



Université Mohamed Khider de Biskra

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

Science de la matière

MÉMOIRE DE MASTER

Domaine des Sciences de la matière

Filière de Chimie

Spécialité de Chimie pharmaceutique

Présenté et soutenu par :

KEDIDI Naouel

GUEBBABI Kamilia

Etude 2D-QSAR DE Dérivés de benzophénone en tant qu'agents antipaludique

Jury :

Dr	Aggoun.sihem	M.A.A	Université Med khider Biskra	Présidente
Dr	Belaidi.Hammel	M.A.A	Université Med khider Biskra	Examineur
Dr	Mme.Ouassaf	M.C.B.	Université Med khider Biskra	Rapporteuse

Année universitaire : 2019/2020



Dédicace

À ma mère

Tous les mots ne sauraient exprimer la gratitude, l'amour, Le respect et la reconnaissance que J'ai pour vous. Votre amour et votre patience m'ont accompagné chacun de mes pas. J'ai conscience de tous les sacrifices que vous avez dû faire pour me permettre de mener mes études dans les meilleures conditions possibles. Mon admiration pour la femme si courageuse que vous êtes est sans limite.

Puisse dieu, le tout puissant, te préserve, te combler de ses grâce infinies, et t'accorder sante, et longue vie et bonheur

À mon père :

Ce travail est le fruit de tes sacrifices que tu as consenti pour mon éducation et ma formation, Je vous remercie pour votre présence et votre soutien tout au long de ma vie dans les meilleurs moments comme dans les mauvais pour devenir ce que je suis, vous étiez ma source d'inspiration.

Puisse dieu, le tout puissant, te préserve, te combler de ses grâce infinies, et t'accorder sante, et longue vie et bonheur.

À mes chères sœurs Mliya, Bassma et Abir, leurs encouragements permanents, et leur soutien moral.

À mes chers frères Ibrahim, Azzedin, Lazhar et Yamine pour leur appui et leur encouragement.

Aux femmes de mes frères Mona, Souad pour encouragement A toute ma famille pour leur soutien tout au long de mon parcours universitaire.

À Mon binôme KEDIDI Naouel qui j'ai trouvé l'entente dont j'avais besoin.

À mes fidèles amis : Ahlam B, Linda B, Assma G, Souhila M, Sawsan Z, Boutheina Z, Khadidja F, Houda D, qui ont m'accompagne tout au long de mes 5 ans. À mes collègues de la chimie pharmaceutique, À tous les gens qui me connaissent, qui m'aident et ceux qui me souhaitent la réussite et le bonheur dans ma vie.

À tous les gens que j'aime sans exception.

GUEBBABI Kamilia



À mon dieu qui m'aide et me donne la patience et le courage durant ce mémoire, et mes études en générale.

Je dédie spécialement mes chers parents, pour tous leurs sacrifices, leur amour, leur tendresse, leur soutien et leur prière tous au long de mes études. Que dieu vous protège, Amin.

A mes chères sœurs rayene et leila manel, pour leurs encouragement permanents, et leur soutien moral.

A mes chères frères, walid et atef spécialement rabbi yerhmou, pour leur appui et leur encouragements, avant et maintenant.

A mon binôme Kamilia.

A toute ma famille KEDIDI et DJARI, pour leur soutien out au long de mon parcours universitaire. Que ce travail soit l'accomplissement de vos vœux tant allégués, et le fruit de votre soutien infailible.

Merci d'être toujours là pour moi.

KEDIDI Naouel

Remerciements

En préambule a ce mémoire nous remerciant ALLAH qui nous a aidés et nous a donné la patience et le courage durant ces années d'étude.

Nous souhaitant adresser nos remerciements les plus sincères aux personnes qui nous ont apporté ont apporté leur aide et qui ont contribué à l'élaboration de ce mémoire ainsi qu'à la réussite de cette formidable année universitaire.

Ces remerciements vont tout d'abord au corps professoral et administratif de la faculté des sciences de la nature et de la vie, pour la richesse et la qualité de leur enseignement et qui déploient de grands efforts pour assurer à leurs étudiants une formation actualisée.

Nous tenant à remercier sincèrement l'encadreur, **Mme OUASSAF mebarka** pour sa patience, ses encouragements, son orientation, et ses conseils précieux, ainsi pour l'inspiration, l'aide et le temp qu'il a bien voulu nous consacrer et sans qui ce mémoire n'aurait jamais vu le jour.

Nous remercions nos parents pour leur contribution bien sûr, leur soutien et leur patience.

Enfin, nous adressons nos plus sincères remerciements à tous nos proches et amis, qui nous ont toujours encouragés au cours de la réalisation de ce mémoire

Merci à tous et à toutes.

GUEBBABI Kamilia

KEDIDI Naouel

TABLEAU DES MATIERES

DEDICACE.....	A
DEDICACE.....	ERREUR ! SIGNET NON DEFINI.
REMERCIEMENTS.....	C
TABLEAU DES MATIERES	D
LISTE DES TABLEAUX	E
LISTE DES FIGURES.....	F
LISTES DES SCHEMAS	F
<i>INTRODUCTION GENERAL :</i>	<i>2</i>
<i>REFERENCES BIBLIOGRAPHIQUES :</i>	<i>5</i>
<i>CHAPITRE I RAPPELS BIBLIOGRAPHIQUE SUE LES ANTIPALADIQUE</i>	
I. LE PALUDISME	7
1. INTRODUCTION :.....	7
2. DEFINITION DE PALUDISME :.....	9
3. AGENT PATHOGENE :.....	9
3.1. Plasmodium falciparum :	9
3.2. Plasmodium vivax :	10
3.3. Paludisme ovale :	10
3.4. Paludisme malaria :	10
3.5. Paludisme knowlesi :	10
4. VECTEUR :.....	11

5. CYCLE :	11
5.1. Chez l'homme :	12
5.1.1. Schizogonie pré-érythrocytaire :	12
5.1.2. Schizogonie érythrocytaire :	13
5.2. Chez l'anophèle femelle :	14
5.3. Modalités de transmission :	14
II. ANTIPALUDIQUE :	15
1. CLASSIFICATION DES ANTIPALUDIQUES :	15
1.1. Les modes d'action des principaux antipaludiques actuels (2, 3) :	15
1.1.1. Les schizonticides électifs :	15
1.1.2. Les inhibiteurs des acides nucléiques :	15
1.1.3. Les nouvelles thérapeutiques antipaludiques : perspectives.....	16
III.LA PROTEINE FARNÉSYLTRANSFÉRASE :	19
1. LA STRUCTURE DE LA FTASE :	19
2. LE MECANISME CATALYTIQUE DE LA FTASE :	21
REFERENCES BIBLIOGRAPHIQUES :	25
<i>CHAPITRE II : ETUDE QUANTITATIVE DES RELATIONS STRUCTURE-ACTIVITÉ (QSAR)</i>	
II.1.INTRODUCTION :	27
II.2.HISTORIQUE :	27
II.3.DEFINITION :	28
II.4.PRINCIPE DE QSAR :	29
II.5.METHODOLOGIE GENERALE D'UNE ETUDE QSPR/QSAR :	30
II.5.1. Les bases de données :	30
II.5.2. Les Descripteurs moléculaires :	30

II.6.METHODES D'ANALYSE DE DONNEES :	31
II.6.1. Analyse par composantes principales :	31
II.6.2. La régression linéaire multiple :	33
II.6.3. La régression non linéaire multiple :	34
II.6.4. La régression des moindres carrés partiels :	35
II.7. TEST DE LA SIGNIFICATION GLOBALE DE LA REGRESSION :	35
II.8. INTERPRETATION ET VALIDATION D'UN MODELE QSAR/QSPR :	37
•Interne :	37
• Y-Randomisation :	38
• Externe :	38
• Domaine d'applicabilité :	39
REFERENCES BIBLIOGRAPHIQUES :	41
<i>CHAPITRE III APPLICATION ET RÉSULTAT ET DISCUSSION</i>	
III.1.INTRODUCTION:	45
III.2.METHODOLOGIE :	45
III.2.1. Base de données :	45
III.2.2. Calcul des descripteurs :	50
III.3. RESULTATS ET DISCUSSIONS :	54
III.3.1. Analyse en composantes principales (ACP) :	54
III.3.2. Répartition des données et développement de modèles :	57
III.3.3. Analyse statistiques de meilleur Modèle MLR :	59
III.3.4.Régression non linéaire multiple (MNLr) :	65
III.4.VALIDATION DU MEILLEUR MODELE QSAR :	66

III.4.1. Validation interne :	66
III.4.2. Validation externe :	67
III.4.3. Y-Randomisation de MLR et MNLR :	68
III.4.4. Domaine d'application :	72
REFERENCES BIBLIOGRAPHIQUES :	76
CONCLUSION GENERAL :	78

LISTE DES TABLEAUX

TABLEAU III.1: NUMEROTATION, STRUCTURE ET ACTIVITE EXPIRATOIRE DE L'ENSEMBLE DE DONNEES.....	46
TABLEAU III.2: DESCRIPTEURS MOLECULAIRES UTILISES DANS L'ETUDE QSAR.	51
TABLEAU III.3: VALEURS PIC50 EXPERIMENTALES AVEC LES DESCRIPTEURS SELECTIONNES.....	53
TABLEAU III.4: LA MATRICE DE CORRELATION (PEARSON (N)) ENTRE LES DIFFERENTS DESCRIPTEURS OBTENUS.....	56
TABLEAU III.5 : MODELES SELECTIONNES ET PARAMETRES STATISTIQUES DES CORRELATIONS ENTRE LES PROPRIETES MOLECULAIRES.	57
TABLEAU III.6: ACTIVITE OBSERVEE VS CALCULEE DES DERIVES DE BENZOPHENONE.....	61
TABLEAU III.7: MATRICE DE CORRELATION.....	62
TABLEAU III.8: PARAMETRES DE VALIDATION CROISEE.	67
TABLEAU III.9: PARAMETRES DE VALIDATION CROISEE.	67
TABLEAU III.10: 100 ITERATIONS D'Y-RANDOMISATION DE MLR.....	69
TABLEAU III.11: 100 ITERATIONS D'Y-RANDOMISATION DE MNLR.....	70
TABLEAU III.12: PRESENTE LES VALEURS DES ACTIVITES EXPERIMENTALES ET PREDITES DES 35 DERIVES ETUDIE ET LEURS RESIDUS.	74

LISTE DES FIGURES

FIGURE I.1: LA REPARTITION MONDIALE DU PALUDISME EN 2010 SELON L'OMS.	8
FIGURE I.2: CYCLE DE PALUDISME.	12
FIGURE I.3: GAMETOCYTE FEMELLE DE P. FALCIPARUM.....	13
FIGURE I.4: FEMELLE DU GENRE ANOPHELES SE GORGEANT.	14
FIGURE I.5: STRUCTURE DE QUELQUES MOLECULES UTILISEES EN TANT QUE MEDICAMENTS ANTIPALUDIQUES.	18
FIGURE I.6 : LA STRUCTURE DEVELOPEE DE LA PROTEINE FTASE.	20
FIGURE I.7: STRUCTURE DE LA FTASE.	20
FIGURE I.8: LIAISONS DU FPP AVEC L'ENZYME EN PRESENCE DU MOTIF CAAX.	23
FIGURE I.9: ROTATION DES DEUX DERNIERES UNITES ISOPRENYLES DU FPP POUR LA FARNESYLATION.....	24
FIGURE II.1 : UNE REPRESENTATION GRAPHIQUE SIMPLE DES BASES DE L'ACP. DANS LE PLOT 1A CI DESSUS, LES DONNEES SONT REPRESENTEES DANS LE SYSTEME DE COORDONNEES XY.	33
FIGURE II.2 : MPORTANCE DE LA VALEUR DE K DANS LA VALIDATION EXTERNE. 39	
FIGURE III.1 : LES PRINCIPALES COMPOSANTES ET LEURS VARIANCES.	54
FIGURE III.2 : CERCLE DE CORRELATION.	55
FIGURE III.3 : CORRELATION ENTRE L'ACTIVITE BIOLOGIQUE EXPERIMENTALE ET PREDITE.	63
FIGURE III.4 : CORRELATION ENTRE L'ACTIVITE BIOLOGIQUE EXPERIMENTALE ET RESIDUELLE.....	64

FIGURE III.5 : VALEURS RESIDUELLES NORMALISEES EN FONCTION DES ACTIVITES PREDITES ET OBSERVEES.....	64
FIGURE III.6: REPRESENTATION GRAPHIQUE DE L'ACTIVITE CALCULEE ET OBSERVEE PAR LE MNLR.....	66
FIGURE III.7 : CORRELATION ENTRE LES VALEURS RESIDUELLES ET L'EXPERIMENTALE DE MODELE MNLR.	66
FIGURE III.8: DIAGRAMME DE WILLIAMS DU MODELE MLR.....	73
FIGURE III.9 : DIAGRAMME DE WILLIAMS DU MODELE MNLR.....	73

LISTES DES SCHEMAS

SCHEMA I.1: STRUCTURES AU COURS DE LA CINETIQUE DE LA FTASE.....	22
SCHEMA II .1 : MODELE D'ETUDE QUANTITATIVE DES RELATIONS STRUCTURE/ACTIVITE (QSAR).....	29

LISTE DES PRINCIPALES ABREVIATIONS

ACP	Analyse en composantes principales
DA	domaine d'application
FPP	Farnésyl Pyphosphate
FTase	Farnésyltransférase
HF	Hartree-Fock
HOMO	Orbitale Moléculaire plus Haut Occupée
LOO	Leave One Out
LUMO	Orbitale Moléculaire plus bas Occupée
MLR	Régression linéaire multiple
MNLR	régression non linéaire multiple
MM	masse molaire moléculaire
MW	masse moléculaire
NHD	Le nombre de donneurs de liaisons hydrogène
NROT	Le nombre de liaisons rotatives
OMS	Organisation Mondiale de la Santé
PFT	protéine Farnésyltransférase
P(e)	Coefficient de polarisabilité
PIC50	log. Inhibiteur concentration of 50%
PLS	Régression des moindres carrés partiels
PRESS	Somme des carrés des erreurs de prédiction
QSAR	Relations quantitatives structure-activités
QSPR	Relations quantitatives structure-propriétés
SPRESS	La capacité prédictive des modèles est évaluée par l'erreur de racine-carrée moyenne.
SSE	Erreur de somme des carrés
SSR	Régression de la somme des carrés
SSY	Somme des carrés des distances des valeurs observées
3D, 2D, 1D	Trois dimensions, deux dimensions, une dimension

Résumé :

Dans ce travail, une relation quantitative structure-activité (QSAR) a été réalisée pour prédire les activités inhibitrices de la protéine farnésyltransférase (PFT) pour une série de 35 dérivés de benzophénone. L'ensemble de données a été divisé en deux sous-ensembles l'ensemble d'apprentissage et de test, les modèles MLR et MNLR ont montré des paramètres de validation internes et externes satisfaisants ($R^2_{\text{ajusté}} = 0.896$, $R^2_{\text{test}}=0.909$) et ($R^2_{\text{ajusté}} = 0.757$, $R^2_{\text{test}} = 0.914$) qui ont passé les critères d'acceptabilité du modèle. Les modèles ont été validés rigoureusement grâce à une validation croisée. En outre, la puissance prédictive externe du modèle proposé a été examinée en tenant compte, des critères de modèle acceptables de Golbraikh, Le domaine d'applicabilité des modèles a été soigneusement défini à l'aide du graphique de Williams. Les modèles montrent que les activités antipaludiques de la benzophénone dépendent des descripteurs logP, bpol, MAXDn et FMF, Ces indicateurs peuvent être utilisés pour concevoir de nouveaux inhibiteurs de la benzophénone PFT et prédire la valeur de leurs activités antipaludiques sur la base des équations MLR et MNLR QSAR.

Mots-clés : QSAR; MLR; MNLR ; Inhibiteurs de la PFT ; benzophénone ; antipaludique.

Abstract:

In this work, a quantitative structure-activity relationship (QSAR) was performed to predict the inhibitory activities of protein farnesyltransferase (PFT) for a series of 35 benzophenone derivatives. The dataset was split into two subsets the training and testing set, the MLR and MNLN models showed satisfactory internal and external validation parameters (adjusted $R^2 = 0.896$, R^2 test = 0.896) and (Adjusted $R^2 = 0.757$, R^2 test = 0.914) that have passed the model acceptability criteria. The models have been rigorously validated through cross validation. In addition, the external predictive power of the proposed model was examined taking into account, Golbraikh's acceptable model criteria. The domain of applicability of the models was carefully defined using the Williams plot. The models show that the antimalarial activities of benzophenone depend on the descriptors logP, bpol, MAXDn and FMF. These indicators can be used to design new inhibitors of benzophenone PFT and predict the value of their antimalarial activities based on the MLR equations and MNLN QSAR.

Keywords: QSAR; MLR; MNLN; PFT inhibitors; benzophenone; antimalarial.



INTRODUCTION GÉNÉRAL

Introduction général :

Le paludisme est dû à des parasites du genre Plasmodium transmis à l'homme par des piqûres de moustiques femelles de l'espèce Anophèles infectés, appelés « vecteurs du paludisme ». Il existe cinq espèces de parasites responsables du paludisme chez l'homme (Plasmodium falciparum, Plasmodium vivax, Paludisme ovale, Paludisme malaria, Paludisme knowlesi), dont deux – P. falciparum et P. vivax – sont les plus dangereuses.

La modification post-traductionnelle des protéines par hydrolyse primaire est un mécanisme important de régulation cellulaire. L'une des enzymes clés de ce mécanisme est la farnésyltransférase, qui est une hétéroprotéine contenant du zinc [1-2] inhibiteurs de cette enzyme sont en cours de développement comme agents anticancéreux [3-4] Plusieurs groupes ont suggéré que la farnésyltransférase pourrait être une cible efficace pour le développement de médicaments contre le paludisme et d'autres maladies parasitaires telles que la trypanosomiase, et nous avons développé des inhibiteurs avec une activité in vitro et dans certains cas in vivo contre Plasmodium falciparum[5-8].

La chimoinformatique est un outil de choix pour diminuer le temps et le coût de développement d'un médicament et constitue le grand exemple issu de ce développement où elle est maintenant reconnue comme subdivision de la chimie et soutien principal à la chimie moderne. Cette discipline peut intervenir à différents niveaux du processus de découverte d'un médicament. Parmi les techniques de chimoinformatique nous pouvons citer le criblage de chimiothèques (techniques de QSAR, de docking et de pharmacophores), la conception de chimiothèques virtuelles, la conception de composés de novo, la prédiction de propriétés et l'étude de l'affinité protéine – ligand.

La chimoinformatique est aujourd'hui présente dans toutes les étapes de développement d'un médicament. La conception de médicament à l'aide d'outils informatiques à partir de la structure de la cible biologique a par exemple contribué à l'introduction en phase clinique d'environ 50 composés et à la mise sur le marché de nombreux médicaments.

Le technique QSAR qui consiste à trouver une corrélation entre une activité biologique mesurée pour un panel de composés et certains descripteurs moléculaires [9].

La comparaison de l'activité biologique de certaines molécules et de leurs structures a permis d'établir dans de nombreux cas des corrélations entre les paramètres structuraux et les propriétés d'une molécule. L'association des variations de l'activité biologique aux paramètres structuraux permet 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 les valeurs des paramètres qui correspondent à une activité biologique maximale et de prédire l'activité des molécules qui n'ont pas encore été synthétisées [10].

La validité d'un modèle QSAR (Quantitative Structure-Activité Relationship) dépendra donc du choix que l'on aura fait sur les paramètres. Il est ainsi judicieux d'identifier et d'évaluer quantitativement les paramètres les plus pertinents à utiliser en fonction de l'activité et des molécules sélectionnées [11-12].

Dans cette thèse, nous rapportons une étude computationnelle de benzophénone et une série de ses dérivés ; donc notre objectif est d'appliquer l'ensemble des méthodes fournies par la chimie informatique pour déterminer les propriétés structurales et électroniques du noyau benzophénone ; l'effet de substitution sur ces propriétés est également discuté et enfin des corrélations entre la structure chimique et l'activité biologique d'une série de molécules bioactives.

Notre travail comporte trois chapitres on commence par **une introduction général** après :

- **Le premier chapitre :** ce manuscrit présente une généralité sur le paludisme, généralités sur les Antipaludiques et l'inhibition de l'enzyme.
- **Dans le deuxième chapitre :** nous développerons un rappel théorique sur des différentes méthodes QSAR et les méthodes statistiques utilisées pour obtenir les modèles QSAR l'analyse en composantes principales ACP, la régression linéaire multiple MLR, la régression non linéaire multiple MNLRCeci 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.
- **Le troisième chapitre** : l'étude quantitative des relations structure-activité d'une série de benzophénone. Nous présentons la méthodologie de l'étude QSAR et les descriptions des méthodes statistiques utilisées.

Enfin, notre travail se termine par **une conclusion générale**.

Références bibliographiques :

- [1] Sousa, S. F. Fernandes, P. A. Ramos, M. J. J. Biol, Inorg. Chem. 3, 2005, 10.
- [2] Casey, P. J.Solski, P. A. Der, C. J.Buss, J. E. Proc, Natl, SU.S.A. Acad. 8323, 1989, 86.
- [3] M.Appels, J.H.Beijnen, Schellens, J. H. Oncologist. 10, 2005, 565.
- [4] H.M.Gelb, Van Voorhis, W. C. Buckner, F. S.Yokoyama, K. Eastman, R. Carpenter, E. P.Panethymitaki, C. Brown, K. A. Smith, D. F. Mol. Biochem, Parasitol. 2003, 126-155.
- [5] R.T.Eastman,J. White,O. Hucke,K. Bauer ,Yokoyama, K. Nallan, L. Chakrabarti, D. Verlinde, C. L.Gelb, M. H. Rathod, P. K. Van Voorhis, W. C. J. Biol. Chem. 13554,2005,280.
- [6] J.Wiesner, Ortmann, R.Jomaa, H.Schlitzer, M. Angew. Chem., Int. Ed. 42, 2003, 5274.
- [7] Buckner, F. S. Eastman, R. T.Yokoyama, K.Gelb, M. H.Van Voorhis, W. C. Curr. Opin. Investig. Drugs 6, 2005,791.
- [8] Kettler, K.Wiesner, J. Ortmann, R. Dahse, H.M. Jomaa, H.Schlitzer, M. Pharmazie.36, 2006, 61.
- [9] S. Belaidi. Thèse de doctorat, Université de Batna, 2002.
- [10] J. Acar, P. Courvalin. La recherche. 314,1998 ,50-52.
- [11] A. G. Maldonado, J. P. Doucet, M. Petijean, et B. T. Fan. Form theory to applications. Mol Divers. 10, 2006, 39-7.
- [12] Z. R. Boissier, J. Asselimean, J. P. Zalta, "Les antibiotiques, structures et exemples de Mode D'action"" Herman. Paris, 1993.



CHAPITRE I

RAPPELS BIBLIOGRAPHIQUE SUE

LES ANTIPALADIQUE

I. Le paludisme

1. Introduction :

Le paludisme (malaria en anglais) est une parasitose due à des hématozoaires du genre Plasmodium, transmise par des moustiques du genre Anophèles. Cette maladie, surtout importante pour les populations vivant en zone d'endémie (zone intertropicale), l'est aussi pour les voyageurs [1].

Importance à l'échelle mondiale :

- En 2009 le paludisme reste la première endémie parasitaire mondiale. On estime que près de la moitié de la population mondiale vit en zone d'endémie.
- Le nombre d'accès palustres survenant chaque année à travers le monde semble diminuer, il est estimé entre 250 à 500 millions, entraînant la mort d'environ 750000 à 1 millions de personnes, parmi lesquelles une majorité de jeunes enfants vivant en Afrique sub-saharienne.
- Le paludisme représente une charge financière énorme pour les populations et par conséquent la maladie constitue un obstacle au développement des pays concernés, notamment en Afrique. Pour toutes ces raisons, la lutte contre le paludisme constitue, avec la lutte contre le SIDA et la tuberculose, un des « Objectifs Du Millénaire » définis par les Nations-Unies, et le « Fond Mondial » est destiné à approvisionner les pays demandeurs en médicaments.

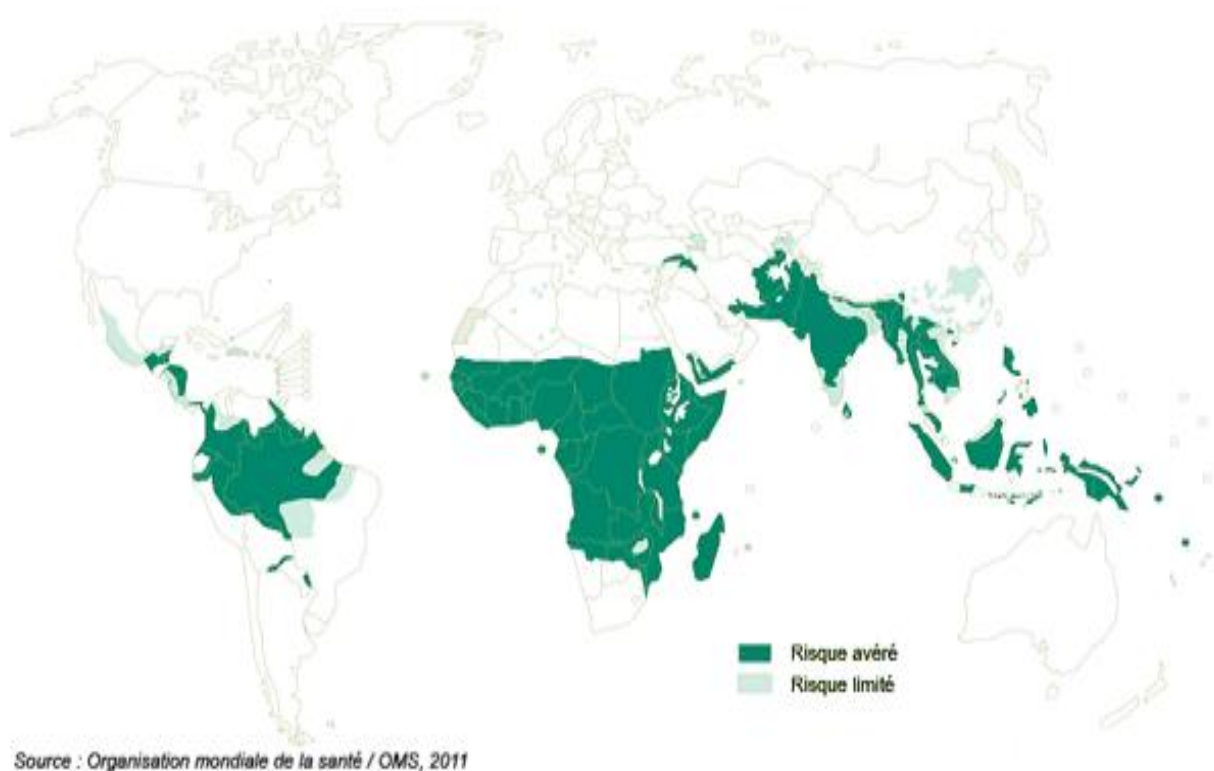


Figure I.1: La répartition mondiale du paludisme en 2010 selon l’OMS.

Importance pour les voyageurs :

A une échelle tout à fait différente, le paludisme est important aussi pour les voyageurs. L’augmentation actuelle des échanges internationaux fait que tout médecin exerçant en France métropolitaine peut être confronté au paludisme dans sa pratique quotidienne. En France, en 2010 le nombre de cas de paludisme d’importation diagnostiqués était estimé à environ 4600. Chaque année, moins de 20 personnes meurent en France du paludisme.

- Le paludisme reste une maladie potentiellement mortelle pour un sujet non-immun (cas des voyageurs).
- Au cours des dernières années, la prévention individuelle est redevenue difficile en raison de l’augmentation des résistances du parasite vis-à-vis des traitements préventifs.
- Actuellement aucun vaccin n’est commercialisé.

2. Définition de paludisme :

Le paludisme ou malaria est une maladie d'origine parasitaire se traduisant par de la fièvre et des troubles digestifs. Elle se transmet à l'Homme essentiellement par la piqure d'un moustique (la femelle du genre *Anophèle*) infecté, et plus rarement lors d'une transfusion sanguine ou par transmission mère-enfant pendant la grossesse.

3. Agent pathogène :

Le paludisme est transmis par un protozoaire appartenant au genre *Plasmodium*. Il existe de très nombreuses espèces de *Plasmodium* (plus de 140), touchant diverses espèces animales mais seulement cinq de ces espèces sont retrouvées en pathologie humaine. Il s'agit de *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale*, *Plasmodium malariae* et *Plasmodium knowlesi*, parasite habituel des singes (macaques) d'Asie qui vient de passer récemment chez l'homme. Les cinq espèces diffèrent par des critères biologiques, cliniques, par leur répartition géographique et par leur capacité à développer des résistances aux antipaludiques. D'emblée il faut différencier *P. falciparum* des autres espèces. En effet *P. falciparum* est celui qui est le plus largement répandu à travers le monde, qui développe des résistances aux antipaludiques et qui est responsable des formes cliniques potentiellement mortelles.

3.1. *Plasmodium falciparum* :

Dans les régions équatoriales, il est transmis toute l'année avec cependant des recrudescences saisonnières. Dans les régions subtropicales, il ne survient qu'en période chaude et humide. Sa transmission s'interrompt lorsque la température tombe en dessous de 18°C. Cela explique aussi que, quelle que soit la latitude, le paludisme n'est plus transmis en altitude (au-dessus de 1500 mètres en Afrique et 2500 mètres en Amérique et en Asie). L'évolution se fait d'un seul tenant après une incubation de 7 à 12 jours. On n'observe pas de rechutes tardives comme avec les autres espèces. Plus de 90% des accès palustres à *P. falciparum* surviennent dans les 2 mois qui suivent le retour du pays d'endémie. *P. falciparum* est responsable des formes cliniques graves, notamment du neuropaludisme. C'est l'espèce la plus fréquemment observée en France, responsable de plus de 80 % des paludismes dit « d'importation », c'est à dire contractés en zone d'endémie mais se révélant en France métropolitaine après le retour.

CHAPITRE I : Rappels Bibliographique Sur Les Antipaludiques

3.2. Plasmodium vivax :

Très largement répandu en Amérique du Sud et en Asie, il est beaucoup plus rarement observé en Afrique. Les érythrocytes du groupe sanguin Duffy négatif (observé chez la majorité des sujets originaires d'Afrique de l'Ouest) ne possèdent pas le récepteur membranaire nécessaire à l'infection par *P. vivax*. Sa transmission s'arrête en dessous de 15°. Sa période d'incubation est de 11 à 13 jours, mais on peut observer des rechutes (accès de reviviscence) pendant 3 à 4 ans. L'affection par *P. vivax* est classiquement considérée comme bénigne (fièvre tierce bénigne, c'est-à-dire due à un cycle érythrocytaire de 48 heures) mais en zone d'endémie il peut avoir des répercussions graves sur l'état de santé des populations, notamment par l'intermédiaire des anémies chez l'enfant. De plus on commence à voir surgir quelques résistances médicamenteuses à *P. vivax* à la chloroquine.

3.3. Paludisme ovale :

Il sévit en Afrique intertropicale du Centre et de l'Ouest (et dans certaines régions du Pacifique) et provoque une fièvre tierce bénigne, comme *P. vivax* dont il est très proche. Son incubation est de 15 jours au minimum mais peut-être beaucoup plus longue, jusqu'à 4 ans. Son évolution est bénigne mais on peut observer, comme avec *P. vivax*, des rechutes tardives (5 ans). Schématiquement on dit que *P. ovale* remplace *P. vivax* là où cette dernière espèce n'existe pas.

3.4. Paludisme malaria :

Il sévit sur les trois continents, de manière beaucoup plus sporadique. Il se différencie des autres espèces par une incubation plus longue (15 à 21 jours), par une périodicité différente de la fièvre (cycle érythrocytaire de 72 heures responsable d'une fièvre quarte) et surtout par sa capacité à entraîner des reviviscences très tardives (jusqu'à 20 ans après le retour de la zone d'endémie). Les mécanismes physiopathologiques responsables de ces reviviscences tardives ne sont pas totalement élucidés, certains évoquent la présence de mérozoïtes latents dans les voies lymphatiques. L'infection est bénigne mais *P. malaria* peut parfois entraîner des complications rénales.

3.5. Paludisme knowlesi :

Il sévit en Asie du Sud-Est (particulièrement en Malaisie, à Bornéo), en zone forestière car il est étroitement lié à la répartition des singes macaques, son hôte habituel, et de son vecteur piquant l'homme et la singe. Il est morphologiquement proche de *P. malaria*. Il se différencie des autres espèces par un cycle érythrocytaire de 24 heures responsable d'une fièvre quotidienne. Il existe de rares formes graves, voire mortelles, avec forte parasité. A ce jour aucune chimiorésistance n'a été observée pour cette espèce.

4. Vecteur :

Le paludisme est transmis à l'homme par la piqure d'un moustique culicidé du genre anophèles au moment de son repas sanguin. Seule la femelle, hématophage, transmet la maladie. Elle ne pique qu'à partir du coucher du soleil avec un maximum d'activité entre 23 heures et 6 heures. Cela explique que l'utilisation des moustiquaires est le moyen de prévention individuelle le plus efficace. Les larves d'anophèles se développent dans les collections d'eau. La nature des sols, le régime des pluies, la température et donc l'altitude, la végétation naturelle ou l'agriculture, rendent les collections d'eau plus ou moins propices au développement des espèces vectrices. Certaines espèces ont ainsi pu s'adapter à des milieux particuliers comme le milieu urbain. Le développement et la longévité des anophèles dépendent de la température avec un optimum entre 20 et 30°C pour une durée de vie de l'ordre de 30 jours.

5. Cycle :

Le cycle se déroule successivement chez l'homme (phase asexuée chez l'hôte intermédiaire) et chez l'anophèle (phase sexuée chez l'hôte définitif). Chez l'homme le cycle est lui-même divisé en 2 phases : - la phase hépatique ou pré-érythrocytaire (= exo-érythrocytaire) : elle correspond à la phase d'incubation, cliniquement asymptomatique. - la phase sanguine ou érythrocytaire : elle correspond à la phase clinique de la maladie.

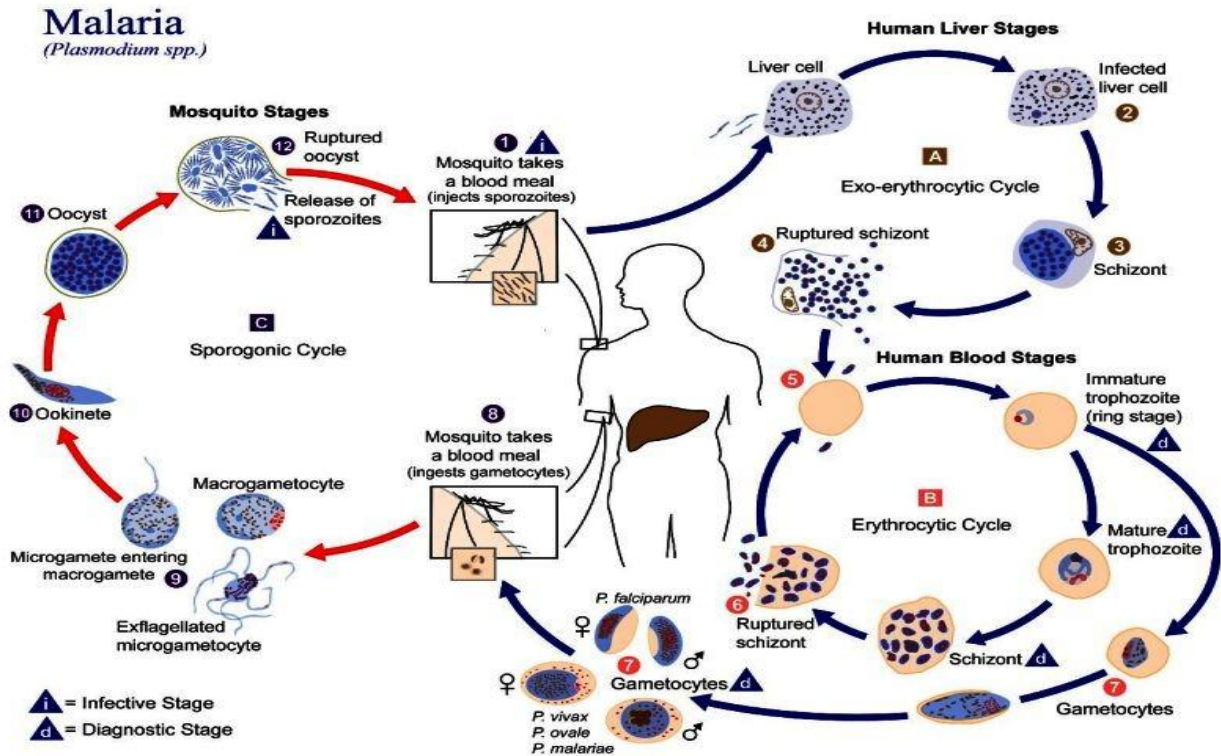


Figure I.2: Cycle de paludisme.

5.1. Chez l'homme :

5.1.1. Schizogonie pré-érythrocytaire :

Les sporozoïtes inoculés par l'anophèle femelle lors de son repas sanguin restent pendant une trentaine de minutes maximum dans la peau, la lymphe et le sang. Beaucoup sont détruits par les macrophages mais certains parviennent à gagner les hépatocytes. Ils se transforment en schizontes pré-érythrocytaires ou « corps bleus » (formes multi nucléées) qui, après 7 à 15 jours de maturation, éclatent et libèrent des milliers de mérozoïtes dans le sang (10 000 à 30 000 mérozoïtes en fonction des espèces). La schizogonie hépatique est unique dans le cycle, la cellule hépatique ne pouvant être infectée que par des sporozoïtes. Dans les infections à *P. vivax* et *P. ovale*, certains sporozoïtes intra-hépatiques restent quiescents (hypnozoïtes) et sont responsables d'une schizogonie hépatique retardée, qui entraîne la libération dans le sang de mérozoïtes plusieurs mois après la piqûre du moustique, expliquant ainsi les reviviscences tardives observées avec ces deux espèces. Les hypnozoïtes n'existent pas dans l'infection à *P. falciparum* (pas de rechute) et ils n'ont pas été mis en évidence non plus dans l'infection à *P. malariae*, malgré l'existence de rechutes tardives, ni-semble-t-il pour *P. knowlesi*.

5.1.2. Schizogonie érythrocytaire :

Très rapidement les mérozoïtes pénètrent dans les globules rouges. La pénétration du mérozoïte dans l'érythrocyte et sa maturation en trophozoïte puis en schizonte prend 24, 48 ou 72 heures (en fonction de l'espèce) et conduit à la destruction du globule rouge hôte et à la libération de 8 à 32 nouveaux mérozoïtes. Des mérozoïtes pénètrent dans de nouveaux globules rouges et débutent un nouveau cycle de réplication. Cette partie du cycle correspond à la phase clinique : la parasitémie s'élève, le sujet devient fébrile, c'est l'accès palustre. En l'absence de traitement, tous les parasites évoluent progressivement au même rythme (on dit qu'ils deviennent synchrones), tous les schizontes érythrocytaires arrivent à maturation au même moment, entraînant la destruction d'un grand nombre de globules rouges de manière périodique, toutes les 24 heures (pour *P. knowlesi*), 48 heures (fièvre tierce de *P. falciparum*, *P. vivax* ou *P. ovale*) ou toutes les 72 heures (fièvre quarte de *P. malaria*). En pratique on n'observe que la fièvre de tierce due à *P. falciparum* et rarement synchrones. Après un certain nombre de cycles érythrocytaires, certains mérozoïtes subissent une maturation d'une dizaine de jours, accompagnée d'une différenciation sexuée : ils se transforment en gamétocytes à potentiel mâle ou femelle, qui vont rester en circulation dans le sang pendant 10 à 15 jours.

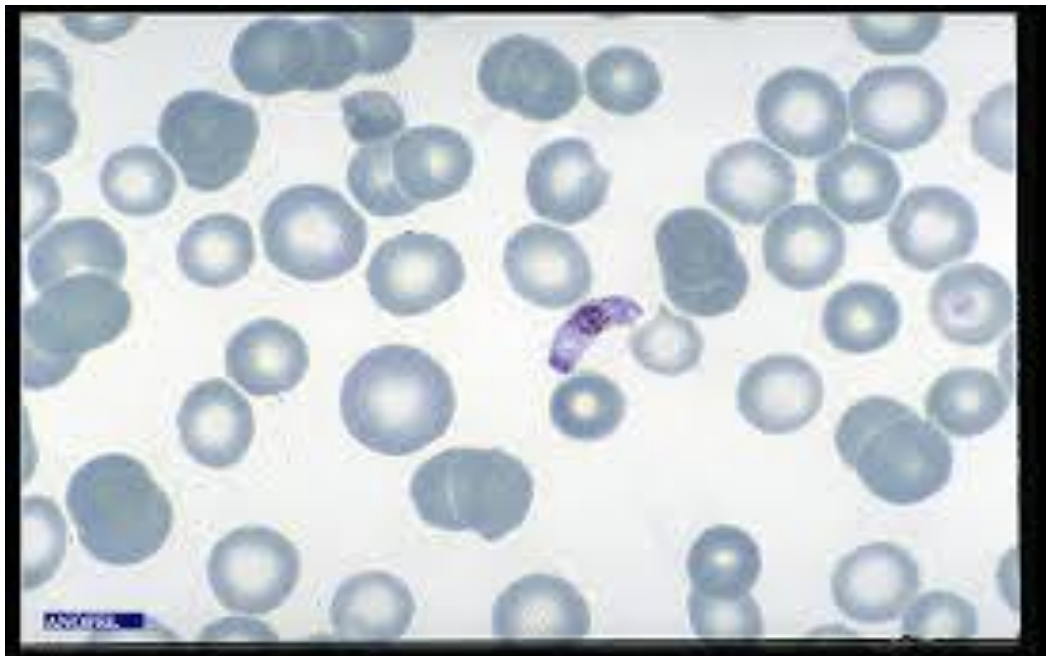


Figure I.3: Gamétocyte femelle de *P. falciparum*.

CHAPITRE I : Rappels Bibliographique Sur Les Antipaludiques

5.2. Chez l'anophèle femelle :

Les gamétocytes, ingérés par le moustique lors d'un repas sanguin sur un sujet infecté, se transforment en gamètes mâles et femelles qui fusionnent en un œuf libre, mobile appelé ookinète. Cet ookinète quitte la lumière du tube digestif, se fixe ensuite à la paroi externe de l'estomac et se transforme en oocyste. Les cellules parasites se multiplient à l'intérieur de cet oocyste, produisant des centaines de sporozoïtes qui migrent ensuite vers les glandes salivaires du moustique. Ces sporozoïtes sont les formes infectantes prêtes à être inoculées avec la salive du moustique, lors d'un repas sanguin sur un hôte vertébré. La durée du développement sporogonique des Plasmodium varie en fonction des conditions climatiques : entre 9 et 20 jours pour *P. falciparum* (entre, respectivement, 30°C et 20°C), un peu plus rapide pour *P. vivax* à températures équivalentes, plus long pour *P. malaria*. (Figure I.3).

5.3. Modalités de transmission :

La connaissance du cycle du paludisme permet de comprendre les modalités de transmission de la maladie. Le paludisme est transmis, pendant la nuit, par la piqûre d'un moustique, l'anophèle femelle (figure I.4). La phase sanguine du cycle rend possible d'autres modes de contamination : transmission congénitale, transfusionnelle, par greffe d'organe ou transmission accidentelle chez des personnels de santé manipulant du sang contaminé. En pratique ces transmissions sont tout à fait exceptionnelles et n'influencent pas l'épidémiologie.



Figure I.4: Femelle du genre Anophèles se gorgeant.

II. 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 obtenus par hémi synthèse qui occupent actuellement une place primordiale lors de paludismes dits récessifs et neurologiques.

1. Classification des antipaludiques :

1.1. Les modes d'action des principaux antipaludiques actuels (2, 3) :

Les principaux antipaludiques actuels peuvent être classés selon leur mode d'action en deux catégories :

1.1.1. Les schizonticides électifs :

Ce groupe comprend les dérivés quinoléiques et les dérivés de l'artémisinine.

- **Les dérivés quinoléiques** : ce sont les amino-4-quinoléines (chloroquine, amodiaquine) et les amino-alcools (quinine, méfloquine, halofantrine, luméfántrine). Ces molécules interférentes avec l'utilisation de l'hémoglobine dans la vacuole nutritive en inhibant la formation de l'hémozoïne.
- **Les dérivés de l'artémisinine (artésunate, artéméther)** : cette nouvelle classe d'antipaludiques de type peroxyde interfère aussi avec l'utilisation de l'hémoglobine, par libération de radicaux libres, toxiques pour le parasite. Les dérivés de l'artémisinine ont une action gamétocytocide, qui réduit la transmission et limite les risques de voir émerger des résistances.

1.1.2. Les inhibiteurs des acides nucléiques :

Ou anti métabolites qui bloquent la division du noyau de l'hématozoaire. Ce groupe comprend les batifolates, les naphtoquinones et les antibiotiques.

- **les antifolates** : ils sont répartis en deux familles, l'anti foliques (sulfamides, dont la sulfadoxine ; sulfones), et les antifoliniques (proguanil, pyriméthamine). Ils agissent au niveau de la voie de synthèse des folates, qui sont essentiels à la biosynthèse des acides nucléiques. Les antifoliques inhibent la dihydroptéroate synthétase (DHPS) qui

CHAPITRE I : Rappels Bibliographique Sur Les Antipaludiques

produit l'acide folique, les antifoliniques inhibent la dihydrofolate réductase (DHFR) qui produit l'acide folinique.

- **Les naphthoquinones** : l'atovaquone est un inhibiteur puissant des fonctions mitochondriales en bloquant la chaîne de transfert d'électrons au niveau de son enzyme-clé, la dihydroorotatedéshydrogénase (DHODase). Elle a peu d'impact thérapeutique lorsqu'elle est utilisée seule. En combinaison avec un antimétabolite (proguanil), on observe une intéressante synergie d'action grâce à une inhibition séquentielle de la synthèse des pyrimidines. Une originalité de l'association atovaquone-proguanil est son action sur les stades hépatocytaires de *P. falciparum*.
- **Les antibiotiques** : les tétracyclines (doxycycline), les macrolides (érythromycine, azythromycine, clindamycine) peuvent inhiber la synthèse protéique par inhibition de certaines fonctions de l'apicoplaste.

1.1.3. Les nouvelles thérapeutiques antipaludiques : perspectives

Les différents éléments des trophozoïtes de *P. falciparum* sont les cibles des nouvelles molécules en cours d'évaluation. Elles agissent : [2]

- Au niveau de la vacuole digestive par inhibition des protéinases vacuolaires qui participent à la protéolyse de l'hémoglobine.
- Au niveau du cytosol par interaction avec le métabolisme du fer : des chélateurs du fer, comme la des ferrioxamine, ont montré leur capacité de diminuer la croissance parasitaire ; la ribonucléotide réductase (RNRase), enzyme qui règle la synthèse d'ADN, pourrait être la cible de la des ferrioxamine.
- au niveau de la membrane plasmique parasitaire par blocage du transporteur de choline qui fournit au parasite un précurseur indispensable pour la synthèse de la phosphatidylcholine, principal phospholipide plasmodial, les phospholipides étant les constituants majeurs des membranes du plasmodium.
- D'autre part, la réversion de la chloroquinorésistance chez *P. falciparum* par des molécules de synthèse et gènes associés aux mécanismes de résistance (Pfcr1 et Pfmdr1) a pour objectif de freiner la fuite de la chloroquine hors de l'hématie parasitée (inhibiteurs calciques vérapamil-like).
- Au niveau de l'apicoplaste, d'autres antibiotiques pourraient être actifs : les fluoroquinolones, qui bloquent des topoisomérases spécifiques et la fosmidomycine, qui est un inhibiteur de la synthèse des isoprénoïdes. [3.4]

Autres antipaludiques et associations antipaludiques :

a. Les aminoalcools :

La méfloquine appartient à ce groupe, elle est active contre les souches multi résistantes de plasmodium et notamment les chloroquino-résistantes. [1]

Leur mécanisme d'action est similaire à celui de la quinine et à la chloroquine. Les effets indésirables sont la psychose aiguë, l'encéphalopathie transitoire avec des convulsions. [5]

b. Quelques Associations antipaludiques :

-Sulfadoxine-Pyriméthamine (Fansidar*)

-Dapsone-Pyriméthamine (Maloprime*)

-Sulfaméthoxazole-Triméthoprime
(Bactrim*)

-Sulfadoxine-Pyriméthamine-Méfloquine
(Fansimef*)

-Quinine-Tétracycline

-Artésunate-Amodiaquine

-Quinine-Doxycycline

-Artéméther-Lumefantrine (Coartem*)

-Quinine Clindamycine

-Proguanil-Atovaquone (Malarone*)

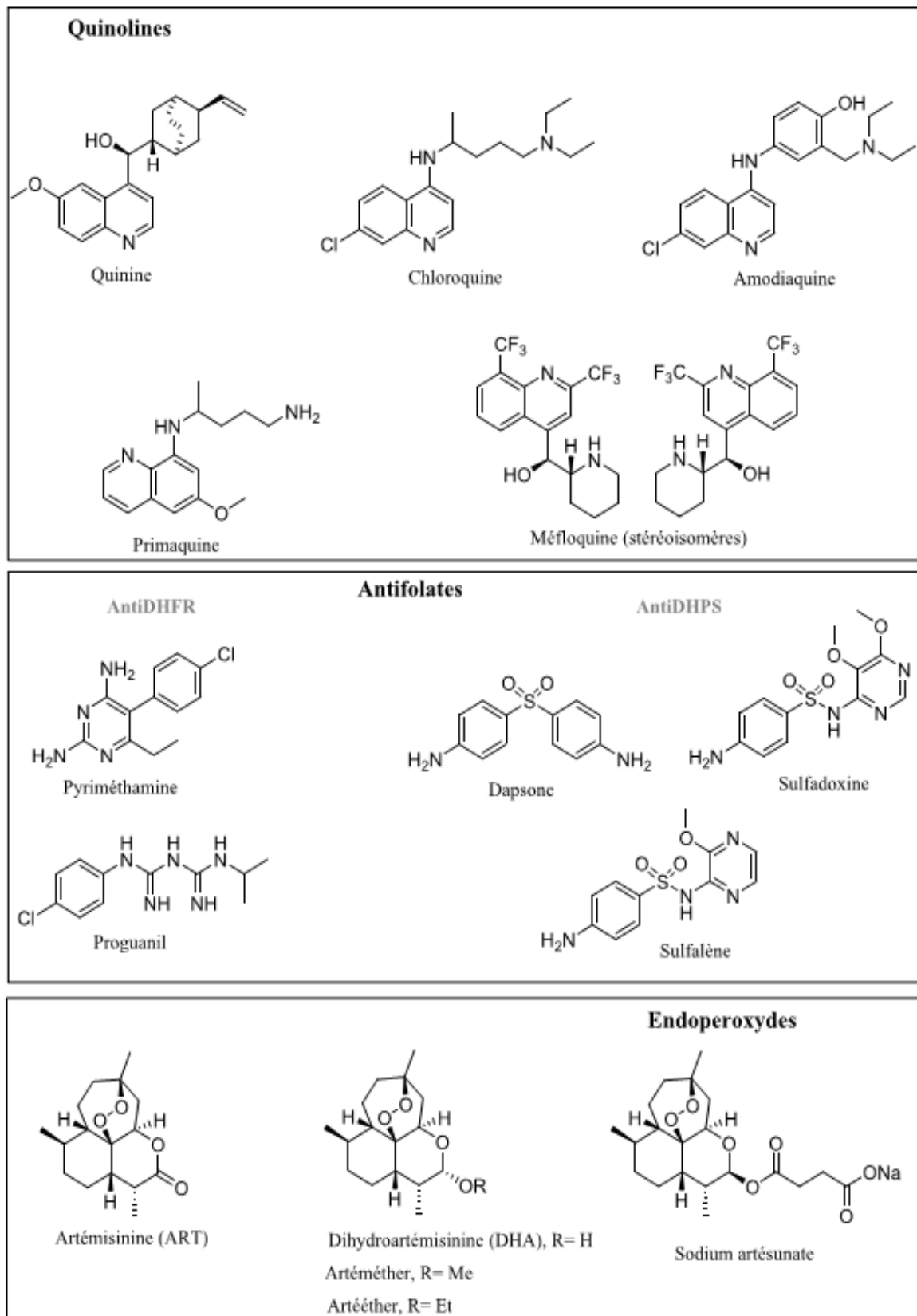


Figure I.5: Structure de quelques molécules utilisées en tant que médicaments antipaludiques.

III. La protéine farnésyltransférase :

De très nombreuses protéines jouant un rôle essentiel dans de nombreux processus cellulaires (croissance, différenciation, morphologie), requièrent, pour être activées, une, voire plusieurs modifications post-traductionnelles. Il existe une centaine de modifications post-traductionnelles, toutes catalysées par des enzymes. L'inhibition de l'une de ces enzymes permettrait d'inactiver leurs protéines substrats, induisant un effet antiprolifératif. Dans le milieu des années 1980 [8].

Cette modification post-traductionnelle consiste à l'ajout d'un groupement farnésyle (15 atomes de carbone) ou géranylgeranyle (20 atomes de carbone) à des protéines par l'intermédiaire de la farnésyltransférase (FTase) ou la géranylgeranyltransférase (GGTase). La prénylation permet la reconnaissance protéine-protéine et la fixation de protéines dans la membrane. Ces protéines peuvent être des protéines G hétérotrimériques et les lamines nucléaires. La famille la plus importante et la plus étudiée pour leur rôle dans la prolifération cellulaire, est la famille des petites protéines G (Ras, Rheb, CENP-E, -F, RhoB...). Les inhibiteurs de FTase (FTis) ont été initialement conçus pour bloquer l'oncoprotéine Ras (mutée dans 30% des cancers), mais leur mode d'action est plus complexe [9].

1. La structure de la Ftase :

La structure de la FTase varie selon les espèces mais certaines caractéristiques restent identiques quels que soient les individus [10]. L'enzyme de rat mais aussi celle de l'Homme a été cristallisée avec ou sans substrats délivrant de nombreuses informations sur sa structure tridimensionnelle. Elle est constituée de deux sous-unités. La sous-unité α (48 kDa chez les mammifères, 65 kDa chez les parasites) est composée de quinze hélices α , de quatre hélices β et d'un feuillet β , le tout organisé en forme de croissant enveloppant partiellement la sous-unité β .

Cette dernière (46 kDa chez les mammifères, 61 kDa chez les parasites) structurée par quatorze hélices α organisées en forme de tonneau α - α , sept petites hélices β et trois feuillets β . Le site actif de l'enzyme situé à la jonction des deux sous-unités comporte un atome de zinc indispensable à l'activité enzymatique (Figure I.6) [10].

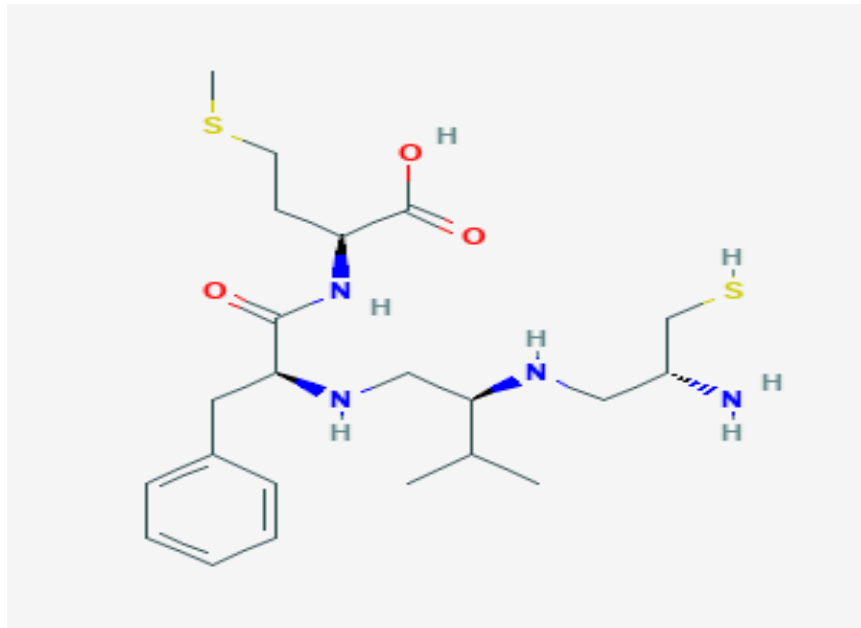


Figure I.6 : La structure développée de la protéine Ftase.

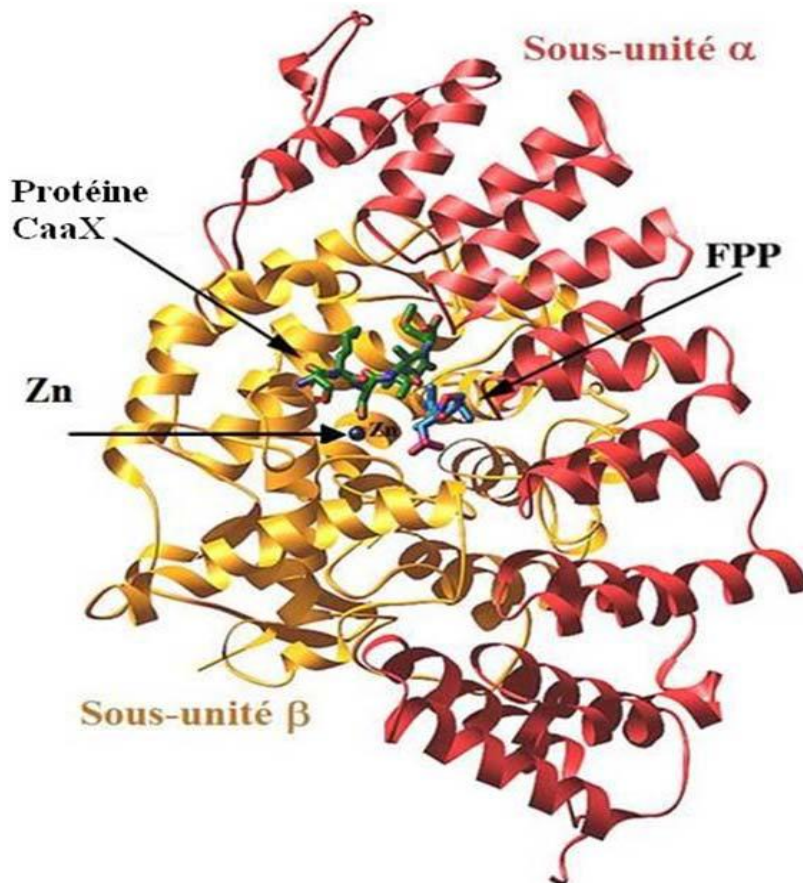


Figure I.7: Structure de la FTase.

Plusieurs études mécanistiques et cristallographiques à haute résolution sur la FTase complexées avec différents substrats, produits et inhibiteurs ont été réalisées dans le but de comprendre le mécanisme catalytique de cette enzyme. Ces études ont montré que l'enzyme

de rat et de l'Homme possède 97 % d'homologies. Ainsi, les données obtenues à partir de la FTase de rat peuvent être extrapolées à l'enzyme humaine.

Les enzymes de parasites tels que *Trypanosomabrucei* (maladie du sommeil), *Plasmodium falciparum* (paludisme) ou *Leishmania major* (leishmaniose) plus volumineuses que celles des mammifères ont montré certaines différences. Dans la séquence linéaire, il a été remarqué des insertions d'acides aminés entraînant des changements conformationnels ayant peu voire pas d'influence sur l'activité catalytique de l'enzyme. Au niveau du site actif, les acides aminés qui constituent la poche hydrophobe dans laquelle se positionne le FPP ainsi que ceux qui entourent le motif Caa du motif CaaX sont identiques chez le rat et le parasite. Cependant, il existe quelques changements dans la séquence qui entoure le site de positionnement de X augmentant entre autre l'hydrophobicité. Ainsi, alors que la FTase de mammifères reconnaît les protéines qui contiennent une sérine, une alanine, une méthionine ou une glutamine, l'enzyme de parasite n'identifie que celles possédant une méthionine ou une glutamine préférentiellement. La différence structurale de ces FTases permet donc d'envisager favorablement le développement d'inhibiteurs de la FTase sélectifs du parasite.

2. Le mécanisme catalytique de la FTase :

La farnésylation fait intervenir deux substrats qui sont le pyrophosphate de farnésyle (FPP), donneur de l'unité farnésyle et le motif CaaX subissant la farnésylation (Schéma I.1).

Il est maintenant clairement accepté que le mécanisme est fonctionnellement ordonné :

- Le FPP se lie à l'enzyme E pour former le complexe binaire E-FPP(1). [9]
- Le substrat CaaX est reconnu et peut se fixer dans le site actif pour former le complexe ternaire réactif (2).
- le soufre de la cystéine va se lier au carbone C1 du FPP formant le produit farnésyle (3) et libérant une molécule de pyrophosphate (PPi).
- Le produit farnésyle est déplacé par une nouvelle molécule de FPP vers une poche de sortie (4). Cette libération constitue l'étape limitant de cette catalyse enzymatique [9].
- Après la sortie du produit farnésyle, le complexe binaire (1) existe en équilibre avec la FTase seule (0).

CHAPITRE I : Rappels Bibliographique Sur Les Antipaludiques

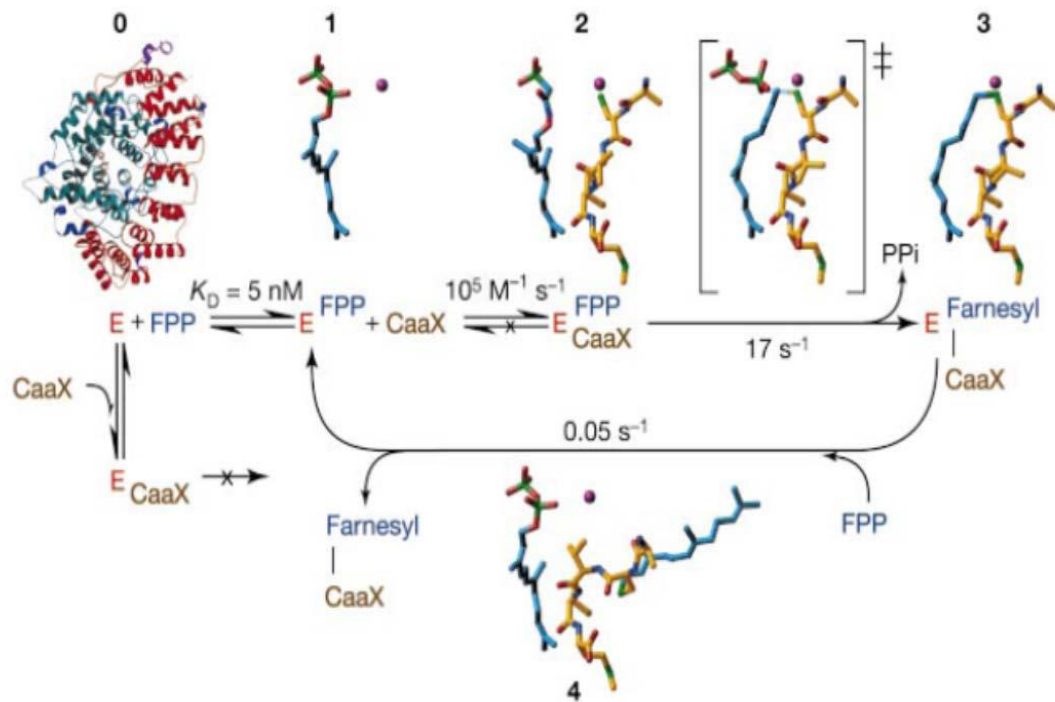


Schéma I.1: Structures au cours de la cinétique de la FTase.

0 : enzyme seule, **1 :** enzyme liée au FPP (en bleu), **2 :** enzyme liée au FPP et au motif CaaX (en jaune), **3 :** enzyme contenant le peptide farnésyle dans le site actif, **4 :** rélargie du produit farnésyle.

Au cours de la catalyse enzymatique, certaines caractéristiques de la FTase varient légèrement. La structure secondaire de la FTase reste inchangée mais il existe en revanche des différences conformationnelles entre l'enzyme non liée et l'enzyme liée aux deux substrats dans le complexe ternaire [9]. La sous-unité β reste majoritairement inchangée alors que la sous-unité α subit la plupart des altérations conformationnelles. Plusieurs structures cristallines de la FTase ont été décrites en présence de petits peptides substrats CaaX et d'analogues du FPP permettant de mettre en évidence deux poches de nature très différente correspondant aux sites de liaison des deux substrats. Ces lieux de fixation se rencontrent au niveau du site actif de l'enzyme où est localisé l'atome de zinc. Ainsi, les études ont montré que le site de liaison de la queue prényle de l'unité farnésyle se trouve sur la sous-unité β au centre de la région hydrophobe. Sa tête diphosphate se lie avec une poche chargée positivement proche de l'interface des deux sous-unités. La protéine CaaX, quant à elle, adopte une conformation allongée dans sa partie C-terminale lors de la formation du complexe ternaire et s'étend dans la cavité du site actif. Elle interagit à la jonction des deux sous-unités α et β avec l'enzyme et avec la chaîne isoprényle du FPP.

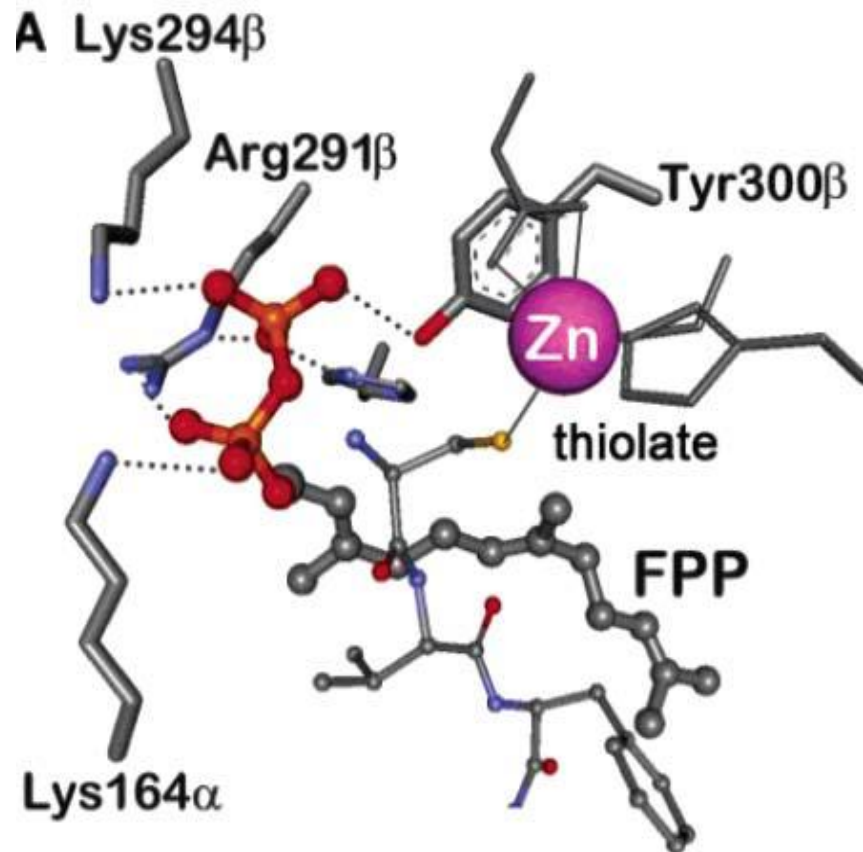


Figure I.8: Liaisons du FPP avec l'enzyme en présence du motif CaaX.

Ces considérations structurales et mécanistiques sont maintenant acceptées mais d'autres aspects sont encore discutés.

Premièrement, dans le complexe ternaire E-analogue FPP peptide, il a été observé que la distance entre le soufre de la cystéine et le carbone du pyrophosphate est de 7,3 Å. Cette distance, trop importante pour que le soufre puisse être alkyle, laisse penser qu'elle doit être réduite au cours de la réaction enzymatique par un déplacement d'au moins un des deux substrats.

Trois hypothèses ont été envisagées et consistent à :

- Un réarrangement conformationnel de l'enzyme. Cette hypothèse est peu fiable étant donné que les seuls réarrangements observés sont très localisés.
- Une dissociation du peptide de l'atome de zinc permettant au motif CaaX de s'approcher du FPP.
- Une rotation entre les deux dernières unités isoprényles qui permettrait de rapprocher le carbone C1 du FPP près du soufre de la cystéine pour atteindre une distance de 2,4 Å (figure I.8).

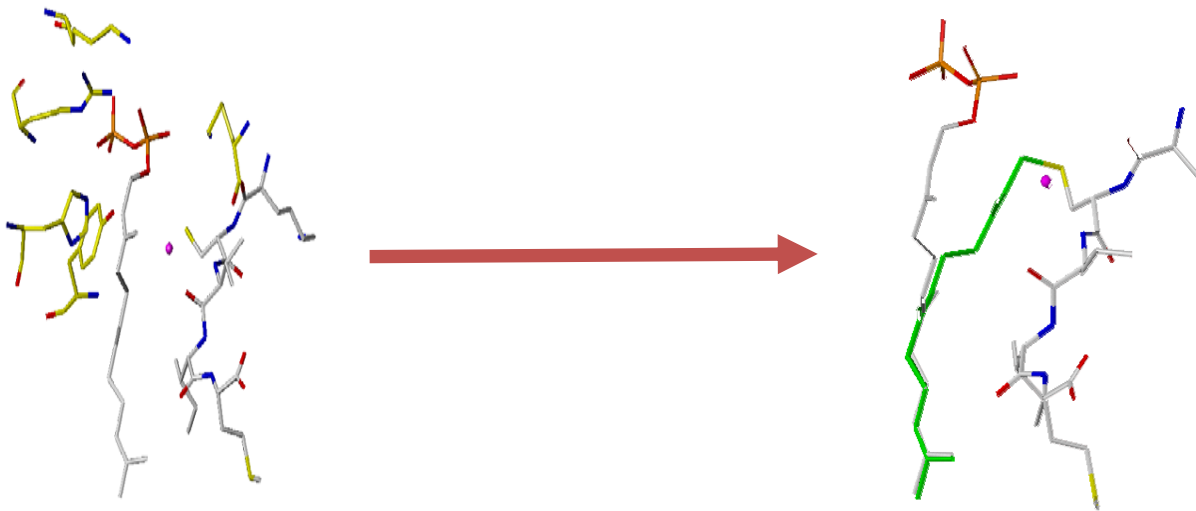


Figure I.9: Rotation des deux dernières unités isoprényles du FPP pour la farnésylation.

Un autre point de controverse concerne la nature de la réaction chimique entre le soufre de la cystéine et le carbone C1 du FPP. Deux hypothèses sont confrontées. La première est un mécanisme électrophile (SN1 ou dissociatif) où un carbocation se formerait au niveau du C1 du FPP. La deuxième hypothèse est un mécanisme nucléophile (SN2 ou associatif) où le thiolate de la cystéine attaquerait directement le C1 du FPP. Pour le moment aucune étude n'a clairement démontré la suprématie de l'une de ces théories car il existe des arguments en faveur de l'une ou l'autre de ces hypothèses.

Une interprétation unifiée existe dans laquelle des charges partielles interviendraient. On parle de mécanisme associatif à caractère dissociatif.

Références bibliographiques :

- [1] CURTIS, SUTTER, WALKER et HOFFMAN, Editions De Boeck, PHARMACOLOGIE INTEGREE De Boeck 2001.
- [2] J.E. Touze, L. Fourcade, B.Pradines, P. Hovette, P.Paule, Ph.Heno, Les modes d'action des antipaludiques. Intérêt de l'association atovaquone-proguanil. *Med. Trop*, 62, 2002, 219-222.
- [3] B.Pradines, H.Vial, P.Olliaro. Prophylaxie et traitement du paludisme : problèmes, récents développements et perspectives. *Med. Trop*, 63, 2003, 79-98.
- [4] M.Danis. Avancées thérapeutiques contre le paludisme en. *Med. Trop*, 63, 2003, 267-270.
- [5] M. GENTILLINI, MEDECINE TROPICALE 5^{ème} édition J.A. Glom set, M.H. Gelb ,C.C. Farnsworth, "Prenyl proteins in eukaryotic cells: A new type of membrane anchor." *Trends Biochem. SCI*, 15, 1990, 139.
- [6] Gelbet, Glomset ont mis en évidence le phénomène de prénylation dans les cellules de mammifère 1990.
- [7] I.M. AHEARN, K.HAIGIS, D.BAR-SAGI, M.R.PHILIPS, Regulating the regulator: post-translational modification of RAS. *Nature Reviews Molecular Cell Biology*, 13, 2012, 39-51.
- [8] M.SHEN, P.PAN, Y.LI, D.LI, H.YU, T.HOU, 2015 Farnesyltransferase and geranyl geranyltransferase I: structures, mechanism, inhibitors and molecular modeling. *Drug Discovery Today*, 20(2), 2015, 267–276.
- [9] S.B. LONG, P.J.CASEY, L.S.BEESE, Reaction path of protein farnesyltransferase at atomic resolution. *Nature*, 2002, 419-645.
- [10] S.B.LONG, P.J.CASEY, L.S.BEESE, 1998.-Co-crystal structure of protein farnesyl transferase with a farnesyl diphosphate substrate. *Biochemistry*, 37 (27), 1998, 9612–9618.



CHAPITRE II

***ETUDE QUANTITATIVE DES RELATIONS
STRUCTURE-ACTIVITÉ (QSAR)***

II.1.Introduction :

Les relations quantitatives structure-activité/propriété (QSAR/QSPR) sont de plus en plus utilisées, du fait de la croissance des moyens de calculs. Très récemment, la mise en place du nouveau règlement européen REACH, qui recommande leur utilisation pour limiter le recours à l'expérience, donne un nouvel essor au développement de tels modèles prédictifs. Dans les dernières années, l'utilisation des méthodes QSAR n'a cessé de progresser. Elle est même devenue indispensable en chimie pharmaceutique et pour la conception de médicaments. Leur développement dans une gamme plus large d'applications leur ouvre d'ailleurs de grandes perspectives (ex : solubilité, points d'ébullition, températures critiques, densité . . . etc.). Il s'agit de présenter ici le principe des modèles QSPR ainsi que ceux des différents outils employés pour leur mise en place et leur évaluation : bases de données expérimentales, descripteurs, outils d'analyse de données [1].

II.2.Historique :

Il y a plus d'un siècle et demi, en 1863, Crois [2] a observé que le point d'ébullition et le point de fusion des alcanes augmente avec le nombre d'atomes de carbone et la masse moléculaire. Il a observé également une diminution de la solubilité dans l'eau des alcools avec l'augmentation du nombre d'atomes de carbone et la masse moléculaire, cela est considéré depuis comme la première formulation générale en QSAR.

Cinq ans après, en 1868, Crum – Brown et Fraser [3] postulèrent que « l'activité biologique d'une molécule est une fonction de sa constitution chimique ».

Quelques décennies plus tard, en 1893, Richet [4] a montré que la cytotoxicité de certains composés organiques était inversement proportionnelle à leur solubilité dans l'eau. A la fin du 19ème siècle, Meyer en 1899 et Overton en 1901 [5-7], ont indépendamment observé « une relation linéaire entre l'activité des narcotiques et leur coefficient de partage huile-eau ».

Six ans après, en 1907, Fühner et Neubauer [8] ont montré pour une série de narcotiques homologues, que l'activité augmentait en fonction de la progression géométrique de la série de composés, ceci montrant l'importance de la contribution d'additivité de groupements fonctionnels pour l'activité biologique.

En 1962, Hansen [9] a montré l'existence d'une corrélation entre la toxicité des acides benzoïques substitués et les constantes électroniques « σ » des substituants.

L'année 1964 est considérée comme le début des méthodes QSAR modernes. Hansch et Fujita ont établi les premières corrélations entre les propriétés physico-chimiques (log P, PKa, paramètres stériques et électroniques) et l'activité biologique (activité enzymatique, pharmacologique). Ces méthodes seront appelées par la suite l'analyse de Hansch et l'analyse de Free Wilson) [10-11]. Sept ans plus tard, Hansch et Lien ont réalisé une étude QSAR sur différentes familles d'antifongiques : benzoquinones, sels d'alkylpyridinium, imidazoles et phénols. Ils ont observé que quels que soient la famille et le champignon utilisé, l'activité antifongique dépend du coefficient de partage Eau-Octanol, expérimental ou calculé [12].

Ces études ont été extrapolées aux techniques séparatives en corrélant les propriétés physico-chimiques des analytes avec les temps de rétention obtenus expérimentalement : c'est l'étude quantitative des relations structure temps de rétention noté QSAR [13].

II.3. Définition :

Une **Relation Quantitative Structure à Activité** (en anglais : *Quantitative Structure-Activity Relations* hipou **QSAR**) est le procédé par lequel une structure chimique est corrélée avec un effet bien déterminé comme l'activité biologique ou la réactivité chimique. Ainsi, l'activité biologique peut être exprimée de manière quantitative, comme pour la concentration de substance nécessaire pour obtenir une certaine réponse biologique. De plus lorsque les propriétés ou structures physicochimiques sont exprimées par des chiffres, on peut proposer une relation mathématique, ou *Relation Quantitative Structure à Activité*, entre les deux. L'expression mathématique obtenue peut alors être utilisée comme moyen prédictif de la réponse biologique pour des structures similaires.

La **QSAR** la plus commune est de la forme : activité = f (propriétés physico-chimiques et/ou structurales).

Pour faciliter l'application pratique des approches (**Q**)**SAR** dans des contextes réglementaires par les autorités et l'industrie et pour améliorer leur acceptation réglementaire, le projet (**Q**)**SAR** de l'**OCDE** a développé divers produits tels que les principes pour la validation de modèles (**Q**)**SAR**, des documents d'orientation ainsi que la Boîte à outils **QSAR**[14].

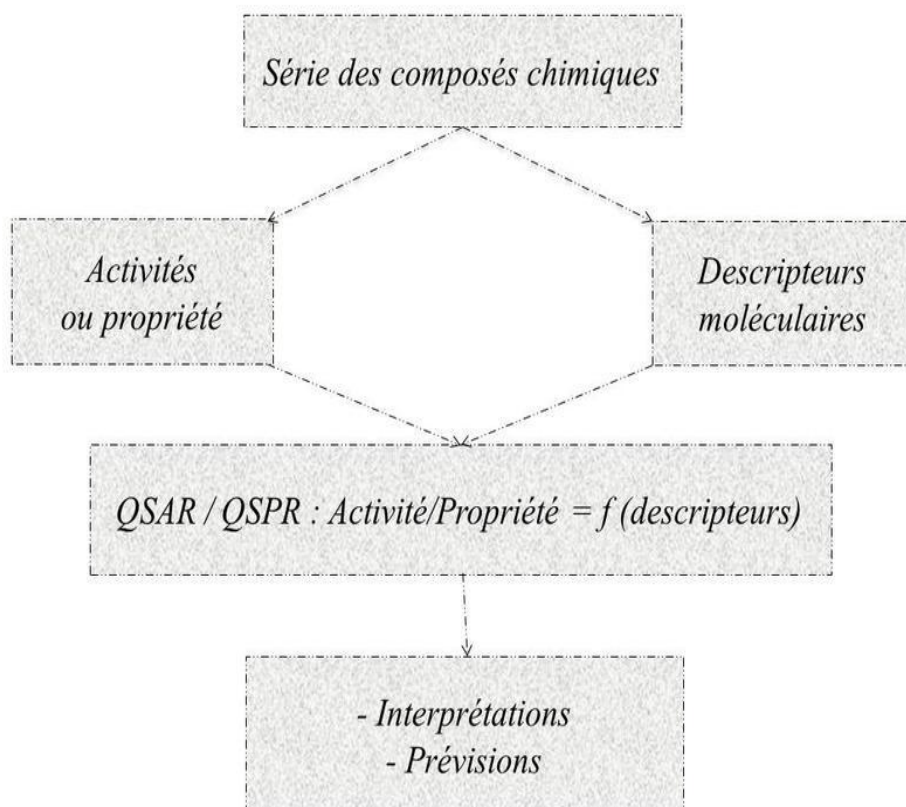


Schéma II .1 : Modèle d'étude quantitative des relations structure/activité (QSAR).

II.4.Principe de QSAR :

Le principe des méthodes QSPR/QSAR est d'établir une relation mathématique reliant de manière quantitative des propriétés moléculaires, appelées descripteurs, avec une observable macroscopique (activité biologique, toxicité, propriété physico-chimique, etc.), pour une série de composés chimiques similaires à l'aide de méthodes d'analyses de données. La forme générale d'un tel modèle est la suivante :

$$\text{Propriété/ Activité} = f (\mathbf{D1}, \mathbf{D2}, \dots \mathbf{Dn},)$$

D1, D2, ...Dn sont des descripteurs des structures moléculaires.

L'objectif d'une telle méthode est d'analyser les données structurales afin de détecter les facteurs déterminants pour la propriété /activité mesurée. Pour ce faire, différents types d'outils statistiques peuvent être employés :

- ✓ Régressions linéaires multiples (MLR).
- ✓ Régressions aux moindres carrées partielles (PLS).
- ✓ Analyse par composantes principales (ACP).

- ✓ Régressions non linéaire multiple (MNLR).

II.5.Méthodologie générale d'une étude QSPR/QSAR :

II.5.1. Les bases de données :

Un modèle QSPR, de par sa construction, est très dépendant des données expérimentales de référence; le choix de la base de données est donc un point critique de son développement.

Dans la plupart des cas, les données expérimentales sont issues de la littérature. Pour être de qualité, une base de données doit être composée de données expérimentales aussi fiables que possible, puisque les barres d'erreurs sur celles-ci se propageront dans le modèle final, étant donné que les paramètres de ce dernier sont ajustés par rapport à ces données. Il est donc important de choisir des données présentant des incertitudes faibles afin de limiter les barres des erreurs expérimentales. En effet, un modèle ne pourra être plus robuste statistiquement que les données théoriques à partir desquelles il a été développé.

Les modèles QSPR sont généralement plus efficaces pour un jeu de molécules similaires. Pareillement, il est important de considérer une base de données dont les molécules se comportent de manière similaire vis-à-vis de la propriété ciblée. De même, afin de réduire la probabilité de rencontrer des molécules très différentes de la majorité des autres, qui peuvent peser significativement sur le modèle et accroître l'incertitude sur les prédictions (Outliers) dans le jeu de données, plus la diversité structurale d'une base de données est importante et plus la taille de la base de données doit être importante [15].

II.5.2. Les Descripteurs moléculaires :

L'activité biologique ne peut pas être reliée directement à la structure moléculaire ; cette dernière est codée par des grandeurs qui représentent d'une manière quantitative les informations contenues dans la structure moléculaire telles que les caractéristiques physicochimiques et structurelles. Ces grandeurs sont appelées descripteurs. Une fois ces derniers disponibles, il est possible d'établir des relations entre ceux-ci et une activité à l'aide des outils de la modélisation [16]. On en dénombre aujourd'hui des milliers, qui peuvent être calculés ou obtenus de manière empirique. De nombreux logiciels ont été développés pour calculer les différents descripteurs moléculaires tels que : *Gaussian*, *ChemOffice*, *ChemSketch*, *Marvin Sketch*, *Dragon*, *QSARIS*.

Dans ce qui suit, nous allons présenter que les descripteurs moléculaires qui ont été utilisés dans l'ensemble de nos travaux.

⇒ **Les descripteurs 0-D** : Ils sont dérivés de la liste d'atomes de la molécule, comme le poids moléculaire.

⇒ **Les descripteurs 1-D** : Ils sont calculés à partir de listes de sous structures de la molécule, comme les groupements fonctionnels, les cycles présents, etc. [17].

⇒ **Les descripteurs 2-D** : Ils sont obtenus à partir des représentations des molécules sous forme de graphes bidimensionnelles [18].

⇒ **Les descripteurs 3-D** : Ils sont calculés à partir de la représentation tridimensionnelle de la molécule, on cite parmi eux, les descripteurs électroniques (moment dipolaire, l'électronégativité, ...), le volume moléculaire, la surface moléculaire superficielle...etc. [18-21].

⇒ **Les descripteurs 4-D** : Obtenus par le calcul des champs d'interactions moléculaire entre deux molécules.

II.6.Méthodes d'analyse de données :

La mise en place des modèles QSAR n'est pas une tâche aisée. La première difficulté réside dans la différence d'échelles existant entre les données à corrélérer, la structure étant à une échelle moléculaire alors que les propriétés à prédire sont à une échelle macroscopique. De plus, on doit tenir compte des problèmes d'incertitudes à la fois au niveau des structures moléculaires (liées niveau de calcul) et des données expérimentales (protocoles de mesures). Le traitement d'une grande quantité de données représente aussi une difficulté majeure dans le processus de mise au point de modèles QSAR. Les corrélations entre un grand nombre de descripteurs d'un grand nombre de molécules avec leur activité peuvent être à analyser. Cependant aucune règle stricte n'existe quant au choix des paramètres structuraux les plus importants parmi le jeu complet de ceux disponibles [22-23].

En fait, de nombreux outils existent pour trouver le moyen le plus adapté pour obtenir un modèle fiable à partir des données disponibles. Selon les cas, plusieurs approches sont envisageables, il faut alors choisir celle permettant de caractériser au mieux le système. Les méthodes de modélisation moléculaire sont des outils informatiques qui permettent d'explorer les relations entre les descripteurs de molécules calculés à partir des structures chimiques, et des propriétés ou des activités déterminées expérimentalement.

II.6.1. Analyse par composantes principales :

L'analyse en composantes principales (ACP), ou *principal component analysis*(PCA) en anglais, qui est sans doute la technique exploratoire la plus répandue pour décrire les données d'entrée, rechercher d'éventuels aberrations et à connaître les corrélations entre les variables

d'entrée. Elle permet d'analyser et de visualiser un jeu de données concernant des individus décrits par plusieurs variables quantitatives. Elle est non supervisée, c'est-à-dire sans phase d'apprentissage [24-26].

C'est une méthode statistique qui permet d'explorer des données dites multi variées (données avec plusieurs variables). Chaque variable pourrait être considérée comme une dimension différente.

Si vous avez plus de 3 variables dans votre jeu de données, il pourrait être très difficile de visualiser les données dans un "hyper-espace" multidimensionnel.

L'analyse en composantes principales est utilisée pour extraire et de visualiser les Informations importantes contenues dans une table de données multi variées. L'ACP Synthétise cette information en seulement quelques nouvelles variables appelées composantes Principales.

Ces nouvelles variables correspondent à une combinaison linéaire des variables originales.

Le nombre de composantes principales est inférieur ou égal au nombre de variables d'origine. L'information contenue dans un jeu de données correspond à la variance ou l'inertie totale qu'il contient. L'objectif de l'ACP est d'identifier les directions (*axes principaux* ou composantes principales) le long desquelles la variation des données est maximale.

En d'autres termes, l'ACP réduit les dimensions de données multi variées à deux ou trois composantes principales, qui peuvent être visualisées graphiquement, en perdant le moins possible d'information. La compréhension des détails de l'ACP nécessite des connaissances de l'algèbre linéaire. Des données.

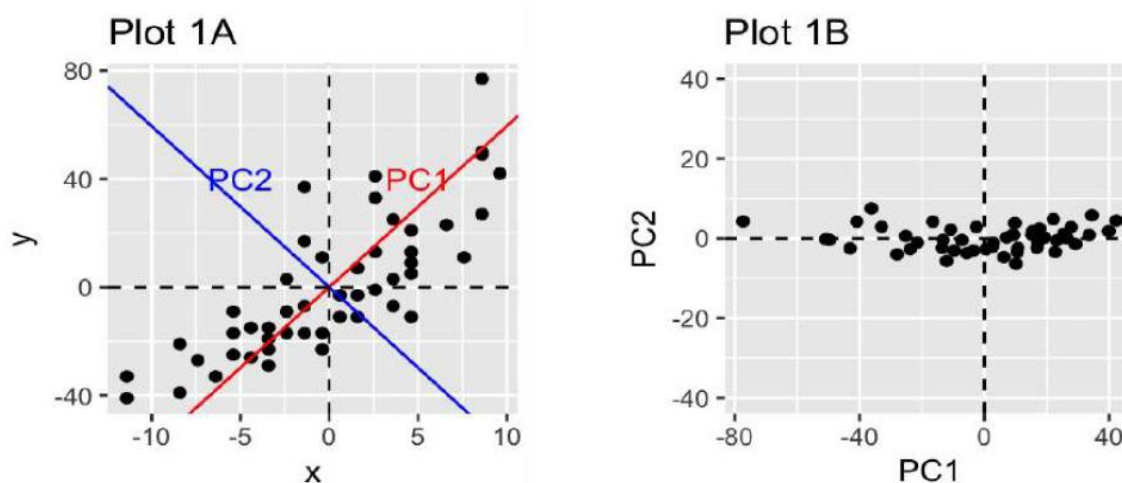


Figure II.1 : Une représentation graphique simple des bases de l'ACP. Dans le Plot 1A ci dessus, les données sont représentées dans le système de coordonnées XY.

La réduction de la dimension est obtenue en identifiant les directions principales, appelées composantes principales, suivant les quelles les données varient. L'ACP suppose que les directions avec les plus grandes variances sont les plus "importantes" (principales).

Dans la figure ci-dessus, l'axe PC1 est le premier axe principal le long duquel les échantillons présentent la plus grande variation. L'axe PC2 est la seconde direction la plus importante et orthogonal à l'axe PC1.

Les dimensions de notre jeu de données peuvent être réduites à une seule dimension en projetant chaque l'ensemble des données sur le premier axe principal (Plot 1B) [27-28].

L'ACP est particulièrement utile lorsque les variables, dans le jeu de données, sont fortement corrélées. La corrélation indique qu'il existe une redondance dans les données. En raison de cette redondance, l'ACP peut être utilisée pour réduire les variables d'origine en un nombre plus petit de nouvelles variables (**composantes principales**), ces dernières expliquant la plus grande partie de la variance contenue dans les variables d'origine. 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.6.2. La régression linéaire multiple :

La régression multilinéaire (MLR, pour *Multiple Linear Regression*) [29] est la méthode la plus simple et la plus communément employée pour le développement de modèles prédictifs. Elle repose sur l'hypothèse qu'il existe une relation linéaire entre une variable dépendante y (ici, la propriété) et une série de n variables indépendantes x_i (ici, les descripteurs). L'objectif est d'obtenir une équation de la forme suivante :

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n \dots\dots(1)$$

Où sont les coefficients de la régression.

La détermination de l'équation (1) se fait alors à partir d'une base de données de p échantillons pour laquelle à la fois les variables dépendantes et la variable indépendante sont connues. Il s'agit donc de considérer un système de p équations.

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 :

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

Avec :

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

X_i: sont les variables indépendantes (explicatives) ;

i : est le nombre de variables explicatives ;

a₀ : est la constante de l'équation du modèle ;

a_i et **b_i** sont les coefficients de descripteurs dans l'équation du modèle.

II.6.4. La régression des moindres carrés partiels :

La régression par les moindres carrés partiels (PLS) est une technique qui diminue le nombre de descripteurs à un plus petit ensemble de composantes non corrélées et qui effectue la régression par les moindres carrés sur ces composantes, plutôt que sur les données initiales. La fonctionnalité PLS est particulièrement utile lorsque les descripteurs sont fortement colinéaires, ou lorsqu'il y a plus de descripteurs que d'observations et que la régression sur les moindres carrés échoue complètement ou produit des coefficients avec des erreurs élevées.

Les moindres carrés partiels (PLS) ne supposent pas que les descripteurs sont fixes, à la différence de la régression multiple. Ainsi, les mesures des descripteurs tolèrent des erreurs, ce qui signifie que l'analyse PLS gère mieux l'incertitude des mesures [30-31].

La régression par les moindres carrés partiels est essentiellement utilisée dans les secteurs chimique, pharmaceutique, alimentaire et plastique. Une application courante consiste à modéliser la relation entre des mesures spectrales (spectre du proche infrarouge, IR UV), qui comprennent de nombreuses variables souvent corrélées les unes avec les autres, et une composition chimique ou d'autres propriétés physiques et chimiques. La régression PLS met l'accent sur le développement de modèles de prévision. Par conséquent, elle n'est généralement pas utilisée pour éliminer les variables qui ne permettent pas d'expliquer la réponse.

II.7. Test de la signification globale de la régression :

- ❖ Le coefficient de détermination R^2 (coefficient de corrélation) est l'indicateur statistique le plus répandu qui évalue la part de la variance de l'activité cible expliquée par le modèle

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

Où :

Y : La valeur expérimentale de l'activité

\hat{Y} : La valeur calculée de l'activité

\bar{Y} : La valeur moyenne des valeurs calculées de l'activité.

Plus la valeur de R^2 sera proche de 1 plus les valeurs calculées et observées sont corrélées.

La valeur de R^2 dépend de la taille de l'échantillon et le nombre de variables prédictives dans le modèle. Il garde la même valeur ou augmente lors de l'ajout d'une nouvelle variable de prédiction, même si la variable ajoutée ne contribue pas à la réduction de la variance inexplicée.

- ❖ Par conséquent, un autre paramètre statistique peut être utilisé, appelé R^2 ajusté (R^2_{adj})

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

Avec : n est le nombre des variables dépendantes (les molécules) ; p est le nombre de variables indépendantes (les descripteurs) ; R^2 est le coefficient de détermination.

- ❖ La déviation standard S est un autre indicateur statistique utilisé

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

Où n est le nombre de données d'apprentissage et p le nombre de descripteurs utilisés dans la construction du modèle.

En particulier, plus la déviation standard est petite et plus la corrélation est meilleure.

- ❖ L'indice de Fisher F est également employé afin de mesurer le niveau de signification statistique du modèle, c'est-à-dire la qualité du choix des descripteurs constituant le modèle.

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

II.8. Interprétation et validation d'un modèle QSAR/QSPR :

Une fois développé, le modèle doit être valide afin de vérifier sa stabilité, son pouvoir explicatif et son pouvoir prédictif.

•Interne :

Quel que soit le but ultime du modèle QSAR, il doit être obligatoirement validé avant d'être interprété ou utilisé à fins prédictives. Il existe différentes solutions pour s'assurer de validité d'un modèle. La validation croisée LOO (leave-one-out) est un processus qui permet d'évaluer la précision prédictive d'un modèle, cette méthode basée sur le calcul de quelques

Paramètres statistiques tels que : *PRESS*, *SSY*, *Spress*, R^2_{CV} et R^2_{adj} .

$$PRESS = \sum (Y_{obs} - Y_{cal})^2$$

$$TSS = \sum (Y_{obs} - \bar{Y})^2$$

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

$$R^2_{CV} = 1 - \frac{PRESS}{TSS}$$

$$S_{PRESS} = \sqrt{\frac{PRESS}{n}}$$

Ces paramètres sont calculés à partir des équations suivantes [32] :

Avec: n : nombre de molécules de la série étudiée.

PRESS: Somme des carrés des écarts entre les valeurs observées et les valeurs prévues à partir du modèle.

SSY: Somme des carrés des distances des valeurs observées pour une variable par rapport à la moyenne de cette variable, la somme des carrés permet de mesurer la variation totale dans une variable.

SPRESS : La capacité prédictive des modèles est évaluée par l'erreur de racine-carrée moyenne.

R^2_{adj} : Estimation de la qualité d'ajustement du modèle à la population.

R^2_{cv} : Est la variation de la statistique du R^2 obtenue en ajoutant ou en enlevant une variable indépendante.

• Y-Randomisation :

Il n'est pas rare d'obtenir des corrélations fortuites (ou « chance corrélation»), c'est-à-dire un modèle affichant de bons résultats statistiques (R^2 , SD) pour l'apprentissage, mais impliquant des descripteurs qui ne sont pas reliés à la propriété/activité modélisée. Ces modèles aléatoires peuvent être détectés par la procédure Y-randomisation [33].

Elle consiste à mélanger aléatoirement les propriétés/activités expérimentales pour la série d'apprentissage en utilisant les mêmes descripteurs, de nouveaux modèles sont obtenus. Ces derniers doivent avoir des performances très faibles.

La validation externe du modèle est indispensable pour vérifier le pouvoir prédictif d'un modèle QSPR/QSAR. D'après certains chercheurs, la validation externe est une norme et une partie obligatoire dans la modélisation QSPR/QSAR [34- 35].

• 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^2 (test) et R^2_{cv} (test). Récemment plusieurs études [36-37] ont montré l'insuffisance des paramètres R^2 (test) et R^2_{cv} (test) pour vérifier la capacité prédictive des modèles QSAR. 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 Tropsha » (Tropshacriteria) [37].

Critères de validation Externe (série de test) :

- $R^2 > 0.7$ (critère 1)
- $R^2_{cv} > 0.6$ (critère 2)
- $R^2 - R^2_0 / R^2 < 0.1$ et $0.85 \leq k \leq 1.15$ (critère 3)
- $R^2 - R'^2_0 / R^2 < 0.1$ et $0.85 \leq k' \leq 1.15$ (critère 4)
- $|R^2 - R^2_0| \leq 0.3$ (critère 5)

Avec

R^2 : coefficient de corrélation pour les molécules de la série de test.

R^2_0 : coefficient de corrélation entre les valeurs prédites et expérimentales pour la série de test.

R^2_0 : coefficient de corrélation entre les valeurs expérimentales et prédites pour la série de test.

K : constante de la droite (à l'origine) de corrélation (valeurs prédites en fonction des valeurs expérimentales)

K' : constante de la droite (à l'origine) de corrélation (valeurs expérimentales en fonction des valeurs prédites).

La figure II.2 montre que R^2 (test) élevée ne signifie pas un bon pouvoir prédictif, il faut que le k soit proche de 1 pour juger que le modèle est prédictif.

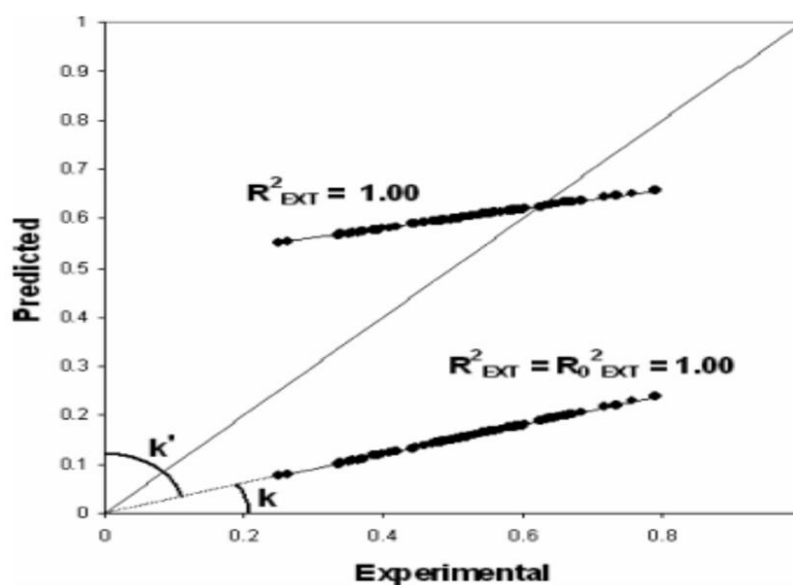


Figure II.2 : importance de la valeur de k dans la validation externe.

• Domaine d'applicabilité :

L'établissement du domaine d'applicabilité pour chaque modèle constitue un des problèmes les plus importants dans l'analyse de QSAR [38]. L'absence de ce domaine admet que chaque modèle de QSAR peut formellement prévoir l'activité d'un produit chimique même si sa structure est complètement différente de ceux inclus dans le sous ensemble (TSET).

Ainsi, l'absence du domaine d'applicabilité comme composant obligatoire de n'importe quel modèle de QSAR mènerait à l'extrapolation injustifiée du modèle dans l'espace de chimie et, en conséquence, avec une probabilité élevée des prévisions imprécises [39].

L'analyse du domaine d'applicabilité est réalisée à l'aide du graphe de Williams. Ce dernier est une représentation graphique des valeurs résiduelles normalisées δ_i , pour chaque composé de l'ensemble (TSET), en fonction de leurs valeurs du levier h_{ii} .

Le levier h_{ii} de l'observation i est lue sur la diagonale principale de la matrice H , dite Ha Matri, et définie de la manière suivante :

$$H = X(X^T X)^{-1} X'$$

$$h_i = x_i^T (X^T X)^{-1} x_i \quad (i=1 \dots n)$$

Où x_i est le rang i du vecteur des descripteurs du composé. X est la matrice du modèle contenant toutes les valeurs liées aux composés de l'ensemble TSET. L'exposant T réfère à la matrice / vecteur transposé.

La valeur critique h^* est définie comme suit :

$$h^* = 3\bar{h} = 3 \frac{\sum_i h_i}{n} = \frac{3p'}{n} \quad (i=1 \dots n)$$

n : nombre des composés dans l'ensemble TSET.

p' : nombre des paramètres ajustés.

Les valeurs h_{ii} peuvent être calculées pour les data TSET du modèle, ainsi que pour des nouvelles structures proposées.

Références bibliographiques :

- [1] A.R. Katritzky, D.C.Fara, R.O.Petrukhin, *Top. Med. Chem.* 2002, 1333-1356.
- [2] A.F.A Cros, *Action de l'alcool amylique sur l'organismell*, thèse de doctorat, faculté de médecine, université Strasbourg, Strasbourg, 1863.
- [3] A.C. Crum-Brown and T.R. Fraser, *On the Connection Between Chemical Constitution and Physiological Action, Part I: On the Physiological Action of the Salts of the Ammonium Bases, Derived from Strychnia, Brucia, Thebia, Codeia, Morphia, Nicotia*", *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, 25, 1868, 151–203.
- [4] M.C. Richet, " *Noté sur le rapport entre la toxicité et les propriétés physiques des corps*", *Comptes rendus des séances de la Société de biologie et de ses filiales*", Paris, 45, 1893, 775–6.
- [5] H. Meyer, *Zur Theorie der Alkoholnarkose. Erste Mittheilung. Welche Eigenschaft der anäs the ticabeding tihrenar kotische Wirkungll*, *Archiv für experimentelle pathologie end Pharmakologie*, 42, 1899, 109–118.
- [6] E. Overton, —*Studien über die Narkos ezugleice in B eitrag zur allgemein en Pharmakologie*, Ed. G. Fischer, Jena, 1901.
- [7] A - R.L. Lipnick, "*Charles Ernest Overton: narcosis studies and a contribution to general pharmacology*", *Trends in Pharmacological Sciences*, 7, 1986, 161–164.
- B - R.L. Lipnick, "Hans Horst Meyer and the lipid theory of narcosis", *Trends in Pharmacological Sciences*, 10(7), 1989, 265–269.
- [8] H. Fühner and E. Neubauer, "*ämolyse durch Sub stanzen homo logen Reihen*", *Archiv fürex perimentelle Pathologie and Pharmakologie*, 56, 1907, 333–345.
- [9] O.R. Hansen, "*Hammett Series with Biological Activity*", *Acta Chemica scan dinavica*, 16, 1962, 1593–1600.
- [10] C. Hansch and T. Fujita, "*p-σ-π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure*", *Journal of the American Chemical Society*, 86(8), 1964, 1616–1626.
- [11] S.M. Free and J.W. Wilson, "*A Mathematical Contribution to Structure-Activity studies*", *Journal of Medicinal Chemistry*, 7(4), 1964, 395–399.
- [12] C. Hansch and E.J. Lien, *Structure-activity relationships in antifungal agents. A survey*, *Journal of Medicinal Chemistry*, 14(8), 1971, 653–670.

- [13] S.Y. Tham and S. Agatonovic-Kustrin, Application of the artificial neural network in quantitative structure-gradient elution retention relationship of phenyl thiocarbamy l amino acids derivatives, *Journal of Pharmaceutical and Biomedical Analysis*, 28(3), 2002, 581-590.
- [15] Goulon-Sigwalt-Abram, 2008.
- [16] J. Wu, Y. Wang, Y. Shen, *Chemomet. Intel. Lab. Syst.* 131, 2014, 22-30.
- [17] A. Goulon-Sigwalt-Abram, *Une nouvelle méthode d'apprentissage de données structurées: applications à l'aide à la découverte de médicaments*, Université Pierre et Marie Curie-Paris VI. 2008.
- [18] R.Todeschini, and V. Consonni, John Wiley & Sons. *Handbook of molecular descriptors*. Vol. 11. 2008.
- [19] A.T.Balaban, *Highly discriminating distance-based topological index*. *Chemical Physics Letters*, 89(5), 1982. 399-404.
- [20] M. Randic, *Characterization of molecular branching*. *Journal of the American Chemical Society*, 97(23), 1975. 6609-6615.
- [21] H. Wiener, *Structural determination of paraffin boiling points*. *Journal of the American Chemical Society*, 69(1), 1947. 17-20.
- [22] H.Van de Waterbeemd, E. Gifford, and ADMET in silico modelling: towards prediction Paradise, *Nat Rev Drug Discov.* 2, 2003, 192-204.
- [23] H. Kasnanen, MJ. Myllymaki, A. Minkkila et al. 3-Heterocycle-phenyl nalkylcarbamates as FAAH inhibitors: design, synthesis and 3D-QSAR studies, *ChemMedChem.* 5, 213-231.
- [24] Abdi, Hervé, and J. Lynne, Williams, *John Wiley and Sons* "Principal Component Analysis." *Inc. WIREsComp Stat* 2, 2010, 433-459.
- [25] Francois Husson, Sebastien Le, and Jérôme Pagès. Chapman; Hall/CRC. Boca Raton, Florida: 2017. <http://factominer.free.fr/bookV2/index.html>.
- [26] Jolliffe, I.T. *Principal Component Analysis*. New York: Springer-Verlag. 2002. <https://goo.gl/SB86SR>.
- [27] Kaiser, F.Henry. "A Note on Guttman's Lower Bound for the Number of common Factors." *British Journal of Statistical Psychology* 14, 1961. 1-2.
- [28] Peres-Neto, R.Pedro, A.Donald, Jackson, and M. Keith. Somers. 2005.
- [29] M. Lejeune, *Statistiques : la théorie et ses applications*, Springer-Verlag, Paris, 2004.
- [30] An Introduction to Partial Least Squares Regression. <http://www.ats.ucla.edu/stat/sas/library/pls.pdf>

- [31] S.Wold, M.Sjostrom, and L.Eriksson Chemometrics and Intelligent laboratory Systems, 58, 2001, 109–130.
- [32] A. Tropsha, P. Gramatica, and V.K. Gombar, *the importance of Being Earnest: Validation is the Absolute Essential for Successful Application and interpretation of QSPR Models, QSAR and Combinatorial Sciences*, 22(1), 2003, 69–77.
- [33] H. Linnan, P.C. Jurs, J. Mol. Graph. Model. 23, 2005, 503.
- [34] A. Tropsha, P. Gramatica, V.K. Gombar, QSAR Comb.Sci. 22, 2003, 69.
- [35] P.P. Roy, S. Paul, I. Mitra, K. Roy, Molecules, 14, 2009, 1660.
- [36] A. Golbraikh, A. Tropsha, J. Mol. Graph. Model. 20, 2002, 269.
- [37] T. M. Martin, P. Harten, D. M. Young, E.N. Muratov, A. Golbraikh, H. Zhu and A.Tropsha, J. Chem. Inf. Model. 52, 2012, 2570.
- [38] E. PAGA; J.C Dearden; P. Gramatica Chemospher 67, 2007, 351-358.
- [39] P. Gramatica; E. Gian i; Paga J. Mol. Graph. Mod. 25,2007, 755-766.

CHAPITRE III
APPLICATION ET RÉSULTAT ET
DISCUSSION

III.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. L'association de 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 paramètres les plus pertinents à utiliser en fonction de l'activité et de la molécule sélectionnées.

Dans ce travail, nous nous intéressons à l'étude des propriétés physico-chimiques de la benzophénone et ses dérivés qui permettent de prévoir les paramètres physico-chimiques qui sont influents sur l'activité biologique, afin de prédire l'activité biologique de nouvelles molécules.

III.2.Méthodologie :

III.2.1. Base de données :

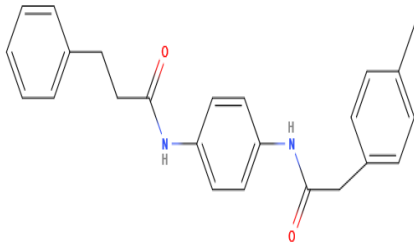
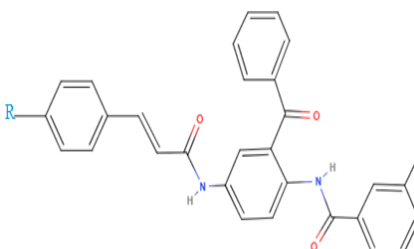
Plusieurs dérivés de la benzophénone présentent des propriétés pharmacologiques importantes. Ils sont utilisés dans le traitement et la prévention d'un grand nombre de maladies.

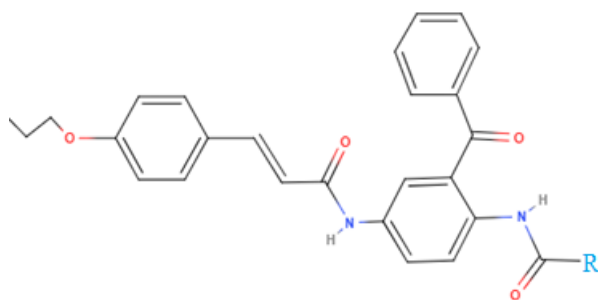
Le choix de la base de données expérimentale initiale est une étape critique pour le développement des modèles QSAR. Elle doit être composée de données expérimentales fiables.

Pour cette étude nous avons utilisé l'activité antipaludique IC_{50} des dérivés de la benzophénone récupérés à partir de la base de données BindingDB [3].

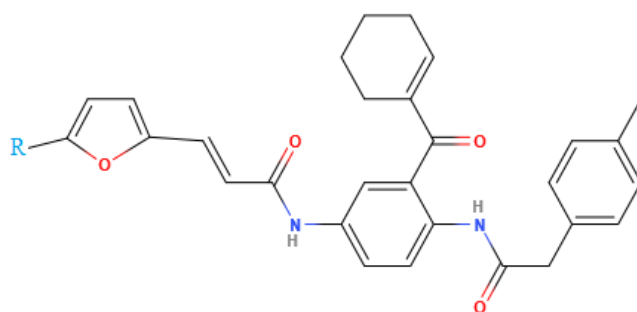
Les structures et l'activité biologique expérimentale des dérivés de la benzophénone sont rapportées dans le tableau III.1.

Tableau III.1: numérotation, structure et activité expiratoire de l'ensemble de données.

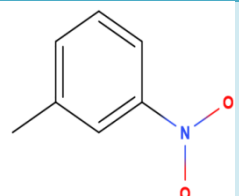
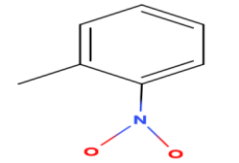
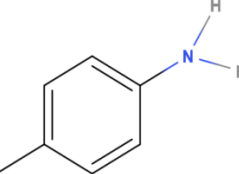
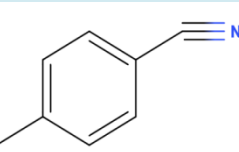
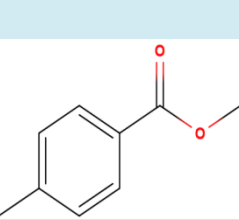
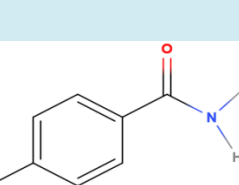
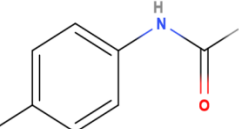
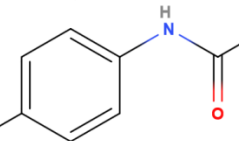
composé	R	PIC50
1		5.57
2		
3	NO ₂	5.19
4	Cl	5.26
5	Br	5.49
6	NH ₂	5.26
7	CH=C (CN) ²	4.37
8	CH ₃	5.85
9	O-CH ₃	5.89
10	CH ₂ -CH ₃	5.92
11	CH (CH ₃) ²	5.92
12	C (CH ₃) ³	5.52
13	O-CH ₂ -CH ₃	6.07
13	O-(CH ₂) ₃ -(CH) ³	5.96



composé	R	PIC50
14		5.60
15		5.89
16		5.62
17		5.55
18		6.92



composé	R	PIC50
19		5.60
20		5.89
21		5.62
22		5.55
23		6.92
24		6.52
25		6.68
26		6.49
27		6.90

28		6.23
29		6.17
30		6.59
31		6.66
32		6.77
33		6.25
34		6.55
35		6.25

Nous avons utilisé le logiciel hyperchem pour représenter les molécules et obtenir les géométries finales. Tous les calculs ont été menés dans le cadre du formalisme RHF (pour restricted Hartree-Fock ou formalisme de Hartree-Fock avec contrainte de spin) sans interaction de configuration. Les structures moléculaires ont été optimisées par la mécanique moléculaire avec un champ de force MM+ à l'aide de l'algorithme Polak- Ribière avec pour critère une racine du carré moyen du gradient égale à $0,01 \text{ kcal/mol}^{-1}$. Puis, à l'aide de la méthode semi-empirique AM1.

III.2.2. Calcul des descripteurs :

La représentation numérique de la structure chimique (descripteurs moléculaires) est une étape importante. Les performances du modèle élaboré et la précision des résultats sont étroitement liées au mode de détermination de ces descripteurs.

Nous avons utilisé PADEL pour trouver des descripteurs, où le programme trouve 1874 descripteurs pour chaque composé exprimant principalement un ensemble de propriétés topologiques, géométriques, et électroniques dans des valeurs numériques de forme 3D hybride, 1D et 2D. Sont des descripteurs 3-dimensionnels, 1-dimensionnels et 2 dimensionnels, respectivement.

Le développement du modèle conçu est passé par les étapes suivantes :

- 1- Supprimer les colonnes sans valeurs spécifiées.
- 2- Supprimer les colonnes avec des valeurs fixes ou semi-constantes.
- 3- Suppression des colonnes ayant un grand coefficient de corrélation entre elles.

Analyse statistique :

Des modèles QSAR ont été générés en utilisant des régressions simples MLR puis MNLR, Sur le programme XLSTAT [4].

XLSTATE :

XLSTAT est la principale solution d'analyse de données et de statistiques pour Microsoft Excel Nous avons choisi ce programme car il possède de nombreuses propriétés:

- ❖ 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.

Les résultats de calculs des paramètres physicochimiques de cette série sont représentés dans le tableau suivant :

Tableau III.2: Descripteurs moléculaires utilisés dans l'étude QSAR.

Descripteurs	Description
ALogP	Coefficient d'eau octanol de Ghose_Crippen
ALogP2	Coefficient d'eau octanol Ghose_Crippen au carré
Bpol	Somme de la valeur absolue de la différence entre les polarisabilités atomiques de tous les atomes liés dans la molécule (y compris les hydrogènes implicites)
TopoPSA	Surface topologique polaire
MAXDN	Variation négative électro topologique maximale (liée à la nucléophilicité de la molécule).
MAXDP	Variation positive électro topologique maximale (liée à l'électrophilie de la molécule).
FMF	Complexité d'une molécule
MW	Masse moléculaire
AMW	Masse moléculaire moyenne (poids moléculaire / nombre total d'atomes)
WPOL	Numéro de polarité de Weiner
WPATH	Numéro de chemin Weiner
XLog P	Coefficient de partage de l'eau d'octanol de Wang
Lipoaffinity Index	Indice d'affinitéLipo
DELS	Variation topologique des électrons moléculaires (mesure du transfert de charge total dans la molécule)
AMR	Réfractivité molaire Ghose_Crippen
Apol	Somme des polarisabilités atomiques (y compris les hydrogènes implicites)

Name	ALogP	ALogp2	AMR	Apol	bpol	Lipoaffin ityIndex	MAXDN	MAXDP	DELS	FMF	MW	AMW	WPATH	WPOL	XLogP	TopoPSA	Pic50	état
1	-0,547	0,299	47,83	80,39	36,71	12,22	2,018	6,357	40,04	0,484	504,2	7,639	5399	56	9,933	84,5	5,89	Train
2	0,077	0,005	54,79	85,78	39,17	13,98	1,988	6,39	37,86	0,45	516,2	7,271	5844	58	12,832	75,27	5,92	Train
3	-0,296	0,088	52,13	83,49	38,9	12,69	2,015	6,378	40,82	0,463	518,2	7,51	5880	57	10,356	84,5	6,07	Train
4	-1,162	1,351	58,57	89,67	43,27	13,88	2,011	6,415	41,85	0,426	546,2	7,283	6960	59	11,283	84,5	5,96	Train
5	-1,296	1,68	63,28	90,48	45,19	13,37	2,028	6,542	46,59	0,407	562,2	7,397	7039	65	9,792	93,7	5,89	Train
6	-0,602	0,363	61,46	89,67	43,27	13,88	2,015	6,516	42,87	0,413	546,2	7,283	6600	64	11,192	84,5	5,62	Train
7	-0,355	0,126	64,56	88,96	41,28	13,29	2,008	6,525	43,26	0,43	610,1	8,474	6600	64	10,459	84,5	5,55	Train
8	-0,759	0,576	39,69	77,83	34,8	11,84	1,901	6,329	36,83	0,5	476,2	7,44	4550	52	11,177	75,2	5,57	Train
9	-0,146	0,021	62,39	89,34	43,6	15,1	5,878	6,498	74,07	0,413	600,2	8,002	7915	69	11,561	84,5	6,92	Train
10	-0,007	5,775	47,64	88,77	38,9	14,01	2,057	6,44	42,24	0,513	554,2	7,697	7278	62	12,081	88,41	6,38	Train
11	-0,007	5,775	47,64	97,14	41,09	16,13	2,061	6,518	44,1	0,525	604,2	7,746	9358	72	8,002	7915	6,7	Train
12	0,634	0,402	53,14	91,86	41,09	14,59	2,056	6,466	43,03	0,493	568,2	7,576	7705	65	12,404	88,41	6,92	Train
13	1,548	2,398	66,77	101,14	47,65	16,65	2,057	6,521	45,17	0,44	610,2	7,265	9331	71	14,087	88,41	6,52	Train
14	0,191	0,036	59,25	95,76	45,19	14,54	2,072	6,495	47,06	0,468	598,2	7,572	8829	68	11,427	97,63	6,52	Train
15	0,079	0,006	54,68	92,55	43,11	14,14	2,433	6,477	54,53	0,486	602,2	7,923	8829	68	11,274	97,63	6,68	Train
16	0,472	0,222	48,51	88,66	39,01	14,28	2,116	6,443	53,34	0,513	572,2	7,947	7705	65	11,407	88,41	6,49	Train
17	0,881	0,776	56,24	91,15	39,1	14,03	2,044	6,471	43,6	0,513	632,1	8,779	7705	65	12,093	88,41	6,9	Train
18	-0,024	5,904	64,15	93,47	43,44	11,66	2,402	6,477	60,91	0,474	603,2	7,733	8749	69	10,877	127,63	6,23	Train
19	-0,024	5,904	64,15	93,47	43,44	11,8	2,402	6,474	60,07	0,474	603,2	7,733	8851	69	10,455	127,63	6,17	Train
20	-1,434	2,057	64,48	93,2	43,71	12,89	2,017	6,471	47,54	0,474	573,2	7,349	7705	65	9,96	110,51	6,59	Train
21	-0,649	0,422	66,46	93,63	42,84	12,86	2,042	6,481	50,08	0,48	583,2	7,574	8210	67	10,549	108,28	6,66	Train
22	0,003	1,296	59,63	95,22	43,96	13,45	2,148	6,499	55,89	0,474	612,2	7,849	9269	71	11,428	114,7	6,77	Train
23	-0,642	0,412	55,08	93,1	40,29	12,83	2,236	6,48	55,15	0,486	597,2	7,858	8717	69	10,379	131,5	6,25	Train

CHAPITRE III: Application et Résultat et Discussion

24	0,351	0,123	46,25	76,43	29,08	11,72	1,977	6,35	37,12	0,542	528,1	8,95	4956	54	9,191	75,27	5,26	Train
25	0,393	0,154	49,12	78,88	32,81	12,22	1,986	6,349	37,019	0,516	552,1	8,904	4956	54	10,81	75,27	5,49	Train
26	-1,245	1,551	44,74	78,26	33,05	11,44	2,001	6,328	39,923	0,5	489,2	7,643	4956	54	9,347	101,28	5,26	Train
27	-0,103	0,01	66,86	88,83	36,12	12,39	2,159	6,448	48,946	0,45	564,2	7,946	7932	63	11,188	122,85	4,37	Train
28	0,146	0,021	46,02	79,59	34,8	12,7	1,99	6,342	36,684	0,492	488,2	7,51	4956	54	11,755	75,27	5,85	Test
29	-0,188	0,035	50,87	82,68	36,99	13,29	1,987	6,367	37,22	0,47	502,2	7,385	5399	56	12,324	75,27	5,92	Test
30	0,0774	0,005	54,79	85,78	39,17	13,98	1,988	6,39	37,86	0,45	516,2	7,271	5844	58	12,832	75,27	5,52	Test
31	-1,244	1,549	55,96	86,58	41,09	13,31	2,015	6,462	42,3	0,43	532,2	7,392	6210	61	10,869	84,5	5,6	Test
32	0,164	0,027	44,41	78,53	32,78	10,62	2,496	6,342	51,61	0,5	519,1	8,112	5844	58	10,264	118,4	5,19	Test
33	-0,059	0,003	54,96	92,66	43	14,07	2,074	6,478	46,27	0,486	584,2	7,687	8245	67	11,004	97,63	6,89	Test
34	-0,711	0,505	56,43	93,1	41,61	13,23	2,11	6,484	50,88	0,486	597,2	7,858	8759	68	10,669	117,5	6,55	Test
35	-0,632	0,4	60,51	96,19	43,8	13,75	2,109	6,502	52,68	0,468	611,2	7,737	9310	69	10,905	117,5	6,25	Test

Tableau III.3: Valeurs pIC50 expérimentales avec les descripteurs sélectionnés.

III.3. Résultats et discussions :

III.3.1. Analyse en composantes principales (ACP) :

L'analyse en composantes principales est utilisée pour extraire et de visualiser les informations importantes contenues dans une table de données multi-variées. L'ACP synthétise cette information en seulement quelques nouvelles variables appelées **composantes principales**. Ces nouvelles variables correspondent à une combinaison linéaire des variables originels. Le nombre de composantes principales est inférieur ou égal au nombre de variables d'origine. L'ACP réduit les dimensions de données multi-variées à deux ou trois composantes principales.

Les 19 descripteurs (variables) codant les 35 molécules ont été soumis à l'analyse des composants principaux et 20 composants ont été obtenus (Figure III.1). Les trois premiers axes, F1, F2 et F3, ont contribué respectivement à 51,92%, 12,42% et 11,58% à la variance totale, et l'information totale a été estimée à 75,92%. Les coefficients de corrélation de Pearson sont résumés dans le tableau III.4, la matrice fournit des informations sur les corrélations positives ou négatives entre les variables; sur la Figure III.2 ces descripteurs sont représentés dans un cercle de corrélation.

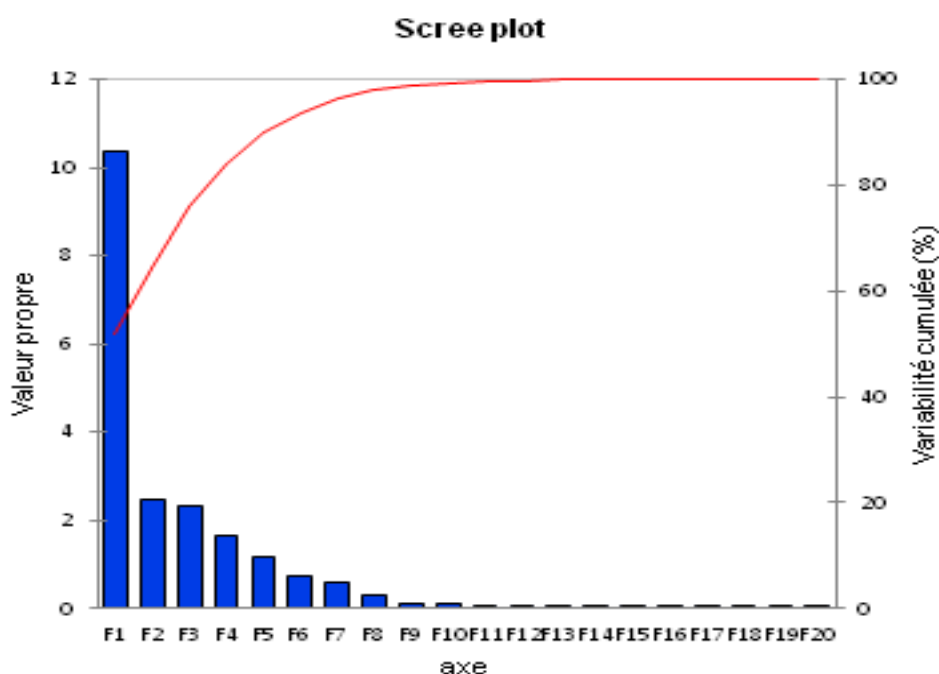


Figure 1 : Les principales composantes et leurs variances.

CHAPITRE III: Application et Résultat et Discussion

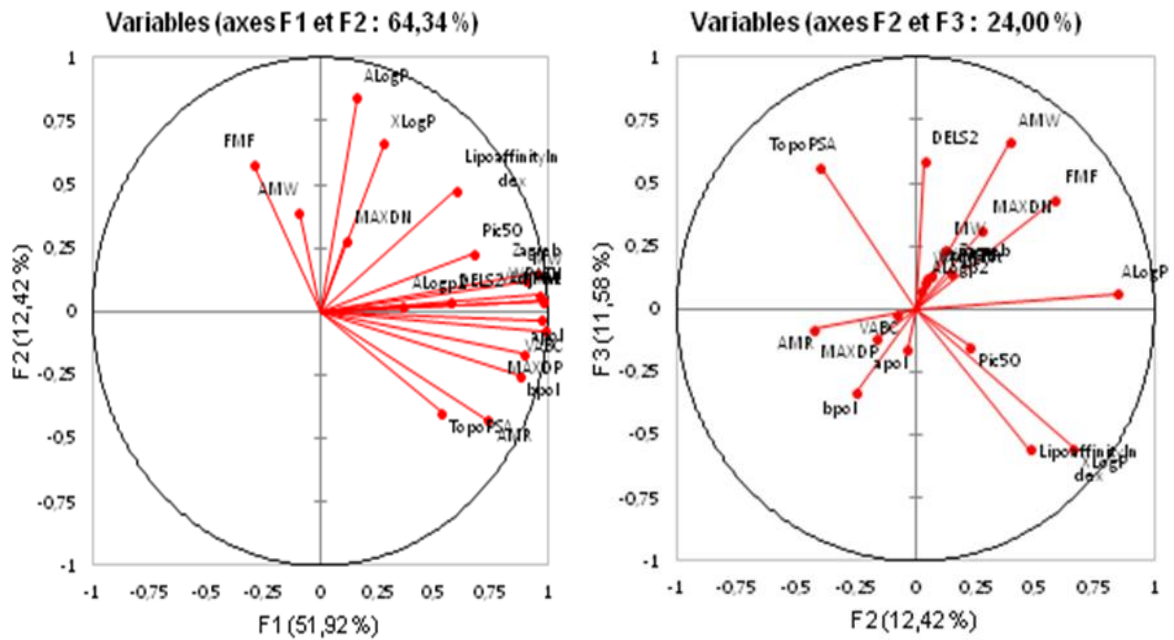


Figure III.1 : cercle de corrélation.

Le résultat le plus important obtenu par l'ACP est la matrice de corrélation (tableau III.4) qui nous permet de déterminer la corrélation entre les 19 descripteurs et par la suite d'éliminer ceux qui sont corrélés. On trouve que :

- VABC, apol : sont corrélés on élimine VABC et on laisse apol.
- VABC, VAdjMat, WPATH, WPOL : sont corrélés, On élimine VABC et on laisse les autres descripteurs.
- VAdjMat, WPATH : sont corrélés On élimine VAdjMat et on laisse WPATH.
- WPATH, WPOL, Zagreb : sont corrélés On élimine Zagreb et on laisse WPOL, WPATH.
- WPOL, Zagreb : sont corrélés On laisse WPOL et on élimine Zagreb.
- Zagreb, VAdjMat : sont corrélés On laisse Zagreb et on élimine VAdjMat

Les descripteurs qui restent pour la suite de cette étude sont : ALog p, Alogp2, AMR, apol, bpol, Lipo affinity Index, MAXDP, WPOL, WPATH, DELS2, FMF, MAXDN, MW, AMW, XLog p, Topo PSA, Pic50.

Variables	ALopP	ALop2	AMR	Apol	bpol	LipoaffinityIn dex	MAXDP	DELS2	FMF	MAXDN	VABC	VAdjMat	MW	AMW	WPATH	WPOL	XLogP	Zagreb	TopoPSA	Pic50
ALopP	1																			
ALop2	0.002	1																		
AMR	-0.09	0.124	1																	
Apol	0.12	0.367	0.724	1																
Bpol	-0.091	0.279	0.771	0.894	1															
LipoaffinityInd	0.387	0.083	0.28	0.628	0.586	1														
MAXDP	0	0.252	0.785	0.882	0.836	0.552	1													
DELS2	0.133	0.182	0.38	0.38	0.36	0.105	0.38	1												
FMF	0.372	0.135	-0.685	-0.288	-0.587	-0.219	-0.46	-0.12	1											
MAXDN	0.177	-0.06	0.017	-0.082	0.013	0.193	-0.042	0.723	-0.128	1										
VABC	0.122	0.328	0.792	0.979	0.875	0.558	0.894	0.52	-0.335	0.033	1									
VAdjMat	0.187	0.359	0.647	0.944	0.794	0.536	0.828	0.617	-0.156	0.116	0.97	1								
MW	0.279	0.266	0.679	0.845	0.683	0.479	0.847	0.574	-0.136	0.121	0.89	0.895	1							
AMW	0.393	-0.158	-0.113	-0.219	-0.421	-0.188	-0.041	0.228	0.411	0.214	-0.124	-0.049	0.315	1						
WPATH	0.2	0.382	0.607	0.937	0.768	0.519	0.785	0.595	-0.099	0.086	0.957	0.994	0.874	-0.063	1					
WPOL	0.182	0.363	0.661	0.933	0.804	0.55	0.875	0.634	-0.203	0.152	0.959	0.982	0.905	-0.026	0.967	1				
XLogP	0.581	0.193	-0.028	0.335	0.232	0.743	0.132	-0.097	0.034	0.101	0.246	0.246	0.147	-0.243	0.263	0.243	1			
Zagreb	0.26	0.407	0.551	0.924	0.737	0.542	0.769	0.597	-0.016	0.107	0.932	0.985	0.876	-0.029	0.989	0.968	0.297	1		
TopoPSA	-0.172	0.319	0.423	0.491	0.334	-0.275	0.368	0.537	0.014	-0.065	0.565	0.62	0.482	-0.029	0.641	0.579	-0.311	0.607	1	
Pic50	0.18	0.185	0.302	0.671	0.641	0.56	0.583	0.275	0.005	0.014	0.615	0.645	0.621	-0.076	0.625	0.644	0.253	0.672	0.127	1

Tableau III.4: La matrice de corrélation (Pearson (n)) entre les différents descripteurs obtenus.

III.3.2. Répartition des données et développement de modèles :

Dans la présente étude, des modèles QSAR ont été construits selon les principes de l'OCDE pour des modèles QSAR acceptables. Cette approche a conduit à la génération de modèles QSAR possédant d'excellentes performances statistiques. Par conséquent, l'ensemble des données a été divisé au hasard en ensembles de d'apprentissage et de test par un bon nombre de modèles MLR avec des performances statistiques, Parmi les produits chimiques de l'ensemble de données, 80% ont été sélectionnés pour l'ensemble d'apprentissage et les 20% restants ont été considérés comme l'ensemble de test. Les modèles qui ne satisfont pas aux principes de l'OCDE [5] ont été sommairement exclus.

13 modèles de régression linéaire multiple (MLR) basés sur la même taille d'ensembles d'apprentissage avec les coefficients de détermination (R^2) Ils sont définis dans le tableau ci-dessous :

Tableau III.5 : Modèles sélectionnés et paramètres statistiques des corrélations entre les propriétés moléculaires.

	Equation	Valeur De Test	R^2	S	VIF	R^2 test
1	$\text{Pic50} = 4,758 + 3,880 \times 10^4 \times \text{WPATH} - 1,416 \times 10^{-02} \times \text{TopoPSA}$	WPATH=6.627 TopoPSA=-2.641	0,686	0.099	WPATH=1.786 TopoPSA=1.786	0,358
2	$\text{Pic50} = -5,792 + 0,154 \times \text{bpol} + 12,135 \times \text{FMF}$	Bpol=6.525 FMF=3.889	0,640	0.175	Bpol=1.710 FMF=1.710	0,645
3	$\text{Pic50} = -4,831 + 0,139 \times \text{bpol} + 11,230 \times \text{FMF}$	Bpol=7,030 FMF=3.704	0,679	0.133	Bpol=1.817 FMF=1.817	0,530
4	$\text{Pic50} = 7,271 + 4,242 \times 10^4 \times \text{WPATH} - 0,1985 \times \text{XLogP} - 2,043 \times 10^{-02} \times \text{TopoPSA}$	WPATH=5.932 XLogP=-2.068 TopoPSA=2.862	0,649	0.119	WPATH=2.470 XLogP=1.891 TopoPSA=3.249	0,434
5	$\text{Pic50} = -9,569 + 0,219 \times \text{bpol} + 17,547 \times \text{FMF} - 2,048 \times 10^0 \times \text{WPATH}$	Bpol=7.049 FMF=6.589 WPATH=-2.908	0,797	0.082	Bpol=19.808 FMF=10.899 WPATH=7.363	0,408

CHAPITRE III: Application et Résultat et Discussion

6	$\text{Pic50} = -4,724 - 7,857 \times 10^2 \times \text{ALogp2} + 0,152 \times \text{bpol} + 10,154 \times \text{FMF}$	$\text{ALogp2} = -1.746$ $\text{Bpol} = 6.375$ $\text{FMF} = 3.614$	0,669	0.117	$\text{ALogp2} = 1.121$ $\text{Bpol} = 1.833$ $\text{FMF} = 1.973$	0,664
7	$\text{Pic50} = -4,184 - 7,062 \times 10^{02} \times \text{ALogp2} + 0,102 \times \text{bpol} + 8,60 \times \text{FMF} + 3,957 \times 10^{-03} \times \text{MW}$	$\text{ALogp2} = -1.185$ $\text{Bpol} = 2,815$ $\text{FMF} = 2.926$ $\text{MW} = 1,662$	0,707	0.103	$\text{ALogp2} = 1.252$ $\text{Bpol} = 4.558$ $\text{FMF} = 2.787$ $\text{MW} = 2.220$	0,578
8	$\text{Pic50} = -5,687 - 4,721 \times 10^{-02} \times \text{ALogp2} + 0,190 \times \text{bpol} + 11,834 \times \text{FMF} - 1,431 \times 10^{-02} \times \text{TopoPSA}$	$\text{ALogp2} = -0.890$ $\text{Bpol} = 8101$ $\text{FMF} = 4.341$ $\text{TopoPSA} = -2.615$	0,755	0.135	$\text{ALogp2} = 1.434$ $\text{Bpol} = 4.911$ $\text{FMF} = 3.130$ $\text{TopoPSA} = 2.811$	0,508
9	$\text{Pic50} = -13,834 + 0,240 \times \text{bpol} + 17,109 \times \text{FMF} + 0,010 \times \text{MW} - 5,136 \times 10^{04} \times \text{WPATH}$	$\text{Bpol} = 7.615$ $\text{FMF} = 6.474$ $\text{MW} = 2.969$ $\text{WPATH} = -4.276$	0,823	0.081	$\text{Bpol} = 19.720$ $\text{FMF} = 7.395$ $\text{MW} = 17.467$ $\text{WPATH} = 33.328$	0,395
10	$\text{Pic50} = -6,023 + 0,192 \times \text{bpol} \times 0,103 \times \text{MAXDN} + 15,931 \times \text{FMF} - 0,140 \times \text{XLogP} \times 1,427 \times 10^{-02} \times \text{TopoPSA}$	$\text{Bpol} = 8.096$ $\text{MAXDN} = -1.183$ $\text{FMF} = 6.104$ $\text{XLogP} = -1.559$ $\text{TopoPSA} = -3.110$	0,798	0.092	$\text{Bpol} = 5.172$ $\text{MAXDN} = 1.251$ $\text{FMF} = 5.251$ $\text{XLogP} = 1.610$ $\text{TopoPSA} = 3.437$	0,477
11	$\text{Pic50} = 12,659 - 0,146 \times \text{apol} \times 1,467 \times \text{MAXDN} + 9,415 \times \text{FMF} + 3,410 \times 10^{-02} \times \text{MW} - 1,817 \times \text{AMW}$	$\text{Apol} = -1,247$ $\text{MAXDN} = -1.602$ $\text{FMF} = 3.418$ $\text{MW} = 1.911$ $\text{AMW} = -1.905$	0,708	0,082	$\text{Apol} = 176.324$ $\text{MAXDN} = 5.195$ $\text{FMF} = 4.602$ $\text{MW} = 205.594$ $\text{AMW} = 52.221$	0,026
12	$\text{Pic50} = -5,301 - 6,164 \times 10^{02} \times \text{ALogp2} + 0,150 \times \text{bpol} + 0,214 \times \text{MAXDN} + 11,871 \times \text{FMF} - 6,286 \times 10^{-03} \times \text{TopoPSA}$	$\text{ALogp2} = -1.564$ $\text{Bpol} = 7.177$ $\text{MAXDN} = 2.692$ $\text{FMF} = 4.847$ $\text{TopoPSA} = -1.449$	0,758	0.082	$\text{ALogp2} = 3.332$ $\text{Bpol} = 11.505$ $\text{MAXDN} = 1.389$ $\text{FMF} = 7.717$ $\text{TopoPSA} = 6.282$	0,555
13	$\text{Pic50} = -13,086 - 6,697 \times 10^{02} \times \text{ALogp2} + 0,246 \times \text{bpol} + 0,358 \times \text{MAXDN} + 21,398 \times \text{FMF} - 2,174 \times 10^{-04} \times \text{WPATH}$	$\text{ALogp2} = -2,973$ $\text{Bpol} = 10.683$ $\text{MAXDN} = 6,132$ $\text{FMF} = 10.874$ $\text{WPATH} = -4,073$	0,916	0.040	$\text{ALogp2} = 1,321$ $\text{Bpol} = 4.590$ $\text{MAXDN} = 1,224$ $\text{FMF} = 3.321$ $\text{WPATH} = 4.175$	0.896

Dans les 13 modèles du tableau III.5, l'activité des agents antipaludiques est fonction de 16 descripteurs (AlogP, AlogP2, AMR, apol, bpol, LipoaffinityIndex, MAXDN, MAXDP,

DELS, FMF, MV, AMW, WPATH, WPOL, XLogP, TopoPSA). Les descripteurs qui se répètent plusieurs fois sont FMF, bpol, WPATH, ALogp2.

Les modèles 1, 3, 4, 5, 7, 8, 9,10, 11 et 12 ont un coefficient de détermination de la validation externe R^2_{test} inférieur à 0,6, et le vif de modèle 5 de descripteurs (Bpol, FMF, WPATH), modèle 9 de descripteurs (Bpol, FMF, WM, WPATH), modèle 10 de descripteurs (Bpol, FMF) et modèle 11 de descripteurs (Apol, AMW, MW, MAXDN) aussi le modèle 12 de descripteur (Bpol, FMF, TopoPSA) sont supérieur à 5 ; Alors ces modèles ne sont pas valide.

Les modèles 2, 6 et 13 ont un R^2 supérieur à 0,5, R^2_{test} supérieur à 0,6 et les valeurs de VIF pour tous descripteurs inférieurs à 5 ; Alors ces modèles sont tous valides.

Parmi les modèles qui sont valides on trouve que le modèle 13 qui présente les valeurs de R^2_{test} ($R^2_{\text{test}} = 0,896$) et R^2 ($R^2 = 0,916$) les plus élevées par rapport aux autres modèles. Donc c'est le meilleur modèle qu'on peut l'utiliser pour la prédiction des activités pour de nouvelles molécules qui peuvent devenir des médicaments antipaludiques.

III.3.3. Analyse statistiques de meilleur Modèle MLR :

Une analyse de régression linéaire multiple des descripteurs moléculaires a été réalisée à l'aide de la stratégie par étapes de XISTATE pour Windows.

La meilleure corrélation est exprimée par la relation mathématique suivante:

$$\text{Pic50} = -13,086-6,697*10^{-2}*A\text{Logp}2+0,246*b\text{pol}+0,358*MAXDN+21,398*FMF-2,174*10^{-4}*WPATH$$

Equation 01

$$N= 27 \quad S= 0.040 \quad R=0.957 \quad R^2 =0.916 \quad R^2_{\text{adj}} =0.896 \quad F= 45.847; \text{ p-value} < 0.0001$$

Le modèle exprimé par l'équation (Equ.01) reliant la variable dépendante pIC50 avec les cinq descripteurs est accompagné par les mesures statistiques citées ci-dessus : coefficient de corrélation multiple R, coefficient de détermination R^2 , coefficient de détermination ajusté R^2_{adj} et l'écart type S.

La valeur F est jugée la signification statistique au niveau de 95%, pour la valeur de F calculées est supérieure par rapport aux valeurs lu dans le tableau du Fischer, nous prenons moins de 0,01% de risque en supposant que l'hypothèse nulle (H_0) est fausse.

CHAPITRE III: Application et Résultat et Discussion

Un bon modèle fiable et robuste nécessite un coefficient de détermination plus élevé et une erreur quadratique moyenne plus faible.

Les valeurs de la fraction de la variance peuvent varier entre 0 et 1. Le modèle QSAR doit considérer un $R^2 > 0.5$ pour qu'il soit valide [6]. La valeur $R = 0,957$ ($R^2 = 0,916$) (modèle 13).

D'après les résultats présents dans le Tableau III.5 il s'avère que le meilleur modèle obtenu avec cinq descripteurs est celui avec LogP2, MAXDN, FMF, Bpol et WPATH.

Le MAXDn démontre la différence d'état intrinsèque négative maximale et peut être associé à la nucléophilicité des molécules.

MAXDn a été proposé sur la base des états intrinsèques de Kier-Hall des atomes [7]. L'état intrinsèque d'un atome est calculé comme le rapport entre l'électronégativité atomique de Kier-Hall et le degré du sommet, de sorte que le nombre de liaisons de l'atome, codant les informations liées aux charges partielles des atomes et à leur topologie négative par rapport à la molécule entière.

Puisque MAXDn a un signe positif dans l'équation linéaire 13, l'augmentation de la valeur des descripteurs via le comportement électrophile des composés entraînera une augmentation des valeurs de pIC50.

FMF est un descripteur lié au concept de topologie moléculaire ; c'est la fraction du cadre moléculaire, qui se compose cycle terminaux et de pont moléculaire. Il a été démontré que la FMF est corrélée aux propriétés ADMET, telles que la solubilité, la perméabilité et l'inhibition de l'isoforme 3A4 du cytochrome P450.

Le troisième descripteur est le descripteur électrostatique bpol Somme de la valeur absolue de la différence entre les polarisabilités atomiques de tous les atomes liés dans la molécule Alors que celle basée sur l'atome.

Ces descripteurs bpol et FMF sont positivement corrélé avec les activités antipaludiques des dérivés de benzophénone. Cela implique qu'une augmentation globale de la polarisabilité et FMF du composé améliore l'activité antipaludique de la benzophénone.

La lipophilicité est un paramètre physico-chimique principal influençant la biodisponibilité, la perméabilité et fréquemment la toxicité des médicaments. Le descripteur logP (o / w) fait référence au log du coefficient de partage octanol / eau de la molécule et est considéré comme une mesure de la lipophilicité d'une molécule.

Le coefficient négatif du descripteur (logp2) dans le modèle MLR suggère que l'augmentation de la lipophilicité globale de la molécule diminuera à son tour l'activité inhibitrice de la PFT des dérivés de la benzophénone.

Tableau III.6: Activité observée vs calculée des dérivés de benzophénone.

	MLR			MNLR	
	Pic50 _{exp}	Pic50 _{préd}	Résid	Pic50 _{préd}	résid
1	5,890	5,853	0,037	5.845	0.045
2	5,920	5,644	0,276	5.639	0.281
3	6,070	5,852	0,218	5.849	0.221
4	5,960	5,817	0,143	5.911	0.049
5	5,890	5,851	0,039	5.908	-0.018
6	5,620	5,685	-0,065	5.707	-0.087
7	5,550	5,571	-0,021	5.607	-0.057
8	5,570	5,848	-0,278	5.801	-0.231
9	6,920	6,889	0,031	6.919	0.001
10	6,380	6,252	0,128	6.395	-0.015
11	6,700	6,598	0,102	6.715	-0.015
12	6,920	6,631	0,289	6.714	0.206
13	6,520	6,628	-0,108	6.450	0.070
14	6,520	6,893	-0,373	6.719	-0.199
15	6,680	6,897	-0,217	6.757	-0.077
16	6,490	6,579	-0,089	6.711	-0.221
17	6,900	6,539	0,361	6.743	0.157
18	6,230	6,332	-0,102	6.194	0.036
19	6,170	6,310	-0,140	6.169	0.001
20	6,590	6,746	-0,156	6.834	-0.244
21	6,660	6,668	-0,008	6.682	-0.022
22	6,770	6,565	0,205	6.548	0.222
23	6,250	6,128	0,122	6.178	0.072
24	5,260	5,306	-0,046	5.114	0.146
25	5,490	5,670	-0,180	5.584	-0.094
26	5,260	5,299	-0,039	5.302	-0.042
27	4,370	4,499	-0,129	4.554	-0.184
28	5,850*	5,658*	0,192*	5.564*	0.286*
29	5,920*	5,629*	0,291*	5.593*	0.327*
30	5,520*	5,644*	-0,124*	5.639*	-0.119*
31	5,600*	5,516*	0,084*	5.655*	-0.055*
32	5,190*	5,319*	-0,129*	5.151*	0.039*

CHAPITRE III: Application et Résultat et Discussion

33	6,890*	6,868*	0,022*	6.825*	0.065*
34	6,550*	6,393*	0,157*	6.444*	0.106*
35	6,250*	6,434*	-0,184*	6.355*	-0.105*

La matrice de corrélation pour pIC50 et les descripteurs sélectionnés pour construire le modèle 2D-QSAR est présentée dans le tableau III.7. Les paramètres utilisés dans le modèle 13 sont presque indépendants, ce qui peut être vu à partir de la matrice de corrélation.

Tableau III.7: Matrice de corrélation.

	ALogp2	Bpol	MAXDN	FMF	WPATH	Pic50
ALogp2	1	0,246	-0,064	0,182	0,390	0,206
Bpol	0,246	1	0,211	-0,613	0,755	0,606
MAXDN	-0,064	0,211	1	-0,331	0,193	0,292
FMF	0,182	-0,613	-0,331	1	-0,140	0,083
WPATH	0,390	0,755	0,193	-0,140	1	0,607
Pic50	0,206	0,606	0,292	0,083	0,607	1

Analyse des valeurs résiduelles normalisées :

Cette étape consiste à examiner les valeurs résiduelles issues du modèle de QSAR obtenu.

Une valeur résiduelle est la différence entre les valeurs observées et prédites de l'activité inhibitrice, présentée par les dérivés des benzophénones.

La justesse de notre modèle est vérifiée par l'analyse de la distribution normale et la linéarité des valeurs résiduelles. Parmi les différentes méthodes utilisées pour l'analyse des valeurs résiduelles nous avons choisi, dans ce travail, d'examiner la linéarité par la représentation graphique des valeurs résiduelles normalisées en fonction des valeurs observées et prédites du pIC50.

L'évaluation de la distribution normale des valeurs résiduelles normalisées est effectuée par la représentation graphique de chaque valeur résiduelle normalisée en fonction de son score. Les résultats de l'analyse obtenus sont illustrés dans les trois figures suivantes (Figures III.3, 4 et 5).

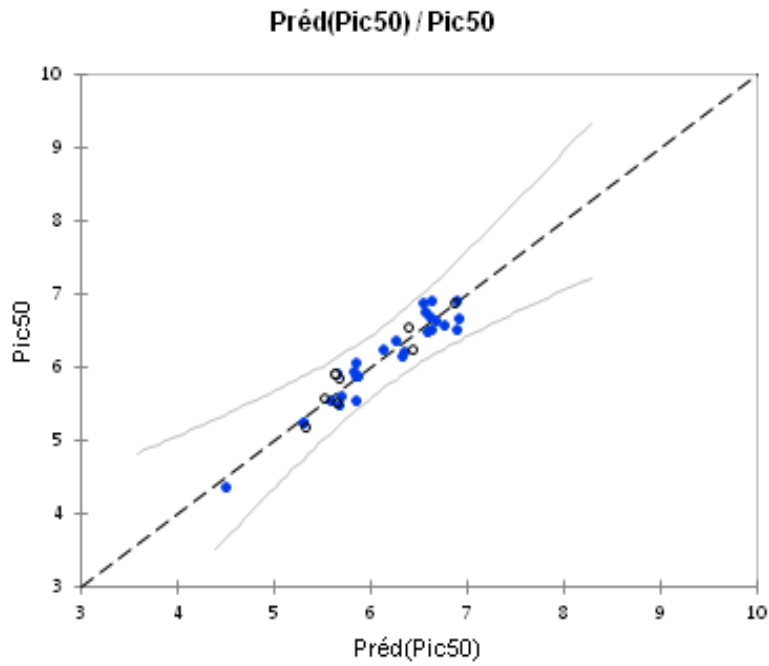
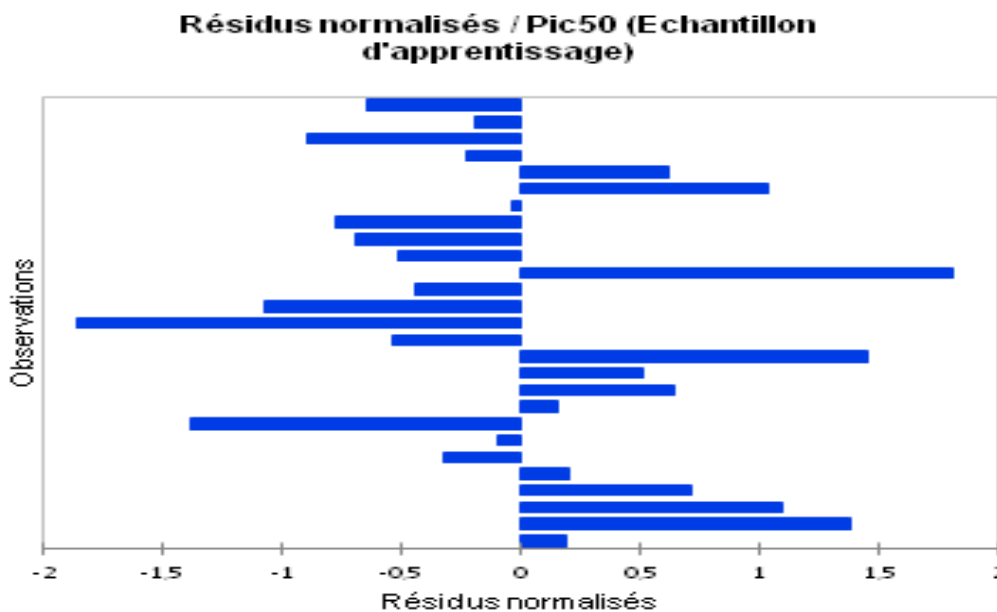


Figure III.3 : Corrélation entre l'activité biologique expérimentale et prédite.

La figure III.3 présente la corrélation entre les valeurs prédites et expérimentales pour l'activité biologique étudiée des dérivés de benzophénone. Du point de vue des performances statistiques, ce modèle présente de coefficient de corrélation R^2 significatif de valeur.

La valeur de R^2 est égale à 0.916, pour le modèle 13. Il indique que ce modèle peut être appliqué avec succès pour prédire l'activité étudiée.



CHAPITRE III: Application et Résultat et Discussion

Figure III.4 : Corrélation entre l'activité biologique expérimentale et résiduelle.

Pour rechercher la présence d'une erreur systématique dans l'élaboration des modèles QSAR, les valeurs résiduelles de l'activité biologique ont été tracées en fonction des valeurs expérimentales, comme le montre la figure III.4.

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

Il indique que ce modèle peut être appliqué avec succès pour prédire l'activité biologique pour l'ensemble des benzophénones utilisées dans le développement de modèle QSAR.

La figure III.5 présente les deux graphes qui correspondent à la représentation graphique des valeurs résiduelles normalisées respectivement en fonction des valeurs prédites et observées du pIC50.

La bonne dispersion des points des deux côtés du zéro, sans tendance d'accroissement ou de diminution, indique que la variance est une constante, alors qu'il n'y a aucun problème de non linéarité de notre modèle.

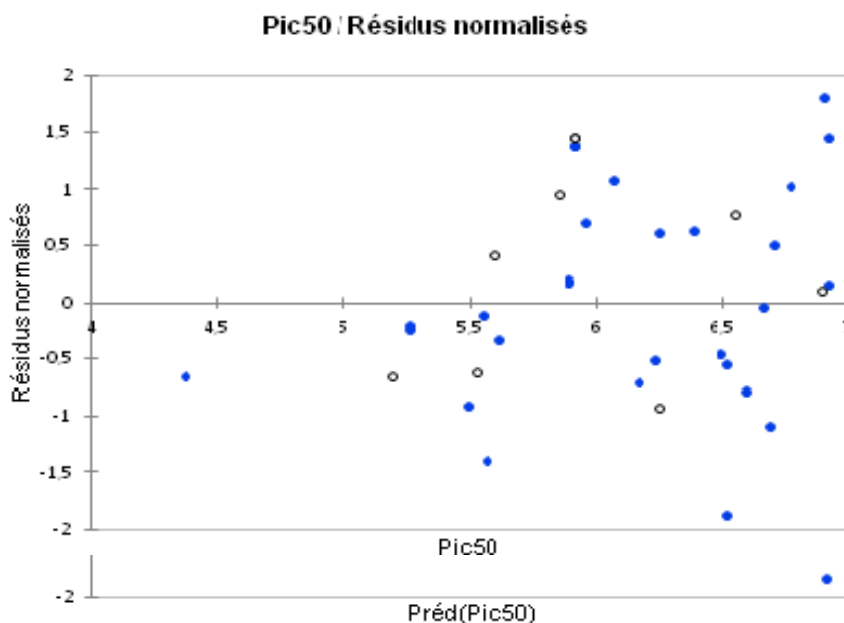


Figure III.5 : Valeurs résiduelles normalisées en fonction des activités Prédites et observées.

III.3.4. Régression non linéaire multiple (MNL) :

Un modèle de régression non linéaire est créé pour améliorer l'activité de la structure de manière quantitative considérant plusieurs paramètres. Ce modèle est l'outil le plus courant pour étudier les multidimensionnelles les données. Nous avons utilisé la matrice de données des descripteurs proposés par le MLR, qui correspond à 35 composés (ensemble d'entraînement).

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

Nous avons utilisé la fonction préprogrammée de XLSTAT suivante :

Où à, b, c, d ... représentent les paramètres et $X_1, X_2, X_3, X_4 \dots$: représentent les variables L'équation résultante est :

$$\begin{aligned} \text{Pic50} = & -13.211 + 0.052 * \text{ALogp2} + 0.543 * \text{bpol} - 3.0641 * \text{FMF} + 0.00008 * \text{WPATH} \\ & 0.191 * \text{MAXDN} - 0.019 * \text{ALogp2}^2 - 0.003 * \text{bpol}^2 + 26.209 * \text{FMF}^2 \\ & 1.805 * 10^2 * \text{WPATH}^2 + 0.070 * \text{MAXDN}^2 \end{aligned}$$

Equation 02

L'équation de modèle a été évaluée par le coefficient de corrélation ajusté. (R^2_{adj}), l'erreur quadratique moyenne (S), le test de Fischer (F) et le coefficient de corrélation croisée (Q^2_{cv}).

$$N = 27 \quad R^2_{\text{adj}} = 0.757 \quad R^2_{\text{cv}} = 0.946 \quad S = 0,034 \quad R^2_{\text{test}} = 0,914 \quad \text{RMSE} = 0,185$$

Les valeurs élevées du R^2_{cv} et R^2_{adj} sont des critères essentiels pour une meilleure qualification des modèles QSAR. Les résultats obtenus de R^2_{cv} et R^2_{adj} pour le modèle MNL QSAR [9] sont 0.946 et 0.757 respectivement.

On peut également utiliser le paramètre ce qui rend compte des variations d'erreur des prédictions.

La robustesse et la puissance prédictive du modèle ont été confirmées par la valeur de R test significative (0,956) des données de l'ensemble de test.

Avec le modèle MNL optimal, les valeurs des activités prédites pIC50 MNL calculées à partir de l'équation 2 et les valeurs observées sont données dans le tableau III.6.

CHAPITRE III: Application et Résultat et Discussion

Les valeurs d'activité pIC50 prédites par ce modèle sont presque similaires à celles observées comme l'illustre la figure III.7.

La relation entre l'estimation, l'observation activités et le graphique résiduel des nombres typiques est montré dans la Figure III.6 et 7. La répartition des résidus dans les deux côtés de l'axe des abscisses indique l'absence de l'erreur systématique.

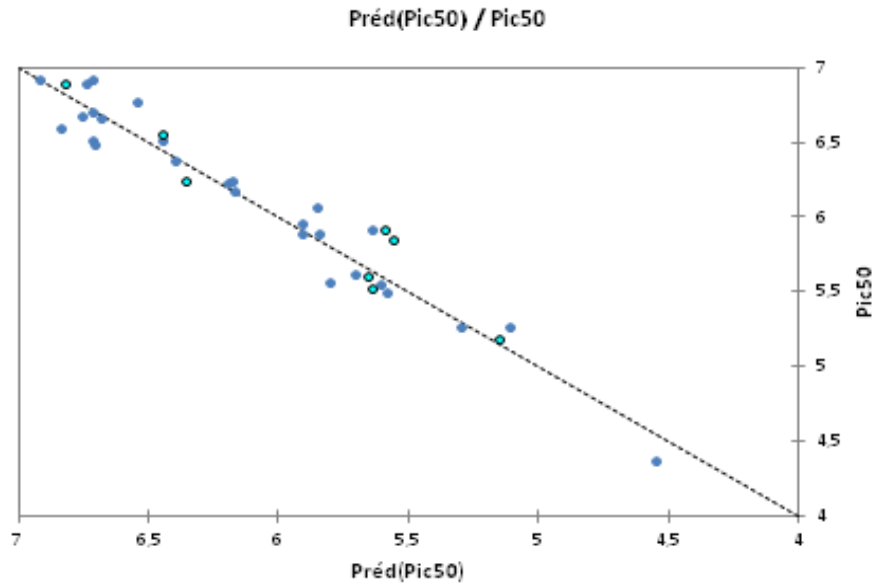


Figure III.6: Représentation graphique de l'activité calculée et observée par le MNLR.

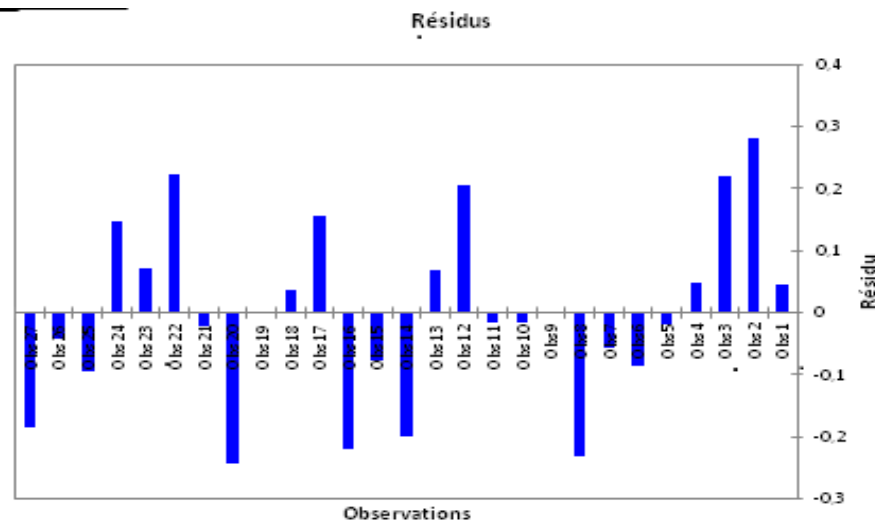


Figure III.7 : Corrélacion entre les valeurs résiduelles et l'expérimentale de modèle MNLR.

III.4. Validation du meilleur modèle QSAR :

III.4.1. Validation interne :

La dernière étape du développement d'un modèle QSAR est la validation du modèle. Afin de tester la validité du pouvoir prédictif du modèle MLR sélectionné, par opposition aux

méthodes de régression traditionnelles, la méthode de validation croisée estime la fiabilité d'un modèle en prédisant les données. Cette méthode utilise moins de paramètres validés de manière croisée:

PRESS (carrés de la somme résiduelle prédite) Sa valeur étant inférieure à SSY souligne que le modèle prédit mieux que le hasard et peut être considéré comme statistiquement important, SSY (somme des carrés de la valeur de réponse) En outre, pour un modèle raisonnable de QSAR, le rapport PESS/TSS doit être inférieur à 0,4, Les résultats présentés dans le tableau III.8 indiquent que ce rapport est de 0,083 pour le modèle 13.

Le modèle QSAR développé présente une faible valeur de Spress (<0.200) ce qui indique que le modèle présente une petite valeur résiduelle entre l'activité biologique observée et prédite.

Tableau III.8: Paramètres de validation croisée.

Modèle	PRESS	SSY	PRESS/SSY	Spres	R ²
13	0.841	10.016	0.083	0.176	0.916

III.4.2. Validation externe :

Cette méthode consiste à prédire la propriété/activité d'une série de molécules appelée généralement série de test qui ne sont pas dans la série de développement du modèle, cette validation est caractérisée par le paramètre R² test.

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 Tropsha».

Les indicateurs statistiques des cinq (5) critères Tropsha de ces deux modèles (MLR et MNLR) de l'ensemble de validation sont donnés dans le tableau III.9.

Tableau III.9: Paramètres de validation croisée.

Modèle	R ² _{test}	R ² ₀	R' ² ₀	(R ₀ -R' ² ₀)	(R ² -R ² ₀)/R ²	(R ² -R' ² ₀)/R ²	K	K'
MLR	0.909	0.995	0.844	0.151	0.094	0.071	1.006	0.993
MNLR	0.903	0.986	0.853	0.133	0.091	0.055	1.0011	0.988

D'après le tableau III.9, qui regroupe les résultats de la validation externe, ils s'avèrent que :

- Les paramètres statistiques pour la série de tests sont très satisfaisants $R^2(\text{test}) = 0,896 > R^2(\text{modèle}), R^2_{cv} = 0,916 >$ (échantillon) pour MLR et MNLR

Toutes les normes Tropsha sont validées.

- La valeur de $k = 1,00$ ($k' = 0,993$) pour MLR et ($0,988$) pour MNLR sont très proches de 1 ce qui indique que l'activité, en utilisant notre meilleure équation de modèle, est très proche de la valeur expérimentale.
- Le résultat ne justifie que le test métrique classique à 100% développé par Roy et al. (2015a) pour un test de biais de modèle QSAR est bon et en bon accord avec d'autres normes énoncées par Golbraikh et Tropsha.
- Les résultats de la vérification externe montrent que le meilleur modèle QSAR (Eq01) est hautement prédictif et peut être utilisé pour prédire l'activité antipaludique, pour composés benzophénone.

III.4.3. Y-Randomisation de MLR et MNLR :

Dans le but d'établir que le modèle obtenu n'est pas dû au hasard ou à une sur-spécification, nous avons appliqué le test de randomisation. Ainsi 100 nouveaux vecteurs de y_{pcic50} ont été générés par permutation des positions des composantes du vecteur réel :

$$y = (y_1, y_2, \dots, y_{27})' \xrightarrow{RND} y_{RND} = (y_8, y_5, \dots, y_2)'$$

Et utilisés comme sources d'observations pour des modèles QSAR dans les conditions optimales établies (2 paramètres).

Les résultats des 100 itérations de l'Y-randomisation sont données dans le tableau III.10, il s'avère que les modèles élaborés avec les valeurs randomisées de Y (équation 1) ont des R^2_r . D'autre part Roy [1++] a proposé un paramètre, nommé R^2_p .

$$R^2_p = R^2 (R^2 - R^2_r)^{\frac{1}{2}}$$

Avec R^2_r est la valeur moyenne de $R^2_r = 0,184$ de modèle obtenu avec l'activité randomisée. Ce paramètre doit être inférieur à $R^2(0,916)$.

Pour le modèle QSAR la valeur de ce paramètre $R^2_p = 0,829 > 0,5$, ce résultat indique que meilleur modèle existe réellement.

Les résultats de l'Y-randomisation pour les 100 itérations sont présentés dans le Tableau III.10. On a constaté que toutes les valeurs de R^2_r des modèles aléatoires sont plus faibles que le R^2 correspondant au modèle d'origine. Ce résultat indique que les modèles obtenus ne sont pas dues à une chance.

Les paramètres statistiques issus du deuxième modèle présenté par l'Equ.2 et des autres modèles sont cités dans le tableau suivant Tableau III.11.

Pour ce modèle le R^2_r est la valeur moyenne de $R^2_r = 0,390$ de modèle obtenu avec l'activité randomisée. Ce paramètre doit être inférieur à $R^2(0,946)$.

Pour le modèle QSAR la valeur de ce paramètre $R = 0,614$ doit être inférieur à $R(0,973)$, et Pour le modèle QSAR la valeur de ce paramètre $R^2_p = 0,733 > 0,5$, ce résultat indique que modèle existe réellement.

Tableau III.10: 100 itérations d'Y-randomisation de MLR.

Modèle	R	R ²	Q ²	Modèle	R	R ²	Q ²
1	0,630	0,397	-3,016	51	0,462	0,214	-0,238
2	0,343	0,117	-0,744	52	0,387	0,150	-1,034
3	0,257	0,066	-16,643	53	0,557	0,311	-0,365
4	0,281	0,079	-2,701	54	0,272	0,074	-5,311
5	0,655	0,429	-0,331	55	0,239	0,057	-1,955
6	0,621	0,385	-0,984	56	0,242	0,059	-0,609
7	0,555	0,308	-0,115	57	0,416	0,173	-1,893
8	0,454	0,206	-0,539	58	0,671	0,450	-0,850
9	0,795	0,633	0,020	59	0,532	0,283	-0,891
10	0,446	0,199	-0,533	60	0,302	0,091	-0,755
11	0,345	0,119	-5,705	61	0,427	0,183	-0,669
12	0,417	0,174	-4,674	62	0,651	0,424	-0,367
13	0,177	0,031	-1,830	63	0,467	0,218	-0,903
14	0,362	0,131	-0,763	64	0,188	0,035	-6,726
15	0,385	0,148	-1,013	65	0,522	0,272	-0,337
16	0,469	0,220	-1,660	66	0,335	0,112	-4,339
17	0,419	0,176	-4,840	67	0,658	0,433	-1,820
18	0,454	0,206	-1,788	68	0,563	0,317	-0,764
19	0,283	0,080	-4,653	69	0,510	0,260	-0,816
20	0,606	0,367	-6,487	70	0,425	0,181	-0,243

CHAPITRE III: Application et Résultat et Discussion

21	0,395	0,156	-0,986	71	0,345	0,119	-4,052
22	0,272	0,074	-6,267	72	0,284	0,080	-0,698
23	0,403	0,162	-7,141	73	0,324	0,105	-2,142
24	0,643	0,414	0,118	74	0,371	0,137	-3,597
25	0,155	0,024	-6,381	75	0,330	0,109	-3,269
26	0,303	0,092	-1,004	76	0,384	0,147	-5,669
27	0,387	0,150	-1,582	77	0,247	0,061	-5,940
28	0,247	0,061	-3,467	78	0,490	0,240	-0,911
29	0,569	0,324	-5,559	79	0,190	0,036	-14,012
30	0,430	0,185	-0,939	80	0,364	0,133	-0,741
31	0,418	0,175	-0,959	81	0,235	0,055	-1,276
32	0,344	0,118	-2,523	82	0,371	0,137	-1,621
33	0,337	0,114	-5,094	83	0,396	0,156	-3,381
34	0,335	0,112	-0,475	84	0,381	0,145	-0,689
35	0,358	0,128	-0,943	85	0,678	0,460	-0,974
36	0,438	0,192	-0,342	86	0,377	0,142	-1,662
37	0,499	0,249	-1,885	87	0,497	0,247	-0,250
38	0,448	0,201	-0,764	88	0,317	0,100	-3,461
39	0,414	0,171	-0,910	89	0,504	0,254	-0,819
40	0,277	0,076	-0,471	90	0,332	0,110	-4,691
41	0,638	0,408	-2,961	91	0,375	0,140	-1,159
42	0,213	0,045	-23,438	92	0,231	0,053	-0,405
43	0,233	0,054	-0,370	93	0,570	0,325	-1,458
44	0,285	0,081	-5,180	94	0,250	0,062	-3,359
45	0,355	0,126	-1,122	95	0,585	0,343	-0,589
46	0,396	0,157	-0,494	96	0,313	0,098	-3,186
47	0,260	0,067	-4,460	97	0,496	0,246	-8,035
48	0,514	0,264	-2,814	98	0,466	0,218	-0,320
49	0,341	0,116	-6,956	99	0,461	0,213	-0,153
50	0,587	0,345	-0,644	100	0,321	0,103	-4,720

Tableau III.11: 100 itérations d'Y-randomisation de MNL.R.

Modèle	R	R ²	Q ²	Modèle	R	R ²	Q ²
1	0,605	0,366	-728,47	51	0,455	0,207	-3249,15
2	0,739	0,546	-635,244	52	0,566	0,321	-3975,25
3	0,564	0,318	-511,993	53	0,537	0,289	-11728,1
4	0,483	0,233	-1376,99	54	0,577	0,333	-2797,07
5	0,664	0,440	-7336,91	55	0,625	0,390	-15,995
6	0,675	0,456	-3,584	56	0,664	0,440	-207,966
7	0,741	0,549	-120,202	57	0,565	0,320	-3488,29
8	0,781	0,610	-2537,7	58	0,770	0,593	-63,444
9	0,701	0,491	-738,561	59	0,691	0,478	-1089,81
10	0,488	0,239	-2068,59	60	0,755	0,570	-73,819
11	0,452	0,205	-102,825	61	0,471	0,222	-4858,93

12	0,725	0,525	-586,024	62	0,713	0,508	-45,287
13	0,531	0,282	-5664,4	63	0,827	0,684	-39,679
14	0,363	0,132	-707,175	64	0,699	0,489	-683,373
15	0,699	0,489	-854,46	65	0,607	0,369	-17,328
16	0,775	0,600	-4041,85	66	0,403	0,163	-1389,36
17	0,686	0,471	-3,135	67	0,657	0,431	-5949,05
18	0,735	0,541	-4068,99	68	0,575	0,331	-412,063
19	0,602	0,362	-19649,1	69	0,503	0,253	-1019,81
20	0,705	0,498	-29,214	70	0,575	0,330	-3478,92
21	0,612	0,375	-4049,32	71	0,482	0,232	-132,801
22	0,642	0,412	-9957,96	72	0,560	0,313	-1790,95
23	0,795	0,632	-1000,5	73	0,418	0,175	-435,004
24	0,853	0,727	-2685,88	74	0,547	0,299	-13347,6
25	0,495	0,245	-3739,13	75	0,792	0,628	-1199,26
26	0,306	0,093	-5534,21	76	0,540	0,291	-208,479
27	0,598	0,357	-536,604	77	0,564	0,318	-2161,34
28	0,628	0,395	-1082,64	78	0,646	0,418	-12195,7
29	0,494	0,244	-11592,7	79	0,750	0,563	-9802,12
30	0,617	0,381	-4645,88	80	0,637	0,406	-1,041
31	0,498	0,248	-5849,85	81	0,571	0,326	-4880,4
32	0,755	0,570	-3,909	82	0,681	0,464	-619,371
33	0,483	0,233	-8906,12	83	0,506	0,256	-7499,55
34	0,432	0,186	-1500,32	84	0,625	0,390	-138,666
35	0,545	0,297	-2274,96	85	0,742	0,551	-215,724
36	0,653	0,427	-7614,75	86	0,502	0,252	-4434,99
37	0,776	0,602	-4865,75	87	0,635	0,403	-113,47
38	0,716	0,513	-2532,53	88	0,522	0,273	-3978,3
39	0,570	0,325	-112,065	89	0,804	0,647	-2437,98
40	0,491	0,241	-7,085	90	0,554	0,307	-331,379
41	0,624	0,389	-886,115	91	0,602	0,363	-3787,72
42	0,758	0,576	-3745,39	92	0,534	0,285	-8213,98
43	0,728	0,530	-609,885	93	0,441	0,195	-10405,3
44	0,650	0,423	-507,368	94	0,535	0,286	-2650,22
45	0,523	0,274	-1,972	95	0,419	0,175	-306,577
46	0,600	0,360	-8,081	96	0,671	0,451	-1360,63
47	0,688	0,474	-12506,5	97	0,601	0,361	-17397,4
48	0,756	0,571	-4743,48	98	0,653	0,427	-236,647
49	0,719	0,517	-1899,15	99	0,626	0,392	-505,909
50	0,549	0,301	-10,246	100	0,715	0,511	-698,544

III.4.4. Domaine d'application :

Ces résultats montrent le grand pouvoir de prédiction des modèles MLR et MNLR; mais même les modèles les plus exhaustifs, dignes de confiance et validés, ne peuvent prédire des propriétés de manière fiable pour l'intégralité des composés chimiques existants.

Le domaine d'application (DA) permet de définir la zone dans laquelle un composé pourra être prédit avec confiance. Le DA correspond donc à la région de l'espace chimique incluant les composés d'apprentissage et les composés similaires, proches dans ce même espace. En effet, un modèle QSAR n'est pas destiné à être employé en dehors de son domaine d'application, c'est-à-dire en dehors de l'espace chimique couvert par son jeu d'entraînement. La détermination des DA est donc d'une grande importance.

Le domaine d'applicabilité du meilleur modèle QSAR MLR obtenu (Eq. 01), est présenté sur la Figure III.8. Ce domaine est une représentation des valeurs des résidus de prédiction standardisés des composés pour les deux sous-groupes (apprentissage et test) en fonction de leurs leviers (leverage) respectifs (La distance entre les valeurs des descripteurs et leur barycentre).

Si un composé a un levier qui dépasse le seuil $h^* = 3p/n$ (ou p est le nombre de descripteurs plus 1 et n le nombre d'observations (training set)), ce composé est considéré comme un composé influent sur modèle élaboré.

Les valeurs de l'effet de levier et des résidus normalisés des observables du modèle utilisé pour développer le domaine d'applicabilité de ce modèle sont présentées dans le tableau III.12.

L'analyse des données du tableau III.12 montre que toutes les observations ont leurs résidus standardisés entre -2 et 2. Les remontées obtenues sont toutes inférieures à la valeur critique $h^* = 0,66$ sauf deux composés (N° 8 et 25) avec un h légèrement supérieur à la valeur critique ($h > h^* = 0,66$). Ce fait est élucidé par le diagramme de Williams (figures III.8)

Après avoir comparé ces composés avec le reste des molécules, nous pouvons remarquer que les seules différences sont liées à la faible valeur de Pic50 pour les molécules (N°8=4.37, N°25=6.92) et à la valeur de descripteur MAXDN élevée (N°8= 2.159, N°25= 5.878) par rapport au reste.

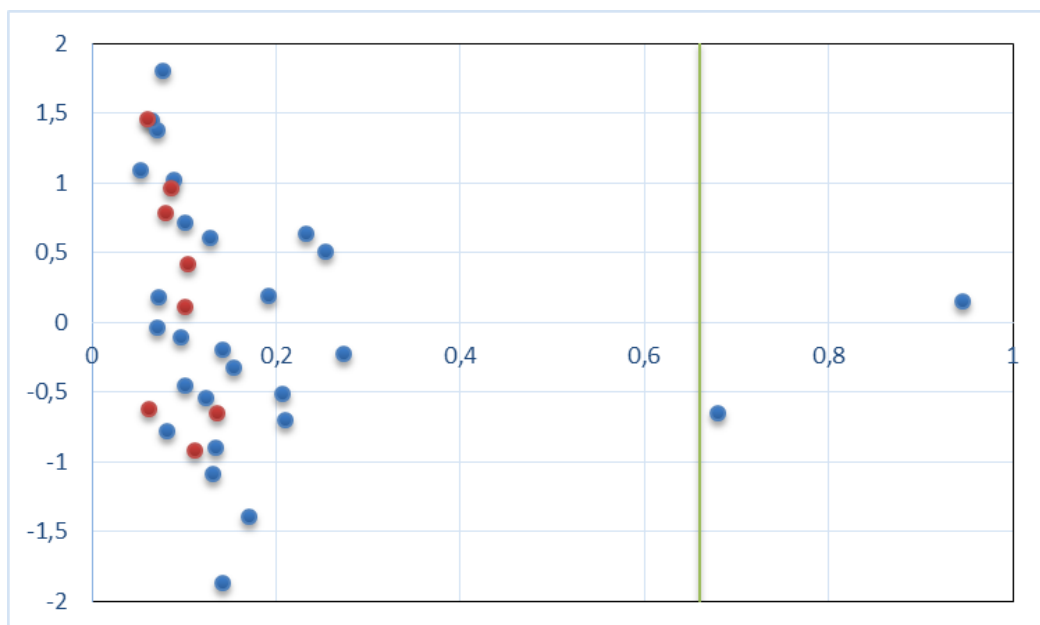


Figure III.8: Diagramme de Williams du modèle MLR.

Ce sont les mêmes observations que l'on peut déduire de l'étude du domaine d'application du modèle MNL. Il y a deux composés hors domaine qui sont toujours 8 et 25.

L'applicabilité du modèle peut être évaluée avec les gammes de descripteurs, valeurs minimales et maximales pour la série de composés modélisés, donnés dans le Tableau III.12.

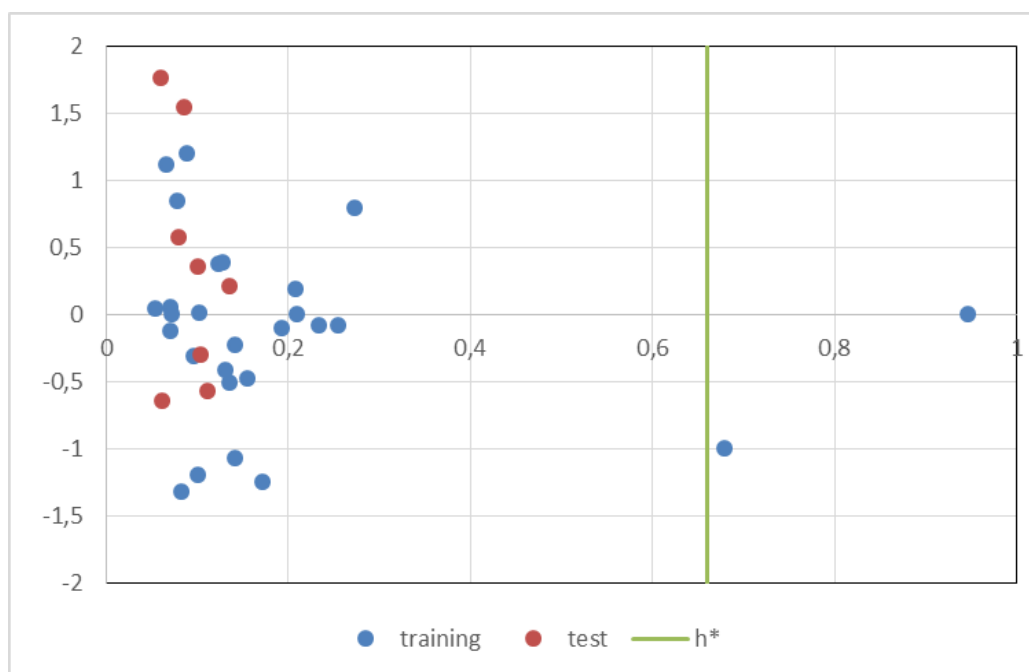


Figure III.9 : Diagramme de Williams du modèle MNL.

Tableau III.12: présente les valeurs des activités expérimentales et prédites des 35 dérivés étudié et leurs résidus.

	N°	MLR		MNLR	
		Résidu	leverage	Résidu	Leverage
Apprentissage	01	0,185	0,072	0,008	0,072
	02	1,378	0,070	0,051	0,070
	03	1,089	0,053	0,040	0,053
	04	0,714	0,101	0,009	0,101
	05	0,195	0,192	-0,096	0,192
	06	-0,323	0,154	-0,471	0,154
	07	-0,105	0,096	-0,310	0,096
	08	-1,391	0,170	-1,248	0,170
	09	0,154	0,946	0,003	0,946
	10	0,639	0,232	-0,083	0,232
	11	0,509	0,253	-0,080	0,253
	12	1,445	0,065	1,115	0,065
	13	-0,537	0,123	0,376	0,123
	14	-1,863	0,141	-1,073	0,141
	15	-1,082	0,131	-0,417	0,131
	16	-0,447	0,100	-1,192	0,100
	17	1,805	0,077	0,850	0,077
	18	-0,511	0,207	0,195	0,207
	19	-0,701	0,209	0,004	0,209
	20	-0,777	0,082	-1,317	0,082
	21	-0,040	0,070	-0,119	0,070
	22	1,024	0,088	1,202	0,088
	23	0,611	0,128	0,389	0,128
	24	-0,229	0,273	0,791	0,273
	25	-0,902	0,134	-0,507	0,134
	26	-0,194	0,141	-0,229	0,141
	27	-0,654	0,679	-0,995	0,679

Groupe de test	28	0,959	0,085	1,547	0,085
	29	1,454	0,060	1,766	0,060
	30	-0,620	0,061	-0,645	0,061
	31	0,418	0,103	-0,298	0,103
	32	-0,654	0,135	0,213	0,135
	33	0,110	0,100	0,352	0,100
	34	0,786	0,079	0,571	0,079
	35	-0,921	0,111	-0,565	0,111

Références bibliographiques :

- [1] V. N. Viswanadhan, M.R.Reddy, R.J.Bacquet et M.D.Erion, *journal of computational Chemistry*, 14, 1993, 1019.
- [2] V. N. Viswanadhan, A.K. Ghoseet, J.N. Weinstein, *biochimica et biophysicaActa*, 356, 1990, 1039.
- [3] T. Liu, et al, *BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities*. *Nucleic acids research*, **35** (1), 2007, 198-201.
- [4] www.xlstat.com.
- [5] OECD Guidance document on the validation of QSAR models Organization for Economic Co-operation & Development, Paris 2007.
- [6] A.Golbraikh, and A.Tropsha. (2002) Beware of QSAR. *Journal of Molecular Graphics and Modelling*, 20. 269-327.
- [7] LB.Kier LB, Hall LH. *Pharm Res* 7, 1990, 801-807.
- [8]MO.Akinwande, HG.Dikko, Samson A *Open J Stat* 05, 2015, 754-767.
- [9] Ajeet, k. bijander, *Int. J. pharm. Fti.* 4, 445.
- [10] P.P. Roy, S. Paul, I. Mitra, K. Roy, *Molecules*. 14, 2009, 1660.

CONCLUSION GENERAL

CONCLUSION GENERAL :

Dans notre travail, Une étude quantitative de la relation structure-activité (QSAR) a été utilisée pour prédire les activités inhibitrices de l'enzyme FTase.

Un ensemble de données composé de 43 composés a été sélectionné, puis différents types de descripteurs moléculaires ont été calculés par le programme PADEL. L'ensemble de données a été divisé en un ensemble d'apprentissage (80% de l'ensemble de données) et un ensemble de test (20% de l'ensemble de données) Dans notre étude en utilisant l'analyse des composants principaux pour les descripteurs les plus significatifs.

Pour prédire l'activité inhibitrice de FTase d'une série de substitués, deux modèles sans ambiguïté ont été développés dans cette étude avec des descripteurs topologiques et électroniques. Une bonne stabilité et une capacité de prédiction ont été démontrées par les modèles MLR et MNLR, sur le même ensemble de descripteurs.

Les meilleurs modèles QSAR ont été sélectionnés avec $r^2 = 0.916$, $F = 45.847$, $s = 0.04$ et $r^2 = 0.946$, $F = 42$, $RMSE = 0.185$ statistiquement significatif pour les deux modèles MLR et MNLR respectivement.

Les meilleurs modèles QSAR obtenus ont montré non seulement une signification statistique, mais aussi une capacité prédictive (R^2_{test} est élevé et tous les critères de la validation externe Tropsha sont vérifiés) et a révélé que les faibles valeurs de pic_{50} combinées avec la grande valeur de descripteur une augmentation de l'activité biologique.

Les meilleurs modèles contiennent également un descripteur de lipophile / hydrophile ($\log P$) ce qui suggère que la différence de solubilité entre les phases, aqueuse et hydrophobe, joue un rôle crucial sur le pouvoir antipaludique de ces composés benzophénone.

Le coefficient de randomisation Y a également été effectué pour les deux modèles MLR et MNLR et ont des valeurs significatives de R^2 , Q^2 respectivement supérieure à 0,5 à l'appui de l'affirmation selon laquelle le modèle généré est puissant et non inféré par hasard.

Pour évaluer la fiabilité de nos modèles QSAR et ses capacités à prédire de nouveaux composés, le domaine d'applicabilité doit être essentiellement défini pour chaque modèle.

À partir de ce graphique, les valeurs de levier (h_i) de tout composé dans les ensembles d'apprentissage et de test sont inférieures à la valeur critique ($h^* = 0,55$) sauf les composés 8 et 25 en acceptant les comme valeurs aberrantes.

CONCLUSION GENERAL :

Toutes les validations (randomisation et domaine d'applicabilité) ont indiqué que les modèles QSAR construits étaient robustes et satisfaisants.

De plus, les résultats obtenus de chaque modèle sur cette série de composés sont assez similaires, aucun des modèles établis n'est considéré meilleur que l'autre. Ainsi, les modèles MLR et MNLR sont considérés comme des outils efficaces pour prédire l'activité inhibitrice de FTase sur la base des descripteurs proposés.

