sont des domaines hautement conservés chez tous les transporteurs ABC. Ces domaines sont hydrophiles et interagissent avec les boucles intracellulaires des TMD. C'est en effet au sein de ces derniers que l'ATP se fixe et s'hydrolyse afin de fournir l'énergie qui permet la translocation des substrats. Plus de 50 structures de NBD ont été déterminées [27]. Ces domaines montrent une structure contenant 2 lobes :

- ✓ **Un lobe large** contenant les motifs Walker A et B et les boucles A, Q et H.
- ✓ **Un lobe plus petit** en hélices, contenant la signature S et la boucle D [21].

### 2.2. Classification :

Les transporteurs ABC se divisent en trois classes majeures selon l'étude phylogénétique par alignement de séquence in silico des 600 protéines ABC répertoriées [28].

- ✓ **La classe 1** contient tous les exportateurs majeurs dont les domaines transmembranaires (TMDs) sont fusionnés aux domaines nucléotidiques (NBDs) (EX : P-gp, MRP1 et ABCG2).
- ✓ **La classe 2** comprend les protéines ABC sans domaine (TMD) connu.
- ✓ **La classe 3** contient toutes les BDP (Binding Dependand Protein), notamment les protéines dont les domaines (TMD) ne sont pas fusionnés aux domaines (NBD).

### 2.3. Mécanisme d'action des transporteurs ABC :

La transduction de l'énergie des NBD vers les TMD va permettre la translocation des substrats d'un côté à l'autre de la membrane plasmique. Des études ont montré que la fixation de l'ATP sur les NBD même sans hydrolyse est suffisante pour induire des changements de conformation dans les TMD d'ABCG2 [10]. Plusieurs modèles permettent d'expliquer ce mécanisme de transport. Le plus couramment utilisé : « L'ATP switch model » [29] (voir Figure 3), dont les étapes sont comme suit :

- ✓ **Etape I** : la fixation du substrat sur les TMD augmente l'affinité du transporteur pour l'ATP et provoque le passage de la conformation ouverte vers la conformation fermée du dimère.
- ✓ **Etape II** : le changement conformationnel expose le site de fixation des xénobiotiques du côté extracellulaire. Le substrat est libéré.
- ✓ **Etape III** : l'ATP est hydrolysée.
- ✓ **Etape IV** : le dimère reprend sa conformation ouverte.

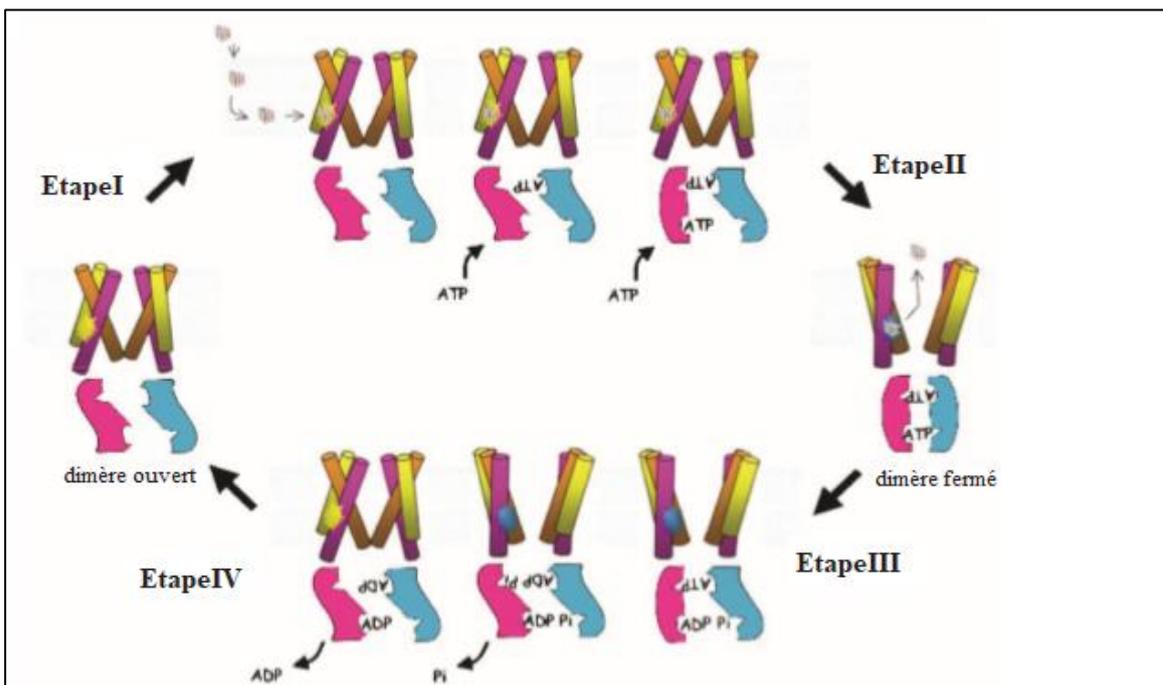


Figure 3 : Modèle de transport des drogues "ATP switch model" [29].

### 2.4. Les transporteurs ABC humains :

Les transporteurs ABC humains sont présents dans toutes les parties de l'organisme. Ils sont classés selon leur position dans l'arbre phylogénétique en 7 sous-familles de « ABCA à ABCG » [11].

- **La sous-famille A** comprend 12 transporteurs entiers. Elle constitue les protéines ABC les plus grandes.
- **La sous-famille B** comprend 4 transporteurs entiers et 7 demi-transporteurs.
- **La sous-famille C** comprend 12 transporteurs entiers impliqués dans diverses fonctions comme le transport des ions et la sécrétion de toxines [30].
- **La sous-famille D** comporte 4 demi-transporteurs présents exclusivement dans le peroxysome fonctionnent comme homo- ou hétéro-dimères pour transporter les acides gras.
- **La sous-famille E et F** font parties de la classe 2 des transporteurs ABC composées seulement de domaine de fixation de nucléotides, ne semblent pas impliqués dans le transport membranaire de substrats.
- **La sous-famille G** comprend 6 demi-transporteurs inversés, signifiant que le domaine TMD est placé derrière le domaine NBD dans la séquence (alors que l'inverse est généralement observé, Figure 4). Elle comprend **ABCG2** (ou BCRP) qui nous intéresse plus particulièrement dans cette étude.

### 2.5. Les transporteurs « Multi Drug Résistance » :

Parmi les transporteurs ABC les plus connus depuis très longtemps, la classe des protéines « Multi Drug Résistances » (MDR) ou la résistance croisée des cellules cancéreuses vis-à-vis des agents anticancéreux. Elle se réfère principalement à une surexpression de transporteurs membranaires dans les cellules tumorales qui sont à l'origine de l'efflux des médicaments et donc de la chimiorésistance, dans ce contexte certaines tumeurs montrent un profil de résistance inhérente, tandis que d'autres développent des résistances au cours du traitement contre un certain nombre de composés chimio thérapeutiques, ce qui rend le traitement inefficace [22-24].

Chez l'homme, les trois principaux types de transporteurs MDR comprennent les sous-familles : **ABCB** (ABCB1/ MDR1/P-glycoprotéine), **ABCC** (ABCC1/ MRP1, ABCC2/ MRP2, probablement aussi ABCC3-6, et ABCC10-11) et **ABCG** (ABCG2/MXR/BCRP) [31, 22-25]. Parmi ces MDR, seuls **ABCB1**, **ABCC10** et **ABCG2** présenteraient un intérêt clinique [22]. En terme fonctionnel, les transporteurs MDR jouent un rôle important de protection à l'échelle cellulaire. En effet, il s'agit de pompe d'efflux membranaire, de type ATPasique, avec un rôle de détoxification. De plus, ces pompes reconnaissent des structures très variées (médicaments, polluants ...) (Voir Figure 4). Elles expulsent vers le milieu extracellulaire ces xénobiotiques entrés passivement ou via d'autres transporteurs membranaires dans les cellules [31,23,24].

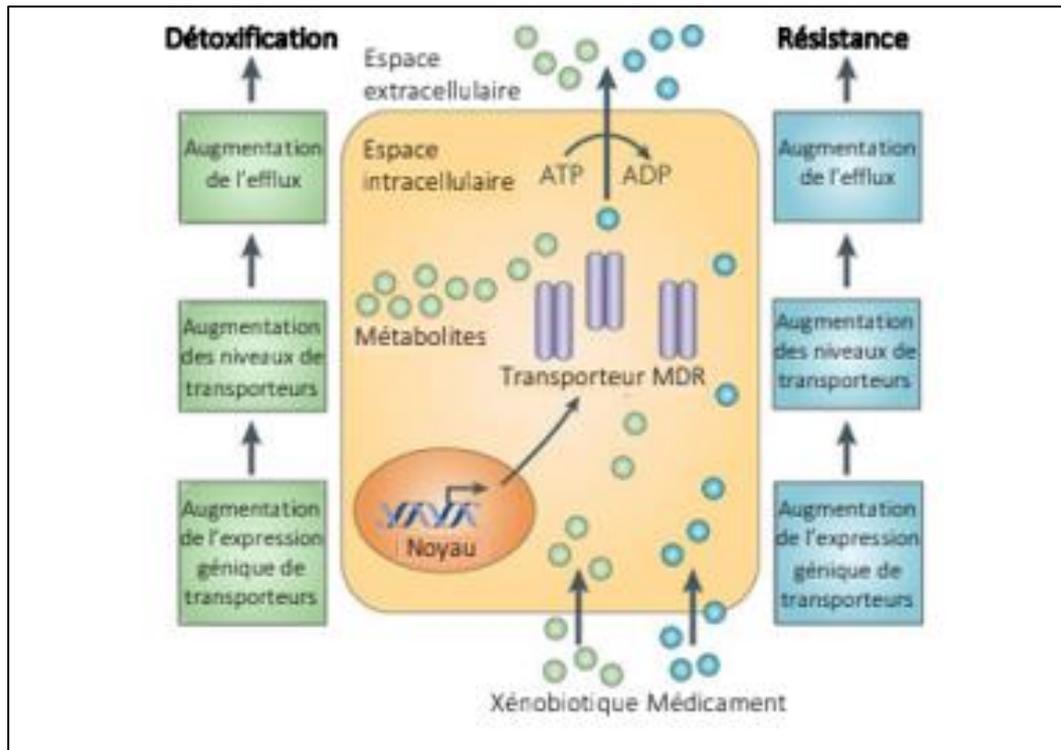


Figure 4 : Les transporteurs ABC et leur implication dans la résistance aux médicaments [32].

## 2.6. Les transporteurs ABCG2 :

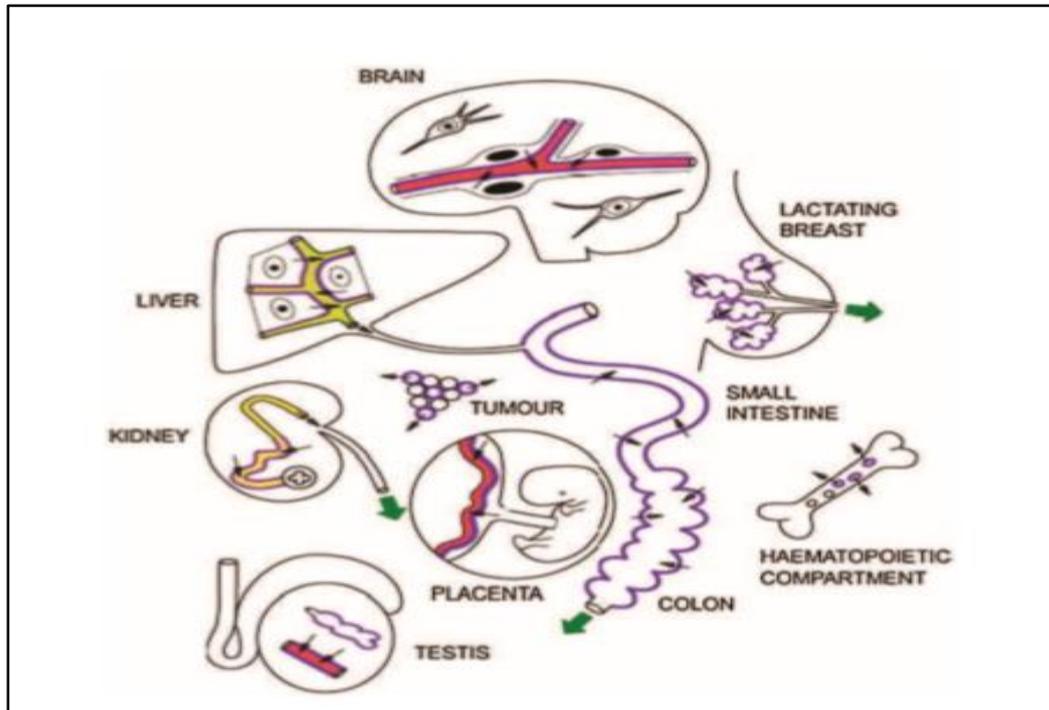
La protéine ABCG2 a été découverte par trois équipes différentes à partir de 1998. En effet, **Doyle** et ses collaborateurs clonent les premiers le gène responsable d'un phénotype de résistance d'une lignée cellulaire de cancer du sein (**MCF-7 Adr/Vp**), nommée la protéine **BCRP** pour « Breast Cancer Resistance Protein » [33]. La même année, **Allikmets** et ses collaborateurs l'ont découvert dans le placenta, d'où son premier nom **ABCP** « ABC transporter in Placenta » ; le gène se trouve sur le chromosome 4q22 [34]. Enfin, **Miyake** et ses collaborateurs ont cloné le gène *Abcg2* dans des cellules cancéreuses du colon (**S1-M1-80**) résistantes à la mitoxantrone ; ils l'ont appelé **MXR** pour « MitoXantrone Resistance-associated gene » [35]. Plus tard, L'analyse phylogénétique a montré que ce gène appartenait à la sous-famille **G** et a été nommé **ABCG2** par la nomenclature HUGO (**H**uman **G**enome **O**rganisation).

### 2.6.1. Localisation :

- **Tissulaire :**

ABCG2 présente physiologiquement dans nombreux tissus de l'organisme. Elle est particulièrement abondante dans le foie, les intestins, le placenta, la barrière hématoencéphalique, la

glande mammaire et les cellules souches [33,36-38]. En revanche, elle est absente dans certains tissus comme le cœur, les poumons et les reins. La Figure 5 récapitule les principaux sites d'expression de cette protéine indiquée en lignes violettes. A chaque site d'expression, les petites flèches indiquent la direction du transport par ABCG2. Les flèches vertes indiquent l'excrétion des substrats de l'organisme.



**Figure 5 :** Vue schématique de l'expression d'ABCG2 dans l'organisme [39].

- **Cellulaire:**

L'ABCG2 est localisée essentiellement à la membrane plasmique, du côté apical des cellules polarisées [40-42] d'autres études ont mis en évidence l'expression de la protéine dans les lysosomes [43] et plus récemment au niveau de la membrane interne de la mitochondrie [44].

### 2.6.2. Rôles physiologiques :

La distribution de BCRP dans les différents tissus est liée à ses nombreux rôles physiologiques [24,39]. En effet, elle limite l'absorption dans le petit intestin des médicaments et de certains composés potentiellement toxiques présents dans l'alimentation donc la protection contre les endo- et xénobiotiques [45-46]. Elle est responsable de la distribution dans les barrières hématoencéphalique et hémato placentaire, ceci suggère un rôle important de protection du fœtus, en favorisant l'efflux des drogues et des toxines du placenta vers la circulation sanguine maternelle [47-49] en plus de faciliter

l'élimination et l'excrétion dans le foie et les reins [50]. Par ailleurs, elle est localisée du côté apical des cellules épithéliales alvéolaires des glandes mammaires, principal lieu de production du lait. Son expression est multipliée par 30 en période de lactation et diminue avec le sevrage, ce qui suggère une implication dans l'apport de nutriments dans le lait maternel [12]. Son rôle est cependant ambivalent dans cet organe, car malgré sa capacité à transporter plusieurs vitamines dans le lait, telle la vitamine B2 ou riboflavine [51], ABCG2 est également responsable de l'accumulation de toxines et de substances médicamenteuses comme la nitrofurantoïne ou la ciprofloxacine [12-14]. Enfin, elle régule l'efflux de l'hème et des porphyrines hors des cellules comme les érythrocytes et les cellules souches [14,52].

### 2.6.3. Substrats d'ABCG2 :

Les substrats cibles de la BCRP sont des molécules hydrophobe ou/et des molécules chargées positivement ou négativement et amphiphiles avec un haut poids moléculaire comme c'est le cas de nombreux anticancéreux. Elle peut prendre en charge l'efflux d'autres xénobiotiques comme des antibiotiques. ABCG2 a la capacité de transporter de nombreuses molécules endogènes et exogènes (Figure 6). Les premiers substrats à avoir été mis en évidence sont les agents chimio thérapeutiques comme la **Mitoxantrone**, le **Flavopiridol** et l'**Irinotécan** [53-54]. D'autres classes pharmacologiques, comme les antiviraux ou les antibiotiques peuvent être transporter également par l'ABCG2 [52,55] ainsi que des composés naturels exogènes, comme les **flavonoïdes** [56] ou le **phéophorbide a** [45,57] et endogènes, tels la **protoporphyrine IX** [58], l'**acide urique** [59] et les **oestrogènes** [60-61].

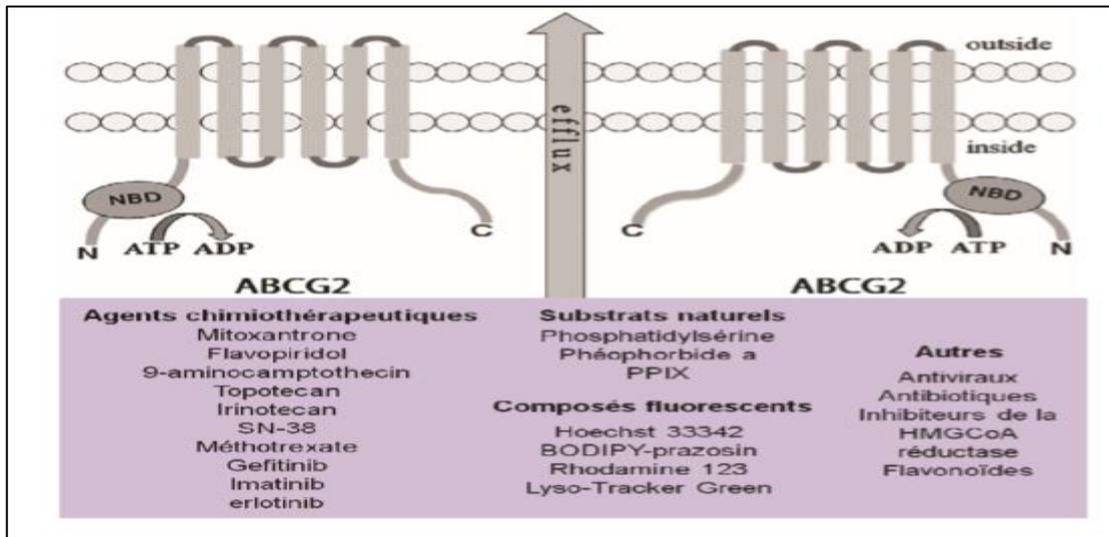


Figure 6 : Schéma d'ABCG2 et liste (non exhaustive) de ses substrats [62].

#### 2.6.4. Inhibiteurs d'ABCG2 :

Plusieurs études concernant les inhibiteurs du transporteur ABCG2 ont été découvertes ces dernières années. Parmi ceux-ci, on distingue deux grands types : d'une part il y'a les inhibiteurs non spécifiques, qui inhibent l'activité d'autres transporteurs ABC ainsi que celle d'ABCG2 et d'autre part les inhibiteurs spécifiques, qui sont actifs exclusivement sur le transporteur ABCG2 tel que le **géfitinib** (ZD1839) [89], l'**EKI-785** ou encore **Erlotinib** [90].

#### 2.7. Les quinazolines :

La diversité des activités biologiques, a incité de nombreux chercheurs à étudier les **quinazolinones** en tant que nouvelles molécules médicamenteuses. Ils ont étudié et rapporté avec succès les propriétés antimicrobiennes et les relations structure-activité (RSA) de plusieurs dérivés de la **quinazolinone**. [63-74]. Certains dérivés de la **quinazoline** ont été approuvés comme médicaments comme la **prazosine** pour traiter l'hyperplasie bénigne de la prostate et la **doxazosine** pour le syndrome de stress post-traumatique [75], et l'**erlotinib** et le **géfitinib** sont utilisés pour la guérison des cancers du poumon et du pancréas [76]. Le **lapatinib** efficace dans le traitement du cancer du sein [77].

##### 2.7.1. Synthèse :

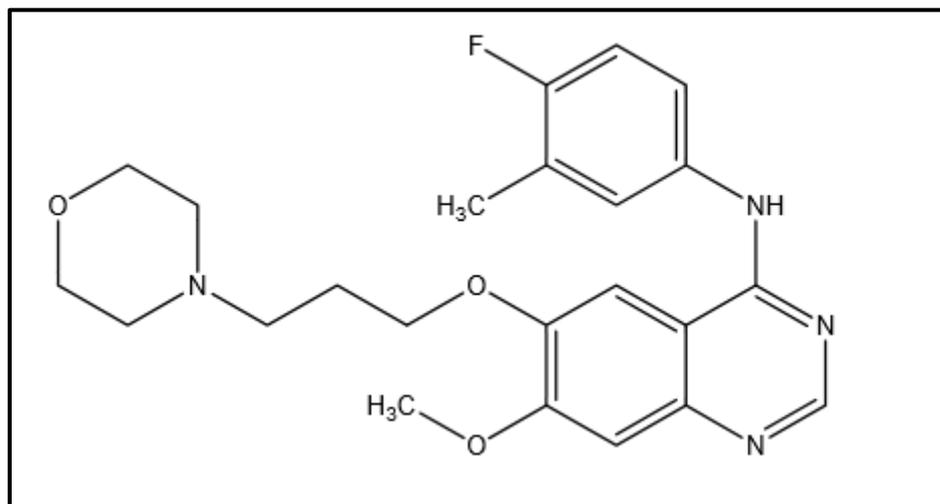
La première synthèse de la quinazoline a été rapportée en 1895 par **August Bischler et Lang** par la décarboxylation du dérivé 2-carboxy (acide quinazoline-2-carboxylique) [78]. En 1903, **Gabriel**

a rapporté la synthèse de la quinazoline mère à partir de l'o-nitrobenzylamine, qui a été réduite avec de l'iodure d'hydrogène et du phosphore rouge en 2-aminobenzylamine, ce dernier réduit se condense avec l'acide formique pour donner la dihydroquinazoline, qui est oxydée en quinazoline [79].

## 2.8. Intérêt biologique et pharmacologique de quelque inhibiteurs d'ABCG2 :

- **Gefitinib :**

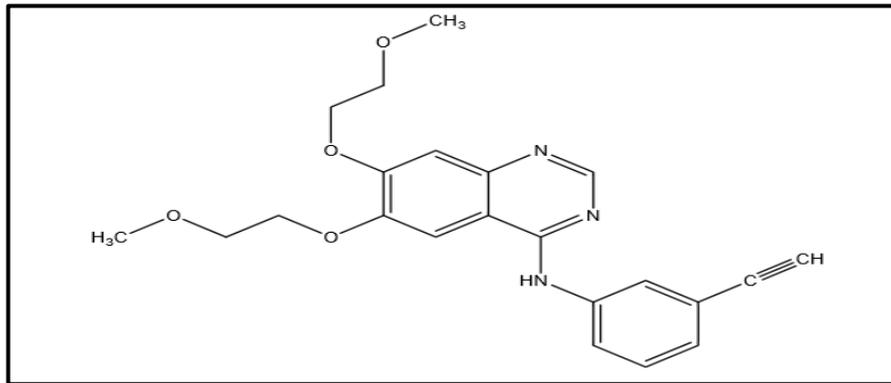
C'est un médicament contenant de la quinazoline (voir figure 7), également connu sous le nom d'Iressa, commercialisé par **AstraZeneca**. Il s'agit d'un médicament utilisé dans le traitement de certains types de cancer. Le gefitinib est un inhibiteur de la protéine kinase du récepteur du facteur de croissance épidermique (EGFR). Il se lie au site de liaison à l'ATP de l'EGFR, inactivant ainsi la cascade de transduction du signal Ras anti-apoptotique, empêchant la croissance des cellules cancéreuses [80-82].



**Figure 7 :** Structure de Gefitinib [89]

- **Erlotinib :**

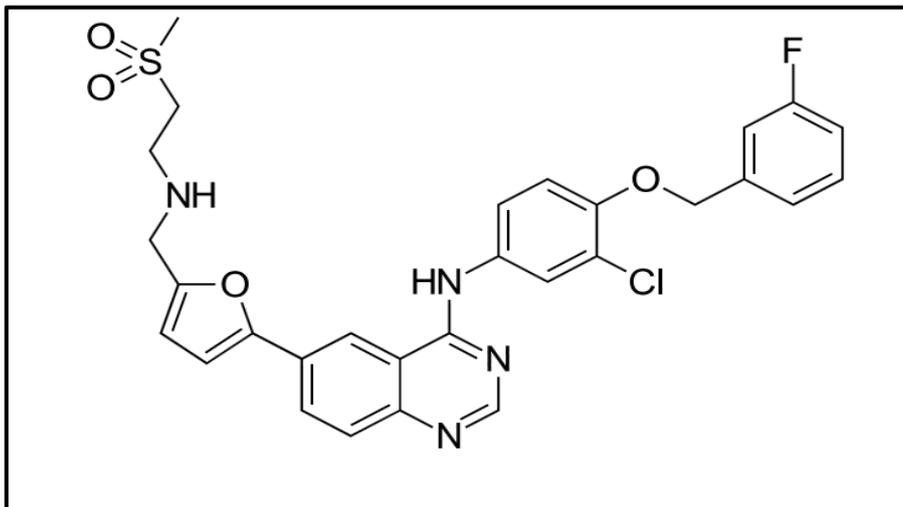
C'est un médicament fabriqué par Astellas, connu par le nom commercial «Tarceva ». Il s'agit d'un inhibiteur de la tyrosine kinase du récepteur du facteur de croissance épidermique (EGFR). La liaison de l'Erlotinib aux sites de liaison à l'ATP des récepteurs de l'EGFR les empêche de produire des résidus de phosphotyrosine, rendant ainsi le récepteur incapable de générer des cascades de signaux pour favoriser la croissance cellulaire (Figure 8) [83-84].



**Figure 8 :** Structure d'Erlotinib [90]

- **Lapatinib :**

C'est un médicament destiné pour traiter le cancer du sein avancé ou métastatique en combinaison avec la capécitabine (voir Figure 9). Le lapatinib élimine la croissance des cellules souches du cancer du sein qui provoquent la croissance de la tumeur. IL s'agit d'un inhibiteur de la protéine kinase du récepteur 2 du facteur de croissance épidermique humain (HER2), il se lie au site de liaison de l'ATP du récepteur (HER2) et inhibe de façon réversible l'activation du mécanisme de signalisation donc une diminution de la croissance cellulaire [85-88].



**Figure 9 :** Structure de Lapatinib



The background features a complex molecular structure graphic. It consists of numerous interconnected nodes (represented by circles of varying shades of gray and black) and edges (represented by thin lines). The structure is dense and multi-layered, with some nodes being significantly larger than others, suggesting a hierarchical or complex network. The overall appearance is that of a molecular model or a network diagram, rendered in a grayscale palette.

**Chapitre II :**  
**La méthodologie**  
**Relation Quantitative**  
**Structure- Activité**  
**(QSAR)**

### 1. introduction :

La relation quantitative entre l'activité biologique et la structure moléculaire (RQAS) connue beaucoup plus sous le terme (QSAR) (en anglais Quantitative structure-activity Relationship) est une méthode qui cherche à établir une relation entre les structures moléculaires, codées en valeurs numériques (dites descripteurs) avec leurs activités. Ce domaine d'étude, utilise des techniques de modélisation afin de prédire l'activité biologique et de proposer des bons modèles pour le développement d'un bon médicament [91-94].

#### 1.1. Historique :

Au 19ème siècle et plus précisément en 1868 **Crum-Brown et Fraser** ont découvert l'existence de relations entre les structures chimiques et les activités physiologiques [95]. Les premiers travaux qui utilisent la méthodologie QSAR sont dus à **Hansch et Free et Wilson** dans lequel **Hansch** a proposé des modèles reliant directement l'activité biologique des composés avec les propriétés hydrophobes, électroniques et stériques à l'échelle moléculaire et **Free et Wilson** ont développé des modèles empiriques pour l'étude de l'activité biologique [96,97]. Au cours de ces dernières décennies, l'utilisation des méthodes QSAR reste à progresser, elle est même devenue indispensable en chimie pharmaceutique, en toxicologie et surtout pour la conception de médicaments [98,99].

#### 1.2. Principe :

Les méthodes QSAR reposent sur le principe d'élaborer modèle mathématique reliant de manière quantitative la structure moléculaire, codée par descripteurs moléculaires avec l'activité biologique de ces derniers en utilisant des méthodes d'apprentissage statistiques pour le développement des modèles prédictifs sous la forme suivante [100] :

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

Pour éviter que cette équation ne soit non significative, il faut suivre les conditions suivantes :

- (Nombre de molécules) / (Nombre de descripteurs) > 5
- Les descripteurs ne doivent pas être corrélés entre eux [122] et en pratique, ils seront alors enlevés lors de l'analyse par la Régression linéaire multiple.

### 1.3. Processus du model QSAR :

Pour prédire l'activité biologique par la méthode QSAR on doit passer par un processus qui est présenté par la Figure 10.

Nous devons Collecter et préparer les structures de chaque composé, pour que la valeur calculée des chaque descripteurs soit optimal et ça se fait à l'aide des logiciels adaptés.

Les descripteurs sélectionnés seront utilisés pour la construction du model QSAR en utilisant l'algorithme génétique (AG), puis évaluer ce modèle par ces paramètres statistiques [101].

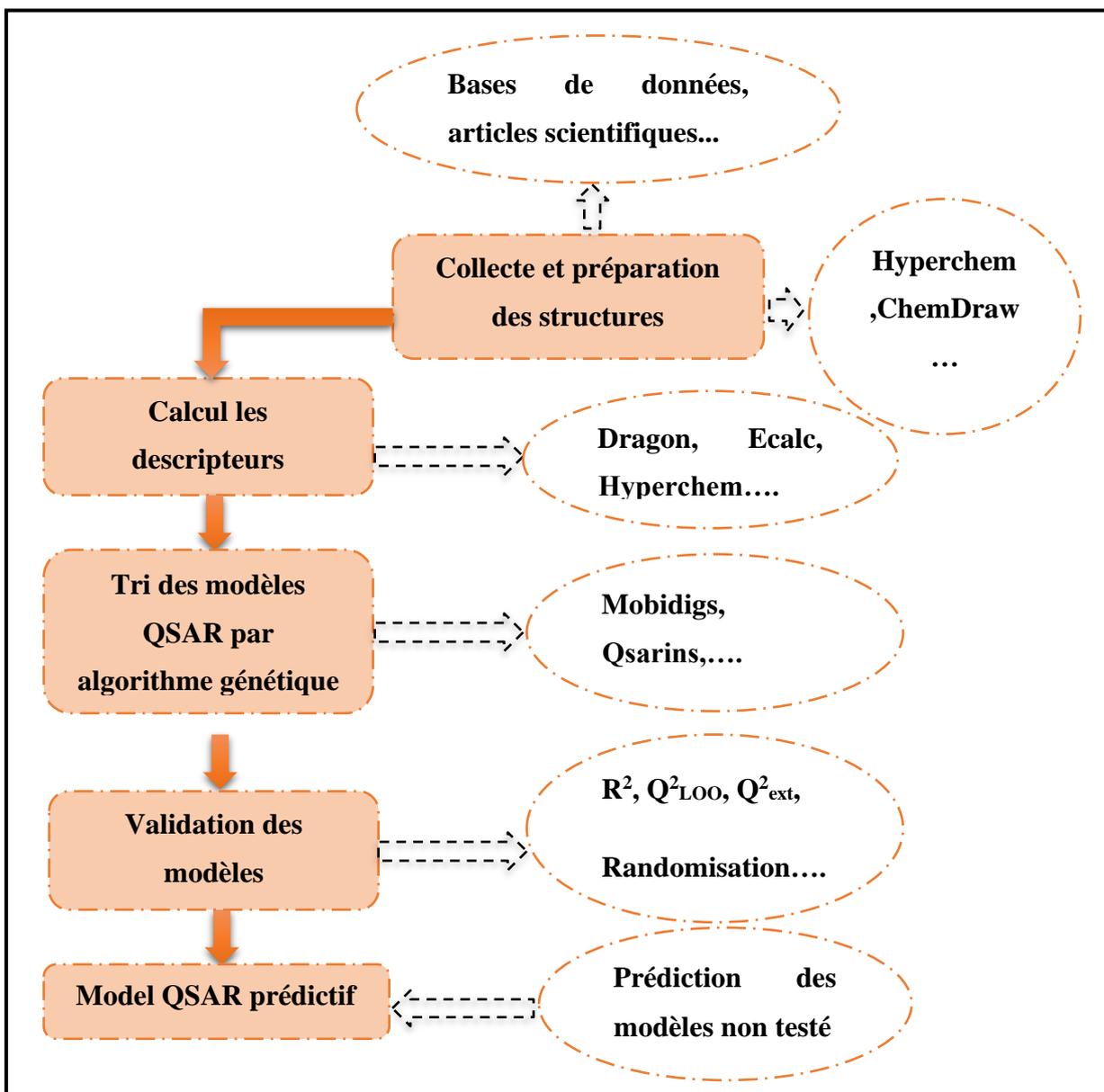


Figure 10 : Processus de la construction d'un model QSAR.

### 1.4. Les descripteurs :

Un descripteur moléculaire est un paramètre ayant une valeur numérique pour une structure chimique donnée, il peut être obtenu de manière empirique ou non-empirique. Les descripteurs préférés sont les descripteurs calculés, et non mesurés puisqu'ils permettent d'effectuer des prédictions sans avoir à synthétiser les molécules pour arriver à la modélisation moléculaire [102].

#### 1.4.1. Types de descripteurs :

Il y a plus de 6000 descripteurs moléculaires classifiés qui décrivent une structure moléculaire [103]. On peut les classer, en fonction de leur origine (constitutionnel, topologique, géométrique, quantique, thermodynamique...), et en fonction de leur dimensionnalité (1D, 2D, 3D ou 4D) [104].

##### 1.4.1.1. Les descripteurs 1-D :

Ces descripteurs représentent des propriétés générales telles que : les pourcentages massiques des atomes, la masse molaire, le poids moléculaire ... Ils sont calculés à partir de la formule brute de la molécule à l'aide de la composition moléculaire et les atomes qui la constituent [105-106].

##### 1.4.1.2. Les descripteurs 2-D :

Les descripteurs 2D sont des descripteurs qui contiennent des informations sur la taille globale du système, sa forme globale et ses ramifications sont obtenus à partir de la structure plane de la molécule. On trouve dans cette catégorie les indices topologiques [107], les indices constitutionnels...

##### 1.4.1.3. Les descripteurs 3-D :

Le calcul de ces descripteurs nécessite la connaissance de la position relative des molécules dans l'espace. Ils sont souvent nécessaires à la modélisation d'activités biologiques [108]. On cite quelques familles de descripteurs 3D, les descripteurs géométriques (volume molaire, la surface accessible, ...), les descripteurs électroniques (HOMO, LUMO...)....

### 1.5. Technique Relation Structure Activité Quantitative :

Le modèle QSAR est une équation relie la valeur de la réponse (variable à expliquer) en fonction des descripteurs (variable explicative), son développement fait appel aux techniques d'analyse de données, a fin de quantifier la relation entre l'activité et la Structure, certaines sont linéaires telles que la régression linéaire multiple (MLR), d'autres sont non linéaires comme les réseaux de neurone.

La régression linéaire multi variée (Multivariate Linear Regression - MLR) est situé parmi les méthodes d'apprentissage statistique bien adaptée et la plus employée [109,111] pour le développement de modèles prédictifs QSAR.

#### 1.5.1. Régression linéaire multiple :

La régression linéaire multiple consiste à trouver une équation mathématique linéaire, reliant la variable à expliquer (Y) et les variables explicatives (X<sub>j</sub>).

Elle a été publiée par **Hansch** en reliant l'activité biologique aux structures expérimentales lipophiliques, électroniques et stériques pour des séries de composés [110].

La méthode MLR est basé sur l'hypothèse que la variable à expliquer (Y) dépend linéairement des différentes variables explicatives, selon la relation (**équation 1**) :

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \dots + \beta_j X_{i,j} + e_i \quad (\text{eq 1})$$

Avec :

**Y<sub>i</sub>** est la i-ème observation de la variable dépendante (à expliquer ou à prédire) ; **X<sub>i,j</sub>** est la i-ème observation de la j-ème variable indépendantes (explicatives) ; *e<sub>i</sub>* est l'erreur du modèle ; **β<sub>0</sub>** est l'ordonné à l'origine de l'équation du modèle ; **β<sub>j</sub>** est le coefficient de descripteurs dans l'équation du modèle; ils indiquent le degré d'influence des descripteurs moléculaires correspondants sur l'activité cible.

**1.6. Paramètres statistiques :**

**1.6.1. Coefficient de détermination :**

Le coefficient de détermination noté  $R^2$  est compris entre 0 et 1 ( $0 \leq R^2 \leq 1$ ), sert à déterminer à quel point l'équation de régression est adaptée pour décrire la distribution des points, il représente la variabilité de y par la corrélation de y avec x.

Le  $R^2$  est calculé selon la formule (**équation 2**) :

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

Où  $\bar{y}$  est la valeur moyenne des valeurs observées pour l'ensemble du calibrage et  $\hat{y}$  est la valeur prédite.

La racine de l'écart quadratique moyen de calcul (RMSE) (**équation 3**) :

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \tag{eq 3}$$

Où  $n$  est le nombre total de composés.

**1.6.2. Le Coefficient de détermination Ajusté noté  $R_{ajusté}^2$  :**

En statistique, le coefficient de détermination, noté  $R_{ajusté}^2$  est une mesure de la qualité de la prédiction d'une régression linéaire. Plus le nombre des variables explicatives augmente plus le nombre de coefficient de détermination augmente, et ça va nous créer un problème, puisque souvent, le nombre élevé des valeurs explicatives permet de produire des modèles non robustes.

Le coefficient  $R_{ajusté}^2$  est calculé selon la formule suivante (**équation 4**) :

$$R_{ajusté}^2 = R^2 - \frac{k(1-R^2)}{n-k-1} = 1 - \frac{(1-R^2)(n-1)}{n-k-1} \tag{eq 4}$$

K : nombre de variables utilisées

Un bon ajustement correspondra à un  $R^2$  proche de l'unité. (Proche de 1)

### 1.6.3. Le test de Fisher ou F :

L'indice de Fisher F est employé afin de mesurer le niveau de signification statistique du modèle, c'est-à-dire la qualité du choix des descripteurs constituant le modèle. La statistique F est lié au coefficient de détermination par la relation suivante (**équation 5**) :

$$F = \frac{R*(N-K-1)}{(1-R^2)*K} \quad (\text{eq. 5})$$

Plus la valeur de F est grande, plus la probabilité pour que la présence de corrélation entre l'ensemble des descripteurs et l'activité biologique est grande.

### 1.7. Validation des modèles :

La validation des modèles est une étape sensible et indispensable pour éviter les erreurs. Plusieurs méthodes de validation sont employées pour évaluer le modèle QSAR et pour déterminer sa pertinence telle que la validation interne, la validation externe et le test de randomisation [112].

#### 1.7.1. Validation interne :

Dans les études QSAR la validation interne est souvent la technique la plus employée pour déterminer la stabilité du modèle et pour tester l'influence de chaque composé de l'ensemble d'apprentissage sur le modèle final [113]. Pour ce faire, on emploie la technique de la validation croisée (cross validation ou CV). On note différentes méthodes de validation croisée :  $Q^2_{LOO}$  (Leave- One-Out (LOO)),  $Q^2_{LMO}$  (Leave-Many-Out), (Bootstrap).

La validation croisée par « leave-one-out » (LOO), consiste à calculer le modèle sur (n-1) objets, et l'utiliser pour prédire la valeur (réponse) du composé éloigné. Cette procédure est répétée pour les n composés de l'ensemble étudié [114] (voir Figure 11). La somme des carrés des erreurs de prédiction désignée par l'acronyme PRESS pour (Predictive Residual Sum of Squares) est une mesure de la dispersion des estimations.

Le PRESS est utilisé pour déterminer le coefficient de prédiction ( $Q^2_{LOO}$ ), et l'écart quadratique moyen de prédiction (ou EQMP) (**équation 6 ,7**):

$$EQMP(SDEP) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_{(i)})^2} = \sqrt{\frac{PRESS}{n}} \quad (\text{eq. 6})$$

$$Q^2_{LOO} = \frac{SCT - PRESS}{SCT} \quad (\text{eq. 7})$$

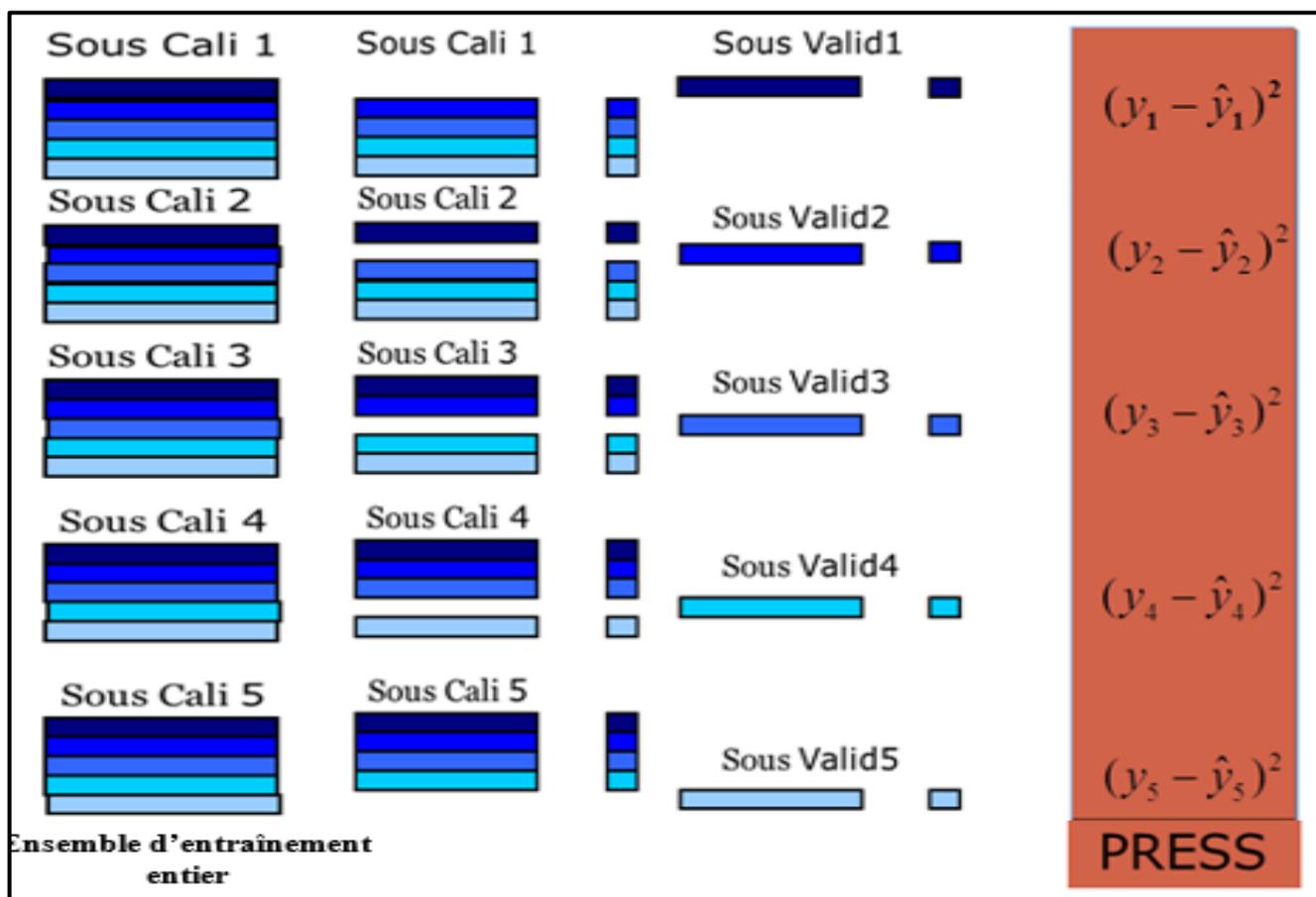


Figure 11 : principe de la validation croisée pour 5 plis

### 1.7.2 Validation externe :

La validation externe est une technique pour juger de manière fiable le pouvoir prédictif du modèle QSAR pour chaque nouvel ensemble de composés issus d'un ensemble n'ayant pas participé à la validation interne, par la division des molécules en deux séries, une série d'entraînement pour la construction du modèle, et une série pour le test [115,116]. La qualité de prédiction est définie par  $Q^2_{\text{ext}}$ , et  $EQMP_{\text{EXT}}$

qui sont définies par l'équation 8 et 9:

$$Q^2_{ext} = 1 - \frac{\sum_{i=1}^{n_{ext}} (y_i - \hat{y}_{(i)})^2 / n_{ext}}{\sum_{i=1}^{n_{tr}} (y_i - \bar{y}_{tr})^2 / n_{tr}} = 1 - \frac{PRESS / n_{ext}}{SCT / n_{tr}} \quad (\text{eq 8})$$

$$EQMP_{ext} (SDEP_{ext}) = \sqrt{\frac{\sum_{i=1}^{n_{ext}} (y_i - \hat{y}_i)^2}{n_{ext}}} \quad (\text{eq. 9})$$

### 1.7.3. Test de randomisation :

Le test de randomisation permet de confirmer que les bonnes corrélations, entre les descripteurs et l'activité, présentées par le modèle QSAR ne sont pas dues au hasard. Pour se faire, les observations sont aléatoirement désorganisées, en changeant itérativement et aléatoirement les valeurs des activités, mais les colonnes des descripteurs restent inchangées. Si la randomisation des observations conduit à des modèles de prévisions élevés cela signifie que les capacités prédictives du modèle QSAR construit sont dues aux corrélations de chance [117,118].

### 1.8. Domaine D'application (DA) :

Le domaine d'application est la région de l'espace chimique définie par les molécules de l'ensemble d'entraînement du modèle [119] lorsqu'une molécule se situe en dehors de cet espace chimique, la prédiction qui est faite sur celle-ci n'est alors plus fiable.

L'espace chimique a été discuté à l'aide du diagramme de Williams [120-121] représentant les résidus de prédiction standardisés en fonction des valeurs des leviers  $h_i$  (la distance entre les valeurs des descripteurs et leur barycentre). L'équation (équation 10) définit le levier d'un composé dans l'espace original des variables indépendantes ( $x_i$ ) :

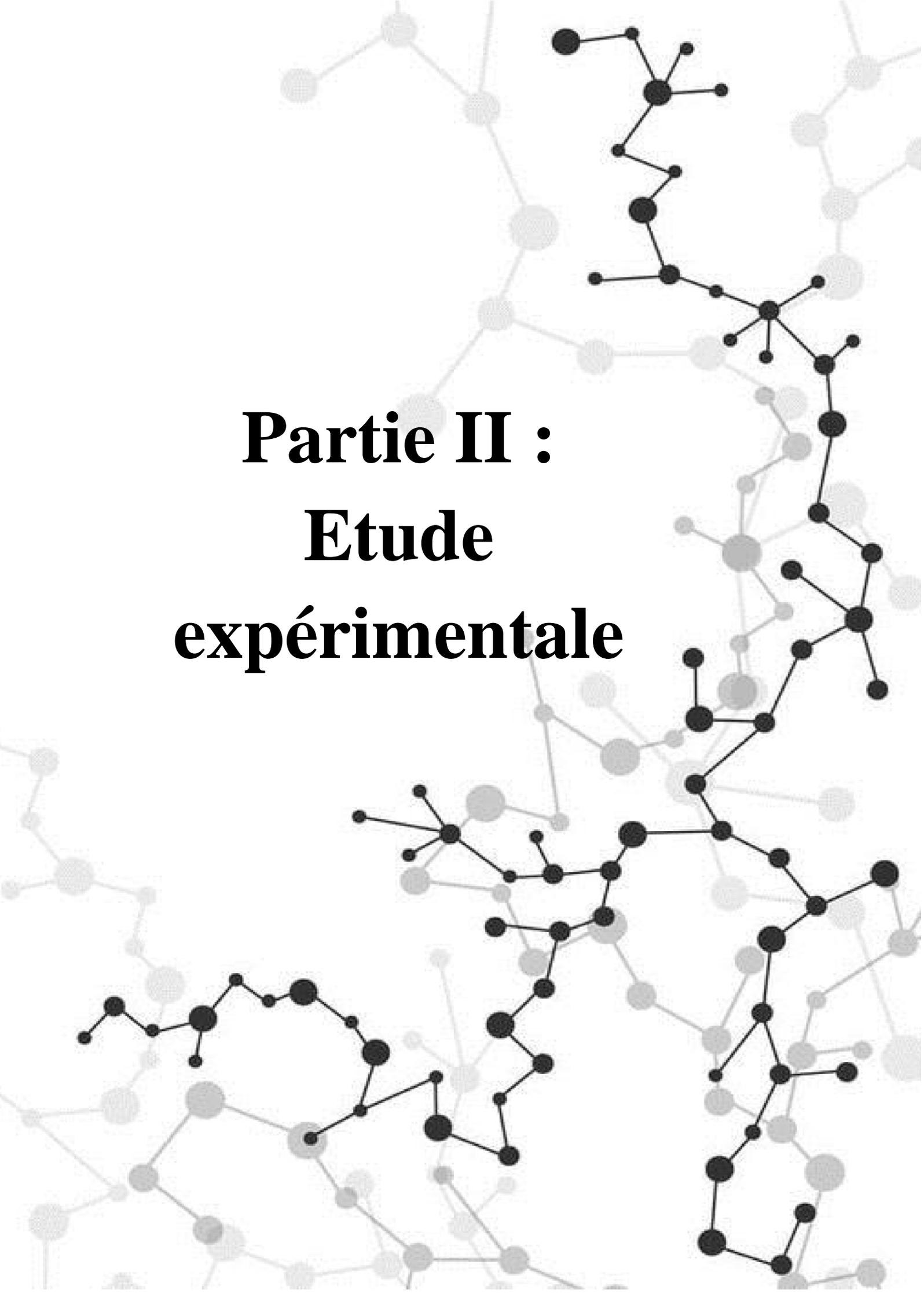
$$h_i = x_i^T (X^T X)^{-1} x_i^T \quad (i = 1, \dots, n) \quad (\text{eq.10})$$

Où  $x_i$  est la vectrice ligne des descripteurs du composé  $i$  et  $X$  ( $n \times p$ ) la matrice du modèle déduite des valeurs des descripteurs de l'ensemble de calibration ; l'indice  $T$  désigne le vecteur (ou la matrice) transposé (e).

La valeur critique du levier ( $h^*$ ) est fixée à  $(3p+1) / n$  ( $p$  est le nombre des descripteurs dans le model). Si  $h_i < h^*$ , seules les prédictions concernant des molécules dans ce domaine peuvent être considérées comme recevables.

Les composés avec  $h_i > h^*$  auront des valeurs prédites douteuses sans pour autant être forcément aberrantes, les résidus pouvant être bas.



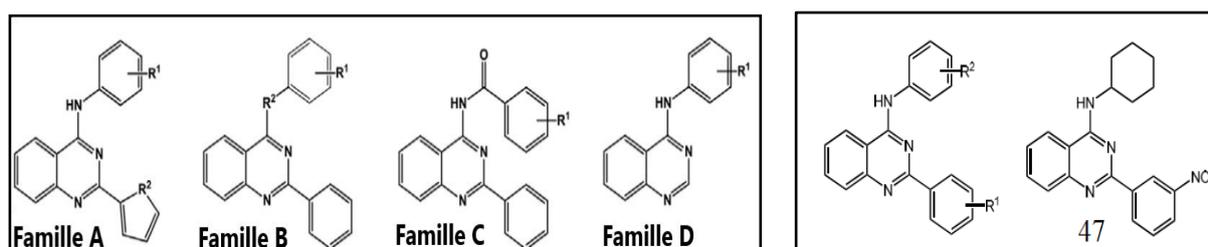
The background of the slide is filled with various molecular structures. Some are rendered in a light gray, semi-transparent style, while others are in solid black. The structures include simple chains, branched molecules, and more complex, multi-ring systems, all set against a white background.

**Partie II :**  
**Etude**  
**expérimentale**

## 1. Méthodologie :

### 1.1. Collecte des données :

Dans notre travail l'homogénéité des données est essentielle, l'activité doit être mesurée par un seul et même protocole, avec les mêmes conditions expérimentales. Un ensemble de données de 87 composés a été choisi à partir de la littérature avec leurs activités. L'activité inhibitrice envers ABCG2 a été déterminée en utilisant le test d'accumulation Hoechst 33342[123,124]. Une fois que le colorant est intégré dans un environnement lipophile comme une membrane cellulaire ou lié à l'ADN, la fluorescence augmente fortement [125]. Étant un substrat d'ABCG2, le colorant peut être utilisé comme indicateur de l'accumulation intracellulaire, qui dépend du degré d'inhibition de corrélé avec le pouvoir inhibiteur d'une substance vis-à-vis de ABCG2. Un ensemble de données de 87 molécules a été prise à partir des articles [129,130], les structures des composés sont présentées Figure 12 et Tableau 1.



**Figure 12 :** Structure de dérivés de quinazoline

**Tableau 1 :** L'activité inhibitrice envers ABCG2 de 87 dérivés de quinazoline

Code	R1	R2	Famille	IC <sub>50</sub> [nM]
<b>1</b>	3-NO <sub>2</sub>	S	A	300
<b>2</b>	4-NO <sub>2</sub>	S	A	335
<b>3</b>	3-CN	S	A	178
<b>4</b>	3,4-Ome	S	A	202

<b>5</b>	3-F	S	A	953
<b>6</b>	3-NHCOCH <sub>3</sub>	S	A	360
<b>7</b>	3-CN	NH	A	156
<b>8</b>	4-Ome	NH	A	2570
<b>9</b>	3,4-Ome	NH	A	652
<b>10</b>	H	O	B	14500
<b>12</b>	3-Br	O	B	3860
<b>13</b>	4-Br	O	B	7270
<b>14</b>	3-CF <sub>3</sub>	O	B	989
<b>15</b>	3-CF <sub>3</sub>	S	B	14800
<b>16</b>	3-Ome	S	B	7400
<b>17</b>	3-NO <sub>2</sub>	N-Me	B	2025
<b>18</b>	4-NO <sub>2</sub>	N-Me	B	3669
<b>20</b>	3-F	N-Me	B	3948
<b>21</b>	H		C	424
<b>23</b>	3-NO <sub>2</sub>		C	54.5
<b>24</b>	4-NO <sub>2</sub>		C	93.1
<b>25</b>	H		D	10700
<b>26</b>	3-NO <sub>2</sub> -4-OH		D	2900
<b>27</b>	4-OH		D	2630
<b>28</b>	4-CN		D	1270
<b>29</b>	3-Ome		D	645
<b>30</b>	3-Sme		D	1350

31	3-F		D	8120
32	3-NHCOCH3		D	28900
33	H	NH	B	882
34	3-NO2	NH	B	130
35	4-NO2	NH	B	69.6
36	3-NO2-4-OH	NH	B	80
37	4-OH	NH	B	204
38	3-CN	NH	B	140
39	4-CN	NH	B	69.9
40	3-Ome	NH	B	1320
41	4-Ome	NH	B	1930
42	3,4-Ome	NH	B	152
43	3-Sme	NH	B	1190
44	3-F	NH	B	355
45	3-NHCOCH3	NH	B	278

Suite Tableau 1

Code	R1	R2	IC <sub>50</sub> [nM]
46	3-NO <sub>2</sub>	H	108 1
47	/	/	2780
48	3-NO <sub>2</sub>	3-NO <sub>2</sub>	236
49	3-NO <sub>2</sub>	3-NO <sub>2</sub> ,4-OH	151
50	3-NO <sub>2</sub>	3-CN	120
51	3-NO <sub>2</sub>	3-OCH <sub>3</sub>	121
52	3-NO <sub>2</sub>	3,4-OCH <sub>3</sub>	55.6
53	3-NO <sub>2</sub>	3-OCH <sub>3</sub> ,4-Br	801
54	3-NO <sub>2</sub>	3,4-OEt	210
55	3-NO <sub>2</sub>	3-[1,4-dioxane]	263
56	3-NO <sub>2</sub>	3-SCH <sub>3</sub>	94.8
57	3-NO <sub>2</sub>	4-SCH <sub>3</sub>	88.6
58	3-NO <sub>2</sub>	3-NH <sub>2</sub>	284
59	3-NO <sub>2</sub>	3-N(CH <sub>3</sub> ) <sub>2</sub>	221
60	3-NO <sub>2</sub>	3-N(Et) <sub>2</sub>	230
61	3-NO <sub>2</sub>	3-NHCOCH <sub>3</sub>	127
63	3-NO <sub>2</sub>	3-CF <sub>3</sub> ,4-OCH <sub>3</sub>	16200
64	3-NO <sub>2</sub>	4-I	1770
65	4-NO <sub>2</sub>	3-NO <sub>2</sub>	166
66	4-NO <sub>2</sub>	3-CN	336
67	4-NO <sub>2</sub>	3-OCH <sub>3</sub>	107
68	4-NO <sub>2</sub>	3,4-OCH <sub>3</sub>	86.7
69	3-CN	3-OH,4-OCH <sub>3</sub>	297
70	3-CN	3-NHCOCH <sub>3</sub>	225
71	2-OCH <sub>3</sub>	3-NO <sub>2</sub>	642
72	3-OCH <sub>3</sub>	3-NO <sub>2</sub>	139
73	3-OCH <sub>3</sub>	4-NO <sub>2</sub>	68
74	3-OCH <sub>3</sub>	3,5-NO <sub>2</sub>	291
75	3-OCH <sub>3</sub>	3-NO <sub>2</sub> ,4-F	1190
76	3-OCH <sub>3</sub>	3-CN	359
77	3-OCH <sub>3</sub>	3-F	1640
78	3-OCH <sub>3</sub>	3-OCH <sub>3</sub>	811
79	3-OCH <sub>3</sub>	3,4-OCH <sub>3</sub>	480
80	4-OCH <sub>3</sub>	3-NO <sub>2</sub>	183
82	4-OCH <sub>3</sub>	3-OCH <sub>3</sub>	2140
83	4-OCH <sub>3</sub>	3,4-OCH <sub>3</sub>	418
84	3,4-OCH <sub>3</sub>	3-NO <sub>2</sub>	82.5
85	3,4-OCH <sub>3</sub>	4-NO <sub>2</sub>	44.2

86	3,4-OCH3	3,5-NO2	610
87	3,4-OCH3	3-NO2,4-OH	305
88	3,4-OCH3	3-CN	136
89	3,4-OCH3	3-CF3,4-OCH3	795
90	3,4-OCH3	3-OCH3	365
92	3,4-OCH3	3-F	462
93	3,4-OCH3	3-SO2F	570

## 1.2. Calcul des descripteurs moléculaires :

### 1.2.1. Préparation de l'ensemble de données :

Les structures de différentes molécules utilisées dans notre travail ont été construites dans le logiciel ChemDraw, utilisé pour dessiner et structurer les composés chimiques.

Tous les calculs d'optimisation géométrique des molécules ont été effectués en utilisant le logiciel de modélisation moléculaire HyperChem, les structures moléculaires ont été pré-optimisées par le champ de force de la mécanique moléculaire MM+, puis la géométrie finale a été obtenue en utilisant la méthode semi empirique PM3. Les conformères les plus stables sont introduites dans le programme DRAGON pour le calcul des descripteurs.

### 1.2.2. Développement et outils de validation du modèle :

Le modèle QSAR a été obtenue par l'analyse de la régression linéaire multiple et la sélection de sous-ensembles de variables par algorithme génétique (AG) moyennant le logiciel Qsarins.

### 1.2.3. Evaluation du modèle :

La validation du modèle est indispensable pour évaluer le modèle. La robustesse et la stabilité du modèle développé a été évaluée à l'aide du coefficient de détermination ( $R^2$ ), Coefficient de détermination ajusté ( $R^2_{adj}$ ) ainsi que le test de Fisher (F), Le test F pour permettre savoir si un modèle est significatif ou non. Le rappel de quelques notions de statistiques de base est préalablement nécessaire

Le modèle a été validé par une validation interne, pour ce faire, on emploie la technique de la validation croisée :  $Q^2_{LOO}$  (Leave- One-Out (LOO) et aussi  $Q_{boot}$  (Bootstrap). La validation interne est insuffisante pour valider la qualité d'un modèle donc une véritable évaluation ne

peut être approuvé que par une validation externe. La validation externe est réalisée à partir de données du coefficient de corrélation externe ( $Q^2_{\text{ext}}$ ).

## 2. Résultats et discussion :

La construction du modèle QSAR a été établie en basant sur cinq descripteurs : MATS2m, MATS7v, MATS5p, RDF010m, EEig12x. (Voir annexe) L'ensemble des données a été réparties en deux séries : une série de 70 composés appelée 'entraînement set' et l'autre de 17 composés appelée 'test set' et l'algorithme génétique a conduit à un bon modèle pour la prédiction des  $IC_{50}$ .

### 2.1. La colinéarité :

La colinéarité crée des difficultés dans l'interprétation des résultats. Les descripteurs ne doivent pas être corrélés entre eux. Si le coefficient de corrélation  $r$  est supérieur à 0.97, les descripteurs seront considérés comme fortement corrélés [122] et en pratique, ils seront alors enlevés dans le procédé de sélection. La matrice de corrélation entre les  $pIC_{50}$  et tous les descripteurs est montrée par le Tableau 2.

**Tableau 2** : Matrice de corrélation entre  $pIC_{50}$  et les descripteurs impliqués dans le modèle QSAR (valeurs de R).

	$pIC_{50}$	MATS2m	MATS7v	MATS5p	RDF010m
MATS2m	-0.123				
MATS7v	0.252	0.479			
MATS5p	0.207	-0.423	-0.286		
RDF010m	0.405	0.538	0.548	-0.284	
EEig12x	0.583	0.283	0.083	-0.209	0.356

Cette matrice montre l'absence de la corrélation entre les descripteurs eux-mêmes. La multi-colinéarité est vérifiée par la règle de quik Q sous l'influence de K, La réponse Y (KXYprédicteurs (KX)) devrait toujours être supérieure à celle mesurée que dans l'ensemble de prédicteurs (KX) c'est-à-dire  $\Delta K (KXY-KX) > 0$  [126-127].

D'après le tableau 3  $\Delta = kxy - kx = 41.46 - 34.35 = 7.11 > 0$  signifie l'absence de multi-colinéarité entre les cinq descripteurs.

## 2.2. Analyse de la régression :

L'activité inhibitrice d'ABCG2 représenté par le cologarithme décimal d' $IC_{50}$  ( $pIC_{50}$ ) de 87 dérivés de quinazoline. Parmi les modèles sélectionnés par l'algorithme génétique nous avons retenu le plus robuste à cinq variables explicatives dont l'équation est :

$$pIC_{50} = -6.13 - 6.38 MATS2m + 2.29 MATS7v + 3.33 MATS5p + 2.81 RDF010m + 0.73 EEIG12X \quad (\text{eq 1})$$

$$n = 87 \quad R^2 = 75.89\%, \quad Q^2 = 70.99\%, \quad F = 40.2982, \quad RMSE = 0.3529 \text{ et } R^2_{adj} = 74.01\%$$

Tableau 3 Statistiques du modèle sélectionné

$R^2$	$Q^2_{LOO}$	$Q^2_{boot}$	$Q^2_{ext}$	$R^2_{adj}$	EQMP	$N_{TR}$		
75.89	70.99	67.91	74.37	74.01	0.37	70		
EQMPext	EQMC	Kx	Kxy	F	RMSE	PRESS	NTEST	
0.348	0.337	34.35	41.46	40.2982	0.3529	9.5901	17	

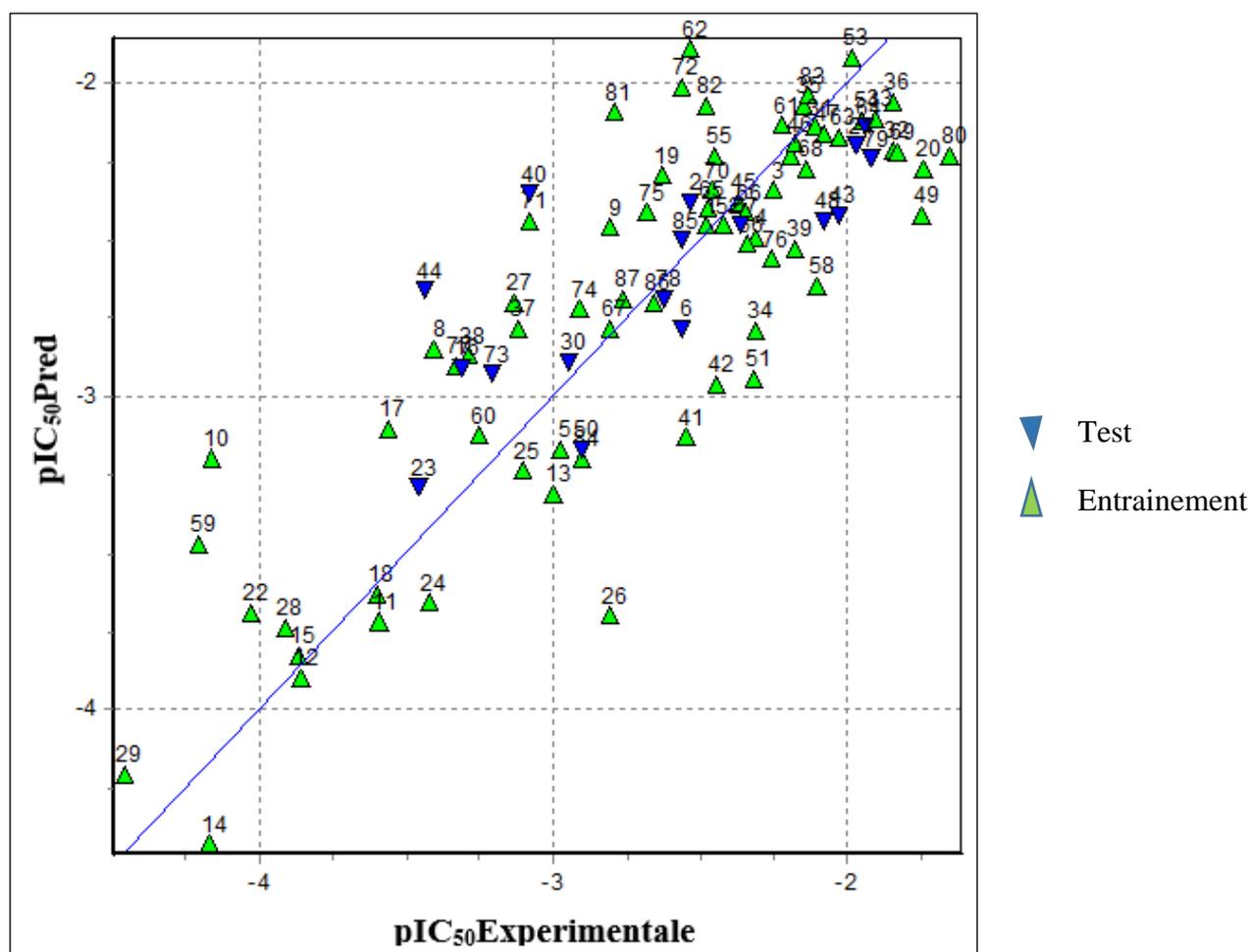
### 2.3. Qualité de l'ajustement :

Les valeurs des paramètres statistiques (Tableau 3) prennent en considération la corrélation entre les cinq descripteurs MATS2m, MATS7v, MATS5p, RDF010m, EEig12x et les  $pIC_{50}$  des 87 dérivés de quinazoline.

La valeur élevée de  $F = 40.30$  et les valeurs de ( $R^2 = 75.89\%$ ) et de ( $R^2_{adj} = 74.01$ ) montrent la bonne qualité de l'ajustement, et la valeur du coefficient de détermination  $R^2$  signifie que 75.89% de la variabilité de  $pIC_{50}$  est expliquée par ces descripteurs.

Les données de l'activité expérimentale est leur prédiction donnée par l'équation (eq.1) pour le entraînement et le test sont présentés dans le Tableau 4 et la Figure 13.

La Figure 13, reproduit les valeurs prédites de l'activité en fonction de celles mesurées. La dispersion des points autour de droite montre que les valeurs prédites sont en adéquation avec les valeurs expérimentales, caractéristique d'un bon ajustement.



**Figure 13 :** La droite de régression des valeurs expérimentales et prédits de la notation  $pIC_{50}$  pour le modèle de QSAR.

**Tableau 4** Les valeurs expérimentales prédit et calculé pour  $pIC_{50}$  pour les 87 dérivés de quinazoline.

N°	Code	Ensemble	$pIC_{50}$ Exp	$pIC_{50}$ Calc	$pIC_{50}$ Pred	Err.Calc.	Err.Pred.
1	1	Entrainement	-2.48	-2.45	-2.45	0.02	0.03
2	2	Test	-2.53	-	-2.38	-	0.15
3	3	Entrainement	-2.25	-2.34	-2.34	-0.09	-0.09
4	4	Entrainement	-2.31	-2.47	-2.49	-0.17	-0.18

<b>5</b>	5	Entrainement	-2.98	-3.16	-3.17	-0.18	-0.19
<b>6</b>	6	Test	-2.56	-	-2.78	-	-0.23
<b>7</b>	7	Entrainement	-2.19	-2.22	-2.23	-0.03	-0.03
<b>8</b>	8	Entrainement	-3.41	-2.91	-2.85	0.5	0.56
<b>9</b>	9	Entrainement	-2.81	-2.49	-2.46	0.33	0.35
<b>10</b>	10	Entrainement	-4.16	-3.27	-3.2	0.89	0.96
<b>11</b>	12	Entrainement	-3.59	-3.71	-3.72	-0.13	-0.14
<b>12</b>	13	Entrainement	-3.86	-3.9	-3.9	-0.04	-0.04
<b>13</b>	14	Entrainement	-3	-3.23	-3.31	-0.24	-0.31
<b>14</b>	15	Entrainement	-4.17	-4.35	-4.43	-0.18	-0.26
<b>15</b>	16	Entrainement	-3.87	-3.84	-3.83	0.03	0.03
<b>16</b>	17	Test	-3.31	-	-2.91	-	0.4
<b>17</b>	18	Entrainement	-3.56	-3.13	-3.1	0.43	0.47
<b>18</b>	20	Entrainement	-3.6	-3.63	-3.63	-0.03	-0.03
<b>19</b>	21	Entrainement	-2.63	-2.3	-2.29	0.32	0.34
<b>20</b>	23	Entrainement	-1.74	-2.24	-2.27	-0.5	-0.53
<b>21</b>	24	Test	-1.97	-	-2.2	-	-0.23
<b>22</b>	25	Entrainement	-4.03	-3.74	-3.69	0.29	0.34
<b>23</b>	26	Test	-3.46	-	-3.29	-	0.17
<b>24</b>	27	Entrainement	-3.42	-3.63	-3.66	-0.21	-0.24

<b>25</b>	28	Entrainement	-3.1	-3.23	-3.24	-0.13	-0.14
<b>26</b>	29	Entrainement	-2.81	-3.6	-3.7	-0.79 *	-0.89 *
<b>27</b>	30	Entrainement	-3.13	-2.87	-2.7	0.26	0.43
<b>28</b>	31	Entrainement	-3.91	-3.76	-3.74	0.15	0.17
<b>29</b>	32	Entrainement	-4.46	-4.26	-4.21	0.21	0.25
<b>30</b>	33	Test	-2.95	-	-2.89	-	0.06
<b>31</b>	34	Entrainement	-2.11	-2.14	-2.14	-0.02	-0.02
<b>32</b>	35	Entrainement	-1.84	-2.19	-2.21	-0.34	-0.37
<b>33</b>	36	Entrainement	-1.9	-2.1	-2.11	-0.2	-0.21
<b>34</b>	37	Entrainement	-2.31	-2.75	-2.79	-0.44	-0.48
<b>35</b>	38	Entrainement	-2.15	-2.08	-2.07	0.07	0.07
<b>36</b>	39	Entrainement	-1.84	-2.04	-2.06	-0.2	-0.21
<b>37</b>	40	Entrainement	-3.12	-2.8	-2.78	0.33	0.34
<b>38</b>	41	Entrainement	-3.29	-2.91	-2.87	0.38	0.41
<b>39</b>	42	Entrainement	-2.18	-2.52	-2.53	-0.33	-0.35
<b>40</b>	43	Test	-3.08	-	-2.35	-	0.73
<b>41</b>	44	Entrainement	-2.55	-3.1	-3.13	-0.55	-0.58
<b>42</b>	45	Entrainement	-2.44	-2.91	-2.96	-0.47	-0.51
<b>43</b>	46	Test	-2.03	-	-2.42	-	-0.39
<b>44</b>	47	Test	-3.44	-	-2.66	-	0.79

<b>45</b>	48	Entrainement	-2.37	-2.38	-2.38	-0.01	-0.01
<b>46</b>	49	Entrainement	-2.18	-2.18	-2.19	-0.01	-0.01
<b>47</b>	50	Entrainement	-2.08	-2.16	-2.16	-0.08	-0.08
<b>48</b>	51	Test	-2.08	-	-2.44	-	-0.36
<b>49</b>	52	Entrainement	-1.75	-2.38	-2.42	-0.63	-0.67
<b>50</b>	53	Test	-2.9	-	-3.17	-	-0.27
<b>51</b>	54	Entrainement	-2.32	-2.87	-2.94	-0.55	-0.62
<b>52</b>	55	Entrainement	-2.42	-2.45	-2.45	-0.03	-0.03
<b>53</b>	56	Entrainement	-1.98	-1.93	-1.92	0.04	0.05
<b>54</b>	57	Entrainement	-1.95	-2.1	-2.12	-0.15	-0.17
<b>55</b>	58	Entrainement	-2.45	-2.24	-2.23	0.21	0.23
<b>56</b>	59	Entrainement	-2.34	-2.5	-2.51	-0.16	-0.17
<b>57</b>	60	Test	-2.36	-	-2.45	-	-0.09
<b>58</b>	61	Entrainement	-2.1	-2.61	-2.65	-0.51	-0.55
<b>59</b>	63	Entrainement	-4.21	-3.58	-3.47	0.63	0.73
<b>60</b>	64	Entrainement	-3.25	-3.13	-3.12	0.12	0.13
<b>61</b>	65	Entrainement	-2.22	-2.14	-2.13	0.08	0.09
<b>62</b>	66	Entrainement	-2.53	-1.93	-1.89	0.6	0.64
<b>63</b>	67	Entrainement	-2.03	-2.17	-2.17	-0.14	-0.14
<b>64</b>	68	Test	-1.94	-	-2.14	-	-0.2

<b>65</b>	69	Entrainement	-2.47	-2.4	-2.4	0.07	0.08
<b>66</b>	70	Entrainement	-2.35	-2.41	-2.41	-0.05	-0.06
<b>67</b>	71	Entrainement	-2.81	-2.78	-2.78	0.02	0.03
<b>68</b>	72	Entrainement	-2.14	-2.27	-2.27	-0.12	-0.13
<b>69</b>	73	Entrainement	-1.83	-2.21	-2.22	-0.37	-0.39
<b>70</b>	74	Entrainement	-2.46	-2.35	-2.34	0.12	0.12
<b>71</b>	75	Entrainement	-3.08	-2.47	-2.44	0.61	0.63
<b>72</b>	76	Entrainement	-2.56	-2.03	-2.01	0.53	0.55
<b>73</b>	77	Test	-3.21	-	-2.93	-	0.28
<b>74</b>	78	Entrainement	-2.91	-2.74	-2.72	0.17	0.19
<b>75</b>	79	Entrainement	-2.68	-2.44	-2.41	0.24	0.27
<b>76</b>	80	Entrainement	-2.26	-2.55	-2.56	-0.28	-0.29
<b>77</b>	82	Entrainement	-3.33	-2.93	-2.9	0.4	0.43
<b>78</b>	83	Test	-2.62	-	-2.69	-	-0.06
<b>79</b>	84	Test	-1.92	-	-2.24	-	-0.32
<b>80</b>	85	Entrainement	-1.65	-2.2	-2.23	-0.55	-0.59
<b>81</b>	86	Entrainement	-2.79	-2.17	-2.09	0.62	0.69
<b>82</b>	87	Entrainement	-2.48	-2.1	-2.07	0.38	0.42
<b>83</b>	88	Entrainement	-2.13	-2.04	-2.04	0.09	0.1
<b>84</b>	89	Entrainement	-2.9	-3.16	-3.2	-0.26	-0.3

85	90	Test	-2.56	-	-2.5	-	0.07
86	92	Entrainement	-2.66	-2.69	-2.7	-0.03	-0.03
87	93	Entrainement	-2.76	-2.7	-2.69	0.06	0.06

## 2.4. Validation :

### 2.4.1 Validation interne :

Dans cette étude l'ensemble des données a été divisé en deux sous-ensemble : sous-ensemble d'entraînement (80% =70 molécules) pour la construction du model et la validation interne, et un sous-ensemble de Test (20% 17 molécules) pour la validation externe. Le coefficient de corrélation de la validation croisée du modèle  $Q^2_{LOO}=70.99\%$ , sa valeur témoigne de la bonne corrélation entre l'activité prédite et l'activité réelle, et reflète la précision du modèle.

La similarité entre EQMP et EQMC (voir Tableau 3) signifie que la capacité de prédiction interne du modèle n'est pas trop dissemblable de son pouvoir d'ajustement.

### 2.4.2 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 n'ont pas été utilisé lors de l'élaboration du modèle.

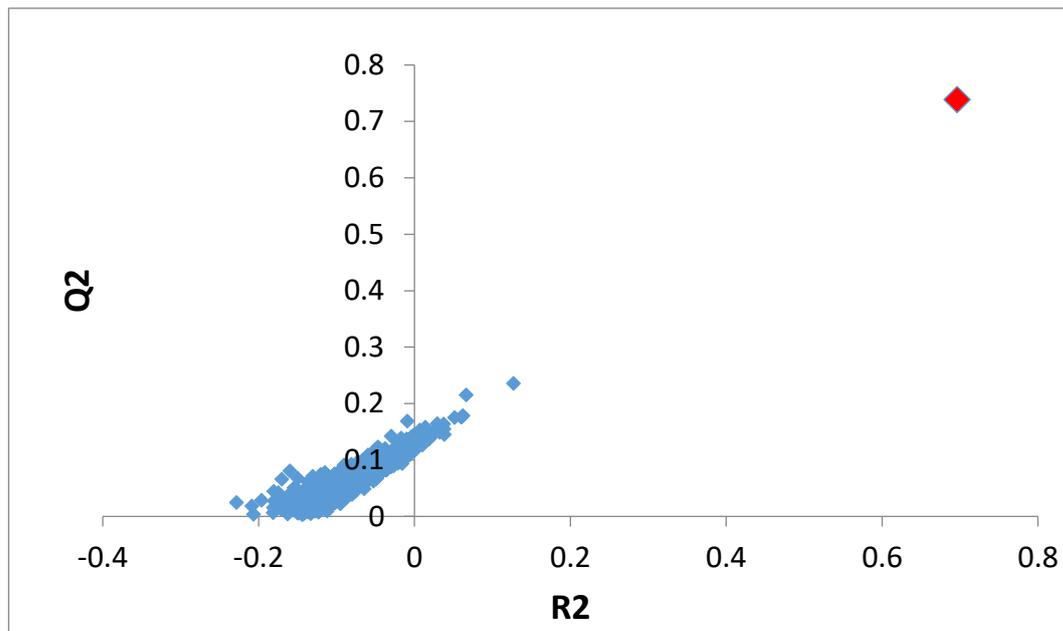
Les paramètres de la validation externe  $Q^2_{ext}=74.37\%$  et  $EQMP_{ext}=0.348$ , confirment la bonne capacité prédictive du modèle pour les composés n'ayant pas participé aux calculs. Il est à noter que le modèle d'équation 1, est à accréditer des meilleures performances.

### 2.4.3 Test de randomisation :

Le test de randomisation (Y- randomisé) permet de vérifier l'existence de corrélations imprévues [128]. Ce test consiste à construire un vecteur de la propriété dont les composantes sont celles du vecteur de la propriété réelle, mais permutées au hasard dans leurs positions. Ce nouveau vecteur d'activité est utilisé comme valeur expérimentale. Ce processus a effectué

1000 itérations, afin de tester le facteur de capacité du modèle à extraire relations structure /activité effectives.

Les losanges en bleu représentent les activités mises au hasard, et le losange rouge correspond aux activités réelles. Le test de randomisation après 1000 itérations, montre que notre modèle n'est pas dû au hasard ceci est justifié par la Figure 14.



**Figure 14** : Test de randomisation associé au modèle QSAR 1000 itérations.

## 2.5. Domaine chimique :

Le diagramme de Williams (voir figure 15) désigne les composés qui se trouvent dans le domaine d'application de l'équation 1. Les valeurs des résidus standardisés sont toutes comprises dans l'intervalle  $\pm 3$  donc le modèle ne présente aucun point aberrant.

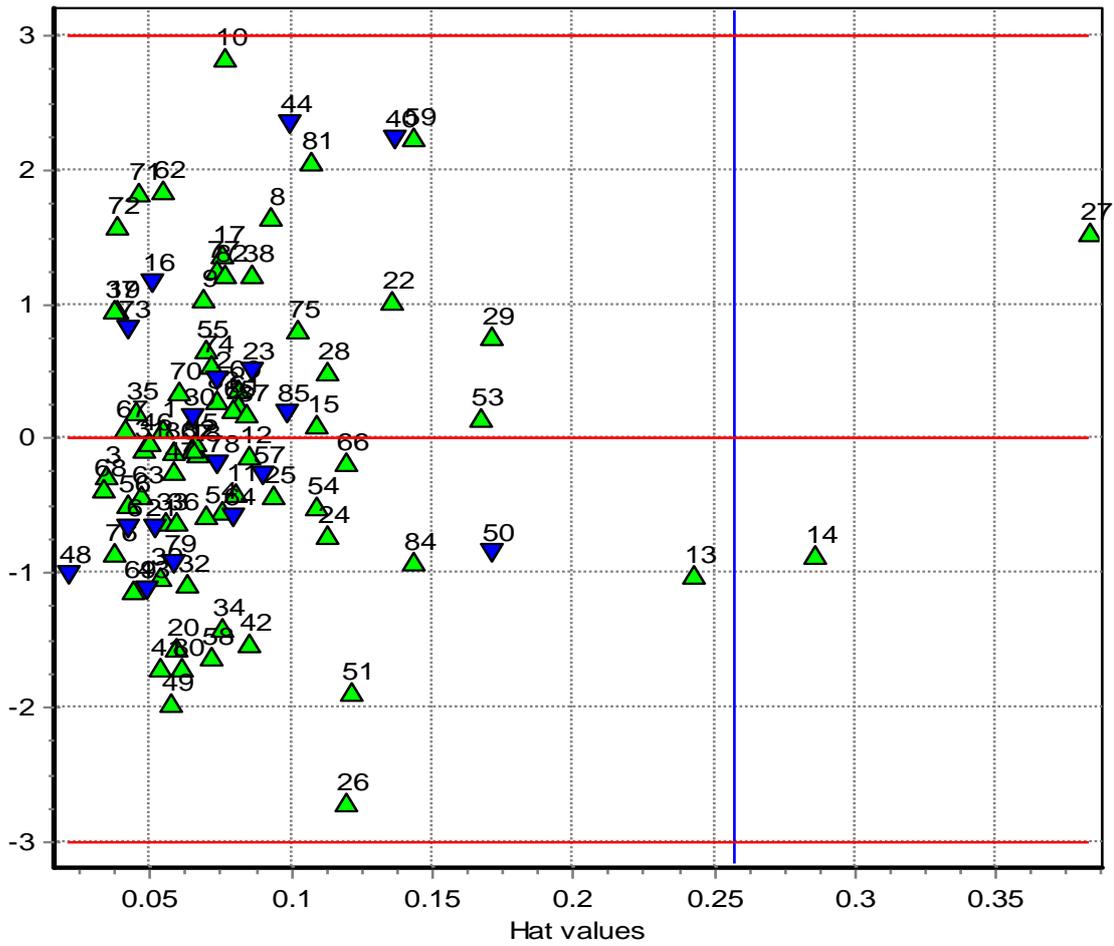
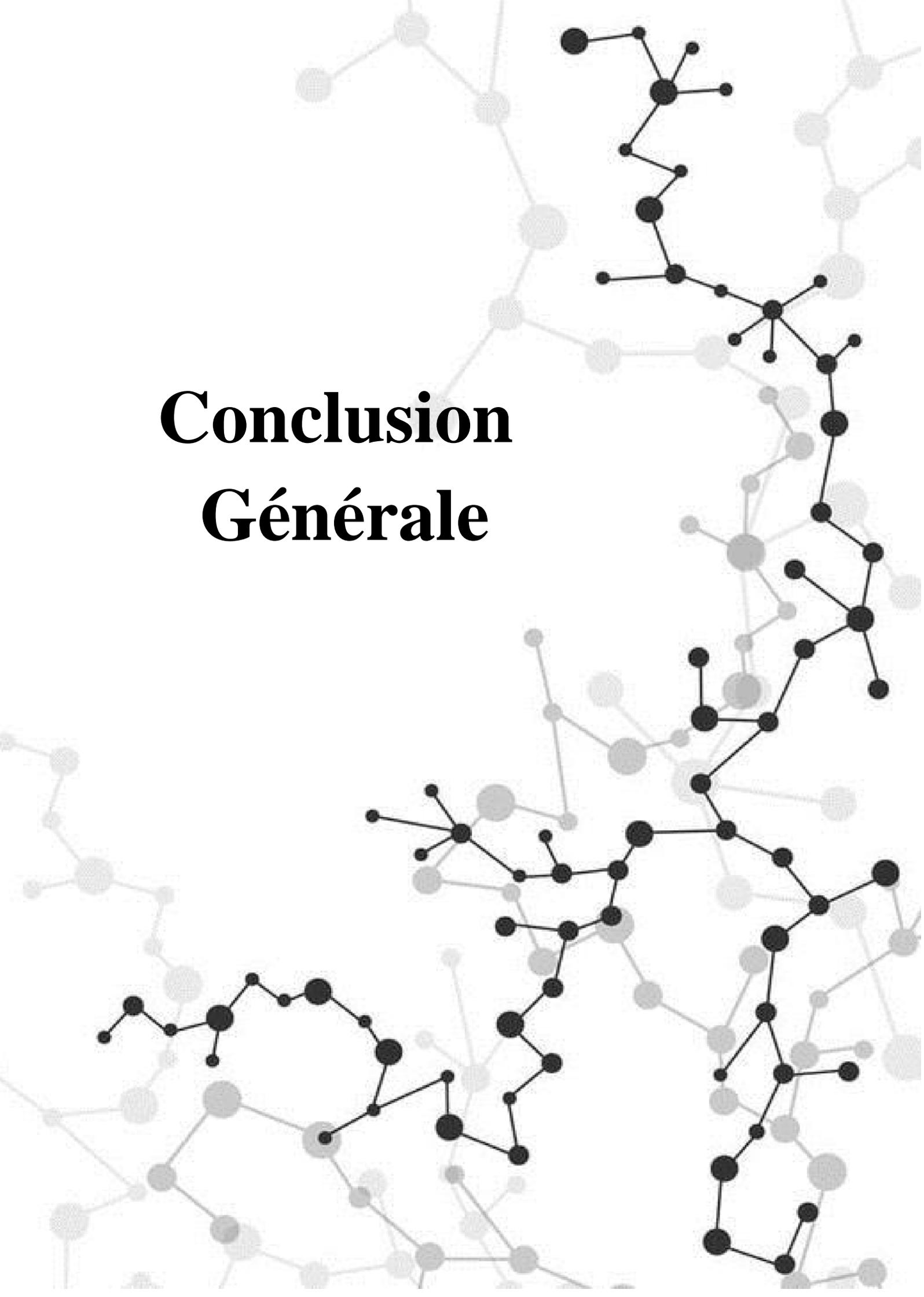


Figure 15 : Diagramme de williams des résidus standardisés

The background features a complex network of molecular structures. It consists of numerous interconnected nodes (represented by circles of varying sizes and shades of gray) and edges (represented by thin lines). Some structures are rendered in a light gray, semi-transparent style, while others are in a solid black color, creating a layered, three-dimensional effect. The overall composition is abstract and scientific, typical of a chemistry or materials science presentation.

# **Conclusion Générale**



## Conclusion générale

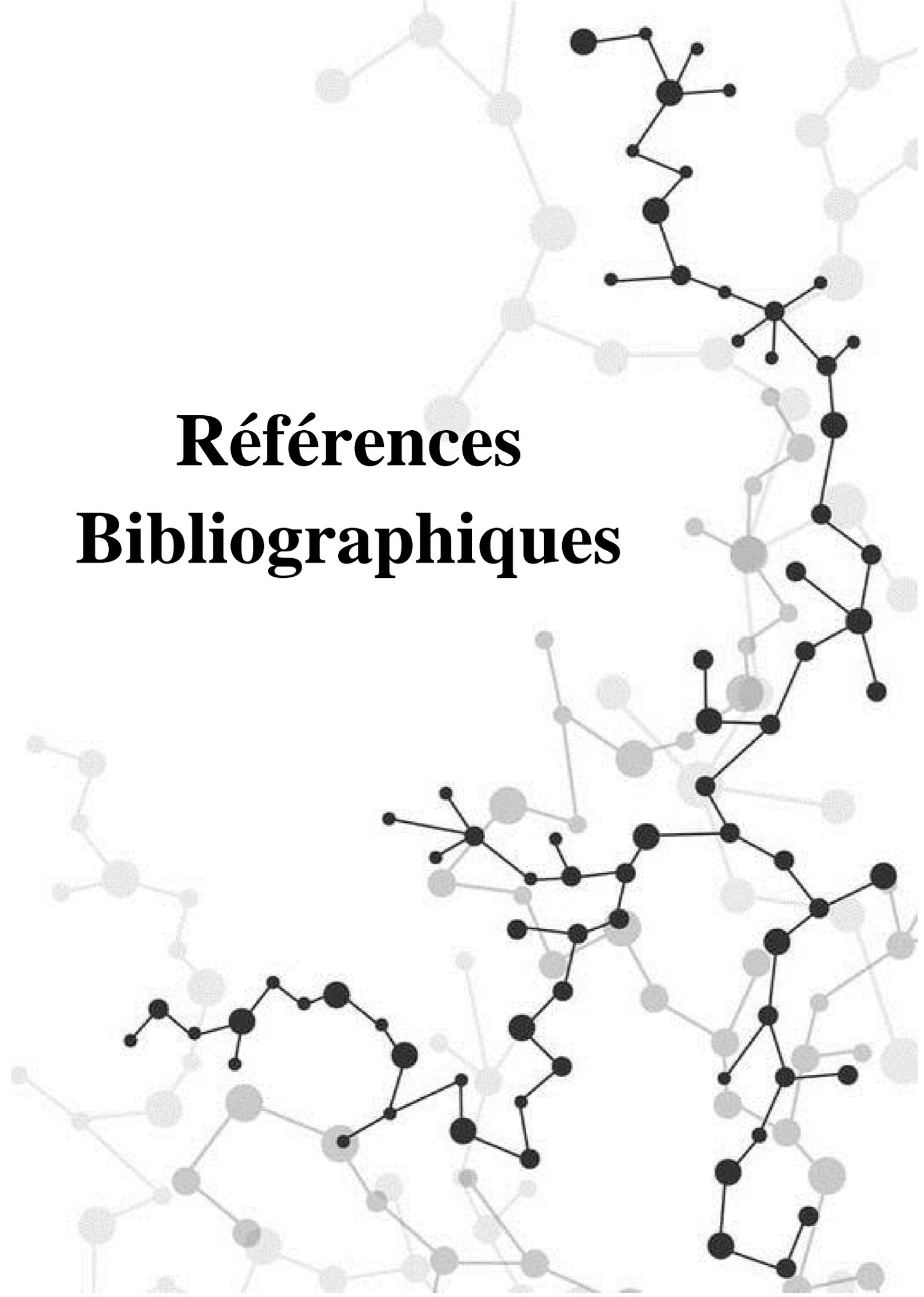
La recherche d'une relation entre la structure moléculaire et l'activité biologique est l'un des enjeux majeurs pour la conception des médicaments. Afin de développer de nouveaux composés anticancéreux. Un modèle QSAR a été construit pour la prédiction de l'activité inhibitrice de la protéine ABCG2.

Pour se faire, le logiciel QSAINS a été utilisé et une analyse de régression linéaire multiple (MLR = Multiple Linear Régression) a été mise en jeu sur 05 descripteurs sélectionnés et sur la base des valeurs d'IC<sub>50</sub> d'une série de 87 dérivés de quinazoline pour former l'équation de QSAR finale. Les principales techniques de validation (validation interne et externe) ont été utilisées pour évaluer le modèle QSAR obtenu.

Dans cette étude nous avons abouti à une corrélation entre pIC<sub>50</sub> qui détermine l'activité inhibitrice de la protéine ABCG2 et cinq descripteurs ; MATS2m, MATS7v, MATS5p, RDF010m, EEig12x. .

Le modèle QSAR trouvé est très robuste, possède des excellents paramètres prédictifs internes et externes. Ce modèle est capable de prédire l'activité étudiée de nouveaux composés.

Ce qui nous permet d'entrevoir des perspectives assez prometteuses dans ce domaine par l'utilisation d'autres méthodes de sélection afin de poursuivre notre chemin de recherche.



# **Références Bibliographiques**

## Références

- 1 Dr. Marc Espié.(janvier2010).’’ Cancer du sein’’, Guide Affection longue durée
- 2 Snodin D.J. (2002) *Toxicology Letters*, 127, 161–168
- 3 Oprea.T. I. (2005) "Chemoinformatics in Drug Discovery" Ed. Wiley-vch Verlag. Allemagne.
- 4 Rekka E. A; Kourounakis P. N..(2008) “Chemistry and Molecular Aspects of Drug Design and Action” Ed. Taylor & Francis Group, LLC. Etats Unies
- 5 Ekins,S. (2006) "Computer Applications in Pharmaceutical Research and Development" Edt. John Wiley & Sons, Inc. Etats Unies,
- 6 Yap, C.W., Xue, Y., Li, Z.R., and Chen, Y.Z. (2006) *Current Topics in Medicinal Chemistry*, 6, 1593–1607.
- 7 Guido, R.V., Oliva, G., and Andricopulo, A.D. (2008) *Current Medicinal Chemistry*, 15, 37–46.
- 8 Go'tte M, Rausch JW, Marchand B, Sarafianos S, Le Grice SF (2010) *Biochim Biophys Acta* 1804:1202
- 9 Hu WS, Hughes SH (2012) *Cold Spring Harbor Perspect Med* 2:a006882
- 10 McDevitt CA, Collins RF, Conway M, Modok S, Storm J, Kerr ID, Ford RC, Callaghan R. (2006) “Purification and 3D structural analysis of oligomeric human multidrug transporter ABCG2.”, *Structure*, 14, 1623-32.
- 11 Dean, M., Rzhetsky, A., and Allikmets, R. (2001) ‘’The human ATP-binding cassette (ABC) transporter superfamily’’. *Genome Res* 11(7), 1156-66.
- 12 Jonker JW, Merino G, Musters S, van Herwaarden AE, Bolscher E, Wagenaar E, Mesman E, Dale TC, Schinkel AH. (2005) “The breast cancer resistance protein BCRP (ABCG2) concentrates drugs and carcinogenic xenotoxins into milk.”, *Nat Med.*, 11, 127-9
- 13 Merino G, Jonker JW, Wagenaar E, van Herwaarden AE, Schinkel AH. (2005) “The breast cancer resistance protein (BCRP/ABCG2) affects pharmacokinetics, hepatobiliary excretion, and milk secretion of the antibiotic nitrofurantoin.”, *Mol Pharmacol.*, 67, 1758-64.

- 14 Merino G, Alvarez AI, Pulido MM, Molina AJ, Schinkel AH, Prieto JG. (2006) "Breast cancer resistance protein (BCRP/ABCG2) transports fluoroquinolone antibiotics and affects their oral availability, pharmacokinetics, and milk secretion.", *Drug Metab Dispos.*, 34, 690-5
- 15 Nuket E, Evrim Y, Baha Z et al. (2015) "Classification moléculaire du cancer du sein: du traditionnel, à l'ancienne à un nouvel âge, et une nouvelle façon", *The journal of Breast Health.* 11(2),59-66.
- 16 <http://www.e-cancer.fr/Patients-et-proches/Les-cancers/Cancer-du-sein/Facteurs-de-risque/Predispositions-genetiques>, (2018) "Le-cancer ; the latest oncologie news", Swiss
- 17 Key TJ, Verkasalo PK,(2001) " Banks E. Epidemiology of breast cancer", *Lancet Oncol*(2),133-40.
- 18 Linton KJ, Higgins CF. (1998) "The Escherichia coli ATP-binding cassette (ABC) proteins.", *Mol Microbiol.*, 5-13
- 19 Dean M, Hamon Y, Chimini G. (2001) "The human ATP-binding cassette (ABC) transporter superfamily.", *J Lipid Res.*, 1007-17.
- 20 Hollenstein K, Dawson RJ, Locher KP.(2007) " Structure and mechanism of ABC transporter proteins", *Curr Opin Struct Biol.* Aug;17(4),412-8.
- 21 Jones P .M, O'Mara M.L and George A.M.(2009) "abc transporteurs : a riddle wrapped in a mystery inside an enigma". *Trends Biochem Sci.* Oct;34(10) ,520-31
- 22 Kathawala RJ, Gupta P, Ashby CR Jr, Chen ZS.(2015) " The modulation of ABC transporter-mediated multidrug resistance in cancer : a review of the past decade". *Drug Resist Updat.* Jan ;29 ,1-17
- 23 Leslie EM, Deeley RG, Cole SP. (2005) "Multidrug resistance proteins: role of P-glycoprotein, MRP1, MRP2, and BCRP (ABCG2) in tissue defense. *Toxicol Appl Pharmacol*". May 1;204(3),216-37.
- 24 Sarkadi, B., Homolya, L., Szakacs, G. & Varadi, A.(2006) "Human multidrug resistance ABCB and ABCG transporters: participation in a chemoiimmunity defense system". *Physiol Rev.* Oct ;86(4) ,1179-236
- 25 Seeger MA et van Veen HW .(2009) " Molecular basis of multidrug transport by ABC transporters. *Biochim Biophys Acta*". May ; 1794(5) ,725-37

- 26 Kos V, Ford RC. (2009) “The ATP-binding cassette family: a structural perspective.”, *Cell Mol Life Sci.*, 3111-26.
- 27 Hung LW, Wang IX, Nikaido K, Liu PQ, Ames GF, Kim SH. (1998) “Crystal structure of the ATP-binding subunit of an ABC transporter.”, *Nature*, 396, 703-7.
- 28 Dassa E, Bouige P. (2001) “The ABC of ABCS: a phylogenetic and functional classification of ABC systems in living organisms.”, *Res Microbiol.*, 152, 211-29.
- 29 Higgins CF, Linton KJ. (2004) “The ATP switch model for ABC transporters.”, *Nat Struct Mol Biol.*, 918-26.
- 30 Quinton, P.M.,(1999)’’ Physiological basis of cystic fibrosis: a historical perspective’’. *Physiol Rev*, 79(1 Suppl), 3-22.
- 31 Glavinias, H. krajcsi P cserepes,J and sarkadi, R. (2004)’’ The role of ABC transporters in drug resistance, metabolism and toxicity’’. *Curr Drug Deliv*, 1(1),27-42.
- 32 Fletcher JI, Haber M, Henderson MJ, Norris MD.(2010)’’ ABC transporters in cancer: more than just drug efflux pumps’’. *Nat Rev Cancer.*;10(2),147-56.
- 33 Doyle, L.A.,. Yang,W. Abruzzo,L,V. Krogmann,T Gao,Y Rishi,A,K and Ross D.D.(1998)’’ A multidrug resistance transporter from human MCF-7 breast cancer cells’’. *Proc Natl Acad Sci U S A*, 95(26), 15665-70
- 34 Allikmets, R., Schriml.L,M Hutchinson,A Romano-Spica V, and Dean M(1998)’’ A human placenta-specific ATP-binding cassette gene (ABCP) on chromosome 4q22 that is involved in multidrug resistance’’. *Cancer Res*, 58(23), 5337-9.
- 35 Miyake, K.,. Mickley., L. Litman , T, Zhan Z., Robey R., Cristensen B., Brangi M,Greenberger , L., Dean M Fojo, T. and Bates S.E(1999)’’ Molecular cloning of cDNAs which are highly overexpressed in mitoxantrone-resistant cells: demonstration of homology to ABC transport genes’’. *Cancer Res*, 59(1),8-13.
- 36 Doyle LA, Ross DD.(2003) “Multidrug resistance mediated by the breast cancer resistance protein BCRP (ABCG2).”, *Oncogene*, 7340-58.
- 37 Faneyte IF, Kristel PM, Maliepaard M, Scheffer GL, Scheper RJ, Schellens JH, van de Vijver MJ. (2002) “Expression of the breast cancer resistance protein in breast cancer.”, *Clin Cancer Res.*, 1068-74
- 38 Zhou S, Schuetz JD, Bunting KD, Colapietro AM, Sampath J, Morris JJ, Lagutina I, Grosveld GC, Osawa M, Nakauchi H, Sorrentino BP. (2001) “The

- ABC transporter Bcrp1/ABCG2 is expressed in a wide variety of stem cells and is a molecular determinant of the side-population phenotype.”, *Nat Med.*, 7, 1028-34.
- 39 Vlaming ML, Lagas JS, Schinkel AH. (2009) “Physiological and pharmacological roles of ABCG2 (BCRP): recent findings in *Abcg2* knockout mice.”, *Adv Drug Deliv Rev.*, 61, 14-25
- 40 Scheffer GL, Maliepaard M, Pijnenborg AC, van Gastelen MA, de Jong MC, Schroeijers AB, van der Kolk DM, Allen JD, Ross DD, van der Valk P, Dalton WS, Schellens JH, Scheper RJ. (2000) “Breast cancer resistance protein is localized at the plasma membrane in mitoxantrone- and topotecan-resistant cell lines.”, *Cancer Res.*, 60, 2589-93.
- 41 Rocchi E, Khodjakov A, Volk EL, Yang CH, Litman T, Bates SE, Schneider E. (2000) “The product of the ABC half-transporter gene ABCG2 (BCRP/MXR/ABCP) is expressed in the plasma membrane.”, *Biochem Biophys Res Commun.*, 271, 42-6.
- 42 Maliepaard M, Scheffer GL, Faneyte IF, van Gastelen MA, Pijnenborg AC, Schinkel AH, van De Vijver MJ, Scheper RJ, Schellens JH. (2001) “Subcellular localization and distribution of the breast cancer resistance protein transporter in normal human tissues.”, *Cancer Res.*, 3458-64.
- 43 Rajagopal A, Simon SM. (2003) “Subcellular localization and activity of multidrug resistance proteins.”, *Mol Biol Cell.*, 14, 3389-99
- 44 Solazzo M, Fantappiè O, D'Amico M, Sassoli C, Tani A, Cipriani G, Bogani C, Formigli L, Mazzanti R. (2009) “Mitochondrial expression and functional activity of breast cancer resistance protein in different multiple drug-resistant cell lines.”, *Cancer Res.*, 69, 7235-42.
- 45 Jonker JW, Buitelaar M, Wagenaar E, Van Der Valk MA, Scheffer GL, Scheper RJ, Plosch T, Kuipers F, Elferink RP, Rosing H, Beijnen JH, Schinkel AH (2002) “The breast cancer resistance protein protects against a major chlorophyll-derived dietary phototoxin and protoporphyria.”, *Proc Natl Acad Sci U S A.*, 99, 15649-54.
- 46 Alfaras I, Pérez M, Juan ME, Merino G, Prieto JG, Planas JM, Alvarez AI. (2010) “Involvement of breast cancer resistance protein (BCRP1/ABCG2) in the bioavailability and tissue distribution of trans-resveratrol in knockout mice.”, *J Agric Food Chem.*, 58, 4523-8

- 47 Jonker JW, Smit JW, Brinkhuis RF, Maliepaard M, Beijnen JH, Schellens JH, Schinkel AH. (2000) "Role of breast cancer resistance protein in the bioavailability and fetal penetration of topotecan.", *J Natl Cancer Inst.*, 92, 1651-6.
- 48 Zhang Y, Wang H, Unadkat JD, Mao Q. (2007) "Breast cancer resistance protein 1 limits fetal distribution of nitrofurantoin in the pregnant mouse.", *Drug Metab Dispos.*, 35, 2154-8.
- 49 Cygalova L, Ceckova M, Pavek P, Staud F. (2008) "Role of breast cancer resistance protein (Bcrp/Abcg2) in fetal protection during gestation in rat.", *Toxicol Lett.*, 178, 176-80.
- 50 Horsey AJ, Cox MH, Sarwat S, Kerr ID.(2016)" The multidrug transporter ABCG2: still more questions than answers". *Biochem Soc Trans*,44(3),824-30.
- 51 Van Herwaarden AE, Wagenaar E, Merino G, Jonker JW, Rosing H, Beijnen JH, Schinkel AH. (2007) "Multidrug transporter ABCG2/breast cancer resistance protein secretes riboflavin (vitamin B2) into milk.", *Mol Cell Biol.*, 27, 1247-53
- 52 Merino G, Jonker JW, Wagenaar E, van Herwaarden AE, Schinkel AH. (2005) "The breast cancer resistance protein (BCRP/ABCG2) affects pharmacokinetics, hepatobiliary excretion, and milk secretion of the antibiotic nitrofurantoin.", *Mol Pharmacol.*, 67, 1758-64.
- 53 Miyake K, Mickley L, Litman T, Zhan Z, Robey R, Cristensen B, Brangi M, Greenberger L, Dean M, Fojo T, Bates SE. (1999) "Molecular cloning of cDNAs which are highly overexpressed in mitoxantrone-resistant cells: demonstration of homology to ABC transport genes.", *Cancer Res.*, 59, 8-13.
- 54 Robey RW, Medina-Pérez WY, Nishiyama K, Lahusen T, Miyake K, Litman T, Senderowicz AM, Ross DD, Bates SE. (2001) "Overexpression of the ATP-binding cassette halftransporter, ABCG2 (Mxr/BCrp/ABCP1), in flavopiridol-resistant human breast cancer cells.", *Clin Cancer Res.*, 7, 145-52.
- 55 Wang X, Furukawa T, Nitanda T, Okamoto M, Sugimoto Y, Akiyama S, Baba M. (2003) "Breast cancer resistance protein (BCRP/ABCG2) induces cellular resistance to HIV-1 nucleoside reverse transcriptase inhibitors.", *Mol Pharmacol.*, 63, 65-72.
- 56 Ahmed-Belkacem A, Pozza A, Muñoz-Martínez F, Bates SE, Castanys S, Gamarro F, Di Pietro A, Pérez-Victoria JM. (2005) "Flavonoid structure-

- activity studies identify 6prenylchrysin and tectochrysin as potent and specific inhibitors of breast cancer resistance protein ABCG2.”, *Cancer Res.*, 65, 4852-60
- 57 Robey RW, Steadman K, Polgar O, Morisaki K, Blayney M, Mistry P, Bates SE. (2004) “Pheophorbide a is a specific probe for ABCG2 function and inhibition.”, *Cancer Res.*, 64, 1242-6.
- 58 Zhou S, Zong Y, Ney PA, Nair G, Stewart CF, Sorrentino BP. (2005) “Increased expression of the Abcg2 transporter during erythroid maturation plays a role in decreasing cellular protoporphyrin IX levels.”, *Blood*, 105, 2571-6.
- 59 Woodward OM, Köttgen A, Coresh J, Boerwinkle E, Guggino WB, Köttgen M. (2009) “Identification of a urate transporter, ABCG2, with a common functional polymorphism causing gout.”, *Proc Natl Acad Sci U S A.*, 106, 10338-42.
- 60 Imai Y, Tsukahara S, Ishikawa E, Tsuruo T, Sugimoto Y. (2002) “Estrone and 17betaestradiol reverse breast cancer resistance protein-mediated multidrug resistance.”, *Jpn J Cancer Res.*, 93, 231-5.
- 61 Imai Y, Asada S, Tsukahara S, Ishikawa E, Tsuruo T, Sugimoto Y. (2003) “Breast cancer resistance protein exports sulfated estrogens but not free estrogens.”, *Mol Pharmacol.*, 64, 610-8.
- 62 Ding XW, Wu JH, Jiang CP. (2010) “ABCG2: a potential marker of stem cells and novel target in stem cell and cancer therapy.”, *Life Sci.*, 86, 631-7.
- 63 Khan I, Zaib S, Batool S, Abbas N, Ashraf Z, Iqbal J, et al.(2016)” Quinazolines and quinazolinones as ubiquitous structural fragments in medicinal chemistry: An update on the development of synthetic methods and pharmacological diversification. *Bioorganic & Medicinal Chemistry*”,24,2361-2381
- 64 Abbas SY, El-Bayouki KA, Basyouni WM, Mostafa EA.(2018)” New series of 4 (3H)-quinazolinone derivatives: Syntheses and evaluation of antitumor and antiviral activities”.*Medicinal Chemistry Research.*,27,571-582
- 65 EI-Hashash MAEAM, Salem MS, Al-Mabrook SAM.(2018)” Synthesis and anticancer activity of novel quinazolinone and benzamide derivatives”. *Research on Chemical Intermediates*,44,2545-2559
- 66 Noolvi MN, Patel HM, Bhardwaj V, Chauhan A.(2011)” Synthesis and in vitro antitumor activity of substituted quinazoline and quinoxaline derivatives:

- Search for anticancer agent. *European Journal of Medicinal Chemistry*,46,2327-2346
- 67 Dohle W, Jourdan FL, Menchon G, Prota AE, Foster PA, Mannion P, et al. (2018)''Quinazolinone based anticancer agents: Synthesis, antiproliferative SAR, antitubulin activity, and tubulin co-crystal structure''. *Journal of Medicinal Chemistry*,61,1031-1044
- 68 Ighachane H, Sedra MH, Lazrek H.(2017)'' Synthesis and evaluation of antifungal activities of (3H)-quinazolin-4-one derivatives against tree plant fungi''. *Journal of Materials and Environmental Science*,8:134-143
- 69 Alaa AM, Abou-Zeid LA, ElTahir KEH, Ayyad RR, Magda AA, El-Azab AS. (2016)''Synthesis, anti-inflammatory, analgesic, COX-1/2 inhibitory activities and molecular docking studies of substituted 2-mercapto-4 (3H)quinazolinones''. *European Journal of Medicinal Chemistry*,121,410-421
- 70 Abuelizz HA, El-Dib RA, Marzouk M, Al-Salahi R.(2018)'' In vitro evaluation of new 2-phenoxy-benzo [g] [1, 2, 4] triazolo [1, 5-a] quinazoline derivatives as antimicrobial agents''. *Microbial Pathogenesis*,117,60-67
- 71 Yang L, Ge S, Huang J, Bao X. (2018)''Synthesis of novel (E)-2-(4-(1H-1,2, 4-triazol-1-yl) styryl)-4-(alkyl/ arylmethyleneoxy) quinazoline derivatives as antimicrobial agents''. *Molecular Diversity*,22,71-82
- 72 Zhang L, Chen Q , Li XQ , Wu SQ , Wan JL, Ouyang GP.(2018)'' Synthesis and antibacterial activity of 2-substitued(3-pyridyl)-quinazolinone derivatives''. *Journal of Heterocyclic Chemistry*,55,743-749
- 73 Al-Salahi R, Abuelizz HA, Ghabbour HA, El-Dib R, Marzouk M.(2016)''Molecular docking study and antiviral evaluation of 2-thioxo-benzo[g] quinazolin-4(3H)-one derivatives''. *Chemistry Central Journal*,19,10-21
- 74 Gupta T, Rohilla A, Pathak A, Akhtar MJ, Haider MR, Yar MS.(2018)'' Current perspectives on quinazolines with potent biological activities: A review''. *Synthetic Communications*,48,1099-1127
- 75 Akduman B, Crawford ED.(2001)'' Terazosin, doxazosin, and prazosin: Current clinical experience''. *Urology*,58,49-54
- 76 McConnell JD et al.(2003)'' The long-term effect of doxazosin, finasteride, and combination therapy on the clinical progression of benign prostatic hyperplasia''. *The New England Journal of Medicine*,349,2387-2398

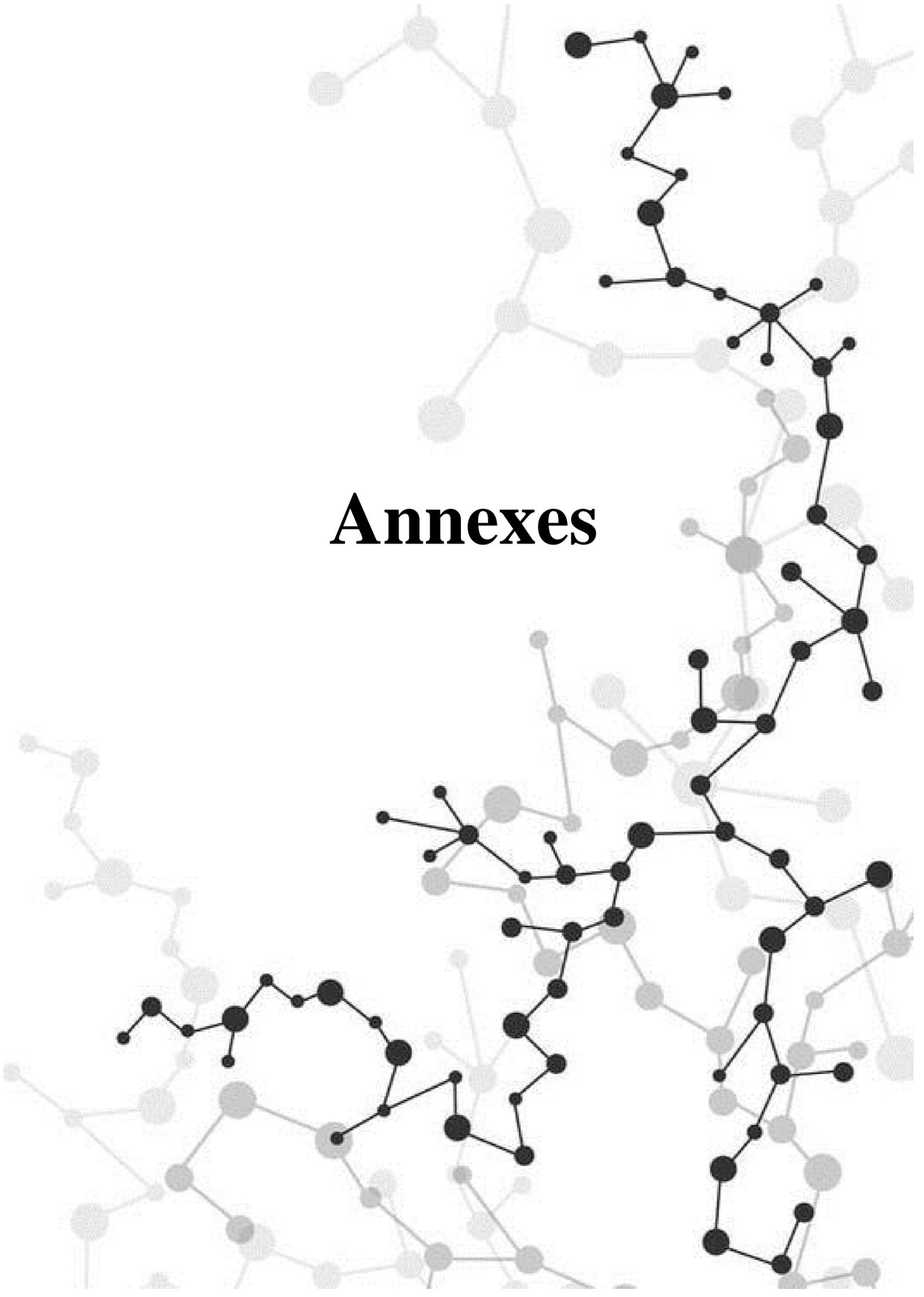
- 77 Johnston S, Pippen J Jr, Pivot X, Lichinitser M, Sadeghi S, Dieras V, et al. (2009)''Lapatinib combined with letrozole versus letrozole and placebo as first-line therapy for postmenopausal hormone receptorpositive metastatic breast cancer''. *Journal of Clinical Oncology*,27,5538-5546
- 78 Asif, M.(2014)'' Chemical Characteristics, Synthetic Methods, and Biological Potential of Quinazoline and Quinazolinone Derivatives'' *International Journal of Medicinal Chemistry*, Article ID 395637.
- 79 Morgan, G.T., ed(1904). *Abstract of Papers*. *Journal of the Chemical Society*. London: Gurney & Jackson, 1904. Print.
- 80 Iressa(Gefitinib)"([http://www.accessdata.fda.gov/drugsatfda\\_docs/label/2003/021399lbl.pdf](http://www.accessdata.fda.gov/drugsatfda_docs/label/2003/021399lbl.pdf)) (PDF). US Food and Drug Administration. 2 May 2003.
- 81 Lynch, Thomas J.; Bell, Daphne W.; Sordella, Raffaella; Gurubhagavatula, Sarada; Okimoto, Ross A.; Brannigan, Brian W.; Harris, Patricia L.; Haserlat, Sara M.; Supko, Jeffrey G.; Haluska, Frank G.; Louis, David N.; Christiani, David C.; Settleman, Jeff; Haber, Daniel A (May 20, 2004). "Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non-Small-Cell Lung Cancer to Gefitinib". *NEJM*. 350 (21): 2129–39.
- 82 Takimoto CH, Calvo E. (2008)"Principles of Oncologic Pharmacotherapy" ([http://www.cancernetwork.com/cancer\\_management11/chapter03/article/10165/1402628](http://www.cancernetwork.com/cancer_management11/chapter03/article/10165/1402628)) in Pazdur R, Wagman LD, Camphausen KA, Hoskins WJ (Eds) *Cancer Management: A Multidisciplinary Approach* (<http://www.cancernetwork.com/cancer-management-11/>). 11 ed.
- 83 Erlotinib"(<https://www.fda.gov/Drugs/InformationOnDrugs/ApprovedDrugs/ucm352317.htm>). US Food and Drug Administration. 14 May 2013.
- 84 Raymond E, Faivre S, Armand J (2000). "Epidermal growth factor receptor tyrosine kinase as a target for anticancer therapy". *Drugs*. 60 Suppl 1: 15–23, discussion41–2
- 85 Lapatinib"(<https://www.fda.gov/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDER/ucm129252.htm>). US Food and Drug Administration. 13 March 2007.
- 86 Wood ER, Truesdale AT, McDonald OB, Yuan D, Hassell A, Dickerson SH, Ellis B, Pennisi C, et al. (2004). "A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib): relationships among protein

- conformation, inhibitor off-rate, and receptor activity in tumor cells". *Cancer Research*. 64 (18),6652–9.
- 87 Rodriguez, A. (April 2008)''New type of drug shrinks primary breast cancer tumors significantly in just six weeks; research provides leads to a new target in\_cancer\_treatmentthe\_cancer\_stem\_cell([https://web.archive.org/web/20081126082044/http://www.eccoorg.eu/News/Pressroom/Pressrelease/page.aspx/439?xf\\_itemId=265&xf\\_catId=27](https://web.archive.org/web/20081126082044/http://www.eccoorg.eu/News/Pressroom/Pressrelease/page.aspx/439?xf_itemId=265&xf_catId=27)).Archived from the original ([http://www.eccoorg.eu/News/Pressroom/Pressrelease/page.aspx/439?xf\\_itemId=265&xf\\_catId=27](http://www.eccoorg.eu/News/Pressroom/Pressrelease/page.aspx/439?xf_itemId=265&xf_catId=27)) on 2008-11-26.
- 88 Nelson MH, Dolder CR (February 2006). "Lapatinib: a novel dual tyrosine kinase inhibitor with activity in solid tumors". *Ann Pharmacother*. 40 (2): 261–9.
- 89 Pao. W., Miller V., and Zakowski, M. (2004)'' EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib''. *Proceedings of the National Academy of Sciences of the United States of America.*, 101 (36), 13306-11.
- 90 Raymond, E., Faivre, S., Armand, J.(2000).''Epidermal growth factor receptor tyrosine kinase as a target for anticancer therapy''. *Drugs* 60 Suppl 1, 15-23.
- 91 Gasteiger J.(2003)''Handbook of Chemoinformatics: From Data to Knowledge (4 Volumes)'' ,Ed. Wiley-VCH Verlag, Weinheim. Allemagne,
- 92 Mannhold R Krogsgaard-Larsen P Timmerman. H. (1993) ''QSAR: Hansch Analysis and Related Approaches'' Ed. Wiley-VCH Verlag, Weinheim. Allemagne,
- 93 He.L(2005) "QSAR and Classification modeling: Prediction of biological activity of organic compounds from molecular structure" ,Thèse de doctorat,, Université de l'état de Pennsylvania. Etats Unies.
- 94 Mosier.P D(2003) "Prediction of chemical properties and biological activities of organic de Pennsylvania. Etats Unies. networks" Thèse de doctorat, Université de l'état de Pennsylvania. Etats Unies.
- 95 Crum Brown, A.Fraser.T.R(1868)'' On the connection between chemical constitution and de l'état de Pennsylvania''. *Etats Unies.chnia, brucia, thebia, codeia, morphia and nicotia*. *Trans. Roy.Soc. Edinbu rgh* ,25, 151-203.
- 96 Hansch,C Leo,A,D (1937)''Hoekmann, Exploring QSAR: hydrophobic, electronic and steric derivatives''. *J. Am. Chem. Soc* . 1937, 59, 96-103.

- 97 Free,S,M ,Wilson,J W A mathematical contribution to structure–activity studies. *J. derivatives. J. Am. Chem. Soc.* . 1937, 59, 96-103.
- 98 Grover,M. Singh,B Bakshi,M Singh,S Quantitative structure- Property relationships derivatives. *J. Am. Chem. Soc.* . 1937, 59, 96-103.
- 99 Katritzky,A R D,Fara, , C. . Petrukhin, R.O Tatham, D.B. Maran, U. Lomaka, A. Karelson, M.The present utility and future Potential for medicinal chemistry of M.Karelson, The present utility and future Potential for medicinal chemistry of QSAR/QSPR
- 100 Zhao, F. Li, H. Wu, L. Li, X. Li, J. W. J. G. M.(2012) Peijnenburg, *Ecotoxico. Environment.Saf.* 80, 273–279.
- 101 Srinivas A Reddy, S. Kumar, R. Garg, *J Mol Graph Model.* (2010),28(8),852–862.
- 102 Karelson,M(2000)’’ Molecular Descriptors in QSAR/QSPR’’, Wiley, New York.
- 103 Todeschini,R Consonni, V. and. Mannhold R (2009)’’Molecular Descriptors for Chemoinformatics’’, *Drug Discovery & Development*, 41(2).
- 104 Dudek,A,Z Arodz,T and.Gálvez,J.(2006)’’Computational methods in developing quantitative structure-activity relationships (QSARs)’’ A review’, *Combinatorial Chemistry & High Throughput Screening*, 9, 213–228.
- 105 Rekkab,S(2014)’’Drug Design et synthèse de nouveaux calix[8]arènes sulfoniques flexibles à activités anticorrosive et anticoagulantel, Thèse de Doctorat, Univ. Constantine, Alger.
- 106 Lipinski,C,A Lombardo, F. DominyB.W, and Feeney. B.W (1997)’’Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings’’, *Advanced Drug Delivery Reviews*, 6(1–3), 3–25.
- 107 Karelson.M(2000)’’Molecular descriptors in QSAR/QSP’’, Wiley, New York.
- 108 Labute.P(2000)’’A widely applicable set of descriptors’’, *Journal of Molecular Graphics and Modelling*, 18, 464–477
- 109 Rakotomalala,R(2011)’’ Pratique de la Régression Linéaire Multiple : Diagnostic et sélection de variables’’, version 2, Université Lumière Lyon 2.
- 110 Confais,J. Le Guen M(2006)’’ premiers pas en regression lineaire avec SAS ‘’, *Revue MODULAD*, 35, 220-363.

- 111 Roy,K. Kar.S and. Narayan R Das.(2015)''Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment, Chapter 6 - Selected Statistical Methods in QSARl'', Academic Press, Boston, 191–229.
- 112 Esposito,E,X Hopfinger, A. J. J. D.(2004)'' Madura, Methods in Molecular Biology'' 275, 131-213
- 113 Golbraikh, A. Tropsha A, Mol.J. Graph. Model. 20, 2002, 269–276
- 114 Hawkins,D,M J. Chem. Inf. (2004),Comput. Sci. 44,1–12
- 115 Shi,L,M Fang,H. Tong, W.J. Chem. (2001),Inf. Comput. Sci. 41, 186–195.
- 116 Nicola N Chirico, P. Gramatica, J. Chem.(2011), Inf. Model. 51(9), 2320–2335.
- 117 Eriksson,L ,Jaworska,J. Worth, A(2003),Environ. Health. Perspect. 111,1361–1375
- 118 Gramatica, P (2007),QSAR. Comb. Sci. 26, 694–701.
- 119 Netzeva, T I. Worth,A P Aldenberg,T . Benigni,R Cronin, M. TD Gramatica, P. a.J.
- 120 Eriksson, L., Jaworska, J., Worth, A., Cronin, M., Mc Dowell, R.M., Gramatica, P.(2003)'' Methods for reliability, uncertainty assessment, and applicability evaluations of regression based and classification QSARs''. Environ. Health Perspect,111(10)
- 121 Tropsha, A., Gramatica, P., Grombar, V.K.(2003)'' The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models''. QSAR & Combi. Sci. 22, 69-76
- 122 Trinajstić,N Nikolić,S. Basak, S C. Lukovits,I (2001) "Distances indices and their hypercounterparts: Intercorrelation and use in the structure-property modeling.''SAR QSAR Environ. Res., 12: 31-54.
- 123 Juvale,K.,Wiese,M.(2012)''4-Substituted-2phenylquinazolines as inhibitors of BCRP'', bioorg, Med. Chem. Lett. 22, 6766–6769.
- 124 Gallus,J Juvale,K. Wiese.M (2014)'' Characterization of 3-methoxy flavones for their interaction with ABCG2 as suggested by ATPase activity'', Biochim. Biophys. Acta 1838, 2929-2938.
- 125 Tawar,U Jain, A K Dwarakanath,B,S Chandra,R Singh,Y Chaudhury,N K Kaitan,D Tandon.V (2003)''Influence of phenyl ring disubstitution on benzimidazole and terbenzimidazole cytotoxicity: synthesis and evaluation as radio- protectors'', J. Med. Chem. 46 ,3785-3792.

- 126 Todeschini, R.; Maiocchi, A.; Consonni, V. (1999), The K Correlation Index: Theory Development and its Application in Chemometrics.. Chemom. Int. Lab. Syst, 46,13-29.
- 127 Todeschini, R.(1997)’’ Data correlation, number of significant principal components and shape of molecules’’ The K correlation index.. Anal. Chim. Acta, 348, 419-430.
- 128 Wold,S Eriksson,L(1995) ‘’Statistical Validation of QSAR Results, Validation Tools in Chemometrics Methods in Molecular Design’’, VCH Publishers, New York,309- 318.
- 129 Michael K. Krapf, Jennifer Gallus, Anna Spindler, Michael Wiese,(2019)’’ Synthesis and biological evaluation of quinazoline derivatives e A SAR study of novel inhibitors of ABCG2’’European Journal of Medicinal Chemistry,161,506-525.
- 130 Michael K. Krapf, Jennifer Gallus, Michael Wiese,(2017)’’ Synthesis and biological investigation of 2,4-substituted quinazolines as highly potent inhibitors of breast cancer resistance protein (ABCG2)’’ European Journal of Medicinal Chemistry,139,587-611.



# Annexes

**Annexes :**

- ❖ **L'incidence du cancer** : est le nombre de nouveaux cas de cancer diagnostiqués au cours d'une période donnée

**Tableau 5** : Signification des descripteurs moléculaires.

Descripteurs	Signification
<b>MATS7v</b>	Moran autocorrelation - lag 7 / weighted by atomic van der Waals volumes
<b>MATS2m</b>	Moran autocorrelation - lag 2 / weighted by atomic masses
<b>MATS5p</b>	Moran autocorrelation - lag 5 / weighted by atomic polarizabilities
<b>RDF010M</b>	Radial Distribution Function 10 / weighted by atomic masses
<b>EEig12x</b>	Eigenvalue 12 from edge adj. matrix weighted by edge degrees

**Tableau 6** : Les valeurs d'IC<sub>50</sub> et les cinq descripteurs.

Code	pIC <sub>50</sub>	MATS2m	MATS7v	MATS5p	RDF010m	EEig12x
1	-2.47712	-0.065	-0.167	0.211	0.496	2.118
2	-2.52504	-0.065	-0.177	0.243	0.493	2.118
3	-2.25042	-0.06	-0.13	0.174	0.561	2.118
4	-2.30535	0.04	0.08	0.157	0.639	1.926
5	-2.97909	-0.052	-0.186	0.219	0.471	1.389
6	-2.5563	0.03	-0.105	0.103	0.671	2.118
7	-2.19312	-0.056	-0.115	0.119	0.664	2.118
8	-3.40993	0.005	-0.163	0.135	0.677	1.746

9	-2.81425	0.053	0.058	0.106	0.743	1.926
10	-4.16137	-0.05	-0.206	0.214	0.455	1.393
12	-3.58659	0	-0.107	0.138	0.421	1.394
13	-3.86153	0	-0.177	0.108	0.421	1.497
14	-2.9952	0.071	-0.152	0.237	0.418	2.375
15	-4.17026	0.09	-0.193	-0.039	0.422	2.375
16	-3.86923	0.069	-0.168	0.145	0.525	1.584
17	-3.30643	0.017	-0.081	0.075	0.528	2.441
18	-3.56455	0.017	-0.199	0.08	0.528	2.476
20	-3.59638	0.031	-0.137	0.056	0.506	1.922
21	-2.62737	-0.056	-0.108	0.132	0.554	2.351
23	-1.7364	-0.051	-0.092	0.085	0.594	2.5
24	-1.96895	-0.051	-0.106	0.169	0.52	2.5
25	-4.02938	-0.095	-0.149	0.166	0.422	0.532
26	-3.4624	-0.057	-0.066	0.186	0.468	0.949
27	-3.41996	-0.061	-0.125	0.18	0.478	0.616
28	-3.1038	-0.087	-0.093	0.173	0.475	0.882
29	-2.80956	-0.008	0.025	0.116	0.496	0.882
30	-3.13033	-0.124	0.129	0.044	0.493	0.882
31	-3.90956	-0.065	-0.127	0.168	0.386	0.82
32	-4.4609	0.044	-0.053	0	0.58	0.884
33	-2.94547	-0.068	-0.172	0.191	0.55	1.393
34	-2.11394	-0.063	-0.149	0.183	0.564	2.376
35	-1.84261	-0.063	-0.159	0.214	0.516	2.384
36	-1.90309	-0.042	-0.134	0.216	0.57	2.386
37	-2.30963	-0.043	-0.158	0.206	0.6	1.497
38	-2.14613	-0.062	-0.115	0.147	0.603	2.376
39	-1.84448	-0.062	-0.161	0.186	0.604	2.383
40	-3.12057	-0.005	-0.062	0.164	0.619	1.584
41	-3.28556	0.015	-0.155	0.155	0.669	1.746
42	-2.18184	0.042	0.06	0.126	0.682	1.926
43	-3.07555	-0.1	-0.013	0.082	0.621	1.584
44	-2.55023	-0.047	-0.166	0.192	0.515	1.394

45	-2.44404	0.015	-0.155	0.155	0.667	1.746
46	-2.03342	-0.063	-0.176	0.147	0.527	2.377
47	-3.44404	-0.02	-0.218	0.172	0.643	2
48	-2.37291	-0.055	-0.152	0.141	0.491	2.599
49	-2.17898	-0.038	-0.135	0.175	0.543	2.603
50	-2.07918	-0.056	-0.119	0.106	0.582	2.598
51	-2.08279	0.001	-0.06	0.128	0.594	2.378
52	-1.74507	0.049	0.067	0.097	0.654	2.392
53	-2.90363	0.016	-0.035	-0.056	0.562	2.389
54	-2.32222	0.07	-0.056	-0.008	0.751	2.397
55	-2.41996	-0.037	-0.076	0.06	0.572	2.483
56	-1.97681	-0.102	-0.012	0.054	0.589	2.378
57	-1.94743	-0.102	-0.137	0.09	0.586	2.391
58	-2.45332	-0.013	-0.149	0.141	0.689	2.378
59	-2.34439	0.067	0.022	0.105	0.681	2.38
60	-2.36173	0.068	-0.017	0.033	0.778	2.536
61	-2.1038	0.041	-0.1	0.044	0.697	2.599
63	-4.20952	0.103	-0.1	-0.008	0.56	2.584
64	-3.24797	0	-0.125	0.044	0.495	2.389
65	-2.22011	-0.055	-0.125	0.217	0.489	2.5
66	-2.52634	-0.056	-0.089	0.19	0.566	2.5
67	-2.02938	0.001	-0.04	0.192	0.593	2.401
68	-1.93802	0.049	0.083	0.149	0.659	2.416
69	-2.47276	0.023	-0.105	0.088	0.737	2.392
70	-2.35218	0.045	-0.068	0.024	0.78	2.591
71	-2.80754	0.001	-0.124	0.068	0.594	2.382
72	-2.14301	0.001	-0.045	0.169	0.594	2.379
73	-1.83251	0.001	-0.056	0.193	0.594	2.387
74	-2.46389	0.011	-0.02	0.131	0.552	2.613
75	-3.07555	0.019	-0.043	0.168	0.56	2.389
76	-2.55509	0.004	-0.012	0.145	0.687	2.379
77	-3.21484	0.014	-0.061	0.178	0.596	1.583
78	-2.90902	0.042	0.014	0.151	0.7	1.584

79	-2.68124	0.079	0.11	0.117	0.764	1.926
80	-2.26245	0.001	-0.153	0.153	0.594	2.41
82	-3.33041	0.042	-0.073	0.136	0.699	1.661
83	-2.62118	0.079	0.042	0.104	0.746	1.927
84	-1.91645	0.049	0.059	0.134	0.663	2.41
85	-1.64542	0.049	0.048	0.153	0.661	2.42
86	-2.78533	0.06	0.085	0.103	0.626	2.801
87	-2.4843	0.065	0.067	0.154	0.714	2.423
88	-2.13354	0.053	0.093	0.116	0.734	2.41
89	-2.90037	0.167	0.072	0.013	0.731	2.426
90	-2.56229	0.079	0.092	0.12	0.766	1.878
92	-2.66464	0.06	0.044	0.142	0.666	1.878
93	-2.75587	0.096	0.081	0.079	0.653	2.41

**Année universitaire : 2019/2020**

**Présenté par :**

**BOUFENNARA FERAL**

**BRIK MERIEM**

**ETUDE CHIMIO METRIQUE DES DERIVES DE QUINAZOLINE POUR LE  
TRAITEMENT DE CANCER DU SEIN**

**Mémoire de fin de cycle en vue de l'obtention du diplôme de Master en Biochimie Appliquée.**

**Résumé :**

Dans notre travail l'outil QSAR de la chimio-métrie a été utilisé pour modéliser l'activité inhibitrice d'ABCG2 représenté par le cologarithme décimal d'IC<sub>50</sub> (pIC<sub>50</sub>) de 87 dérivés de quinazoline. Un model QSAR a été construit à partir de 5 descripteurs MATS2m, MATS7v, MATS5p, RDF010m, EEig12xc corrélés avec pIC<sub>50</sub>.

Le modèle QSAR obtenu a donné des meilleurs paramètres statistiques ( $R^2=75.89$ ,  $Q^2_{LOO}=70.99$ ,  $Q^2_{ext}=74.37$ ,  $F=40.2982$ ) pour prédire l'activité inhibitrice. Le modèle proposé a prouvé sa robustesse, sa bonne capacité prédictive (Interne et externe) ainsi qu'une bonne stabilité. Il pourrait être utilisé efficacement pour estimer les molécules non testés expérimentalement.

**Mots clés :** Quinazoline, QSAR, ABCG2, pIC<sub>50</sub>, Chimio-métrie

**Laboratoire de recherche :** laboratoire de l'informatique N°18 au niveau du centre de recherche de la biotechnologie (C.R.B.T), Constantine

**Jury d'évaluatin :**

**Président du jury :** Mr. BENSEGUENI ABDERRAHMANE Pr. UFM, Constantine 1

**Encadreur:** Mr. DEMS MOHAMED ABDESSELEM. MRA CRBt Constantine

**Examineur :** Mr. MOKRANI ELHASSEN MAA UFM, Constantine 1

**Date de soutenance :** 30/09/2020