



République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Med Khider Biskra

Faculté des Sciences Exactes et des Sciences de la Nature et de la Vie

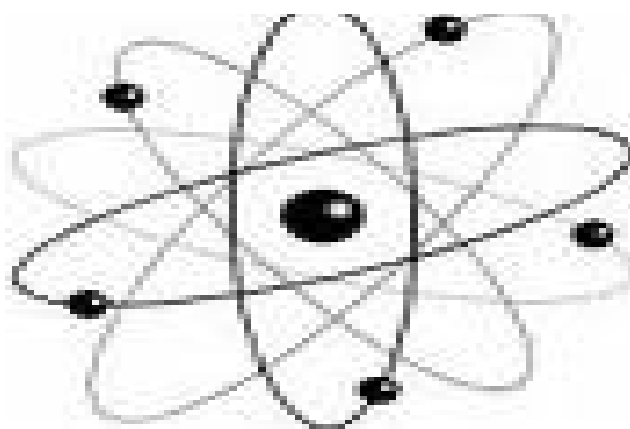


Département des Sciences de la Matière

Domaine : Sciences de la Matière

Filière : Chimie

Spécialité : Chimie pharmaceutique



*Mémoire de fin d'étude en Master  
Intitulé*

***ETUDE QSAR DE CERTAIN DERIVES 7-SUBSTITUÉ  
4-AMINOQUINOLINE COMME DES AGENTS  
ANTIPALUDIQUES.***

*Présenté par :*

Aied Maroua

*Devant le jury :*

Mr : Belaidi Salah  
Mme. Ouassaf Moubarka  
Mca: Mellaoui Malika

*Président  
Rapporteur  
Examineur*

Année Universitaire 2017-2018



# Remerciements

*Avant toute chose, Nous remercions  
du plus profond de notre cœur dieu le tout puissant de nous avoir donné le courage et la  
volonté d'achever ce travail. d'achever ce travail.*

*Je remercie mon encadreur*

*Pr.OUSSAF MOUBARAKA pour son aide précie*

*Ses conseils, ses orientations, ses orientations, et sonpatience.*

*Nous adressons notre sincères remerciements aux membres du jury qui ont accepté de  
juger ce travail : M.BELAIDI SALAH Professeur à l'Université de Biskra et Mellaoui  
Malika mca l'Université de Biskra*

*Je remercie également tous les professeurs du département de chimie*

*Enfin, je remercie aussi tous cequi a contribué de loin ou de qui a contribué de loin ou de près  
à la réalisation de ce travail. près à la réalisation de ce travail*



# Dédicace

*Je dédie ce travail en signe de reconnaissance.  
A celui qui a lutté et sacrifié pour m'offrir les conditions propices à  
ma*

*réussite  
Mon très cher père  
MOHAMED*

*A celle qui m'a étreint de tendresse et d'affection et qui a constitué la  
première école de mon existence*

*Ma très précieuse, chaleureuse et aimable mère  
SAFIA*

*Grace à mes parents que j'ai pu faire mes études et gravir les pentes  
qui me  
semblaient infranchissables*

*. A mes frères*

*\* A ma famille*

*\* A tous mes copains*

*A tout le groupe de' chimie.*

# Sommaire

Titre	Page
Remercîments.....	-
Dédicace .....	-
Sommaire .....	-
Liste des figures .....	1
Liste des tableaux .....	2
Liste des principales abréviations .....	3
Introduction .....	4
<b>Chapitre I</b>	
<b>Généralités sur le Paludisme et les Antipaludique</b>	
I.1 le Paludisme .....	7
I.1.1 Introduction .....	7
I.1.2 .1.une époque clinique .....	8
I.1.2 .2.une époque des découvertes thérapeutiques .....	8
I.1.2 .3.une époque des découvertes Biologiques .....	8
I.1.3. Agent pathogène et vecteur .....	8
I.2 Antipaludique .....	10
I.2.1. L'histoire de la pharmacie antipaludique .....	10
I.2 .2. Classification biologique des antipaludéens .....	11
I.2 3. Structure chimique des antipaludéens .....	12
<b>Chapitre II</b>	
<b>Bases théoriques des méthodes de calculs utilisée</b>	
II.1.Introduction. ....	20
II.2.Histoire Des QSAR .....	20
II.3. Principe des méthodes QSPR/QSAR .....	22
II .4. Méthodologie générale d'une étude QSPR/QSAR .....	22
II .5.Les descripteurs physico-chimiques .....	24
II 5.1.Le coefficient de partage Octanol/Eau :(logp) .....	24
II .5.2.La réfractivité moléculaire :(Ref) .....	24

II.5.3. La polarisabilité,(pol) .....	25
II .5.4.Surface moléculaire, (SAG) .....	25
II.5.5. Volume moléculaire : (V) .....	25
II .6.Interprétation et validation d'un modèle QSAR/QSPR .....	26
II.6.1. Validation interne .....	26
II .6.2. Coefficients et tests statistiques standards .....	27
II .6.2.1. Coefficient de corrélation r (et coefficient de détermination $r^2$ ) .....	27
II .6.2.2. Le coefficient de détermination ajusté $r^2_{adj}$ .....	28
II .6.2.5. Le test de Fisher F .....	28
II. 6. 3. Validation externe. ....	29
II.7.Logiciel « XLSTAT» .....	29
II.8. Méthodes statistiques .....	30
II.8.1. La statistique descriptive .....	30
II.8.2. La statistique décisionnelle ou prédictive .....	31
<b>Chapitre III</b>	
<b>Résultats et discussions</b>	
III .1. Introduction .....	37
III.1.1. Méthodologie .....	37
III.1.1.1. Base de données .....	37
III.1.1.2.Sélection des descripteurs et méthodes de calcul .....	37
III.2. Structures chimiques des dérivés de 7-substituted -4-aminoquinolin	38
III.3..Étude des propriétés physico-chimiques des derive 7-substitued -4aminoquinoline .....	41
III.2.1.Interprétation des resultants .....	42
III.4. Modèles QSAR et analyse statistiques. ....	41
III.4.1. Résultats et discussions .....	41
III.4.1. 1. Analyse en composantes principales (acp) .....	41
III. 4.1.2..Régression linéaire multiple (MLR) .....	47
III.4.1.3 Régression non linéaire multiple (MNLR) .....	49
III.4.1.4: régression par les moindres carrés partiels (PLS) .....	50
Conclusion .....	54

## Liste des figures

<b>Figure</b>	<b>Titre</b>	<b>Page</b>
<b>Figure I. 1:</b>	La répartition mondiale du paludisme en 2010 selon l’OMS [29]	<b>7</b>
<b>Figure I.2</b>	Anophèle femelle	<b>9</b>
<b>Figure I. 3</b>	Cycle de vie du parasite	<b>10</b>
<b>Figure I. 4</b>	Chronologie de la découverte des médicaments antipaludiques majeurs (Hudson, 1993 ; Wells, 2011 ; Krafts, 2012 ; Patil, 2014)	<b>11</b>
<b>Figure I.5</b>	Structure de quelques molécules utilisées en tant que médicaments antipaludiques	<b>12</b>
<b>figure I.6</b>	Structure de L'amodiaquine_(C <sub>20</sub> H <sub>22</sub> ClN <sub>3</sub> O), 4-(7-chloro-4-quinolyamino)-2- diéthylaminométhylphénol	<b>13</b>
<b>figure I.7</b>	Constantes d’acidité (pKa) de la chloroquine et de l’amodiaquine (* valeur calculée)	<b>13</b>
<b>Figure I.8</b>	Structure chimique de la chloroquine	<b>15</b>
<b>Figure. II. 1</b>	Modèle d'étude quantitative des relation structure/ Activité (QSAR)	<b>20</b>
<b>Figure .II .2</b>	Procédure d’obtention et de validation d’un modèle QSAR	<b>23</b>
<b>Figure .III .1</b>	Les principaux composants et leurs variances	<b>46</b>
<b>Figure. III 2</b>	Cercle de corrélation	<b>46</b>
<b>Figure. III .3</b>	Corrélation entre l’activité biologique expérimentale et prédite pour les modèle MLR	<b>48</b>
<b>Figure. III .4.</b>	Corrélation entre les valeurs résiduelles et l’expérimentales de modèle MLR	<b>48</b>
<b>Figure III .5</b>	Corrélation entre l’activité biologique expérimentale et prédite pour les	<b>50</b>
<b>Figure III 6</b>	Corrélation entre les valeurs résiduelles et expérimentales de modèle MNLR	<b>50</b>
<b>figure 7</b>	Corrélations de pIC <sub>50</sub> observées et prédites par PLS	<b>51</b>

## Liste des tableaux

Tableau	Titre	Page
<b>Tableau III.1</b>	Structures chimiques des dérivés 7-substitué -4-aminoquinoline	<b>38</b>
<b>Tableau III.2</b>	Paramètres physicochimiques des dérivés 7-substitué -4-aminoquinoline	<b>41</b>
<b>Tableaux .III.3</b>	Matrice de corrélation du modèle QSAR	<b>46</b>
<b>Tableau. III. 5</b>	Activité antipaludique expérimentale, prédite et résiduelle des dérivés de 7-substitué -4-aminoquinoline pour les trois modèles (MLR. MNLR. PLS)	<b>51</b>

## Liste des principales abréviations

QSAR: Quantitative structure –Activité Relationship.

ACP : Analyse en composantes principales

RLM: Régression linéaire multiple

RNLM: Régression non linéaire multiple

PLS: Partial Least Squares

AQ : L'amodiaquine

CQ : La chloroquine

AQQI: l'amodiaquine quinone-imine

GSH: glutathione

L'OMS: Organisation Mondiale de la Santé

G6PD : glucose-6-phosphate déshydrogénase

LOO : Leave-One-Out

LMO : Leave-Many-Out

PIC50: log. Inhibitory concentration of 50%

CAH : classification ascendante hiérarchique

MM : Mécanique Moléculaire

EQM: l'erreur quadratique moyenne

MW : Molar Weight

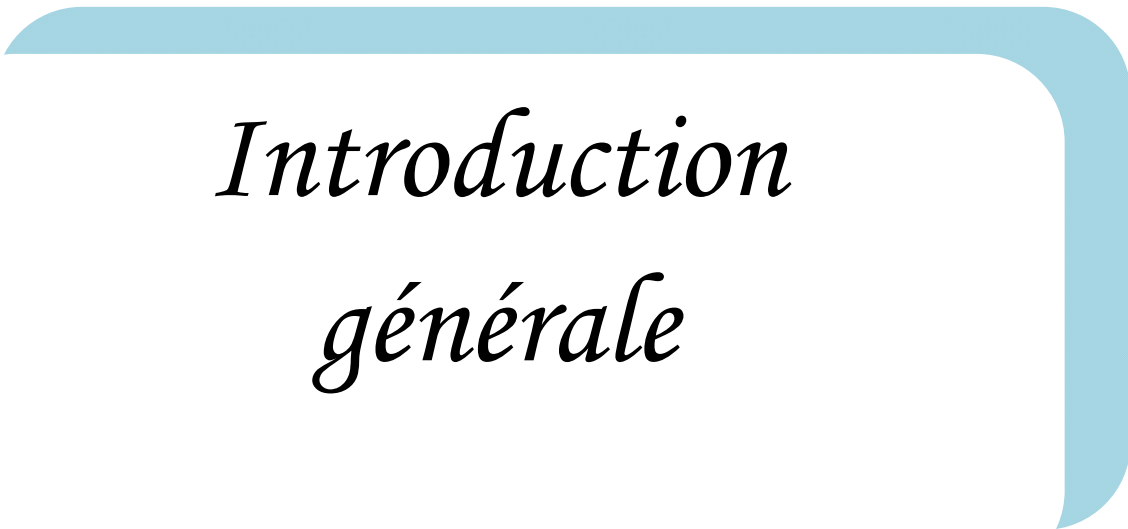
HE : d'hydratation Pol Polarisabilité

Log p :Coefficient de partage

SAG: Surface Area Grid

V : Volume moléculaire





*Introduction  
générale*

L'outil informatique a été mis en œuvre dans le domaine chimique dans le but de modéliser les structures moléculaires. Depuis cette époque, l'augmentation de la puissance de calcul des ordinateurs a permis à la modélisation moléculaire de s'appliquer à des structures de plus en plus complexes de manière de plus en plus rigoureuse et réaliste. Ainsi, petit à petit, la modélisation moléculaire est devenue une nouvelle technique de compréhension des phénomènes chimiques et un outil de travail à part entière dans le domaine de la chimie structurale. Donc, si on s'intéresse aux structures et conformations moléculaires, l'emploi de cette méthode peut être la source d'une bien meilleure compréhension des géométries observées ou attendues [1].

La découverte de médicament est un processus long et complexe. Cette discipline peut intervenir à différents niveaux du processus de découverte d'un médicament. Parmi les techniques de chémo-informatique nous pouvons citer les techniques de QSAR qui consiste à trouver une corrélation entre une activité biologique mesurée pour un panel de composés et certains descripteurs moléculaires[2].

L'objectif principal des études QSAR est une règle empirique ou une fonction dépendante des descripteurs structuraux des composés étudiés aux bio activités. Cette règle ou cette fonction est idéale pour prédire les mêmes bio activités des composés non impliqués dans l'apprentissage à partir de leurs descripteurs structurels. Le fait que les bio activités soient caractérisées par une grande précision de la performance de la méthode d'analyse des données multivariées, à condition que la propriété soit liée aux descripteurs. Beaucoup d'analyse de données multivariées

Des méthodes telles que l'analyse des propriétés principales (ACP) et le réseau neuronal artificiel (ARN) ont été utilisées dans les études QSAR. ANN offre une précision supérieure dans la plupart des cas mais tend à sur ajuster les données d'entraînement. Il existe un grand nombre de descripteurs moléculaires pouvant être utilisés dans les études QSAR. Une fois validés, les résultats peuvent être utilisés pour prédire les activités de composés non testés. Récemment, la conception assistée par ordinateur de médicaments basée sur QSAR a été employée avec succès pour développer de nouveaux médicaments pour le traitement du cancer et d'autres maladies

Le développement continu de la chimie hétérocyclique désormais incontournable en synthèse organique, l'hétérocycle constitue le squelette de base dans une grande variété de composés d'intérêt chimique, biologique, pharmacologique et industriel. Les hétérocycles

contenant un atome d'azote, tels que: les quinoléines, les dizocilpines, les pyrimidines, les acridines, les phénothiazines, les indoles..., se trouvent dans de nombreux produits naturels et sont parmi les éléments les plus présents dans le domaine de la chimie médicinale. La recherche de nouveaux hétérocycles azotés à potentiel biologique est cruciale pour le développement de nouveaux composés répondent à une demande toujours croissante de molécules originales [4].

Le principal objectif de travail à l'aide des calculs effectués une relation entre structure et activité et comment amélioré la structure pour avoir une meilleure activité. En utilisant les méthodes SAR et QSAR, pour chercher un modèle relie les paramètres calculés d'une série de des dérivés d'ainoquinoline à leurs activités antipaludiques et peut aussi la prédire

Le présent travail comporte trois chapitres. Le premier chapitre nous allons présenter une généralité sur le paludisme et généralités sur les Antipaludique

Dans le deuxième chapitre, nous développerons un rappel théorique sur les méthodes QSAR et les méthodes statistiques utilisées pour obtenir les modèles QSAR l'analyse en composantes principales ACP, la régression linéaire multiple RLM, la régression non linéaire multiple RNLM, la régression par les moindres carrés partiels PLS, ceci est très important dans le cadre des développements des relations mathématiques à partir des données, suivie par la représentation du formalisme de la méthodologie choisie, le diagramme (les étapes) de travail effectué

Dans troisième chapitre est consacré à la méthodologie du travail réalisé dans cette thèse, aux résultats obtenus et à leurs discussions

## References

- [1] A. Albright; K. Burdett.H. Whangbo, Orbital interactions in chemistry, John Wiley & Sons, **2013**.
- [2]. TABET Rim; Etude des propriétés QSAR d'une série des isoindoles à activité thérapeutique, Mémoire Master, Université Med Khider Biskra,2013/2014.
- [3]. Samir CHTITA, Modélisation de molécules organiques hétérocycliques biologiquement actives par des méthodes QSAR/QSPR.Recherche de nouveaux médicaments, THÈSE de Doctorat, Université de maroc .2016/2017.

# *Chapitre I*

*Généralités sur le Paludisme et  
les Antipaludique*

## I.1 le Paludisme

### I.1.1 Introduction

Paludisme ou la fièvre des marécages (marécage : terrain humide où l'on trouve des marais) ou fièvre intermittente ou malaria est une maladie parasitaire qui est une érythrocytopathie fébrile et hémolysante due au développement et à la multiplication chez l'homme d'hématozoaires du genre Plasmodium [1]. Les données parasitologiques obtenues lors de différentes études en Afrique subsaharienne, montrent que cinq espèces plasmodiales sont rencontrées : Plasmodium falciparum (P.falciparum), Plasmodium malariae (P.malariae), Plasmodium ovale (P.ovale), Plasmodium vivax (P.vivax) et plasmodium knowlesi (P.knowlesi). Le P. falciparum est la forme la plus répandue, et la seule meurtrière [2]. Les régions endémiques intertropicales représentent 90 % de l'incidence mondiale du paludisme et la grande majorité des décès, concernent principalement les enfants de moins de 5ans [3].

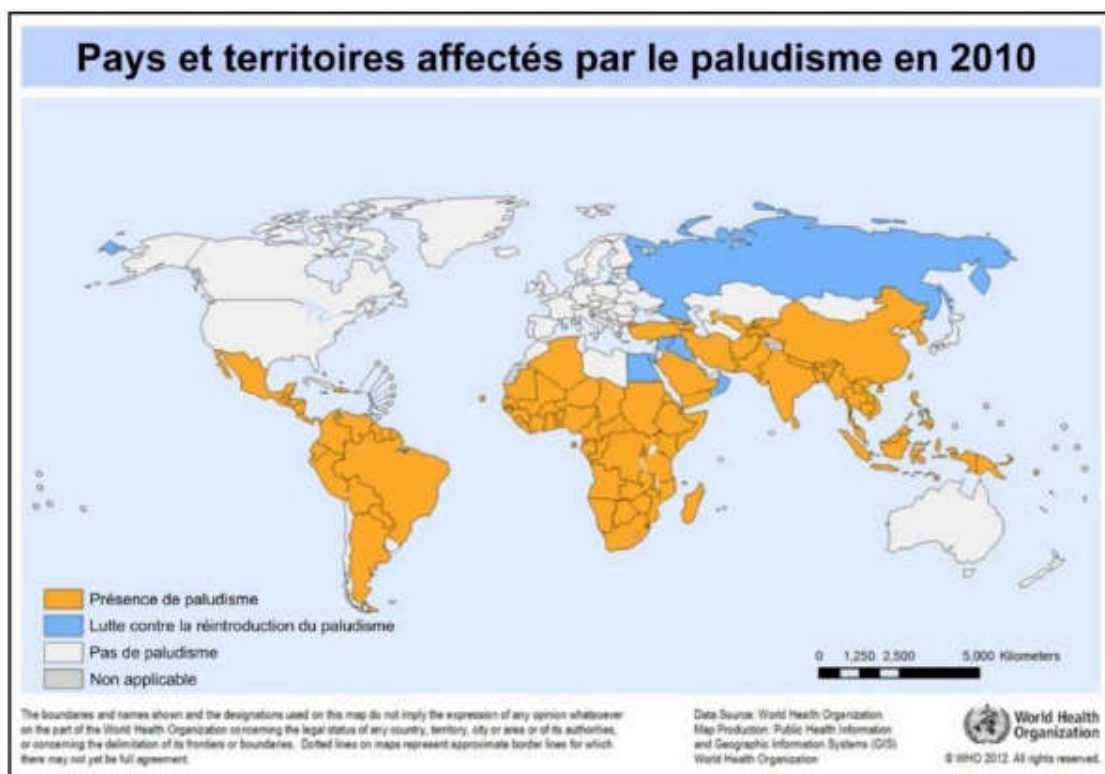


Figure I. 1: La répartition mondiale du paludisme en 2010 selon l'OMS [29]

### I.1.2 Historique :

Le paludisme est l'une des plus vieilles maladies que l'humanité ait connue. Son histoire peut être divisée en trois étapes [4]

**I.1.2 .1.une époque clinique :**

Les fièvres intermittentes ont été évoquées dans les textes des médicastres chinois et égyptiens Hippocrate, dans son traité des airs des eaux, opposait fièvre périodique et fièvre intermittente.

La notion de fièvre intermittente se dégageait dans les régions où l'on trouvait des eaux stagnantes, des marécages (Palu) et dans les zones où l'air était vicié (malaria) d'où le nom de paludisme ou malaria.

**I.1.2 .2.une époque des découvertes thérapeutiques :**

En 1630, Don Francisco Lopez apprend des indiens du Pérou les vertus de l'écorce du quinquina et à partir de cette époque, les fièvres ont été divisées selon leur sensibilité à cette drogue.

**I.1.2 .3.une époque des découvertes Biologiques :**

En 1820, Pelletier et Caventou isolent du quinquina l'alkaloïde actif : la quinine. L'agent pathogène est découvert par Laveron en 1880 à Constantine.

Marchiafava celli et Golgi distinguent bientôt trois espèces de parasites chez l'Homme : Plasmodium falciparum ; Plasmodium vivax et Plasmodium malariae.

De 1895 à 1897, la transmission de cette affection par un moustique du genre anophèle est soupçonnée et confirmée par Ross et Grassi en 1898.

**I.1.3. Agent pathogène et vecteur :**

Cinq espèces plasmodiales sont inféodées à l'homme. Il s'agit de:

1. Plasmodium falciparum : responsable de la quasi-totalité des décès dus au paludisme. Il représente 85 à 90% de la formule parasitaire au Mali,
2. Plasmodium malaria: 10 à 14%
3. Plasmodium ovale : avec moins de 1% [4]
4. Plasmodium vivax : sa présence a été confirmée en transmission autochtone au nord du Mali, dans les populations leucodermes en 1988[4].
5. plasmodium knwolesi découvert récemment en Malaisie

Le cycle de développement de Plasmodium peut être divisé en deux parties (Figure I.3):

- Le cycle asexué ou schizogonie se produit chez l'être humain (hôte intermédiaire).

- Le cycle sexué ou sporogonie se produit chez le moustique Anophèle (vecteur et hôte définitif).



Figure I.2 : Anophèle femelle

### a) Cycle asexué (chez l'homme)

Le cycle de vie de Plasmodium est subdivisé en deux phases chez l'homme : une phase hépatique et une phase érythrocytaire.

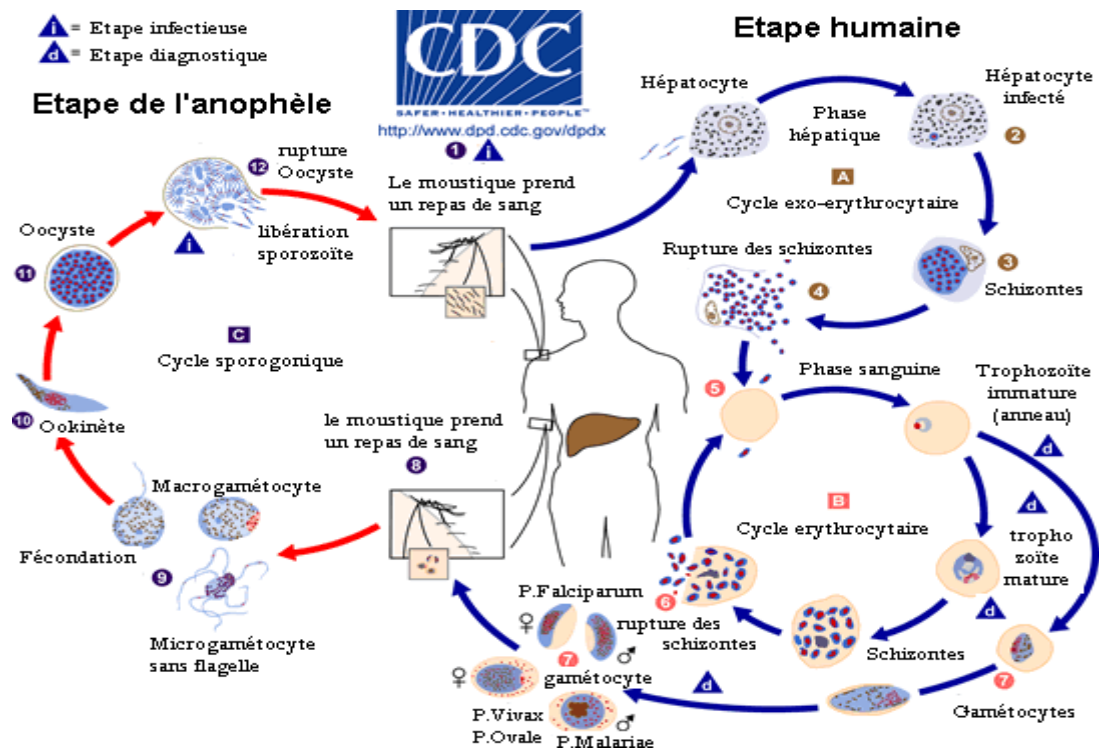
- ✓ **Phase hépatique** : Lorsque le moustique (Anophèle femelle) pique l'homme pour prélever son repas sanguin, il lui transmet le parasite sous forme sporozoïte. Les sporozoïtes gagnent les cellules du foie (hépatocytes) où ils se transforment en schizontes pré-érythrocytaires (formes multinucléées) qui éclatent pour libérer les mérozoïtes dans le sang. Le foie est un lieu de développement pour le parasite, mais aussi, un lieu où le parasite peut résider (hypnozoïtes) et réapparaître plus tard lorsqu'il s'agit d'une infection par *P. vivax* ou *P. ovale*.<sup>6, 7</sup>
- ✓ **Phase érythrocytaire** : Très rapidement (48 heures pour *P. falciparum* et *P. vivax*, contre 72 heures pour *P. malariae*), les mérozoïtes envahissent les globules rouges où ils se multiplient et provoquent l'éclatement des érythrocytes. Les mérozoïtes libérés pénètrent dans d'autres globules rouges et débutent un nouveau cycle de réplication. Cette partie du cycle correspond à la phase clinique et la personne infectée devient fébrile. En l'absence de traitement, tous les mérozoïtes évoluent progressivement au même rythme (ils deviennent synchrones) et provoquent l'éclatement d'un grand nombre de globules rouges de manière périodique. Après un certain nombre de cycles érythrocytaires, certains mérozoïtes se transforment en gamétocytes et restent en circulation dans le sang pendant 10 à 15 jours.

### b. Le cycle sexué (sporogonique) chez l'anophèle

Lorsque le moustique (Anophèle femelle) pique une personne infectée pour prélever son repas sanguin, il ingère des gamétocytes. Ces derniers se transforment en gamètes mâles et femelles



qui fusionnent en un œuf libre et mobile appelé ookinète. Cet ookinète évolue ensuite en oocyste dans lequel se multiplient plusieurs cellules parasites produisant des sporozoïtes. Ces derniers migrent vers les glandes salivaires du moustique qui peut infecter à nouveau une personne saine et débiter ainsi un cycle de transmission homme – moustique – homme.



Source: [http://www.uni-tuebingen.de/biometry/me/me\\_e\\_malaria\\_cycle.html](http://www.uni-tuebingen.de/biometry/me/me_e_malaria_cycle.html)

Figure I.3: Le cycle de développement de *Plasmodium*

## I.2 Antipaludique

Deux antipaludiques extraits des plantes médicinales sont toujours parmi ceux les plus utilisés et les plus efficaces. Il s'agit de la quinine, qui demeure un antipaludique majeur préconisé dans le traitement des cas de paludisme grave ou d'urgence et l'artémisinine (qinghaosu) ainsi que ses dérivés (artéether, artésunate et artéméther) obtenus par hémisynthèse qui occupent actuellement une place primordiale lors de paludismes dits récessifs et neurologiques.

### I.2.1. L'histoire de la pharmacie antipaludique

L'histoire de la pharmacie antipaludique démarre en 1820 avec la découverte de la quinine par Pierre-Joseph Pelletier et Joseph Bienaime Caventou (Pelletier, 1820). La découverte du bleu de méthylène, un composé redox, marque une nouvelle étape en 1891 (Guttman, 1891). Le papyrus Ebers d'Egypte datant de 1600 av. J.-C. a recommandé l'utilisation de l'*Artemisia*

annua (ou wormwood ou qinghao) contre les symptômes du paludisme (Baker, 2001), mais ce n'est qu'en 1971 que l'artémisinine est isolée par des scientifiques chinois (Qinghaosu Antimalaria Coordinating Research Group, 1979 ; Tu, 2011). Depuis de nombreux nouveaux composés ont été synthétisés de novo ou isolés de plantes. [4]

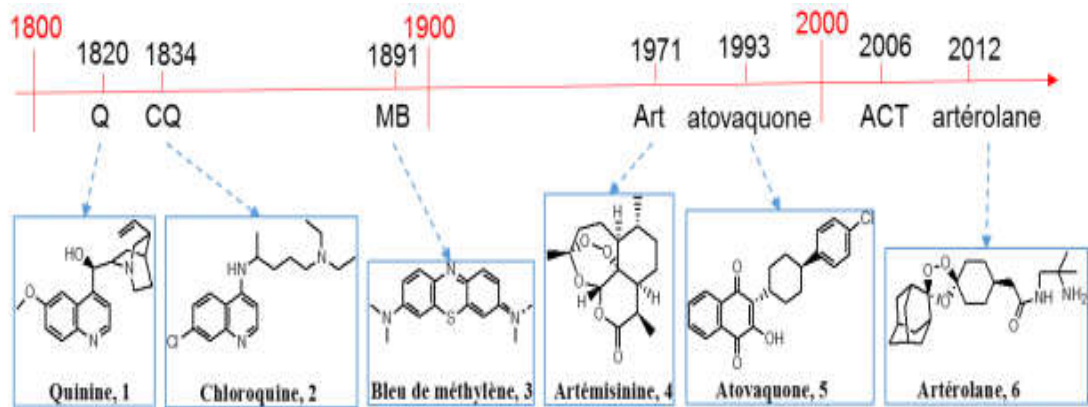


Figure I. 4. Chronologie de la découverte des médicaments antipaludiques majeurs (Hudson, 1993 ; Wells, 2011 ; Krafts, 2012 ; Patil, 2014)

## I.2 .2. Classification biologique des antipaludéens

### 2.1. Les schizontocides sanguins.

#### 2.1.1. Groupe 1: Les schizontocides naturelle

- ✓ Alcaloïdes du quinquina

Quinine

- ✓ Sesquiterpènes lactoniques de *Artemisia annua*

Artémésinine

#### 2.1.2. Groupe 2: Les schizontocides hémisynthétiques

- ✓ Artémether
- ✓ Artésunate
- ✓ Artéether (Artemotil)

### 2.2. Les gamétocytocides

- ❖ Dérivés de l'Artémisinine: DHA, Artésunate, Artémether, Artéether
- Amino-8-quinoléines: Primaquine, Tafénoquine (en cours de développemet

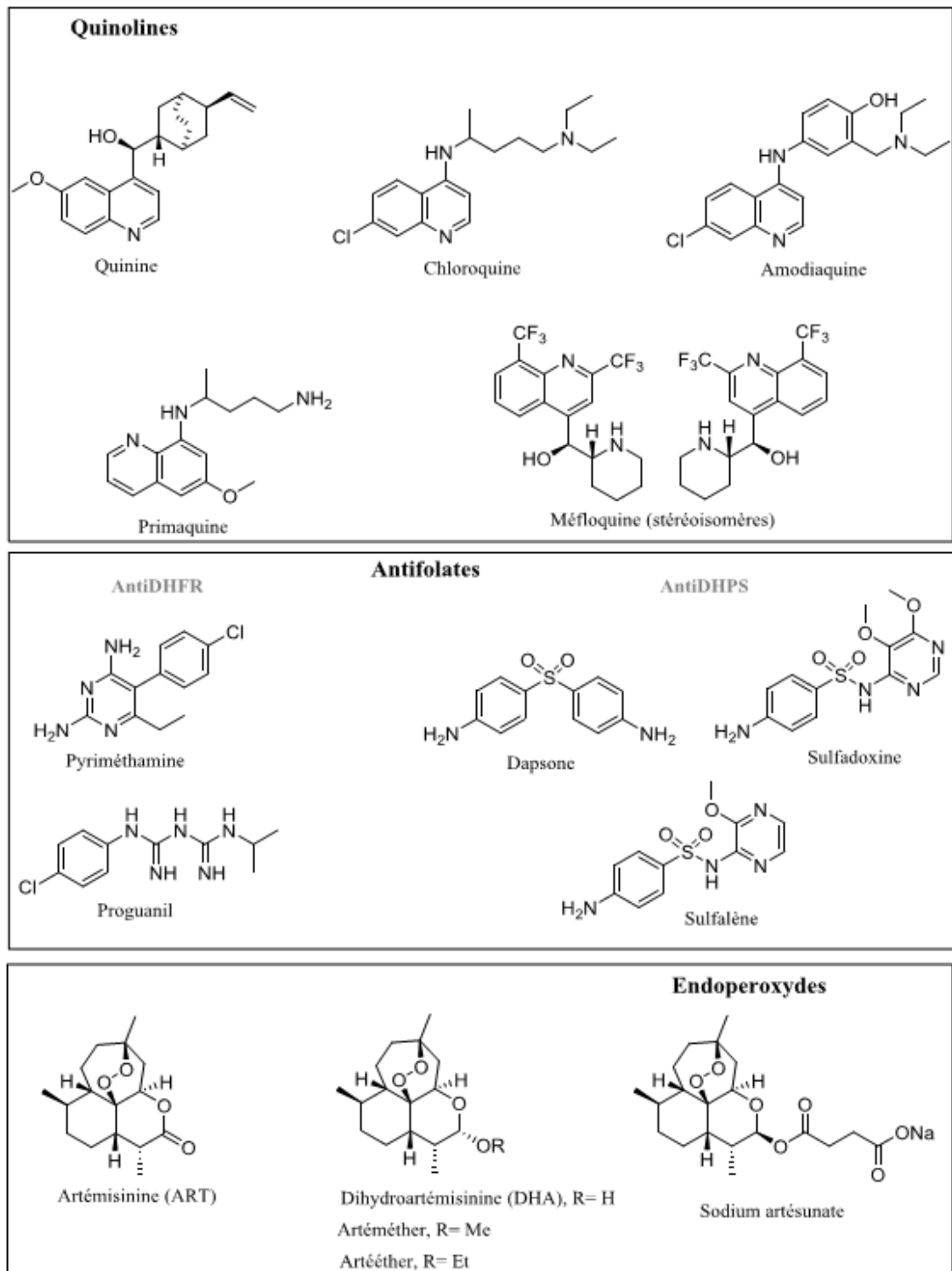


Figure I.5.. Structure de quelques molécules utilisées en tant que médicaments antipaludiques

### I.2 3. Structure chimique des antipaludéens

A l'exception de l'artémisinine et de ses dérivés, la structure chimique des molécules antipaludiques s'organise en général autour d'un noyau aromatique. Il peut être le benzène ou

un hétérocycle mono ou polycyclique contenant un à trois hétéro-atomes;. La présence fréquente d'un atome d'azote dans le cycle semble être indispensable à l'activité anti malarique

### I.2 3.1.L'Amodiaquine

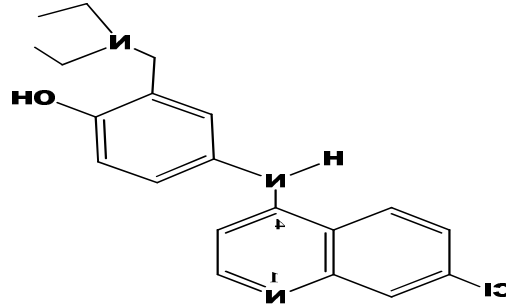


Figure I.6 . Structure de L'amodiaquine\_(C<sub>20</sub>H<sub>22</sub>ClN<sub>3</sub>O), 4-(7-chloro-4-quinolyamino)-2- diéthylaminométhylphénol.

L'amodiaquine (AQ) est un médicament de type 4-aminoquinoléine, considéré une alternative meilleure que la CQ en spécial en Afrique.[6] Certaines limitations dans l'utilisation de ce médicament ont été imposées en raison de la résistance croisée potentielle avec la CQ et aux effets toxiques associés principalement au traitement prophylactique ou aux supra-dosages[.7,8]

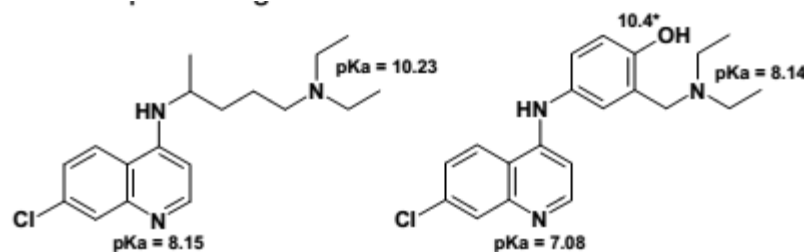


Figure I.7. Constantes d'acidité (pKa) de la chloroquine et de l'amodiaquine (\* valeur calculée)

### I.2. 3.2. Effets indésirables et toxicité de l'amodiaquine

L'amodiaquine a été délaissée à cause de sa grande toxicité quand elle est utilisée en prophylaxie. Cependant, l'amodiaquine semble être un bon médicament pour traiter le paludisme en Afrique. Elle est efficace sur les parasites présentant une résistance modérée à la chloroquine. Elle est moins chère et n'a pas le goût amer de la chloroquine, un avantage pour son utilisation pédiatrique [9, 10]. Utilisée en prophylaxie pendant de nombreuses années, l'amodiaquine s'est montrée très toxique avec des cas mortels de toxicité hépatique et de la moelle osseuse [11].

L'agranulocytose survient en général entre cinq et 14 semaines de prophylaxie et a été associée à une hépatite chez certains voyageurs. Des cas d'agranulocytoses et d'hépatites ont été signalés lors de l'administration prolongée d'amodiaquine en prophylaxie [12, 13]. Le risque de développer l'agranulocytose est estimé à 1 sur 2 000 – 2 200, avec un risque de mort de 1 cas sur 31 300, et 1 sur 15 650 pour des atteintes hépatiques sérieuses. L'amodiaquine peut aussi entraîner une neutropénie dans de rares occasions quand elle est utilisée pour le traitement des arthrites rhumatoïdes.

L'amodiaquine semble induire directement une toxicité hépatique à travers la production d'intermédiaires imino de quinine ou pourrait interagir avec des IgG anti-amodiaquines [14]. L'amodiaquine semble ne pas avoir la toxicité sérieuse cardiaque de la chloroquine [15,16]. Cependant, elle ne doit pas être recommandée chez la femme enceinte ou allaitante à l'absence de données suffisantes. A cause de l'insuffisance de données sur la sécurité de l'amodiaquine au cours de la grossesse, il existe un besoin urgent d'évaluer sa tolérance et sa toxicité au cours de la grossesse avant son éventuelle utilisation dans le traitement préventif intermittent du paludisme au cours de la grossesse [17]. Une récente étude au Ghana (900 femmes enceintes, 225 par bras) a démontré que malgré quelques effets secondaires (pas de toxicité hépatique ni d'anomalie du nombre des leucocytes), l'amodiaquine est efficace pour le traitement du paludisme au cours de la grossesse [18]. L'amodiaquine ne doit pas être utilisée en prophylaxie. Une réaction allergique antérieure est une contre-indication de l'amodiaquine.

Les principaux effets secondaires associés à l'amodiaquine sont l'hépatotoxicité et l'agranulocytose qui ont été essentiellement observées chez les personnes utilisant l'amodiaquine en prophylaxie pendant plusieurs années [9, 13]. Le métabolite associé à ces effets est l'amodiaquine quinone-imine (AQQI) qui est le résultat de la métabolisation oxydative de l'amodiaquine. Ce métabolite de l'amodiaquine lie de manière covalente les groupements thiol de GSH intracellulaire ou des protéines et peut ainsi soit augmenter le stress oxydatif, soit inactiver des protéines essentielles. Les effets indésirables sont similaires à ceux provoqués par la chloroquine chez les patients avec le paludisme. Les effets indésirables les plus fréquemment rencontrés sont la nausée, les vomissements, les douleurs abdominales, la diarrhée, les démangeaisons et souvent des bradycardies [9]. L'amodiaquine en utilisation prophylactique est hépatotoxique et peut induire une agranulocytose. Cette toxicité est due aux propriétés immunogénétique de la quinone-imine produite par auto-oxydation de la drogue mère. Du fait que l'amodiaquine se concentre dans le foie, elle devrait

être utilisée avec beaucoup de précautions chez les patients alcooliques et présentant une maladie hépatique d'où l'existence d'un risque d'aggravation d'une pathologie hépatique sous-jacente. Ainsi, l'amodiaquine est contre-indiquée en cas de maladies graves du foie. Cette même contre-indication est valable pour les patients sous traitement d'un ou plusieurs médicaments avec des effets hépatotoxiques. De même, l'amodiaquine est contre-indiquée chez les patients présentant un dysfonctionnement du tissu sanguin, en cas de risque d'aggravation d'une agranulocytose existante qui peut être fatale pour le patient. Elle est aussi contre-indiquée chez les personnes ayant des antécédents d'hypersensibilité aux amino-4-quinoléines, notamment à l'amodiaquine. La toxicité sévère est observée chez les sujets après un traitement à long cours (par exemple, cas de neuromyopathie, de rétinopathie) ou comme une réaction idiosyncratique (exemple, érythème multiforme, toxicité de la moelle osseuse). Ces réactions sont très rares et ont été observées chez les personnes soumises à une chimioprophylaxie au long cours [19,20]. Les effets toxiques des amino-4-quinoléines varient énormément et ceci d'un auteur à un autre. Les effets toxiques ici présentés sont issues de deux études faites au Sénégal et en Côte d'Ivoire pays d'Afrique de l'Ouest. Les résultats issus de l'étude faite à propos de 35 cas chez les enfants en Côte d'Ivoire sont présentés ci-dessous [21];

### I.2 .3.2.La chloroquine

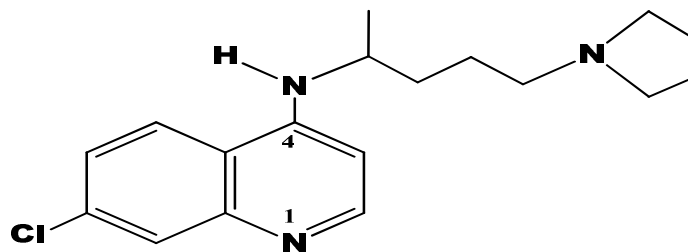


Figure I.8. Structure chimique de la chloroquine.

Les amino-4-quinoléines sont des dérivés de la chloro-7-quinoléine substitués en 4 par une chaîne de type dialkylamine. La chloroquine a pour formule chimique brute  $C_{18}H_{26}ClN_3$  [N-(7-chloroquinolin-4-yl)-N,N-diéthyl-pentane-1,4-diamine]. Les 5 atomes de carbones entre les deux atomes d'azotes de la chaîne latérale semblent conférer à la chloroquine sa grande activité antipaludique. Il existe dans la chaîne latérale un centre de chiralité qui permet l'obtention de deux énantiomères avec des activités différentes. Le produit utilisé est le mélange racémique des deux énantiomères.

**a) Indications**

La Chloroquine est utilisée en :

- **Parasitologie** : pour le traitement curatif et préventif du paludisme
- **Rhumatologie** : pour le **traitement** symptomatique d'action lente de la polyarthrite rhumatoïde ;
- **Dermatologie** : contre le lupus érythémateux discoïde, contre le lupus érythémateux subaigu, dans le traitement d'appoint ou prévention des rechutes du lupus systémiques, pour la prévention des lésions cutanées. Aucun cas de résistance de *P. ovale* et de *P. malariae* n'a été détecté avec la chloroquine, bien qu'une étude provenant de Sumatra en Indonésie portant sur plusieurs patients présentant une infection persistante de *P. malariae* pendant huit jours malgré les concentrations adéquates sanguines de chloroquine, suggère qu'une possible résistance de *P. malariae* pourrait se développer à la chloroquine [22].

**b) Contre-indications**

La chloroquine a peu de contre-indications qui sont surtout limitées aux patients allergiques ou ayant eu une dose exagérée à la chloroquine lors d'une utilisation antérieure. Cependant, la chloroquine est contre-indiquée en cas d'épilepsie, de maladies rénales et/ou hépatiques graves. L'OMS conseille également de ne pas utiliser la chloroquine en cas de psoriasis. Certains praticiens surtout dans les pays anglophones déconseillent de donner la chloroquine aux sujets présentant un déficit en glucose-6-phosphate déshydrogénase (déficit en G6PD), bien qu'il n'existe aucune évidence d'hémolyse provoquée par la chloroquine chez ces déficients en G6PD. La chloroquine inhibe aussi la libération et l'action du TNF- $\alpha$  [23]. La chloroquine doit être utilisée avec prudence en cas de maladies hépatiques ou dans de graves désordres gastro-intestinaux, neurologiques ou sanguins. Si de tels désordres surviennent au cours du traitement le médicament doit être arrêté. Pour les patients recevant un traitement à long terme, un contrôle ophtalmologique et neurologique est recommandé tous les trois à six mois.

**I.2 3.2.2. Effets indésirables et toxicité de la chloroquine**

La large diffusion de la résistance de *P. falciparum* à la chloroquine est hélas devenue un facteur limitant son emploi. En plus des problèmes de prurit plus volontiers rencontrés chez les sujets à peau noire, des troubles oculaires peuvent parfois apparaître après une très longue

utilisation. Ce prurit est un facteur limitant l'adhérence au traitement, qui débute 6 – 24 heures après absorption de la chloroquine et touche tout le corps, particulièrement les palmes des mains et des pieds, et peut souvent durer plusieurs jours jusqu'à l'élimination complète de la molécule de l'organisme. Les antihistaminiques ont peu d'effets sur le prurit provoqué par la prise de chloroquine. Des symptômes mineurs tels que maux de tête, malaise, vision trouble, difficulté de se concentrer, douleurs épigastriques, peuvent survenir après une prise de la chloroquine [11]. Il est postulé que l'effet de l'ingestion chronique de la chloroquine aggravée par la toxicité chronique de la lumière pourrait être responsable de la dégénérescence maculaire observée chez les adultes [24]. La chloroquine a une marge thérapeutique très faible et ses effets toxiques sont très liés à la dose ingérée. Une dose unique de 20 mg/kg est considérée comme toxique et des doses aussi faibles que 30 mg/kg ont été toujours fatales [16, 19]. Une dose de chloroquine de 5 g est un facteur prédictif fatal chez un adulte. Le diazépam est un antidote spécifique de la chloroquine. La chloroquine administrée par voie parentale peut provoquer une hypotension létale. Elle ne doit jamais être administrée par injection intraveineuse. La chloroquine est très bien tolérée chez la femme enceinte en traitement curatif ou préventif. Elle peut être utilisée pendant les trois trimestres que dure la grossesse. La chloroquine passe dans le lait maternel en petite quantité avec un temps de demi-vie de 8,8 jours [25]. La dose qu'un enfant allaité reçoit au sein est estimée à 0,7% de la dose quotidienne maternelle, et 4,2% pendant 9 jours [26, 27]. La chloroquine peut être utilisée sans conséquence par la femme qui allaite. Bien que la rétinopathie de la chloroquine soit bien connue, les effets secondaires cardiaques sont rarement signalés, mais dans certains cas peuvent être graves et irréversibles. Ainsi deux cas de transplantation d'organes ont été décrits dans la littérature [28]. Ils comprennent des troubles de la conduction (bloc de branche, bloc auriculo-ventriculaire) et la cardiomyopathie - souvent avec l'hypertrophie, la physiologie restrictive et une insuffisance cardiaque congestive [29].



**RÉFÉRENCES**

- [1]. Bruce-Chwatt L.J. Epidemiology of malaria. IN “Essential malarialogy”. Second Edition: William Heinemann Medical Books, London. 1985.
- [2]. OMS.Grands lignes du plan d’action de l’OMS pour la lutte contre le paludisme. 1993 - 2000. Conférence ministérielle sur le paludisme.Amsterdam. 27 octobre 1992.
- [3]. Malvy D, Djossou F, Thiebaut R, Le Bras M. Plasmodies-Malaria.Formes cliniques, diagnostic. Encycl. Méd Chir (Editions Scientifiques et Médicales Elsevier), Maladies Infectieuses, 2000 ; 8-507-A-20, 16 p.
- [4]. **Archibald HM.** Malaria in South-western and North-wester Nigeria communities. Bull WHO, 1956; 15:695-709.
- [6]. Olliaro, P.; Nevill, C.; LeBras, J.; Ringwald, P.; Mussano, P.; Garner, P.; Brasseur, P. *Lancet.* **1996**, 348, 1196- 1201.
- [7]. Phillips-Howard, P.A.; West, L.J. *J. Royal Soc. Med.* **1990**, 83, 82-85.
- [8]. Staedke, S.G.; Dorsey, G.; Rosenthal, P.J. *Lancet.* **2003**, 361, 1229-1230
- 9]. Olliaro P, Nevill C, LeBras J, Ringwald P, Mussano P, Garner P, et al. Systematic review of amodiaquine treatment in uncomplicated malaria. *Lancet* 1996; 348(9036):1196-201.
- [10]. Brasseur P, Guiguemde R, Diallo S, Guiyedi V, Kombila M, Ringwald P, et al. Amodiaquine remains effective for treating uncomplicated malaria in west and central Africa. *Trans R Soc Trop Med Hyg* 1999; 93(6):645-50.
- [11]. Taylor WR, White NJ. Antimalarial drug toxicity: a review. *Drug Saf.* 2004;27(1):25-61.
- [12]. O'Neill PM, Bray PG, Hawley SR, Ward SA, Park BK. 4-Aminoquinolines--past, present, and future: a chemical perspective. *Pharmacol Ther* 1998; 77(1):29-58.
- [13]. Olliaro P, Mussano P. Amodiaquine for treating malaria. *Cochrane Database Syst Rev* 2003(2):CD000016.
- [14]. Gil JP. Amodiaquine pharmacogenetics. *Pharmacogenomics* 2008; 9(10):1385-90.
- [15]. Ngouesse B, Basco LK, Ringwald P, Keundjian A, Blackett KN. Cardiac effects of amodiaquine and sulfadoxine-pyrimethamine in malaria-infected African patients. *Am J Trop Med Hyg* 2001; 65(6):711-6.
- [16]. White NJ. Cardiotoxicity of antimalarial drugs. *Lancet Infect Dis* 2007; 7(8):549- 58.
- [17]. Thomas F, Erhart A, D'Alessandro U. Can amodiaquine be used safely during pregnancy? *Lancet Infect Dis* 2004; 4(4):235-9.

- [18]. Tagbor H, Bruce J, Browne E, Randal A, Greenwood B, Chandramohan D. Efficacy, safety, and tolerability of amodiaquine plus sulphadoxine-pyrimethamine used alone or in combination for malaria treatment in pregnancy: a randomised trial. *Lancet* 2006; 368(9544):1349-56.
- [19]. White NJ, Miller KD, Churchill FC, Berry C, Brown J, Williams SB, et al. Chloroquine treatment of severe malaria in children. Pharmacokinetics, toxicity, and new dosage recommendations. *N Engl J Med* 1988; 319(23):1493-500.
- [20]. Taylor WR, White NJ. Antimalarial drug toxicity: a review. *Drug Saf* 2004; 27(1):25-61. al. Chloroquine-resistant *Plasmodium malariae* in south Sumatra, Indonesia. *Lancet* 2002; 360(9326):58-60
- [23]. Picot S, Peyron F, Vuillez JP, Polack B, Ambroise-Thomas P. Chloroquine inhibits tumor necrosis factor production by human macrophages in vitro. *J Infect Dis* 1991; 164(4):830.
- [24]. Obikili AG. A type of macular degeneration in adult Nigerians; a possible role of chloroquine. *East Afr Med J* 1990; 67(9):614-21.
- [25]. Law I, Ilett KF, Hackett LP, Page-Sharp M, Baiwog F, Gomorrai S, et al. Transfer of chloroquine and desethylchloroquine across the placenta and into milk in Melanesian mothers. *Br J Clin Pharmacol* 2008; 65(5):674-9.
- [26]. Osadchy A, Ratnapalan T, Koren G. Ocular toxicity in children exposed in utero to antimalarial drugs: review of the literature. *J Rheumatol* 2011; 38(12):2504-8.
- [27]. Klinger G, Morad Y, Westall CA, Laskin C, Spitzer KA, Koren G, et al. Ocular toxicity and antenatal exposure to chloroquine or hydroxychloroquine for rheumatic diseases. *Lancet* 2001; 358(9284):813-4.
- [28]. Tonnesmann E, Kandolf R, Lewalter T. Chloroquine cardiomyopathy - a review of the literature. *Immunopharmacol Immunotoxicol* 2013; 35(3):434-42
- [29]. Tonnesmann E, Stroehmann I, Kandolf R, Wolburg H, Strach K, Musshoff F, et al. Cardiomyopathy caused by longterm treatment with chloroquine: a rare disease, or a rare diagnosis? *J Rheumatol* 2012; 39(5):1099-103



# *Chapitre II*

*Bases théoriques des méthodes  
de calculs utilisée*

## II.1.Introduction.

Le modèle QSAR (quantitative structure-activité Relationship) étudie la corrélation basée sur un calcul statistique, de paramètres physico-chimiques avec l'activité biologique [1]. La comparaison de l'activité biologique de certaines molécules et de leur structure a permis d'établir dans de nombreux cas des corrélations entre les paramètres structuraux et les propriétés d'une molécule. La variation de l'activité aux paramètres a permis d'obtenir un système d'équations qui donne, pour une série chimique donnée et pour une activité définie, une équation de corrélation.

L'intérêt essentiel de cette équation est qu'elle doit permettre de déterminer la valeur des paramètres qui correspondent à une activité maximale et ainsi de prédire l'activité des molécules qui n'ont pas encore été synthétisées [2].

La validité d'un modèle QSAR dépendra donc des choix que l'on aura faits sur les paramètres. Il est ainsi judicieux d'identifier et d'évaluer quantitativement les sélectionnées.

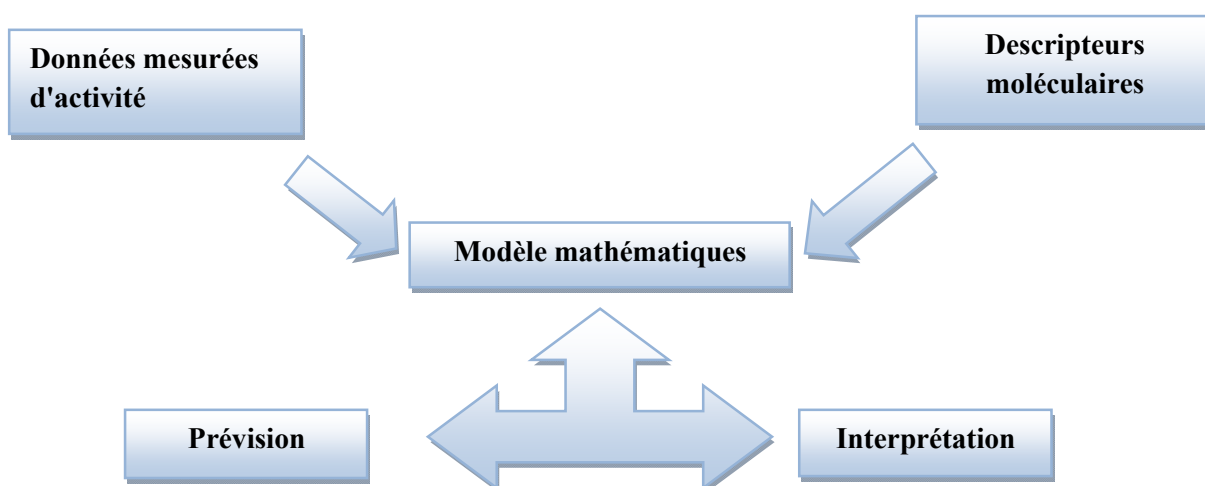


Figure. II. 1 : modèle d'étude quantitative des relations structure/Activité(QSAR)

## II.2.Histoire Des QSAR :

Les premiers essais de modélisation d'activités de molécules datent de la fin du 19<sup>ème</sup> siècle, lorsque Crum-Brown et Frazer [3] postulèrent que l'activité biologique d'une molécule est une fonction de sa constitution chimique  $c$

$$\phi = f(c) \quad (\text{eq.01})$$

Richet [3] a découvert que la toxicité des composés organiques suit inversement leur solubilité dans l'eau. Un tel rapport correspond à l'éq.01, où les  $\Delta\phi$  représentent des

différences entre les valeurs des activités biologiques, que sont causes par les changements de ces propriétés chimiques et particulièrement les propriétés physico-chimiques  $\Delta c$

$$\Delta\phi=f(\Delta c) \quad (\text{eq.02})$$

Il n'y a aucune méthode qui applique l'eq. 01 pour traiter les données biologiques. Toutes les équations QSAR correspondent à l'eq. 02, parce que les différences dans l'activité biologique sont seulement quantitativement corrélées avec des changements de lipophilie et/ou d'autres propriétés physico-chimiques des composés.

On peut considérer l'année 1964 comme l'année de naissance de la méthodologie QSAR moderne. Deux articles ont été publiés, un par Hansch et Fujita intitulé "method for the correlation of biological activity and chemical structure" [4], l'autre par Free et Wilson portant pour titre "A mathematical contribution to structure activity studies" [5]. Les deux contributions ont commencé par élaborer deux nouvelles méthodes pour quantifier la relation Activité Biologique / Structure (QSAR) appelée "Hansch analysis" (linear free energy-related approach, extrathermodynamic approach) et Free "Wilson analysis", respectivement.

L'approche QSAR résulte de la combinaison de différents paramètres physicochimiques de façon linéaire additive (eq. 03 ;  $\log(1/C)$  est le logarithme de l'inverse de la dose molaire qui produit ou empêche une certaine réponse biologique,  $\log P$  est le logarithme du coefficient de partage de n-octanol/ eau). D'autres méthodes utilisent un paramètre connu sous le nom de paramètre de lipophilie calculée  $\pi$  (eq.03) il est employé au lieu des valeurs mesurées de  $\log P$  (comme les valeurs  $\sigma$  de Hammett sont employées au lieu des constantes d'équilibre des réactions organiques), et la formulation d'une équation parabolique pour la description quantitative non-linéaire des rapports lipophilie-activité (eq. 04) [6, 7].

$$\log 1/C = a \log P + b\sigma + \dots + \text{const.} \quad (\text{eq.03})$$

$$\pi_x = \log P_R - X - \log P_R - H \quad (\text{eq. 04})$$

$$\log 1/C = a(\log P)^2 + b \log P + c \sigma + \dots + \text{const} \quad (\text{eq.05})$$

D'après la contribution signifiée par Fujita et Ban [8], le modèle libre de Wilson est défini par l'eq. 06, où  $a_{ij}$  est la contribution du groupe substituant **XI** en position  $j$ ,  $\mu$  est la valeur de l'activité biologique (théorique) d'un composé référence dans la série ; toutes les contributions de groupe  $a_{ij}$  des différents substituants **XI** se rapportent aux substituants correspondants (le plus souvent l'hydrogène) pour ce composé de référence.

$$\text{Log } 1/c = \sum a_{ij} + \mu \quad (\text{eq.06})$$

### II.3. Principe des méthodes QSPR/QSAR

Le principe des méthodes QSAR est, comme leur nom l'indique, de mettre, en place une relation mathématique reliant de manière quantitative des propriétés moléculaires aussi bien électroniques que géométriques, appelées descripteurs, avec une observable macroscopique (activité biologique, toxicité, propriété physico-chimique,.....ect), pour une série de composés chimiques similaires à l'aide de méthodes d'analyses de données. Aussi la forme générale d'un tel modèle est la suivante :

$$\text{Propriété} = f(\text{Descripteurs})$$

Donc le principe de ces études est d'établir une corrélation entre des données structurales de la molécule et leur activité biologique.

Les technique QSAR s'appuient sur le concept postulant que des structures similaires ont des propriétés similaires [9], plus les molécules sont différentes, plus il est difficile à corréler les propriétés physico-chimiques et l'activité biologique, alors que le contraire est plus aisé. Ce type d'étude permet d'une part, d'expliquer les paramètres moléculaires impliqués dans l'activité biologique, une propriété physico-chimique ou l'élution sur une phase stationnaire et de prévoir d'autre part, l'influence de certaines modifications structurales dans l'activité biologique, une propriété physicochimique ou la relation d'un composé sur une colonne.

### II .4. Méthodologie générale d'une étude QSPR/QSAR

La méthodologie générale d'une étude QSAR/QSPR est la suivante :

1. Collecte d'une base de données.
2. Recherche de descripteurs adéquats pour l'activité étudiée
3. Le choix d'une méthode d'analyse des données
4. Validation du modèle.

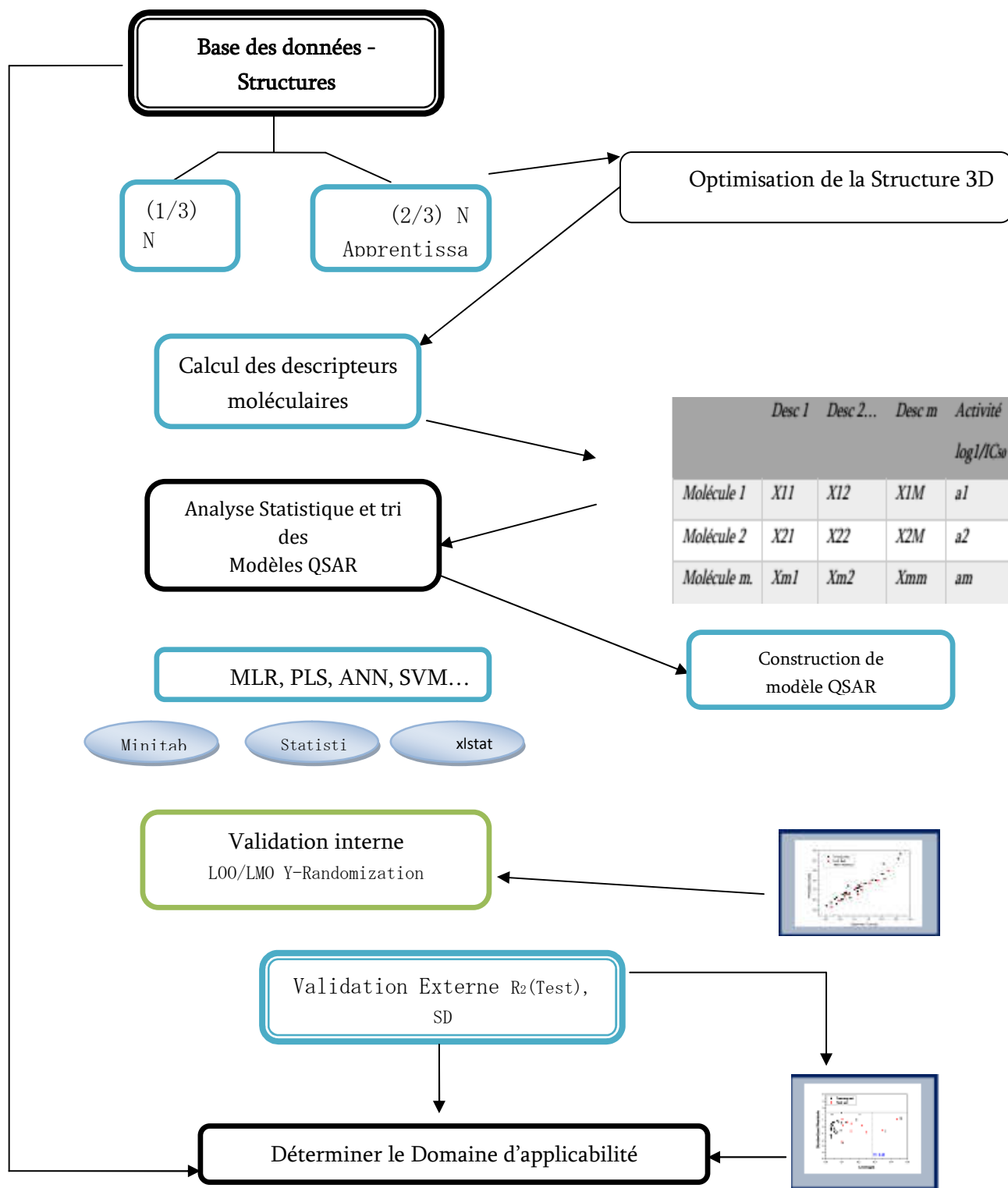


Figure .II .2. : Procédure d'obtention et de validation d'un modèle QSAR

## II .5.Les descripteurs physico-chimiques

Les descripteurs physicochimiques, (ou indices physicochimiques) certains d'entre eux reflètent la composition moléculaire du composé (le nombre et le type d'atomes et de liaisons présents dans la molécule, le nombre de cycle, les propriétés donneur/accepteur de liaison H, cation, anion, etc....) [10]. D'autres représentent le caractère hydrophile ou lipophile de la molécule généralement évalué à partir du coefficient de partage Octanol/eau représenté par le log P [11]. Parmi ceux que nous avons utilisé dans nos travaux, on trouve :

### II 5.1.Le coefficient de partage Octanol/Eau :(logp)

Le transport, le passage à travers les membranes et l'activité pharmacologique d'une molécule peuvent être conditionnés par son partage entre une phase lipidique et une phase aqueuse, c'est-à-dire son caractère hydrophile. Celui-ci peut être quantifié par le coefficient de partage Octanol-Eau, noté (log P), qui mesure la solubilité différentielle d'un soluté dans ces deux solvants non miscibles [12]. C'est une mesure importante pour l'identification de la similarité médicamenteuse, selon la règle de Lipinski, les médicaments délivrés par voie orale doivent avoir des valeurs de log P supérieures ou égales à -2 et inférieures ou égales à 5) [13].

Il est défini par la formule suivante :

$$\log P = \log \frac{[\text{Octanol}]}{[\text{H}_2\text{O}]}$$

[Octanol] et [H<sub>2</sub>O] sont les concentrations du soluté dans l'Octanol et l'eau.

Les composés qui ont les valeurs de log P > 0 sont dites lipophiles, et les composés qui ont les valeurs de log P < 0 sont dites hydrophiles. Si le Log P est positif et très élevé, cela exprime le fait que la molécule est plus soluble dans l'Octanol que dans l'eau, ce qui reflète son caractère lipophile, et inversement, si le Log P est négatif cela signifie que la molécule est hydrophile. Un Log P nul signifie que la molécule est aussi soluble dans un solvant que dans l'autre.

### II .5.2.La réfractivité moléculaire :(Ref)

Notée (MR), en m<sup>3</sup>/mol, est le volume de la substance absorbée par mole de cette substance. Elle est définie par Lorentz-Lorenz [14] par la formule suivante :

$$MR = \frac{n^2 - 1}{n^2 + 2} \frac{MW}{d} = \frac{n^2 - 1}{n^2 + 2} MV$$

Où :

**MW**: est le poids moléculaire ;

**d** : est la densité ; n est l'indice de réfraction ;



**MV**: est le volume molaire.

La réfractivité moléculaire est également proportionnelle à la polarisabilité  $\alpha_e$ , par la relation suivante [15] :

$$MR = 4/3\pi NA \alpha_e$$

Où : NA est le nombre d'Avogadro qui est, le nombre de molécules dans une mole de substance,

$$NA = 6.022 \cdot 10^{23}.$$

### II.5.3. La polarisabilité, (pol):

Notée ( $\alpha_e$ ), en (m<sup>3</sup>), est l'aptitude à la déformation du nuage électronique de la molécule sous l'influence d'un champ électrique uniforme. C'est l'un des paramètres qui traduisent les propriétés moléculaires liées à l'hydrophobie et par conséquent aux activités biologiques [16-17].

Elle est calculée à partir de la réfractivité molaire ou du volume molaire comme suit :

$$\text{pol} = 0.3964308 \times MR = 0.3964308 \times n^2 - 1 n^2 + 2 MV$$

### II .5.4.Surface moléculaire, (SAG):

C'est une enveloppe entourant les atomes localisés à la périphérie, qui explique la surface de contact Ligand-Récepteur c'est-à-dire l'affinité de liaison L-R [18,19].

L'HyperChem permet de calculer la surface grille (SAG), c'est une méthode de grille où une méthode plus rapide plus approximative; quelque soit la zone accessible au solvant ou la surface de van der Waals [20].

Dans cette théorie, chaque atome de la molécule est représenté par une sphère. La surface extérieure de toutes les sphères atomiques définit la surface de Van der Waals [21].

### II.5.5. Volume moléculaire : (V)

C'est un volume occupé par un corps, c'est un volume spécifique (à température et pression normales). Son calcul est très similaire à celui de la Surface (Grid), il emploie une méthode de grille décrite par Bodor et al.

Le volume est défini par la relation suivante [22] :

$$V = \frac{W}{d}$$

Où: **W**: est la masse moléculaire.

**d**: est la densité.

L'hydratation est la formation d'une solution implique l'interaction du soluté avec des molécules de solvant, différents liquides peuvent être utilisés comme solvants, mais l'eau est le solvant le plus couramment utilisé. Lorsque l'eau est utilisée comme solvant, le processus de dissolution est appelé hydratation [23].

Dans les milieux, les molécules polaires s'entourent pas des molécules d'eau, ce qui fait apparaître des liaisons hydrogène entre eux; évidemment, les sites donneurs de proton interagissent avec l'atome d'oxygène de l'eau et les sites accepteurs de proton avec l'atome d'hydrogène [24].

L'énergie d'hydratation est un facteur déterminant la stabilité des différentes conformations moléculaires dans les solutions aqueuses [25].

Le calcul de l'énergie d'hydratation est basé sur la surface exposée qui dépend du type d'atome des groupements moléculaires qui peuvent être donneurs des liaisons hydrogène tels que : O-H, N-H, P-H... ou bien des groupements accepteurs qui portent des doublets libres tels que : O, N, S, P...

## II.6. Interprétation et validation d'un modèle QSAR/QSPR

Une fois développé, le modèle doit être interprété en analysant tous les paramètres statistiques de ce modèle, sa qualité doit être aussi étudiée, cette qualité est vérifiée par ce que l'on appelle validation. Sa robustesse, c'est-à-dire l'influence des composés de la série d'apprentissage sur le modèle, est estimée par des méthodes de validation interne. Afin d'estimer son pouvoir prédictif, des données expérimentales supplémentaires sont nécessaires afin de déterminer la capacité du modèle à prédire ces valeurs c'est ce que l'on appelle validation externe. Enfin, il est important de savoir quel type de molécules utilisées avec quel modèle. On parle alors de domaine d'applicabilité.

### II.6.1. Validation interne

La validation interne d'un modèle QSAR/QSPR a été réalisée en utilisant la validation croisée LOO (Leave-One-Out) ou LMO (Leave-Many-Out) qui est quantifiée par le coefficient  $R^2_{cv}$ . Ce processus consiste à extraire un certain nombre  $k$  de molécules du jeu initial à  $N$  molécules et à construire un nouveau modèle avec les  $(N-k)$  molécules restantes à l'aide des descripteurs choisis (seules les constantes de la régression changent). Ce processus est ensuite réitéré pour retirer et prédire les valeurs de toutes les molécules de la série d'apprentissage. En fonction du nombre de molécules retirées à chaque itération, on parlera de Leave-One-Out (LOO) ou de Leave-Many-Out (LMO) selon qu'une ou plusieurs molécules est (sont) retirée(s) [26].

Cependant, la validation interne est insuffisante pour étudier le pouvoir prédictif d'un modèle. Pour cette raison la validation externe du modèle est devenue une norme et une partie obligatoire dans la modélisation QSAR/QSPR [27,28].

## II .6.2. Coefficients et tests statistiques standards

Afin de déterminer la qualité d'un modèle, différents paramètres statistiques sont employés, tels que les erreurs quadratiques moyennes (*Mean Square Errors*), les coefficients de corrélation qui sont régulièrement utilisés dans les études QSAR/QSPR, sont décrits en détail dans cette partie.

### II .6.2.1. Coefficient de corrélation $r$ (et coefficient de détermination $r^2$ )

C'est l'indicateur statistique le plus répandu est le coefficient de corrélation qui évalue la part de la variance de l'activité / la propriété cible expliquée par le modèle.

$$r = \sqrt{1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}}$$

Avec :

$r$ : est le coefficient de corrélation ;

$y_i$  et  $\hat{y}_i$ : sont, respectivement, les valeurs observées et calculées de la variable dépendante ;

$\bar{y}$ : est la valeur moyenne des valeurs observées.

Ces coefficients ne sont pas affectés par l'unité de mesure choisie et traduisent une bonne corrélation entre l'activité cible et l'activité initiale si  $r^2$  est proche de 1 (cas idéal).

Le jugement sur la valeur de  $r$  ou  $r^2$  est très subjectif. Bien que ce coefficient soit très facile à comprendre, il faut se garder d'y attacher trop d'importance car il est loin de fournir un critère suffisant pour juger la qualité d'une régression. Il n'est pas recommandé d'utiliser  $r^2$  pour comparer des modèles avec un nombre différent de descripteurs, le coefficient  $r^2$  nous dira toujours de choisir le modèle avec le plus grand nombre de descripteurs car son  $r^2$  sera plus important (on projette sur un espace plus grand), même si les variables sont sans effets sur la réponse (l'activité ou la propriété étudiée).

La valeur de  $r^2$  dépend de la taille de l'échantillon et le nombre de variables prédictives dans l'équation. Il garde la même valeur ou augmente lors d'une nouvelle variable de prédiction est ajoutée à l'équation de régression, même si la variable ajoutée ne contribue pas à la réduction de la variance inexpliquée. Par conséquent, un autre paramètre statistique peut être utilisé, appelé  $r^2$  ajusté ( $r^2_{adj}$ ). Bien entendu, un autre indicateur est l'erreur quadratique moyenne (*MSE*, pour *Mean Square Error*), à laquelle est parfois préférée la déviation standard  $s$ .

### II .6.2.2. Le coefficient de détermination ajusté $r^2_{adj}$

Ce coefficient est utilisé en régression multiple par ce qu'il tient compte du degré de liberté :

$$r^2_{adj} = \frac{r^2(n - 1) - p}{n - p - 1}$$

Avec :

**n** est le nombre des observations (les molécules) ;

**p** est le nombre de variables indépendantes (les descripteurs) ;

**r<sup>2</sup>** est le coefficient de détermination du modèle.

### II .6.2.3. L'erreur quadratique moyenne « MSE » et l'erreur type résiduel « s »

$$MSE = \frac{\sum (y_i - \hat{y}_i)^2}{n}$$

Ou encore, l'erreur type résiduel « s » :

$$s = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n - p - 1}}$$

Avec :

**Y<sub>i</sub>** et **ŷ<sub>i</sub>**: sont, respectivement, les valeurs observées et calculées de la variable dépendante ;

**n**: est le nombre des observations ;

**p**: est le nombre de variables indépendantes.

Ces paramètres mesurent la variation de l'activité cible non expliquée par le modèle RQSA/RQSP. En particulier, plus la déviation standard est petite et plus la corrélation est meilleure. Sa valeur est toujours fonction de l'unité de mesure de l'activité cible et tient également compte des erreurs expérimentales ce qui explique qu'une valeur trop petite n'ait aucune signification.

### II .6.2.5. Le test de Fisher F

L'indice de Fisher *F-test* est employé afin de mesurer le niveau de signifiante statistique du modèle à « x% » (le niveau usuel est 95%), c'est-à-dire la qualité du choix du jeu de paramètres. La conclusion obtenue ne doit pas nous faire penser que la corrélation a « x % » de chances d'être vraie mais seulement que la corrélation est vérifiée pour « x% » des composés pris pour référence et qu'une abstraction est faite pour les autres.

Hypothèses :

**H0**: les variances des échantillons sont homogènes

**H1**: les variances des échantillons ne sont pas homogènes

La valeur à calculer est :

On calcule le F (observé) à partir de la formule :

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

Avec :

**F**: est l'indice de Fisher ;

**y<sub>i</sub>** et **ŷ<sub>i</sub>** : sont, respectivement, les valeurs observées et calculées de la variable dépendante ;

**ȳ**: est la valeur moyenne des valeurs prédites ;

**n**: est le nombre des observations (les molécules) ;

**p**: est le nombre de variables indépendantes (les descripteurs).

Après le calcul de F (observé) on le compare avec le F théorique obtenu à partir des tables statistiques usuelles (la table de Fisher).

Si F observé est plus grand que le F théorique : refus de l'hypothèse nulle H0 et cela signifie que les variances des échantillons sont trop différentes pour être considérées comme homogènes.

Si F observé est plus petit que le F théorique : acceptation de l'hypothèse nulle H1 et cela signifie que les deux variances ont des valeurs suffisamment proches pour qu'on accepte l'idée qu'elles soient homogènes. [29]

### II. 6. 3. Validation externe.

Cette méthode consiste à prédire la propriété/activité d'une série de molécules appelée généralement série de test qui ne sont pas dans la série de développement du modèle, cette validation est caractérisée par les paramètres R<sup>2</sup>test . Récemment plusieurs études [30, 31] ont montré l'insuffisance des paramètres R<sup>2</sup>, R<sup>2</sup>CV pour vérifier la capacité prédictive des modèles QSAR/QSPR. Par conséquent, d'autres paramètres doivent être vérifiés pour cet objectif. Ces paramètres sont connus sous le nom « critères de validation externe » ou souvent appelés « critères de Trophsa » (Trophsa criteria) [30].

**II.7. Logiciel « XLSTAT »:** On choisit ce logiciel lorsque contient plusieurs caractéristiques :

- XLSTAT c'est un logiciel très fiable et très stable.
- nous permet de réorganiser nos données et de les structurer afin de faciliter leur analyse.
- Avec XLSTAT l'analyse en composantes principales (ACP) et l'analyse des correspondances (AFC) deviennent faciles, de même que la régression linéaire (simple ou multiple),
- nous donne accès à un grand nombre de tests. Les résultats des tests comprennent une interprétation explicite.

- Il offre une large variété de tests statistiques pour que nous permette de tester des hypothèses.
- nous permet d'évaluer la qualité de nos données, et détermine les erreurs de calcul (écarts + résidus), ce que nous aide d'améliorer notre travail.
- Les résultats des essais incluent une première conclusion de niveau pour soulager notre interprétation.

Il inclut quelques outils qui facilitent la visualisation des données.

## II.8. Méthodes statistiques

Faire de la statistique suppose que l'on étudie un ensemble d'objets équivalents sur lesquels on observe des caractéristiques appelées « variables ». Dans notre cas, les objets (ou individus) sont les molécules et les variables sont les descripteurs moléculaires précédemment décrits dans ce chapitre.

Après le recueil des descripteurs, la démarche statistique consiste à traiter et interpréter les informations recueillies sur ces molécules. Cette démarche comporte deux grandes classes : la statistique descriptive et la statistique décisionnelle ou prédictive.

### II.8.1. La statistique descriptive

La statistique descriptive (appelée aussi l'analyse des données), a pour but d'extraire le maximum de l'information contenue dans les données d'une façon efficace, simple et compréhensible. Elle permet de résumer les caractéristiques essentielles du phénomène étudié et de suggérer des hypothèses pour des études plus sophistiquées. Elle utilise pour cela des représentations de données sous forme de graphiques, de tableaux et d'indicateurs statistiques. Elle est utilisée aussi pour diviser et classer les données dans des classes homogènes.

Dans l'ensemble de nos travaux, nous avons principalement utilisé l'analyse en composantes principales (ACP) comme technique pour l'analyse des données, et la méthode du partitionnement en k-moyennes (ou k-means en anglais) et la classification ascendante hiérarchique (CAH) pour la classification des données.

#### ❖ L'analyse en composantes principales

L'analyse en composantes principales (ou ACP) [31-32], est une méthode très efficace d'analyse de données quantitatives utilisée pour réduire la dimension de l'espace de représentation des données. Les variables initiales sont remplacées par de nouvelles variables, appelées composantes principales, deux à deux non corrélées, et telles que les projections des données sur ces composantes soient de variance maximale.

Considérons un ensemble de  $M$  observations, représentées chacune par  $N$  données. Ces observations forment un nuage de  $M$  points dans  $\mathbb{R}^N$ . Le principe de l'ACP est d'obtenir une représentation approchée des variables dans un sous-espace de dimension  $K$  plus faible, par projection sur des axes bien choisis ; ces axes principaux sont ceux qui maximisent l'inertie du nuage projeté. La maximisation de l'inertie permet de préserver au mieux la répartition des points. Par conséquent, les  $N$  composantes principales peuvent être représentées dans l'espace sous-tendu par ces axes, par une projection orthogonale des  $N$  vecteurs d'observations sur les  $K$  axes principaux.

Ces composantes peuvent être classées par ordre d'importance. Puisqu'elles sont des combinaisons linéaires des variables initiales, l'interprétation du rôle de chacune de ces composantes reste possible. Il suffit en effet de déterminer quels descripteurs d'origine leur sont le plus fortement corrélés. Les variables obtenues peuvent ensuite être utilisées en tant que nouvelles variables du modèle.

L'analyse en composantes principales est généralement utilisée pour visualiser et analyser rapidement les corrélations entre les variables et pour visualiser et analyser les observations initialement décrites par les variables sur un graphique à deux ou trois dimensions, construit de manière à ce que la dispersion entre les données soit aussi bien préservée que possible...

Les limites de l'ACP viennent du fait que c'est une méthode de projection, et que la perte d'information induite par la projection peut entraîner des interprétations erronées.

### **II.8.2. La statistique décisionnelle ou prédictive**

Contrairement à la statistique descriptive, dans ce type de statistiques les probabilités jouent un rôle fondamental. Cette statistique a pour but de prendre des décisions et de faire des prévisions au vu des observations. En général, il faut pour cela proposer des modèles probabilistes du phénomène aléatoire étudié et savoir gérer les risques d'erreurs. Dans notre cas, il faut rechercher une relation approximative entre une activité ou propriété et plusieurs variables quantitatives (descripteurs moléculaires), la forme de cette relation peut être linéaire ou non linéaire.

Dans l'ensemble de nos travaux, nous avons utilisé la régression linéaire multiple MLR, la régression non linéaire multiple MNLR, la régression des moindres carrés partiels PLS et les réseaux de neurones artificiels ANN pour la construction des modèles QSAR/QSPR.

#### **a. La régression linéaire multiple**

La régression linéaire multiple MLR est l'une des méthodes de modélisation les plus populaires grâce à sa simplicité d'utilisation et facilité d'interprétation. L'avantage important

de la régression linéaire multiple est qu'elle est très transparente, puisque l'algorithme est disponible, et que les prédictions peuvent être réalisées facilement. Dans la plupart de nos travaux, cette méthode a été utilisée aussi pour la sélection des descripteurs moléculaires utilisés dans les autres méthodes statistiques

La méthode MLR se base sur l'hypothèse que la propriété  $y$  dépend linéairement des différentes variables (les descripteurs)  $x_1, x_2, \dots, x_i$ , selon la relation :

$$y = a_0 + \sum_{i=1}^n a_i x_i$$

Avec :

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

**X<sub>i</sub>**: sont les variables indépendantes (explicatives) ;

**n** :est le nombre de variables explicatives ;

**a<sub>0</sub>**: est la constante de l'équation du modèle ;

**a<sub>i</sub>**: sont les coefficients de descripteurs dans l'équation du modèle ;

La taille de ces coefficients indique le degré d'influence des descripteurs moléculaires correspondants sur l'activité/propriété cible. Un coefficient positif indique que le descripteur moléculaire correspondant contribue positivement à la cible, tandis qu'un coefficient négatif suggère la contribution négative.

On distingue divers types de MLR, les plus utilisés sont :

- La MLR progressive ascendante, qui consiste à incorporer les variables au modèle une à une, en sélectionnant, à chaque étape, la variable dont la corrélation partielle avec la grandeur modélisée est la plus élevée. À l'inverse, lors de MLR progressive descendante, on débute la modélisation avec l'ensemble des descripteurs, en les éliminant un par un jusqu'à obtenir le meilleur jeu de composantes, c'est-à-dire l'obtention d'un modèle valide (voir la partie validation) ayant la bonne corrélation.
- La MLR pas à pas (Stepwise), est une combinaison des deux méthodes évoquées précédemment. Les variables sont incorporées une à une dans le modèle, par sélection progressive. Cependant, à chaque étape, on vérifie que les corrélations partielles des variables précédemment introduites sont encore significatives.

### **b. La régression non linéaire multiple**

La régression non linéaire multiple MNLR est une méthode non linéaire (exponentielle, logarithmique, polynomiale, ...) qui permet de déterminer le modèle mathématique qui permet d'expliquer non-linéairement au mieux la variabilité d'une propriété ou d'une activité  $y$  en fonction des descripteurs moléculaires. Dans l'ensemble de nos travaux nous avons utilisé



le modèle polynomial en nous basant sur les descripteurs proposés par le modèle linéaire qui seront élevés à la puissance 2 selon l'équation suivante :

$$y = a_0 + \sum_{i=1}^n a_i x_i + b_i x_i^2$$

Avec :

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

**x<sub>i</sub>**: sont les variables indépendantes (explicatives) ;

**i**: est le nombre de variables explicatives ;

**a<sub>0</sub>**: est la constante de l'équation du modèle ;

**a<sub>i</sub>** et **b<sub>i</sub>** sont les coefficients de descripteurs dans l'équation du modèle ;

### c. La régression des moindres carrés partiels PLS

La régression des moindres carrés partiels PLS, est une généralisation de la régression linéaire multiple, elle peut être utilisée lorsque le nombre de descripteurs est élevé et que ceux-ci sont fortement corrélés [33, 34]. Cette méthode utilise à la fois des principes de l'ACP et de la régression multilinéaire. Elle permet de trouver par une transformation linéaire, les axes qui représentent au mieux les données dans l'espace. En d'autres termes, cette méthode va permettre de trouver les axes qui expliquent au mieux la dispersion du nuage de points. Si les données sont représentées en fonction de n descripteurs, la PLS va donc permettre de trouver au maximum n axes classés en fonction de la variance qu'ils représentent. Cette méthode consiste à remplacer une matrice des données prédictives X comprenant n lignes et m colonnes, par une nouvelle matrice, dérivée de X, qu'on désigne par T, comprenant le même nombre de lignes (molécules) que X, mais un nombre de colonnes k très inférieur à m. On impose, de plus, que les colonnes de la matrice T soient des combinaisons linéaires des variables d'origine. Sous forme matricielle, la relation s'écrit :

$$T = XW$$

Avec :

**W(m\*k)**: est la matrice des coefficients définissant les combinaisons linéaires ;

**T**: est la nouvelle matrice dont les colonnes forment des « variables artificielles », obtenues par combinaison linéaire des variables d'origine ;

Après cette transformation, la régression linéaire multiple est appliquée sur le tableau T à la place de X.

### Références

- [1] .N. Allinger, K. Chen, J. Lii, *J. Comp. Chem.*, 17, 642 (1996).
- [2]. G. Jones, B. Chapman, *Synthesis*, 475 (1995).
- [3]. A. Crum-Brown, T. Frazer, On the connection between chemical constitution and physiological action. Transactions of the Royal Society of Edinburgh, 25,151-203 (1868-69).
- [4]. M. C. Richet, *Compt. Rend. Soc. Biol. (Paris)* 45, 775-776 (1893)
- [5]. C. Hansch, T. Fujita, , *J. Am. Chem. Soc.*86, 1616-1626 (1964)
- [6]. Jr. Free, S. M, J. W. Wilson,. *Med. Chem.* 7, 395-399 (1964)
- [7]. C. Hansch, *Ace. Chem. Res.* 2, 232-239 (1969)
- [8].C. Hansch , J. M. Clayton, *J. Pharm. Sci.* 62, 1-21 (1973)
- [9].T.Fujita, T.Ban, , *J. Med. Chem.*44, 148-152 (1971)
- [10] .R. Bosque, J. Sales, E. Bosch, M. Rosès, M.C. Garcia-Alvarez-Coque, and J.R. Torres-Lapasio, “A QSPR study of the p-solute polarity parameter to estimate retention in HPLC”, *Journal of Chemical Information and Modeling*, 43, **2003**, 1240–1247.
- [11] .V.N. Viswanadhan, A.K. Ghose, G.R. Revankar and R.K. Robins, “Atomic physicochemical parameters for 3D structure directed quantitative structure-activity relationships: Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics”, *Journal of Chemical Information and Modeling*, 29, **1989**, 163–172.
- [12]. P.J. Taylor, “Hydrophobic Properties of Drugs, In Quantitative Drug Design”, *Pergamon Press, Oxford (UK)*, 4, **1990**, 241–294.
- [13] C.A. Lipinski, F. Lombardo, B.W. Dominy, and P.J. Feeney, “Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings”, *Advanced Drug Delivery Reviews*, 6(1–3), **1997**, 3–25.
- [14] H.A. Lorentz, “Über die Beziehungzwischen der Fortpflanzungsgeschwindigkeit des Lichtes der K orperdichte”, *Wiedemann's Annalen der Physik*, 9, **1880**, 641–665

- [15] C. Hansch, B.R. Telzer, and L.T. Zhang, “Comparative QSAR in toxicology: examples from teratology and cancer chemotherapy of aniline mustards”, *Critical Reviews in Toxicology*, 25, **1995**, 67–89.
- [16] A. Cammarata, “An Apparent Correlation between the in Vitro Activity of Chloramphenicol Analogs and Electronic Polarizability”, *Journal of Medicinal Chemistry*, 10, **1967**, 525–527.
- [18].Hugo Kubinyi, Gerd Folkers, Yvonne C. Martin, 3D QSAR in Drug Design, Springer, ( 1998).
- [19]. Abhilash Thakur. Mamta Thakur, Total QSAR study on Phenyl Acridine Derivatives in Reference to DNA Binding Affinity, 3rd International conference on Medical, Biological and Pharmaceutical Sciences, (2013).
- [20]. HyperChem help.
- [21].Ovidiu Ivanciuc, 3D QSAR MODELS, Department of Organic-chemistry, Faculty of Chemical Technology University "Politehnica" of Bucharest, Romania.
- [22].Y. Cohen, Pharmacologie moléculaire, Masson, (1978).
- [23].Site internet : [aterloo.ca/science](http://aterloo.ca/science).
- [24].S. Belaidi, A. Dibi, and M. Omari, A conformational exploration of dissymmetric macrolides antibiotics, *Turkish Journal of Chemistry*, vol. 26, pp. 491–500,( 2002).
- [25]. N. Melkemi and S. Belaidi, Structure-Property Relationships and QSAR Modeling of Detoxication Properties of some 1,2-Dithiole-3-thione Derivatives, *International journal of chemical research* vol 4, pp. 134-139( 2012).
- [26]. L. Zhang, H. Zhu, T.I. Oprea, A. Golbraikh, A. Tropsha, QSAR Modeling of the Blood-Brain Barrier Permeability for Diverse Organic Compounds, *Pharm. Res.* 25 (2008) 1902–1914.
- [27]. L. He, P.C. Jurs, Assessing the reliability of a QSAR model's predictions, *J. Mol. Graph. Model.* 23 (2005) 503-523.

- [28]. A. Tropsha, P. Gramatica, V.K. Gombar, The importance of being Earnest: Validation is the absolute essential for successful application and interpretation of QSPR models, *QSAR Comb. Sci.* 22 (2003) 69-77.
- [29]. P.P. Roy, S. Paul, I. Mitra, K. Roy, On Two Novel Parameters for Validation of Predictive QSAR models, *Molecules* 14 (2009) 1660-1701.
- [30]. Samir CHTITA, Modélisation de molécules organiques hétérocycliques biologiquement actives par des méthodes QSAR/QSPR. Recherche de nouveaux médicaments, THÈSE de Doctorat, Université de maroc .2016/2017.
- [31] I.T. Jolliffe, "Principal Component Analysis", New-York, NY: Springer, 2<sup>ème</sup> édition, **2002**.
- [32]. A. Morineau and T. Aluja-Banet, "Analyse en composantes principales Centre international de statistique et d'informatique appliquée", Saint-Mandé CISIA-CERESTA, 1998.
- [33]. T. Puzyn, J. Leszczynski and M.T. Cronin, "Recent Advances in QSAR Studies: Methods and Applications: Part I Theory of QSAR", *Challenges and Advances in Computational Chemistry and Physics*, 8, **2010**.
- [34]. R.D. Tobias, "An Introduction to Partial Least Squares Regression", Statistical Analysis System Institute Inc., Cary, USA, **2002**.



# *Chapitre III*

*Résultats et discussions*

## III .1. Introduction

Au cours de ces dernières années, l'importante évolution dans le domaine informatique et les progrès réalisés dans les méthodes de calcul de chimie quantique permettent de donner une bonne prédiction et une bonne description des propriétés d'une entité donnée, ces propriétés sont reliées à la nature et la forme de la molécule.

Dans les chapitres précédents, on a parlé sur la conception des méthodes prédictives et leur impotences dans le calcul de propriétés physico-chimiques, on a détaillé la méthodologie de développement de la méthode « **QSAR** » par l'utilisation des outilles comme la régression. Dans ce chapitre on va présenter quelques résultats de régression des données expérimentales obtenu par le logiciel XLSTAT. Effectués à chaque block des descripteurs utilisés dans ce travail. Les résultats sont présentés en deux formes:

- Résultats sous forme des tableaux.
- Résultats sous forme des figures.

Nous avons commencé à étudier la relation entre les propriétés physico-chimiques et l'activité chimique

### III.1.1. Méthodologie

#### III.1.1.1. Base de données

Une série de 30 dérivés de 7-substitué-4-aminoquinoline a été utilisée pour développer des modèles QSAR. Les valeurs de l'activité biologique sont décrites dans le travail de Martyna

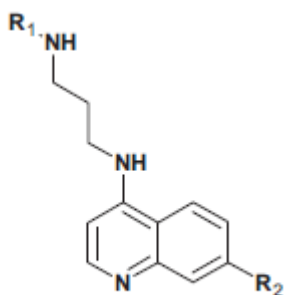
Les structures et l'activité biologique expérimentale des dérivés de 7- substitué-4-aminoquinolie sont rapportées dans le tableau III.1. L'activité antipaludique a été exprimée par PIC50.

#### III.1.1.2.Sélection des descripteurs et méthodes de calcul

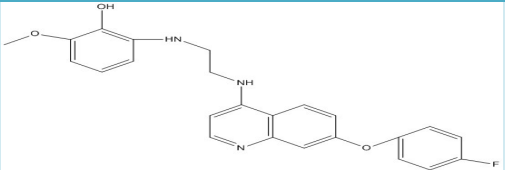
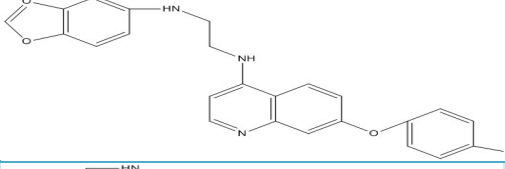
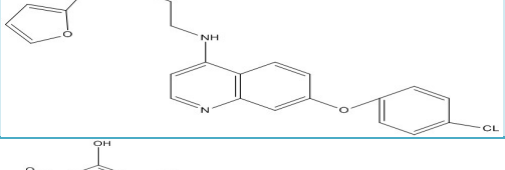
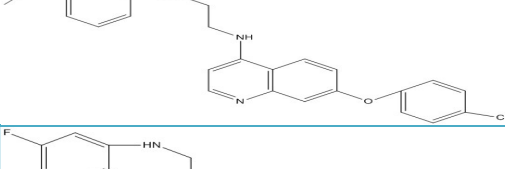
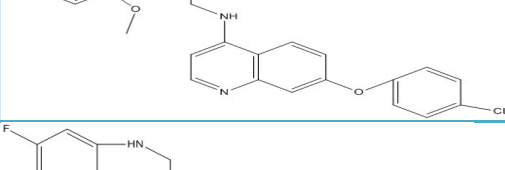
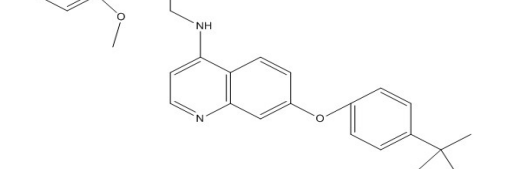
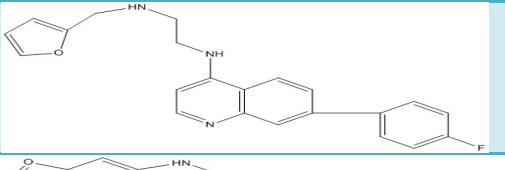
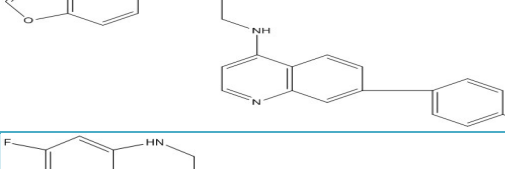
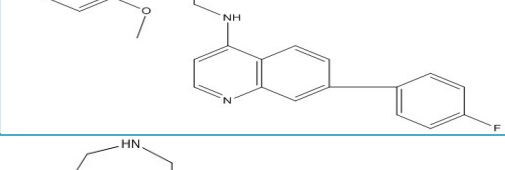
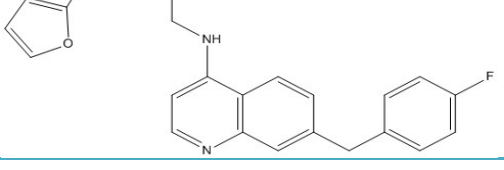
Le choix des descripteurs dépend des outils dont on dispose, de la nature des composés décrits et de la propriété ciblée. Les molécules de la série de 7-substitué-4-aminoquinoline ont été pré-optimisées au moyen de la mécanique moléculaire, champ de force (MM+) en utilisant le logiciel HyperChem 8.07[ 1].

### III.2. Structures chimiques des dérivés de 7-substitué -4-aminoquinolin

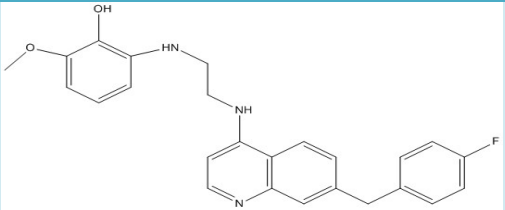
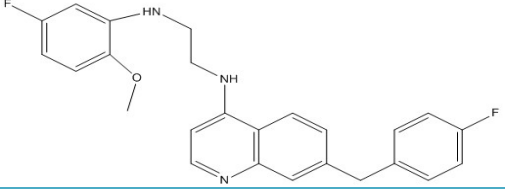
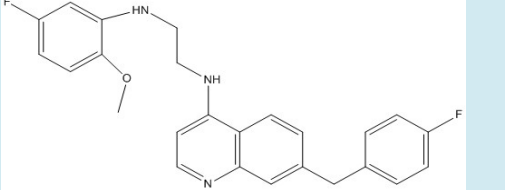
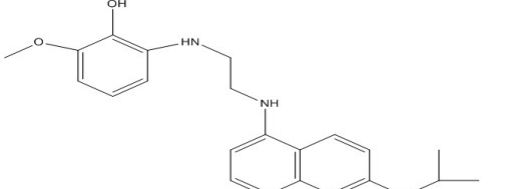
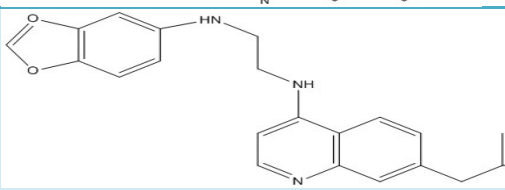
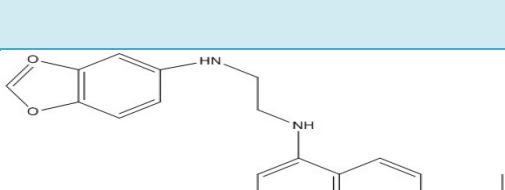
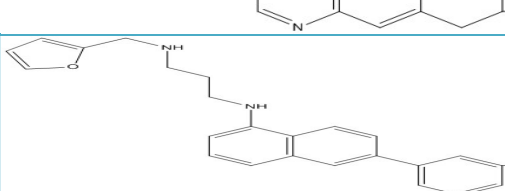
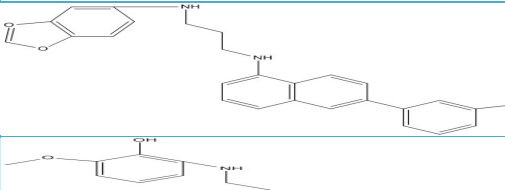
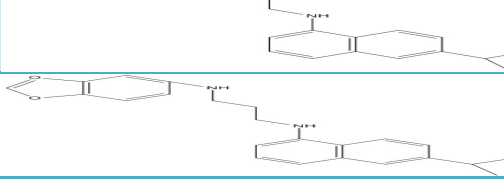
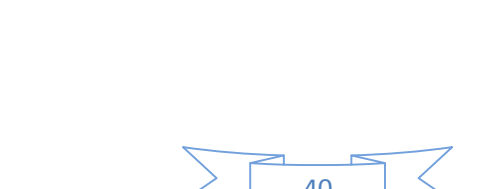
Tableau III.1 : Structures chimiques des dérivés 7-substitué -4-aminoquinolin

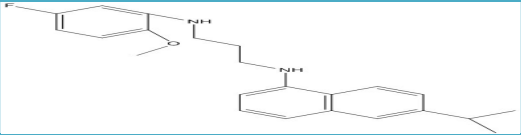
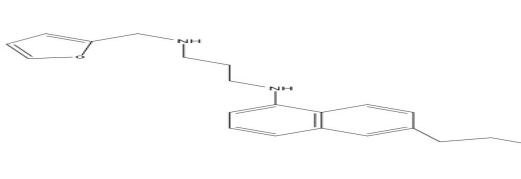


Composé	Structure	Pic50
1		7.8239
2		7.7447
3		8.0458
4		7.8539
5		7.6576
6		8.0458
7		7.7447
8		7.5850

9		7.9586
10		7.8239
11		7.3979
12		7.9586
13		7.509
14		6.975
15		7.495
16		7.475
17		7.244
18		7.244



19		7.824
20		7.745
21		7.125
22		7.658
23		7.569
24		7.569
25		7.337
26		7.886
27		7.699
28		7.181

29		7.051
30		7.066

### III.3..Étude des propriétés physico-chimiques des dérivé 7-substitué -4-aminoquinoline

. Les résultats de calculs des paramètres physicochimiques de cette série sont représentés dans le tableau III.2

**Tableau III.2 :** Paramètres physicochimiques des dérivés 7-substitué -4-aminoquinoline

Composé	Masse moléculaire (uma)	Polarisabilité $\epsilon$ ( $\text{\AA}^3$ )	Réfractivité molaire ( $\text{\AA}^3$ )	Volume moléculaire ( $\text{\AA}^3$ )	Surface moléculaire ( $\text{\AA}^2$ )	Énergie d'hydratation kcal/mol)	La lipophilie logP	Masse moléculaire (uma)
1	429.52	49.02	139.46	1219.34	678.17	-14.78	-1.45	429.52
2	433.55	48.82	133.85	1208.97	662.75	-8.77	0.38	433.55
3	459.54	51.49	145.84	1314.60	745.20	-15.24	-2.44	459.54
4	457.53	50.71	143.45	1307.81	750.13	-13.59	-2.22	457.53
5	461.54	50.76	144.36	1306.55	727.76	-8.93	-2.02	461.54
6	403.48	45.54	128.71	1138.08	615.87	-10.05	-2.15	403.48
7	459.54	51.49	145.84	1313.60	733.83	-15.69	-2.44	459.54
8	391.44	42.97	122.40	1100.43	629.63	-10.70	-1.76	391.44
9	445.41	48.15	137.20	1253.40	718.16	-13.50	-1.83	445.41
10	447.51	48.15	139.59	1275.39	742.44	-14.26	-2.05	447.51
11	407.90	44.99	127.05	1144.57	656.64	-9.23	-1.38	407.90
12	461.95	50.17	141.79	1221.01	665.09	-12.67	-1.45	461.95
13	465.95	50.22	142.70	1263.17	711.70	-8.13	-1.24	465.95
14	485.64	56.83	158.40	1400.7	757.93	-5.02	1.83	485.64
15	379.48	42.72	118.67	1068.07	598.13	-5.06	0.30	379.48
16	429.49	47.51	134.75	1230.55	723.97	-11.84	-0.72	429.49

17	433.50	47.56	135.67	1166.04	135.67	-5.10	-0.51	433.50
18	371.48	44.26	124.49	1110.01	635.34	-8.38	0.34	371.48
19	427.55	50.21	141.62	1268.64	721.22	-12.17	0.06	427.55
20	443.52	49.35	139.35	1237.67	701.25	-9.80	-0.32	443.52
21	447.53	49.39	140.39	1251.67	706.20	-5.36	-0.12	447.53
22	258.22	22.68	60.81	258.22	177.70	-0.03	0.65	258.22
23	431.52	48.38	137.02	1127.97	642.41	-10.70	-0.34	431.52
24	425.56	49.44	129.66	1190.35	677.49	-7.50	0.28	425.56
25	409.55	47.17	129.66	1203.04	662.63	-3.77	1.15	409.55
26	371.48	44.26	124.17	1125.99	644.90	-8.32	0.10	371.48
27	413.52	48.38	137.02	1169.33	651.06	-11.63	-0.34	413.52
28	393.53	47.97	126.53	1152.03	630.22	-8.34	0.33	393.53
29	391.51	45.28	124.14	1204.42	707.76	-9.06	0.55	391.51
30	429.52	49.02	139.46	1219.34	678.17	-14.78	-1.45	429.52

### III.2.1. Interprétation des résultats:

On remarque que l'ordre des valeurs de la polarisabilité est généralement proportionnel aux valeurs de la surface, du volume et de la réfractivité des dérivés 7-substitué - 4-aminoquinoline étudiés (Tableau III.2). Alors, ces résultats sont en accord avec la formule de Lorentz-Lorenz [2].

Cette relation montre que la réfractivité molaire et la polarisabilité augmentent avec le volume et la masse moléculaire. Par exemple, les composés 3.4.5. et 7.14. portent des substituant volumineux ont des valeurs importantes de la polarisabilité ( $51.49 \text{ \AA}^3$ ), ( $50.71 \text{ \AA}^3$ ), ( $50.76 \text{ \AA}^3$ ). et ( $51.49 \text{ \AA}^3$ ), ( $56.83 \text{ \AA}^3$ ) de la réfractivité molaire, ( $145.84 \text{ \AA}^3$ ), ( $143.45 \text{ \AA}^3$ ), ( $144.32 \text{ \AA}^3$ ). et ( $145.84 \text{ \AA}^3$ ), ( $158.40 \text{ \AA}^3$ ) respectivement. Contrairement, les composés 15 et 22 sont des petites molécules dans la série étudiée, 7-substitué - 4-aminoquinoline elles ont des petites valeurs des composés 15 et 22. respectivement de polarisabilité (42.71) et (22.68) de réfractivité molaire (118.6) et (60.82)

L'énergie d'hydratation en valeurs absolues, les plus importantes sont celle des composés 22 (-0.3 Kcal/mol) et les petites valeurs est celle de composé 3 (-15.24 Kcal/mol), 7 (-15.69 Kcal/mol)

En effet, dans les environnements biologiques les molécules polaires sont entourées par des molécules d'eau. Elles sont les liaisons hydrogène établies entre une molécule d'eau et ces molécules. Les emplacements de distributeur du proton agissent l'un sur l'autre avec l'atome d'oxygène de l'eau et les emplacements d'accepteur du proton agissent l'un sur l'autre avec l'atome d'hydrogène.

Les groupements hydrophobes dans les structures des dérivés 7-substitué - 4-aminoquinoline induit une diminution de l'énergie d'hydratation; toutefois, la présence des groupements hydrophiles comme dans les composés 22 qui possèdent 3 sites donneurs des protons (2 NH) et (OH), et sites accepteurs des protons (N) et (O) entraînent l'augmentation de l'énergie d'hydratation (**Tableau III.2**).

La lipophile est un facteur important dans les processus de solubilité, d'absorption, de distribution, de métabolisme et d'excrétion, ainsi que, l'activité pharmacologique. Hansch et Leo ont démontrés que les molécules très lipophiles se réparties et conservées à l'intérieur des couches lipidiques des membranes cellulaires [3]

Pour une bonne biodisponibilité orale, le log P doit être supérieur à zéro et inférieur à 3 ( $0 < \log P < 3$ ). En effet, Pour un log P trop élevé, le médicament présente une faible solubilité aqueuse. Dans le cas d'un log P très faible, le médicament présente une difficulté de se pénétrer dans les bicouches lipidiques des membranes cellulaires [4]. A l'opposé de l'énergie d'hydratation, la lipophilie augmente proportionnellement avec le caractère hydrophobe des groupes substituant.

A l'opposé de l'énergie d'hydratation, la lipophilie augmente proportionnellement avec le caractère hydrophobe des groupes substituant

Les composés possèdent des valeurs élevées de log P (1.83, 1.15 respectivement)

Ces composés possèdent généralement une bonne absorption intestinale, en raison d'un bon équilibre entre la solubilité et la perméabilité par diffusion passive; l'élimination métabolique est faible, en raison de faibles liaisons avec les enzymes métaboliques. En outre, ces molécules ont tendance à être libéré par les reins, en raison de leur polarité.

Les composés 3 et 4 présentent des faibles coefficients de partition -2.44 et -2.15, respectivement. Ces composés donnent une bonne solubilité, mais une faible absorption et pénétration, en raison de leur mauvaise perméabilité par la diffusion passive. Ces composés ont une tendance élevée d'éliminer par les reins [4].

### III.4. Modèles QSAR et analyse statistiques.

L'objectif de l'analyse quantitative de la relation structure-activité (QSAR) est de dériver des modèles empiriques qui relient l'activité biologique des composés à leurs structures chimiques

Dans cette analyse QSAR, des descripteurs quantitatifs sont utilisés pour décrire la structure chimique et les résultats d'analyse dans un modèle mathématique décrivant la relation entre la structure chimique et l'activité biologique. Pour expliquer la relation structure-activité, ces 8 descripteurs sont calculés pour 30 molécules

Les descripteurs quantitatifs des antipaludique substitués sont étudiés par des méthodes statistiques basées sur l'analyse en composantes principales (ACP) [5] en utilisant le logiciel XLSTAT version 2017 [6].

L'analyse par régression linéaire multiple (MLR) avec sélection descendante et élimination des variables est utilisée pour modéliser la relation structure-activité. C'est une technique mathématique qui minimise la différence entre les valeurs réelles et prédites. Il est également utilisé pour sélectionner les descripteurs utilisés comme paramètres d'entrée dans la régression non linéaire multiple (MNLR) et les pls.

pour prédire pIC50 Les équations sont justifiées par le coefficient de corrélation (R), l'erreur quadratique moyenne (ESM), l'indice de Fisher (F) . Pour la validation interne, le processus de validation croisée «leave-one-out» ( $R^2$  CV) a été utilisé pour évaluer la stabilité et la capacité interne des modèles proposés dans la présente étude.

#### III.4.1. Résultats et discussions

##### III 4.1.1. Analyse en composantes principales (acp)

L'analyse en composantes principales (ACP) est une méthode d'analyse multi-variée permettant l'étude simultanée d'un grand nombre de variables dont l'information totale ne peut pas être visualisée à cause d'un espace à plus de trois dimensions. Cette méthode permettrait de préciser les relations entre les variables et les phénomènes à l'origine de ces relations. L'objectif est d'avoir une information concentrée sur un minimum d'axe [7].

La matrice de corrélation donne une première idée des associations existantes entre les différentes variables telles que la polarisabilité et la réfractivité, et le masse, volume, et surface.

Ces paramètres sont relativement bien corrélés entre eux (Tableau 1).

L'histogramme des valeurs propres montre que le premier plan factoriel, constitué des axes F1 et F2, F3 représente 84.52% 30.6% et 13.5% de l'inertie totale. Il se révèle suffisant pour traduire l'essentiel de cette inertie. Car si l'on observe la répartition des valeurs propres, on constate que le décrochement principal se fait juste après les deux premières valeurs ce qui

signifie que les deux premiers axes sont suffisants pour représenter l'information dans sa globalité.

Les tableaux III.3 montre que les descripteurs obtenus sont corrélés entre eux et avec l'activité biologique.

les descripteurs logp et eh sont des paramètres les plus importants dans la corrélation entre les descripteurs sélectionnés et l'activité Antipaludique des dérivés de 7-substitué 4-aminoquinolin.

La totalité des 8 descripteurs (variables) codant les 30 molécules ont été soumise à une analyse en composantes principales. 8 composants principaux ont été obtenus (Figure.III.1). L'histogramme des valeurs propres montre que le premier plan factoriel, constitué des axes F1 et F2, contribuant respectivement 62.21% et 22.73% à la variance totale, l'information totale est estimée à un pourcentage de 84.94%. Il se révèle suffisant pour traduire l'essentiel de cette inertie. Car si l'on observe la répartition des valeurs propres, on constate que le décrochement principal se fait juste après les deux premières valeurs ce qui signifie que les deux premiers axes sont suffisants pour représenter l'information dans sa globalité.

Les coefficients de corrélation de Pearson sont résumés dans le tableau III.3. La matrice obtenue fournit des informations sur la corrélation négative ou positive entre les variables. L'analyse en composantes principales (ACP) a été réalisée pour identifier le lien entre les différentes variables. Les corrélations entre les 7 descripteurs sont présentées dans le tableau 3 comme une matrice de corrélation et sur la figure 3, ces descripteurs sont représentés dans un cercle de corrélation.

Afin de réduire le nombre de paramètres non significatifs et selon la matrice et le cercle de corrélations (**figur.III.2**) nous avons observé les corrélations suivantes :

- pol, ref et m sont fortement corrélés ( $r(\text{pol}, \text{MV}) = 0,943$ ,  $r(\text{ref}, \text{m}) = 0,962$ ).
- logp, eh et pic50 sont corrélés ( $r(\text{logp}, \text{pic50}) = 0,658$ ,  $r(\text{eh}, \text{logp}) = 0,630$ ).

Les tableaux. III.3: : Matrice de corrélation du modèle QSAR

Variables	m	logp	eh	S	v	ref	pol	pic50
<b>M</b>	<b>1</b>							
<b>Logp</b>	-0,378	<b>1</b>						
<b>Eh</b>	-0,500	0,699	<b>1</b>					
<b>S</b>	0,578	-0,225	-0,560	<b>1</b>				
<b>V</b>	0,881	-0,270	-0,533	0,725	<b>1</b>			
<b>Ref</b>	0,962	-0,311	-0,526	0,642	0,953	<b>1</b>		
<b>Pol</b>	0,943	-0,215	-0,479	0,656	0,956	0,985	<b>1</b>	
<b>pic50</b>	0,226	-0,658	-0,630	0,151	0,045	0,155	0,120	<b>1</b>

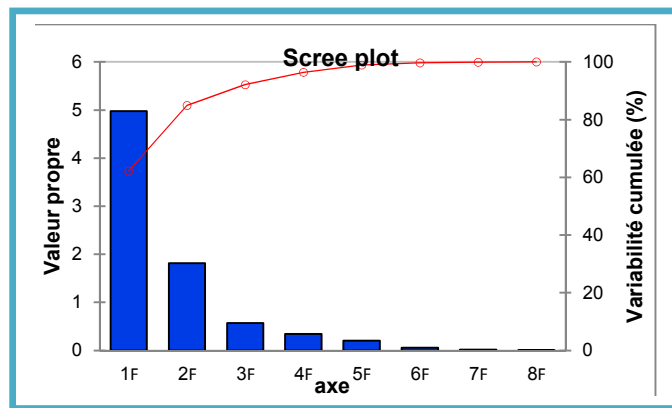


Figure .III.1: Les principaux composants et leurs variances

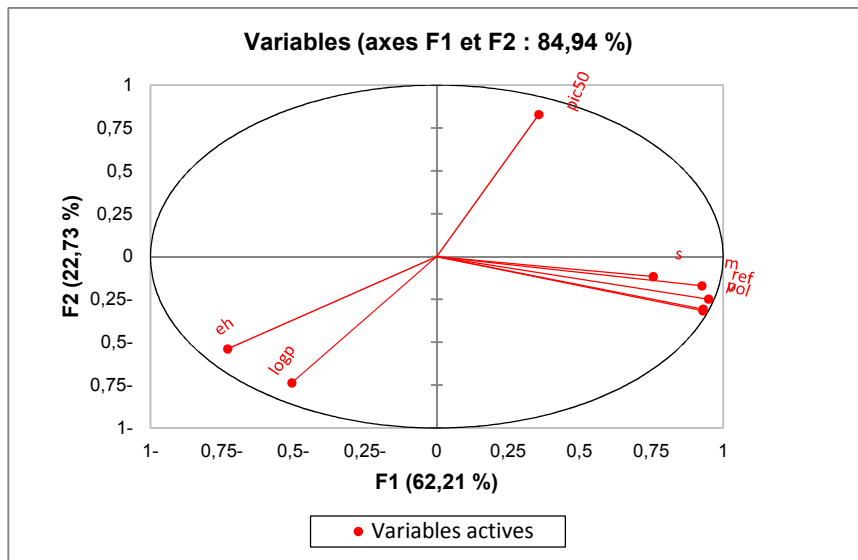


Figure. III .2 : Cercle de corrélation

### III. 4.1.2..Régression linéaire multiple (MLR)

Afin de proposer un modèle mathématique reliant les descripteurs et l'activité, et d'évaluer quantitativement les effets physicochimiques du substituant sur l'activité de la totalité de l'ensemble de ces 30 molécules, nous avons présenté la matrice de données qui sont les variables physicochimiques correspondantes molécules à une analyse de régression linéaire multiple. Cette méthode a utilisé les coefficients R,  $R^2$ , MSE et les valeurs F pour sélectionner la meilleure performance de régression.

Où

**R**: est le coefficient de corrélation;  
 **$R^2$** : est le coefficient de détermination;  
**MSE**: est le erreur quadratique moyenne;  
**F** :est l'indice de Fisher.

Le traitement par régression linéaire multiple est plus précis car il permet de relier les descripteurs structuraux pour chaque activité de 30 molécules afin d'évaluer quantitativement l'effet du substituant. Les descripteurs sélectionnés sont : logp, eh. v. pol

Les corrélations entre l'activité biologique et les descripteurs utilisées sont exprimés par la relation mathématiques suivante :

$$pic50 = 6.766 - 0.114logp - 0.053EH - 0.003V + 0.073POL$$

(Equation 1)

$$N = 30 \quad R^2 = 0.752. \quad R^2 \text{ ajusté} = 0.708. \quad mse=0.031 \quad R^2cv = 0.601$$

Le modèle QSAR doit considérer un  $R^2 > 0.6$  pour qu'il soit valide [8].

Dans notre modèles QSAR obtenus  $R^2 = 0,752$  .

La valeur F est jugée la signification statistique au niveau de 95%, pour toutes les valeurs de F calculées sont supérieures par rapport aux valeurs lus dans la table du Fischer.

Le modèle obtenu a été validé en interne par la technique de validation croisée leave-one-out; le coefficient de validation croisée  $r^2cv$  pour le modèle a été déterminé en fonction de la capacité prédictive du modèle. La valeur de  $r^2cv$  est supérieure à 0,5 ( $r^2cv = 0,729$ ); indique une bonne prévisibilité de modèle.

Dans l'équation du modèle QSAR, les coefficients négatifs de : log p, eh et V montrent que toute augmentation dans les valeurs de ces paramètres entraine une diminution de l'activité biologique.

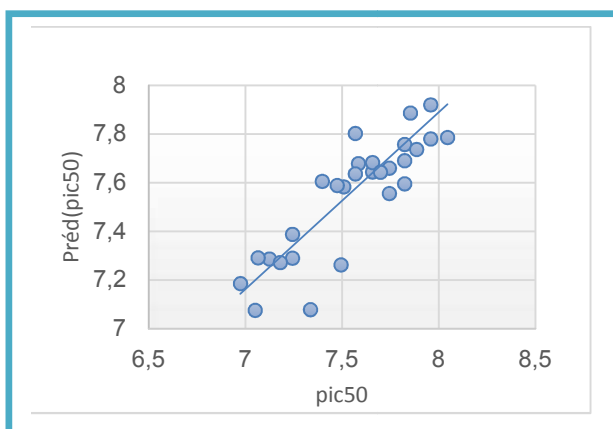


Les coefficients positives de pol, montre que toute augmentation dans la valeur de ce paramètre entraîne une augmentation de l'activité antipaludique

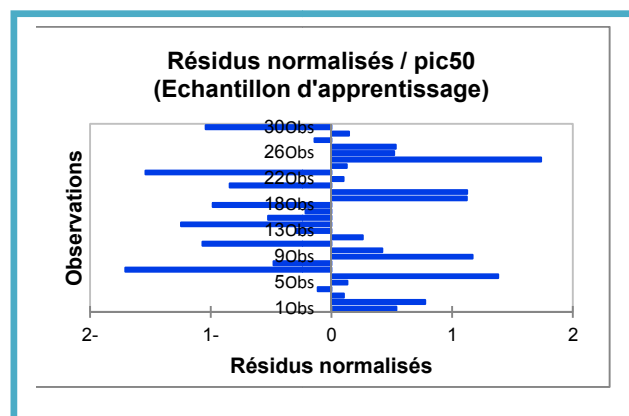
Avec le modèle MLR optimal, les valeurs des activités prédites pIC50 MLR calculées à partir de l'équation 1 et les valeurs observées sont données dans le (tableau 5) Les corrélations des activités expérimentale et observées sont illustrées à la (figure. III. 3)

Le résidu est dû d'une part à l'incertitude de la mesure expérimentale et d'autre part à l'imperfection du modèle, donc si l'activité prédite est plus proche de l'activité expérimentale, le modèle peut être applicable [9]

Le courbe obtenu présente une très bonne adéquation Il indique que le modèle peut être appliqué avec succès pour prédire l'activité biologique.



**Figur III .3:** Corrélation entre l'activité biologique expérimentale et prédite pour les modèle MLR



**Figur III .4** Corrélation entre les valeurs résiduelles et l'expérimentales de modèle MLR

QSAR, les valeurs résiduelles de l'activité biologique ont été tracées en fonction des valeurs expérimentales (**Figure.III. 4**)

Pour étudier la présence de l'erreur systématique dans l'élaboration des modèles QSAR, les valeurs résiduelles de l'activité biologique ont été tracées en fonction des valeurs expérimentales (**Figure.4**)

La répartition des résidus dans les deux côtés de l'axe des abscisses indique l'absence de l'erreur systémique [10]. Il indique que ce modèle peut être appliqué avec succès pour prédire l'activité antipaludique pour la série de molécules utilisée dans le développement de modèle QSAR.

Les descripteurs proposés dans l'équation (1) par MLR sont ensuite utilisés comme paramètres d'entrée dans les régressions non linéaires multiples (MNL) et PLS

#### III.4.1.3 Régression non linéaire multiple (MNL)

Nous avons également utilisé la technique du modèle de régression non linéaire pour améliorer la relation structure-activité évaluer quantitativement l'effet du substituant. Nous avons appliqué à la matrice de données constituée de manière évidente

Descripteurs proposés par MLR correspondant aux 30 molécules. Les coefficients R, R<sup>2</sup> et les valeurs F sont utilisés pour sélectionner la meilleure performance de régression. Nous avons utilisé une fonction pré-programmée de XLSTAT suivant:

$$Y = a + (bX_1 + cX_2 + dX_3 + eX_4 \dots) + (fX_1^2 + gX_2^2 + hX_3^2 + iX_4^2 \dots)$$

Où a, b, c, d ... représentent les paramètres et X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub>, X<sub>4</sub> ...: représentent les variables

L'équation résultante est la suivante

$$\text{pic50} = -0,3751 + 1,5805E-02M - 0,123\log p - 7,071E-02EH + 1,049E-02s - 1,627E-02v + 0,438\text{pol} - 2,19348150288456E-05m^2 - 2,591E-02\log p^2 - 7,249E-04eh^2 - 1,294E-05S^2 + 8,336E-06v^2 - 4,376E-03\text{pol}^2$$

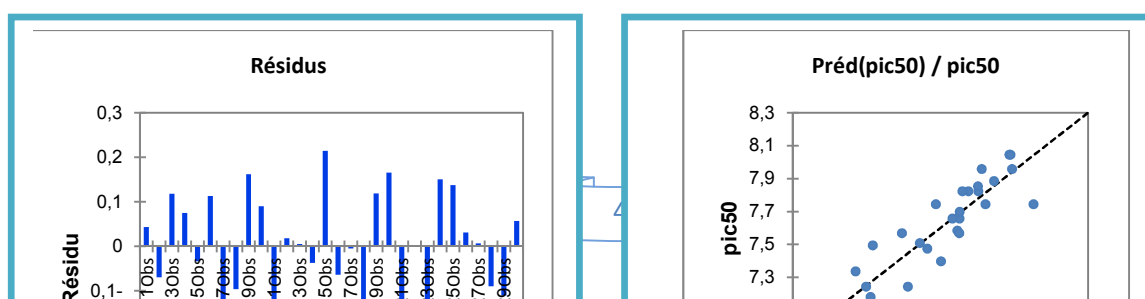
L'éq02

$$N = 30 ; R = 0.914 ; R^2 = 0.835 ; \text{mse} = 0.027 ; R^2_{cv} = 0.853$$

Avec le modèle MNLR optimal, les valeurs des activités prédites pIC50 MNLR calculées à partir de l'équation 2 et les valeurs observées sont données dans le tableau 5. Les corrélations des activités prédites et observées sont illustrées à la (**figure. III. 5**). La corrélation entre les activités MNLR calculées et les activités expérimentales est très significative, comme l'illustre la (**figure .III .6**) et comme indiqué par le valeur R.

La valeur élevée du R<sup>2</sup><sub>cv</sub> est un critère essentiel pour une meilleure qualification de modèle QSAR . Dans notre étude on constate que R<sup>2</sup><sub>cv</sub> = 0.853 Il ressort clairement des indicateurs statistiques clés, du coefficient de détermination R<sup>2</sup>, et l'erreur quadratique moyenne MSE, que la capacité prédictive de ce modèle est meilleure que celle du modèle linéaire (MLR). L'amélioration de la capacité prédictive était due à l'implication des termes au carré dans le modèle non linéaire

La répartition des résidus dans les deux côtés de l'axe des abscisses indique l'absence de l'erreur systématique.



**Figure.III.6** :Corrélation entre les valeurs résiduelles et l'expérimentale de modèle MNLR

**Figure .III .5:** Corrélation entre l'activité biologique expérimentale et prédite pour les Modèle MNLR

#### III.4.1.4: régression par les moindres carrés partiels (PLS) .

La régression par les moindres carrés partiels (PLS) est une méthode efficace et optimale pour une méthode basée sur la covariance. Il est recommandé dans les cas où le nombre de variables est élevé et où il est probable que les variables explicatives sont corrélées.

Nous avons soumis la matrice de données constituée évidemment à partir des descripteurs proposés par MLR correspondant aux 30 molécules, aux moindres carrés partiels (PLS). Cette méthode a utilisé les coefficients  $R^2$  et les valeurs mse pour sélectionner la meilleure performance de régression.

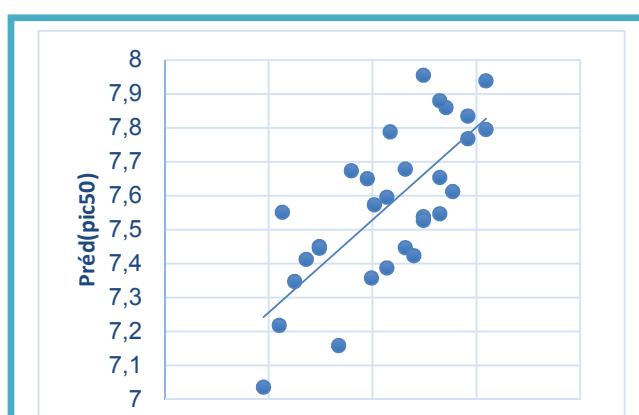
Les modèles QSAR construits à l'aide de la méthode des moindres carrés partiels (PLS) sont représentés par l'équation suivante

$$\text{pic50} = 7,648 + 1,796E-04M - 0,119\log p - 3,426E-02eh - 6,501E-05s - 2,710E-04v - 3,916E-03\text{pol}$$

l'éq. 3

$$N = 30 \quad R = 0.742 \quad R^2 = 0.552. \quad \text{mse} = 0.042 \quad R^2_c = 0.853$$

Les valeurs de pIC50 prédites calculées à partir de l'éq. 3, et les valeurs observées sont données dans le (tableau.III.5). Les corrélations des valeurs de pIC50 prédites et observées sont illustrées sur la (figure III 7).



**Figure. III 7:** Corrélations de *pic50* observées et prédites par PLS.

Ce mauvais résultat obtenu  $R^2$  inférieure 0.6 montre que le modèle proposé par le PLS n'est pas capable de prédire l'activité antipaludique de façon satisfaisante

**Tableau. III 5:** Activité antipaludique expérimentale, prédite et résiduelle des dérivés de 7-substitue -4aminoquinolin pour les trois modèles (MLR.MNLR.PLS)

		MLR		MNLR		PLS	
	<b>pic50<sub>exp</sub></b>	<b>Pic50<sub>préd</sub></b>	<b>Résid</b>	<b>Pic50<sub>préd</sub></b>	<b>résid</b>	<b>Pic50<sub>préd</sub></b>	<b>résid</b>
1.	7,824	7,740	0,084	7,781	0,043	7,655	0,169
2.	7,745	7,599	0,146	7,814	-0,069	7,539	0,206
3.	8,046	8,019	0,027	7,928	0,118	7,939	0,107
4.	7,854	7,877	-0,023	7,779	0,075	7,861	-0,007
5.	7,658	7,659	-0,001	7,691	-0,033	7,679	-0,021
6.	8,046	7,813	0,233	7,933	0,113	7,796	0,250
7.	7,745	8,036	-0,291	8,042	-0,297	7,956	-0,211
8.	7,585	7,686	-0,101	7,681	-0,096	7,789	-0,204
9.	7,959	7,757	0,202	7,797	0,162	7,835	0,124
10.	7,824	7,756	0,068	7,734	0,090	7,881	-0,057
11.	7,398	7,608	-0,210	7,604	-0,206	7,674	-0,276
12.	7,959	7,907	0,052	7,941	0,018	7,769	0,190
13.	7,509	7,585	-0,076	7,504	0,005	7,574	-0,065
14.	6,975	7,213	-0,238	7,012	-0,037	7,037	-0,062
15.	7,495	7,261	0,234	7,280	0,215	7,358	0,137
16.	7,475	7,577	-0,102	7,539	-0,064	7,651	-0,176
17.	7,244	7,287	-0,043	7,249	-0,005	7,451	-0,207
18.	7,244	7,430	-0,186	7,447	-0,203	7,446	-0,202
19.	7,824	7,619	0,205	7,705	0,119	7,548	0,276
20.	7,745	7,556	0,189	7,580	0,165	7,528	0,217
21.	7,125	7,310	-0,185	7,302	-0,177	7,348	-0,223

22.	7,658	7,668	-0,010	7,658	0,000	7,447	0,211
23.	7,569	7,835	-0,266	7,690	-0,121	7,596	-0,027
24.	7,569	7,556	0,013	7,419	0,150	7,388	0,181
25.	7,337	7,066	0,271	7,199	0,138	7,159	0,178
26.	7,886	7,784	0,102	7,855	0,031	7,613	0,273
27.	7,699	7,604	0,095	7,692	0,007	7,424	0,275
28.	7,181	7,227	-0,046	7,270	-0,089	7,413	-0,232
29.	7,051	7,066	-0,015	7,161	-0,110	7,219	-0,168
30.	7,066	7,272	-0,206	7,009	0,057	7,552	-0,486

## Références

[1].www. HyperChem.com

[2]. V. Dimitrov, T. Komatsu, An interpretation of optical properties of oxides and oxide glasses in terms of the electronic ion polarizability and average single bond strength, Journal of the University of Chemical Technology and Metallurgy vol 45, pp. 219-250,( 2010).

[3]. L. B. Kier, Molecular Orbital Theory in Drug Research, Academic Press, New York, USA (1981).

[4]. A. K. Ghose, G. M. Crippen, J. Chem. Inf. Comput. Sci.27:21 (1987).

[5]. G. Burgot, N. Bona, M. V. Christen, J. L. Burgot, Int. J. Pharm.129:295 (1996).

[6].www.xlstat.com

[7]- 17. Ayadi, I., H,Habib., D,yasmine., International Water Ressources Association.XIIIth World Water Conress 1 (4 Septembre 2008). Montpellier, France.

[9]- M.Clark, and R. D. Cramer, III, The probability of chance correlation using partial Least squares (PLS). Quantitative Structure-Activity Relationships, 12 (2): pp. 45,137,(1993).

[10] M. Jalali-Heravi and A. Kyani, J. Chem. Inf. Comput. Sci. 44, 1328, (2004).



# *Conclusion*

## Conclusion

Ce travail est une étude propriétés de la relation structure-activité d'une série aminoquinolie bioactive que basé sur la méthode QSAR, consiste à trouver une corrélation entre une activité biologique mesurée pour un panel des composés et certains descripteurs moléculaires.

Une analyse quantitative de la relation structurepropriété (QSAR) a été réalisée sur 30 molécules dérivées du 7-substitué-4-aminoquinoline dérivé. Un modèle QSAR a été établi en utilisant la régression linéaire multiple (MLR), les moindres carrés partiels (PLS) et la régression non linéaire multiple (MNLR).

Dans ce travail, la technique PCA est utilisée pour obtenir une vue d'ensemble des similarités et des dissimilarités des composés examinés, pour éliminer les descripteurs indépendants (la masse et réfractivité et surface sont éliminés) et de sélectionner des descripteurs qui montrent une corrélation élevée avec l'activité Étudié ( $\log p; \text{eh}; \text{pol}; \text{v}$ ) qui donnent un poids supplémentaire parce qu'ils seront plus efficaces pour les prédictions.

L'évaluation de la qualité des modèles MLR, PLS et MNLR a montré que la capacité prédictive de MNLR était nettement meilleure que celle des autres méthodes ( $R^2 = 0.835$ ). Le pouvoir prédictif du modèle obtenu a été confirmé par la validation croisée LOO  $R^2_{cv}=0.853$ . Une forte corrélation a été observée entre les valeurs expérimentales et prédites des activités biologiques, qui indiquaient la validité et la qualité du modèle QSAR développé dans ce travail. Nous concluons que la découverte la plus importante de cette recherche est que nous avons pu concevoir et prédire de nouveaux composés avec des valeurs plus élevées ou plus faibles que les composés existants (**Tableau 5**) en ajoutant des substituants appropriés en calculant leur propriété à l'aide des équations LRM, RNLM et PLS. Ainsi, les modèles proposés permettront de réduire le temps, le coût, et aussi la mobilisation humaine.



## **Abstract:**

The quinoline moiety is one of the widely studied scaffolds for generating derivatives with various pharmacophoric groups due to its potential antimalarial activities. In the present study, a series of 7-substituted-4-aminoquinoline derivatives were selected to understand their antimalarial properties computationally by molecular modeling techniques including 2D QSAR, and find a correlation between the different physic-chemical parameters (descriptors) of these compounds and its biological activity, using principal components analysis (PCA), multiple linear regression (MLR), *to a regression partial least squares (PLS)*.multiple non-linear regression (MNLR)). The principal component analysis (PCA) has been used to select descriptors that show a high correlation with activities. Based on our result, a comparison of the quality of, MLR, PLS, and MNLR models shows that the MNLR (R = **0.914** and R<sup>2</sup> = **0.835**;R<sub>CV</sub>= **0.853**) models have substantially better predictive capability because the MNLR approach gives better results than MLR(R = 0.835 and R<sup>2</sup> = 0,752;R<sub>CV</sub>=0.601)), PLS (R = **0.742** and R<sup>2</sup> = **0.552**; R<sub>CV</sub>=**0.550**) . The model of MNLR gave statistically significant results and showed good stability to data variation in leave-one-out cross-validation.

**KEYWORDS** :Antimalarial; 7-substituted-4-aminoquinoline; QSAR; PCA; MLR ;MNLR; MLR

## **Résumé:**

La fraction de quinoléine est l'un des échafaudages largement étudiés pour générer des dérivés avec différents groupes pharmaceutiques en raison de leurs activités antipaludiques potentielles. Dans cette étude, une série de dérivés de 7-substituted-4-aminoquinoline 1 a été choisie pour comprendre leurs propriétés antipaludiques par des techniques de modélisation moléculaire incluant QSAR 2D,

Et pour trouver une relation entre les différents paramètres physico-chimiques (descripteurs) de ces composés et leur activité biologique, en utilisant l'analyse en composantes de base (ACP), régression linéaire multiple (MLR), à la régression sous-unités minimales (PLS) Régression non linéaire (MNLR). L'analyse élémentaire (ACP) a été utilisée pour déterminer quels attributs sont les plus pertinents pour les activités. Sur la base de nos résultats, une comparaison de la qualité des modèles MLR, PLS et MNLR montre que les modèles MNLR (R = **0.914** and R<sup>2</sup> = **0.835**;R<sub>CV</sub>= **0.853**) La prévisibilité est nettement meilleure parce que l'approche MNLR donne de meilleurs résultats que MLR(R = 0.835 and R<sup>2</sup> = 0,752;R<sub>CV</sub>=0.601)), PLS (R = **0.742** and R<sup>2</sup> = **0.552**; R<sub>CV</sub>=**0.550**) Le modèle MNLR a donné des résultats statistiquement significatifs et a montré une bonne stabilité pour le changement des données lors de la vérification croisée de l'un à l'autre

**Mots Clés:** antipaludiques; 7-substituted-4-aminoquinoline; QSAR; PCA; MLR ;MNLR; MLR