

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي

BADJI MOKHTAR- ANNABA UNIVERSITY

UNIVERSITE BADJI MOKHTAR- ANNABA



جامعة باجي مختار- عنابة

Année 2018

Faculté des Sciences
Département de Chimie

THÈSE

Présentée pour obtenir le diplôme de

Doctorat Option : Chimie organique

***Étude théorique pour l'identification des
molécules biologiquement actives (approche
QSAR, DOCKING, et VIRTUEL SCREENING)***

THÈME

Par :

Mr. Mohamed AISSAOUI

Devant le jury :

Mr. Sallah-Eddine DJILANI	Président	Professeur	Université d'Annaba
Mr. Abdelhafid DJEROUROU	Directeur de thèse	Professeur	Université d'Annaba
Mr. Farhi HALAIMIA	Examineur	Professeur	Université d'Annaba
Mr. Malek BENDJABALLAH	Examineur	MCA	Université de Guelma
Mr. Abdelmalek KHORIEF NACEREDDINE	Examineur	MCA	ENSET de Skikda

ÉTUDE THÉORIQUE POUR L'IDENTIFICATION DES MOLÉCULES
BIOLOGIQUEMENT ACTIVES (APPROCHE QSAR, DOCKING, ET VIRTUEL
SCREENING)

Mohamed AISSAOUI

©date d'édition 10 juillet 2018

10 juillet 2018

Dédicaces

A :

Mes Parents ;

Mes sœurs ;

Mes collègues de laboratoire ;

Mes chères amis ;

Tous ceux qui m'ont soutenue pendant ces années ;

Je dédie ce travail

Remerciements

Le travail présenté dans ce manuscrit a été réalisé en collaboration entre le Laboratoire de Synthèse et Biocatalyse Organique (LSBO) à l'Université Badji Mokhtar Annaba-Algérie, et le laboratoire de Modélisation Moléculaire et Virtuel Screening (MMVSL) à l'Université de Pise-Italie.

Je tiens tout d'abord à remercier Monsieur DJEROUROU Abdelhafid pour m'avoir tant apporté durant cette aventure. Je le remercie de m'avoir encadré et soutenu pendant ces quatre années. Merci de la confiance que tu m'as accordé, de m'avoir fait partager ta passion pour la chimie, de ta grande disponibilité, de ta patience et de tous les conseils que tu as pu me donner.

Je remercie vivement le Pr. Salah-Eddine DJILANI pour sa participation à ce jury en tant que président. J'exprime également ma reconnaissance aux Pr. Farhi HALAIMIA, Pr. Malek BENDJABALLAH et le Dr. Abdelmalek KHORIEF NACEREDDINE pour avoir accepté de juger ce travail.

Je remercie vivement monsieur Tiziano Tuccinardi, professeur à l'université de Pise, pour m'avoir accueilli dans son laboratoire de Modélisation Moléculaire et Virtuel Screening (MMVSL) durant mon stage et ma bourse d'étude en Italie .

Un grand Merci à tous mes amis, ALI RACHEDI Sofiane, LOUZIM Habiba, Mostafa MO-KHNANE, Margherita LAPPILO, Zahra LACHTAR, Farida ZEGHAD, Houria BENTOUMI, Nawal CHETIBI, Salah LAKROUT. Je leur souhaite à tous bon courage pour la suite.

J'adresse mes vifs remerciements à tous mes collègues du Laboratoire de Synthèse et Biocatalyse Organique. Nombreuses sont les personnes qui m'ont aidé à l'élaboration de ce travail. C'est aussi à elles que s'adressent mes remerciements et ma sympathie.

Et enfin un grand merci tout spécial à ma mère qui partage ma vie et qui endure au quotidien les difficultés et les incertitudes de la vie d'un doctorant chercheur notamment sa faible disponibilité et son humeur inégale.

Abbreviations

0D	Zérounidimensionnels
1D	Unidimensionnels
2D	Bidimensionnels
3D	Tridimensionnels
4D	Quadimensionnels
ACP	Analyse en composantes principales
ACO	Ant Colony Optimization
ADN	Acide désoxyribonucléique
AG	Algorithmes Génétiques
AM1	Austin Model 1
AMP	Adénosine monophosphate
ARG 99	Arginine 99
ARG 112	Arginine 112
ASP52	Asparagine 52
ATP	Adénosine triphosphate
Ar	Aryle
BMR	Bactéries multi-résistantes
CA	Carbonyle alpha
CMI	Concentration Minimale Inhibitrice
CNN	Computational Neural Network
CoMFA	Comparative Molecular Field Analysis
DFT	Density Functional Theory
DM	Dynamique moléculaire
<i>E. coli</i>	<i>Escherichia coli</i>
ECBU	Examen Cytobactériologique des Urines
EMME	L'éthoxyméthylènemalonate de diéthyle
Equ	Equation
FAD	La flavine adénine dinucléotide
Fdg	Fluorodésoxy-glucose
FS	Forward Stepwise
FT	Facteur de température
GLIDE	Grid-based Ligand Docking with Energetics
GLY 97	Glycine 97
GOLD	Genetic Optimisation for Ligand Docking
gyrA	Gyrase A
gyrB	Gyrase B
HTS	High-Throughput Screening
ILE 116	Ileucine 116

LDH	La Lactate Déshydrogénase
Lig	Ligand
LMO	leave-many out
LOO	leave-one out
M	Masse moléculaire
NADH	Nicotinamide Adénine Dinucléotide
ns	nano seconde
PDB	Proteine Data Bank
PET	Emission de positon
PfLDH	La lactate déshydrogénase de <i>Plasmodium falciparum</i>
PHE 119	Phenylalanine 119
PLANTS	Protein-Ligand ANT System
PSET	Prédiction Set
QRDR	Quinolone resistance determining region
QSAR	Quantitative Structure-Activity Relationships
QSPR	Quantitative Structure Propriety Relationships
®	Signe accompagnant les marques de commerce spécifique
RLM	Régression linéaire multiple
$Q^2_{Loo_CV}$	Le coefficient de corrélation issu du cross validation
RMSD	Root mean square deviation
s	l'écart type
SD	standard deviation
SP	<i>standard precision</i>
SVM	support machines vector
TSET	Training Set
VAL 53	Valine 53
VIF	Variance Inflation Factor
VSDA	analyse de l'espace vectoriel de descripteurs
VS	Virtual Screening
XP	<i>extra precision</i>

الملخص:

لقد تم تقديم دراستين مستقلتين في هذا العمل من الأطروحة.

يخصص الجزء الأول لتطوير منهجية تستخدم للحصول على علاقات تربط النشاط البيولوجي بهيكلية وكمية المركبات، ولتطوير نموذج QSAR باستعمال مجموعة من الجزئيات.

ينقسم هذا الجزء إلى ثلاثة فصول: الفصل الأول يعبر عن عموميات على الكينولونات كعوامل مضادة للجراثيم وأيضاً مقاومة البكتيريا بالنسبة لهذه العائلة، أما الفصل الثاني يقوم بوصف منهجية QSAR بمختلف مراحلها، فيما يخص الفصل الثالث فقد خصص لتطبيق منهجية QSAR لنمذجة النشاط المضاد للبكتيريا *E. Coli* المعبر عنه بالعلاقة $-\text{Log MIC}$ باستعمال 85 مركب من مشتقات الكينولونات، وذلك باستخدام الانحدار الخطي المتعدد والخوارزميات الجينية التي تستخدم في تطوير النموذج والاختيار على التوالي.

وقد تم تطوير نموذج QSAR باستخدام جميع الواصفات المتوفرة في برنامج DRAGON 6. النتائج التي تم الحصول عليها بعد التحقق وتحليل القيم القياسية المتبقية تثبت صحة واستقرار وثبات النموذج الذي تم الحصول عليه وامكانيته من توفير 90 % من التباين الملاحظ في قيم النشاط البيولوجي.

ويتناول الجزء الثاني دراسة تصف غريلة افتراضية على أساس التفاعل مع الموقع النشط للبروتين. يخصص هذا الجزء من الأطروحة لإيجاد أدوية جديدة مضادة للسرطان وذلك عن طريق إيجاد مثبطات لأنزيم نازعة هيدروجين اللاكتات.

على مدى السنوات القليلة الماضية، يعتبر انزيم نازعة هيدروجين اللاكتات على نحو متزايد هدفاً محتملاً للعوامل العلاجية، بما في ذلك الأدوية المضادة للملاريا والأورام السرطانية.

ينقسم هذا الجزء إلى ثلاثة فصول: الفصل الأول يعبر عن العموميات على أنزيم نازعة هيدروجين اللاكتات ودورها في آلية التحلل وإبراز بعض الأمثلة التي وجدت فيما سبق كمثبطات لهذا الإنزيم، أما الفصل الثاني هو الجانب النظري في الكيمياء الحسابية والبرمجيات المستخدمة خلال منهجية الغريلة الافتراضية، الفصل الثالث يهدف لتحليل هيكل البروتين الخاص بدراستنا (PDB CODE: 4QSM) وفي الأخير تطبيق إجراء الغريلة الافتراضية باستخدام نهج التفاعل مع الموقع الفعال للأنزيم وذلك من خلال استعمال قاعدة البيانات التجارية Enamine والتي تحتوي على أكثر من مليوني مركب، وبالنسبة لاختيار المركبات فهو يعتمد على عدة معايير.

دراسات باستخدام المحاكاة الجزيئية الديناميكية أدت في نهاية المطاف إلى إيجاد 9 جزيئات مستقرة، مع تفاعلات مثيرة للاهتمام مع الموقع النشط. وقد اعتبرت 5 جزيئات واعدة بشكل خاص للقيام بالفحوصات الأنزيمية.

الكلمات المفتاحية: QSAR، الانحدار الخطي المتعدد، الخوارزميات الجينية، الكينولونات، بكتيريا *E. Coli*، مثبط، نازعة هيدروجين اللاكتات، *Molecular Docking*، الغريلة الافتراضية، أنزيم.

Résumé

Deux études indépendantes ont été présentées dans ce travail, la première est consacrée à la mise au point d'une méthodologie employée pour obtenir des relations quantitatives, structure-activité (QSAR) et pour le développement d'un modèle QSAR à partir d'un ensemble de molécules.

Cette partie rapporte des généralités sur les quinolones comme des agent antibactériens et la résistance bactérienne vis-à-vis cette famille, la méthodologie de QSAR et à la fin l'application de la méthodologie de QSAR, pour modéliser l'activité antibactérienne vis-à-vis E.Coli, exprimée par Log CMI à partir d'un ensemble de 85 molécules, dérivés des quinolones en utilisant la méthode de régression linéaire multiple et les algorithmes génétiques qui sont utilisés dans le développement de modèle en tant que méthode d'apprentissage et de sélection respectivement.

Un modèle QSAR a été développé par l'utilisation de la totalité des descripteurs issus de logiciel DRAGON 6. Les résultats obtenus de la validation et l'analyse des valeurs résiduelles normalisées prouvent la validité, la stabilité et la robustesse de modèle obtenu. Il peut expliquer 90% de la variance des valeurs de l'activité biologique observées.

La seconde partie portera sur le Virtuel Screening à base de modèle pharmacophore (pharmacophore-based). Cette partie de thèse est consacrée à l'identification des nouveaux anticancéreux, des inhibiteurs de l'enzyme lactate déshydrogénase (LDH). Cet enzyme a été de plus en plus considéré comme une cible potentielle pour les agents thérapeutiques, y compris les agents antipaludéens et antitumoraux.

Cette partie rapporte des généralités sur le lactate déshydrogénase, son rôle dans le mécanisme de glycolyse et à quelques inhibiteurs dans la littérature de cet enzyme, un aspect théorique sur la chimie computationnelle et les logiciels utilisés pendant la procédure de Virtuel Screening, l'analyse de notre structure protéique de code PDB : 4QSM et l'application d'une procédure de Virtuel Screening à base de modèle pharmacophore sur le database commercialisé Enamine en passant par plusieurs critères pour la sélection des composés.

Des études utilisant la simulation dynamique moléculaire conduisant éventuellement à l'obtention de 9 molécules stables, avec des interactions intéressante avec le site actif. 5 molécules ont été considérées comme particulièrement prometteuses et seront achetées pour être soumises à des dosages enzymatiques.

Mots-Clés : QSAR, RLM; Algorithmes Génétique; quinolones; E.Coli; LDH; Docking moléculaire; Virtuel Screening; Modèle pharmacophore; enzyme.

Abstract

Two independent studies have been presented in this thesis work.

The first part is devoted to the development of a methodology used to obtain quantitative structure activity relationships, and for the development of a QSAR model from a set of molecules.

This part reports generalities on quinolones as antibacterial agents and bacterial resistance to this family, the methodology of QSAR and the application of the QSAR methodology for modeling the antibacterial activity against *E. Coli*, expressed by Log MIC from a set of 85 quinolone derivatives, using the multiple linear regression and genetic algorithms that are used in the development of the model and selection respectively.

A QSAR model has been developed using all the descriptors from DRAGON 6 software. The results obtained from the validation and the analysis of the standardized residual values prove the validity, the stability and the robustness of the obtained model. It can provide 90% of the observed variance in biological activity values.

The second part deals with pharmacophore-based Virtual Screening. This part of the thesis is devoted to the identification of new anticancer drugs Such as lactate dehydrogenase (LDH) inhibitors. Over the past few years, the LDH enzyme has been increasingly considered as a potential target for therapeutic agents, including antimalarial and antitumor agents.

This part reports generalities on lactate dehydrogenase, its role in the glycolysis mechanism and some inhibitors of this enzyme in the literature, a theoretical aspect on computational chemistry and the software used during the Virtual Screening procedure, at the end analyze of the protein structure (PDB CODE : 4QSM) and the application of Virtual Screening procedure using pharmacophores approach with the commercialized Enamine database (more than 2 million compounds), passing through several criteria for the selection of compounds.

Studies using dynamic molecular simulation eventually leading to nine stable molecules, with interesting interactions with the active site. Five molecules have been considered particularly promising and will be purchased for enzymatic assays.

Keywords : QSAR ;MLR ; Genetic Algorithm ; quinolones ; *E.Coli* ; inhibitor ; LDH ; Molecular Docking ; Virtual Screening ; Phamacophore model ; enzyme.

Table des figures

1	Les différentes étapes de la modélisation moléculaire.	2
1.1	Structure de 7-chloroquinoline.	8
1.2	Structure d'acide 4-oxo-1,4-dihydroquinoline-3-carboxylique.	9
1.3	La découverte de l'acide nalidixique.	9
1.4	Des structures de quelques dérivés quinolones à partir de 7-chloroquinoline.	10
1.5	Structure de Norfloxacin.	10
1.6	Structure de Ciprofloxacine.	10
1.7	Structure commun des quinolones.	11
1.8	Relation structure activité des quinolones.	12
1.9	Structure d'acide 1-éthyl-7-méthyl-4-oxo-1,4-dihydro-1,8-naphthyridine-3-carboxylique.	13
1.10	Structure d'acide 8-éthyl-5-oxo-2-(pyrrolidin-1-yl)-5,8-dihydropyrido [2,3- <i>d</i>]pyrimidine-6- carboxylique.	13
1.11	Structure d'acide 1-cyclopropyl-6-fluoro-4-oxo-7-(pipérazin-1-yl)-1,4-dihydroquinoline-3- carboxylique.	14
1.12	Structure d'acide (3 <i>S</i>)-9-fluoro-3-méthyl-10-(4-méthylpiperazin-1-yl)-7-oxo-2,3-dihydro-7 <i>H</i> -[1,4] oxazino[2,3,4- <i>ij</i>]quinoline- 6-carboxylique.	14
1.13	. Structure d'acide 1-cyclopropyl-6-fluoro-7-((4 <i>aS</i> ,7 <i>aS</i>)-hexahydro-1 <i>H</i> -pyrrolo[3,4- <i>b</i>]pyridin-6(2 <i>H</i>)- yl)-8-methoxy-4-oxo-1,4-dihydroquinoline-3-carboxylique.	15
1.14	Structure d'acide 7-((1 <i>R</i> ,5 <i>S</i> ,6 <i>s</i>)-6-amino-3-azabicyclo[3.1.0]hexan-3-yl)-1-(2,4-difluorophenyl)- 6-fluoro-4-oxo-1,4-dihydro-1,8-naphthyridine-3-carboxylique.	15
1.15	Structure commune des quinolones	16
1.16	Acide 1-éthyl-7-méthyl-4-oxo-1,4-dihydro-1,8-naphthyridine-3-carboxylique.	16
1.17	Acide 8-éthyl-5-oxo-2-(piperazin-1-yl)-5,8-dihydropyrido[2,3- <i>d</i>]pyrimidine-6-carboxylique.	16
1.18	Dénomination commune des quinolones.	17
1.19	a) Structure de fluméquine ; b) Structure d'ofloxacin	18
1.20	a) Structure d'Amifloxacine ; b) de Miloxacin.	19
1.21	La structure de Cinoxacin.	19
1.22	Structure de 1-méthyl-1 <i>H</i> -tétrazol-5-yl.	19
1.23	Synthèse de l'acide nalidixique selon la méthode de GOULD et JACOBS.	21
1.24	Synthèse de la Ofloxacin.	22
1.25	Synthèse des quinolones selon la cyclisation de DIECKMANN.	23
1.26	Synthèse des quinolones selon la méthode de GROHE.	24
1.27	Mécanisme d'action des quinolones.	27

TABLE DES FIGURES

1.28 Mécanisme de résistance bactérienne aux quinolones.	28
2.1 Rôle des Descripteurs moléculaires en relation quantitative structure activité.	34
2.2 Présentation des étapes aboutissant à QSAR.	35
2.3 Les principales étapes de la construction du modèle de QSAR.	36
2.4 L'importance des descripteurs moléculaire pour l'information structurelle.	37
2.5 Rôle clé des descripteurs moléculaires dans la recherche scientifique.	38
2.6 Schéma représentatif d'un réseau de neurones artificiels.	45
3.1 Les pourcentages d'examen microbiologique d'urine.	54
3.2 Structures développées des dérivés des quinolones	56
3.2 Structures développées des dérivés des quinolones <i>continuation...</i>	57
3.2 Structures développées des dérivés des quinolones <i>continuation...</i>	58
3.2 Structures développées des dérivés des quinolones <i>continuation...</i>	59
3.3 Le nombre optimal de descripteurs du modèle optimal obtenu par GA	62
3.4 Valeurs résiduelles normalisées en fonction de la valeur CMI expérimentale.	68
3.5 Valeurs prédites en fonction des valeurs observées pour la série d'apprentissage(TSET) et la série de prédiction(PSET).	71
3.6 Graphe de Williams obtenu avec les cinq descripteurs.	71
3.7 La contribution des descripteurs dans le modèle.	72
3.8 Comparaison des potentiels électrostatiques moléculaires (MEP) entre les composés N°30 et N°26.	73
3.9 Comparaison structurale entre les composés N°20 et N°48.	74
4.1 La réaction catalysée par la lactate déshydrogénase.	83
4.2 Domaines de liaison du LDH-A; en jaune le domaine du cofacteur, avec le NADH, en vert le domaine du substrat, avec le pyruvate (chaîne B, cristal 4JNK).	85
4.3 Chevauchement de chaîne B cristal 1H10 (boucle fermée, en bleu) et chaîne C cristal 4JNK (boucle ouverte, rouge). En rose l'inhibiteur oxydé; en jaune le cofacteur NADH.	86
4.4 Lactate (bleu) dans le site catalytique avec le cofacteur (vert).	87
4.5 Structure de Gossipol.	92
4.6 Structure d'inhibiteur FX11.	93
4.7 Structure d'Oxamate.	93
4.8 Dérivés de N-hydroxyindole.	93
4.9 Structure d'inhibiteurs bivalents d'AstraZeneca.	94
4.10 Complexation du cuivre (a) et du zinc (b).	95
4.11 (a) Quinoléine et (b) 1,4-dihydro-4-quinolones.	95
4.12 Acide 4-hydroxy-1, 2, 5-oxadiazole-3-carboxylique.	96
4.13 Acide (R)-3-(5-amino-6-((1-phenylethyl)amino)pyrazin-2-yl)-4-chlorobenzoïque; code PDB : 22Y.	96
4.14 2-((5-cyano-4-(3,4-dichlorophenyl)-6-oxo-1,6-dihydropyrimidin-2-yl)thio)-N-(4-sulfamoylphenyl)propanamide; Code PDB : ZHK.	97
4.15 2',6'-dichloro-4-((2-chlorophenyl)thio)-5-hydroxy-1,6-dihydro-[1,1'-biphenyl]-3(2H)-one; Code PDB : 36U.	97
5.1 Interface du logiciel Chimera.	105
5.2 Interface du logiciel Maestro.	106

TABLE DES FIGURES

5.3	Interface du logiciel Ligandscout.	108
5.4	Principe général d'un programme de Docking.	109
5.5	Interface d'Autodock 4.	110
5.6	Interface du logiciel Glide.	112
5.7	Interface du logiciel Dock6.	116
6.1	Les différentes étapes du procédure de Virtuel Screening.	124
6.2	Modélisation moléculaire des inhibiteurs identifiés avec le ligand de référence (composé 2) : (A) composé 5; (B) composé 6; (C) composé 7; (D) composé 8; (E) composé 9; (F) ligand de référence (38K).	125
6.3	Structures développées des ligands identifiés avec le ligand référence.	126
6.4	La structure cristallographique du 4QSM : chaîne A (gris), B (bleu), C (violet), D (Vert clair), E (vert foncé), F (jaune), G (orangé), H (rouge).	127
6.5	Chevauchement des huit chaînes, détail sur le site catalytique. chaîne A (gris), B (bleu), C (violet), D (Vert clair), E (vert foncé), F (jaune), G (orangé), H (rouge).	128
6.6	Les interactions 3D du composé 2 avec le site actif.	129
6.7	Les interactions 2D du composé 2 avec le site actif.	130
6.8	Les interactions 3D de composé 9 avec le site actif.	131
6.9	Les interactions 2D de composé 9 avec le site actif.	131
6.10	a) Le modèle pharmacophore avec les deux ligands actifs; b) le modèle pharmacophore seul.	135
6.11	Virtual screening utilisant le logiciel Ligandscout.	136
6.12	Graphe d'enrichissement calculé sur la valeur du XP_Gscore de deux ligands actifs vis à vis 73 218 inactifs.	137
6.13	Résultats du Docking pour les 73 218 molécules utilisant Glide_XP.	138
6.14	L'analyse visuelle des résultats de Docking utilisant Chimera.	139
6.15	Graphe du RMSD des carbones α de la protéine (en bleu), du ligand de référence (en vert) et du ligand 9 (en orange).	141
6.16	Graphe de RMSD pour la molécule VS_f17_Z1127125184 pendant la DM.	145
6.17	Exemple d'une DM : en gris la conformation initiale, en orange les différentes étapes de la DM; étape 1 (a), étape 28 (b), étape 100 (c).	145
6.18	Graphe de RMSD pour la molécule VS_f17_Z1317967696 pendant la DM.	147
6.19	DM VS_f17_Z1317967696; comparaison entre l'étape 1 (à gauche) et l'étape 100 (à droite). En gris la conformation initiale de référence; en violet la molécule pendant la dynamique.	148
6.20	Graphe de RMSD pour la molécule VS_f17_Z226451846 pendant la DM.	148
6.21	DM VS_f17_Z226451846; comparaison entre l'étape 1 (à gauche) et l'étape 100 (à droite). En gris la conformation initiale de référence; en violet la molécule pendant la dynamique.	149
6.22	Graphe de RMSD pour la molécule VS_f17_Z16507440 pendant la DM.	150
6.23	DM VS_f17_Z16507440; comparaison entre l'étape 1 (à gauche) et l'étape 100 (à droite). En gris la conformation initiale de référence; en violet la molécule pendant la dynamique.	151
6.24	Graphe de RMSD pour la molécule VS_f17_Z16507442 pendant la DM.	151
6.25	DM VS_f17_Z16507442; comparaison entre l'étape 1 (à gauche) et l'étape 100 (à droite). En gris la conformation initiale de référence; en violet la molécule pendant la dynamique.	152
6.26	Graphe de RMSD pour la molécule VS_f17_Z16507998 pendant la DM.	153

TABLE DES FIGURES

6.27	DM VS_f17_Z16507998 ; comparaison entre l'étape 1 (à gauche) et l'étape 100 (à droite). En gris la conformation initiale de référence ; en violet la molécule pendant la dynamique.	154
6.28	Graphe de RMSD pour la molécule VS_f17_Z979261714 pendant la DM.	154
6.29	DM VS_f17_Z979261714 ; comparaison entre l'étape 1 (à gauche) et l'étape 100 (à droite). En gris la conformation initiale de référence ; en violet la molécule pendant la dynamique.	155
6.30	Graphe de RMSD pour la molécule VS_f17_Z16508305 pendant la DM.	156
6.31	DM VS_f17_Z16508305 ; comparaison entre l'étape 1 (à gauche) et l'étape 100 (à droite). En gris la conformation initiale de référence ; en violet la molécule pendant la dynamique.	157
6.32	Graphe de RMSD pour la molécule VS_f17_Z358317798 pendant la DM.	157
6.33	DM VS_f17_Z358317798 ; comparaison entre l'étape 1 (à gauche) et l'étape 100 (à droite). En gris la conformation initiale de référence ; en violet la molécule pendant la dynamique.	158
6.34	Graphe de RMSD pour la molécule VS_f17_Z17766169 pendant la DM.	159
6.35	DM VS_f17_Z17766169 ; comparaison entre l'étape 1 (à gauche) et l'étape 100 (à droite). En gris la conformation initiale de référence ; en violet la molécule pendant la dynamique.	160
6.36	La structure développée des 9 molécules stables après la DM.	160

Liste des tableaux

1.1	Effet de la N_1 substitution sur l'activité antibactérienne (<i>Klebsiellapneumoniae</i>) d'analogues de l'acide nalidixique..	18
1.2	Les caractéristiques physico-chimiques de Lévofoxacine.	25
1.3	Les caractéristiques physico-chimiques de Ciprofloxacine.	25
2.1	Liste des blocs de descripteurs fournis par le logiciel Dragon.	40
3.1	Liste des cinq descripteurs sélectionnés par les AG.	60
3.2	Série des modèles optimums obtenus à différentes dimensions.	61
3.3	Valeurs CMI expérimentales et prédites pour TSET et PSET calculé par le modèle RLM-GA. . .	63
3.3	Valeurs CMI expérimentales et prédites pour TSET et PSET calculé par le modèle RLM-GA. . .	64
3.3	Valeurs CMI expérimentales et prédites pour TSET et PSET calculé par le modèle RLM-GA. . .	65
3.4	Résultats de l'analyse de la variance (ANOVA).	66
3.5	Matrice de corrélation de l'équation.	67
3.6	Valeurs des critères VIF et TF pour les descripteurs significatifs.	68
3.7	R^2 et Q_{cv-100}^2 issus du test de randomisation.	70
6.1	Les valeurs du dosage enzymatique et score de liaison total.	126
6.2	Les valeurs de FT de Arg99 dans la chaîne A et B.	128
6.3	Les logiciels de Docking utilisés.	132
6.4	Les logiciels de Docking utilisés.	133
6.5	Résultats du programme Glide_XP sur les deux ligands actifs d'après la littérature.	136
6.6	Les interactions du ligand de référence pendant les différentes étapes de DM.	141
6.7	Les interactions du ligand 9 pendant les différentes étapes de DM.	142
6.8	Les interactions du ligand VS_f17_Z1317967696 pendant les différentes étapes de DM.	147
6.9	Les interactions du ligand VS_f17_Z226451846 pendant les différentes étapes de DM.	149
6.10	Les interactions du ligand VS_f17_Z16507440 pendant les différentes étapes de DM.	150
6.11	Les interactions du ligand VS_f17_Z16507442 pendant les différentes étapes de DM.	152
6.12	Les interactions du ligand VS_f17_Z16507998 pendant les différentes étapes de DM.	153
6.13	Les interactions du ligand VS_f17_Z979261714 pendant les différentes étapes de DM.	155
6.14	Les interactions du ligand VS_f17_Z16508305 pendant les différentes étapes de DM.	156
6.15	Les interactions du ligand VS_f17_Z358317798 pendant les différentes étapes de DM.	158
6.16	Les interactions du ligand VS_f17_Z17766169 pendant les différentes étapes de DM.	159
6.17	Liste des descripteurs constitutionnels générés par le logiciel Dragon.	167

6.18 Présentation de quelques descripteurs moléculaires du bloc de groupements fonctionnels.	169
6.19 Quelques descripteurs topologiques fournis par logiciel Dragon.	170

Table des matières

Introduction générale	2
------------------------------	----------

I ÉTUDE QSAR SUR LES QUINOLONES ET LA RESISTANCE BACTÉRIENNE

1 Généralités sur les quinolones et la résistance bactérienne	7
1.1 Introduction	8
1.2 Définition	8
1.3 Historique	8
1.4 Les aspects généraux des quinolones	11
1.4.1 Structure chimique	11
1.4.2 Classification	12
1.4.3 Nomenclature	16
1.4.4 Analyse fonctionnelle du squelette quinolone	17
1.5 Synthèse chimique	20
1.5.1 Méthode de GOULD et JACOB	20
1.5.2 Cyclisation de DIECKMANN	22
1.5.3 Cyclisation d'une enamine-ester par la méthode de GROHE	22
1.6 Propriétés physicochimiques	24
1.6.1 Les caractéristiques physico-chimiques de Lévofoxacine	24
1.6.2 Les caractéristiques physico-chimiques de Ciprofloxacine	25
1.7 Mécanisme d'action des quinolones	25
1.7.1 Inhibition d'ADN gyrase	26
1.7.2 Inhibition des ADN topo-isomérases IV	26
1.8 Résistance Bactérienne aux quinolones	27
1.8.1 Définition de la résistance	27
1.8.2 Mécanisme de la résistance bactérienne	27
1.9 Conclusion	29
2 Aspect théorique sur la méthodologie de QSAR et les descripteurs moléculaires	33
2.1 Introduction	33
2.2 Présentation de la méthode de QSAR	35
2.2.1 Introduction du Data Set	35
2.2.2 Calcul et optimisation	35

2.2.3	Développement du modèle	41
2.2.4	Validation du modèle	45
2.3	Conclusion	49
3	Modélisation de la Concentration minimale Inhibitrice (CMI) des quinolones vis-à-vis la bactérie <i>Escherichia coli</i> (<i>E.Coli</i>)	53
3.1	Introduction	53
3.2	Méthodologie	55
3.2.1	Base de données	55
3.2.2	Introduction et optimisation des structures	59
3.2.3	Génération des descripteurs	59
3.2.4	Analyse statistique	59
3.3	Résultats et discussion	61
3.3.1	Sélection du modèle optimale	61
3.3.2	Analyse de la justesse du modèle optimal	62
3.3.3	Tests de colinéarité et multicollinéarité	66
3.3.4	Analyse des valeurs résiduelles normalisées	68
3.3.5	Validation interne et externe du modèle de QSAR obtenu	69
3.3.6	Domaine d'applicabilité	70
3.3.7	Analyse de la contribution des descripteurs dans le modèle	72
3.4	Conclusion	75
II	STRATEGIE DE VIRTUEL SCREENING A BASE DE PHARMACOPHORE POUR L'IDENTIFICATION DES NOUVEAUX INHIBITEURS DU L'ENZYME LACTATE DÉSHYDROGÉNASE	79
4	Généralité sur l'enzyme Lactate Déshydrogénase et ses inhibiteurs	80
4.1	Introduction	81
4.2	Métabolisme du glucose	81
4.2.1	Les trois voies métaboliques pour le pyruvate	81
4.3	Lactate déshydrogénase	83
4.4	Caractéristiques de l'isotype LDH-A	84
4.5	Mécanisme catalytique	85
4.6	Métabolisme du cancer	86
4.6.1	Effet Warburg	88
4.6.2	Mécanismes moléculaires	90
4.7	Autres possibilités thérapeutiques	90
4.7.1	Lactate déshydrogénase et paludisme	91
4.8	Les inhibiteurs du LDH dans la littérature	92
4.8.1	Gossipol et ses dérivés	92
4.8.2	Oxamate	92
4.8.3	Dérivés de N-hydroxyindole	93
4.8.4	Les Inhibiteurs bifonctionnels	94
4.8.5	Galloflavine	94

4.8.6	Complexes de cuivre (II) et de zinc (II)	94
4.8.7	Quinoléine et 1,4-dihydro-4-quinolones	95
4.8.8	Les azoles	95
4.8.9	Récents découvertes de la société Genetech Inc	96
4.9	Conclusion	98
5	Aspect théorique sur la chimie Computationnelle et les logiciels utilisés	103
5.1	Introduction	103
5.2	Programmes de visualisation	104
5.2.1	UCSF CHIMERA	104
5.2.2	MAESTRO	105
5.2.3	LIGANDSCOUT	106
5.3	Programmes de Docking	108
5.3.1	Définition du Docking	108
5.3.2	Les algorithmes génétiques	109
5.3.3	AUTODOCK	110
5.3.4	AutoDock VINA	111
5.3.5	GLIDE	112
5.3.6	GOLD	114
5.3.7	DOCK6	115
5.3.8	PLANTS	115
5.4	La simulation dynamique moléculaire	116
5.4.1	Principes théoriques	116
5.4.2	Calcul de dynamique moléculaire	117
5.4.3	Programme pour la dynamique moléculaire	117
5.5	Conclusion	119
6	Virtuel Screening à base du pharmacophore pour l'identification des inhibiteurs de LDH	123
6.1	Introduction	123
6.2	Analyse de la structure protéique 4QSM (X-Ray structure), de l'inhibiteur de référence et de ses dérivés actifs	125
6.2.1	Modèle de référence (Template)	125
6.2.2	Caractéristiques de la structure cristallographique code : 4QSM	127
6.3	Analyse du self_Docking	132
6.4	Screening de Database Enamine	133
6.4.1	Database Enamine	133
6.4.2	Construction d'un modèle pharmacophore basé sur les ligands 38K et le composé 9	134
6.4.3	Virtuel Screening utilisant le software Ligandscout	135
6.4.4	Le Docking avec le programme Glide	136
6.5	Simulation dynamique moléculaire	138
6.5.1	Calibration de la procédure	139
6.5.2	Simulation dynamique moléculaire	142
6.5.3	Analyse des résultats	144
6.5.4	Analyse visuelle	145

6.6 Conclusion	161
Conclusions générales et perspectives	164
Annexe	167
6.7 Les descripteurs constitutionnels	167
6.8 Descripteurs de groupements fonctionnels	168
6.9 Les descripteurs topologiques	170
6.10 Les descripteurs géométriques	171
article	173
□	

Introduction générale

Découvrir de nouveaux médicaments de la manière la plus efficace et la moins coûteuse possible constitue un enjeu majeur pour les années à venir. Il est admis qu'en moyenne, pour une molécule qui arrive sur le marché en tant que médicament innovant, 10 000 molécules sont synthétisées et testées [1]. De plus, le développement d'un médicament demande généralement entre 10 et 15 ans de recherches. Il s'agit en effet de trouver une molécule qui doit à la fois présenter des propriétés thérapeutiques particulières, et posséder le minimum d'effets secondaires indésirables. Le prix de revient d'un médicament est essentiellement dû à ces synthèses longues et coûteuses. Les dépenses pour la recherche et développement (Research and development) étaient estimées entre 500 et 600 millions de Dollars les années quatre-vingt pour dépasser les 900 millions de Dollars de nos jours, dont deux-tiers du coût sont dépensés pour la découverte des précurseurs potentielles (leads compounds) qui ne surpassent malheureusement pas l'étape des tests précliniques[2].

Aujourd'hui, après plusieurs années de développement et d'amélioration, l'ordinateur est devenu un outil indispensable dans les différents modes de vie. Sa technologie de pointe et ses utilisations répandues ont accélérés considérablement son développement. De ce fait, plusieurs nouvelles disciplines sont apparues dans presque tous les domaines scientifiques(Figure 1) [3].

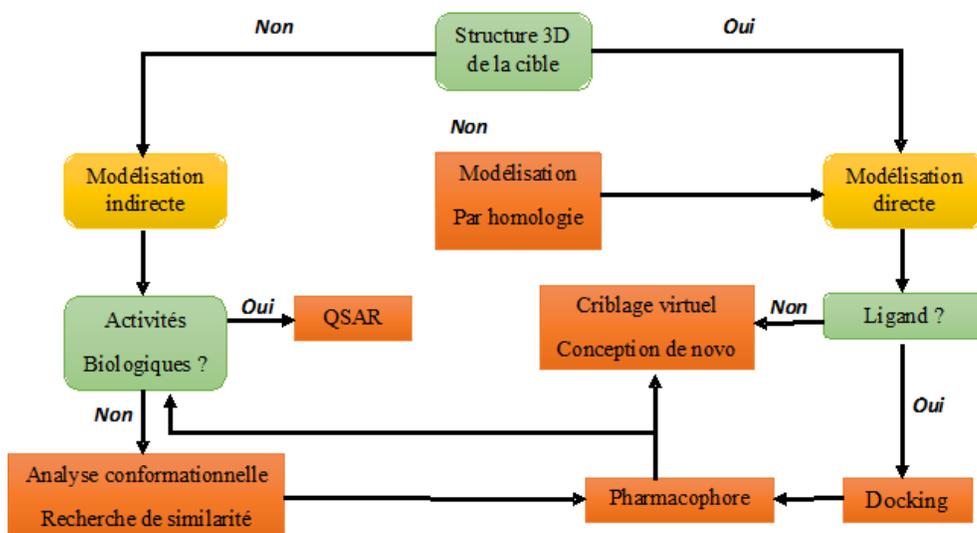


FIGURE 1 – Les différentes étapes de la modélisation moléculaire.

La chimie informatique constitue le grand exemple issu de ce développement où elle est maintenant reconnue comme subdivision de la chimie et un soutien principal à la chimie moderne. Spécialement dans la chimie des médicaments, plusieurs méthodes de recherche utilisant l'ordinateur ont été créées. Ces méthodes ont été développées pour aider les chimistes

médicinales (medicinal chemist) pour identifier et concevoir des ligands et des précurseurs qui pourraient agir plus favorablement avec un récepteur donné augmentant ainsi le nombre de composés présentant une affinité puissante[4][5].

Pour cette raison, l'industrie pharmaceutique s'oriente vers des nouvelles méthodes de recherche, qui consistent à prédire les propriétés et l'activité des molécules avant même que celles-ci ne soient synthétisées. Deux disciplines se sont développées en réponse à ce besoin : les relations structure-activité ou QSAR (Quantitative Structure-Activity Relationships) et la chimie computationnelle (Docking, Virtuel Screening, Modèle Pharmacophore, Simulation Dynamique).

Dans ce contexte, le présent travail s'articule autour de deux thèmes. Le premier thème correspond à une étude QSAR sur les quinolones, le second quant à lui, traite une stratégie de Virtuel Screening (Docking, la Simulation Moléculaire, Modélisation Pharmacophore) pour l'identification des inhibiteurs du LDH (Lactate Déshydrogénase). Le présent manuscrit est organisé sous forme de deux parties, chaque partie comprend trois chapitres.

Dans la première partie, nous nous sommes intéressés à l'étude QSAR et à la résistance bactérienne aux quinolones. Dans ce but, nous avons fait recours à la méthodologie de QSAR qui va nous amener à produire un modèle mathématique pour modéliser la CMI (concentration minimal inhibitrice) par une série des molécules à base de motif quinolone.

La présentation générale de cette partie du manuscrit sera répartie comme suit :

- **Chapitre 1.** Nous présentons ici, des notions générales sur les quinolones, leurs modes d'action, et leurs intérêts biologiques, ainsi qu'un aperçu général sur la résistance bactérienne. Ce choix est justifié par le succès marqué par cette famille des antibactériens à travers des années, ainsi que la résistance bactérienne qui pose une problématique d'après des recherches récentes qui montrent son évolution avec le temps ;
- **Chapitre 2.** Il s'agit d'un aspect théorique sur la méthodologie de QSAR, et les descripteurs générés par le software Dragon 6, leurs efficacités a été démontré lors de leurs utilisations dans plusieurs domaines de recherches : environnement, médecine ;
- **Chapitre 3.** Ce chapitre est consacré à la modélisation de la CMI (concentration minimal inhibitrice) par les dérivés de quinolone. L'étude QSAR de notre série des molécules choisies basé sur le développement d'un modèle mathématique qui lie la structure chimique des molécules au CMI (concentration minimal inhibitrice). Dans ce chapitre plusieurs logiciels ont été utilisés, le Dragon.6 pour la génération des descripteurs, Gaussian.09 pour l'optimisation des molécules, Xlsat et Minitab pour la génération de modèle QSAR.

Dans la seconde partie, le travail sera consacré pour une stratégie de Virtual Screening (criblage virtuel) pour l'identification des nouveaux inhibiteurs de l'enzyme LDH. Au cours des dernières années, cette enzyme a été de plus en plus considérée comme une cible potentielle pour les agents thérapeutiques, y compris les agents antipaludiques et antitumoraux ce qui fait appel à des approches théoriques (Docking, Simulation Dynamique Moléculaire), permettant de prédire le mode d'interaction de ces molécules avec la cible biologique (récepteur), ne tenant compte que des critères structuraux.

Le travail que nous présentons dans cette partie comprend trois chapitres :

- **Chapitre 1.** Généralité sur l'enzyme Lactate Déshydrogénase et leurs inhibiteurs ;
- **Chapitre 2.** Il s'agit d'un aspect théorique sur la chimie computationnelle, (Virtual Screening, Docking, Simulation Dynamique Moléculaire), et tous les logiciels utilisés dans notre étude, qui concerne l'étude d'interaction ligand-récepteur (Chimera), dessin et l'optimisation des molécules (Maestro), Virtual Screening (LigandScout), Docking (Glide), et Simulation Dynamique (Amber) ;
- **Chapitre 3.** La validité de LDH comme une cible anticancéreuse, rendent les techniques de VS de plus en plus attractives pour des projets d'identification de molécules bioactives sur cette cible, passant par une stratégie de Virtual Screening, qui contient plusieurs étapes de filtrage à base d'interaction ligand-récepteur. Nous détaillerons les différentes interactions entre les molécules choisies et l'enzyme cible LDH.

Enfin, une conclusion générale suivra les résultats de cette thèse et des perspectives seront proposées.

Bibliographie

- [1] Tudor I Oprea. *Chemoinformatics in drug discovery*, volume 23. Wiley Online Library, 2005.
- [2] EA Rekka and PN Kourounakis. *Chemistry and molecular aspects of drug design and action*. CRC Press, 2008.
- [3] Raymond C Rowe, Elizabeth A Colbourn, Robert Harrison, Matthew K McGowan, Richard J McGowan, Martin Romacker, Nicolas Grandjean, Pierre Parisot, Olivier Kreim, Daniel Cronenberger, et al. *Computer applications in pharmaceutical research and development*. 2006.
- [4] Felix Deanda Jr. *Development and application of software tools for computer-assisted drug design*. The University of Texas at Austin, 1999.
- [5] Hans-Joachim Böhm. New computational approaches to predict protein-ligand interactions. *Structure-Based Ligand Design*, pages 129–142, 1998.

Première partie

**ÉTUDE QSAR SUR LES QUINOLONES
ET LA RESISTANCE BACTÉRIENNE**

Chapitre 1

Généralités sur les quinolones et la résistance bactérienne

Sommaire

1.1 Introduction	8
1.2 Définition	8
1.3 Historique	8
1.4 Les aspects généraux des quinolones	11
1.4.1 Structure chimique	11
1.4.2 Classification	12
1.4.3 Nomenclature	16
1.4.4 Analyse fonctionnelle du squelette quinolone	17
1.5 Synthèse chimique	20
1.5.1 Méthode de GOULD et JACOB	20
1.5.2 Cyclisation de DIECKMANN	22
1.5.3 Cyclisation d'une enamine-ester par la méthode de GROHE	22
1.6 Propriétés physicochimiques	24
1.6.1 Les caractéristiques physico-chimiques de Lévofoxacine	24
1.6.2 Les caractéristiques physico-chimiques de Ciprofloxacine	25
1.7 Mécanisme d'action des quinolones	25
1.7.1 Inhibition d'ADN gyrase	26
1.7.2 Inhibition des ADN topo-isomérases IV	26
1.8 Résistance Bactérienne aux quinolones	27
1.8.1 Définition de la résistance	27
1.8.2 Mécanisme de la résistance bactérienne	27
1.9 Conclusion	29

1.1 Introduction

Parmi les agents antimicrobiens, la classe des quinolones suscite depuis sa découverte en 1962 un intérêt considérable avec l'acide nalidixique, et sont introduites en usage clinique en 1963, marquant ainsi le début de cinq décennies de développement et d'utilisation. Depuis, des progrès non négligeables ont été accomplis quant à notre compréhension des mécanismes moléculaires d'action des quinolones vis-à-vis des bactéries pathogènes, de l'induction de résistance parmi ces organismes, et de la capacité de chacun de ces composés à induire des effets toxiques chez les différents patients traités[1].

Avec un spectre d'activité antibactérienne étendu et des propriétés pharmacocinétiques intéressantes, les quinolones représentent une des principales classes d'agents antimicrobiens actuels[2].

1.2 Définition

Les quinolones sont les premiers agents antibactériens de synthèse utilisés chez l'homme. Bactéricides, ils forment une classe d'antibiotiques qui a pris une place importante dans l'arsenal thérapeutique. Depuis le tout premier agent : l'acide nalidixique, jusqu'aux fluoroquinolones de quatrième génération, plusieurs molécules ont été synthétisées avec chacune ses particularités[3].

1.3 Historique

En 1938 : des chercheurs dans le laboratoire de Sterling en Rochester, NY ; s'aperçoivent, lors de recherches sur les antipaludéens, de l'activité bactéricide d'un produit secondaire obtenu lors de la synthèse de chloroquine : la 7-chloroquinoline (Figure 1.1) [4].

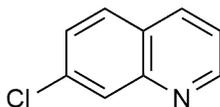


FIGURE 1.1 – Structure de 7-chloroquinoline.

En 1962 : Introduit en thérapeutique dès 1962 sous le nom de NEGRAM, l'acide -nalidixique est le chef de file historique des quinolones ; série apparentée à la 4-benzopyridone et de structure générale dérivant de l'acide 4-oxo-1,4-dihydroquinoline-3-carboxylique (Figure 1.2) [5].

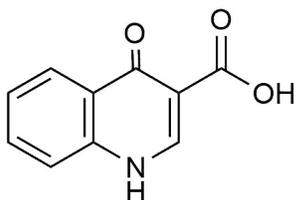


FIGURE 1.2 – Structure d'acide 4-oxo-1,4-dihydroquinoline-3-carboxylique.

L'atome d'azote y est généralement substitué de même que le cycle benzénique qui peut être remplacé par un hétérocycle aromatique (Figure 1.3).

La synthèse par LESHER de cet antibactérien a fait suite à la mise en évidence de l'activité antimicrobienne de la 7-chloroquinoline : ce dérivé mono-chloré est un sous-produit de la préparation de la 4,7-dichloroquinoléine qui est l'intermédiaire carrefour de la synthèse conduisant soit à la chloroquine (NIVAQUINE[®], Antimalarique); soit à la glafénine (GLIFANAN[®]; antalgique périphérique).

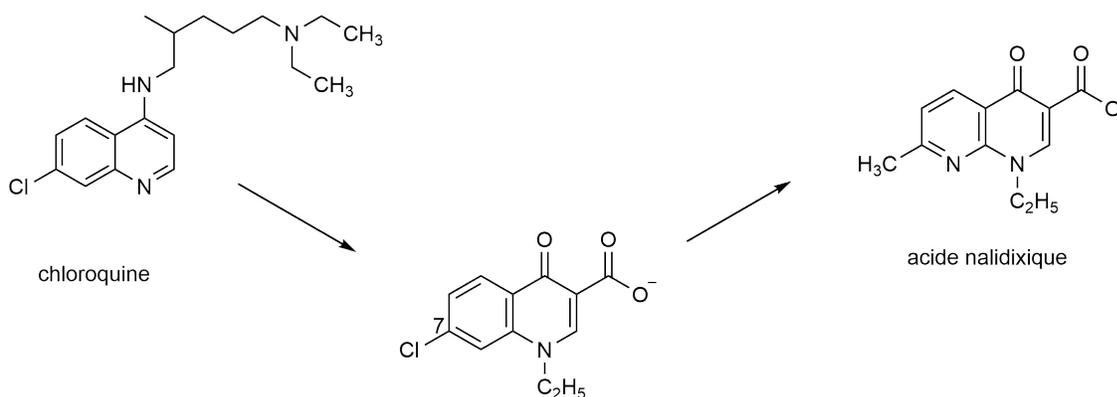


FIGURE 1.3 – La découverte de l'acide nalidixique.

Les années 60-70 : Les travaux et le développement des quinolones à partir du premier composés (la 7-chloroquinoline)(Figure 1.4) : la cinoxacine, l'acide piromidique, l'acide pipemidique, l'acide oxolinique et la flumequine[6].

Les années 80 : la KyorinCompany crée la Norfloxacin[®], première vraie fluoroquinolone, par adjonction au noyau de base de la pipérazine diamine cyclique trouvée sur l'acide pipemidique et du 6-fluor trouvé sur la flumequine[7](Figure 1.5).

En 1987 : début de la commercialisation de la Ciprofloxacin[®](Figure 1.6), le premier composé à avoir une vraie activité sur les bactéries Gram (+)[8].

En 1991 : L'apparition de l'Ofloxacin[®] cette molécule découverte en 1980 par des chercheurs japonais[9] est commercialisée sous le nom de marque Ofloce[®] par les laboratoires Aventis (en Belgique : Tarivid[®] - Sanofi-Aventis). La molécule est commercialisée au Ca-

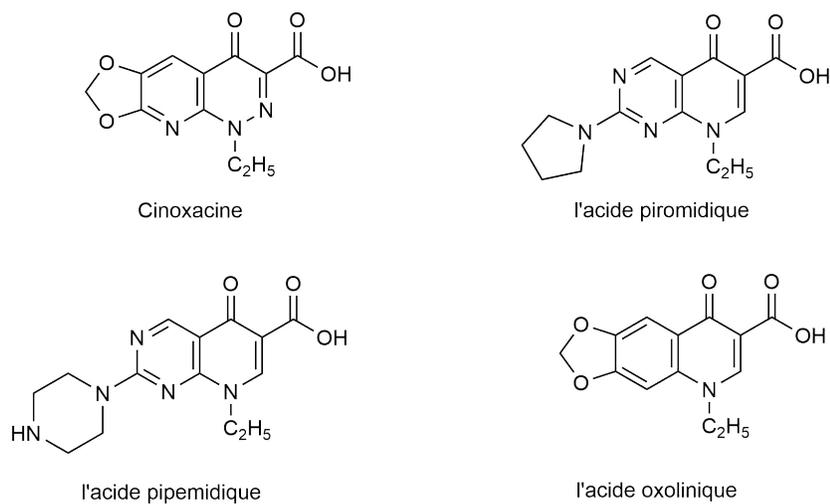


FIGURE 1.4 – Des structures de quelques dérivés quinolones à partir de 7-chloroquinoline.

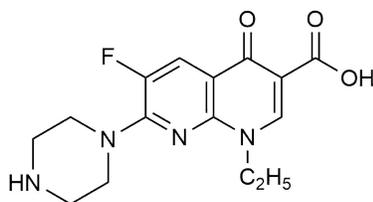


FIGURE 1.5 – Structure de Norfloxacin.

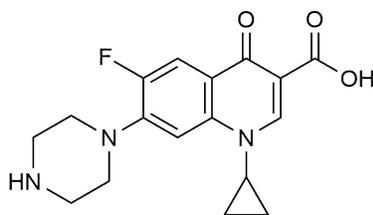


FIGURE 1.6 – Structure de Ciprofloxacin.

nada sous le nom de Floxin® par Janssen-Ortho. Elle fait partie de la liste des médicaments essentiels de l'Organisation mondiale de la santé (liste mise à jour en avril 2013)[10].

Les Années 90 : L'apparition de la troisième génération et les fluoroquinolones anti-pneumococciques, dont la Lévofoxacin® forme le lévogyre de l'Ofloxacin est le chef de file[11].

Les années 2000 : quatrième génération avec un spectre élargi aux anaérobies strictes (Gémifloxacin®).

Les premiers temps ont vu la production de composés à spectre restreint aux bactéries gram (-) qui étaient utiles seulement pour que les infections urinaires. Aujourd'hui, il existe des composés qui couvrent presque tous les pathogènes connus, avec une diffusion tissulaire et intracellulaire remarquable[12].

Les recherches ont eu pour objectifs :

- D'exalter le pouvoir antibactérien (la CMI à l'égard des entérobactéries est abaissée en moyenne de 100 à 250 fois, notamment pour les plus récents dérivés introduits en thérapeutique : Ofloxacin[®] et Ciprofloxacine) et d'élargir le spectre d'action aux germes à Gram positif, à certaines bactéries à développement intracellulaire ;
- D'améliorer les caractéristiques pharmacocinétiques (biodisponibilité, distribution tissulaire) conférant aux quinolones récentes une activité systémique.

1.4 Les aspects généraux des quinolones

1.4.1 Structure chimique

Toutes les quinolones ont un noyau de base commun(Figure 1.7)[13], Il s'agit des molécules planes qui possèdent :

- Deux cycles aromatiques accolés. Le premier est un cycle pyridine, le deuxième est un hétérocycle (pyridine ou pyrimidine) ou un benzène ;
- Une fonction cétone en position 4 ;
- Une fonction carboxylique COOH en position 3.

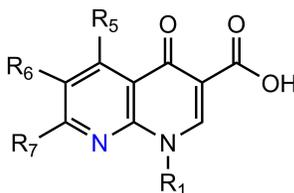


FIGURE 1.7 – Structure commun des quinolones.

Les molécules de ce groupe se distinguent par (Figure 1.8)[14] :

- La nature du substituant R_1 porté par l'azote hétérocyclique de la quinoléine, responsable de l'activité antibactérienne globale ;
- La nature du substituant R_6 : une racine fluor dans les fluoroquinolones, responsable de la pénétration pariétale et de la liaison à l'ADN gyrase ;
- Le substituant R_7 porté par le noyau quinoléine, qui définit le spectre antibactérien de la molécule ;
- Le remplacement du noyau phényle condensé avec un cycle pyridine dans le noyau quinoléine par un autre hétérocycle aromatique, pyridine ou pyrimidine.

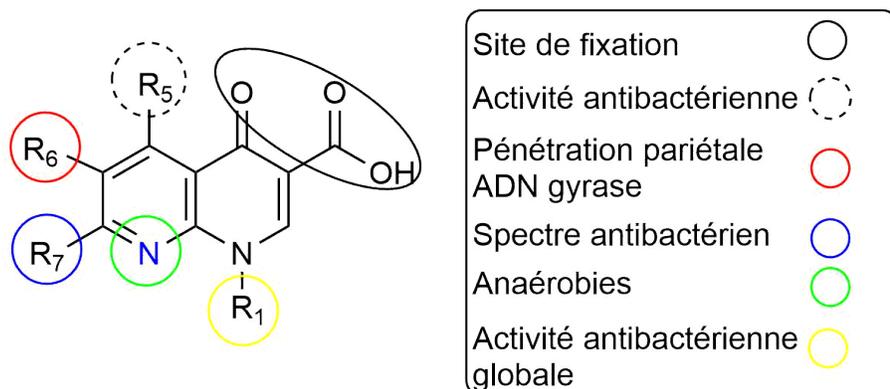


FIGURE 1.8 – Relation structure activité des quinolones.

1.4.2 Classification

Tout en conservant le noyau de base, la famille des quinolones s'est vue enrichie à travers les années de molécules de plus en plus performantes et avec des propriétés thérapeutiques plus larges.

On peut classer les quinolones en fluorées et non fluorées, ou en générations selon la chronologie d'apparition, mais cette classification n'a aucun intérêt pour le praticien. En 1997, une classification qui prend en compte en plus de la chronologie, le spectre d'action bactérien, les propriétés pharmacocinétiques et les indications communes a été proposée[15].

1.4.2.1 Première génération

Les quinolones de la première génération ont un spectre étroit et limité contre les bactéries et aussi contre les germes de type Gram négatif et comprend :

- L'acide Nalidixique (Negram[®], Negram Forte[®]) (Figure 1.9);
- La Cinoxacine (Cinobac[®]);
- L'acide Pipemidique (Pipram[®], Pipram Forte[®]);
- L'acide Oxolinique (Uroxin[®]);
- La flumequine (Flubactin[®], usage vétérinaire);

Les deux premiers sont des antiseptiques urinaires purs en raison de leur mauvaise diffusion dans les tissus. Cette génération comprend un seul produit fluoré : la fluméquine.

Les quinolones de première générations sont actives sur les germes à Gram négatif incluant ou non le bacille pyocyanique (*Pseudomonas aeruginosa*).

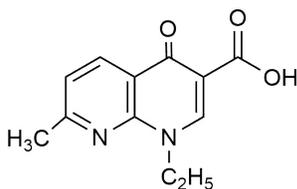


FIGURE 1.9 – Structure d'acide 1-éthyl-7-méthyl-4-oxo-1,4-dihydro-1,8-naphthyridine-3-carboxylique.

DCI : acide nalidixique NEGRAM[®] 1965

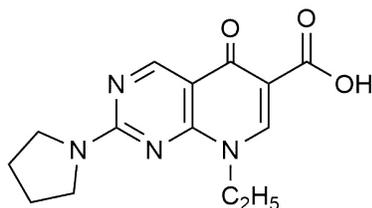


FIGURE 1.10 – Structure d'acide 8-éthyl-5-oxo-2-(pyrrolidin-1-yl)-5,8-dihydropyrido [2,3-*d*]pyrimidine-6-carboxylique.

DCI : acide piromidique PURIM[®] 1981

1.4.2.2 Deuxième génération

La deuxième génération a vu l'apparition des fluoroquinolones, avec un spectre plus élargi sur les Gram (-) et quelque Gram (+), et une meilleure diffusion tissulaire donc de nouvelles indications en thérapeutique.

- Norfloxacin (Noroxine[®]);
- Ciprofloxacin (CifloxR, Cipro[®])(Figure 1.11);
- Ofloxacin (Oflocet[®], Monoflocet[®]);
- Lomefloxacin (Logiflox[®], Decalogiflox[®]);
- Péfloxacin (Peflacine[®]);
- Enoxacin (Enoxor[®]).

Cette génération comprend des produits administrables par voie parentérale (l'Ofloxacin et la Ciprofloxacin). Toutefois, elle est demeurée dénuée d'activité sur le streptocoque.

Ces Nouvelles quinolones bénéficient de plus :

- D'une bonne absorption digestive autorisant une antibiothérapie par voie orale, ce qui est particulièrement nouveau et intéressant dans le champ d'utilisation de ces anti-infectieux ;
- D'exceptionnelles distributions et concentrations tissulaires permettant le traitement d'infections pulmonaires, O.R.L., cutanées, osseuses, etc.

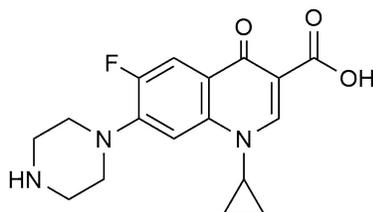


FIGURE 1.11 – Structure d'acide 1-cyclopropyl-6-fluoro-4-oxo-7-(pipérazin-1-yl)-1,4-dihydroquinoline-3- carboxylique.

DCI : ciprofloxacin CIFLOX[®] 1988

1.4.2.3 Troisième génération

La troisième génération a été conçue pour couvrir plus de germes Gram (+), surtout le *streptocoque pneumonie* et pyogènes, en plus du spectre des générations précédentes. Elle est également active sur des germes atypiques (*Mycoplasmapneumoniae* et *Chlamydia pneumo- niae*).

- Lévofloxacine isomère de l'Ofloxacine (Tavanic[®])(Figure 1.12);
- Moxifloxacine (Izilox[®] parfois rapportée comme de 4^{me} génération)(Figure 1.13);
- Sparfloxacine (Zagam[®]);
- Pazufloxacine (Pasil[®], Pazucross[®]);
- Balofloxacine (Baloxin[®]);
- Tosufloxacine (Ozex[®], Tosacin[®]);
- Grepafloxacine (Raxar[®] retirée du marché);
- Gatifloxacine (Tequin[®], retirée du marché);
- Temafloxacine (Omniflox[®], retirée du marché).

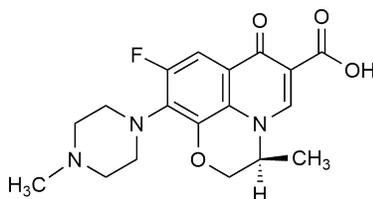


FIGURE 1.12 – Structure d'acide (3S)-9-fluoro-3-méthyl-10-(4-méthylpiperazin-1-yl)-7-oxo-2,3-dihydro-7H-[1,4] oxazino[2,3,4-ij]quinoline- 6-carboxylique.

DCI : Lévofloxacine Tavanic[®] 1997

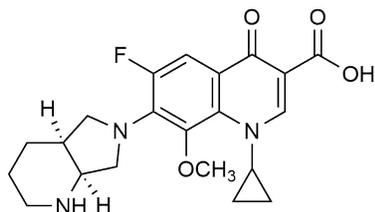


FIGURE 1.13 – . Structure d'acide 1-cyclopropyl-6-fluoro-7-((4aS,7aS)-hexahydro-1H-pyrrolo[3,4-b]pyridin-6(2H)-yl)-8-methoxy-4-oxo-1,4-dihydroquinoline-3-carboxylique.

DCI : Moxifloxacin eIzilox[®] 1999

Cette génération présente en outre un meilleur profil pharmacocinétique avec la possibilité de prise journalière unique. Cependant certaines molécules ont été retirées du marché après commercialisation suite à des effets secondaires graves.

1.4.2.4 Quatrième génération

La quatrième génération a élargi son spectre sur les germes anaérobies stricts.

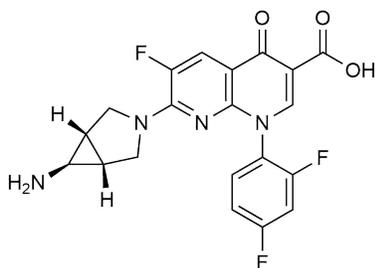


FIGURE 1.14 – Structure d'acide 7-((1R,5S,6s)-6-amino-3-azabicyclo[3.1.0]hexan-3-yl)-1-(2,4-difluorophenyl)-6-fluoro-4-oxo-1,4-dihydro-1,8-naphthyridine-3-carboxylique.

DCI : Trovafloxacin Trovan[®] 1998

Cette classe est la plus récente d'un point de vue commercial mais souffre cependant des retraits successifs des molécules ou de leurs limitations d'usage. A titre d'exemple : la Trovafloxacin(Figure 1.14) n'est plus commercialisée en Europe en raison de l'apparition de cas graves de toxicité hépatique (aux Etats-Unis elle est réservée à l'usage hospitalier et seulement quand le pronostic vital est en jeu en l'absence d'autre alternative).

1.4.3 Nomenclature

1.4.3.1 Dénomination scientifique

Dans toute quinolone, il est possible d'identifier un noyau commun A (acide 1-alkyl-4-oxo-1,4-dihydropyridine-3-carboxylique) orthocondensé avec un cycle B de nature aromatique (Figure 1.15). Bien entendu, la prééminence d'un noyau sur l'autre oriente le sens de numérotation du polycycle ainsi formé[16].

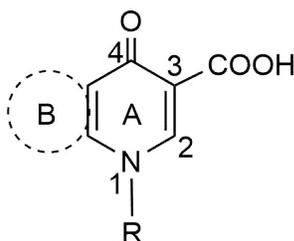


FIGURE 1.15 – Structure commune des quinolones

— 1^{ier} exemple : Acide nalixidique

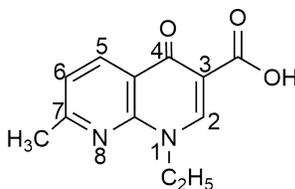


FIGURE 1.16 – Acide 1-éthyl-7-méthyl-4-oxo-1,4-dihydro-1,8-naphthyridine-3-carboxylique.

On notera que cette molécule (Figure 1.16), qui est à l'origine du nom générique des quinolones, ne présente pas l'architecture quinoléine mais un squelette 1,8-naphthyridine.

— 2^{me} exemple : Acide pipémidique (Figure 1.17).

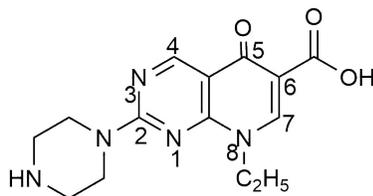


FIGURE 1.17 – Acide 8-éthyl-5-oxo-2-(piperazin-1-yl)-5,8-dihydropyrido[2,3-d]pyrimidine-6-carboxylique.

1.4.3.2 Dénominations communes

Les dérivés de quinolones de première génération sont généralement dénommés acides (Figure 1.18); vraisemblablement d'une part, afin de rappeler la présence d'une fonction acide carboxylique sur le cycle A[17], qui participe avec le groupement β -carbonylé à un des mécanismes d'action (chélation des métaux) de ces antibiomimétiques, et d'autre part, afin de souligner le caractère acide de ces premières quinolones : $pK_{a1} = 5.6$ à 6.4 (COOH).

Les dérivés de quinolones de deuxième génération par la présence en position 7 d'un cycle pipérazinique bénéficient au contraire d'un caractère amphotère, ce qui explique la commercialisation de certains d'entre eux sous forme de sels, et justifie l'abandon du terme "acide" pour un nouveau mode de dénomination, utilisant un segment-clef "oxacine" en suffixe.

Ciprofloxacine introduite en thérapeutique sous forme de chlorhydrate (CIFLOX). acide 1-cyclopropyl-6-fluoro-4-oxo-7-(pipérazin-1-yl)-1,4-dihydroquinoline-3-carboxylique, chlorhydrate, monohydraté.

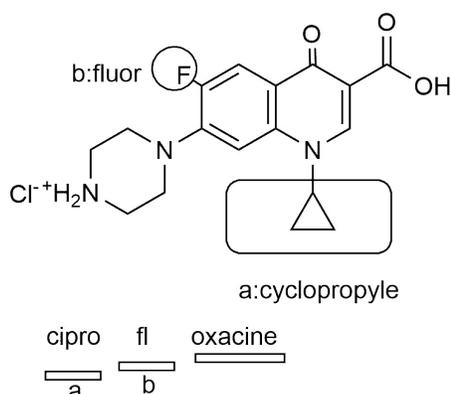


FIGURE 1.18 – Dénomination commune des quinolones.

1.4.4 Analyse fonctionnelle du squelette quinolone

1.4.4.1 Pharmacomodulations affectant le cycle A

Dans tous les composés actuellement utilisés en thérapeutique, le cycle A est toujours associé à un cycle B aromatique. La nature des groupements fonctionnels portés par ce noyau commun à toutes les quinolones, autorise les manipulations moléculaires suivantes :

1.4.4.1.1 N_1 -substitution (alkylation, cycloalkylation, arylation)

C'est actuellement le seul site de pharmacomodulation possible, sans altération fondamentale des propriétés du modèle[18]. Dans le (Tableau 1.1), sont indiquées les variations du pouvoir bactéricide in vivo de dérivés N_1 substitués de l'acide nalidixique, à l'égard de *Klebsiel-*

lapneumoniae.

Le maximum d'activité est donc obtenu pour un substituant éthyle ($\text{CH}_2\text{-CH}_3$) qui, dans le cas de la fluméquine et de l'ofloxacine (Figure 1.19) est engagé dans un système cyclique supplémentaire, conférant à ces molécules, d'une part une grande stabilité métabolique, et un centre de chiralité d'autre part.

Par exemple, dans l'ofloxacine le C_3 implique l'existence de deux énantiomères, dont l'isomère lévogyre est à l'égard de divers germes, de 8 à 128 fois plus actif que son inverse optique. Le produit commercialisé (OFLOCET) est le racémique : l'acide (R, S) 9-fluoro-3-méthyl-10-(4-méthylpipérazin-1-yl)-7-oxo-2,3-dihydro-7H-[1,4]oxazino[2,3,4-ij] quinoline-6-carboxylique [19].

TABLE 1.1 – Effet de la N_1 substitution sur l'activité antibactérienne (*Klebsiellapneumoniae*) d'analogues de l'acide nalidixique..

Substituant (Z)	Activité bactéricide <i>Klebsiellapneumoniae</i>	Substituant (Z)	Activité bactéricide <i>Klebsiellapneumoniae</i>
H	-	C_6H_{13}	++
CH_3	++	$\text{CH}_2\text{-CH=CH}_2$	+++
C_2H_5	++++	$\text{CH}_2\text{-C}\equiv\text{CH}_3$	+
C_3H_7	++++	$\text{CH}_2\text{-CH}_2\text{OH}$	++
C_4H_9	+++	$\text{CH}_2\text{-CH}_2\text{-O-C}_2\text{H}_5$	+

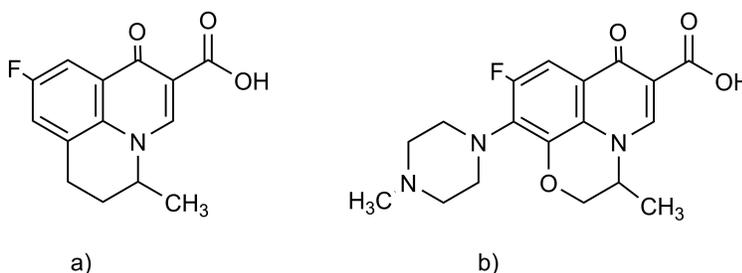


FIGURE 1.19 – a) Structure de fluméquine ; b) Structure d'ofloxacine

Enfin, le remplacement isostérique du groupement méthylénique (CH_2) par NH (Amifloxacine) ou O (Miloxacine) (Figure 1.20) assure aux quinolones obtenues une meilleure biodisponibilité [20][21].

Ce concept a également été appliqué au sommet CH (isostère de N) du cycle A, donnant naissance à la cinoxacine (Figure 1.21) [22].

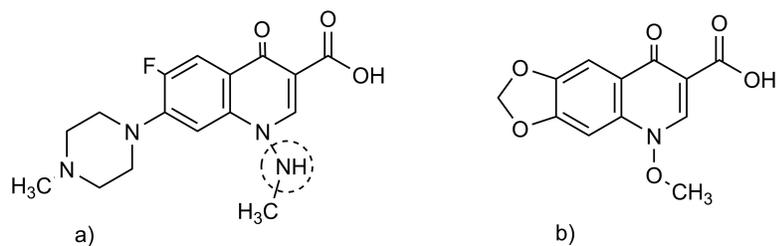


FIGURE 1.20 – a) Structure d’Amifloxacin ; b) de Miloxacin.

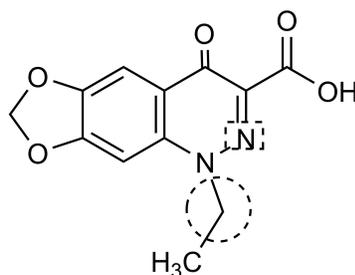


FIGURE 1.21 – La structure de Cinoxacin.

1.4.4.1.2 C₂-substitution

Toute substitution à ce niveau engendre des composés inactifs ; seul le remplacement isostérique de CH (acide oxolinique) par N exalte l’activité bactéricide du modèle et conduit à la cinoxacin particulièrement appréciée pour le traitement des pyélonéphrites.

1.4.4.1.3 Modification du 3-carboxyle

Ce groupement fonctionnel a été soumis aux pharmacomodulations suivantes :

- **Soit la transformation en fonctions** : éther, nitrile, alcool, aldéhyde. Toutes ces modifications diminuent ou annulent le pouvoir antibactérien du modèle. Toutefois l’estérification, qui est un moyen classique de création de formes dites retard, a été mise à profit dans le métioxate[23].
- **Soit la substitution par un radical** : 1-méthyl-1H-tétrazol-5-yl, d’acidité considérée comme comparable au carboxyle (même utilisation en série céphalosporine).

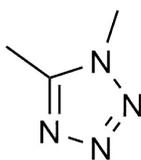


FIGURE 1.22 – Structure de 1-méthyl-1H-tétrazol-5-yl.

Cet hétérocycle (Figure 1.22) annule toute activité antibactérienne des quinolones sur lesquelles il est greffé. Compte tenu, que l'on observe les mêmes effets en remplaçant le carboxyle par un groupement fonctionnel sulfonique ($-SO_3H$), il apparaît que seul le groupement COOH peut participer à un mécanisme d'action de type complexation vis-à-vis de certains métaux divalents (Cu^{2+} , Co^{2+} , Ni^{2+} , Fe^{2+})[24].

1.4.4.2 Pharmacomodulation affectant le cycle B

Ces noyaux aromatiques, de nature benzénique (rosoxacine, l'acide oxolinique, norfloxacine, péfloxacin), pyridinique (l'acide nalidixique) ou pyrimidique (acide pipémidique, acide piromidique, fluméquine, ofloxacine) sont liés entre eux par des relations d'isostérie.

En ce qui concerne les quinolones de 2ème génération, la double substitution de ce cycle :

- **En position 6** : Par un atome de fluor, accroît considérablement le pouvoir bactéricide.
- **En position 7** : Par un noyau pipérazinique accroît considérablement le pouvoir bactéricide à l'égard de *Pseudomonas aeruginosa*[24].

1.5 Synthèse chimique

Quelle que soit la complexité du polycycle à créer, la stratégie de synthèse des quinolones consiste à construire progressivement le cycle A, au départ d'une arylamine primaire ou d'un chlorure d'O-chlorobenzoyl convenablement fonctionnalisé, préfigurant le cycle aromatique B[24]. La nature des réactifs mis en œuvre et le mode de création du cycle A permettent de distinguer principalement trois groupes de méthodes.

1.5.1 Méthode de GOULD et JACOB

Le principe de cette méthode, mise à profit par G. LESHER pour synthétiser l'acide nalidixique [25], consiste à condenser une arylamine (1) avec l'éthoxy-méthylène-malonate de diéthyle (EMME)(2) en aniline-méthylène-malonate de diéthyle (3) selon une réaction de Michaël (Figure 1.23). Ce diester éthylique, soumis à un chauffage prolongé vers 250°C, conduit par cyclisation intramoléculaire aux systèmes bicycliques (4), (5) et (6) qu'il convient d'abord d'alkyler en position 1 (7), puis d'hydrolyser au niveau de la fonction ester en acide carboxylique (8).

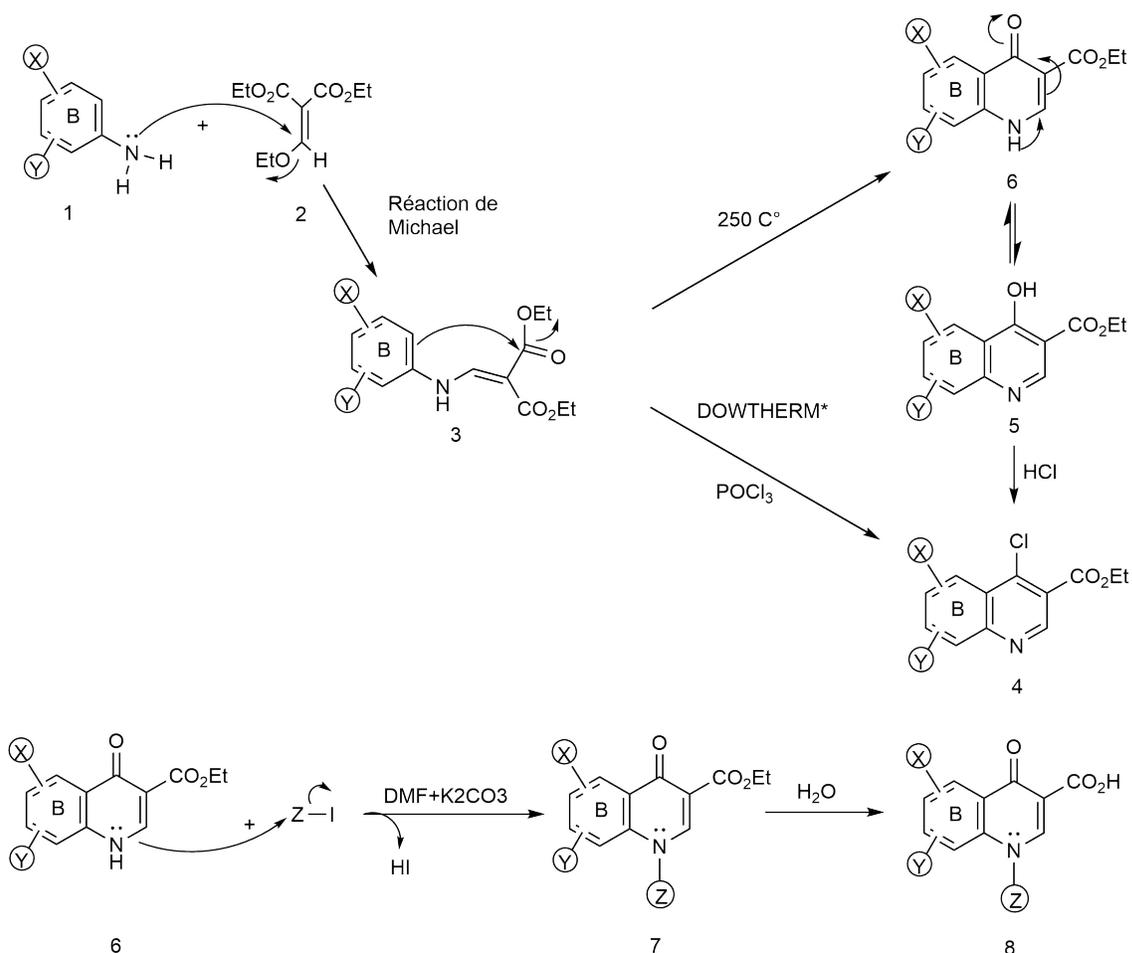


FIGURE 1.23 – Synthèse de l'acide nalidixique selon la méthode de GOULD et JACOBS.

La cyclisation de (3) est réalisée, soit au sein de solvant suffisamment stables à haute température DOWTHERM®), soit en présence d'agents de cyclisation (POCl₃, acide polyphosphorique : PPA ou son ester éthylique : PPE) catalysée par divers sels (ZnCl₂, NiSO₄). L'équilibre tautomère (5)⇌(6) permet d'expliquer l'obtention par alkylation de dérivés O-alkylés et N-alkylés (7); l'emploi de solvants aprotiques (DMF) et de bases (K₂CO₃) favorise la N-alkylation recherchée. Par ailleurs, la nature et la position des substituants X et Y sur l'intermédiaire réactionnel (3), oriente la fermeture du cycle. La voie (a) de cyclisation étant dans cette exemple très favorisée, conduit à la quinolone recherchée.

Les quinolones de 2^{ème} génération, substituées en position 7 par un groupement pipérazinique, sont obtenues en mettant à profit l'activation de la position 7 par l'ensemble 4-pyridone.

La synthèse de l'Ofloxacin(Figure 1.24) utilise les différents réactifs et principes précédemment évoqués :

— Construction du cycle supplémentaire (oxacine), au départ d'un synton préfigurant le

- cycle B ;
- Création du noyau A ;
- Substitution du système tricyclique par un reste pipérazinyle N-méthylé.

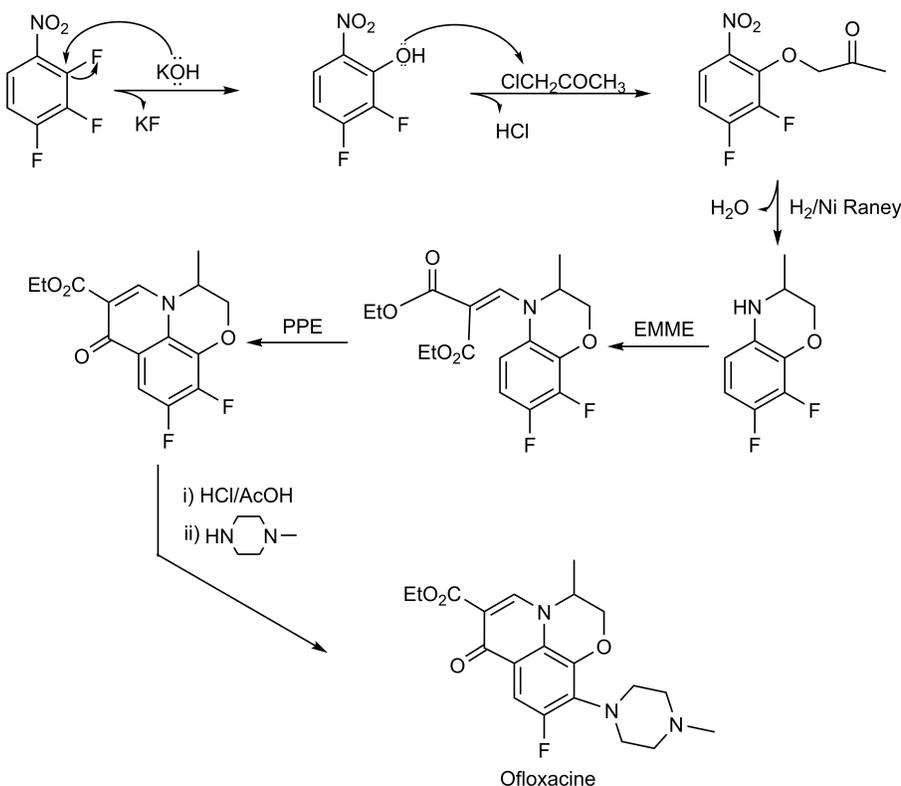


FIGURE 1.24 – Synthèse de la Ofloxacin.

1.5.2 Cyclisation de DIECKMANN

L'intermédiaire (3) obtenu au départ d'*o*-amino (1) ou *o*-chlorobenzoate d'éthyle (2), est cyclisé sous l'action d'alcoolates alcalins, la création d'une double liaison en C₂-C₃ (6) consiste d'abord à bromer le C₃ (en alpha des deux carbonyles) (5), puis à éliminer HBr. Une hydrolyse finale de la fonction ester conduit à la quinolone attendue. Cette méthode est surtout intéressante pour obtenir les quinolones dans lesquelles le cycle B est de nature pyridinique ou pyrimidinique : acide pipémidique[26].

- **Exemple** : X=Y=N dans l'acide pipéimidique

1.5.3 Cyclisation d'une enamine-ester par la méthode de GROHE

Ce troisième mode de formation du cycle A permet enfin d'obtenir les quinolones diversément substituées sur l'atome d'azote N₁(Figure 1.26)[27]. Ainsi la préparation de la Cipro-

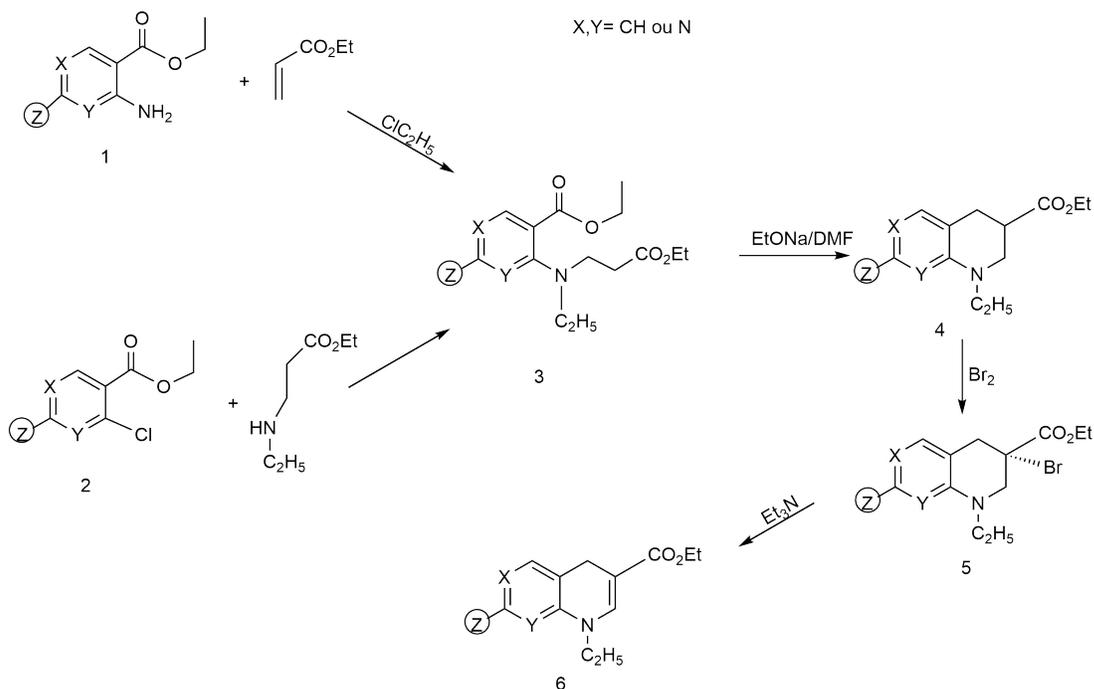


FIGURE 1.25 – Synthèse des quinolones selon la cyclisation de DIECKMANN.

floxacine, répond au schéma de synthèse de GROHE résumé comme suit :

La condensation du chlorure de benzoyle polyhalogéné (1) avec le malonate de diéthyle en présence d'éthylate de magnésium, conduit au benzoylmalonate (2), qui est alors partiellement hydrolysé et décarboxylé au moyen de l'acide p-toluènesulfonique en benzoylacétate d'éthyle (3). Le reflux de ce β -cétoester avec l'orthoformiate d'éthyle, au sein de l'anhydride acétique, donne naissance à l'intermédiaire(4) qui est alors traité par la cyclopropylamine. La création du cycle A est obtenue par l'hydru de sodium au sein du dioxane à reflux et conduit à la quinolone N1 substituée (6), qu'il suffit de condenser à chaud avec la pipérazine.

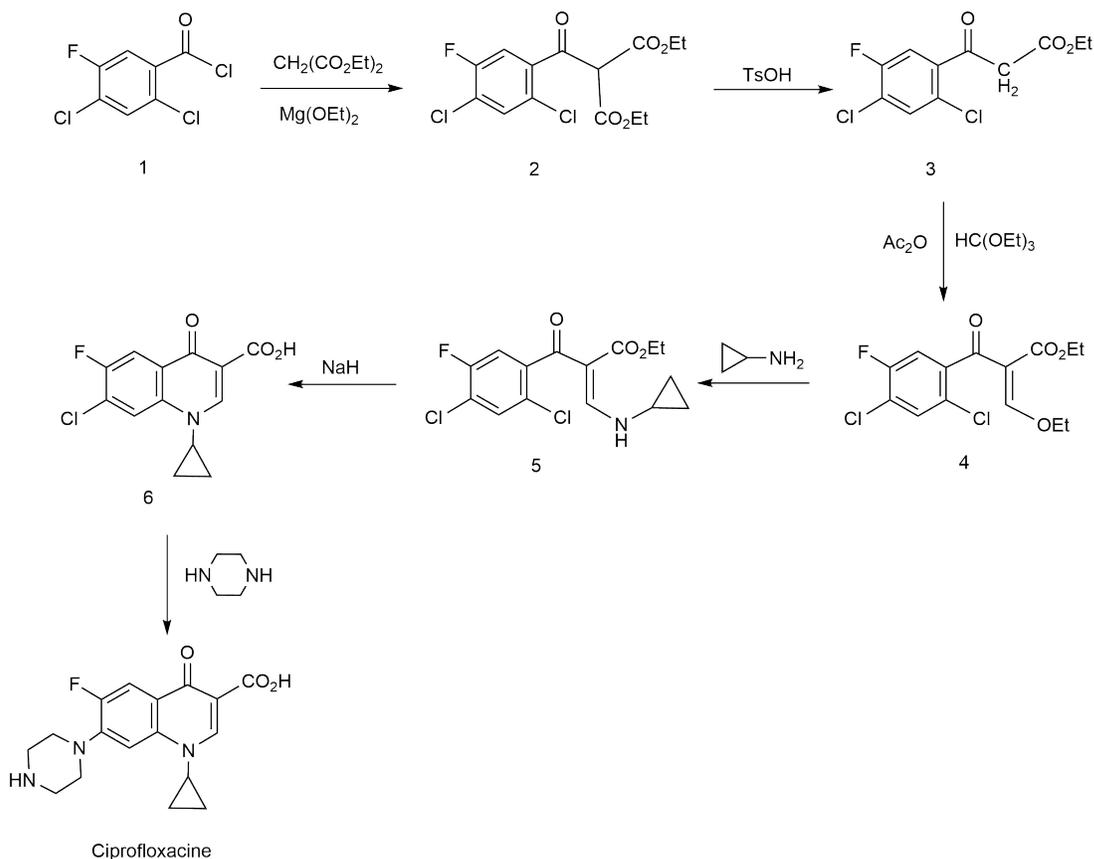


FIGURE 1.26 – Synthèse des quinolones selon la méthode de GROHE.

1.6 Propriétés physicochimiques

1.6.1 Les caractéristiques physico-chimiques de Lévofoxacine

Lévofoxacine est un antibiotique de la famille des fluoroquinolones, commercialisé en Europe par le laboratoire Sanofi-Aventis sous le nom Tavanic®. La Lévofoxacine est aussi un bactéricide très efficace sur les bactéries gram-positives et gram-négatives. Son action est d'empêcher la réplication bactérienne en bloquant les enzymes ADN gyrases nécessaires à l'ouverture de la double hélice d'ADN bactérienne, utilisé généralement pour le traitement des infections pulmonaires. Ce composé est stéréo-chimiquement stable c'est-à-dire, ne subit pas des inversions chirales ni dans le plasma ni dans l'urine. L'ensemble de ces propriétés physico-chimiques sont regroupés dans le Tableau 1.2[28].

TABLE 1.2 – Les caractéristiques physico-chimiques de Lévofoxacine.

Nom chimique	Chlorhydrate d'acide 1-cyclopropyl-6-fluoro-4-oxo-7-(piperazine-1-yl)-1,4-dihydroquinoline-3-carboxylique monohydraté
Formule brute	$C_{18}H_{20}FN_3O_4$
Masse molaire (g/mol)	361,36
Point de fusion (°C)	255-257
Solubilité	25,3 g.L ⁻¹ à 25 °C dans l'eau
pka	6,1 - 8,7
Description	Poudre de coloration jaune claire
Médicamentn	TAVINIC [®]

1.6.2 Les caractéristiques physico-chimiques de Ciprofloxacine

La Ciprofloxacine est un antibiotique de synthèse créé et commercialisé par le laboratoire Bayer sous le nom de Ciflox[®]. Elle appartient à la famille de la deuxième génération des fluoroquinolones. Elle neutralise la réplication des enzymes bactériennes et empêche toutes les multiplications cellulaires. Elle agit sur les ADN gyrases et particulièrement sur la topoisomérase IV. Ces propriétés physico-chimiques sont regroupées dans le Tableau 1.3[29].

TABLE 1.3 – Les caractéristiques physico-chimiques de Ciprofloxacine.

Nom chimique	acide 1-cyclopropyl-6-fluoro-4-oxo-7-piperazin-1-yl-quinoline-3-carboxylique
Formule brute	$C_{17}H_{18}FN_3O_3$
Masse molaire (g/mol)	331,34
Point de fusion (°C)	255-257
Solubilité	3.10 mg.L ⁻¹ à 20 °C dans l'eau
pka	6,09
Description	Poudre cristalline de coloration jaune pâle
Médicamentn	Ciflox [®] , Cirolon [®] et Cifloxin [®]

1.7 Mécanisme d'action des quinolones

Les quinolones sont des inhibiteurs de topo-isomérase de type II (l'ADN gyrase et de l'ADN topo-isomérase IV). La cible principale est la gyrase chez les bactéries à Gram négatif alors que l'action sur les germes à Gram positif est plus particulièrement liée à la topoisomérase IV (Figure 1.27).

1.7.1 Inhibition d'ADN gyrase

Les quinolones exercent une inhibition sélective de la synthèse de l'ADN bactérien, et on sait depuis longtemps qu'une topoisomérase dénommée ADN gyrase constitue une cible intracellulaire des quinolones. Les topo-isomérasas sont des enzymes capables de modifier la topologie de l'ADN, c'est-à-dire le degré de torsion de la double hélice : celle-ci peut en effet se trouver dans un état relâché ou bien surenroulé positivement ou négativement (c'est-à-dire, en fait, sous-enroulé) et la gyrase est la seule topo-isomérase bactérienne capable de surenrouler négativement l'ADN. Elle possède une structure tétramérique, avec deux sous-unités A (GyrA) et deux sous-unités B (GyrB), codées respectivement par les gènes *gyrA* et *gyrB*. L'activité enzymatique de la gyrase est ATP-dépendante et peut être décomposée en trois étapes :

1. Coupure double-brin de l'ADN ;
2. Passage d'un autre segment d'ADN par la coupure ;
3. Refermeture de la coupure.

Au moment de la formation de celle-ci, l'ADN et la gyrase sont liés de manière covalente. Les quinolones sont capables de se fixer sur ce complexe covalent ADN-enzyme, le rendant irréversible. La nature précise des interactions moléculaires existant au sein du complexe ternaire ADN-gyrase-quinolone n'est cependant pas bien connue. Le blocage de l'enzyme sur l'ADN empêcherait la progression de l'ADN polymérase bactérienne au cours de la réplication, d'où l'inhibition des synthèses d'ADN et donc de la croissance bactérienne. Mais les quinolones sont des antibiotiques bactéricides, pas seulement bactériostatiques, et d'autres mécanismes doivent entrer en jeu. On pense que les coupures d'ADN double-brin stabilisées par les quinolones constituent des lésions non réparables de l'ADN, capables de déclencher l'activation de certaines synthèses protéiques, responsables de la bactéricidie[30].

1.7.2 Inhibition des ADN topo-isomérasas IV

Comme la gyrase, la topoisomérase IV peut passer un double brin d'ADN par un autre double brin d'ADN, modifiant ainsi le nombre de liaisons d'ADN par deux dans chaque étape enzymatique. Cette enzyme a été découverte en 1990 chez *E. coli*[31] et a été retrouvée par la suite chez d'autres espèces bactériennes. Elle a comme la gyrase, une structure tétramérique C2E2 : deux gènes, *parC* et *parE*. La topoisomérase IV semble avoir un rôle spécifique, le désenchevêtrement des ADN fils en fin de réplication. Les quinolones sont capables d'inhiber *in vitro* et *in vivo* l'activité de cette enzyme.

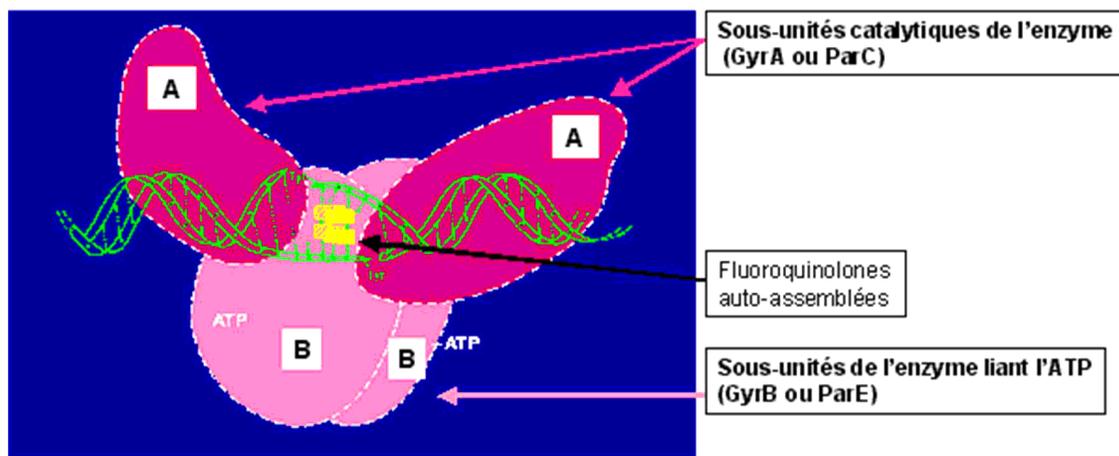


FIGURE 1.27 – Mécanisme d'action des quinolones.

1.8 Résistance Bactérienne aux quinolones

La résistance bactérienne aux agents antimicrobiens est un problème d'importance croissante en pratique médicale[32]. La dissémination des bactéries résistantes est à l'origine d'une augmentation considérable de la mortalité, de la morbidité ainsi que du coût des traitements. La résistance aux antibiotiques peut être soit une résistance naturelle ou innée (liée au spectre clinique d'un antibiotique) soit une résistance acquise. La résistance acquise est un problème de santé publique mondiale[33]. Les bactéries font de plus en plus de la résistance, s'adaptent aux thérapeutiques antibactériennes et sont responsables d'échecs de traitement.

1.8.1 Définition de la résistance

Une souche résistante :

- La concentration d'antibiotique qu'elle est capable de supporter est nettement plus élevée que la concentration qu'on peut atteindre in vivo ;
- Supporte une concentration d'antibiotique nettement plus élevée que celle inhibe le développement de la majorité des autres souches de la même espèce ;
- Pour une souche résistante, il existe une probabilité d'échec thérapeutique quel que soient les traitements et la dose d'antibiotique.

1.8.2 Mécanisme de la résistance bactérienne

1.8.2.1 Modifications des topoisomérases

Des mutations siégeant en général au niveau des gènes de la gyrase (gyrA ou plus rarement gyrB) et la topoisomérase IV(Figure 1.28), entraînent une élévation des CMI qui concerne, à

des degrés divers, l'ensemble des quinolones. La fréquence de ces mutations est assez élevée, de sorte que la sélection de mutants résistants au cours d'un traitement par les quinolones est un phénomène courant. Différentes mutations peuvent survenir successivement au niveau des gènes des topoisomérases et entraîner une augmentation de résistance par paliers.

Chez les bactéries à Gram négatif, l'ADN gyrase est plus sensible à l'inhibition médiée par les quinolones et les mutations responsables de résistance surviendront d'abord au niveau de *gyrA*, alors que chez les germes Gram positifs, la topoisomérase IV représente habituellement la cible primaire de l'antibiotique, et les mutations responsables de résistance se produiront d'abord au niveau de *parC*. En fait, les mutations impliquent des substitutions en acides aminés qui apparaissent dans une région particulière, nommée QRDR (quinolone resistance determining region).

1.8.2.2 Diminution de la concentration intracellulaire

Il existe chez les bactéries des systèmes permettant d'excréter certains antibiotiques. Ces systèmes jouent un rôle dans la résistance naturelle. Sous l'effet de mutations, leur niveau d'expression peut augmenter et faire apparaître une résistance acquise pouvant toucher simultanément plusieurs familles d'antibiotiques (par exemple fluoroquinolones et β -lactamines). Le phénomène a été décrit surtout chez les bactéries à Gram négatif. La résistance aux tétracyclines est due le plus souvent à l'acquisition d'un gène responsable d'un mécanisme d'efflux. La résistance aux macrolides peut être due à un mécanisme d'efflux[33].

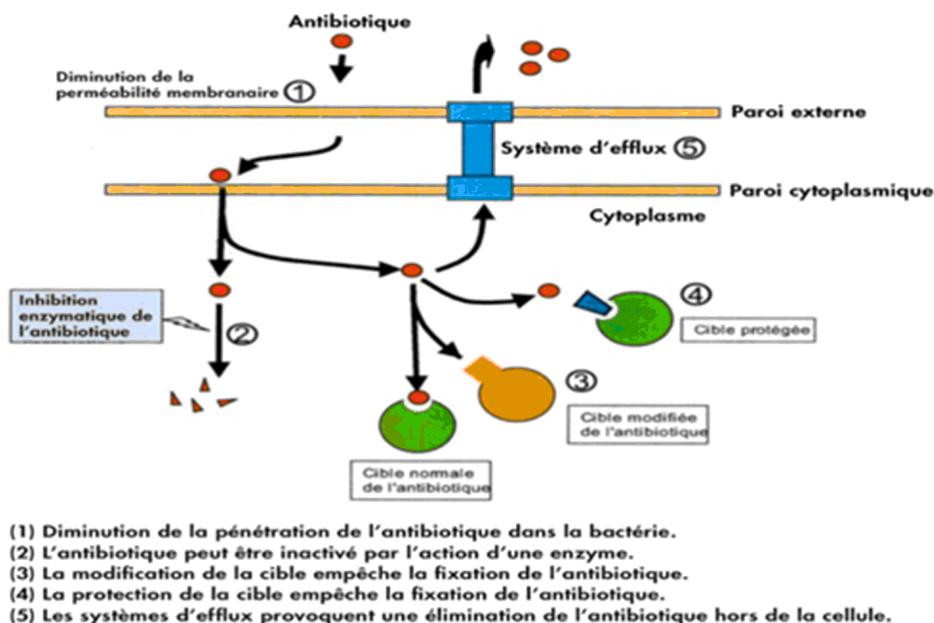


FIGURE 1.28 – Mécanisme de résistance bactérienne aux quinolones.

1.9 Conclusion

On a montré dans ce chapitre que les quinolones sont une classe importante et largement utilisée d'agents antimicrobiens en médecine clinique. La résistance s'est toutefois répandue dans un certain nombre de pathogènes humains, en partie en raison de l'utilisation de quinolones chez l'homme. La chimie computationnelle est venue dans le but d'obtenir une meilleure compréhension des paramètres contrôlant cette résistance, par l'identification des nouveaux chefs de file capables de résoudre ce type de problème.

Bibliographie

- [1] Michelle J Cole, Stephanie A Chisholm, Steen Hoffmann, Angelika Stary, Catherine M Lowndes, Catherine A Ison, European Surveillance of Sexually Transmitted Infections Network, et al. European surveillance of antimicrobial resistance in neisseria gonorrhoeae. *Sexually transmitted infections*, 86(6) :427–432, 2010.
- [2] John M Domagala and Susan E Hagen. Structure-activity relationships of the quinolone antibacterials in the new millennium : some things change and some do not. In *Quinolone Antimicrobial Agents, Third Edition*, pages 3–18. American Society of Microbiology, 2003.
- [3] M Grobbel, A Lübke-Becker, LH Wieler, R Froyman, S Friederichs, and S Filios. Comparative quantification of the in vitro activity of veterinary fluoroquinolones. *Veterinary microbiology*, 124(1) :73–81, 2007.
- [4] CK Dresser. Chloroquine dose for malaria prevention. *Canadian Family Physician*, 32 :231, 1986.
- [5] J Thomas Cross. Fluoroquinolones. In *Seminars in Pediatric infectious diseases*, volume 12, pages 211–223. Elsevier, 2001.
- [6] W Michael Scheld. Maintaining fluoroquinolone class efficacy : review of influencing factors. *Emerging infectious diseases*, 9(1) :1, 2003.
- [7] P Turgeon, J Dubois, and S Stoffel. In vitro activity of norfloxacin against 4003 clinical isolates : a multicentre study in canada. *Current therapeutic research*, 38(1) :169–172, 1985.
- [8] HE Pichler, G Diridl, K Stickler, and D Wolf. Clinical efficacy of ciprofloxacin compared with placebo in bacterial diarrhea. *The American journal of medicine*, 82(4A) :329–332, 1987.
- [9] Joichi Kumazawa and Morimasa Yagisawa. The history of antibiotics : the japanese story. *Journal of infection and chemotherapy*, 8(2) :125–133, 2002.
- [10] World Health Organization et al. Who model list of essential medicines : 18th list, april 2013. 2013.
- [11] A Bauernfeind. Comparison of the antibacterial activities of the quinolones bay 12-8039, gatifloxacin (am 1155), trovafloxacin, clinafloxacin, levofloxacin and ciprofloxacin. *The Journal of antimicrobial chemotherapy*, 40(5) :639–651, 1997.

- [12] Ann Allen, Elizabeth Bygate, Stuart Oliver, Martin Johnson, Christopher Ward, Ae-Jin Cheon, Youn Sung Choo, and In-Chull Kim. Pharmacokinetics and tolerability of gemifloxacin (sb-265805) after administration of single oral doses to healthy volunteers. *Antimicrobial agents and chemotherapy*, 44(6) :1604–1608, 2000.
- [13] PC Appelbaum and PA Hunter. The fluoroquinolone antibacterials : past, present and future perspectives. *International journal of antimicrobial agents*, 16(1) :5–15, 2000.
- [14] N Saravana Kumar, D Dhivya, and B Vijayakumar. A focus on quinolones and its medicinal importance. *Inter. J Novel Trends Pharm. Sci*, 1(1) :23–29, 2011.
- [15] GS Tillotson. Quinolones : structure-activity relationships and future predictions. *Journal of medical microbiology*, 44(5) :320–324, 1996.
- [16] Alan R Katritzky, Mati Karelson, and Philip A Harris. Prototropic tautomerism of heteroaromatic compounds. *ChemInform*, 22(32), 1991.
- [17] AR Katritzky and JM Lagowski. Prototropic tautomerism of heteroaromatic compounds : Ii. six-membered rings. *Advances in Heterocyclic Chemistry*, 1 :339–437, 1963.
- [18] DTW Chu, IM Lico, AK CLAIRBORNE, JJ Plattner, and AG Pernet. Structure-activity relationship of quinolone antibacterial agents : the effects of c-2 substitution. *Drugs under experimental and clinical research*, 16(5) :215–224, 1990.
- [19] MASAZUMI Imamura, SEIICHI Shibamura, ISAO Hayakawa, and YASUAKI Osada. Inhibition of dna gyrase by optically active ofloxacin. *Antimicrobial agents and chemotherapy*, 31(2) :325–327, 1987.
- [20] Harold C Neu and Pornpen Labthavikul. Antibacterial activity of amifloxacin (win 49, 375), a new quinolone agent. *Diagnostic microbiology and infectious disease*, 3(6) :469–478, 1985.
- [21] Robert M Lumish and Carl W Norden. Cinoxacin : in vitro antibacterial studies of a new synthetic organic acid. *Antimicrobial agents and chemotherapy*, 7(2) :159–163, 1975.
- [22] Akira Ito, Keiji Hirai, Matsuhisa Inoue, Hiroshi Koga, Seigo Suzue, Tsutomu Irikura, and Susumu Mitsuhashi. In vitro antibacterial activity of am-715, a new nalidixic acid analog. *Antimicrobial Agents and Chemotherapy*, 17(2) :103–108, 1980.
- [23] Armand Trousseau and Hermann Pidoux. *Traité de thérapeutique et de matière médicale*. Société typographique belge, 1843.
- [24] Jacques Bompard. *Synthèse de nouvelles quinolones : relations structure-activité*. PhD thesis, 1986.

- [25] Jie Jack Li. Gould–jacobs reaction. *Name Reactions*, pages 263–265, 2009.
- [26] Makoto Shimizu, Yuki Takao, Haruna Katsurayama, and Isao Mizota. Synthesis of indolin-3-ones and tetrahydro-4-quinolones from α -imino esters. *Asian Journal of Organic Chemistry*, 2(2) :130–134, 2013.
- [27] K Grohe. The chemistry of the quinolones : Methods of synthesizing the quinolone ring system. In *Quinolone antibacterials*, pages 13–62. Springer, 1998.
- [28] British Pharmacopoeia Commission. *British Pharmacopoeia 2007*. Number vol. 1 in British Pharmacopoeia (Incl.BP Veterinary) 6 Vols+online. Stationery Office, 2006.
- [29] Kathleen Parfitt and James EF Reynolds. *Martindale : The extra pharmacopoeia*.
- [30] Chang-Rung Chen, Muhammad Malik, Michael Snyder, and Karl Drlica. Dna gyrase and topoisomerase iv on the bacterial chromosome : quinolone-induced dna cleavage. *Journal of molecular biology*, 258(4) :627–637, 1996.
- [31] Jun-ichi Kato, Yukinobu Nishimura, Ryu Imamura, Hironori Niki, Sota Hiraga, and Hideho Suzuki. New topoisomerase essential for chromosome segregation in e. coli. *Cell*, 63(2) :393–404, 1990.
- [32] Christian G Giske, Dominique L Monnet, Otto Cars, Yehuda Carmeli, et al. Clinical and economic impact of common multidrug-resistant gram-negative bacilli. *Antimicrobial agents and chemotherapy*, 52(3) :813–821, 2008.
- [33] David C Hooper and John S Wolfson. *Quinolone antimicrobial agents*. american Society for Microbiology, 1993.

Chapitre 2

Aspect théorique sur la méthodologie de QSAR et les descripteurs moléculaires

Sommaire

2.1 Introduction	33
2.2 Présentation de la méthode de QSAR	35
2.2.1 Introduction du Data Set	35
2.2.2 Calcul et optimisation	35
2.2.3 Développement du modèle	41
2.2.4 Validation du modèle	45
2.3 Conclusion	49

2.1 Introduction

La connaissance des propriétés et des activités est d'une importance capitale pour pouvoir classer et utiliser les composés chimiques. La caractérisation expérimentale complète est difficile, voire impossible, pour des raisons de temps, de coût, de dangerosité de certains essais ou d'éthique (limitations des essais sur les animaux) pour cela, l'utilisation des méthodes alternatives à l'expérience est devenue plus qu'indispensable[1] (Figure 2.1).

Parmi ces méthodes de modélisation les plus utilisées, on peut citer la méthode QSAR (Quantitative Structure-Activity Relationships), cette méthode s'appuie sur le principe que l'activité biologique des molécules qui dépend fortement de leurs structures chimiques (propriétés moléculaires). L'étude de la relation existante entre ces propriétés et la structure moléculaire des composés chimiques constitue le parcours idéal dans la recherche d'un modèle capable de prédire ces propriétés[2].

Du fait que les propriétés sont exprimées en nombres, alors que la structure moléculaire ne peut pas l'être, le codage de la structure moléculaire présente le premier défi pour modéliser la

relation structure-propriétés. En conséquence, l'approche la plus pratique pour résoudre cette difficulté est l'utilisation des descripteurs moléculaires.

L'outil QSAR est devenu avantageux et précieux, notamment dans la recherche pharmaceu-

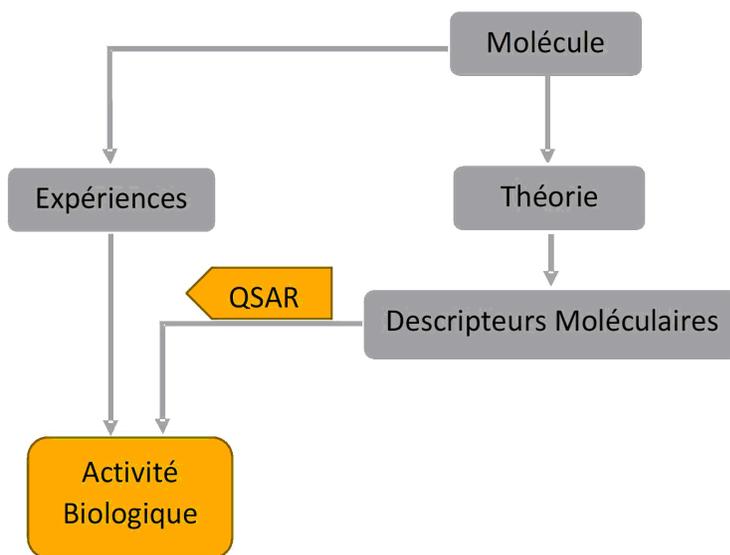


FIGURE 2.1 – Rôle des Descripteurs moléculaires en relation quantitative structure activité.

tique industrielle, plus particulièrement dans le cas où la disponibilité des échantillons est limitée et/ou les mesures expérimentales sont excessives, longues et chères. Les prévisions fournies par QSAR permettent de proposer une bonne stratégie pour le développement d'un bon pharmacophore, elles peuvent être ainsi un support adéquat pour valider ou rejeter une hypothèse[3][4]. Sans l'utilisation de grands instruments analytiques, QSAR peut également fournir des informations utiles sur les caractéristiques structurales d'un composé responsable d'une activité biologique. La dérivation de la relation directe avec la structure moléculaire n'est pas facile. Il est cependant possible d'identifier plusieurs facteurs structurels connus comme descripteurs moléculaires qui influent sur la propriété moléculaire choisie.

Comme il est bien illustré dans la Figure 2.2 présentant le problème de QSAR, proposé par le groupe des auteurs[5][6], l'activité ne peut pas être obtenue directement à partir de la structure moléculaire. Cette dernière doit d'abord être codée en tant que descripteurs numériques traduisant les propriétés physico-chimiques influant sur l'activité visée, tels que : la taille, la forme, le nombre de radicaux, la polarité et la capacité d'établir des liaisons d'hydrogène.

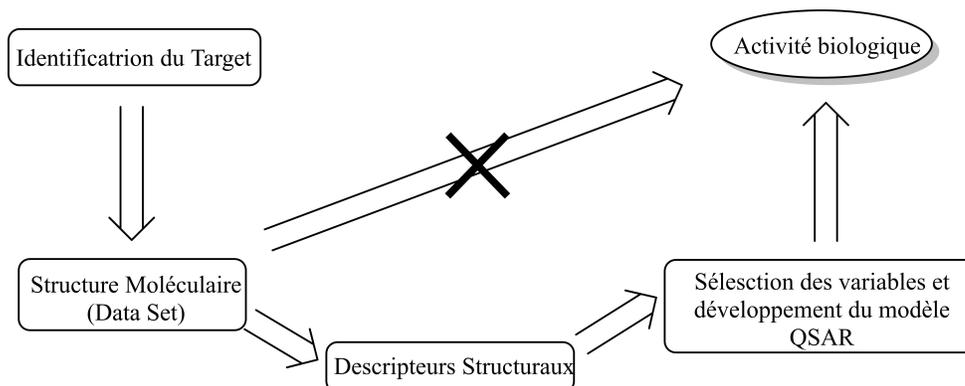


FIGURE 2.2 – Présentation des étapes aboutissant à QSAR.

2.2 Présentation de la méthode de QSAR

En général, pour prédire l'activité biologique par la méthode QSAR, la construction du modèle mathématique est effectuée en 4 étapes de base (Figure 2.3). L'ensemble des données (l'ensemble des molécules) utilisées dans l'étude de QSAR sont collectées soit directement à partir des mesures expérimentales, extraites de la littérature ou à partir de bases de données. Les données collectées sont divisées en premier lieu aléatoirement en deux parties :

- Les données d'apprentissage «Training Set (TSET)» : utilisées pour la construction des modèles finaux ;
- Les données de prédiction «Prédiction Set (PSET)» : utilisées pour la validation des modèles formés. Ces données ne subissent en revanche aucun traitement durant la construction des modèles et elles sont réservées uniquement pour tester la fiabilité des modèles.

2.2.1 Introduction du Data Set

La représentation des structures consiste en la représentation structurale de tous les composés, existant dans l'ensemble des données, sous forme de croquis bidimensionnel. Plusieurs programmes spécifiques peuvent être employés pour le dessin des structures moléculaires des composés organiques on cite à titre d'exemple : ChemBioDraw ultra 16.0 ou Hyperchem[7][8].

2.2.2 Calcul et optimisation

2.2.2.1 Optimisation géométrique des molécules du Data set

Avant de calculer numériquement les descripteurs structuraux pour le développement de notre modèle QSAR, il faut être sûr que la géométrie des molécules correspond à un minimum d'énergie potentielle.

La structure 2D, des 85 molécules de dérivées de quinolones ont été exportés vers Gaussview

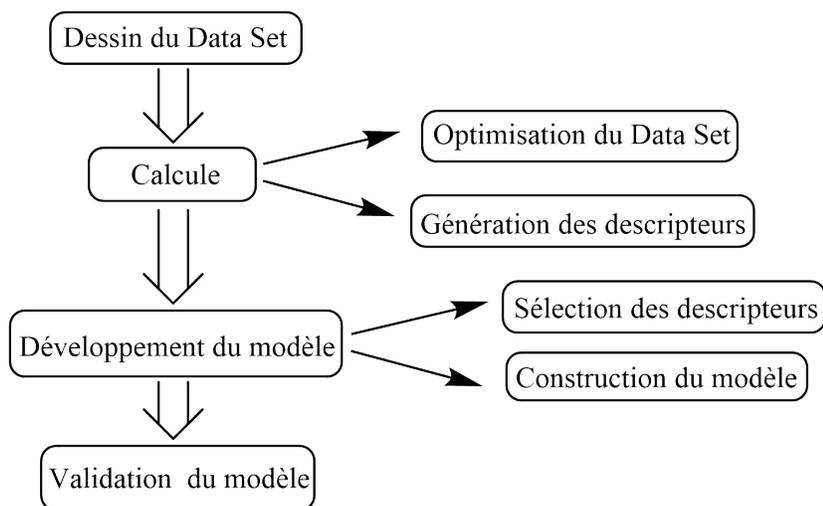


FIGURE 2.3 – Les principales étapes de la construction du modèle de QSAR.

5.0.9 software. Les structures tridimensionnelles de toutes les molécules ont été générées et leurs géométries ont été optimisées par prélèvement avec la méthode semi-empirique AM1[9], puis en utilisant la théorie de densité fonctionnel, DFT (La fonctionnelle de la densité)[10].

2.2.2.2 Génération des descripteurs

Les descripteurs sont des valeurs numériques qui encodent les caractéristiques physico-chimiques des molécules à partir de leurs représentations structurales. La génération des descripteurs constitue l'étape cruciale lors de l'analyse de QSAR et la qualité d'un modèle dépendra essentiellement de la nature des descripteurs générés. La plupart des descripteurs, actuellement plus de 4000, peuvent être calculées théoriquement à partir de différents logiciels tels que : Flap[11]; Phase[12]; AutoQSAR[13] et DRAGON[14] etc.

2.2.2.2.1 Aspect théorique sur les descripteurs moléculaire

"The molecular descriptor is the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment." [15]

"Le descripteur moléculaire est le résultat final d'une procédure logique et mathématique qui transforme des informations chimiques codées dans une représentation symbolique d'une molécule en un nombre utile ou le résultat d'une expérience standardisée."

Les descripteurs moléculaires jouent un rôle fondamental en chimie, en sciences pharmaceutiques, en politique de protection de l'environnement, en sciences de l'alimentation, en re-

cherche en santé et en contrôle qualité, obtenu lorsque les molécules sont transformées en une représentation moléculaire permettant un traitement mathématique. De nombreux descripteurs moléculaires ont été proposés à partir de théories et d'approches différentes dans le but de prédiction des propriétés biologiques et physicochimiques des molécules.[15]

Le contenu informationnel d'un descripteur moléculaire dépend du type de la représentation moléculaire utilisée et de l'algorithme défini pour son calcul (Figure 2.4). Il existe de simples descripteurs moléculaires dérivés par comptage de certains types d'atomes ou fragments structuraux dans la molécule, d'autres dérivés d'algorithmes appliqués à une représentation topologique (graphique moléculaire) et habituellement appelés descripteurs topologiques ou 2D. Il existe également des descripteurs moléculaires dérivés d'une représentation géométrique appelés descripteurs géométriques ou 3D.

Les descripteurs moléculaires sont des nombres capables d'extraire de petits morceaux d'in-

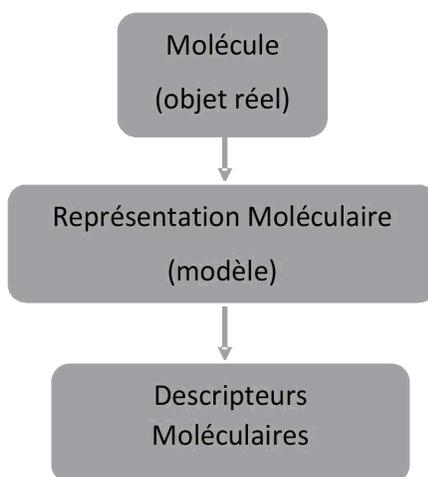


FIGURE 2.4 – L'importance des descripteurs moléculaire pour l'information structurelle.

formations chimiques provenant des différentes représentations moléculaires. Tous les descripteurs moléculaires doivent contenir, à des degrés divers, des informations chimiques, doivent satisfaire certaines propriétés invariantes de base et des exigences générales, et doivent être dérivés de procédures bien établies qui permettent de calculer les descripteurs moléculaires pour n'importe quel ensemble de molécules. Il est évident qu'un seul descripteur ou un petit nombre de descripteurs ne peut pas représenter entièrement la complexité moléculaire ou modéliser toutes les réponses physico-chimiques et les interactions biologiques. En conséquence, bien que nous devions nous habituer à vivre avec des modèles approximatifs, nous devons garder à l'esprit que "approximatifs" n'est pas un synonyme de "inutile".

Le domaine des descripteurs moléculaires est fortement interdisciplinaire et implique une masse de théories différentes. Pour la définition des descripteurs moléculaires, une connaissance d'algèbre, théorie des graphes, théorie de l'information, chimie computationnelle, théo-

ries de la réactivité organique et de la chimie physique est habituellement requise, bien qu'à des niveaux différents. Pour l'utilisation des descripteurs moléculaires, une connaissance des statistiques, la chimiométrie et les principes des approches QSAR/QSPR est nécessaire outre la connaissance spécifique du problème (Figure 2.5).

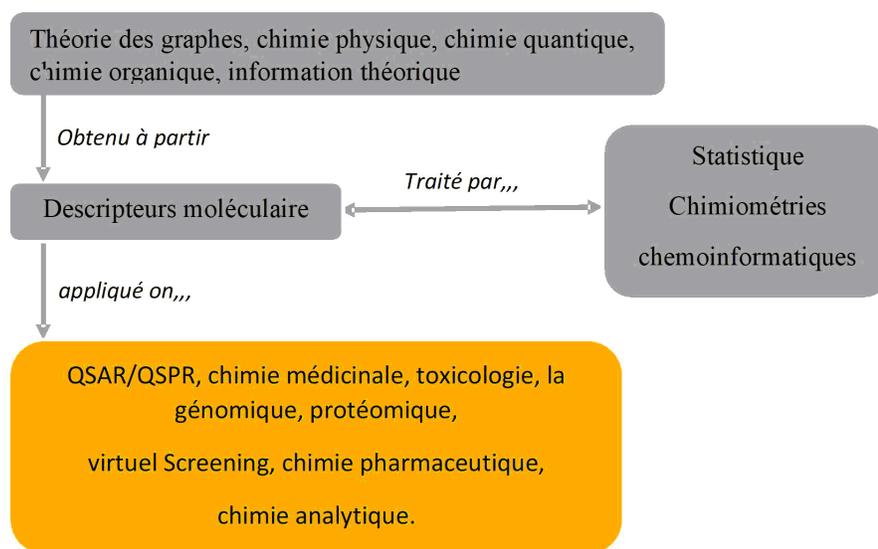


FIGURE 2.5 – Rôle clé des descripteurs moléculaires dans la recherche scientifique.

2.2.2.2 Classification des descripteurs moléculaires

Il existe à ce jour plus de 5000 descripteurs répertoriés, qui sont classés en 4 types :

- **Les descripteurs 0-D** : Les descripteurs moléculaires obtenus à partir de la formule chimique peuvent être considérés comme étant de dimension zéro (0D) parce qu'aucune information concernant les liaisons n'est prise en compte.
- **Les descripteurs 1-D** : La présence/l'absence ou le comptage de différents groupes fonctionnels ou sous-structures peuvent être définis comme des descripteurs unidimensionnels (1D) car la connectivité entre atomes est exprimée dans chaque sous structure.
- **Les descripteurs 2-D** : A partir d'un graphe moléculaire, il est possible de calculer des descripteurs qui prennent en compte la connectivité entre atomes, tels que les indices topologiques et les indices de connectivité, les paires d'atomes à une certaine distance topologique et d'autres. Par conséquent, cette représentation moléculaire et les descripteurs calculés à partir de celle-ci sont considérés comme bidimensionnels (2D).
- **Les descripteurs 3-D** : La représentation géométrique (3D) d'une molécule permet la dérivation de descripteurs qui codent des informations sur la distribution spatiale des atomes. Ces descripteurs sont communément appelés descripteurs 3D ou géométriques.

Comme une représentation géométrique implique la connaissance des positions relatives des atomes dans l'espace tridimensionnel, c'est-à-dire les coordonnées atomiques (x, y, z), les descripteurs géométriques fournissent généralement une plus grande quantité d'informations que les descripteurs 2D. La spécification supplémentaire de la structure électronique permet le calcul de descripteurs chimiques quantiques. Malgré leur contenu informatif élevé, les descripteurs géométriques présentent également certains inconvénients. Tout d'abord, ils nécessitent une géométrie moléculaire à définir (typiquement, la conformation la plus stable). Deuxièmement, certains descripteurs utilisés dans les analyses QSAR basées sur la grille nécessitent des règles d'alignement, nécessaires pour obtenir des résultats comparables. De plus, les géométries obtenues au moyen de différentes méthodes ou logiciels peuvent différer.

Par conséquent, un modèle QSAR basé sur des descripteurs 3D doit également implémenter l'algorithme d'optimisation de la géométrie utilisé, pour la préparation de l'ensemble de données, ou obliger les utilisateurs à suivre les mêmes étapes.

- **Les descripteurs 4-D** : Enfin, les descripteurs 4D prennent en compte l'aspect de flexibilité de la représentation structurelle 3D. Cette classe de descripteurs comprend également les représentations stéréo-électroniques caractérisant les interactions électroniques d'une molécule avec son environnement. Ce concept est à la base des techniques QSAR basées sur la grille telles que l'analyse comparative des champs moléculaires CoMFA (Comparative Molecular Field Analysis)[16][17].

Quatre propriétés ont été reconnues comme des exigences de base que les descripteurs moléculaires doivent posséder[18] :

- Invariance à la numération atomique ;
- Invariance à la rotation et à la translation de la molécule ;
- Une définition claire, calculable au moyen d'algorithmes ;
- Valeurs dans un intervalle numérique approprié.

Des caractéristiques supplémentaires que de bons descripteurs moléculaires devraient posséder ont été discutées par Randić[19].

Actuellement, des milliers de descripteurs moléculaires ont été définis et peuvent être calculés au moyen d'un logiciel dédié. Dans cette étude, différents types de descripteurs moléculaires ont été étudiés pour l'étalonnage des modèles QSAR. Compte tenu des enjeux liés à la description géométrique de la structure moléculaire (temps nécessaire à l'optimisation, cohérence des résultats d'optimisation, conforme utilisé pour la modélisation), des descripteurs zéro, unidimensionnel, bidimensionnel et tridimensionnel ont été utilisés.

2.2.2.2.3 Descripteurs générés par Dragon software

Grâce à son grand nombre de descripteurs, le logiciel DRAGON est l'un des outils les plus utilisés pour le calcul des descripteurs moléculaires. C'était l'outil principal de calcul des descripteurs moléculaires utilisés dans ce travail. Il calcule près de 5000 descripteurs moléculaires.

Pour faciliter la méthode de calcul aux utilisateurs, les descripteurs sont classés en 29 blocs logiques (Tableau 2.1) de groupes connus tels que les indices constitutionnels, les indices topologiques, les descripteurs géométriques, les paires d'atomes 2D et 3D, les groupes fonctionnels et les états atomiques de type E. En outre, le calcul de plusieurs propriétés moléculaires importantes telles que logP, les surfaces polaires topologiques, les surfaces de Van der Waals ainsi que certains indices de type-médicament tels que la règle de Lipinski de 5 (rule-of-five) est également fourni[20].

Ce logiciel supporte tous les formats moléculaires couramment utilisés et effectue une vérification préliminaire des structures, c'est-à-dire que les structures erronées et déconnectées sont habituellement rejetées. En plus du calcul des descripteurs moléculaires, ce logiciel permet d'effectuer une analyse préliminaire des descripteurs calculés avant la phase de modélisation. Les corrélations par paire, l'analyse en composantes principales (ACP), l'analyse graphique et l'importation de variables externes sont d'autres facilités fournies par DRAGON.

Nous décrivons dans la partie d'annexe ces différents blocs de descripteurs et nous nous limiterons surtout aux descripteurs les plus utilisés.

TABLE 2.1 – Liste des blocs de descripteurs fournis par le logiciel Dragon.

N°	Nom du bloc	N°	Nom du bloc
1	Constitutional descriptors	16	RDF descriptors
2	Ring descriptors	17	3D-MoRSE descriptors
3	Topological indices	18	WHIM descriptors
4	Walk and pathcounts	19	GETAWAY descriptors
5	Connectivity indices	20	Randic molecular profiles
6	Information indices	21	Functional group counts
7	2D matrix-based descriptors	22	Atom-centred fragments
8	2D autocorrelations	23	Atom-type E-state indices
9	Burdeneigenvalues	24	CATS 2D
10	<i>P_VSA – likedescriptors</i>	25	2D Atom Pairs
11	ETA indices	26	3D Atom Pairs
12	Edgeadjacency indices	27	Charge descriptors
13	Geometrical descriptors	28	Molecular properties
14	3D matrix-based descriptors	29	Drug-like indices
15	3D autocorrelations		

2.2.3 Développement du modèle

2.2.3.1 Sélection des descripteurs

La recherche de l'ensemble des descripteurs qui forment le bon modèle, exprimant l'activité biologique avec un coût raisonnable de calcul, constitue l'étape déterminante, car le calcul de tous les modèles possibles n'est pas pratique eu égard au nombre élevé des descripteurs (plus de 4500 descripteurs). Afin d'éviter la formation de modèles dus à la chance un contrôle rigoureux est exigé sur la taille de l'ensemble des descripteurs. Ainsi, l'approche rationnelle de la sélection des variables permet d'éviter les redondances, de diminuer le coût calculatoire et de trouver les meilleurs sous-ensembles de descripteurs.

La procédure de sélection des variables peut être divisée en deux étapes[21] :

— **Sélection objective :**

Elle consiste à la sélection des variables en réduisant le nombre de descripteurs sans faire participer la variable dépendante (la réponse biologique) afin de diminuer les corrélations entre les descripteurs. La suppression d'un descripteur, ayant un pourcentage élevé de valeurs identiques pour l'ensemble des composés, aura lieu au début de la sélection. Par exemple, le nombre d'atomes d'halogène ne fournira aucune information distinctive utile si uniquement l'halogène est présent sur un seul composé sur 100 existant dans l'ensemble (TSET). Par conséquent, ce descripteur est éliminé de l'ensemble total des descripteurs. Il en est de même pour les descripteurs fournissant des informations superflues. Ensuite, le coefficient de corrélation (R) entre les descripteurs est calculé par paires. Un des deux descripteurs est supprimé si leur combinaison possède un coefficient de détermination supérieur au seuil requis ($R > 0.90$). Cette valeur numérique de R est le seuil utilisé pour toutes les applications de sélection en réduisant le nombre de descripteurs sans pour autant perdre de l'information.

Il est parfois nécessaire, dans quelques cas, de réduire encore plus le nombre de descripteurs. A cet effet, l'analyse de l'espace vectoriel de descripteurs (VSDA) peut être utilisée. Dans cette procédure, un descripteur de l'ensemble est choisi au hasard comme descripteur (vecteur) de base initiale. Le procédé d'orthogonalisation de Gram Schmidt est appliqué pour la recherche du descripteur le plus orthogonal au descripteur de base de l'ensemble. Ensuite, ce descripteur est ajouté à l'ensemble de base (2D). Alors, le descripteur de l'ensemble le plus orthogonal au plan défini par les deux descripteurs de base est, à la fois, choisi et ajouté à l'ensemble de base. Ce procédé itératif est poursuivi jusqu'au rangement de la totalité des descripteurs selon ce principe d'orthogonalité. La formation de l'ensemble final, réduit, se fera en choisissant à partir de la base ordonnée, le sommet principal des descripteurs choisi par l'utilisateur. Le nombre de descripteurs choisis, x , sera déterminé par la taille de l'ensemble TSET.

— **Sélection subjective :**

La sélection subjective des variables de l'ensemble des descripteurs déjà réduit est alors employée dans la recherche de sous-ensembles de descripteurs optimaux en utilisant l'activité visée. A cet effet, diverses méthodes d'optimisations sont utilisées pour parcourir, d'une manière efficace, la dimension du modèle (le nombre optimal de descripteurs par modèle) sans pour autant examiner toutes les combinaisons possibles. Vu le nombre élevé de combinaisons possibles des sous-ensembles de descripteurs, des techniques statistiques et informatiques, sont employées pour explorer la dimension du modèle. Par la suite, on expose deux méthodes de sélection de descripteurs différents :

1. **Sélection par les Algorithmes génétiques** Les algorithmes génétiques (AG) sont des méthodes d'optimisation stochastiques inspirées des principes évolutionnaires[22]. L'aspect distinctif des algorithmes génétiques est qu'elles découvrent plusieurs solutions simultanément, dont chacune de ces solutions explore différentes régions de l'espace.

La première étape consiste en la création d'une population choisie au hasard de N chromosomes, dont chaque chromosome est une solution candidate pour ce problème (modèle). Dans notre cas, une représentation appropriée d'un chromosome peut être une chaîne de caractère numérique codant une combinaison particulière de certains descripteurs moléculaires. Le classement des chromosomes dans la population se fait suivant une fonction d'évaluation. Cette fonction est d'origine statistique tels que : le coefficient de détermination R^2 ; l'écart type s , le test de Fisher F , et le coefficient de corrélation issu du cross validation $Q_{Loo_CV}^2$; ce qui reflète la qualité de chaque solution candidate.

L'étape suivante est la reproduction, création de nouveaux chromosomes à partir de la génération existante. Selon l'opérateur de sélection, les meilleurs chromosomes peuvent proliférer préférentiellement. Aussi, chaque chromosome a l'occasion d'échanger de l'information avec les autres via les procédures de croisement et de mutation. Enfin, une affiliation plus convenable peut apparaître dans la prochaine génération si des mutations intéressantes ont lieu. La recherche collective, qui a commencé au début d'une manière aléatoire, commence à gagner plus de précision et se déplace vers des régions plus optimales, car le système recueille plus de connaissance au sujet de l'espace original.

2. **Sélection Ascendante pas à pas (Forward Stepwise)** Comme son nom l'indique, il s'agit d'une technique incrémentale qui consiste à repérer dans chaque étape la variable proposant une valeur absolue de t (Student) la plus élevée et de l'ajouter dans

l'ensemble (pool) courant si le coefficient est significatif, et de continuer ainsi tant que les ajouts sont possibles[23].

On commence par p régressions simples. Si une variable a été ajoutée, on poursuit avec $p-1$ régressions à 2 variables, etc. L'ajout d'une variable dépend de la significativité du coefficient de la variable choisie.

2.2.3.2 Construction du modèle

Le but d'une modélisation QSAR est de trouver une fonction permettant de relier de manière quantitative ou qualitative une propriété (ou activité) étudiée avec les descripteurs. Plusieurs méthodes sont utilisées pour relier les descripteurs aux propriétés parmi lesquelles il y a la régression linéaire multiple (MLR), les réseaux de neurones (CNN) et la régression par les machines à vecteur de support (SVM).

Concrètement, pour un ensemble de n molécules et p descripteurs, stockés dans la matrice \mathbf{X} , un modèle QSAR détermine une fonction qui relie \mathbf{X} au vecteurs de propriétés \mathbf{y} ($\mathbf{y} = (y_1, y_n)^T$).

$$\begin{pmatrix} y_1 \\ y_1 \\ \cdot \\ \cdot \\ y_n \end{pmatrix} = f \begin{pmatrix} X_{11} & X_{12} & \cdot & \cdot & X_{1p} \\ X_{21} & X_{22} & \cdot & \cdot & X_{2p} \\ X_{31} & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ X_{n1} & X_{n2} & \cdot & \cdot & X_{np} \end{pmatrix}$$

2.2.3.2.1 Régression Linéaire Multiple

La régression linéaire multiple (MLR) est une méthode reliant les descripteurs à la variable à modéliser y . Elle détermine les coefficients \mathbf{a} du vecteur \mathbf{A} et les erreurs \mathbf{b} du vecteur \mathbf{B} dans l'équation $\mathbf{Y} = \mathbf{XA} + \mathbf{b}$ pour une série de points définis par leurs coordonnées (x, y) . \mathbf{X} représente la matrice d'attributs, c'est-à-dire l'ensemble des descripteurs pour l'ensemble de données (molécules)[24]. La méthode MLR établit que la propriété y d'une molécule dépend linéairement des descripteurs x_1, x_2, \dots, x_p selon la relation :

$$y = a_0 + \sum_{i=1}^p a_i x_i \quad (2.1)$$

Pour un ensemble de données, cela revient à calculer un ensemble d'équations :

$$y_1 = a_0 + a_1 x_{11} + a_2 x_{12} + \dots a_p x_{1p} + b_1$$

$$y_2 = a_0 + a_1 x_{21} + a_2 x_{22} + \dots a_p x_{2p} + b_2$$

$$y_n = a_0 + a_1x_{n1} + a_2x_{n2} + \dots a_px_{np} + b_n \quad (2.2)$$

Au final, la méthode MLR résout l'équation :

$$Y = XA + b \quad (2.3)$$

Où Y, X, A et B représentent respectivement le vecteur de propriétés, la matrice des attributs (descripteurs), la matrice des coefficients et la matrice des erreurs de régression.

$$Y = \begin{pmatrix} y_1 \\ y_1 \\ \cdot \\ \cdot \\ y_n \end{pmatrix} \quad X = \begin{pmatrix} X_{11} & X_{12} & \cdot & \cdot & X_{1p} \\ X_{21} & X_{22} & \cdot & \cdot & X_{2p} \\ X_{31} & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ X_{n1} & X_{n2} & \cdot & \cdot & X_{np} \end{pmatrix} \quad A = \begin{pmatrix} a_1 \\ a_1 \\ \cdot \\ \cdot \\ a_n \end{pmatrix} \quad B = \begin{pmatrix} b_1 \\ b_1 \\ \cdot \\ \cdot \\ b_n \end{pmatrix}$$

La méthode des moindres carrés permet de simplifier (L'équation 2.3 en minimisant la somme des carrés des erreurs b. L'équation devient donc : $Y = XA$, la matrice X n'étant pas carrée, il faut multiplier chaque membre de l'équation par la transposée X^T , pour arriver à $X^T.X.A = X^T.Y$. Connaissant X et la propriété Y, on en déduit que :

$$A = (X^T.X)^{-1}X^T.Y$$

2.2.3.2.2 Réseau de neurones artificiels

L'approche par réseau de neurones (CNN : Computational Neural Network) est analogue aux systèmes de neurones biologiques. Les neurones biologiques permettent de véhiculer et de traiter des informations en faisant circuler des messages électroniques dans un réseau constitué d'axones. Les informations sont transmises d'un neurone à l'autre, de manière unidirectionnelle, par l'intermédiaire de points de jonction appelés synapses. En modélisation, un neurone possède une couche d'entrée par laquelle les données arrivent. Le neurone renvoie une valeur +1 ou -1 selon que la somme pondérée dépasse un certain seuil. Les poids de chaque neurone sont ajustés au cours de l'apprentissage[25].

Un réseau de neurones est constitué de multiples neurones, une couche d'entrée, une ou plusieurs couches cachées et une couche de sortie. Les neurones d'une couche sont interconnectés aux neurones de la couche voisine. La figure 3 représente un réseau de neurones à 3 couches, avec :

- **La couche d'entrée** : elle est constituée d'autant d'entrées que de descripteurs pour l'ensemble de données entier. Chaque neurone est connecté aux neurones de la couche cachée;

- **La couche cachée** : elle est constituée d'un nombre variable de neurones. Pour chaque neurone, le réseau effectue une opération de somme pondérée avec les différents poids de chaque neurone d'entrée. Un réseau de neurones peut avoir plusieurs couches cachées. Un poids est associé à chaque neurone de la couche cachée ;
- **La couche de sortie** : le nombre de neurones est égale au nombre de propriétés modélisées. Pendant la phase d'apprentissage du modèle par un réseau de neurones, les molécules sont présentées une par une aux neurones de la couche d'entrée. Le système change itérativement les poids associés aux neurones d'entrées et de la couche cachée de façon à minimiser l'erreur entre la propriété calculée et la propriété expérimentale. A cette fin, il ajuste les poids de chaque neurone.

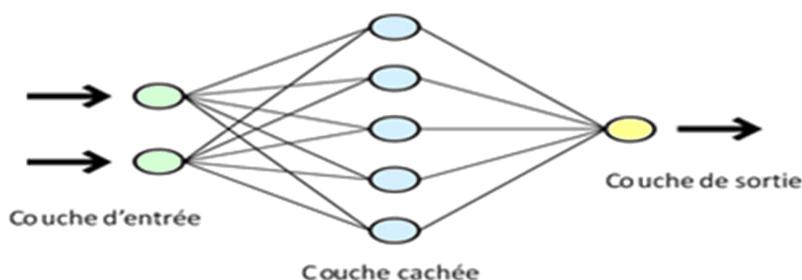


FIGURE 2.6 – Schéma représentatif d'un réseau de neurones artificiels.

2.2.4 Validation du modèle

Une fois développé, le modèle doit être interprété en analysant tous les paramètres statistiques de ce modèle, sa qualité doit être aussi étudiée, cette qualité est vérifiée par ce que l'on appelle validation[26]. Sa robustesse, c'est-à-dire l'influence des composés de la série d'apprentissage sur le modèle, est estimée par des méthodes de validation interne. Afin d'estimer son pouvoir prédictif, des données expérimentales supplémentaires sont nécessaires afin de déterminer la capacité du modèle à prédire ces valeurs c'est ce que l'on appelle validation externe. Enfin, il est important de savoir quel type de molécules utilisées avec quel modèle. On parle alors de domaine d'applicabilité.

2.2.4.1 Validation interne

Dans le passé, la validation interne d'un modèle QSAR a été réalisée en utilisant la validation croisée LOO (leave-one out) ou LMO (leave-many out) qui est quantifiée par le coefficient Q_{cv-loo}^2 . Ce processus consiste à extraire un certain nombre k de molécules du jeu initial à N molécules et à construire un nouveau modèle avec les $(N-k)$ molécules restantes à l'aide des

descripteurs choisis (seules les constantes de la régression changent). Ce processus est ensuite réitéré pour retirer et prédire les valeurs de toutes les molécules de la série d'apprentissage. En fonction du nombre de molécules retirées à chaque itération, on parlera de LOO ou de LMO selon qu'une ou plusieurs molécules est (sont) retirée(s). Dans ces dernières années, d'autres méthodes sont utilisées pour faire la validation interne, telle que la hasardisation de la réponse (Y-Randomization)[27].

La performance des modèles de régression est estimée avec les paramètres statistiques de la validation croisée s_{cv-loo} (l'écart type de la validation croisée), et Q_{cv-loo}^2 (coefficient de détermination de validation, il représente la capacité de la prévision du modèle QSAR). Q_{cv-loo}^2 est calculé à partir de l'équation suivante :

$$Q_{cv-loo}^2 = \frac{\sum_{i=1}^N (y_i - \bar{y}_i)^2 - \sum_{i=1}^N (y_i - \bar{y}_{icv})^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} \quad (2.4)$$

Avec :

y_{icv} : est la valeur prévue de la propriété y par la méthode de la validation croisée ;

\bar{y}_{icv} : est la moyenne arithmétique des y ;

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

2.2.4.2 Test de randomisation

Le test de randomisation permet d'affirmer que la corrélation de chance ne joue aucun rôle durant le développement du modèle. Ces tests sont exécutés quantitativement avec les modèles de QSAR et qualitativement avec les modèles issus de la classification. Les observations sont aléatoirement désorganisées dix fois. C'est-à-dire que la colonne des observations (propriétés) sera changée aléatoirement, en revanche la colonne des descripteurs reste inchangée. A la fin, on obtient dix modèles avec des caractéristiques statistiques spécifiques[28].

La supposition sous-jacente de l'essai de randomisation est la suivante : si les capacités prédictives du modèle ne sont pas dues aux corrélations de chance, alors la désorganisation aléatoire des observations conduira à des modèles (quantitatifs ou qualitatifs) de prévisions faibles, et vice versa.

Cependant, la validation interne est insuffisante pour étudier le pouvoir prédictif d'un modèle. Pour cette raison la validation externe du modèle est devenue une norme et une partie obligatoire dans la modélisation QSPR/QSAR[29][30].

2.2.4.3 Validation externe

Il n'est pas suffisant de développer un modèle de QSAR avec une excellente qualité d'ajustement et de prévision, mais il est également nécessaire de généraliser ces prévisions en l'appliquant sur un échantillon externe (PSET)[31].

Cette méthode consiste à prédire l'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 $R^2(\text{test})$ $Q_{cv-loo}^2(\text{test})$. Récemment plusieurs études ont montré l'insuffisance de ces deux paramètres pour vérifier le pouvoir prédictif des modèles QSAR[32][33]. Par conséquent, d'autres paramètres doivent être vérifiés pour cet objectif. Ces paramètres sont connus sous le nom « critères de validation externe » ou souvent appelés « critères de Trophsa ». [32]

Critères de validation Externe (série de test) :

$$R^2 > 0.6 \quad (\text{critère1}) \quad (2.5)$$

$$Q_{cv-loo}^2 > 0.5 \quad (\text{critère2}) \quad (2.6)$$

$$\frac{R^2 - R_0^2}{R^2} < 0.1 \quad \text{et} \quad 0.85 \leq k \leq 1.15 \quad (\text{critère3}) \quad (2.7)$$

$$\frac{R^2 - R_0'^2}{R^2} < 0.1 \quad \text{et} \quad 0.85 \leq k' \leq 1.15 \quad (\text{critère4}) \quad (2.8)$$

$$|R^2 - R_0^2| \leq 0.3 \quad (\text{critère5}) \quad (2.9)$$

Avec :

R^2 : Coefficient de corrélation pour les molécules de la série de test ;

R_0^2 : Coefficient de corrélation entre les valeurs prédites et expérimentales pour la série de test ;

$R_0'^2$: Coefficient de corrélation entre les valeurs expérimentales et prédites pour la série de test ;

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

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

2.2.4.4 Domaine d'aplicabilité

L'établissement du domaine d'aplicabilité pour chaque modèle constitue un des problèmes les plus importants dans l'analyse de QSAR[34]. 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[35].

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_{ij} .

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

$$A = x_i^T (X^T . X)^{-1} . X_i$$

($i=1\dots n$)

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 :

$$h^* = 3\bar{h} = 3 \frac{\sum_i h_i}{n} = \frac{3p'}{n} \quad (i = 1\dots n) \quad (2.10)$$

Avec :

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.

2.3 Conclusion

Dans ce chapitre nous avons présenté la méthode de QSAR en décrivant ses différentes étapes ; la génération des descripteurs moléculaires, la sélection des variables, l'obtention et la validation des modèles de QSAR. Les notions des méthodes de sélection des variables subjective et objective, des méthodes d'apprentissage, de la validation et de domaine d'applicabilité sont détaillées. Les modèles établissant des relations entre une propriété (ou activité biologique) et des descripteurs représentant les structures chimiques par une fonction d'apprentissage. Le modèle est alors validé par des procédés de validation interne, validation externe et test de randomisation.

Bibliographie

- [1] Frank Jensen. *Introduction to computational chemistry*. John Wiley & Sons, 2017.
- [2] Hans Matter. Computational medicinal chemistry for drug discovery. *Drug Discovery Today*, 9(8) :350, 2004.
- [3] Johann Gasteiger et al. *Handbook of chemoinformatics*, volume 1. Wiley Online Library, 2003.
- [4] Raimund Mannhold, Povl Krosgaard-Larsen, and Hendrik Timmerman. *QSAR : Hansch analysis and related approaches*, volume 1. John Wiley & Sons, 2008.
- [5] Linnan He. *Qsar and classification modeling : Prediction of biological activity of organic compounds from molecular structure*. 2005.
- [6] Philip D Mosier and Peter C Jurs. Qsar/qspr studies using probabilistic neural networks and generalized regression neural networks. *Journal of chemical information and computer sciences*, 42(6) :1460–1470, 2002.
- [7] M Froimowitz. Hyperchem : a software package for computational chemistry and molecular modeling. *Biotechniques*, 14(6) :1010–1013, 1993.
- [8] Robert E Buntrock. Chemoffice ultra 7.0. *Journal of chemical information and computer sciences*, 42(6) :1505–1506, 2002.
- [9] Michael JS Dewar, Eve G Zebisch, Eamonn F Healy, and James JP Stewart. Development and use of quantum mechanical molecular models. 76. am1 : a new general purpose quantum mechanical molecular model. *Journal of the American Chemical Society*, 107(13) :3902–3909, 1985.
- [10] Chengteh Lee, Weitao Yang, and Robert G Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Physical review B*, 37(2) :785, 1988.
- [11] Massimo Baroni, Gabriele Cruciani, Simone Sciabola, Francesca Perruccio, and Jonathan S Mason. A common reference framework for analyzing/comparing proteins and ligands. fingerprints for ligands and proteins (flap) : theory and application. *Journal of chemical information and modeling*, 47(2) :279–294, 2007.
- [12] Steven L Dixon, Alexander M Smondyrev, Eric H Knoll, Shashidhar N Rao, David E Shaw, and Richard A Friesner. Phase : a new engine for pharmacophore perception, 3d qsar model development, and 3d database screening : 1. methodology and preliminary results. *Journal of computer-aided molecular design*, 20(10-11) :647–671, 2006.

- [13] Steven L Dixon, Jianxin Duan, Ethan Smith, Christopher D Von Bargen, Woody Sherman, and Matthew P Repasky. Autoqsar : an automated machine learning tool for best-practice quantitative structure–activity relationship modeling. *Future medicinal chemistry*, 8(15) :1825–1839, 2016.
- [14] Andrea Mauri, Viviana Consonni, Manuela Pavan, and Roberto Todeschini. Dragon software : An easy approach to molecular descriptor calculations. 56 :237–248, 01 2006.
- [15] Roberto Todeschini and Viviana Consonni. *Handbook of molecular descriptors*, volume 11. John Wiley & Sons, 2008.
- [16] M Salahinejad and JB Ghasemi. 3d-qsar studies on the toxicity of substituted benzenes to tetrahymena pyriformis : Comfa, comsia and volsurf approaches. *Ecotoxicology and environmental safety*, 105 :128–134, 2014.
- [17] Matthew Clark, Richard D Cramer, Dumont M Jones, David E Patterson, and Perry E Simeroth. Comparative molecular field analysis (comfa). 2. toward its use with 3d-structural databases. *Tetrahedron Computer Methodology*, 3(1) :47–59, 1990.
- [18] Viviana Consonni and Roberto Todeschini. Molecular descriptors. In *Recent advances in QSAR studies*, pages 29–102. Springer, 2010.
- [19] Milan Randić. Molecular bonding profiles. *Journal of Mathematical Chemistry*, 19(3) :375–392, 1996.
- [20] Christopher A Lipinski. Lead-and drug-like compounds : the rule-of-five revolution. *Drug Discovery Today : Technologies*, 1(4) :337–341, 2004.
- [21] Michal Dereziński and Manfred K Warmuth. Unbiased estimates for linear regression via volume sampling. *arXiv preprint arXiv :1705.06908*, 2017.
- [22] Charles Darwin. On the origin of species by means of natural selection. 1859. *London : Murray Google Scholar*, 1968.
- [23] F Guillaume Blanchet, Pierre Legendre, and Daniel Borcard. Forward selection of explanatory variables. *Ecology*, 89(9) :2623–2632, 2008.
- [24] Sanford Weisberg. *Applied linear regression*, volume 528. John Wiley & Sons, 2005.
- [25] James Devillers. *Neural networks in QSAR and drug design*. Academic Press, 1996.
- [26] Alexandre Varnek and Alex Tropsha. *Cheminformatics approaches to virtual screening*. Royal Society of Chemistry, 2008.
- [27] Patrick Bultinck, Hans De Winter, Wilfried Langenaeker, and Jan P Tollenare. *Computational medicinal chemistry for drug discovery*. CRC Press, 2003.

- [28] Tao Wang, Mian-Bin Wu, Jian-Ping Lin, and Li-Rong Yang. Quantitative structure–activity relationship : promising advances in drug discovery platforms. *Expert opinion on drug discovery*, 10(12) :1283–1300, 2015.
- [29] Alexander Tropsha, Paola Gramatica, and Vijay K Gombar. The importance of being earnest : validation is the absolute essential for successful application and interpretation of qspr models. *Molecular Informatics*, 22(1) :69–77, 2003.
- [30] Partha Pratim Roy, Somnath Paul, Indrani Mitra, and Kunal Roy. On two novel parameters for validation of predictive qsar models. *Molecules*, 14(5) :1660–1701, 2009.
- [31] Suhas J Patankar. Prediction of enzyme inhibition and receptor antagonist properties from molecular structure, and, development of radial basis function neural networks for the analysis of inhibitor binding. 2004.
- [32] Alexander Golbraikh and Alexander Tropsha. Beware of q^2 ! *Journal of molecular graphics and modelling*, 20(4) :269–276, 2002.
- [33] Todd M Martin, Paul Harten, Douglas M Young, Eugene N Muratov, Alexander Golbraikh, Hao Zhu, and Alexander Tropsha. Does rational selection of training and test sets improve the outcome of qsar modeling? *Journal of Chemical Information and Modeling*, 52(10) :2570–2578, 2012.
- [34] E Papa, JC Dearden, and P Gramatica. Linear qsar regression models for the prediction of bioconcentration factors by physicochemical properties and structural theoretical molecular descriptors. *Chemosphere*, 67(2) :351–358, 2007.
- [35] John C Dearden, Patricia Rotureau, and Guillaume Fayet. Qspr prediction of physicochemical properties for reach. *SAR and QSAR in Environmental Research*, 24(4) :279–318, 2013.

Chapitre 3

Modélisation de la Concentration minimale Inhibitrice (CMI) des quinolones vis-à-vis la bactérie *Escherichia coli* (*E.Coli*)

Sommaire

3.1 Introduction	53
3.2 Méthodologie	55
3.2.1 Base de données	55
3.2.2 Introduction et optimisation des structures	59
3.2.3 Génération des descripteurs	59
3.2.4 Analyse statistique	59
3.3 Résultats et discussion	61
3.3.1 Sélection du modèle optimale	61
3.3.2 Analyse de la justesse du modèle optimal	62
3.3.3 Tests de colinéarité et multicollinéarité	66
3.3.4 Analyse des valeurs résiduelles normalisées	68
3.3.5 Validation interne et externe du modèle de QSAR obtenu	69
3.3.6 Domaine d'applicabilité	70
3.3.7 Analyse de la contribution des descripteurs dans le modèle	72
3.4 Conclusion	75

3.1 Introduction

Pendant plus de soixante-dix ans, les antibiotiques ont été l'un des plus grands progrès médicaux. De nos jours, la résistance bactérienne aux antibiotiques est l'une des plus graves

menaces pesant sur la santé publique[1]. Son implication dans la morbidité et la mortalité rend les choix thérapeutiques plus complexes, affectant sérieusement la qualité des soins médicaux. Elle réduit notre gamme thérapeutique sans donner aux laboratoires pharmaceutiques le temps de répondre aux nouveaux besoins avec des produits plus actifs. Par conséquent, le contrôle de la résistance bactérienne aux antibiotiques est un problème de santé majeur pour notre pays en dépit des mesures prises depuis plusieurs années à la fois dans le suivi de la résistance et la prévention de la transmission des bactéries résistantes dans les établissements de santé afin de promouvoir une meilleure utilisation des antibiotiques.

Cette partie de notre travail est commencée par une observation de la résistance bactérienne aux quinolones au niveau du service d'urologie au Centre Hospitalo-Universitaire Ibn Rochd Annaba (Algérie). Il s'agit d'une étude rétrospective réalisée sur une période de trois ans allant du 1 janvier 2011 au 31 décembre 2013. Nous avons étudié 1506 dossiers médicaux des patients admis au service d'urologie du C.H.U. Ibn Rochd Annaba pour différentes pathologies dont 30% des tumeurs de la vessie, 28% des hydrocèles, 14% des tumeurs rénales et autres (Lithiase, rétrécissement urétral, syndrome de jonction pyélo-urétérale, syndrome irritatif du bas appareil et kyste rénale).

On s'est intéressé à la résistance des bactéries isolées vis-à-vis l'acide nalidixique, la ciprofloxacine, la lévofloxacine et l'ofloxacine (quinolones présentes régulièrement dans les antibiogrammes des ECBU). On a trouvé que *E.Coli* est le germe le plus incriminé responsable d'environ 3/4 des résultats positifs de l'analyse d'urine, d'autres bactéries ont également été isolés, comme le montre la Figure 3.1 Concernant la résistance aux quinolones, une augmen-

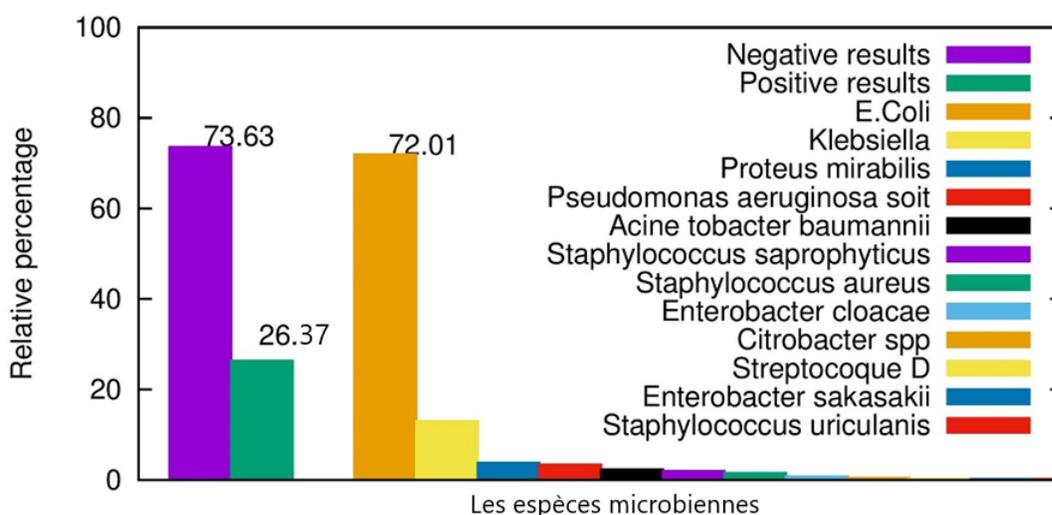


FIGURE 3.1 – Les pourcentages d'examen microbiologique d'urine.

tation inquiétant a été noté pour l'Ofloxacin[®] : passer de 23% en 2011 à 30% en 2013. Une autre quinolone : La Levofloxacin[®] semble plus active avec un taux de résistance de 10% en 2011, légèrement en augmentation jusqu'à 13% en 2013. Par conséquent, des conclusions ont été tirées ; la résistance bactérienne dans ce service est en augmentation, principalement pour *E.Coli* le germe le plus incriminé avec 72.01%.

Sur la base de ces données, l'étude QSAR que nous avons menée s'intéresse aux quinolones par l'établissement d'une relation entre leurs structures et leurs activités inhibitrice vis-à-vis *E.Coli*, le germe le plus incriminé. Ce chapitre est consacré à l'étude QSAR qui est limitée aux propriétés exhibées contre *E.Coli* en raison de l'augmentation incroyable de la résistance de ces bactéries[2][3], utilisant les dérivés de quinolones comme database pour cette étude. Nous avons fait appel à la régression linéaire multiple et aux algorithmes génétiques en tant que méthodes d'apprentissage et de sélection de variables respectivement.

3.2 Méthodologie

3.2.1 Base de données

Dans cette étude, l'ensemble à modéliser est constitué de 85 molécules dérivées de quinolones, qui ont montré une activité inhibitrice vis-à-vis la bactérie *E.coli*[4][5][6][7][8][9]. Leurs structures moléculaires sont illustrées dans la Figure 3.2.

L'activité biologique inhibitrice est rapportée en terme CMI : concentration minimal inhibitrice d'une drogue. Pour notre cas nous avons exprimé l'activité antibactérienne par l'expression $\log CMI$ (mol/ml).

Les molécules de cet ensemble contiennent dans leurs structures des atomes de C, N, O, H, S et un halogène (F) avec un poids moléculaire variant de 347.34 à 531.54 g/mol. Ces molécules présentent une gamme d'activité biologique pour un CMI =0.6 à 4.7. Les valeurs CMI observées et prédites des structures moléculaires sont listées dans le Tableau 3.3.

L'ensemble des molécules est divisé aléatoirement en deux sous-ensembles : un sous ensemble de 27 molécules (environ 30%) choisies au hasard, en tant qu'ensemble de prédiction (PSET), pour évaluer le pouvoir prédictif des modèles proposés. Le reste des molécules 58 (environ 70%) constitue l'ensemble d'apprentissage (Training Set : TSET) utilisé dans la construction du modèle de QSAR.

Figure 3.2: Structures développées des dérivés des quinolones

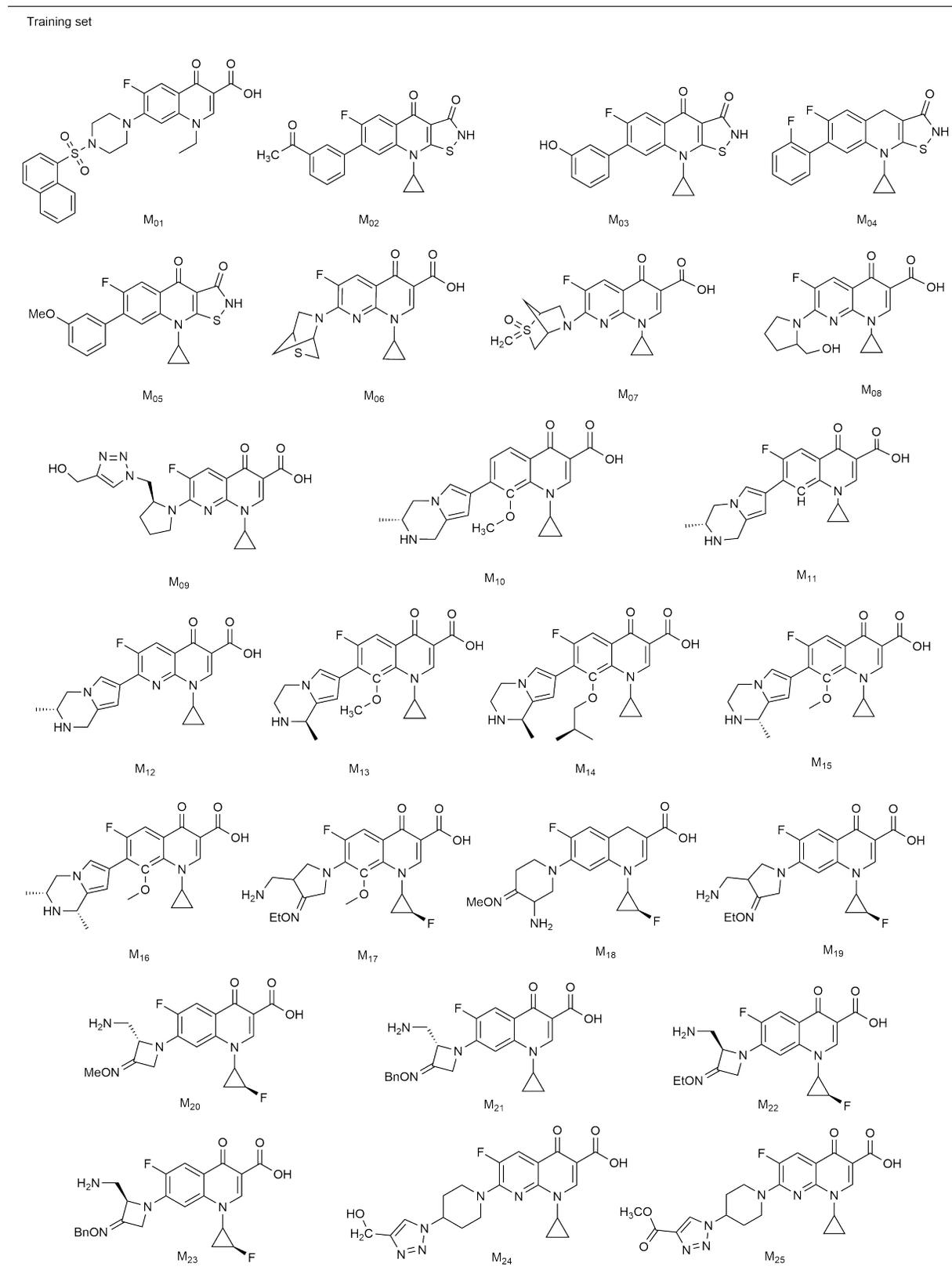


Figure 3.2: Structures développées des dérivés des quinolones *continuation...*

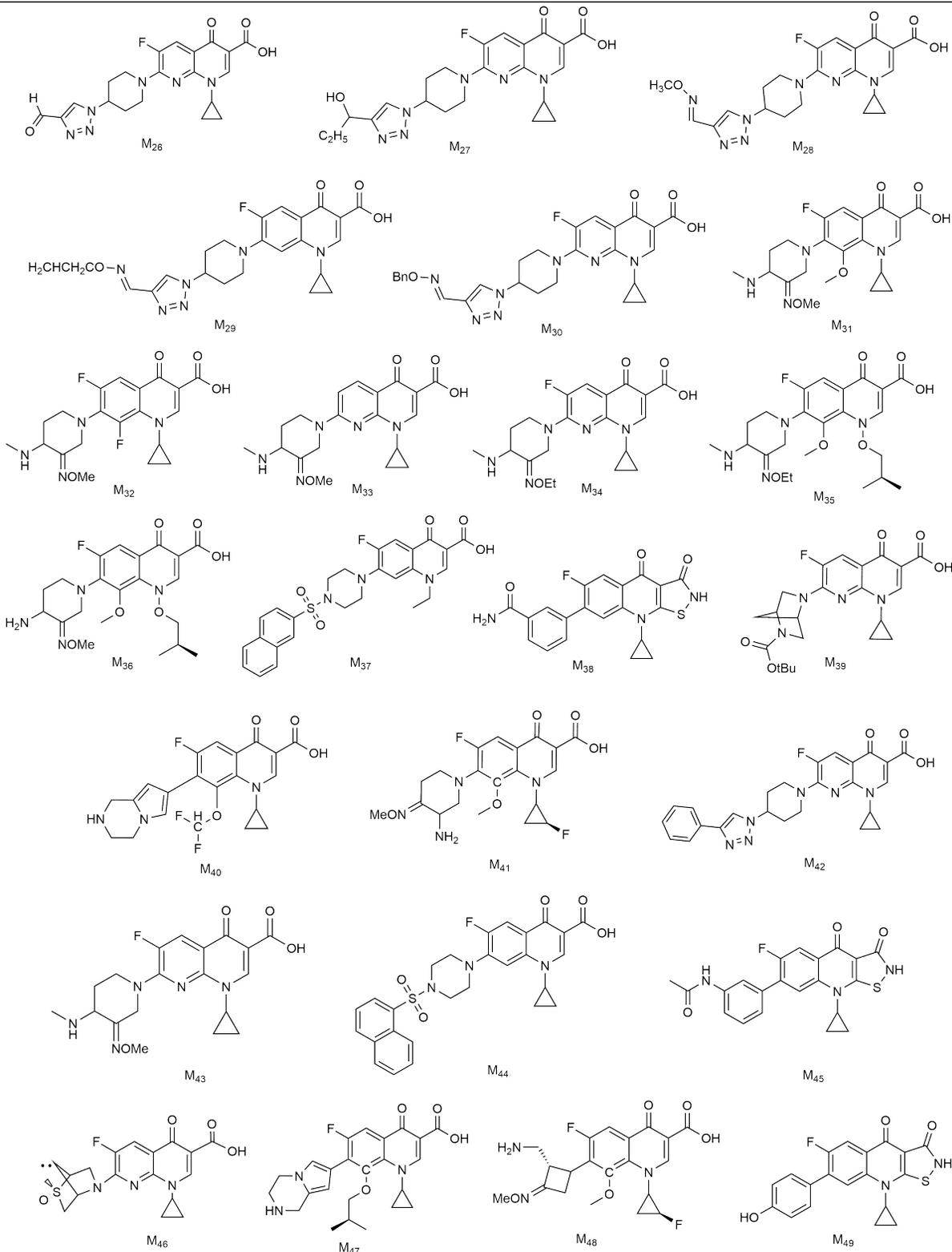


Figure 3.2: Structures développées des dérivés des quinolones *continuation...*

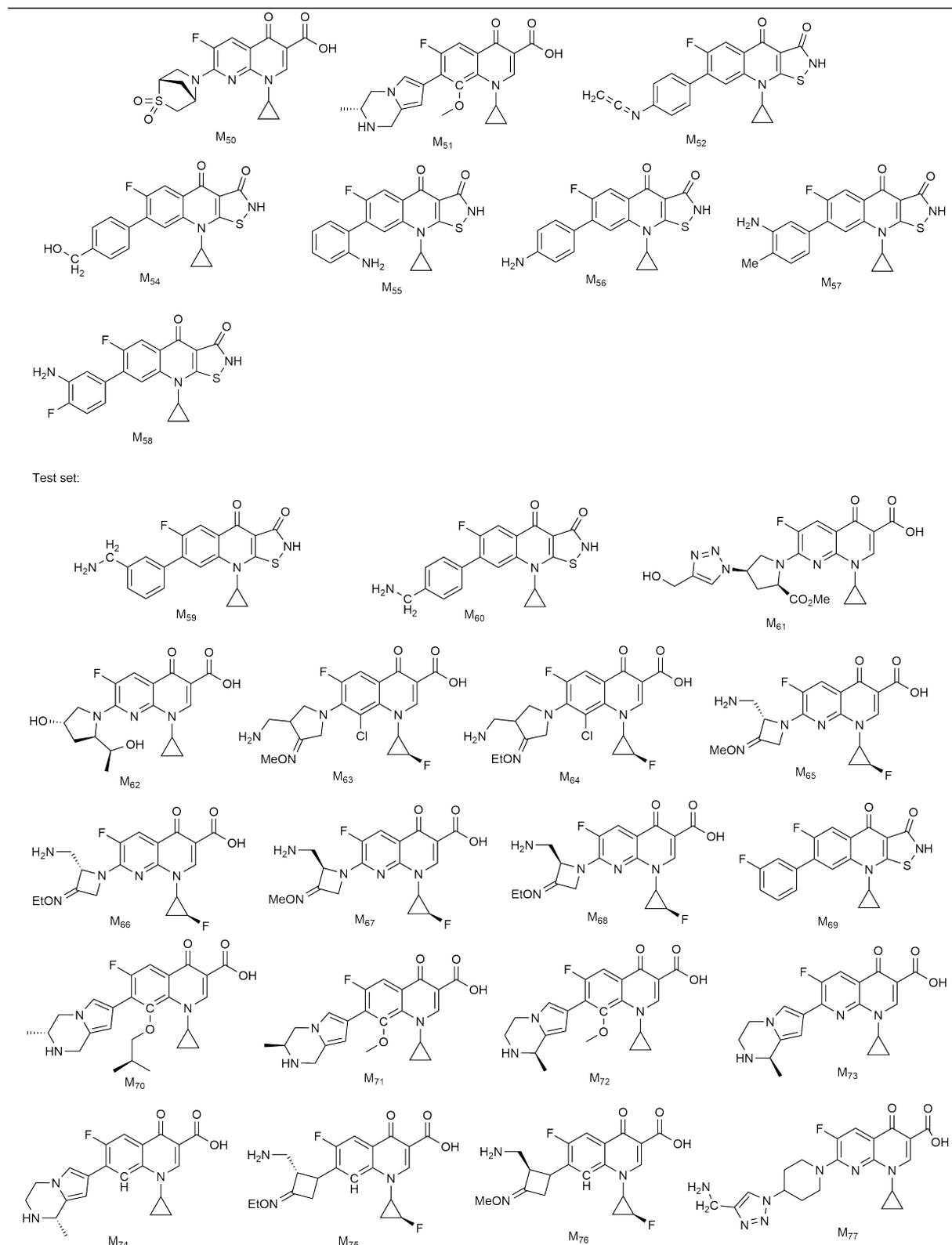
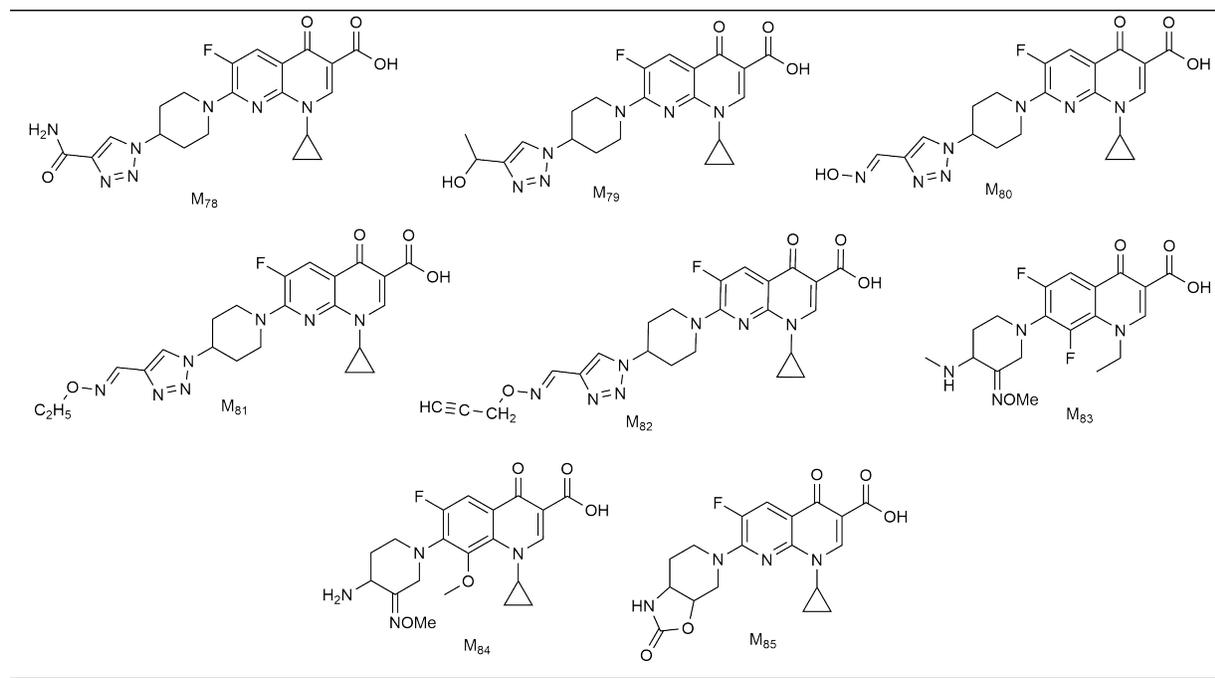


Figure 3.2: Structures développées des dérivés des quinolones *continuation...*



3.2.2 Introduction et optimisation des structures

Les représentations 2D ont été réalisées à l'aide du programme ChemBioDraw ultra 12.0 et ont été exportées vers les logiciels Gaussview 5.0.9 et Gaussian 09. Les structures tridimensionnelles de toutes les molécules ont été générées, et leurs géométries ont été optimisées préliminairement avec la méthode semi-empirique AM1, puis en utilisant la méthode DFT de la chimie quantique inclus dans le logiciel Gaussian 09. Après optimisation, les coordonnées atomiques (x, y, z) de la conformation énergétique minimale pour chaque molécule peuvent être déterminées.

3.2.3 Génération des descripteurs

Les descripteurs moléculaires utilisés sont générés à l'aide du logiciel Dragon 6, programme utilisé pour le calcul des descripteurs moléculaires, qui comprend 4885 descripteurs répartis en 29 blocs. Les cinq descripteurs moléculaires sélectionnés par les AG sont rassemblés dans le Tableau 3.1.

3.2.4 Analyse statistique

La sélection des variables constitue une partie cruciale lors du développement du modèle de QSAR. Elle sert à réduire le nombre de descripteurs en deux étapes successives : sélection

TABLE 3.1 – Liste des cinq descripteurs sélectionnés par les AG.

Descripteur	Description	Bloc
<i>Dp</i>	D total accessibility index / weighted by polarizability	WHIM descriptors
<i>RDF055u</i>	Radial Distribution Function -055/unweighted	RDF descriptors
<i>RCI</i>	ring complexity index	Ring descriptors
<i>Mor17i</i>	signal 17 / weighted by ionization potential	3D-MoRSE
<i>CATS2D 04 AL</i>	CATS2D Acceptor-Lipophilic at lag 04	CATS 2D

objective et subjective.

Initialement, plus de 4485 descripteurs ont été calculés pour chaque molécule dans l'ensemble de données, une sélection caractéristique a été utilisée pour réduire le nombre de descripteurs. La sélection objective est utilisée pour sélectionner un sous-ensemble de descripteurs qui codent le mieux la propriété des dérivés de quinolone en utilisant seulement les variables indépendantes (descripteurs), la réduction des variables a été réalisée par le logiciel Dragon sans employer la variable dépendante. Ce procédé implique :

- Tous les descripteurs présentant les mêmes valeurs pour plus de 90% du data set ont été éliminés ;
- Ensuite, on classe par ordre décroissant les descripteurs en fonction de leurs variances. Dragon peut calculer la matrice de corrélation de tout l'ensemble de descripteurs. A partir de cette matrice, le software élimine un des deux descripteurs ; celui qui prouve un coefficient de corrélation au-dessus de 0,90[10][11].

Après élimination avec sélection objective, nous avons obtenu que 2255 descripteurs, le nombre a été réduit une autre fois par la sélection subjective en cherchant un sous ensemble optimal riche en informations. Dans cette étape on prend en considération la variable dépendante CMI pour le choix des descripteurs, et pour cela nous avons fait appel aux AG comme méthode de réduction en combinaison avec RLM.

Le but général derrière la procédure RLM est de construire un modèle de régression linéaire multiple à partir d'un ensemble de variables indépendantes (descripteurs) en saisissant et en supprimant le descripteur jusqu'à ce qu'il n'y ait plus de raison valable d'entrer ou de retirer. Tous les descripteurs moléculaires sont utilisés pour construire un modèle QSAR par analyse RLM, tel que la mise en œuvre dans le logiciel Xlstat[12].

Les conditions de simulation utilisant les AG étaient : 10000 itérations, changement de longueur d'équation de 3 à 8 descripteurs par modèle pour modéliser l'ensemble des dérivés de quinolone à l'activité, nombre de croisements était de 5000, le facteur de smoothness (douceur) était de 1, probabilité de mutation de 0,5 et nombre initial d'équations générés était 500. AG a été exécuté plusieurs fois (10-15 fois) jusqu'à ce qu'une solution optimisée soit trou-

vée. Le programme des algorithmes génétiques est exécuté sur un PC de processeur *i7* (3.2 GHz). A la fin les modèles sont évalués par la valeur R^2 , le R^2_{adj} , la valeur de SD, (racine de la moyenne des carrés des erreurs) et la valeur F (Fischer statistique), le nombre d'observations N est également noté. Le pouvoir prédictif est évalué par le coefficient de la validation croisée (Q^2_{cv-100}) et par la validation externe.

3.3 Résultats et discussion

Lors de la sélection d'un modèle de QSAR statistiquement fiable, la tâche consiste à concevoir, ou plus souvent à sélectionner un algorithme pour décrire mathématiquement la relation entre les descripteurs et l'activité biologique.

L'application des algorithmes à l'ensemble des descripteurs nous a permis d'obtenir une série de modèles optimaux de différentes dimensions. La détermination du nombre de descripteurs par le modèle de QSAR est importante, car elle empêche dans une certaine mesure les corrélations de chance entre les descripteurs qui constituent le modèle.

3.3.1 Sélection du modèle optimale

Pour déterminer le nombre optimal de descripteurs dans le modèle de QSAR, nous avons utilisé une méthode simple : le point de rupture (breacking point). Nous avons établi consécutivement plusieurs modèles de différentes dimensions allant de 3 jusqu'à 8 descripteurs par modèle. Le modèle optimal correspondant au point d'arrêt est obtenu par l'analyse de la représentation graphique des coefficients de détermination R^2 et de validation (Q^2_{cv-100}) en fonction du nombre de descripteurs[13]. Le Tableau 3.2 présente la série des modèles optimaux obtenus à différentes dimensions allant de 3 à 8 descripteurs par modèle.

La Figure 3.3 montre la projection de R^2 et Q^2_{cv-100} en fonction du nombre des descripteurs

TABLE 3.2 – Série des modèles optimaux obtenus à différentes dimensions.

Nombre des descripteurs	Les descripteurs par modèle	R^2	Q^2_{cv-100}
3	Dp; RDF055u; CATS2D 04 AL	0.713	0.70
4	Dp; RDF055u; CATS2D 04 AL; RDCHI	0.811	0.799
5	Dp; RDF055u; RCI; Mor17i; CATS2D 04 AL	0.91	0.90
6	Dp; RDF055u; RCI; Mor17i; CATS2D 04 AL; R2m	0.93	0.919
7	Dp; RDF055u; RCI; Mor17i; CATS2D 04 AL; R2m; MATS1m	0.935	0.935
8	Mor17i; JGI4; MATS1m; Eig01 _{EA} ; ChiA _G ; DF015e; Mor29e; R8m+	0.935	0.935

en utilisant les algorithmes génétiques. D'après la technique du point d'arrêt (breaking point) le nombre optimal de descripteurs est égale à cinq, avec $R^2= 0.91$ et $Q^2_{cv-100}= 0.90$.

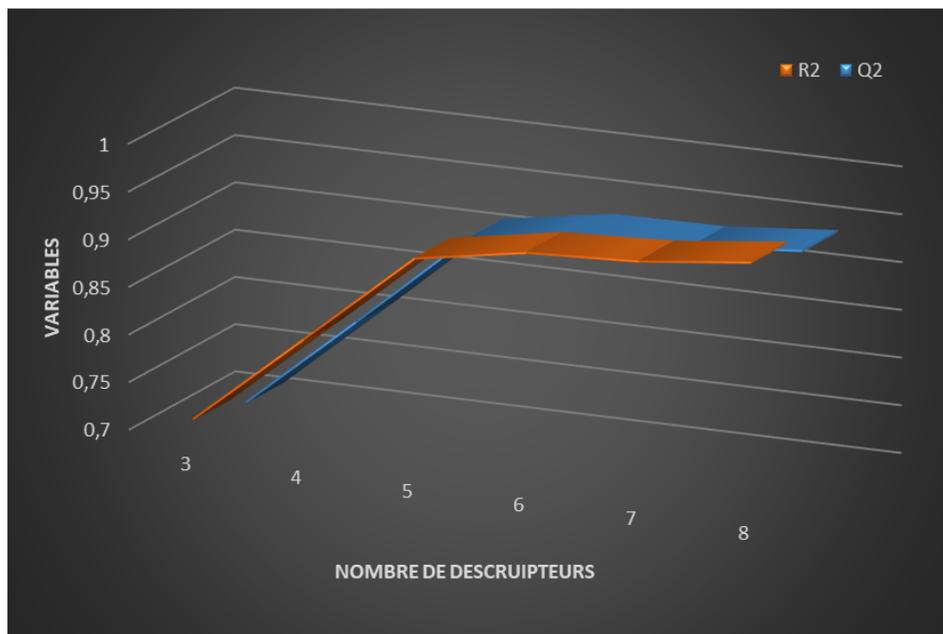


FIGURE 3.3 – Le nombre optimal de descripteurs du modèle optimal obtenu par GA

3.3.2 Analyse de la justesse du modèle optimal

Nous avons établi une relation structure-activité antibactérienne sur *E.coli* avec 85 molécules dérivées de quinolones en utilisant la RLM couplée avec les AG en tant que méthode de sélection de variable. Cinq descripteurs ont été sélectionnés (cités dans le Tableau 3.1) en tant que facteurs ayant une influence significative sur l'activité antibactérienne. Nous avons construit le modèle optimal, à l'aide du logiciel Xlstat (équation 3.1). Les résultats exprimés à l'aide de ce modèle sont rassemblés dans le Tableau 3.3.

$$CMI = 5.723 - 8.60 Dp - 0.03385 RDF055u - 6.808 RCI - 1.022 Mor17i - 0.2049 CATS2D04AL \quad (3.1)$$

Avec :

$$N = 58; R^2 = 0.91; R_{adj}^2 = 0.91; SD = 0.2743; F = 120.41$$

Le modèle exprimé par l'équation 3.1 reliant la variable dépendante CMI avec les cinq descripteurs est accompagné par les mesures statistiques citées ci-dessus : coefficient de détermination R^2 , coefficient de détermination ajusté R_{adj}^2 , l'écart type SD (standard deviation), et la constante de Fischer.

En terme de variabilité, le coefficient de détermination et le coefficient de détermination ajusté

montrent, que 91% de la variance des valeurs observées est expliqué par les cinq descripteurs, avec une valeur élevée de constante de Fischer ($F = 120.41$) et une faible déviation standard ($SD = 0,2743$) ce qui confirme que le modèle optimal explique l'activité antibactérienne (variable dépendante) d'une manière statistiquement significative et satisfaisante.

TABLE 3.3: Valeurs CMI expérimentales et prédites pour TSET et PSET calculé par le modèle RLM-GA.

No.	$-LogCMI_{Exp}$ ($\mu\text{mol/ml}$)	$-LogCMI_{Pred}$ ($\mu\text{mol/ml}$)	Dp	RDF055u	RCI	Mor17i	CATS2D_04_AL	Leverage (hi)
Training Set								
1	0.60	0.74	0.380	28.397	-0.091	-1.066	12.000	0.25067135
2	3.50	3.55	0.345	10.716	-0.086	-1.566	5.000	0.25962742
3	3.79	4.10	0.314	9.572	-0.113	-1.228	3.000	0.06509144
4	3.47	3.58	0.295	8.198	-0.019	-1.134	3.000	0.09549577
5	3.48	3.43	0.322	9.096	-0.093	-1.157	5.000	0.05317436
6	3.46	3.34	0.325	11.623	-0.014	-1.901	6.000	0.27592132
7	2.60	2.71	0.327	17.098	-0.017	-1.863	8.000	0.27394688
8	2.24	1.88	0.310	18.826	0.003	-0.895	7.000	0.16156481
9	1.13	1.24	0.329	28.471	0.023	-1.284	9.000	0.17257736
10	2.63	2.93	0.320	24.247	-0.028	-1.979	7.000	0.07082532
11	3.50	3.24	0.309	12.863	0.058	-1.790	4.000	0.14068250
12	3.18	3.07	0.309	14.710	0.047	-1.811	5.000	0.11444626
13	2.92	2.83	0.342	22.960	0.018	-1.731	4.000	0.05801818
14	2.92	2.98	0.348	38.413	0.036	-2.561	4.000	0.21743082
15	2.32	2.58	0.332	24.797	0.018	-1.469	4.000	0.06486185
16	2.63	2.67	0.344	30.443	-0.016	-1.615	4.000	0.06958903
17	2.05	2.10	0.361	31.690	-0.034	-0.917	3.000	0.20257029
18	3.21	3.54	0.289	22.881	-0.095	-1.025	3.000	0.10438099
19	3.85	3.32	0.321	21.233	-0.066	-1.216	3.000	0.05038511
20	4.69	4.48	0.285	20.024	-0.148	-1.460	3.000	0.13707945
21	2.64	2.98	0.330	31.095	-0.115	-1.768	7.000	0.14241815
22	4.70	4.22	0.284	19.092	-0.135	-1.253	3.000	0.10910278
23	3.27	2.73	0.348	21.720	-0.115	-1.360	7.000	0.06447378
24	2.93	2.83	0.325	15.535	-0.026	-1.256	5.000	0.03874262
25	2.66	2.89	0.368	15.313	-0.054	-1.482	5.000	0.06484982

Continuer dans la page suivante

TABLE 3.3: Valeurs CMI expérimentales et prédites pour TSET et PSET calculé par le modèle RLM-GA.

No.	$-LogCMI_{Exp}$ ($\mu\text{mol/ml}$)	$-LogCMI_{Pred}$ ($\mu\text{mol/ml}$)	Dp	RDF055u	RCI	Mor17i	CATS2D_04_AL	Leverage (hi)
26	3.23	3.02	0.339	14.457	-0.035	-1.464	5.000	0.03153311
27	2.66	2.40	0.361	16.897	-0.006	-1.911	8.000	0.06316960
28	2.36	2.70	0.377	15.577	-0.031	-1.533	5.000	0.28750275
29	2.08	2.01	0.405	17.715	-0.014	-1.474	6.000	0.15183708
30	1.82	1.66	0.414	19.303	-0.017	-1.846	9.000	0.13406368
31	3.54	3.16	0.365	30.427	-0.070	-1.906	4.000	0.11804362
32	4.15	3.87	0.360	18.795	-0.113	-1.689	3.000	0.13063692
33	2.92	2.97	0.342	19.481	-0.108	-1.923	9.000	0.12224275
34	2.01	2.40	0.383	21.149	-0.055	-1.716	7.000	0.25903428
35	1.47	1.48	0.369	38.565	0.031	-1.847	7.000	0.12157641
36	1.75	1.61	0.348	28.655	0.065	-1.692	7.000	0.10470795
37	0.91	0.87	0.425	27.258	-0.052	-1.190	9.000	0.20343004
38	3.50	3.73	0.296	9.319	-0.100	-1.186	5.000	0.06889501
39	2.05	2.26	0.402	28.763	-0.013	-2.673	9.000	0.18387913
40	3.22	3.20	0.307	21.240	-0.036	-1.990	7.000	0.22526157
41	3.85	3.25	0.325	23.785	-0.061	-1.300	3.000	0.05582371
42	1.17	1.35	0.376	18.490	-0.025	-1.541	11.00	0.12495147
43	3.51	3.14	0.357	19.955	-0.071	-2.075	7.000	0.07228965
44	1.21	1.04	0.375	27.245	-0.067	-1.439	12.00	0.18091071
45	3.51	3.75	0.349	11.730	-0.122	-1.589	5.000	0.08634628
46	2.92	2.92	0.335	16.078	-0.017	-2.103	8.000	0.09257504
47	2.64	2.60	0.328	36.111	0.067	-2.152	4.000	0.17339414
48	3.23	3.77	0.300	30.370	-0.106	-1.519	3.000	0.14764938
49	4.08	4.31	0.306	8.642	-0.101	-1.414	3.000	0.06714957
50	3.50	3.50	0.323	16.047	-0.027	-2.104	6.000	0.07766224
51	2.21	2.58	0.334	22.232	-0.051	-1.613	4.000	0.09129070
52	4.29	3.87	0.295	9.758	-0.083	-1.446	5.000	0.06670504
53	3.82	3.61	0.332	10.493	-0.091	-1.476	5.000	0.04814234
54	3.82	4.05	0.307	9.992	-0.075	-1.383	3.000	0.05245198
55	3.46	3.57	0.310	10.480	-0.091	-0.852	3.000	0.07675033
56	3.78	3.98	0.284	8.658	-0.094	-0.946	3.000	0.08288749
57	3.48	3.64	0.311	9.843	-0.058	-1.122	3.000	0.05990887

Continuer dans la page suivante

TABLE 3.3: Valeurs CMI expérimentales et prédites pour TSET et PSET calculé par le modèle RLM-GA.

No.	$-LogCMI_{Exp}$ ($\mu\text{mol/ml}$)	$-LogCMI_{Pred}$ ($\mu\text{mol/ml}$)	Dp	RDF055u	RCI	Mor17i	CATS2D_04_AL	Leverage (hi)
58	3.48	3.50	0.304	9.503	-0.101	-0.635	3.000	0.10936951
Test Set								
59	4.4	4.04	0.324	11.910	-0.102	-1.404	3.000	0.16794792
60	3.82	4.08	0.284	9.576	-0.068	-1.250	3.000	0.22267215
61	0.97	1.17	0.347	28.622	-0.054	-1.181	7.000	0.27527671
62	1.67	2.04	0.357	18.471	-0.064	-0.796	6.000	0.27688470
63	3.54	3.05	0.345	19.473	-0.074	-0.841	2.000	0.27251481
64	2.96	2.85	0.344	22.528	-0.062	-0.820	2.000	0.27840659
65	3.81	3.38	0.287	19.446	-0.036	-1.338	4.000	0.13920913
66	3.51	3.29	0.288	19.427	-0.019	-1.362	4.000	0.15194296
67	3.19	3.17	0.297	21.033	-0.036	-1.269	4.000	0.11544441
68	2.92	2.60	0.325	25.584	-0.019	-1.202	4.000	0.12669077
69	3.77	4.06	0.309	9.101	-0.101	-1.208	3.000	0.1753176
70	2.96	2.56	0.361	38.159	0.069	-2.476	4.000	0.25828249
71	2.95	2.63	0.335	21.783	0.051	-1.661	4.000	0.19939862
72	2.89	3.31	0.342	14.502	0.021	-1.948	4.000	0.24221746
73	2.88	3.29	0.316	16.515	0.008	-1.883	5.000	0.13524025
74	3.18	3.38	0.314	16.040	0.021	-1.825	4.000	0.17348944
75	4.13	4.38	0.286	22.151	-0.135	-1.534	3.000	0.30026026
76	3.89	4.26	0.282	18.356	-0.148	-1.166	3.000	0.30520997
77	2.43	2.69	0.332	20.012	-0.018	-1.375	5.000	0.04948037
78	2.24	2.51	0.342	16.272	-0.014	-1.190	5.000	0.08904159
79	2.64	2.24	0.350	15.798	0.019	-1.598	7.000	0.20680778
80	2.34	2.88	0.330	15.452	-0.024	-1.357	5.000	0.06155180
81	2.06	2.68	0.395	17.214	-0.036	-1.685	5.000	0.28856971
82	2.07	2.16	0.343	17.514	-0.015	-1.090	6.000	0.13723289
83	2.91	3.32	0.356	18.897	-0.124	-1.051	3.000	0.24274444
84	2.52	2.81	0.349	24.323	-0.032	-1.485	4.000	0.09370465
85	2.59	2.55	0.324	13.671	0.026	-1.659	7.000	0.23920959

Les valeurs élevées des coefficients mentionnés ci-dessus résument la corrélation des cinq

descripteurs avec l'activité biologique. Cette corrélation est vérifiée en examinant le tableau d'ANOVA (Analyse Of Variance). A partir du Tableau 3.4 d'ANOVA, la présence ou l'absence de corrélation entre l'ensemble des descripteurs et l'activité biologique est vérifiée en examinant la statistique de Fisher Fobs. Pour la réaliser nous avons fait appel aux deux hypothèses : hypothèse nulle et hypothèse alternative[14].

— **L'hypothèse nulle** : « aucun descripteur n'est lié à l'activité biologique ».

$$B_j = 0 \text{ avec } j = 0; 1; 2; 3; 4; 5.$$

B_j : Coefficient correspond au descripteur j .

Cette hypothèse est acceptée si la valeur de la statistique de Fischer observée est inférieure à la valeur $F_{(0.05;5;58)} = 5.23$.

— **L'hypothèse alternative** : « il existe au moins un descripteur corrélé avec l'activité biologique ».

Cette hypothèse est acceptée lorsque la valeur de la statistique de Fischer observée est supérieure de la valeur $F_{(0.05;58)} = 5.23$.

D'après le Tableau d'ANOVA, la statistique de Fischer observée ($F_{obs}=120.418$) est supérieure à $F_{(0.05;5;58)} = 5.23$, ce qui nous permet d'accepter l'hypothèse alternative et de confirmer qu'il existe au moins un coefficient différent de zéro c'est-à-dire un descripteur corrélé avec l'activité antibactérienne expliquée par CMI.

TABLE 3.4 – Résultats de l'analyse de la variance (ANOVA).

Source	Dégré de liberté	Somme des carrés	Moyenne des carrés	F	Pr > F
Modèle	5	10.316	2.063	120.418	0.0001
Erreur Résiduel	53	3.914	0.075		
Total corrigé	58	12.230			

3.3.3 Tests de colinéarité et multicolinéarité

La colinéarité et la multicolinéarité sont expliquées par la présence d'une forte corrélation d'un descripteur avec un autre (colinéarité) ou avec un ensemble de descripteurs (multicolinéarité). Les descripteurs corrélés entre eux contiennent une grande partie de la même information et provoque ainsi des problèmes calculatoires lors de la génération de la matrice inverse

$(X'X)^{-1}$.

Le problème de colinéarité entre les descripteurs inclus dans le modèle final de QSAR est testé

TABLE 3.5 – Matrice de corrélation de l'équation.

Variabiles	Dp	RDF055u	CATS2D_04_AL	Mor17i	RCI	CMI
Dp	1.000	0.397	0.227	-0.325	0.595	-0.691
RDF055u	0.397	1.000	0,356	-0.383	0.273	-0.538
CATS2D_04_AL	0.227	0.356	1.000	-0.414	0.172	-0.484
Mor17i	-0.325	-0.383	-0.414	1.000	-0.299	0.093
RCI	0.595	0.273	0.172	-0.299	1.000	-0.736
CMI	-0.691	-0.538	-0.484	0.093	-0.736	1.000

par l'examen de la matrice de corrélation, en calculant le coefficient de corrélation pour toutes les combinaisons paires possibles des cinq descripteurs. Les valeurs élevées du coefficient de corrélation $R \geq 0,9$ correspondent aux fortes corrélations entre les descripteurs du modèle. Les résultats obtenus sont récapitulés dans le Tableau 3.5. L'examen de la matrice de corrélation confirme l'absence du problème de colinéarité entre les descripteurs du modèle, expliqué par les faibles valeurs des coefficients de corrélation ($R < 0,9$).

Nous avons ainsi examiné la multicollinéarité par l'approche de la facteur d'inflation de la variance (VIF : Variance Inflation Factor)[15] exprimée par l'équation.

$$VIF(x_k) = \frac{1}{1 - r_k^2} \quad (3.2)$$

Avec $VIF(x_k)$ est la valeur du facteur d'inflation de la variance pour le descripteur x_k ($k=1;2;3;4;5$), et r_k^2 est le coefficient de corrélation carré résultant de la régression du descripteur x_k sur le reste des descripteurs.

Une valeur de $VIF \geq 10$ signifie la présence d'une forte colinéarité entre les descripteurs, l'absence de la multicollinéarité est signifiée par une valeur de $VIF < 10$. On peut aussi confirmer la présence ou l'absence de la forte multicollinéarité par une autre grandeur, similaire au VIF, appelée facteur de tolérance TF donnée par l'équation 3.3 :

$$TF(x_k) = \frac{1}{VIF(x_k)} \quad (3.3)$$

Les valeurs de TF varient dans l'intervalle $0 < TF < 1$.

D'après le Tableau 3.6, les cinq descripteurs ne présentent aucun problème de multicollinéarité avec une valeur maximale de $VIF = 1.724$ et les valeurs de TF toutes inférieures à 1.

TABLE 3.6 – Valeurs des critères VIF et TF pour les descripteurs significatifs.

Statistique	Dp	RDF055u	CATS2D_04_AL	Mor17i	RCI
TF	0.580	0.736	0.782	0.728	0.634
VIF	1.724	1.359	1.279	1.374	1.578

3.3.4 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 de quinolones vis-à-vis la bactérie *E.coli*. 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 expérimentales du CMI. L'évaluation de la distribution normale des valeurs résiduelles normalisées est effectuée par la représentation graphique de chaque valeur résiduelle normalisée en fonction des valeurs observées du CMI. La Figure 3.4 présente 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.

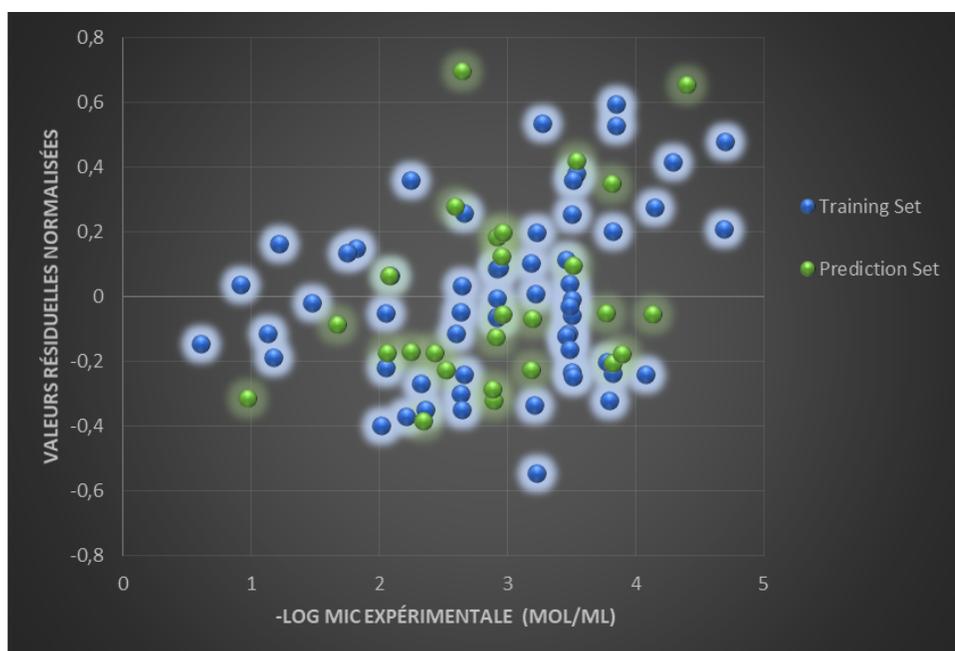


FIGURE 3.4 – Valeurs résiduelles normalisées en fonction de la valeur CMI expérimentale.

3.3.5 Validation interne et externe du modèle de QSAR obtenu

Afin de vérifier la fiabilité et la stabilité du meilleur modèle QSAR (Equation 3.1), nous avons utilisé trois méthodes de validations : la validation interne (validation croisée et test de randomisation) et la validation externe.

La validation croisée est effectuée par la procédure leave-one-out, on retire successivement une molécule de l'ensemble d'apprentissage. Cette procédure est répétée n fois (n est le nombre des molécules qui constituent l'ensemble d'apprentissage) afin de prédire les propriétés de toutes les molécules. La valeur élevée du coefficient de détermination ($Q_{cv-100}^2 = 0.90$) et la petite valeur de l'écart type ($SD_{cv-100} = 0.2886$) de la validation croisée du modèle obtenu prouvent la puissance prédictive de cette approche et la stabilité du modèle. La bonne mesure de prévisibilité du modèle de QSAR obtenu par la validation interne (validation croisée) ne suffit pas. En effet, nous devons en plus généraliser ces prévisions en les appliquant sur un échantillon externe.

Pour cela, nous avons utilisé l'équation du modèle obtenu pour prédire l'activité antibactérienne d'un ensemble externe de molécules (ensemble de prédiction) mentionné au départ. Les valeurs prédites de l'activité des molécules de l'ensemble externe sont affichées dans la troisième colonne du Tableau 3.3. Selon Tropsha et les autres[16][17], un modèle de QSAR ne possède une puissance prédictive acceptable sauf si les conditions suivantes sont satisfaites :

$$Q_{cv-100}^2 = 0.90; \text{ accepter (valeur de seuil } Q^2 > 0.5)$$
$$R^2 = 0.719; \text{ accepter (valeur de seuil } R^2 > 0.6)$$
$$\frac{R^2 - R_0^2}{R^2} < 0.00051; \text{ accepter (valeur de seuil } \frac{R^2 - R_0^2}{R^2} < 0.1)$$
$$K = 1.035; \text{ accepter (valeur de seuil } 0.85 < K < 1.15)$$

Avec :

R^2 : Coefficient de détermination entre les valeurs observées et celles prédites par le modèle (seulement pour l'ensemble d'apprentissage ; 58 molécules).

R_0^2 : Coefficient de détermination issu de la régression des activités observées sur les activités prédites pour tout l'ensemble (85 molécules).

K : est la pente de ligne de régression qui passe par l'origine.

Afin de tester la robustesse du modèle obtenu, nous avons ensuite procédé au test de randomisation. Ce test consiste à la permutation aléatoire des composantes du vecteur de la réponse Y sans toucher à la matrice des descripteurs X, suivie par une reconstruction du modèle. Ce procédé est répété dix fois. Les paramètres statistiques issus de chaque reconstruction du modèle. Sont cités dans le Tableau 3.7 :

TABLE 3.7 – R^2 et Q_{cv-100}^2 issus du test de randomisation.

Statistique	Dp	RDF055u
CATS2D_04_AL	Mor17i	RCI
Model	R^2	Q_{cv-100}^2
Original model	0.920	0.903
Random 1	0.149	-0.075
Random 2	0.052	-0.236
Random 3	0.088	-0.106
Random 4	0.077	-0.117
Random 5	0.049	-0.225
Random 6	0.028	-0.205
Random 7	0.067	-0.131
Random 8	0.101	-0.103
Random 9	0.054	-0.190
Random 10	0.065	-0.155

La représentation par le nuage de points des valeurs du coefficient Q_{cv-100}^2 en fonction du coefficient de détermination R^2 (Figure 3.5) confirme que le modèle présenté, est plus robuste et n'est pas dû à la chance.

La Figure 3.5 montre la corrélation linéaire entre les valeurs observées et les valeurs prédites de l'activité antibactérienne, pour la série d'apprentissage et de test. Les résultats obtenus montrent que la technique linéaire de RLM combinée avec les AG en tant que procédure de sélection de variable est adéquate pour produire un modèle efficace de QSAR capable de modéliser et de prédire l'activité biologique des dérivés de quinolones.

3.3.6 Domaine d'applicabilité

L'objectif principal de la recherche QSAR est de développer des modèles qui fournissent des prédictions précises pour le plus de structures chimiques possible dans l'univers, en particulier pour celles qui n'ont pas été testées ou pour lesquelles des données expérimentales fiables ne sont toujours pas disponibles. Cependant, le modèle QSAR doit toujours être vérifié en ce qui concerne son domaine d'applicabilité afin de produire des données prédites qui peuvent être considérées fiables sauf pour des composés chimiques qui présentent une grande similarité structurale.

Le domaine d'applicabilité du modèle a été analysé à l'aide d'un graphique de Williams (Figure 3.6)[18], où la ligne verticale est la valeur de levier critique ($h^* = 0,31$), et les lignes horizontales sont la valeur de coupure pour l'espace Y . Ce dernier est une représentation gra-

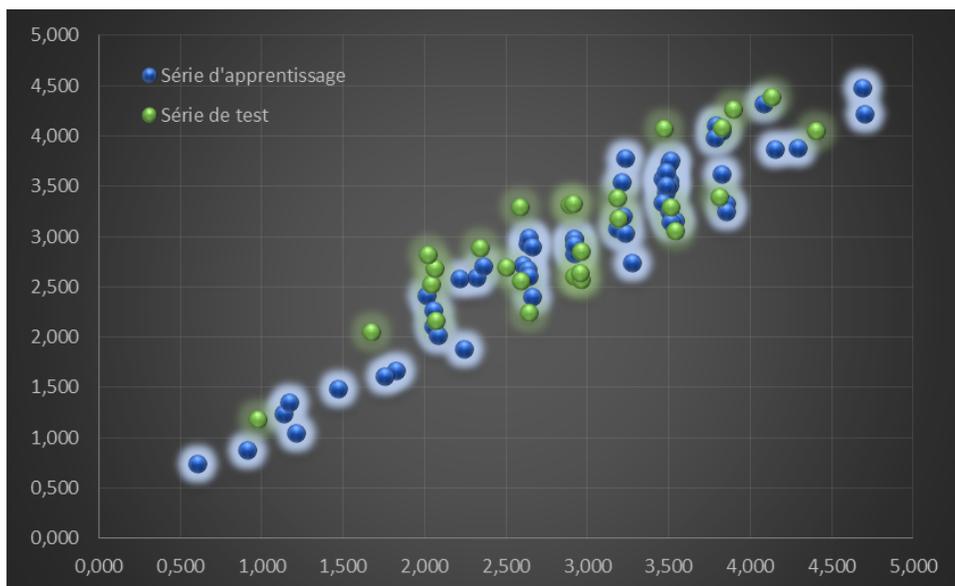


FIGURE 3.5 – Valeurs prédites en fonction des valeurs observées pour la série d'apprentissage(TSET) et la série de prédiction(PSET).

phique des valeurs résiduelles normalisées en fonction des valeurs de levier de chaque molécule à partir du quel on peut déterminer les molécules correspondantes aux points aberrants ainsi que les produits chimiques influents sur le modèle.

D'après la Figure 3.6 nous remarquons que tous les composés de l'ensemble de données se

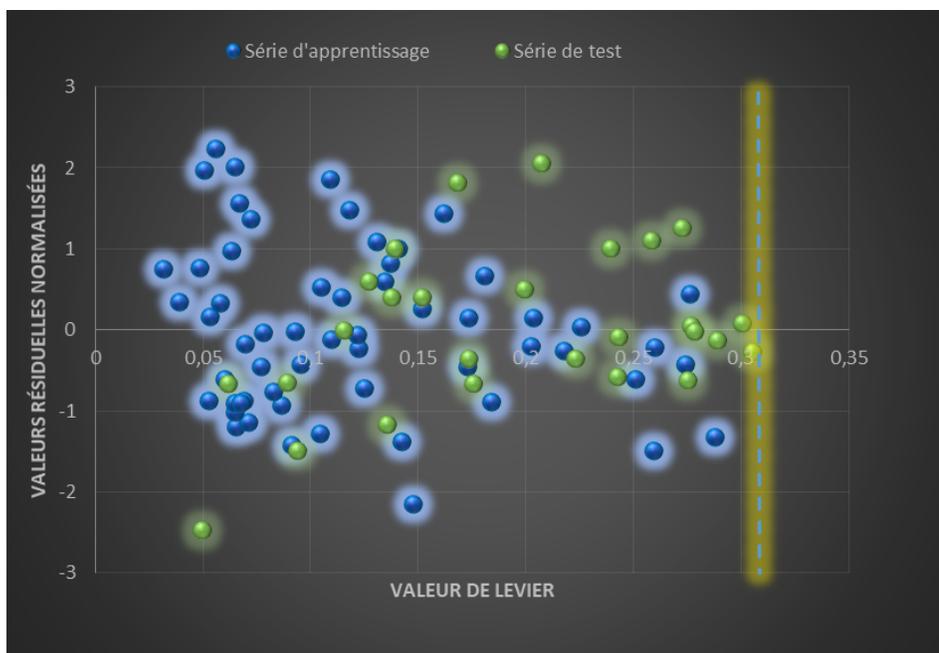


FIGURE 3.6 – Graphe de Williams obtenu avec les cinq descripteurs.

trouvent à l'intérieur de cette zone. Apparemment, aucune donnée influente ou aberrante n'a été détectée. Par conséquent, le modèle présenté par l'équation affiche les meilleurs paramètres statistiques, une bonne prédiction et une applicabilité.

3.3.7 Analyse de la contribution des descripteurs dans le modèle

Pour voir l'importance de chaque descripteur pour la prédiction de l'activité antibactérienne, les contributions relatives de 5 descripteurs au modèle RLM-GA ont été déterminées par l'équation suivante[19][20] :

$$C_i = \frac{100 * \Delta m_i}{\sum \Delta m_i} \quad (3.4)$$

Avec :

C_i : La contribution du descripteur i .

Δm_i : La valeur moyenne de l'écart absolu entre les valeurs observées et les valeurs prédites de toutes les molécules.

Les résultats obtenus sont représentés dans la Figure 3.7. D'après la contribution, les cinq descripteurs sont classés en fonction de leur contribution comme suit :

$$Dp > RCI > Mor17i > CATS2D_04_AL > RDF055u$$

.

Interpréter un modèle QSAR en termes de contribution spécifique des substituants et d'autres

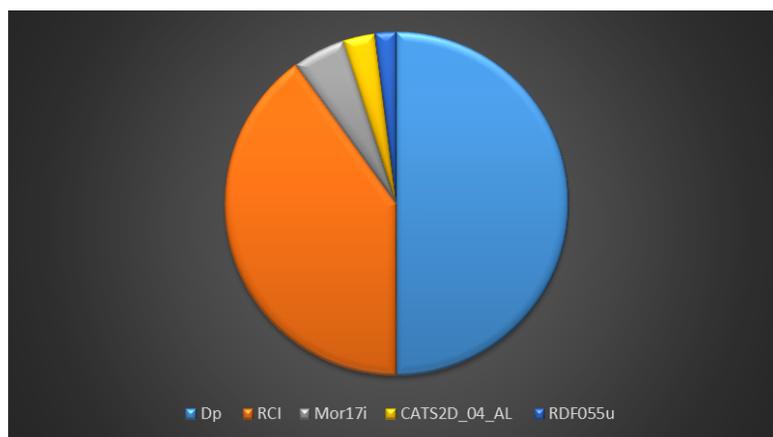


FIGURE 3.7 – La contribution des descripteurs dans le modèle.

caractéristiques moléculaires à l'activité biologique modélisée reste toujours une énigme difficile à résoudre. Dp et RCI sont les variables les plus importantes dans cette équation, car elles

présentent la plus grande contribution dans notre modèle. Pour montrer l'importance de ces deux descripteurs, une comparaison a été faite entre les molécules, N°26 et N°30, N°20 et N°48.

Comme on peut le voir, la différence entre les composés N°30 et N°26 est dans le radical en position 7 de cycle quinolone, alors que le radical de la molécule N°26 contient 4 atomes d'azote et un atome d'oxygène, et pour la molécule N°30 le radical contient 5 atomes de l'azote et l'oxygène (augmentation de la polarisabilité), cette différence conduit à une augmentation des valeurs de D_p de 0,339 à 0,414 et par conséquent diminuer la valeur CMI comme l'indique la Figure 3.8.

D'autre part, comme l'indique l'équation, l'inhibition des dérivés de quinolone dépendait du

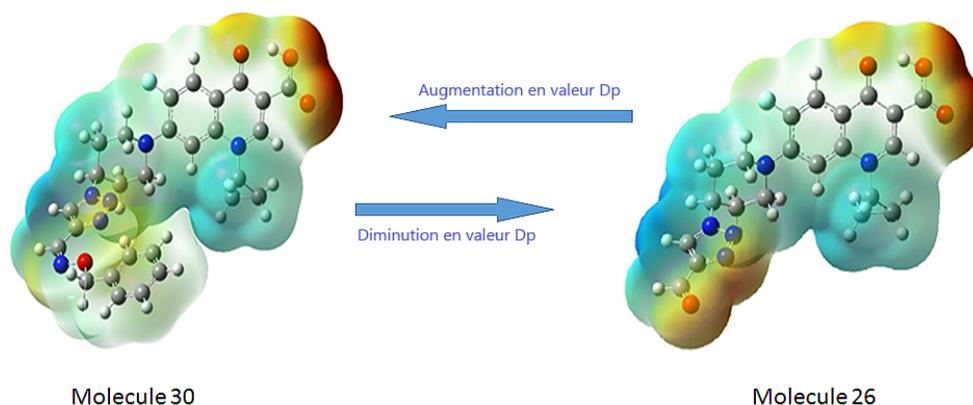


FIGURE 3.8 – Comparaison des potentiels électrostatiques moléculaires (MEP) entre les composés N°30 et N°26.

paramètre stérique RCI (ring complexity index). En effet, la structure analogue du 2-benzoylacrylique dans la molécule de quinolone consiste en un gros système conjugué contenant 12 électrons π . Les atomes de carbone et les hétéroatomes fluorine, azote et oxygène fournissent des électrons à paire unique. Une étude précédente a indiqué que le transfert d'électrons peut se produire lorsque les quinolones interagissent avec l'ADN ou la protéine[21]. La position et le type de substituant pouvant affecter le mécanisme d'action du pharmacophore (2-benzoylacrylique) avec l'ADN. C'est pourquoi une autre comparaison a été faite entre les composés N°48 et N°20; la seule différence entre ces deux composés est le type de substituant dans le pharmacophore de ces dérivés en position 8. Ainsi, une différence apparemment faible conduit à une augmentation des valeurs de RCI de -0.148 à -0.106 et par conséquent diminue la valeur de CMI comme le montre la Figure 3.9.

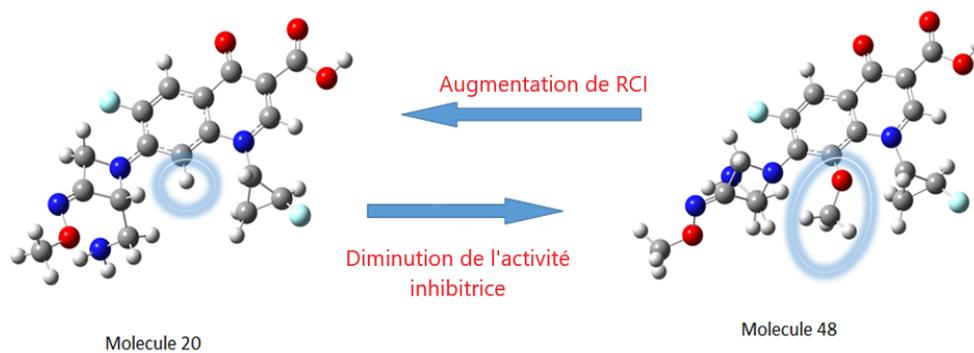


FIGURE 3.9 – Comparaison structurale entre les composés N°20 et N°48.

3.4 Conclusion

Au cours de ce travail nous avons appliqué la méthode de régression linéaire multiple dans le développement d'un modèle décrivant la relation quantitative structure-activité inhibitrice de la bactérie *E.coli*. Pour cela nous avons fait appel à la méthode de régression linéaire multiple en tant que méthode d'apprentissage en combinaison avec les algorithmes génétiques comme méthode de sélection. Cinq descripteurs moléculaires sélectionnés par les AG sont : *Dp*; *RDF055u*; *RCI*; *Mor17i*; *CATS2D_04_AL*. L'étude de la contribution des cinq descripteurs sélectionnés a montré que les descripteurs moléculaires *Dp*, et *RCI*, exerces l'effet le plus important sur le pouvoir prédictif de l'activité antibactérienne.

Bibliographie

- [1] Jackson E Fowler Jr and Thomas A Stamey. Studies of introital colonization in women with recurrent urinary infections. vii. the role of bacterial adherence. *The Journal of urology*, 117(4) :472–476, 1977.
- [2] W Goettsch, W Van Pelt, N Nagelkerke, MGR Hendrix, AGM Buiting, PL Petit, LJM Sabbe, AJA Van Griethuysen, and AJ De Neeling. Increasing resistance to fluoroquinolones in *Escherichia coli* from urinary tract infections in the Netherlands. *Journal of antimicrobial chemotherapy*, 46(2) :223–228, 2000.
- [3] TW Maddox, PD Clegg, NJ Williams, and GL Pinchbeck. Antimicrobial resistance in bacteria from horses : Epidemiology of antimicrobial resistance. *Equine veterinary journal*, 47(6) :756–765, 2015.
- [4] Jason A Wiles, Qiuping Wang, Edlaine Lucien, Akihiro Hashimoto, Yongsheng Song, Jijun Cheng, Christopher W Marlor, Yangsi Ou, Steven D Podos, Jane A Thanassi, et al. Isothiazoloquinolones containing functionalized aromatic hydrocarbons at the 7-position : synthesis and in vitro activity of a series of potent antibacterial agents with diminished cytotoxicity in human cells. *Bioorganic & medicinal chemistry letters*, 16(5) :1272–1276, 2006.
- [5] Xiaoguang Huang, Dongliang Chen, Ning Wu, Aiqin Zhang, Zhenhua Jia, and Xingshu Li. The synthesis and biological evaluation of a novel series of c7 non-basic substituted fluoroquinolones as antibacterial agents. *Bioorganic & medicinal chemistry letters*, 19(15) :4130–4133, 2009.
- [6] Bin Zhu, Brett A Marinelli, Raul Goldschmidt, Barbara D Foleno, James J Hilliard, Karen Bush, and Mark J Macielag. Synthesis and antibacterial activity of 7-(1, 2, 3, 4-tetrahydropyrrolo [1, 2-a] pyrazin-7-yl) quinolones. *Bioorganic & medicinal chemistry letters*, 19(17) :4933–4936, 2009.
- [7] Xiaoguang Huang, Aiqin Zhang, Dongliang Chen, Zhenhua Jia, and Xingshu Li. 4-substituted 4-(1h-1, 2, 3-triazol-1-yl) piperidine : Novel c7 moieties of fluoroquinolones as antibacterial agents. *Bioorganic & medicinal chemistry letters*, 20(9) :2859–2863, 2010.
- [8] Kai Lv, Jinwei Wu, Jian Wang, Mingliang Liu, Zengquan Wei, Jue Cao, Yexin Sun, and Huiyuan Guo. Synthesis and in vitro antibacterial activity of quinolone/naphthyridone derivatives containing 3-alkoxyimino-4-(methyl) aminopiperidine scaffolds. *Bioorganic & medicinal chemistry letters*, 23(6) :1754–1759, 2013.

- [9] Hongmin Liu, Ju Huang, Jiayang Wang, Minghua Wang, Mingliang Liu, Bin Wang, Huiyuan Guo, and Yu Lu. Synthesis, antimycobacterial and antibacterial evaluation of 1-[(1r, 2s)-2-fluorocyclopropyl] fluoroquinolone derivatives containing an oxime functional moiety. *European journal of medicinal chemistry*, 86 :628–638, 2014.
- [10] Roberto Todeschini, Viviana Consonni, and R Mannhold. Methods and principles in medicinal chemistry. *Kubinyi H, Timmerman H (Series eds) Handbook of molecular descriptors*. Wiley-VCH, Weinheim, 2000.
- [11] María Jimena Martínez, Ignacio Ponzoni, Mónica F Díaz, Gustavo E Vazquez, and Axel J Soto. Visual analytics in cheminformatics : user-supervised descriptor selection for qsar methods. *Journal of cheminformatics*, 7(1) :39, 2015.
- [12] XLSTAT Addinsoft. Data analysis and statistical software for microsoft excel. *Paris, France*, 2010.
- [13] Lennart Eriksson, Joanna Jaworska, Andrew P Worth, Mark TD Cronin, Robert M McDowell, and Paola Gramatica. Methods for reliability and uncertainty assessment and for applicability evaluations of classification-and regression-based qsars. *Environmental health perspectives*, 111(10) :1361, 2003.
- [14] Kevin P Weinfurt. Repeated measures analysis : Anova, manova, and hlm. 2000.
- [15] Edward R Mansfield and Billy P Helms. Detecting multicollinearity. *The American Statistician*, 36(3a) :158–160, 1982.
- [16] Alexander Golbraikh and Alexander Tropsha. Predictive qsar modeling based on diversity sampling of experimental datasets for the training and test set selection. *Molecular diversity*, 5(4) :231–243, 2000.
- [17] Alexander Golbraikh and Alexander Tropsha. Beware of q^2 ! *Journal of molecular graphics and modelling*, 20(4) :269–276, 2002.
- [18] Tatiana I Netzeva, Andrew P Worth, Tom Aldenberg, Romualdo Benigni, Mark TD Cronin, Paola Gramatica, Joanna S Jaworska, Scott Kahn, Gilles Klopman, Carol A Marchant, et al. Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. *ATLA*, 33 :155–173, 2005.
- [19] Rajarshi Guha and Peter C Jurs. Interpreting computational neural network qsar models : a measure of descriptor importance. *Journal of chemical information and modeling*, 45(3) :800–806, 2005.
- [20] Fang Zheng, Ersin Bayram, Sangeetha P Sumithran, Joshua T Ayers, Chang-Guo Zhan, Jeffrey D Schmitt, Linda P Dwoskin, and Peter A Crooks. Qsar modeling of mono-and

bis-quaternary ammonium salts that act as antagonists at neuronal nicotinic acetylcholine receptors mediating dopamine release. *Bioorganic & medicinal chemistry*, 14(9) :3017–3037, 2006.

- [21] Jianying Hu, Wanfeng Wang, Zhou Zhu, Hong Chang, Feng Pan, and Binle Lin. Quantitative structure- activity relationship model for prediction of genotoxic potential for quinolone antibacterials. *Environmental science & technology*, 41(13) :4806–4812, 2007.

Deuxième partie

**STRATEGIE DE VIRTUEL
SCREENING A BASE DE
PHARMACOPHORE POUR
L'IDENTIFICATION DES NOUVEAUX
INHIBITEURS DU L'ENZYME
LACTATE DÉSHYDROGÉNASE**

Chapitre 4

Généralité sur l'enzyme Lactate Déshydrogénase et ses inhibiteurs

Sommaire

4.1 Introduction	81
4.2 Métabolisme du glucose	81
4.2.1 Les trois voies métaboliques pour le pyruvate	81
4.3 Lactate déshydrogénase	83
4.4 Caractéristiques de l'isotype LDH-A	84
4.5 Mécanisme catalytique	85
4.6 Métabolisme du cancer	86
4.6.1 Effet Warburg	88
4.6.2 Mécanismes moléculaires	90
4.7 Autres possibilités thérapeutiques	90
4.7.1 Lactate déshydrogénase et paludisme	91
4.8 Les inhibiteurs du LDH dans la littérature	92
4.8.1 Gossipol et ses dérivés	92
4.8.2 Oxamate	92
4.8.3 Dérivés de N-hydroxyindole	93
4.8.4 Les Inhibiteurs bifonctionnels	94
4.8.5 Galloflavine	94
4.8.6 Complexes de cuivre (II) et de zinc (II)	94
4.8.7 Quinoléine et 1,4-dihydro-4-quinolones	95
4.8.8 Les azoles	95
4.8.9 Récentes découvertes de la société Genetech Inc	96
4.9 Conclusion	98

4.1 Introduction

Qui n'a jamais craqué devant des pâtisseries ou des bonbons ? Pourtant les méfaits des aliments sucrés sont connus : surpoids, caries...etc. Cette mauvaise réputation est d'ailleurs un peu exagérée car les sucres sont également les carburants de notre organisme. Le glucose est le substrat métabolique prédominant pour la plupart des cellules animales, son catabolisme dans le cytoplasme et/ou les mitochondries fournit l'énergie et les molécules de départ pour la prolifération et la croissance cellulaire[1].

4.2 Métabolisme du glucose

La première partie du métabolisme du glucose est une séquence de réactions chimiques catalysées par de multiples enzymes, appelée glycolyse. Le résultat final de la glycolyse, qui se produit entièrement dans le cytoplasme, est de deux molécules de pyruvate et quatre molécules d'ATP (Adénosine triphosphate) pour une seule molécule de glucose, avec réduction simultanée de deux molécules de NAD^+ (nicotinamide adénine dinucléotide)[2][3].



Avec :

ATP : Adénosine triphosphate ;

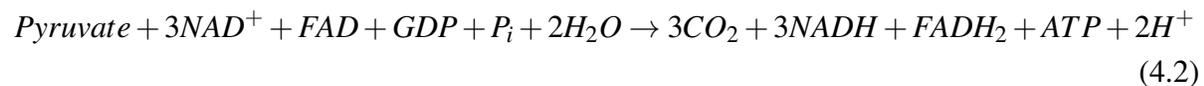
NADH ; NAD^+ : Nicotinamide adénine dinucléotide.

4.2.1 Les trois voies métaboliques pour le pyruvate

Les molécules de pyruvate ainsi obtenues peuvent suivre trois voies métaboliques :

4.2.1.1 En présence d'oxygène (conditions aérobies) :

- **Oxydation complète du dioxyde de carbone** : dans ce cas, le pyruvate est transporté vers les mitochondries, subit une décarboxylation oxydative catalysée par le complexe enzymatique de pyruvate déshydrogénase en générant une molécule d'acide acétique qui forme l'acétyl-CoA pour se lier à la coenzyme-A. L'acétyl-CoA participe ensuite au cycle de Krebs (également connu sous le nom de cycle de l'acide citrique ou du cycle de l'acide tricarboxylique), dont les produits finaux sont le CO_2 , H_2O .



Avec :

FAD; $FADH_2$: Flavine adénine dinucléotide ;

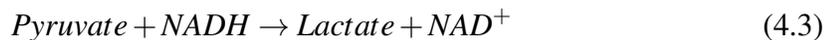
GDP : Protéine G.

P_i : phosphate inorganique.

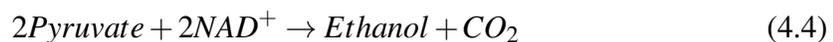
Les électrons issus de l'oxydation du glucose dans le cytoplasme et les mitochondries sont envoyés dans des complexes de chaînes respiratoires pour générer le gradient électrochimique des protons mitochondriaux. Le retour des protons dans la matrice mitochondriale à travers le canal H^+ /ATP est utilisé comme force motrice pour la phosphorylation oxydative dans la synthèse de l'ATP à partir de l'ADP et du phosphate.

4.2.1.2 En l'absence d'oxygène (conditions anaérobies) :

- **Réduction partielle en acide lactique** : Sans quitter le cytoplasme, le pyruvate est réduit par l'enzyme lactate déshydrogénase (LDH) avec oxydation simultanée de NADH en NAD^+ . Le processus prend le nom de fermentation lactique, il est typique pour de nombreux micro-organismes et cellules d'organismes supérieurs lorsque la quantité d'oxygène disponible est limitée.



- **Réduction partielle en éthanol et dioxyde de carbone** : Comme la fermentation lactique, le pyruvate est d'abord réduit en acétaldéhyde puis en éthanol par action du pyruvate décarboxylase et d'alcool déshydrogénase, avec oxydation simultanée de NADH en NAD^+ . Le processus prend le nom de fermentation alcoolique, il est typique pour certains types de levure et de micro-organismes.



Le bilan final d'ATP de la phosphorylation oxydative est de 32-36 moles d'ATP par mole de glucose contre 2 moles d'ATP résultant de la fermentation associée à la glycolyse. La phosphorylation oxydative (OXPHOS) fournit une quantité d'énergie presque 20 fois plus grande que les voies de fermentation, mais nécessite de l'oxygène pour atteindre la phase finale. Par conséquent, dans des conditions normales, la fermentation lactique et alcoolique n'est effectuée que dans des conditions de faible disponibilité d'oxygène (hypoxie) ou de détérioration génétique des enzymes mitochondriales.

4.3 Lactate déshydrogénase

La lactate déshydrogénase (LDH) est une enzyme oxydoréductase capable de catalyser la conversion du pyruvate en lactate et vice-versa[4]. Plusieurs types de cette enzyme existent, qui diffèrent suivant la nature du cofacteur de la réaction et le stéréoisomère du lactate formé. Le lactate possède en effet deux isomères D et L et il existe des lactates déshydrogénases capables de former soit l'un soit l'autre des produits.

On la trouve dans une grande diversité d'organismes, aussi bien végétaux qu'animaux, avec des niveaux élevés d'homologie structurale[5]. Alors que pour des nombreuses bactéries, la LDH est la seule façon possible de glycolyse, elle est donc essentielle pour leur survie. Dans les cellules eucaryotes, le rôle de la lactate déshydrogénase est de reléguer la situation transitoire de manque d'oxygène au niveau cellulaire, par exemple typique des cellules musculaires sous réserve d'une activité intense.

Chez l'homme, la lactate déshydrogénase peut être constituée de deux sous-unités principales :

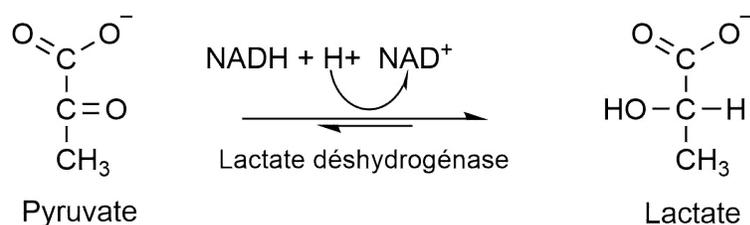


FIGURE 4.1 – La réaction catalysée par la lactate déshydrogénase.

- M(muscle) : principalement présente dans les cellules musculaires et le foie ;
- H(cœur) : présente dans le cœur, la rate, les reins, le cerveau et les érythrocytes.

Par la suite, une troisième sous-unité a été identifiée :

- X : exclusivement présente au niveau du testicule mature.

L'association tétramérique des deux sous-unités principales conduit à cinq isoformes possibles pour la lactate déshydrogénase :

- **LDH-1** (H_4) : également appelée **LDH-H** ou **LDH-B** ; est principalement présente dans les cellules musculaires squelettiques, le myocarde et les globules rouges ainsi que dans l'écorce du rein ;

- **LDH-2** (M_1H_3) : est principalement distribuée dans le myocarde et les globules rouges, mais également présente dans le pancréas, l'écorce de rein, le poumon et le muscle squelettique ;
- **LDH-3** (M_2H_2) : On la trouve dans les poumons, le placenta, le muscle squelettique et le pancréas ;
- **LDH-4** (M_3H_3) : est présente dans la moelle osseuse du rein, du muscle squelettique, du poumon et du placenta ;
- **LDH-5** (M_5) : également appelée **LDH-M** ou **LDH-A**, à prédominance hépatique et musculaire. Elle est également présente dans la moelle osseuse rénale et le pancréas.

La sous-unité X ou C forme une sixième isoforme homotétramère[6], apparemment impliquée dans la production de spermatozoïdes.

- **LDH-X** : également appelée **LDH-C4** ; est principalement présente au niveau du testicule.

La sous-unité M a une plus grande affinité pour le pyruvate et une plus forte V_{max} (vitesse initiale maximale) pour la réduction du pyruvate que la sous-unité H. Ainsi, la sous-unité M, particulièrement le tétramère LDH-A, catalyse en préférence la conversion du pyruvate en lactate avec production de NAD^+ . En revanche, la sous-unité H oxyde de préférence le lactate en pyruvate, ce qui permet aux cellules d'utiliser le lactate comme source de nourriture pour le métabolisme oxydatif (tissus cardiaques et neuronaux) et/ou de la gluconéogenèse (dans le foie et les reins)[7].

Les sous-unités H et M présentent un bon degré de conservation structurelle, avec une identité d'acides aminés de 75%. Les résidus du site catalytique sont particulièrement bien conservés, ce qui rend très difficile l'obtention de molécules inhibant sélectivement une seule des sous-unités.

4.4 Caractéristiques de l'isotype LDH-A

Ma thèse focalise sur l'étude de l'isoforme LDH-A ou LDH-5. La LDH-A est une protéine homotétramère, dans laquelle chaque chaîne a une séquence primaire de 331 acides aminés. Le gène codant est LDHA, il se localise dans le chromosome 11 ; des mutations à ce niveau ont été liées au développement de myoglobulinurie par la souche[8].

D'un point de vue structurel, deux domaines peuvent être reconnus, comme montre la Figure 4.2 :

- **Domaine de liaison pour le cofacteur NADH** : se compose des résidus 18-160 et 246-264. La région du cofacteur est constituée de quatre alpha-hélices et d'un ensemble de

six feuillets parallèles bêta. C'est un domaine structural récurrent dans la déshydrogénase, appelé enveloppe de Rossman, découvert la première fois Michael Rossman[9]. Ce type d'enroulement représente peut-être un site de liaison primordial pour les dinucléotides ;

- **Domaine de liaison pour le substrat** : constitué des résidus 161-245. C'est un domaine beaucoup plus petit, et se compose de quatre feuilles bêta et trois hélices alpha. Dans ce domaine il y a une boucle composée de résidus 97-108 et extrêmement mobile. La boucle mobile est capable de se fermer sur la coenzyme après la liaison avec le substrat, bloquant ainsi le site actif du pyruvate.

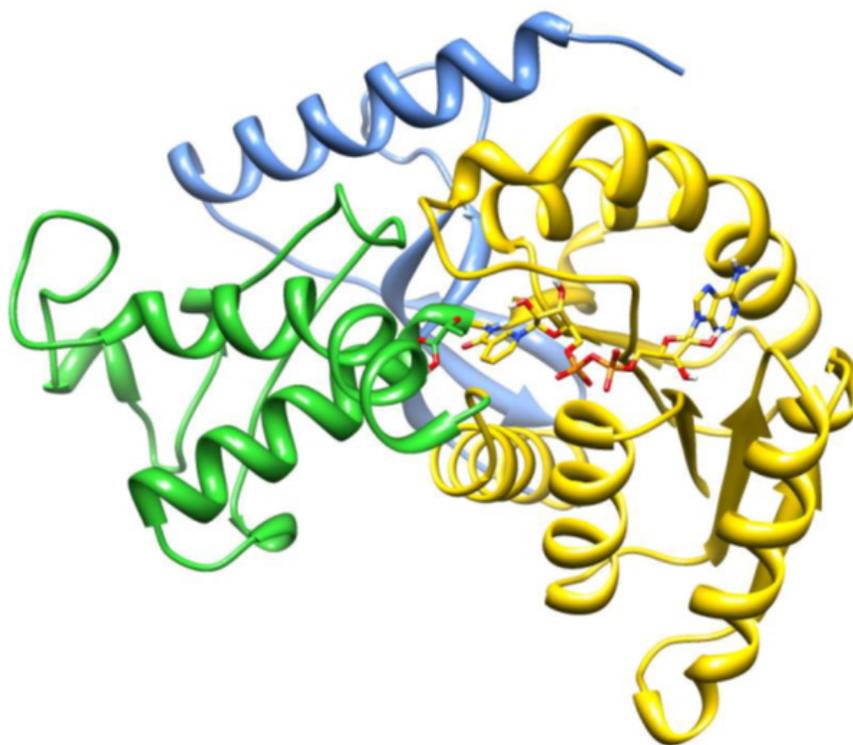


FIGURE 4.2 – Domaines de liaison du LDH-A ; en jaune le domaine du cofacteur, avec le NADH, en vert le domaine du substrat, avec le pyruvate (chaîne B, cristal 4JNK).

la Figure 4.3 montre la comparaison entre les deux configurations de la boucle (cavité) mobile : ouverte (rouge) et fermée (bleue). Le rose est la structure de l'oxytamate, un analogue du substrat.

4.5 Mécanisme catalytique

Les phases d'oxydation du pyruvate catalysées par la LDH-5 sont :

1. Asp165 interagit avec His192 protonée, stabilisant la forme cationique ;

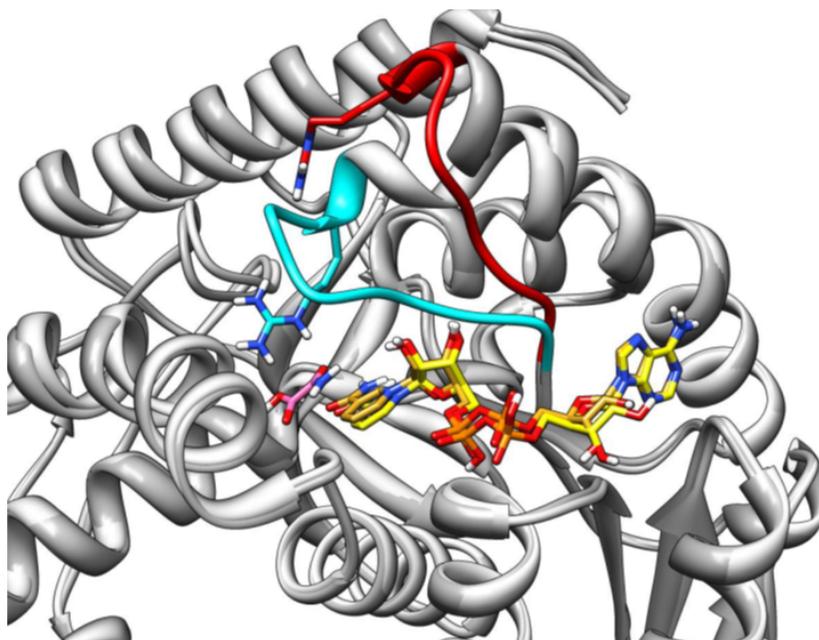


FIGURE 4.3 – Chevauchement de chaîne B cristal 1I10 (boucle fermée, en bleu) et chaîne C cristal 4JNK (boucle ouverte, rouge). En rose l'inhibiteur oxydé ; en jaune le cofacteur NADH.

2. Attaque de NADH au domaine enzymatique 1 ; le noyau nicotinamide est stabilisé par des interactions avec la chaîne hydrophobe de Ile247 ;
3. Entrée de pyruvate et interaction d'ions avec Arg168 dans le second domaine ;
4. Interaction du pyruvate avec His 192 et NADH ;
5. Fermeture de la boucle mobile et déplacement de la base de l'hélice alpha-D ;
6. L'acide aminé Arg105 de la boucle mobile, polarise le groupement carbonyle pyruvate, favorisant le transfert d'un proton du NADH au carbone ;
7. Transfert stéréospécifique d'un ion hydrure (H^-) de carbone du groupement carbonyle du cycle dinucléonotinicamide NADH au carbone du groupement carbonyle du pyruvate ;
8. HIS 192 donne un proton d'oxygène du carbonyle.

4.6 Métabolisme du cancer

La cancérogénicité est un processus qui consiste en plusieurs étapes, dans lesquelles elle reconnaît huit caractéristiques principales qui sont développées et qui confèrent aux cellules cancéreuses leurs caractéristiques malignes :

- Résistance à l'apoptose ;
- Potentiel répliatif infini ;

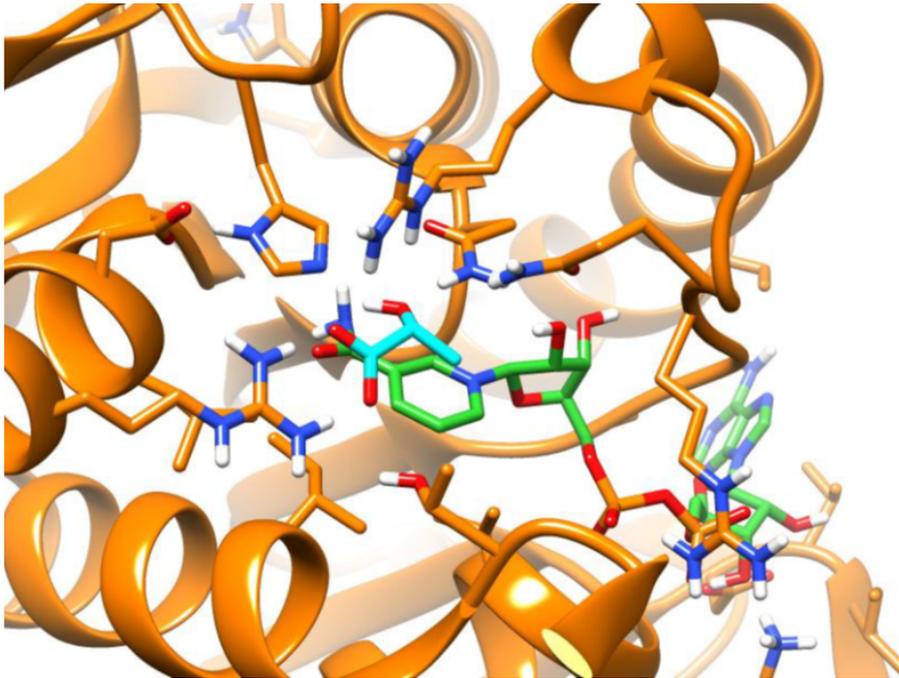


FIGURE 4.4 – Lactate (bleu) dans le site catalytique avec le cofacteur (vert).

- Proliférer même en l'absence de facteurs de croissance ;
- Grandir en présence de signaux d'arrêt ;
- Echapper au système immunitaire ;
- Promouvoir l'angiogenèse ;
- Invasion tissulaire et métastase ;
- Reprogrammer votre métabolisme[10][11].

Ce développement a d'abord été comparé à un processus évolutif de 1976 par Nowell[12] et confirmé par de nombreuses études. Nowell a émis l'hypothèse que la cancérogénicité passerait par les phases d'expansion des clones, dans lesquelles les cellules avec des mutations qui augmentent le taux de croissance ou les capacités de survie se reproduiront plus que les cellules environnantes avec des mutations moins favorables.

Selon ce modèle, la présence commune d'une propriété phénotypique dans les populations cancéreuses suggère qu'elles devraient conférer un avantage sélectif à la croissance.

Dans ce travail de dissertation, le métabolisme des cellules cancéreuses, et en particulier les altérations qui donnent un avantage évolutif tumoral, ont été analysés en détail et en même temps peuvent être une cible thérapeutique utile.

L'énorme quantité d'études cliniques sur le métabolisme du glucose est principalement due à l'application de la technologie d'imagerie connue sous le nom de tomographie par émission

de positons (PET) en utilisant un analogue du glucose, le fluorodésoxy-glucose (FdG) comme traceur. En utilisant FdG-PET, il a été possible de montrer de façon univoque que la plupart des tumeurs humaines, primaires et métastatiques ont une augmentation significative de l'absorption du glucose. Ces études ont associé à plusieurs reprises un pronostic flou et une agressivité accrue avec une augmentation de l'absorption du glucose.

4.6.1 Effet Warburg

Selon le paragraphe précédent, il semblerait paradoxal de constater qu'une curieuse, mais commune, propriété tumorale invasive est un métabolisme du glucose altéré, qui est fortement déplacé vers la glycolyse couplée à la fermentation lactique, même en présence d'oxygène. Cette particularité des cellules cancéreuses est connue depuis 1920 et porte le nom du nom de Warburg, du nom du lauréat du prix Nobel de biochimie allemand Otto Warburg qui l'a décrite pour la première fois[13]. La première hypothèse de Warburg que l'on appelle la glycolyse aérobie résultant des dommages mitochondriaux aux cellules cancéreuses a alors été refusée. À première vue, cette hypothèse semble être en contraste avec le modèle évolutif de la cancérogenèse, car l'avantage prolifératif du phénotype glycolytique n'est pas immédiatement apparent. Les principales contre-indications de la glycolyse aérobie sont :

- Le métabolisme du glucose anaérobie est inefficace, car il ne produit que 2 molécules d'ATP par molécule de glucose, alors que, comme mentionné ci-dessus, une oxydation complète à travers OXPHOS produit 38 molécules d'ATP par molécule de glucose ;
- Les produits métaboliques de glycolyse, tels que les ions H^+ , provoquent une acidification importante de l'espace extracellulaire, ce qui peut provoquer une toxicité cellulaire.

Intuitivement, il semblerait que les forces darwiniennes qui prévalent au cours de l'évolution somatique des tumeurs invasives devraient aller à l'encontre d'un phénotype métabolique qui est un ordre moins qu'efficace de ses concurrents et rend l'environnement toxique.

Alors, quels sont les avantages possibles de l'effet Warburg ?

4.6.1.1 Survivre à l'hypoxie

Le stade zéro de la cancérogenèse, ou développement du cancer, se caractérise par la prolifération de cellules anormales qui n'ont pas de phénotype infiltrant et ne peuvent donc vaincre la membrane basale et entrer en contact avec les vaisseaux lymphatiques et les vaisseaux sanguins.

Cela signifie que l'apport des nutriments nécessaires à la croissance (principalement le glucose et l'oxygène) passe obligatoirement par un processus de diffusion à travers la membrane basale et les différentes couches de cellules précancéreuses. Une faible concentration en oxygène

semble être le premier substrat limitant pour les cellules cancéreuses, car les modèles de réponse au motif ont montré que la pO_2 diminue plus rapidement avec la distance des vaisseaux sanguins que les niveaux de glucose.

La prolifération du carcinome détermine donc invariablement le développement des régions hypoxiques, où ils seront les cellules préférées qui ont augmenté le métabolisme glycolytique.

4.6.1.2 Générer de la biomasse

Pour se répliquer par mitose, une cellule proliférative doit répliquer tout son contenu cellulaire. Cela nécessite beaucoup de nucléotides, d'acides aminés et de lipides. Bien que l'ATP soit cruciale pour générer de l'énergie, les réactions biochimiques nécessaires à la réplication de la biomasse ont d'autres exigences, telles que les molécules de NADPH, ou des intermédiaires de glycolyse pour la synthèse d'acides aminés non essentiels. Il est évident que dans ce cas, la conversion de tout le glucose en CO_2 va à l'encontre des besoins de la cellule proliférative.

Les métabolites obtenus par la glycolyse et leur destinée principale :

- Le glucose-6-phosphate utilisé dans la voie des pentoses phosphates et essentiel pour la préparation du ribose-5-phosphate pour la biosynthèse des nucléotides et du glycogène ;
- Le phosphate de dihydroxyacétone utilisé dans la synthèse des triglycérides et des phospholipides membranaires ;
- Pyruvate utilisé pour la synthèse de malate et de l'alanine ;
- Acétyl-CoA utilisé pour la synthèse d'acides gras ;

4.6.1.3 Augmenter la capacité d'infiltration

La plupart des acides lactiques produits par les cellules cancéreuses sont excrétés à l'extérieur de la cellule, avec une réduction conséquente du pH environnant. Cela semble faciliter l'infiltration de la tumeur et en même temps empêcher le système immunitaire de reconnaître la cellule et l'attaquer[14]. Il a été montré comment des fortes concentrations de lactate peuvent également avoir une action inhibitrice sur les cellules du système immunitaire[15].

4.6.1.4 Résistance aux agents oxydants

L'interconnexion entre la glycolyse et la voie des pentoses phosphates permet aux cellules de produire de grandes quantités de NADPH, qui agit comme une défense contre les oxydatifs stimuli, tels qu'un microenvironnement antagoniste ou des agents chimiothérapeutiques[16]. Pour confirmer cette observation, des études ont montré que l'expression du transaldolase 1,

qui transforme le glucose-6-phosphate de la glycolyse en le premier substrat de la voie des pentoses phosphatés, a augmenté dans de nombreux cas des tumeurs[17].

4.6.2 Mécanismes moléculaires

Un régulateur clé de la réponse tumorale à l'hypoxie, avec l'enchaînement des mutations dans le métabolisme de la glycolyse, est le facteur de transcription HIF-1-alpha (facteur inducible par l'hypoxie, Hypoxia Inducible Factor). HIF-1 est une protéine hétérodimérique composé d'une sous-unité alpha (HIF-1 alpha), régulé par la pression O_2 et une sous-unité bêta (HIF-1-bêta), constituée de manière expressive à un niveau nucléaire. Dans des conditions normales de pression d' O_2 , HIF-1-alfa est rapidement hydroxylé par l'enzyme proline hydroxylase-2 et par la suite dégradée à l'aide de protéasome (un grand complexe d'enzymes protéolytiques). Pour être en mesure d'accomplir ses activités, la prolyl hydroxylase-2 nécessite la présence de plusieurs facteurs, parmi lesquels la base d'oxygène. Cela signifie que dans des conditions hypoxiques, la proline hydroxylase-2 ne catalysera pas l'hydroxylation de HIF-1-alpha, qui s'accumulera alors dans la cellule et migrera dans le noyau pour rejoindre la sous-unité bêta.

Après stabilisation de HIF-1-alpha, la protéine hétérodimérique est capable d'interagir avec l'ADN au niveau des éléments sensibles à l'hypoxie (hypoxia-responsive elements), et activer la transcription de plusieurs gènes essentiels pour la survie dans des conditions hypoxiques, y compris[18] :

- LDH-A ;
- Transporteurs de glucose (GLUT1 et GLUT3) ;
- Acteurs d'angiogenèse (VEGF, angiopoïétine-2, oxyde nitrique synthétase) ;
- L'hexokinase I et II ;
- Facteurs hématopoïétiques (c.-à-d. transferrine et érythropoïétine) ;
- Les extrudeuses de lactate (MCT-4) ;
- La pyruvate déshydrogénase kinase (PDK), qui inhibe le complexe du pyruvate déshydrogénase. De cette façon, HIF-1 inhibe la conversion du pyruvate en citrate, la première étape du cycle de Krebs.

4.7 Autres possibilités thérapeutiques

Les molécules qui inhibent la lactate déshydrogénase peuvent être utilisées comme :

- Antipaludéens ;
- Agent anti-carie ;
- Agents anti-infertilité masculine ;

4.7.1 Lactate déshydrogénase et paludisme

Le paludisme est la maladie la plus répandue des parasites et la deuxième maladie infectieuse dans le monde, après la tuberculose dans la propagation et taux de mortalité. Le paludisme provoque de 200 millions d'infections par an, et à partir de 500 000 à 1 million de décès[19].

L'agent étiologique du paludisme est un parasite appartenant au genre *Plasmodium*, dont quatre sous-espèces principales sont connues ; Parmi eux, le plus connu est *Plasmodium falciparum*. Dans son stade à l'intérieur des cellules sanguines, le *Plasmodium* fonde son métabolisme presque exclusivement sur la glycolyse couplée à la fermentation homolactique, probablement parce que le parasite est incapable d'utiliser un cycle fonctionnel de Krebs (ou de l'acide citrique) pour une grande partie de son cycle de vie. La demande énergétique pour les érythrocytes avec des parasites, telle que l'absorption de glucose est jusqu'à 100 fois plus élevée que les érythrocytes normaux, et presque tous sont dirigés vers la production de lactate.

Étant donné qu'un approvisionnement constant en NADH est une condition préalable à la glycolyse, et que la LDH agit comme principale source de régénération du NADH par NAD^+ , on a émis l'hypothèse que l'inhibition de la LDH pouvait stopper la production d'ATP entraînant la mort de *Plasmodium falciparum*. Tout composé capable de bloquer la LDH est un agent antipaludéen potentiel avec un mécanisme d'action différent à celui des médicaments existants. Par ce biais, la lactate déshydrogénase de *Plasmodium falciparum* (PfLDH) a été proposée comme cible par plusieurs auteurs.

La séquence d'acides aminés prédite par séquençage génomique et DNA indique que les résidus essentiels sont conservés (en utilisant la numérotation standard : His195, Arg109 et Arg171). Une comparaison entre les structures cristallines de la LDH de *Plasmodium falciparum* et l'humaine montre ces deux différences clés :

- Positionnement différent du cofacteur NADH, qui se traduit par des changements dans la poche de liaison avec déplacement de l'anneau de nicotinamide d'environ 1,2 Å ;
- Changements dans la séquence (qui incluent l'insertion de 5 résidus) et dans la sous-structure de la région de boucle qui se ferme sur le site actif pendant la catalyse

Ces changements contribuent à une augmentation du volume de la fente du site actif dans *Plasmodium falciparum* par rapport à son homologue humain. En plus de ces variations structurales, il existe également des différences cinétiques significatives entre les deux enzymes ; en fait, en utilisant un dérivé de NADH, les différences cinétiques entre la lactate déshydrogénase humaine et parasitaire sont si élevées, que l'on peut utiliser l'activité de la LDH comme un indicateur du parasitisme in vivo[20].

4.8 Les inhibiteurs du LDH dans la littérature

4.8.1 Gossipol et ses dérivés

Le gossipolo (du nom latin de la plante de coton, *Gossypium*, dont il a été extrait) est un flavonoïde trouvé dans les graines de nombreuses plantes appartenant à la famille des Malvacées avec fonction insecticide naturel[21].

De nombreuses études ont révélé le large spectre d'action, allant du mâle anticonceptionnel à l'antitumorale, avec une bonne activité in vitro également en tant qu'antioxydant, antitumoral et pesticide. Il a ensuite été précisé que le gossipol est capable d'inhiber la déshydrogénase et en particulier la LDH en déplaçant le cofacteur NADH. L'absence de sélectivité explique le large spectre d'action car elle affecte à la fois la LDH-A humaine (activité antitumorale) et la LDH-C4 (activité non spermicide) et non humaine (telle que l'isoforme de *P. falciparum*).

Bien que le gossipol a montré une activité inhibitrice très prometteuse (valeurs de K_i entre $0,7 \mu$ pour la P.fLDH et $1,9 \mu$ pour la LDH-5), il s'agissait d'une molécule hautement réactive pouvant causer une toxicité non précisée et des dommages graves tels que arythmies, hypochimie, insuffisance rénale, faiblesse musculaire et parfois paralysie.

Pour cette raison, des dérivés ont été développés qui imitent l'action sans avoir les mêmes

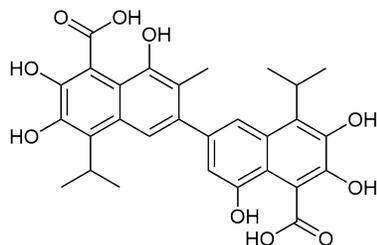


FIGURE 4.5 – Structure de Gossipol.

effets secondaires, principalement en remplaçant les groupes aldéhydes et phénoliques qui peuvent chélater les métaux et réagir avec les groupes amine.

Parmi les dérivés de Gossipolo, nous trouvons FX11, qui a une sélectivité d'action pour l'isoforme humaine LDH-5 avec des résultats prometteurs dans des études in vitro et in vivo. Il est à noter cependant qu'il n'a pas été possible d'identifier un mécanisme de liaison unique pour le FX11, ce qui confirme le fait que la molécule a une action non spécifique due à la partie catécholique hautement active.

4.8.2 Oxamate

Oxamate est un isostère structural de pyruvate qui peut rivaliser pour son site actif et inhiber LDH-5 avec K_i égal à 136μ . Bien que l'un des inhibiteurs les plus étudiés, il est plutôt non

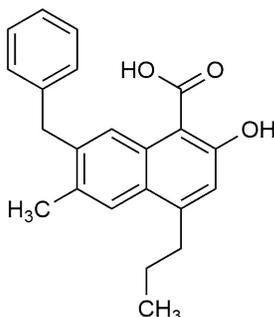


FIGURE 4.6 – Structure d'inhibiteur FX11.

spécifique, a une faible puissance et pénètre avec difficulté les membranes cellulaires, et est donc actif dans des études in vitro seulement à des concentrations assez élevées[22].

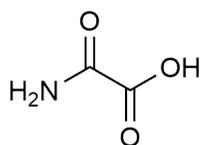


FIGURE 4.7 – Structure d'Oxamate.

4.8.3 Dérivés de N-hydroxyindole

Cette classe d'inhibiteurs agit avec un mécanisme compétitif à la fois contre le pyruvate et le NADH, avec des valeurs micromolaires faibles de K_i , possédant maintenant le noyau structural OH-COOH de lactate et d'autres inhibiteurs constituent une classe innovante de composés hétérocycliques[23].

Des études subséquentes utilisant des expériences de résonance plasmonique de surface ont donné des résultats peu communs, conduisant à l'hypothèse d'une liaison pas entièrement spécifique[24]. Un développement récent de cette classe de composés conduit à la détection d'un ester méthylique conjugué à une molécule de glucose avec un pouvoir inhibiteur significativement plus faible ($K_i = 37,8 \mu$) mais une excellente perméabilité cellulaire grâce au transport actif à travers les transporteurs GLUT.

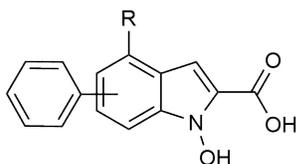


FIGURE 4.8 – Dérivés de N-hydroxyindole.

4.8.4 Les Inhibiteurs bifonctionnels

Il s'agit d'une classe d'inhibiteurs qui contient deux portions qui peuvent interagir avec la poche de liaison pyruvate et NADH, connectées via un lien. Un premier exemple d'un inhibiteur bifonctionnel conjugué entre l'acide glycolique et le NADH (GA-NADH), qui a démontré une activité inhibitrice élevée sur les isoformes de LDH isolées dans le muscle cardiaque des bovins et des lapins ($K_i \simeq 4nm$)[25].

Un exemple récent de cette catégorie se compose des molécules illustrées à la Figure , obtenues en 2012 par les laboratoires d'AstraZeneca par une approche basée sur des fragments[24]. Des études in vitro de certains acides carboxyliques prometteurs et des correspondants esters, ont montré des bonnes caractéristiques et ont également été caractérisées par des études de cristallographie aux rayons X (par exemple code PDB : 4AJP).

Malheureusement, ces molécules montrent une faible perméabilité cellulaire.

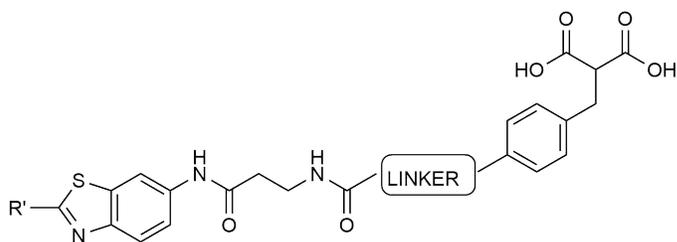


FIGURE 4.9 – Structure d'inhibiteurs bivalents d'AstraZeneca.

4.8.5 Galloflavine

En 2012[26], la galloflavine a été sélectionnée et synthétisée à la fin d'une étude de virtuel screening à l'université de Bologne. Il n'y a pas de sélectivité entre les isoformes H et M de la hLDH Avec de faibles valeurs d'inhibition micromolaires ($K_i = 5,46 \mu$ par LDH-5 isoforme et $K_i = 15,1 \mu$ pour LDH-1). A partir des études d'interactions, nous avons trouvé que la galloflavine inhibe les deux isoformes humaines LDH en liant préférentiellement l'enzyme libre, sans interaction avec le substrat ou le cofacteur NADH.

Des études in vitro subséquentes sur différentes cultures tumorales (carcinome hépatocellulaire et cancer du sein) ont donné des résultats prometteurs.

4.8.6 Complexes de cuivre (II) et de zinc (II)

Ces complexes ont été synthétisés et testés sur la LDH d'origine murin[27], où ils montraient une faible capacité inhibitrice, plus accentuée par rapport au dérivé complexé avec le

zinc. Des expériences subséquentes ont montré que ce dérivé est capable d'inhiber la LDH seulement dans le foie, les reins et le coeur, sans effet sur la rate, le cerveau et les tissus musculaires squelettiques. Au contraire, le dérivé complexe avec le cuivre montre une faible capacité d'inhiber l'activité de la LDH dans tous les tissus étudiés. Les tests effectués par le groupe de recherche suggèrent que ces dérivés sont capables d'inhiber sans discernement les 5 isoformes de la LDH.

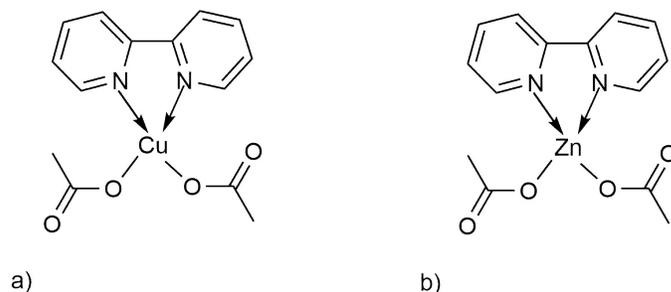


FIGURE 4.10 – Complexation du cuivre (a) et du zinc (b).

4.8.7 Quinoléine et 1,4-dihydro-4-quinolones

Au début des années 70, le squelette quinoléique a été utilisé pour synthétiser des molécules qui avaient la capacité d'inhiber les quatre principales déshydrogénases actives dans le métabolisme du glucose des cellules cancéreuses (lactate déshydrogénase, glutamate déshydrogénase, Glycéraldéhyde-3-phosphate déshydrogénase et malate déshydrogénase)[28].

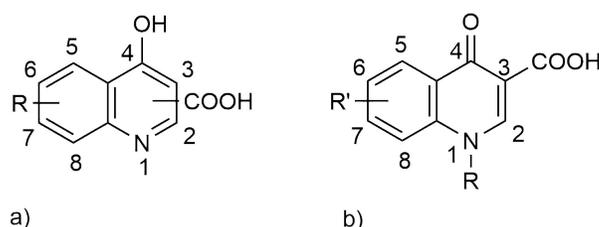


FIGURE 4.11 – (a) Quinoléine et (b) 1,4-dihydro-4-quinolones.

4.8.8 Les azoles

Cette classe d'inhibiteurs a été identifiée à la suite d'un criblage à haut débit pour les nouveaux médicaments antipaludiques, mais ont également montré qu'elles ont une bonne activité pour l'isoforme humaine de LDH[29].

Parmi les molécules ayant une activité micromolaire, il a été possible d'obtenir la structure cristallographie du LDH-A complexée avec un dérivé acide 4-hydroxy-1, 2, 5-oxadiazole-3-

carboxylique. Malheureusement, la structure simple de ces molécules de faible poids moléculaire les rend potentiellement capables d'interagir avec de nombreuses autres cibles biologiques.

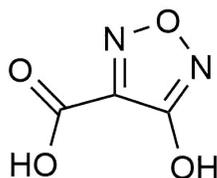


FIGURE 4.12 – Acide 4-hydroxy-1, 2, 5-oxadiazole-3-carboxylique.

4.8.9 Récentes découvertes de la société Genetech Inc

En 2013, une étude visant à identifier des nouveaux inhibiteurs du LDH à travers la base de données des composés de Roche et Genentech. A la fin d'étude, deux molécules inhibitrices ont été identifiées, qui ont fait l'objet d'une étude approfondie de la relation structure-activité et de la caractérisation cristallographique des deux composés les plus actifs.

- Dérivés de 2-amino-5-aryl-pyrazine[30] : (Code PDB :4M49) avec $IC_{50}=0.5 \mu$. Ils montrent une bonne sélectivité pour l'isoforme LDH-5 par rapport à la LDH-1 ou à la malate déshydrogénase. Les structures cristallographiques obtenues montrent que ces molécules interagissent avec les principaux résidus de la poche du substrat, avec une haute affinité en présence du cofacteur NADH. La Figure 4.13 illustre l'inhibiteur du code PDB 22Y.

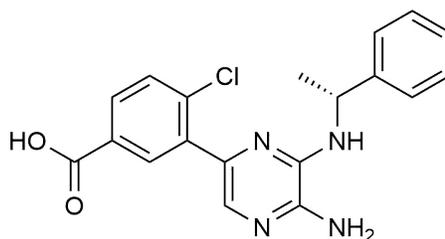


FIGURE 4.13 – Acide (R)-3-(5-amino-6-((1-phenylethyl)amino)pyrazin-2-yl)-4-chlorobenzoïque; code PDB : 22Y.

- Dérivés de 2-thio-6-oxo-1,6-dihydropyrimidines[31] : (code PDB : 4JNK). Montre une sélectivité semblable à l'autre classe, mais se lie à des résidus autres que ceux du site catalytique, (ça sera approfondi dans les chapitres suivants). La Figure 4.14 montre l'inhibiteur du code PDB ZHK.

Un autre article a ensuite été publié par le même groupe :

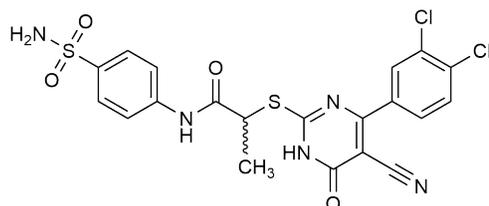


FIGURE 4.14 – 2-((5-cyano-4-(3,4-dichlorophenyl)-6-oxo-1,6-dihydropyrimidin-2-yl)thio)-N-(4-sulfamoylphenyl) propanamide; Code PDB : ZHK.

— Dérivés de 3-hydroxy-2-mercaptopyclohex-2-enones[32] : (code PDB : 4QO8). Sélectivité remarquable pour l'isoforme LDH-5 ($IC_{50} = 0,87 \mu$). La structure cristallographique montre qu'elle se lie à l'enzyme seulement en présence de NADH. Il contient un groupe énolique ionisable en milieu physiologique, ce que lui permet d'être capable d'imiter le groupe carboxylique du substrat, tandis que le groupe cétone forme des liaisons hydrogène avec deux résidus importants du site catalytique (His192 et Asn137). La Figure montre l'inhibiteur avec le code 36U.

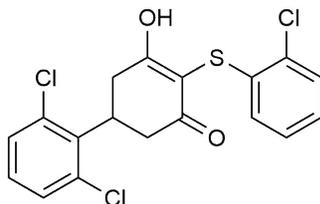


FIGURE 4.15 – 2',6'-dichloro-4-((2-chlorophenyl)thio)-5-hydroxy-1,6-dihydro-[1,1'-biphenyl]-3(2H)-one ; Code PDB : 36U.

4.9 Conclusion

Dans ce chapitre, on a montré l'importance du lactate déshydrogénase dans la voie de transformation de sucre en énergie. Plusieurs études ont mis en évidence l'implication directe de cette enzyme dans la genèse et la prolifération de certains types de tumeurs (effet Warburg), et constitue donc une cible attractive dans le domaine de l'oncologie. L'effet Warburg a été très étudié mais sa nature précise reste floue, ce qui freine l'émergence de projets ayant pour but d'explorer son potentiel thérapeutique, c'est pourquoi la chimie computationnelle est venue avec tous ses outils pour une compréhension complète et détaillée de cette enzyme au niveau des cellules (mécanisme réactionnelle), qui nous permettra d'identifier des nouveaux inhibiteurs de cette enzyme, c'est ce que nous recherchons dans les prochaines parties de notre étude.

Bibliographie

- [1] Philipp Mergenthaler, Ute Lindauer, Gerald A Dienel, and Andreas Meisel. Sugar for the brain : the role of glucose in physiological and pathological brain function. *Trends in neurosciences*, 36(10) :587–597, 2013.
- [2] David L Nelson, Michael M Cox, and Albert L Lehninger. Principles of biochemistry. *WH Freeman and Company, New York, fourth edition edition*, 1(1.1) :2, 2005.
- [3] Jeremy M Berg, Lubert Stryer, and John L Tymoczko. *Stryer Biochimie*. Springer-Verlag, 2015.
- [4] Virginia Zewe and Herbert J Fromm. Kinetic studies of rabbit muscle lactate dehydrogenase. *J Biol Chem*, 237 :1668–1675, 1962.
- [5] JH Wilkinson and Wendy A Withycombe. Organ specificity and lactate-dehydrogenase activity. some properties of human spermatozoal lactate dehydrogenase. *Biochemical Journal*, 97(3) :663, 1965.
- [6] Erwin Goldberg and Jerrold Lerum. Pregnancy suppression by an antiserum to the sperm specific lactate dehydrogenase. *Science*, 176(4035) :686–687, 1972.
- [7] David M Dawson, Theodore L Goodfriend, and Nathan O Kaplan. Lactic dehydrogenases : functions of the two types. *Science*, 143(3609) :929–933, 1964.
- [8] Takashi Kanno, Kayoko Sudo, Masato Maekawa, Yodshiro Nishimura, Minoru Ukita, and Katsuhiko Fukutake. Lactate dehydrogenase m-subunit deficiency : a new type of hereditary exertional myopathy. *Clinica chimica acta*, 173(1) :89–98, 1988.
- [9] Michael G Rossman, Anders Liljas, Carl-Ivar Brändén, and Leonard J Banaszak. 2 evolutionary and structural relationships among dehydrogenases. In *The enzymes*, volume 11, pages 61–102. Elsevier, 1975.
- [10] Douglas Hanahan and Robert A Weinberg. The hallmarks of cancer. *cell*, 100(1) :57–70, 2000.
- [11] Douglas Hanahan and Robert A Weinberg. Hallmarks of cancer : the next generation. *cell*, 144(5) :646–674, 2011.
- [12] Peter C Nowell. The clonal evolution of tumor cell populations. *Science*, 194(4260) :23–28, 1976.
- [13] Otto Warburg. The metabolism of carcinoma cells. *The Journal of Cancer Research*, 9(1) :148–163, 1925.

- [14] Pawel Swietach, Richard D Vaughan-Jones, and Adrian L Harris. Regulation of tumor pH and the role of carbonic anhydrase 9. *Cancer and Metastasis Reviews*, 26(2) :299–310, 2007.
- [15] Karin Fischer, Petra Hoffmann, Simon Voelkl, Norbert Meidenbauer, Julia Ammer, Matthias Edinger, Eva Gottfried, Sabine Schwarz, Gregor Rothe, Sabine Hoves, et al. Inhibitory effect of tumor cell-derived lactic acid on human t cells. *Blood*, 109(9) :3812–3819, 2007.
- [16] Robert A Gatenby and Robert J Gillies. Why do cancers have high aerobic glycolysis? *Nature Reviews Cancer*, 4(11) :891–899, 2004.
- [17] Martha Földi, Elmar Stickeler, Lidija Bau, Oliver Kretz, Dirk Watermann, Gerald Gitsch, Gian Kayser, Axel zur Hausen, and Johannes F Coy. Transketolase protein tktl1 overexpression : A potential biomarker and therapeutic target in breast cancer. *Oncology reports*, 17(4) :841–845, 2007.
- [18] Gregg L Semenza. Hypoxia-inducible factor 1 (hif-1) pathway. *Sci. Stke*, 2007(407) :cm8–cm8, 2007.
- [19] World Health Organization. *Global tuberculosis report 2013*. World Health Organization, 2013.
- [20] Manuel S Gomez, Robert C Piper, Lucy A Hunsaker, Robert E Royer, Lorraine M Deck, Michael T Makler, and David L Vander Jagt. Substrate and cofactor specificity and selective inhibition of lactate dehydrogenase from the malarial parasite p. falciparum. *Molecular and biochemical parasitology*, 90(1) :235–246, 1997.
- [21] Jerzy W Jaroszewski, Thorbjørn Strøm-Hansen, Steen Honoré Hansen, Ole Thastrup, and Helmer Kofod. On the botanical distribution of chiral forms of gossypol. *Planta medica*, 58(05) :454–458, 1992.
- [22] Seoung-ryoung Choi, Aaron B Beeler, Anupam Pradhan, E Blake Watkins, John M Rimoldi, Babu Tekwani, and Mitchell A Avery. Generation of oxamic acid libraries : antimalarials and inhibitors of plasmodium falciparum lactate dehydrogenase. *Journal of combinatorial chemistry*, 9(2) :292–300, 2007.
- [23] Carlotta Granchi, Sarabindu Roy, Chiara Giacomelli, Marco Macchia, Tiziano Tuccinardi, Adriano Martinelli, Mario Lanza, Laura Betti, Gino Giannaccini, Antonio Lucacchini, et al. Discovery of n-hydroxyindole-based inhibitors of human lactate dehydrogenase isoform a (ldh-a) as starvation agents against cancer cells. *Journal of medicinal chemistry*, 54(6) :1599–1612, 2011.

- [24] Richard A Ward, Claire Brassington, Alexander L Breeze, Alessandro Caputo, Susan Critchlow, Gareth Davies, Louise Goodwin, Giles Hassall, Ryan Greenwood, Geoffrey A Holdgate, et al. Design and synthesis of novel lactate dehydrogenase a inhibitors by fragment-based lead generation. *Journal of medicinal chemistry*, 55(7) :3285–3306, 2012.
- [25] Alexander B Kotlyar, Antonio Randazzo, Norman Honbo, Zhu-Qui Jin, Joel S Karliner, and Gary Cecchini. Cardioprotective activity of a novel and potent competitive inhibitor of lactate dehydrogenase. *FEBS letters*, 584(1) :159–165, 2010.
- [26] Marcella Manerba, Marina Vettrai, Luigi Fiume, Giuseppina Di Stefano, Andrea Sarti, Elisa Giacomini, Rosa Buonfiglio, Marinella Roberti, and Maurizio Recanatini. Galloflavin (cas 568-80-9) : a novel inhibitor of lactate dehydrogenase. *ChemMedChem*, 7(2) :311–317, 2012.
- [27] Raj Kumar Koiri, Surendra Kumar Trigun, Santosh Kumar Dubey, Santosh Singh, and Lallan Mishra. Metal cu (ii) and zn (ii) bipyridyls as inhibitors of lactate dehydrogenase. *Biometals*, 21(2) :117–126, 2008.
- [28] BR Baker and Ray R Bramhall. Irreversible enzyme inhibitors. 189. inhibition of some dehydrogenases by derivatives of 4-hydroxyquinoline-2-and-3-carboxylic acids. *Journal of medicinal chemistry*, 15(3) :230–233, 1972.
- [29] Angus Cameron, Jon Read, Rebecca Tranter, Victoria J Winter, Richard B Sessions, R Leo Brady, Livia Vivas, Anna Easton, Howard Kendrick, Simon L Croft, et al. Identification and activity of a series ofazole-based compounds with lactate dehydrogenase-directed anti-malarial activity. *Journal of Biological Chemistry*, 279(30) :31429–31439, 2004.
- [30] Benjamin P Fauber, Peter S Dragovich, Jinhua Chen, Laura B Corson, Charles Z Ding, Charles Eigenbrot, Anthony M Giannetti, Thomas Hunsaker, Sharada Labadie, Yichin Liu, et al. Identification of 2-amino-5-aryl-pyrazines as inhibitors of human lactate dehydrogenase. *Bioorganic & medicinal chemistry letters*, 23(20) :5533–5539, 2013.
- [31] Peter S Dragovich, Benjamin P Fauber, Laura B Corson, Charles Z Ding, Charles Eigenbrot, HongXiu Ge, Anthony M Giannetti, Thomas Hunsaker, Sharada Labadie, Yichin Liu, et al. Identification of substituted 2-thio-6-oxo-1, 6-dihydropyrimidines as inhibitors of human lactate dehydrogenase. *Bioorganic & medicinal chemistry letters*, 23(11) :3186–3194, 2013.
- [32] Peter S Dragovich, Benjamin P Fauber, Jason Boggs, Jinhua Chen, Laura B Corson, Charles Z Ding, Charles Eigenbrot, HongXiu Ge, Anthony M Giannetti, Thomas Hunsaker, et al. Identification of substituted 3-hydroxy-2-mercaptocyclohex-2-enones as potent

inhibitors of human lactate dehydrogenase. *Bioorganic & medicinal chemistry letters*, 24(16) :3764–3771, 2014.

Chapitre 5

Aspect théorique sur la chimie Computationnelle et les logiciels utilisés

Sommaire

5.1 Introduction	103
5.2 Programmes de visualisation	104
5.2.1 UCSF CHIMERA	104
5.2.2 MAESTRO	105
5.2.3 LIGANDSCOUT	106
5.3 Programmes de Docking	108
5.3.1 Définition du Docking	108
5.3.2 Les algorithmes génétiques	109
5.3.3 AUTODOCK	110
5.3.4 AutoDock VINA	111
5.3.5 GLIDE	112
5.3.6 GOLD	114
5.3.7 DOCK6	115
5.3.8 PLANTS	115
5.4 La simulation dynamique moléculaire	116
5.4.1 Principes théoriques	116
5.4.2 Calcul de dynamique moléculaire	117
5.4.3 Programme pour la dynamique moléculaire	117
5.5 Conclusion	119

5.1 Introduction

Dans les années 1980, des technologies de criblage à haut débit (High-Throughput Screening ou HTS) ont été développées dans le but de trouver de nouvelles molécules ayant des activités potentielles sur des cibles d'intérêt thérapeutiques majeures. Cependant, le nombre

de cibles identifiées et le nombre de molécules chimiques à tester dans les banques de données augmentent continuellement, ce qui a pour conséquence une diminution de l'efficacité de la méthode, accompagnée d'une augmentation des coûts. Le faible taux d'identification de hits par rapport aux coûts de la méthode[1] a conduit à la fin des années 1990 au développement des méthodes informatiques permettant de ne sélectionner qu'une partie de la base de données enrichie en molécules susceptibles d'interagir avec la cible protéique. Ces méthodes ont été dans un premier temps basées sur la structure des ligands ayant une activité biologique (approches de type pharmacophore et QSAR), en absence d'un nombre significatif de structures protéiques dans la PDB. Avec l'augmentation continue du nombre de ces structures, l'approche de type Docking a été de plus en plus utilisée. L'application de ces techniques sur des chimiothèques de taille est appelé criblage virtuel (virtual screening).

La recherche en biologie ne peut, actuellement, se passer des outils informatiques pour traiter le flot de données produites et optimiser ses avancées. La visualisation en trois dimensions de molécules constitue une partie importante de la stéréochimie. Depuis maintenant près de deux siècles, les chimistes à la suite des cristallographes[2], ont rivalisé d'imagination afin de représenter les molécules et cristaux pour les besoins de l'enseignement et de la recherche. Les progrès accomplis en informatique depuis une vingtaine d'années permettent à tout chimiste de visualiser des molécules en s'aidant d'un micro-ordinateur grâce à des programmes dont certains appartiennent au domaine public ou sont libres d'utilisation. Loin de se limiter à une simple représentation statique, ces programmes permettent de déplacer la molécule en trois dimensions, d'afficher et de calculer les principaux paramètres moléculaires ; distances interatomiques, angles entre les liaisons, rayons de Van der Waals, interaction moléculaire etc.

5.2 Programmes de visualisation

5.2.1 UCSF CHIMERA

Le software Chimera[3], développé par l'Université de Californie, San Francisco, est un programme multifonctionnel principalement utilisé pour la visualisation graphique et l'analyse de petites molécules, protéines et acides nucléiques. Chimera peut être utilisé par exemple pour visualiser les alignements des séquences, les résultats du Docking, les trajectoires obtenues à partir d'études de simulation dynamique moléculaire, ou pour calculer les liaisons hydrogène, les interactions hydrophobes et visualiser les surfaces. Une fonction utile de Chimera est ViewDock, utilisé pour l'analyse de multi-fichiers en format MOL2, comme les résultats d'un docking moléculaire pour un très grand nombre de molécules. Pour effectuer ce type d'analyse, vous devez :

- Ouvrir le fichier de protéine d'intérêt ou le complexe de ligand-protéine du menu *File/open* ;
- Sélectionner la fonction ViewDock de menu *Tools/Surface-Binding Analysis/ViewDock*

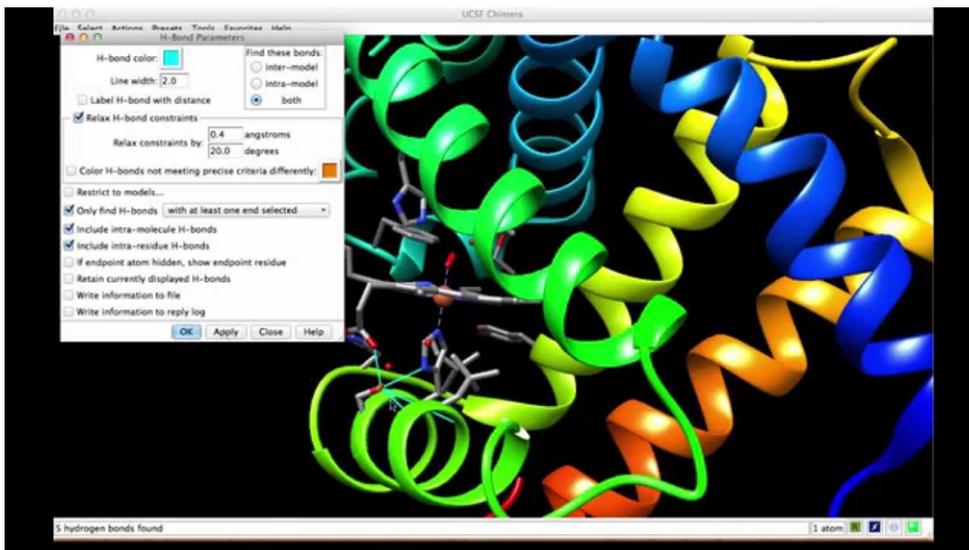


FIGURE 5.1 – Interface du logiciel Chimera.

et ouvrir les résultats en format MOL2 ;

- De la fenêtre de *ViewDock*, si ouverte, il sera donc possible d’afficher les ligands à la fois dans la poche du récepteur ;

Les molécules, initialement marquées comme toutes visibles, peuvent ensuite être masquées par l’option *Purged* ou directement supprimées avec l’option *Deleted*.

Pour faciliter l’analyse *ViewDock*, on peut dire qu’il peut également être utilisé pour calculer les liaisons hydrogène totales de chaque complexe ligand-protéine en utilisant l’option *FindH-Bond*.

5.2.2 MAESTRO

Maestro[4] est une plate-forme unifiée pour l’utilisation des outils de Schrodinger, et permet ainsi une grande variété d’utilisations possibles. Parmi les caractéristiques du Maestro, nous trouvons : la génération de modèles et la construction de molécules de toutes sortes ; possibilité de visualisation de petites molécules à des complexes biomoléculaires, à la fois dans des modèles 3D et 2D et analyse structurale quantitative. Maestro comprend des instruments pour la mesure précise des caractéristiques structurales des molécules, ainsi que la possibilité de superposition des structures pour la comparaison ; gestion et analyse des données, calcul des propriétés moléculaires. En outre, les principaux programmes de la suite Schrodinger sont intégrés dans Maestro, y compris les plus pertinents pour mon travail de thèse :

- GLIDE, pour le docking moléculaire ; (approfondi en section III.5) ;

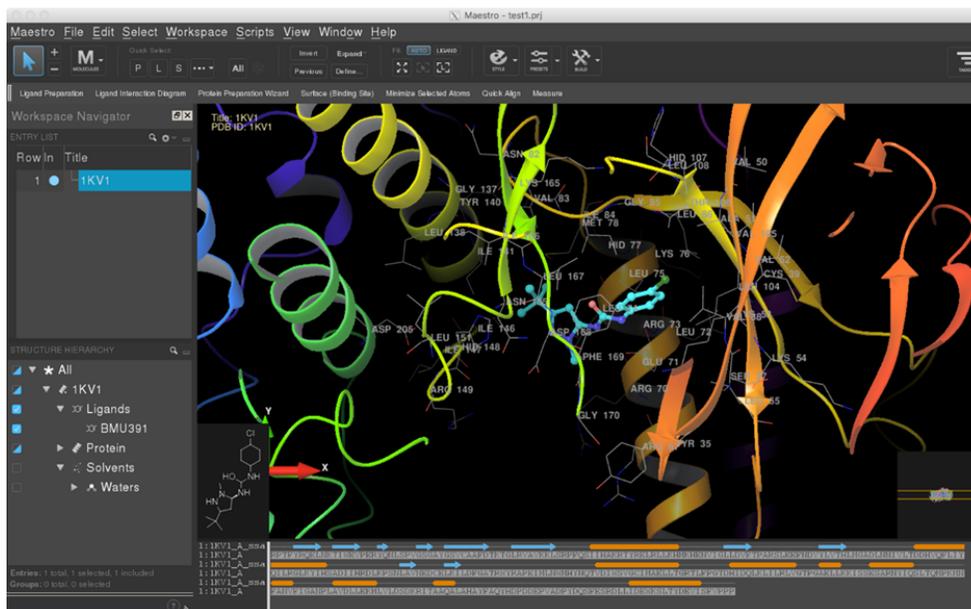


FIGURE 5.2 – Interface du logiciel Maestro.

— Ligprep, pour générer des structures moléculaires d'énergie 3D minimales. LigPrep applique également des règles sophistiquées pour corriger les structures de Lewis et éliminer les erreurs possibles. A partir de molécules 2D à travers ligprep, il est possible de :

1. Ajouter des hydrogènes ;
2. Retirer les contre-ions ou les molécules d'eau ;
3. Ioniser les molécules sur la base des conditions de pH spécifiées (dans mon travail de thèse, le pH a été fixé à 7,0) ;
4. Générer des tautomères ;
5. Supprimer les structures endommagées ;
6. Optimiser la géométrie tridimensionnelle des structures ;
7. Définir le nombre d'états de tautomères / d'ionisation pour chaque composé ; (la valeur d'une structure unique pour chaque molécule a été définie).

5.2.3 LIGANDSCOUT

LigandScout[5], il s'agit d'une plate-forme intégrée pour la conception moléculaire et des procédures de virtuel screening (Criblage Virtuel), avec une interface graphique simple et intuitive. Il permet de tirer parti des principales stratégies de virtuel screening, c'est-à-dire la construction de modèles à base de ligands ou de récepteurs, la génération des conformères pour la création de base des données pour le virtuel screening. Une fonction a également été

implémentée pour effectuer des calculs d'amarrage avec l'Autodock et l'Autodock Vina.

LigandScout présente quatre écrans différents :

- *Receptor based* : pour la visualisation des complexes ligand-protéine et la création d'un récepteur à base de pharmacophore. Dans cet écran, vous pouvez également charger des bibliothèques de molécules qui peuvent être ancrées (Docking) dans le site actif ou alignées avec une structure qui possède un modèle pharmacophore. Il est également possible de calculer un score des molécules par rapport à un pharmacophore sans les aligner ;
- *Ligand Based* : Pour l'analyse des molécules et la création de modèles pharmacophore de type à base de ligands ;
- *Alignement* : pour comparer les structures et les modèles pharmacophore ;
- *Virtual Screening* : Pour importer ou créer des bases de données et exécuter le filtrage.

Un modèle pharmacophore à base de récepteurs peut être créé en procédant comme suit :

1. Charger un complexe protéine-ligand en tapant le code pdb pour télécharger la structure correspondante à partir de la Banque de données des protéines ou en ouvrant un fichier PDB ou MOL2. Les ligands sont automatiquement reconnus et interprétés ;
2. Cliquer sur le ligand pour agrandir le site de liaison, dans lequel l'interprétation du ligand peut être révisée et facilement corrigée (par exemple, un type de liaison est modifié en sélectionnant la liaison et en tapant un numéro pour la nouvelle Valence) ;
3. Créer un modèle pharmacophore avec des fonctionnalités complémentaires (*Features*) entre macromolécule et ligand en appuyant sur l'icône appropriée ou en utilisant les touches Ctrl + F9 ;
4. Optimiser le modèle pharmacophore en changeant, en ajoutant ou en supprimant les features (les interactions du modèle) proposées par le programme, en choisissant dans le menu *pharmacophores/Features*.

Ainsi les modèles pharmacophores créés peuvent être affichés sur l'écran d'alignement, où ils peuvent être alignés pour générer un pharmacophore qui contient des caractéristiques communes ou une somme de plus de pharmacophore. Pour effectuer une procédure de virtuel screening avec LigandScout, vous devez préparer la base de données en format LDB, qui contient toutes les informations sur les possibles conformations structurelles des molécules. Le screening peut être effectué de deux manières :

1. De l'interface graphique (plus lent) ;
2. De la ligne de commande en appelant le programme Iscreen (plus rapide).

À la fin de la procédure, le fichier de sortie au format SDF peut être ouvert dans l'écran *Receptor-Based* ; Pour chaque molécule sélectionnée, vous pouvez voir la valeur du score attribué au modèle pharmacophore, un tableau montrant les features choisies pour chaque molécule, et les autres caractéristiques calculées.



FIGURE 5.3 – Interface du logiciel Ligandscout.

5.3 Programmes de Docking

5.3.1 Définition du Docking

Le Docking est une opération utilisée pour explorer la conformation de liaison entre un complexe de deux molécules[6]. Une procédure de Docking est utilisée comme guide pour identifier l'orientation la plus probable pour un ligand avec une macromolécule (Protéine). Le ligand garanti la possibilité de subir de petits changements de conformation pour éviter la répulsion stérique et de produire des interactions favorables avec le récepteur.

Dans une procédure d'amarrage (Docking) automatisée, avec la structure 3D d'une protéine, le programme est capable d'effectuer le Docking, c'est-à-dire, adapter un ligand à une cavité dans une conformation plausible. Au départ, vous aurez une orientation approximative, qui sera ensuite optimisée pour tirer le meilleur parti des interactions favorables[7]. Le calcul se termine lorsque vous obtenez un minimum d'énergie pour le complexe.

Une simulation du Docking comprend essentiellement deux étapes : le docking proprement dit et le scoring (Figure 5.4).

- La première (le Docking) est l'étape de sélection, consistant à placer le ligand dans le site actif de la protéine et à échantillonner les conformations, positions et orientations (poses) possibles, en ne retenant que celles qui représentent les modes d'interactions les plus favorables.

- La deuxième (le scoring) est l'étape de classement, qui consiste à évaluer l'affinité entre le ligand et la protéine et de donner un score aux poses obtenues lors de la phase de docking. Ce score permettra de retenir la meilleure pose parmi toutes celles proposées[6].

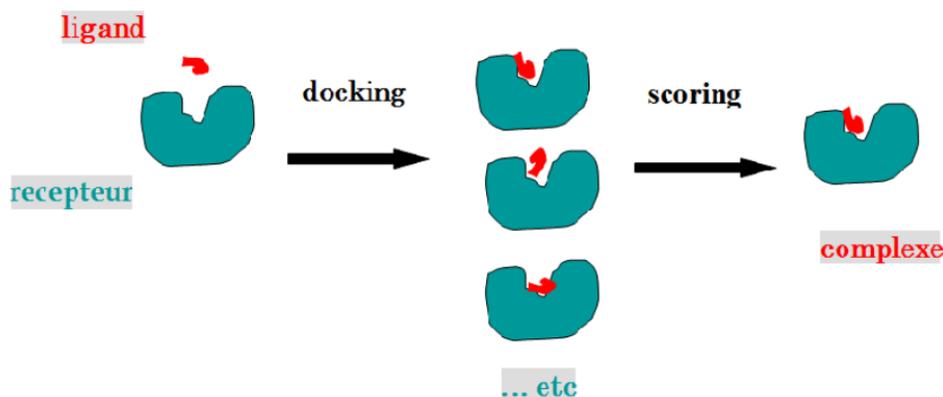


FIGURE 5.4 – Principe général d'un programme de Docking.

Les types de programmes de Docking automatisés peuvent être répartis-en :

- **Exhaustive** : un grand nombre de conformations sont générés pour chaque ligand, et tous sont évalués (ex : Fred);
- **Stochastique** : seules quelques solutions sont évaluées ; les résultats peuvent ne pas être reproductibles.

5.3.2 Les algorithmes génétiques

Dans ce type de calcul, la conformation d'un ligand est calculée en tenant compte des règles de la génétique classique. Un algorithme génétique est caractérisé par l'évolution d'une population de solutions possibles (conformations) à travers des opérateurs génétiques tels que des mutations, des recombinaisons et des translocations pour optimiser une fonction de fitness prédéterminée. Pour chaque conformation initiale, un code génétique est attribué qui permet l'attribution d'un score basé sur une fonction de notation, et ce code peut subir des mutations pour générer de nouvelles solutions.

Les phases d'un calcul par algorithme génétique sont :

1. **Phase d'initialisation** : Générer une population de solutions possibles ;
2. **Phase de sélection** : À chaque cycle d'algorithme, un pourcentage de la population existante est disséqué pour générer une nouvelle génération. Les solutions individuelles sont sélectionnées à travers un processus basé sur la forme physique (basé sur les énergies d'interaction des solutions individuelles) ;

- Lecture :** Cette étape génère une deuxième génération de solutions à partir des solutions sélectionnées. Pour chaque nouvelle solution, une paire de solutions «parentes» est sélectionnée dans le groupe précédemment sélectionné. En produisant une solution «fille» à travers les méthodes de mutation génétique précédemment nommées, il est possible d'obtenir une nouvelle génération. Le processus est répété pour générer de nouvelles générations.

Des exemples de programmes qui utilisent des algorithmes génétiques sont AutoDock et GOLD.

5.3.3 AUTODOCK

AutoDock[8] est une suite d'outils de Docking automatisés, qui comprend deux programmes principaux : Autodock 4 et Vina Autodock. Il existe une interface graphique (AutoDockTools ou ADT), mais vous pouvez également lancer les deux programmes de l'interpréteur de commandes (The shell). L'algorithme d'exploitation d'Autodock est de type génétique, nommé

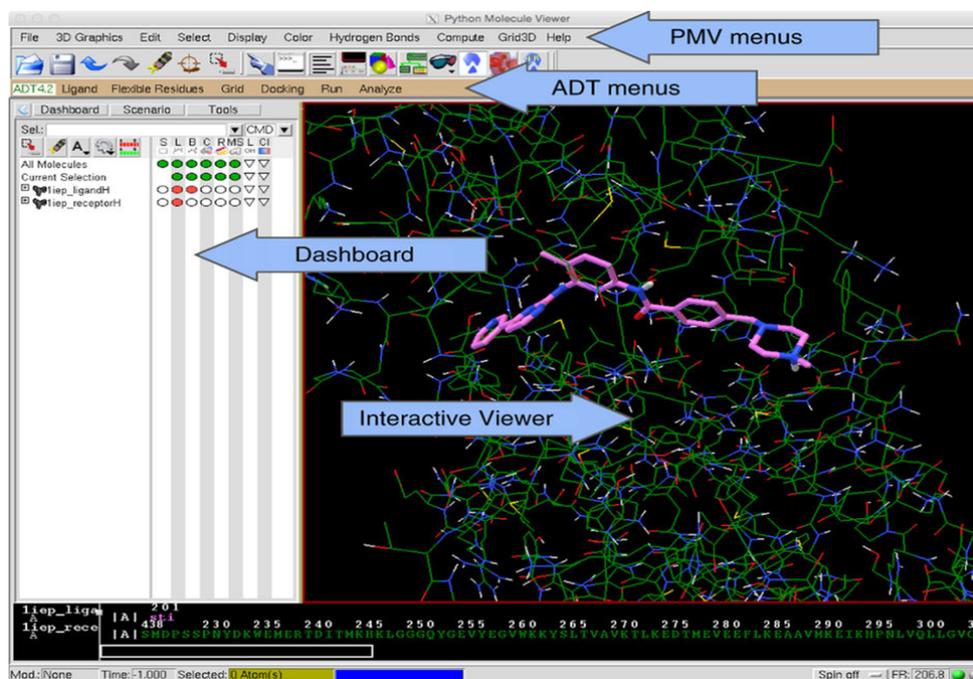


FIGURE 5.5 – Interface d'Autodock 4.

Lamarckian Genetic Algorithm. D'une manière similaire à la théorie d'évolution naturelle de Lamarck, chaque fois qu'une nouvelle population est générée, il sera possible d'optimiser les solutions en considérant «l'énergie autour»; De cette façon, le «génotype», c'est-à-dire la chaîne binaire à laquelle chaque ligand correspond sera influencée par des facteurs externes. En plus de l'Algorithme Génétique Lamarckien (LGA), des algorithmes de recuit transgénique sont également disponibles.

Le point de départ est une population initiale à partir de laquelle l'algorithme sélectionne un ensemble de solutions basées sur la fonction de notation, qui donnera lieu à une nouvelle population de solutions enfants, à partir de laquelle un cycle de génération supplémentaire va commencer. Les populations de solutions sont obtenues par des opérateurs génétiques (mutations, croisements et migrations) qui imitent les produits biologiques. Les degrés de liberté sont codés dans des gènes ou des chaînes binaires, et aux gènes et chromosomes sont attribués une valeur basée sur l'aptitude de la fonction de notation. Les opérations de mutation provoquent des changements dans la valeur d'un gène, tandis que le crossover déplace un ensemble de gènes d'un chromosome «parent» à un autre ; La migration déplace plutôt les gènes individuels d'une sous-population à une autre.

Autodock 4 utilise une fonction de notation d'énergie libre pour évaluer les conformations durant le Docking. Le champ de force est semi empirique, il a été paramétré avec un grand nombre de complexes inhibiteurs-protéines pour lesquels la structure et la constante d'inhibition (K_i), sont connus. L'interaction entre le ligand et la protéine est évaluée en deux phases. La première évalue la variation de l'énergie intramoléculaire dans la transition de la conformation de ligand libre à celle liée ; La deuxième phase tient compte de la variation de l'énergie intermoléculaire du complexe protéine-ligand dans la conformation liée.

5.3.4 AutoDock VINA

Vina[9] représente la nouvelle version d'AutoDock, et présente en fait de nombreuses similitudes avec son prédécesseur mais, en même temps, de nombreuses différences ; Une différence importante est la rapidité de calcul, puisque Vina est très peu coûteuse à cet égard ; Une autre différence fondamentale est que lors du calcul de la grille, Vina calcule automatiquement les cartes de grille (*grid maps*), en comparaison avec Autodock. C'est un grand avantage en termes de facilité et de rapidité d'exécution. De plus, les fonctions de scoring et les algorithmes utilisés dans ce type d'analyse sont complètement différents de ceux de son prédécesseur[10][11], ce qui conduit à considérer Vina comme un logiciel autonome. En outre, ce programme, qui utilise un algorithme génétique pour les calculs de Docking, nécessite des fichiers avec une extension *PDBQT*, contenant ainsi les paramètres de coordonnées déjà observés pour AUTODOCK. Ils seront créés de la même manière que la suite précédente. La grille (The grid) requise pour le Docking est beaucoup plus simple : grâce à une commande shell, vous obtiendrez un fichier modifiable appelé config.txt qui contiendra toutes les informations et les paramètres nécessaires pour créer une grille appropriée.

5.3.5 GLIDE

Grid-based Ligand Docking with Energetics (GLIDE)[12]. Il s'agit d'un programme développé par la société Schrödinger qui utilise une série de filtres hiérarchiques pour identifier les conformations possibles d'un complexe ligand-protéine. La forme et les propriétés du récepteur sont représentées par une grille à travers différents types de champs qui garantissent une précision progressivement plus élevée dans le calcul de la notation des poses de ligand. Chaque ligand est subdivisé en une région de noyau et un certain nombre de groupes de roulement. Chaque groupe est attaché au noyau par une liaison roulante, mais ne contient aucune liaison roulante supplémentaire. Pendant la génération de la conformation, chaque noyau est représenté par une série de conformations de base (typiquement 500), dont le nombre dépend du nombre de liaisons de rotation et d'autres paramètres limitant géométriquement. Pour chaque conformation, une recherche approfondie des orientations possibles au sein du site protéique actif est réalisée.

GLIDE vous permet de choisir entre 3 fonctions de score :

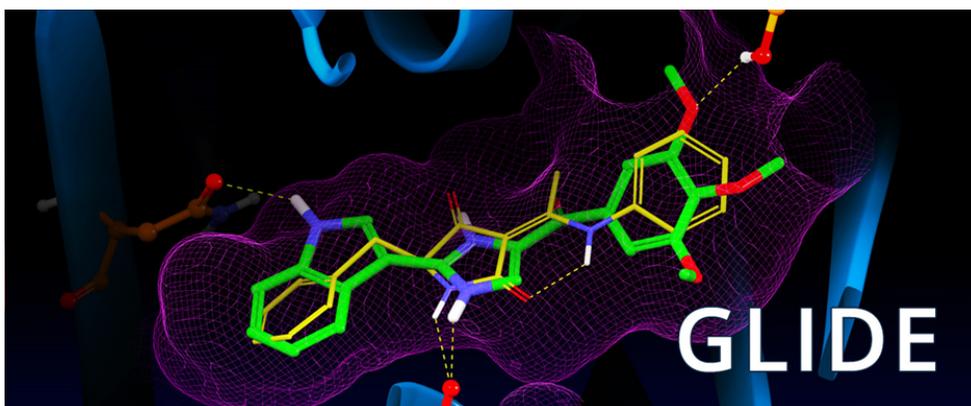


FIGURE 5.6 – Interface du logiciel Glide.

- **HTVS (high-throughput virtual screening)** : extrêmement rapide mais peu fiable. Convient pour le screening préliminaire de bases de données à grande échelle ;
- **SP (standard precision)**[12] : basée sur la fonction de score *ChemScore*. C'est une fonction de score qui cherche à diminuer les faux négatifs et est particulièrement adaptée au criblage de bases de données importantes. Elle est plus permissive que la seconde fonction de score de Glide (XP). La fonction *GlideScore SP* est différente par rapport à *ChemScore* de plusieurs façons :
 1. Le terme correspondant à l'énergie des liaisons hydrogènes est divisé en trois parties, en fonction de la charge des partenaires (neutre-neutre, neutre-chargé et chargé-chargé). Ces trois composantes sont pondérées par des coefficients qui rendent compte

de la stabilité de la liaison en fonction de la charge des partenaires, la première étant plus stabilisante que la dernière ;

2. Le terme correspondant aux interactions métal/ligand est modifié de telle sorte que seuls les atomes anioniques accepteurs soient pris en compte et que seule l'interaction métal/ligand la plus forte soit prise en compte lorsque plusieurs sont possibles. La charge du métal dans la protéine seule est calculée et conditionne la préférence pour les atomes anioniques accepteurs (prise en compte seulement lorsque le métal est chargé positivement) ;
 3. Un terme rendant compte des interactions de Coulomb et de van der Waals est ajouté ;
 4. Un terme de solvation est également ajouté. Il rend compte des interactions de différents groupements chimiques du ligand avec des molécules d'eau explicitement dockées dans le site actif après le ligand. Cependant, son influence est limitée par un poids de terme assez faible, dans un souci de vitesse et d'applicabilité de la méthode à des criblages de bases de données importantes ;
- **XP (*extra precision*)**[13] : Elle est plus contraignante et pénalise plus fortement les positions de Docking en désaccord avec les principes physico-chimiques comme l'orientation par rapport au solvant de groupements chargés ou fortement polaires. Cette fonction de score a été développée pour diminuer au maximum les faux positifs sur un faible jeu de données, typiquement pour améliorer l'activité des molécules à visée thérapeutique. L'ensemble des termes de cette fonction de score est calculé d'une manière minutieuse, ce qui fait qu'elle n'est pas adaptée pour cribler des bases de données importantes.

Cette fonction de score est également basée sur *ChemScore*, dont elle reprend les éléments principaux en les modifiant, en particulier le calcul des interactions entre atomes hydrophobes. Des éléments de *GlideScore SP* sont également repris et modifiés pour augmenter leur précision, notamment le calcul des trois composantes des interactions des liaisons hydrogènes, mais également le calcul de l'énergie de solvation. D'autres termes sont enfin ajoutés, comme l'énergie d'interaction due à l'empilement π - π ou π cation, et un terme pénalisant qui rend compte de l'énergie de torsion du ligand.

Toute la procédure de Docking a été réalisée à l'aide de l'interface graphique implémentée dans le programme Maestro. Pour effectuer un calcul Docking avec Glide, la première étape consiste à sélectionner les "site points", c'est-à-dire les points égaux à une distance de 2 Å appartenant à la grille de site actif, et qui seront utilisés pour le placement des atomes. Dans une deuxième étape, les conformations qui ont trop de répulsions stériques avec le récepteur sont écartées, afin de sélectionner les orientations possibles du diamètre du ligand (compris comme la liaison reliant les deux atomes les plus éloignés entre eux). À ce stade, les interactions de tous les atomes capables de fournir des liaisons hydrogène sont évaluées, et un score leur est attribué ; si c'est assez bon alors un score de toutes les interactions avec le récepteur est donné.

Pour ce faire, on utilise la fonction de *Chemscore*, qui reconnaît les interactions hydrogène et hydrophobes en tant que des interactions favorables, tout en pénalisant la répulsion de la Stérique.

Seul un petit nombre de poses est capable de passer à la troisième étape, où la minimisation de l'énergie est appliquée sur la base des grilles électrostatiques, des interactions van der Waals, et de force-field OPLS-AA. Enfin, les poses minimisées sont réévaluées à l'aide de *GlideScore*, une fonction de notation basée sur *ChemScore* mais qui diffère pour certains termes, tels que le calcul des déséquilibres électrostatiques. Le choix de la meilleure pose pour chaque ligand est donné par la fonction *Emodel*, qui combine la partition de la grille d'énergie, celle de *GlideScore*, et l'énergie interne de la molécule.

5.3.6 GOLD

Genetic Optimisation for Ligand Docking (GOLD)[14] C'est un algorithme génétique pour le placement de ligands flexibles dans un site de liaison avec une protéine, et qui contient quatre différentes fonctions de forme physique :

- **GoldScore** ; La fonction de score *GoldScore* est la fonction de score historique de GOLD. Pour évaluer les interactions entre le ligand et les atomes du site actif de la protéine, cette fonction de score prend en compte l'énergie des liaisons hydrogène, l'énergie électrostatique et de van derWaals, les interactions éventuelles avec des métaux et une pénalité en fonction de la torsion du ligand.
- **Chemscore** ; La fonction de score *ChemScore* a été introduite en 2003 dans GOLD. Cette fonction de score est dérivée de l'adaptation par Baxter et al[15] de la fonction de score développée par Eldridge et al[16] obtenue empiriquement par les analyses de multiples régressions linéaires effectuées sur un ensemble de 82 complexes protéine-ligand issus de la PDB. Cette fonction de score prend en compte : les liaisons hydrogènes, les liaisons avec les métaux, les contacts entre atomes lipophiles et les conflits entre atomes non-impliqués dans une liaison hydrogène. Elle intègre également une énergie de torsion interne au ligand qui dépend de sa conformation. Une valeur d'énergie covalente peut également être utilisée dans le cas où le ligand se lie de cette façon à la protéine. Cette fonction de score a l'avantage d'être jusqu'à trois fois plus rapide que *GoldScore* en conservant une précision comparable.
- **Astex Statistical Potential (ASP)**[17] ; La fonction de score *ASP* est de type *knowledge-based*, elle est comparable aux fonctions du même type comme par exemple *PMF*[18] et *DrugScore*[19]. C'est une fonction de score provenant de l'analyse statistique des distances entre paires d'atomes entre protéine et ligand qui favorisent les interactions. Ces distances sont issues d'une base de données de complexes de protéine-ligand.

- **Piecewise Linear Potential (CHEMPLP)**[20]; Elle est de type empirique et utilise le terme énergétique des liaisons hydrogène de la fonction *ChemScore*, ainsi que plusieurs potentiels linéaires pour rendre compte des interactions de type van der Waals et des termes de répulsion. C'est la fonction la plus rapide, supérieure à *ChemScore* et jusqu'à 5 fois plus rapide que *GoldScore*. Elle est en outre généralement plus précise que les autres méthodes pour le docking et le criblage virtuel. C'est la fonction de score utilisée par défaut dans les dernières versions du logiciel.

Les configurations de base de GLOD peuvent être modifiées par l'utilisateur en fonction des besoins du projet. GOLD peut en effet prendre en compte la flexibilité du récepteur, et vous permet de choisir entre différentes vitesses de calcul. Il est également possible d'appliquer une grande variété de contraintes pour garantir par exemple, que les liaisons hydrogène clés sont maintenues, ou pour orienter les résultats de l'amarrage vers un modèle de liaison connu.

5.3.7 DOCK6

DOCK6[21] est un programme développé par l'Université de Californie à San Francisco, qui utilise un algorithme stochastique qui prend le nom d'anchor-and-growth ou anchor-first (premier ancrage). L'algorithme de dock génère automatiquement autant d'orientations et de conformations que possible pour un ligand dans une poche du récepteur. La forme de la poche du récepteur est décrite par des sphères, et le centre de chaque sphère est considéré comme un point potentiel du Docking pour un atome du ligand. En utilisant la stratégie anchor-first, on prend en compte la flexibilité du ligand : d'abord, DOCK détecte les plus grandes sous-structures rigides des ligands, qui sont orientées de manière rigide vers le site actif au centre des sphères. Les différents «ancrages» sont évalués et classés en fonction de leur contribution énergétique à travers une grille de notation, et par la suite les portions flexibles du ligand sont construites autour des conformations les plus probables de la structure rigide. Le suite de programmes comprend :

- **sphgen** : Pour l'identification du site actif, et génération des sphères de Docking ;
- **grid** : Pour générer la grille (The grid) de scoring ;
- **dock6** : Pour effectuer le calculer de Docking.

5.3.8 PLANTS

Protein-Ligand ANT System (PLANTS)[22] un logiciel de Docking développé en 2006. L'algorithme d'amarrage PLANTS est basé sur une classe d'algorithmes stochastiques pour l'optimisation appelée Ant Colony Optimization (ACO), inspiré par le comportement des fourmis à trouver le chemin le plus court entre elles, et la bonne source de nourriture. Les fourmis utilisent un moyen de communication indirecte sous la forme de traces de phéromones qui

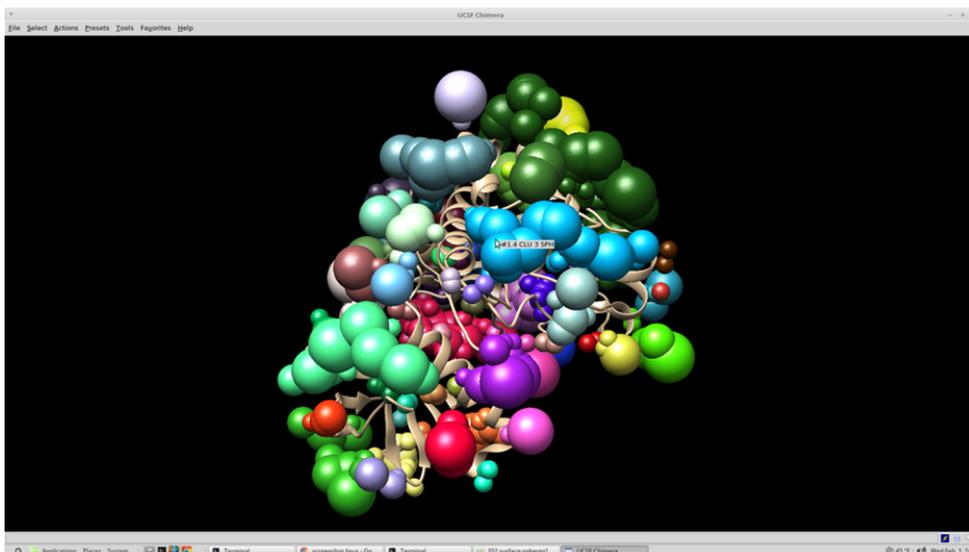


FIGURE 5.7 – Interface du logiciel Dock6.

marquent des voies entre la fourmi et la source de la nourriture. Dans le cas d'un amarrage de ligand-protéine, une colonie de fourmis artificielles est utilisée pour trouver la conformation d'énergie la plus basse du ligand dans son site de liaison. Les fourmis virtuelles sont utilisées pour imiter le comportement des vraies fourmis et marquer les conformations à basse énergie d'un ligand avec des traces de phéromones artificielles. Cette information est ensuite modifiée en itérations pour générer des configurations à haute probabilité et faible énergie.

5.4 La simulation dynamique moléculaire

5.4.1 Principes théoriques

La dynamique moléculaire est une approche de simulation qui prend en compte les interactions entre atomes et molécules pendant une période de temps, à une certaine température, permettant ainsi une évaluation des mouvements atomiques[23]. On se trouve alors en présence d'un assemblage de masses, de ressorts et de vitesses et donc le problème peut être résolu en intégrant la seconde équation de mouvement de Newton.

$$\vec{F}_i = m_i \vec{a}_i = m_i \frac{d^2 r_i(t)}{dt^2} \quad (5.1)$$

Avec :

\vec{F}_i : Vecteur force agissant sur l'atome i ;

m_i : Masse de l'atome i ;

\vec{a}_i : Vecteur accélération de l'atome i ;

Les vitesses sont appliquées de façon aléatoire sur les atomes (sphères) selon la température choisie pour l'expérience. En utilisant la DM (dynamique moléculaire), on peut étudier la flexibilité ou la rigidité des biomolécules, mesurer les interactions intermoléculaires entre la protéine et les ligands ou d'autres biomolécules, calculer l'énergie libre ou bien étudier l'effet du solvant sur la structure des biomolécules. Parmi les algorithmes intégrateurs les plus souvent implémentés dans les logiciels se trouvent les algorithmes de Verlet et de Verlet-Leapfrog[24].

5.4.2 Calcul de dynamique moléculaire

Une simulation de dynamique moléculaire se réalise généralement en quatre étapes :

- Une étape de minimisation de la structure initiale destinée à éliminer les contacts stériques ;
- Une étape dite de thermalisation au cours de laquelle les vitesses des atomes sont augmentées progressivement afin d'atteindre la température finale choisie ;
- Une étape d'équilibration pendant laquelle les vitesses ne sont plus modifiées. L'énergie cinétique se répartit sur toute la molécule afin d'atteindre son état d'équilibre ;
- Enfin, pendant la dernière étape, appelée dynamique productive, les coordonnées et les vitesses sont sauvegardées pour une analyse de la dynamique.

Deux points importants doivent être pris en compte :

La température d'une molécule ou agitation thermique correspond en réalité à l'énergie cinétique des atomes de la molécule puisque :

$$E_{cintique} = \frac{1}{2}m_i v_i^2 = \frac{3}{2}NK_bT \quad (5.2)$$

Avec :

v_i^2 : Valeur de la moyenne carrée de la vitesse ;

N : Nombre d'atomes du système ;

K_b : Constante de Boltzmann ;

T : Température absolue.

5.4.3 Programme pour la dynamique moléculaire

5.4.3.1 Introduction à Amber

Des études de dynamique moléculaire ont été menées à l'aide du logiciel AMBER14[25]. Il s'agit en fait d'une suite de programmes qui permettent d'effectuer des simulations de dynamique moléculaire, notamment sur des biomolécules. La suite AMBER comprend deux parties :

- AmberTools, Un ensemble de programmes pour la finition, tels que Sander, et pour la simulation PMEMD ;
- AMBER 14 (2015).

Dans cette suite, on peut distinguer trois catégories de programmes :

1. Programmes de préparation :

- **LEaP** : Représente le programme principal pour créer des fichiers d'entrée avec des coordonnées et une topologie pour les macromolécules ;
- **ANTECHAMBER** : fait partie de la suite ANTECHAMBER, qui est le programme principal. Il est utilisé pour préparer les fichiers d'entrée LEaP dans le cas où le système contient des molécules autres que des macromolécules ;

2. Programmes de simulation :

- **SANDER** : acronyme de Simulated Annealing with NMR-Derived Energy Restraints Le programme de minimisation énergétique et de dynamique moléculaire. Ce programme est capable de proposer une conformation suffisamment stable, agissant sur les atomes se déplaçant itérativement le long du gradient d'énergie pour obtenir une structure plus relâchée (stable). Pour la dynamique moléculaire, il intègre les équations de mouvement de Newton pour générer des configurations de système. De cette façon, la molécule est capable de surmonter les petites barrières énergétiques et donc d'assumer des conformations non réalisables avec la minimisation mais également stable. L'utilisation de Sander conduit à des résultats plus précis que l'utilisation de la force Gall, car le calcul est caractérisé par une plus grande finesse ;
- **PMEMD** : acronyme de Particle Mesh Ewald Molecular Dynamics, est une version de SANDER qui vous permet de travailler plus rapidement, permet également d'utiliser le GPU en plus du CPU ;

3. Programmes d'analyse :

- **CPPTRAJ** : est un programme d'analyse et de traitement des fichiers de trajectoire de sortie des processus dynamiques ;
- **M-PBSA** : est une script qui automatise l'analyse d'énergie instantanée d'une simulation de dynamique moléculaire. Il existe deux versions de ce programme : le plus ancien MMPBSA.pl (écrit en Perl) et le dernier MMPBSA.py (écrit en Python), généralement plus facile à configurer. Ce programme n'a pas été utilisé dans mon travail de thèse.

5.5 Conclusion

Dans ce chapitre, nous avons décrit des méthodes utilisables pour analyser l'affinité d'une molécule organique ou synthétique pour une cible de nature protéique (Modélisation Moléculaire). En particulier, nous avons montré l'intérêt du Docking pour lequel il est primordial de considérer la flexibilité du ligand. Nous avons donné plusieurs exemples de logiciels utilisés à ce stade en expliquant le principe de chacun d'entre eux. Nous avons également introduit notion de dynamique moléculaire en amont du Docking afin d'introduire l'information de variabilité structurale de la cavité. Par ailleurs, les fonctions de scoring ont été décrites ainsi que la manière dont l'information doit en être interprétée.

Bibliographie

- [1] Chakkodabyly S Ramesha. How many leads from hits?—comment. *Drug discovery today*, 5(2) :43–44, 2000.
- [2] René-Just Haüy. Essai d’une théorie sur la structure des cristaux. *Goué & Née, Paris*, 1784, 1784.
- [3] Eric F Pettersen, Thomas D Goddard, Conrad C Huang, Gregory S Couch, Daniel M Greenblatt, Elaine C Meng, and Thomas E Ferrin. Ucsf chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13) :1605–1612, 2004.
- [4] Schrödinger Release. 3 : Maestro, version 9.6. *Schrödinger, LLC, New York, NY*, 2013.
- [5] Gerhard Wolber and Thierry Langer. Ligandscout : 3-d pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *Journal of chemical information and modeling*, 45(1) :160–169, 2005.
- [6] Gregory L Warren, C Webster Andrews, Anna-Maria Capelli, Brian Clarke, Judith LaLonde, Millard H Lambert, Mika Lindvall, Neysa Nevins, Simon F Semus, Stefan Senger, et al. A critical assessment of docking programs and scoring functions. *Journal of medicinal chemistry*, 49(20) :5912–5931, 2006.
- [7] Dina Duhovny, Ruth Nussinov, Haim J Wolfson, et al. Efficient unbound docking of rigid molecules. *Lecture notes in computer science*, 2452 :185–200, 2002.
- [8] Garrett M Morris, Ruth Huey, William Lindstrom, Michel F Sanner, Richard K Belew, David S Goodsell, and Arthur J Olson. Autodock4 and autodocktools4 : Automated docking with selective receptor flexibility. *Journal of computational chemistry*, 30(16) :2785–2791, 2009.
- [9] Oleg Trott and Arthur J Olson. Autodock vina : improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2) :455–461, 2010.
- [10] Renxiao Wang, Luhua Lai, and Shaomeng Wang. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *Journal of computer-aided molecular design*, 16(1) :11–26, 2002.
- [11] Renxiao Wang, Xueliang Fang, Yipin Lu, and Shaomeng Wang. The pdbbind database : Collection of binding affinities for protein- ligand complexes with known

- three-dimensional structures. *Journal of medicinal chemistry*, 47(12) :2977–2980, 2004.
- [12] Richard A Friesner, Jay L Banks, Robert B Murphy, Thomas A Halgren, Jasna J Klicic, Daniel T Mainz, Matthew P Repasky, Eric H Knoll, Mee Shelley, Jason K Perry, et al. Glide : a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *Journal of medicinal chemistry*, 47(7) :1739–1749, 2004.
- [13] Richard A Friesner, Robert B Murphy, Matthew P Repasky, Leah L Frye, Jeremy R Greenwood, Thomas A Halgren, Paul C Sanschagrín, and Daniel T Mainz. Extra precision glide : Docking and scoring incorporating a model of hydrophobic enclosure for protein- ligand complexes. *Journal of medicinal chemistry*, 49(21) :6177–6196, 2006.
- [14] Marcel L Verdonk, Jason C Cole, Michael J Hartshorn, Christopher W Murray, and Richard D Taylor. Improved protein–ligand docking using gold. *Proteins : Structure, Function, and Bioinformatics*, 52(4) :609–623, 2003.
- [15] Carol A Baxter, Christopher W Murray, David E Clark, David R Westhead, and Matthew D Eldridge. Flexible docking using tabu search and an empirical estimate of binding affinity. *Proteins : Structure, Function, and Bioinformatics*, 33(3) :367–382, 1998.
- [16] Matthew D Eldridge, Christopher W Murray, Timothy R Auton, Gaia V Paolini, and Roger P Mee. Empirical scoring functions : I. the development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *Journal of computer-aided molecular design*, 11(5) :425–445, 1997.
- [17] Wijnand Mooij and Marcel L Verdonk. General and targeted statistical potentials for protein–ligand interactions. *Proteins : Structure, Function, and Bioinformatics*, 61(2) :272–287, 2005.
- [18] Ingo Muegge and Yvonne C Martin. A general and fast scoring function for protein–ligand interactions : a simplified potential approach. *Journal of medicinal chemistry*, 42(5) :791–804, 1999.
- [19] Holger Gohlke, Manfred Hendlich, and Gerhard Klebe. Knowledge-based scoring function to predict protein-ligand interactions. *Journal of molecular biology*, 295(2) :337–356, 2000.

- [20] Oliver Korb, Thomas Stutzle, and Thomas E Exner. Empirical scoring functions for advanced protein- ligand docking with plants. *Journal of chemical information and modeling*, 49(1) :84–96, 2009.
- [21] William J Allen, Trent E Balias, Sudipto Mukherjee, Scott R Brozell, Demetri T Moustakas, P Therese Lang, David A Case, Irwin D Kuntz, and Robert C Rizzo. Dock 6 : impact of new features and current docking performance. *Journal of computational chemistry*, 36(15) :1132–1156, 2015.
- [22] Oliver Korb, Thomas Stützle, and Thomas E Exner. Plants : Application of ant colony optimization to structure-based drug design. In *International Workshop on Ant Colony Optimization and Swarm Intelligence*, pages 247–258. Springer, 2006.
- [23] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with namd. *Journal of computational chemistry*, 26(16) :1781–1802, 2005.
- [24] Philip J Davis and Philip Rabinowitz. *Methods of numerical integration*. Courier Corporation, 2007.
- [25] Timothy J Giese, Maria T Panteva, Haoyuan Chen, and Darrin M York. Multipolar ewald methods, 2 : Applications using a quantum mechanical force field. *Journal of chemical theory and computation*, 11(2) :451–461, 2015.

Chapitre 6

Virtuel Screening à base du pharmacophore pour l'identification des inhibiteurs de LDH

Sommaire

6.1 Introduction	123
6.2 Analyse de la structure protéique 4QSM (X-Ray structure), de l'inhibiteur de référence et de ses dérivés actifs	125
6.2.1 Modèle de référence (Template)	125
6.2.2 Caractéristiques de la structure cristallographique code : 4QSM	127
6.3 Analyse du self_Docking	132
6.4 Screening de Database Enamine	133
6.4.1 Database Enamine	133
6.4.2 Construction d'un modèle pharmacophore basé sur les ligands 38K et le composé 9	134
6.4.3 Virtuel Screening utilisant le software Ligandscout	135
6.4.4 Le Docking avec le programme Glide	136
6.5 Simulation dynamique moléculaire	138
6.5.1 Calibration de la procédure	139
6.5.2 Simulation dynamique moléculaire	142
6.5.3 Analyse des résultats	144
6.5.4 Analyse visuelle	145
6.6 Conclusion	161

6.1 Introduction

Le travail que j'avais réalisé pour identifier de nouveaux inhibiteurs potentiels du lactate déshydrogénase est schématisé comme suit :

- Analyse de la structure cristallographique de 4QSM (X-Ray structure), de l'inhibiteur de référence 38K (composé 2) et de ses dérivés ;
- Procédure SelfDocking ;
- Screening du database Enamine ;
- Simulation dynamique moléculaire.

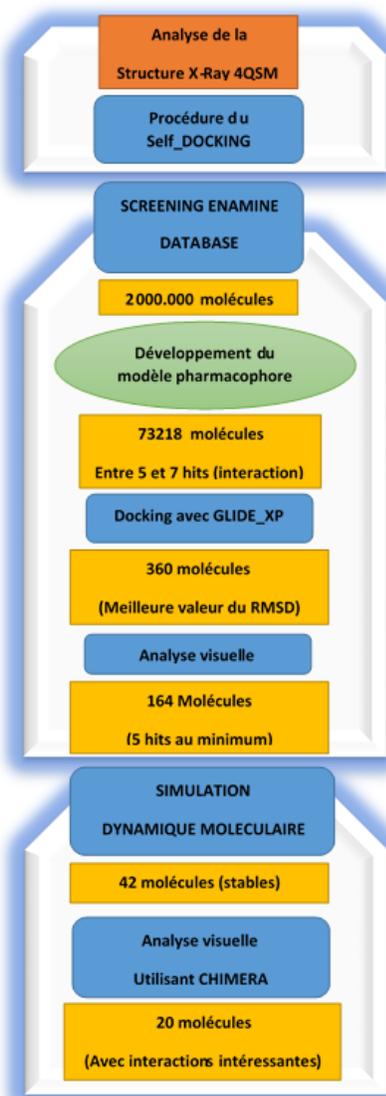


FIGURE 6.1 – Les différentes étapes du procédure de Virtuel Screening.

6.2 Analyse de la structure protéique 4QSM (X-Ray structure), de l'inhibiteur de référence et de ses dérivés actifs

6.2.1 Modèle de référence (Template)

Le point de départ de mon travail dans cette partie était l'analyse du travail de groupe du département d'orthopédie, Hôpital de la province du Heilongjiang, en Chine l'année 2016, qui a permis l'identification d'une nouvelle classe d'inhibiteurs de la LDH-5 2-((3-cyanopyridin-2-yl)thio)acetamides[1]. Cette classe chimique a été sélectionnée à la fin d'une étude VS sur une collection des ligands commercialisés de zinc database. Wei Cui et all ont considéré la protéine du code PDB : 4QSM comme un template pour leur travail de VS. Cette structure protéinique qui a été identifiée en 2014, par le groupe de biologie moléculaire de l'université de Simon Fraser à Cannada[2], montre une bonne résolution de 3 Å°. Cette étude a aboutit à l'identification de cinq ligands (5,6,7,8,9) qui présente des interactions intéressantes avec le site actif (Figure 6.2).

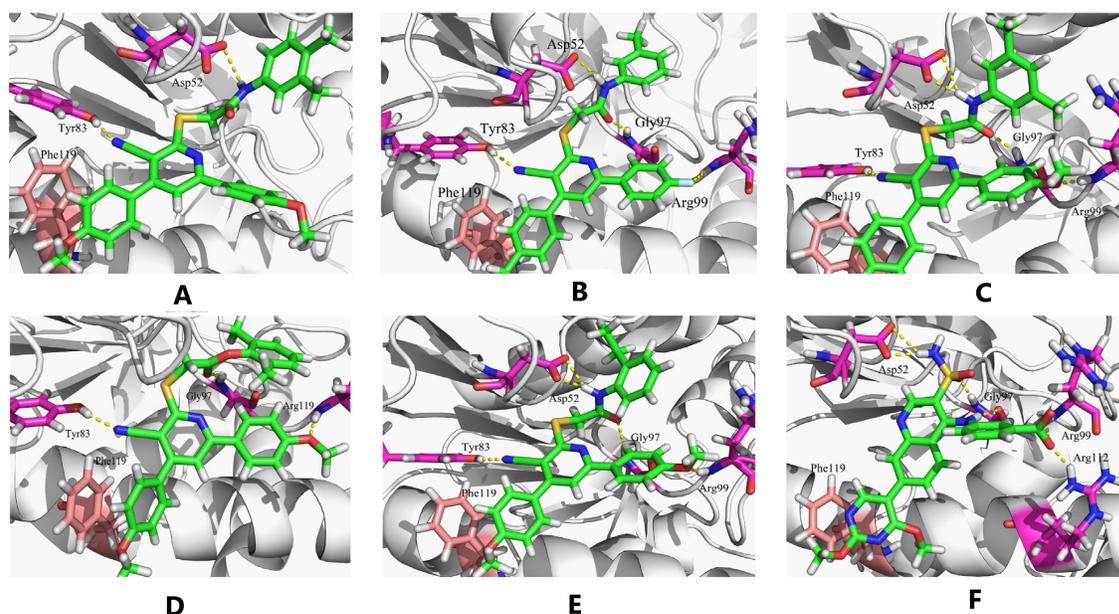


FIGURE 6.2 – Modélisation moléculaire des inhibiteurs identifiés avec le ligand de référence (composé 2) : (A) composé 5 ; (B) composé 6 ; (C) composé 7 ; (D) composé 8 ; (E) composé 9 ; (F) ligand de référence (38K).

Ainsi, les 5 composés ont été testés pour une validation biologique supplémentaire, dans le but d'obtenir des informations sur cette série de composés. Le dosage enzymatique a été déterminé par mesure de la disparition du NADH à 340 nm lors de la conversion du pyruvate en lactate (Tableau 6.1).

TABLE 6.1 – Les valeurs du dosage enzymatique et score de liaison total.

Composés	Inhibition du LDHA $CI_{50}(\mu M)$	Score de liaison total
5	>100	5.71
6	3.40 ± 0.41	7.80
7	6.59 ± 0.77	7.31
8	22.47 ± 0.40	6.21
9	1.24 ± 0.18	7.29
2 (code 38K)	7.40 ± 0.88	5.43

Par rapport au composé 2 (code : 38K) ($CI_{50} = 7,40 \mu M$), la plupart des composés identifiés montraient de meilleures puissances inhibitrices. L'activité inhibitrice accrue contre la LDHA pourrait être partiellement expliquée par les preuves obtenues à partir de leur étude de Docking (Figure 6.3). Semblable au composé 2, les inhibiteurs sélectionnés 6, 7 et 9 ont conservé l'interaction des liaisons hydrogènes avec les résidus ASP 52, le Gly 97 et Arg 99, en outre, ils ont réussi à capter l'interaction hydrogène-liaison avec le résidu Tyr 83, ce qui peut contribuer à la puissance inhibitrice accrue.

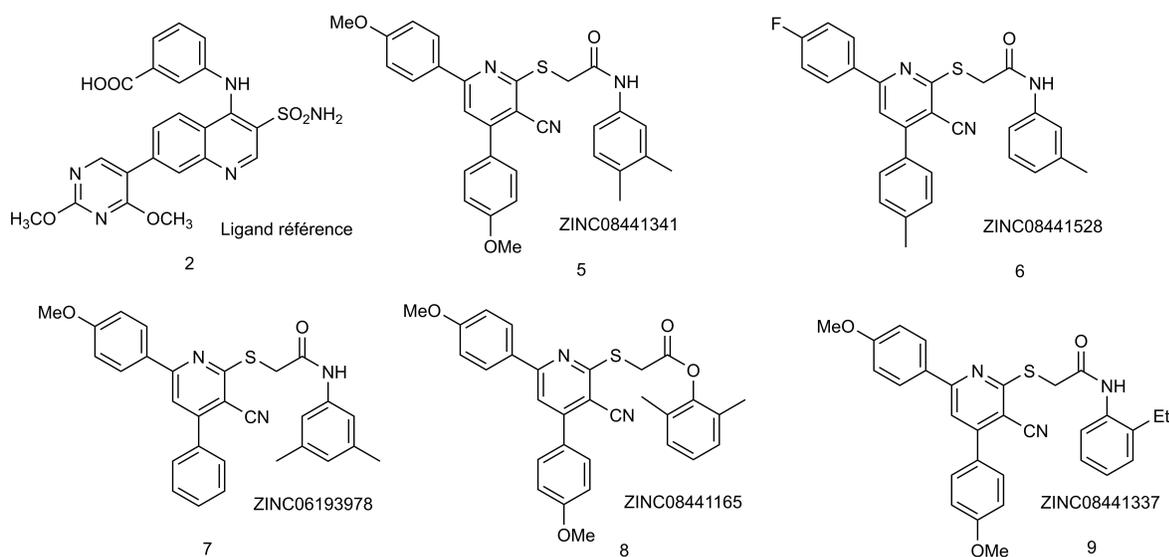


FIGURE 6.3 – Structures développées des ligands identifiés avec le ligand référence.

En résumé, Wei Cui et ces collaborateurs ont identifié après une étude VS des inhibiteurs de LDH humaine avec motif 2 - ((3-cyanopyridin-2-yl) thio) acétamide, utilisant le PDB du code 4QSM comme un template de travail.

Le composé 9 présente une activité inhibitrice plus puissante que celle du composé 2

.parce qu'il peut faire une autre interaction de plus avec Tyr 83, ce qui le met en tête de file, et nous permet l'identification d'autre dérivés plus puissants, plus anti-proliférative. Dans ce contexte mon travail était basé sur le développement d'un modèle qui réunit les interactions des deux composés actifs le composé 2 et le composé 9, ce qu'on l'appel un modèle pharmacophore.

6.2.2 Caractéristiques de la structure cristallographique code : 4QSM

La structure cristallographique a été déposée dans la Banque de Données des Protéines (Proteine Data Bank)[3] en 2014 avec une résolution de 3.0 Å. Les huit chaînes du 4QSM

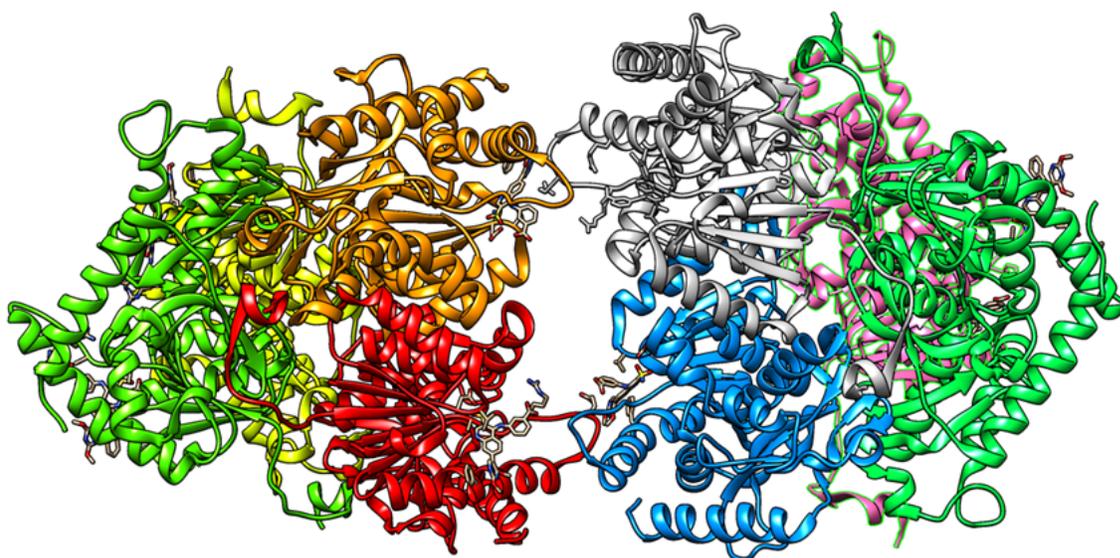


FIGURE 6.4 – La structure cristallographique du 4QSM : chaîne A (gris), B (bleu), C (violet), D (Vert clair), E (vert foncé), F (jaune), G (orangé), H (rouge).

ont les mêmes caractéristiques : complètes, boucle fermée, cristallisé avec l'inhibiteur du code 38K(Figure 6.4; Figure 6.5).

Le composé 2 et le composé 9, ont été analysés pour leur capacité d'inhiber la LDHA en mesurant la diminution de l'absorbance lorsque le NADH est converti en NAD⁺, conversion du pyruvate en lactate catalysée par LDHA.

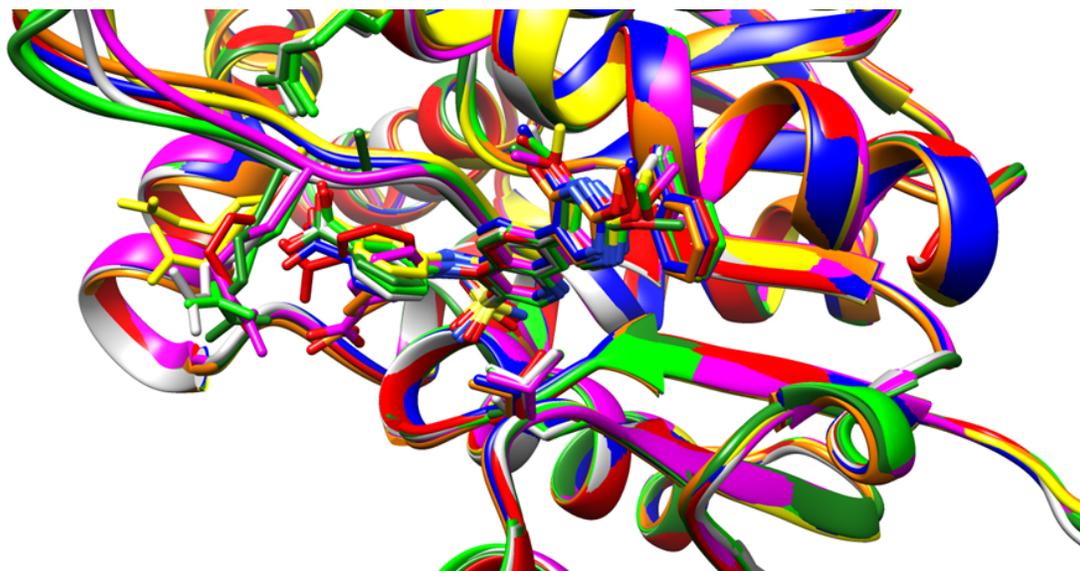


FIGURE 6.5 – Chevauchement des huit chaînes, détail sur le site catalytique. chaîne A (gris), B (bleu), C (violet), D (Vert clair), E (vert foncé), F (jaune), G (orangé), H (rouge).

Dans le cadre de ce travail, la chaîne A a été sélectionnée, en prenant comme référence le facteur de température (facteur de température ou FT) du résidu extrêmement mobile Arg99. FT est un facteur utilisé en cristallographie, qui corréle la précision de la position d'un résidu dans la protéine ; plus sa valeur est grande, plus la possibilité que la position du résidu soit incertaine. Le tableau 6.2 montre des extraits du fichier PDB code : 4QSM ; liés aux valeurs de FT de Arg99 dans la chaîne A et B.

Les interactions principales des deux composé actifs avec la chaîne A sont montrées ;

TABLE 6.2 – Les valeurs de FT de Arg99 dans la chaîne A et B.

ATOM	1610	N	ARG	A	105	30.68	ATOM	8592	N	ARG	B	105	39.59
ATOM	1611	CA	ARG	A	105	38.31	ATOM	8593	CA	ARG	B	105	45.68
ATOM	1612	C	ARG	A	105	32.48	ATOM	8594	C	ARG	B	105	33.92
ATOM	1613	O	ARG	A	105	31.59	ATOM	8595	O	ARG	B	105	31.02
ATOM	1614	CB	ARG	A	105	43.96	ATOM	8596	CB	ARG	B	105	57.04
ATOM	1615	CG	ARG	A	105	51.18	ATOM	8597	CG	ARG	B	105	70.98
ATOM	1616	CD	ARG	A	105	53.80	ATOM	8598	CD	ARG	B	105	74.65
ATOM	1617	NE	ARG	A	105	49.54	ATOM	8599	NE	ARG	B	105	78.72
ATOM	1618	CZ	ARG	A	105	44.02	ATOM	8600	CZ	ARG	B	105	82.53
ATOM	1619	NH ₁	ARG	A	105	39.71	ATOM	8601	NH ₁	ARG	B	105	81.16
ATOM	1620	NH ₂	ARG	A	105	39.71	ATOM	8602	NH ₂	ARG	B	105	85.59

la visualisation 3D est montrée à travers le software UCSF CHIMERA et un schéma 2D avec le software Flap.

Ce qui suit sont les différentes interactions de composé 2 avec le récepteur (Figure 6.6; Figure 6.7) :

- Cinq liaisons hydrogènes avec les résidus : ASP 52 ; GLY 97 ; ARG 99 ; ARG 112 ;
- Trois interactions hydrophobes avec les résidus : VAL 53 ; ILE 116 ; PHE 119.

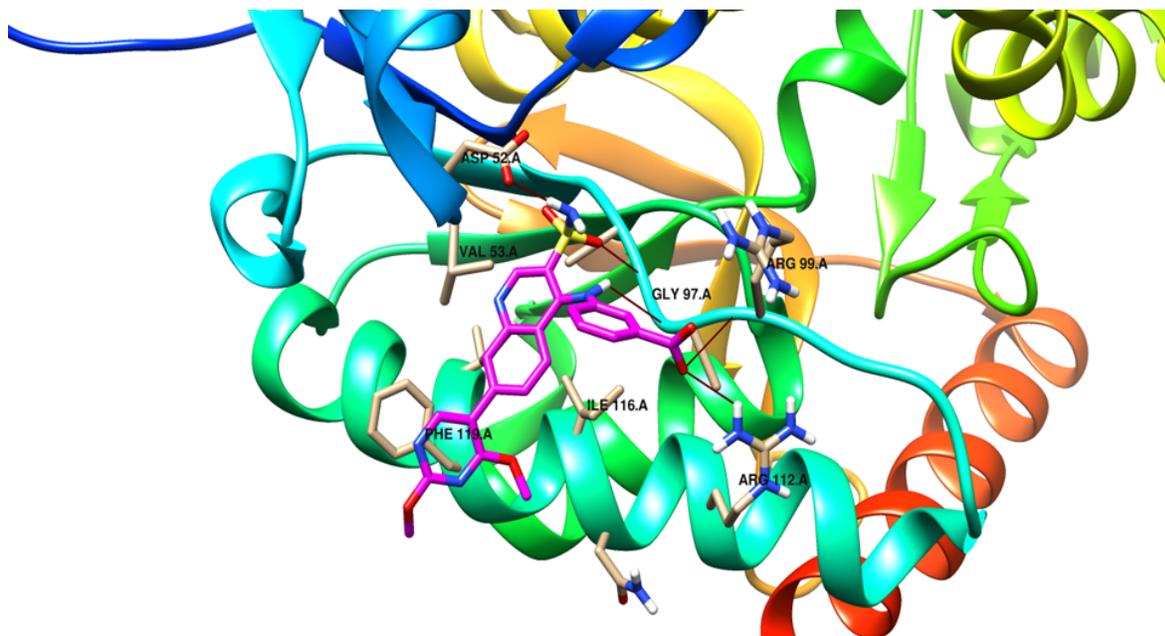


FIGURE 6.6 – Les interactions 3D du composé 2 avec le site actif.

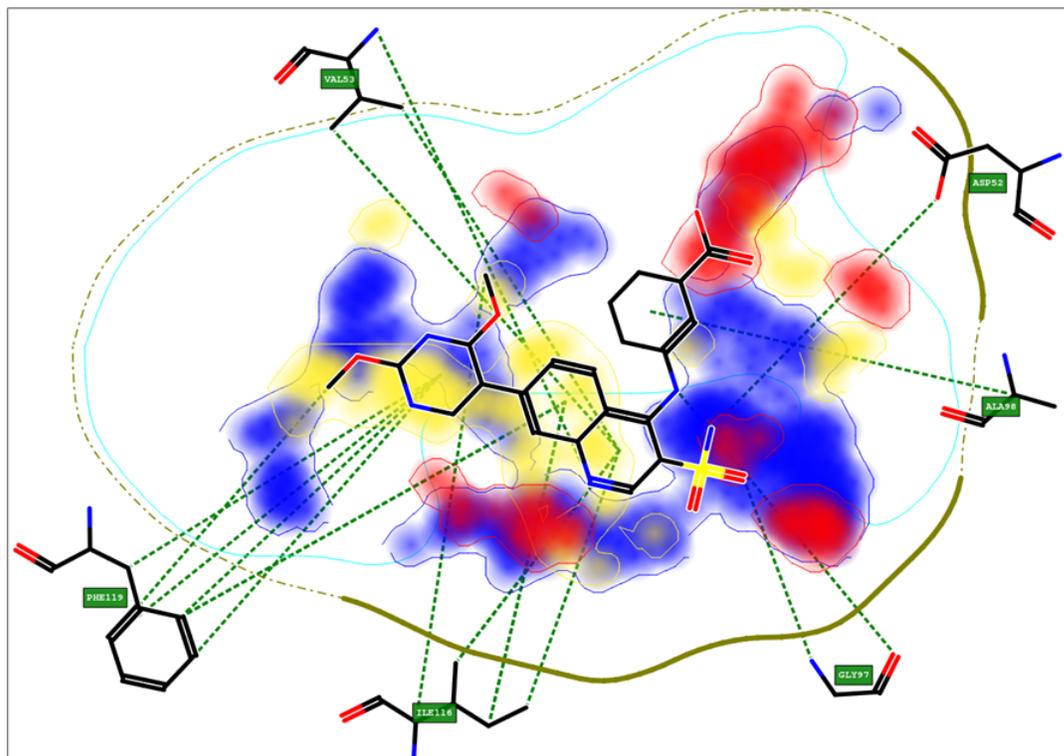


FIGURE 6.7 – Les interactions 2D du composé 2 avec le site actif.

Ce qui suit sont les différentes interactions de composé 9 avec le récepteur (Figure 6.8; Figure 6.9) :

- Cinq liaisons hydrogènes avec les résidus : ASP 52 ; GLY 97 ; ARG 99 ; TYR 83.
- Deux interactions hydrophobes avec les résidus : ILE 116 ; PHE 119.

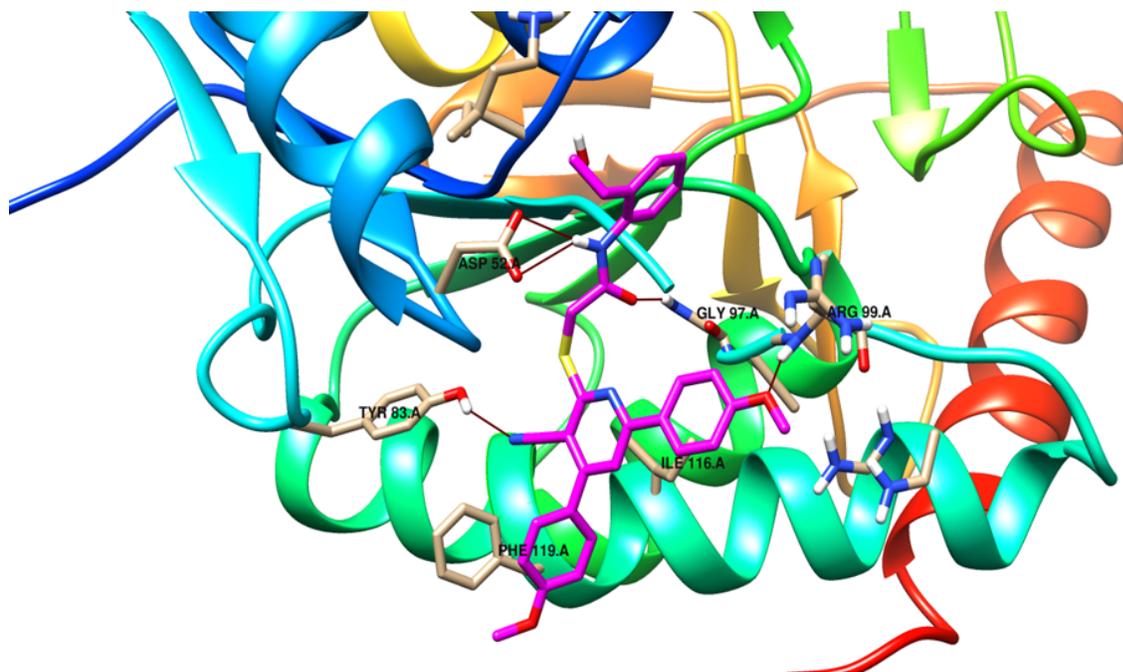


FIGURE 6.8 – Les interactions 3D de composé 9 avec le site actif.

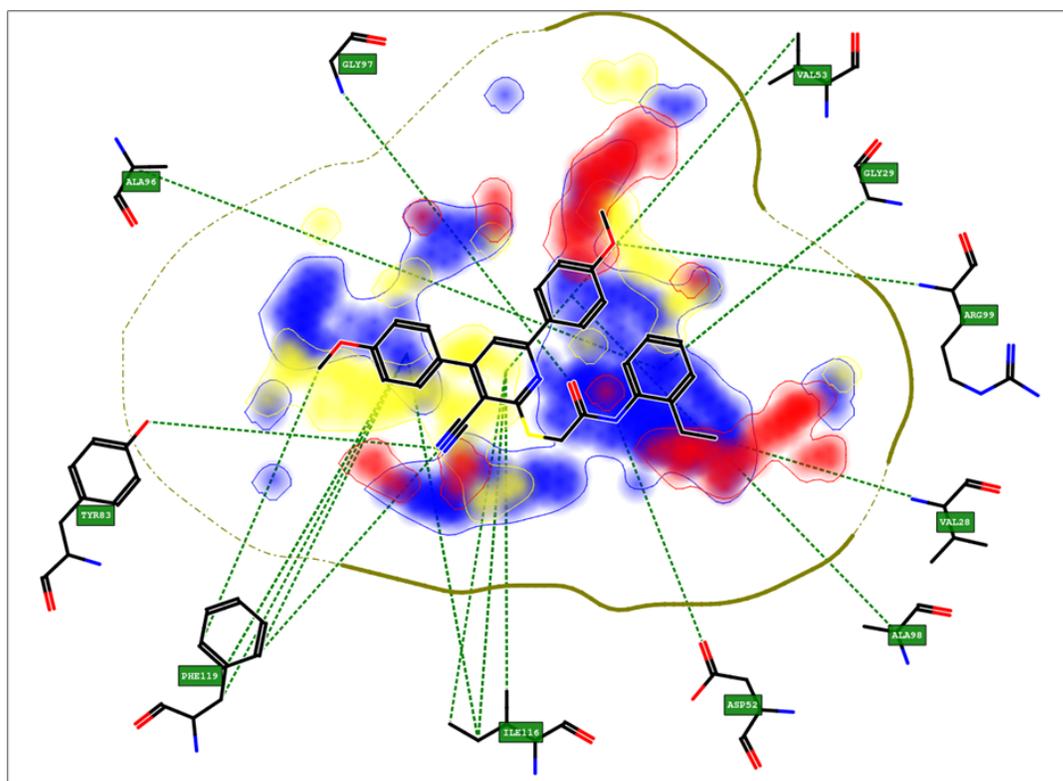


FIGURE 6.9 – Les interactions 2D de composé 9 avec le site actif.

6.3 Analyse du self_Docking

Une procédure self_Docking est l'évaluation de la capacité d'un programme de Docking pour prédire la conformation d'un ligand de référence avec sa protéine correspondante.

La procédure de self_Docking a été utilisée pour évaluer les 12 programmes disponibles dans le laboratoire (Laboratoire de Modélisation Moléculaire et Virtuel Screening, Université de Pise, Italie), résumés dans le Tableau 6.3 :

TABLE 6.3 – Les logiciels de Docking utilisés.

	Programme	Fonction du Score
1		AutoDock
2	ADT	VINA
3	DOCK6	DOCK score
4	FRED	ChemGauss4
4	FRED	DOCK score
5	GLAMDOCK	Monte-Carlo
6		Standard Precision (SP)
7	GLIDE	Extra Precision (XP)
8		ASP
9		Chemscore
10	GOLD	Goldscore
11		PLP
12	PLANTS	ChemPLP

Cette étude a été faite pour les deux ligands actifs (composé 2 et 9), le Tableau 6.4 montre les valeurs du RMSD obtenus.

Cette phase de l'analyse est particulièrement importante, car le logiciel de Docking choisis, va être utilisée comme seuil pour la partie suivante de ma thèse.

A partir de cette première analyse, le potentiel de la fonction GLIDE_XP intégré en Glide software a été mis en évidence en tant qu'outil permettant de filtrer la base de données ENAMINE, car elle donne les meilleurs résultats de Self_Docking avec les deux composés.

TABLE 6.4 – Les logiciels de Docking utilisés.

Programme	Composé 2	Composé 9
ADT	2.3	3.0
VINA	1.0	2.2
FRED	2.4	3.6
GLAMDOCK	2.03	1.7
GLIDE_SP	1.9	1.5
GLIDE_XP	1.0	1.1
GLOD_ASP	4.2	4.1
GLOD_CS	3.4	1.3
GLOD_GS	2.3	7.02
GLOD_PLP	5.8	1.4
PLANTS	1.7	1.1
DOCK6	6.1	3.8
RMSD $\leq 1.5\text{\AA}^\circ$		
RMSD $> 1.5\text{\AA}^\circ$		

6.4 Screening de Database Enamine

l'idée crucial de cette partie a été l'analyse d'une grande base de données à l'aide d'une étude virtuel screening, comme la base de données Enamine. Comme ça était indiqué dans la partie introductive de ce chapitre, j'ai utilisé une stratégie bien déterminée pour atteindre les résultats finaux, passant par le développement d'un modèle pharmacophore, pour effectuer une présélection sur la base de données, et finissant par le Docking par le logiciel potentiel pour mon étude GLIDE_XP.

6.4.1 Database Enamine

La société Enamine Ltd[4] possède la plus grande base de données des molécules pour le screening à haut débit (HTS), et pour des études de criblage virtuel (Virtual Screening). En effet, il est également possible de télécharger des bases de données virtuelles de molécules pour des études in silico. Enamine Ltd offre les possibilités suivantes :

- Advanced Collection (mise à jour mars 2016), 278 365 composés ;
- HTS Collection (mise à jour mars 2016), 1 720 909 composés ;
- Screening Collection (mise à jour mars 2016), 36 061 composés ;

Pour les études du criblage virtuel de notre laboratoire, toutes les collections HTS sont utilisées, téléchargées en fichier de format SDF avec des informations uniquement sur la structure bidimensionnelle des molécules.

Afin de préparer la base de données pour les études suivantes, il était donc nécessaire d'utiliser le programme ligprep[5], qui fait partie de la suite Schrodinger, dans le but de générer des structures 3D avec une énergie minimisée, d'appliquer aussi des règles sophistiquées pour corriger les structures de Lewis et éliminer les erreurs dans le but de réduire les erreurs de calcul.

6.4.2 Construction d'un modèle pharmacophore basé sur les ligands 38K et le composé 9

Cette partie de ma thèse a été entièrement réalisée en utilisant le programme Ligand-Scout. Au lieu de charger le cristal 4QSM directement à partir de la DataBank de protéines, j'ai utilisé une structure de chaîne A précédemment optimisée au format MOL2, où le ligand 38K et le ligand 9, présentent l'ordre de liaison correct et toutes les autres molécules sont supprimées, que se soient les molécules d'eau ou les cofacteurs.

À la suite de la procédure décrite au chapitre 2, j'ai créé plusieurs modèles pharmacophores communs (les interactions de 38K et composé 9), en essayant de garder les interactions les plus importantes avec la protéine et en même temps créer une structure, qui est en fait une projection pour un criblage approprié d'une grande base de données. Pour cette raison, j'ai fait plusieurs tests, en construisant initialement une petite bibliothèque de composés (28 composés) tous structurellement liés aux deux ligands actifs en littérature. Sur cette bibliothèque j'ai fait plusieurs tests pour évaluer la différence entre les différentes options offertes par le programme. De cette façon, il était possible de choisir la construction de pharmacophores avec des interactions (des features) obligatoires plus un deuxième groupe des interactions facultatives plutôt que de laisser au programme la possibilité d'omettre une interaction non spécifiée. Ce type de pharmacophore a été mis à l'épreuve d'un screening virtuel par un plus petit paquet de dababase Enamine (environ 60 000 molécules), afin d'avoir une estimation sur le nombre et les molécules de la proposition qui ont pu filtrer à travers ce modèle. Le modèle pharmacophore qui a donné les meilleurs résultats était le numéro 17, il a été ensuite utilisé pour filtrer le reste de la base de données Enamine sa structure du modèle est représentée sur la Figure 6.10 :

Le modèle pharmacophore se compose donc de 3 interactions obligatoires :

- Accepteur d'une liaison hydrogène avec le résidu ARG 99;

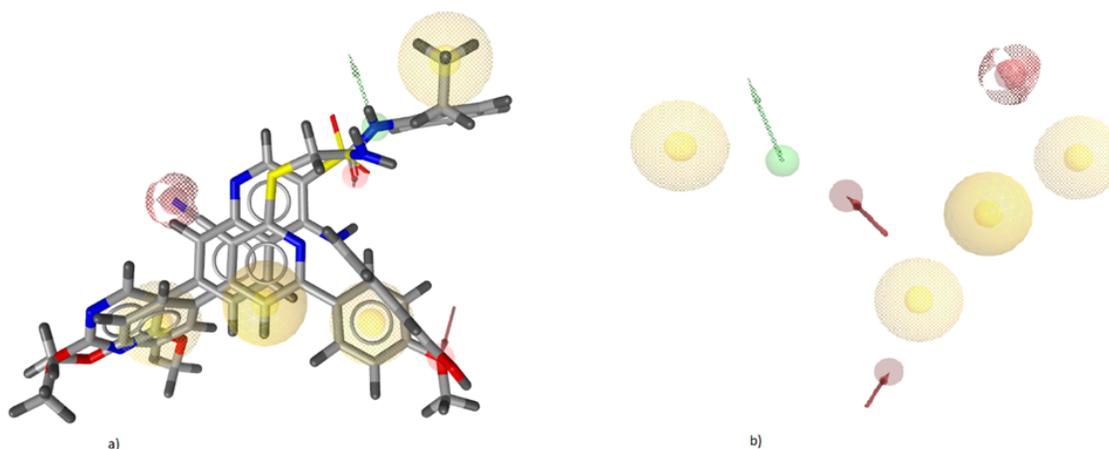


FIGURE 6.10 – a) Le modèle pharmacophore avec les deux ligands actifs ; b) le modèle pharmacophore seul.

- Accepteur d'une liaison hydrogène avec le résidu GLY 97 ; ;
- Interaction hydrophobique avec les deux résidus VAL 53 et ILE 116.

Et aussi 5 interactions optionnelles, qui peuvent être supprimées par le programme lors de la recherche :

- Donneur d'une liaison hydrogène avec le résidu ASP 52 ;
- Accepteur d'une liaison hydrogène avec le résidu TYR 83 ;
- Interaction hydrophobique avec ALA 98 ;
- Interaction hydrophobique avec le résidu PHE 119 ;
- Interaction hydrophobique avec le résidu LYS 57.

6.4.3 Virtuel Screening utilisant le software Ligandscout

Le criblage de la base de données Enamine utilisant la fonction *Iscreen* de Ligand-Scout a conduit à l'obtention de 73 218 molécules à partir de 3 million molécules.

Cette étape nous a permis de réduire le nombre de molécules jusqu'à 73 218 molécules, en prenant que les molécules qui possèdent aux minimum 5 interactions ou plus avec notre modèle pharmacophore, donc les interactions dans notre cas doivent être comprises entre 5 et 8 (nombre d'interactions total dans le modèle égal 8).

aléatoire, ou dans ce cas, l'enrichissement d'une base de données de molécules inactives (*decoys*) à partir de molécules dites actives (les deux molécules actifs). La formule de calcul du facteur d'enrichissement est :

$$FE = \left[\frac{t_p}{t_p + f_n} \right] \left(\frac{NC_{tot}}{NC} \right) \quad (6.1)$$

Avec

t_p : Nombre de vrais positifs (molécules actives identifiées comme actives) ;

f_n : Nombre de faux négatifs (molécules actives identifiées comme inactives) ;

NC_{tot} : Nombre total des molécules de database ;

NC : Nombre total des molécules considérées comme actives.

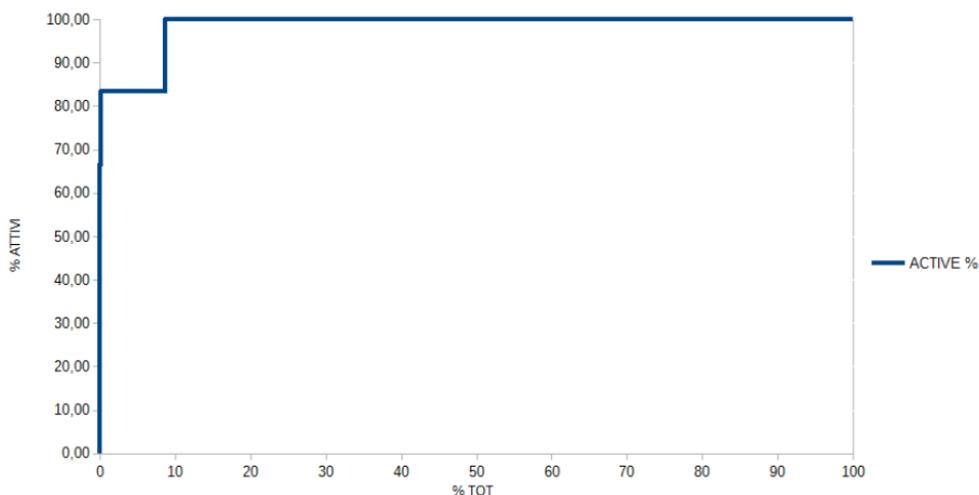


FIGURE 6.12 – Graphe d'enrichissement calculé sur la valeur du XP_Gscore de deux ligands actifs vis à vis 73 218 inactifs.

Pour la construction du graphique d'enrichissement (*enrichment plot*) correspondant à la Figure 6.12, j'ai comparé les valeurs de XP_Gscore des 73218 molécules obtenues dans l'étape précédente, après le screening à base de pharmacophore avec celles obtenues dans la phase de self_Docking.

D'après ce graphe, il est clair que Glide_XP permet d'avoir des niveaux élevés de FE, puisque dans les premiers 5% des molécules totales en ordre avec les valeurs de XP_Gscore (sur l'axe Y), plus de 80% des molécules actives sont déjà incluses.

De plus GLIDE_XP est absolument le programme le plus approprié parmi ceux disponibles car il donne les meilleures valeurs du RMSD, et par conséquent, grâce à sa précision et aux bonnes valeurs de FE obtenues, il a été choisi pour le Docking des 73 218 molécules obtenues après le filtrage à base de pharmacophore.

Row	In	Title	Stars	glide_gscore	glide_rmsd_to_input	Entry ID	Date Added	Date Modified
1	<input type="checkbox"/>	Z71394704	☆☆☆	-7.16	1.842	8	13:02	13:02
2	<input type="checkbox"/>	Z218034438	☆☆☆	-6.74	1.780	5	13:02	13:02
3	<input type="checkbox"/>	Z16508305	☆☆☆	-6.73	1.838	7	13:02	13:02
4	<input type="checkbox"/>	Z362557984	☆☆☆	-6.62	2.372	52	13:02	13:02
5	<input type="checkbox"/>	Z1528343122	☆☆☆	-6.57	1.284	3	13:02	13:02
6	<input type="checkbox"/>	Z49592896	☆☆☆	-6.56	1.467	4	13:02	13:02
7	<input type="checkbox"/>	Z1172314765	☆☆☆	-6.42	1.640	0	13:02	13:02
8	<input type="checkbox"/>	Z15661695	☆☆☆	-6.24	2.083	09	13:02	13:02
9	<input type="checkbox"/>	Z16507442	☆☆☆	-6.19	2.115	14	13:02	13:02
10	<input type="checkbox"/>	Z88363798	☆☆☆	-6.16	2.481	63	13:02	13:02
11	<input type="checkbox"/>	Z16506294	☆☆☆	-6.10	2.341	49	13:02	13:02
12	<input type="checkbox"/>	Z16508390	☆☆☆	-6.10	2.260	39	13:02	13:02
13	<input type="checkbox"/>	Z45701089	☆☆☆	-6.06	1.997	8	13:02	13:02
14	<input type="checkbox"/>	Z239635716	☆☆☆	-6.05	1.909	0	13:02	13:02
15	<input type="checkbox"/>	Z449585644	☆☆☆	-6.04	1.791	7	13:02	13:02
16	<input type="checkbox"/>	Z371950944	☆☆☆	-6.01	2.127	18	13:02	13:02
17	<input type="checkbox"/>	Z16508182	☆☆☆	-6.00	2.380	53	13:02	13:02
18	<input type="checkbox"/>	Z738056752	☆☆☆	-5.97	1.543	0	13:02	13:02
19	<input type="checkbox"/>	Z458808666	☆☆☆	-5.95	1.990	4	13:02	13:02
20	<input type="checkbox"/>	Z16507440	☆☆☆	-5.91	1.766	32	13:02	13:02
21	<input type="checkbox"/>	Z114189580	☆☆☆	-5.91	1.544	2	13:02	13:02
22	<input type="checkbox"/>	Z31723825	☆☆☆	-5.90	2.468	62	13:02	13:02
23	<input type="checkbox"/>	Z16507821	☆☆☆	-5.80	2.185	26	13:02	13:02
24	<input type="checkbox"/>	Z362664344	☆☆☆	-5.78	2.386	54	13:02	13:02

FIGURE 6.13 – Résultats du Docking pour les 73 218 molécules utilisant Glide_XP.

A la fin de Docking le nombre total est réduit de 73 218 molécules jusqu'à 360 molécules en fonction des valeurs du RMSD, où on a pris que les molécules qui ont une valeur de RMSD inférieure à 3 Å° comme il est montré sur la Figure 6.13.

6.4.4.2 Analyse visuelle

Comme on a mentionné dans la section précédente, que les molécules qui possèdent plus de 5 interactions avec le modèle pharmacophore ont été sélectionnées, mais parfois le logiciel Ligandscout exagère en donnant des interactions qui n'existent pas au niveau de notre modèle, donc il est nécessaire d'effectuer une analyse visuelle pour les molécules obtenues.

Dans cette étape en fait appel au logiciel de visualisation Chimera qui nous permet de voir les molécules avec notre modèles pharmacophore en 3D (Figure 6.14). Les 360 molécules sont analysées une par une pour un choix approprié de nos molécules. Après cette étape le nombre des molécules est réduit encore plus jusqu'à 164 molécules, des molécules qui présentent des fortes interactions avec notre modèle pharmacophore.

6.5 Simulation dynamique moléculaire

Dans cette dernière phase, grâce au logiciel AMBER14 (Assisted Model Building with Energy Refinement), une étude de dynamique moléculaire a été réalisée.

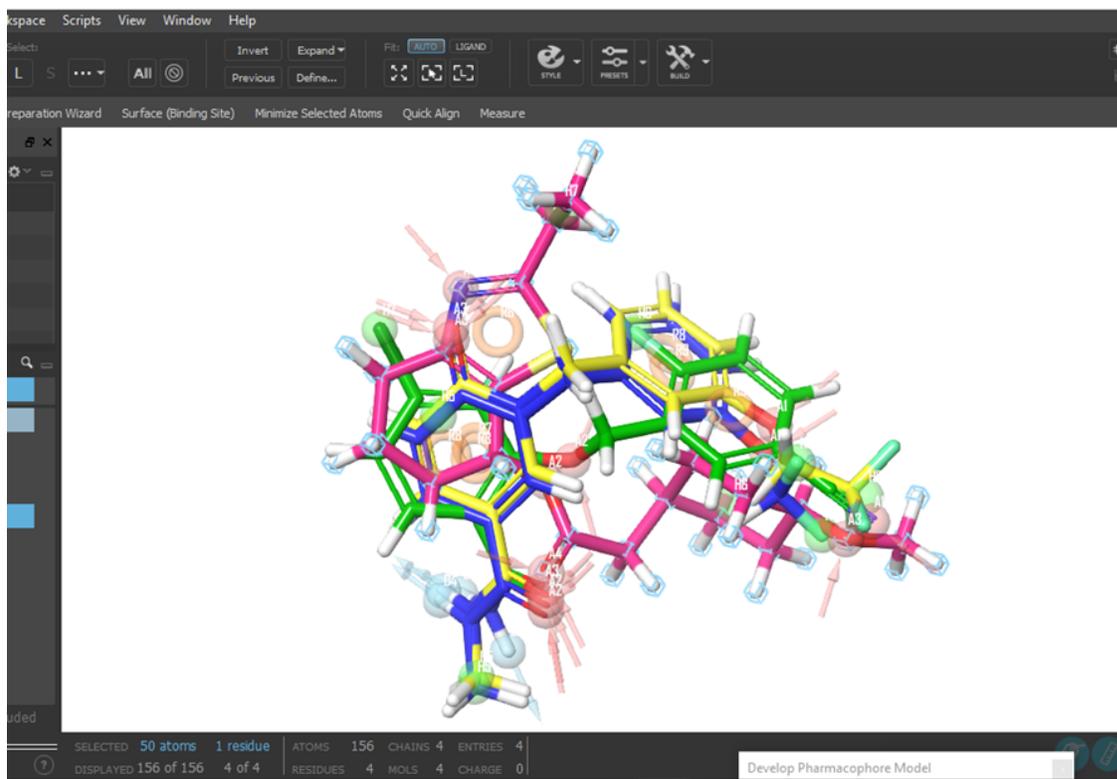


FIGURE 6.14 – L'analyse visuelle des résultats de Docking utilisant Chimera.

Grâce à la dynamique moléculaire, il est possible de vérifier la stabilité d'un système biologique en fournissant de l'énergie[7]. Contrairement au Docking, où les modèles considèrent que le ligand est en mouvement (flexible), dans la DM la protéine acquiert également une certaine capacité de mouvement, obtenant ainsi une reproduction des conditions plus similaires au système biologique. Si les interactions établies ne sont pas assez fortes, le complexe pourra rencontrer la dissociation (la perte des interactions entre le ligand et la protéine).

Dans notre cas, une étude de ce type a été réalisée dans le but de vérifier lesquelles des 164 molécules obtenues à la fin de filtrage, étaient suffisamment stables pour pouvoir maintenir les interactions importantes avec la LDH.

6.5.1 Calibration de la procédure

Avant d'effectuer la dynamique sur les 164 ligands obtenus en complexe avec le récepteur (l'enzyme LDH), il a été nécessaire de valider la procédure de calcul. A cette fin, une simulation préliminaire de 10 ns du complexe contenant les deux ligands actifs (le ligand référence et le composé 9) a été réalisée.

Pour évaluer ce qui pourrait être la procédure la plus appropriée, plusieurs tests ont

été effectués qui peuvent être résumés comme suit :

- Procédure "classique"[8] : le ligand est chargé avec le logiciel Antichambre, suivi par une minimisation et une DM qui est divisée en trois étapes ;
- Utilisation du programme Gaussian Accelerated[9] pour charger le ligand et le cofacteur ;
- Utilisation d'une procédure étape par étape pour basculer entre les étapes de la dynamique du carbone-alpha bloquée et libre :
 - (a) Première étape de minimisation (comme ci-dessus) ;
 - (b) Première étape de la dynamique du volume constant (comme avant) ;
 - (c) Deuxième étape de la dynamique à température constante (0,5 ns), où la valeur de la contrainte sur les carbones-alpha est fixée à 10 kcal/mol ;
 - (d) Troisième étape de dynamique à température constante (0,5 ns), où la valeur de contrainte est portée à 5 kcal / mol ;
 - (e) Quatrième étape de la dynamique à température constante (0,5 ns), où la valeur de la contrainte est portée à 1 kcal/mol ;
 - (f) Cinquième et dernière étape de la dynamique, avec les carbones-alpha ne sont plus fixés (16 ns).

Résultats de la procédure "classique" :

C'est la procédure indiquée au chapitre 2 de cette partie ; après chargement du ligand référence (38K) et le composé 9 dans le programme Antichambre, le schéma dynamique est le suivant :

- Première étape de minimisation, dans laquelle seuls les atomes de carbone-alpha de la protéine sont fixés (afin de minimiser le ligand et les chaînes latérales) ;
- Première étape de la dynamique réelle, d'une durée de 0,25 ns, à volume constant, dans laquelle l'énergie est fournie au système en maintenant le carbone-alpha de la protéine fixé (la température passe progressivement de 0 K à 300 K) ;
- Deuxième étape dynamique de 1.5 ns à pression et température constantes (P = 1 atm et T = 300 K). Les carbones-alpha restent fixés ;
- Troisième et dernière étape de la dynamique, d'une durée de 8,25 ns. Les carbones-alpha ne sont plus fixés, maintenant tout le complexe est libre durant le reste de la simulation DM.

Le graphe ci-dessous montre le RMSD des carbones-alpha de la protéine (en bleu), du ligand de référence (en vert) et du ligand 9 (en orange) en fonction du nombre d'étapes (100 étapes équivalent à 10 ns de dynamique).

Il est évident que la DM a donné de très bons résultats, puisque les carbonés-alpha ont une valeur RMSD moyenne d'une valeur égale à 1,6 Å, le ligand de référence et le ligand 9 tous les deux présentent une stabilité remarquable (Figure 6.15). Le pic observé autour de l'étape numéro 40 ne crée pas de soucis car il se produit simultanément avec le passage de la dynamique verrouillée à la dynamique libre. Le graphe des ligands se stabilise ensuite et reste stable pour le reste de la dynamique.

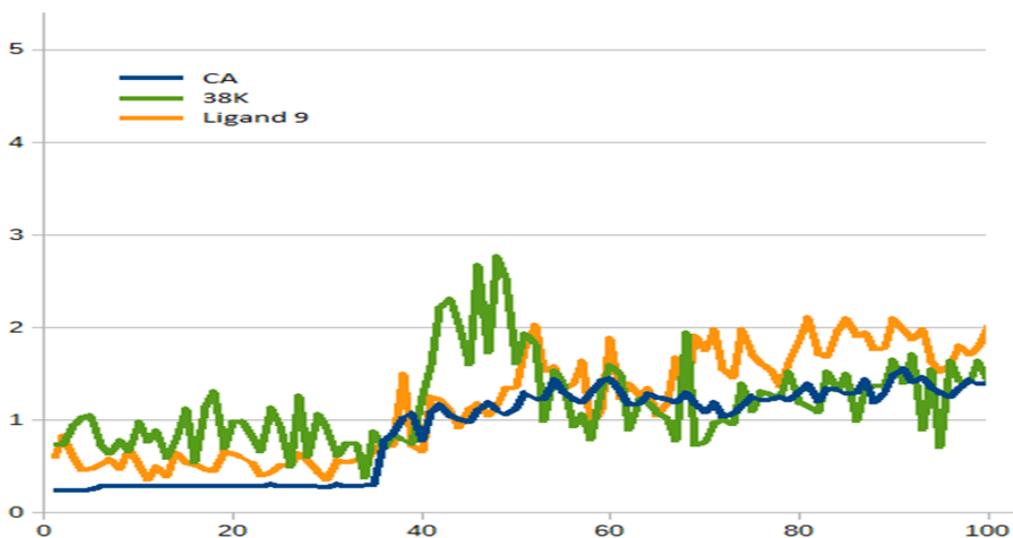


FIGURE 6.15 – Graphe du RMSD des carbonés α de la protéine (en bleu), du ligand de référence (en vert) et du ligand 9 (en orange).

TABLE 6.6 – Les interactions du ligand de référence pendant les différentes étapes de DM.

Liaison hydrogènes		
Accepteur	Donneur	N° d'étape de DM (frame)
Lig_333	ARG_99	95
GLY_97	Lig_333	73
ASP_52	Lig_333	52
Lig_333	ARG_112	40
ASP_52	Lig_333	37

Cette procédure a été considérée comme la meilleure et la plus facile à appliquer compte tenu du nombre de molécules à tester.

De plus, les données obtenues par le software *Cpptraj* sur les valeurs RMSD et sur les liaisons hydrogène, pour les deux ligands actifs, ont été prises en compte comme référence dans les analyses ultérieures.

TABLE 6.7 – Les interactions du ligand 9 pendant les différentes étapes de DM.

Liaison hydrogènes		
Accepteur	Donneur	N° d'étape de DM (frame)
Lig_333	TYR_83	96
Lig_333	ARG_99	70
GLY_97	Lig_333	65
ASP_52	Lig_333	44
ASP_52	Lig_333	36

6.5.2 Simulation dynamique moléculaire

Pour chacune des 164 molécules sélectionnées dans la phase du criblage virtuel, la même procédure considérée comme la meilleure dans la phase de calibrage a été appliquée.

L'étude pour chaque molécule était donc :

- Ajout des charges avec le programme Antichambre ;
- Création de complexe protéine et ligand ;
- Création d'un cube d'eau de 20 Å ;
- Équilibrage des charges par addition d'ions Cl⁻ ;
- La minimisation du complexe ;
- Dynamique moléculaire de 10 ns.

La procédure a été optimisée pour être lancée automatiquement à travers le script montré ici :

```

! /bin/csh -f
cd /home/moham/nfs/AMBER/docking_ff14
foreach code ('cat lista_maggiore7')
mkdir /home/marghe/nfs/AMBER/docking_ff14/code
cd /home/marghe/nfs/AMBER/docking_ff14/code
end
cd /home/moham/nfs/AMBER/dynamic_simulation_2
foreach code ('cat lista_nuovi_codici')
cd /home/moham/nfs/AMBER/dynamic_simulation_2/code
echo 'cd /home/moham/nfs/AMBER/dynamic_simulation_2/'code' » script_MD_mohamed
echo 'parmchk -i'code'_ac.mol2 -f mol2 -o frcmod_'code'' >> script_MD_mohamed
echo 'tleap -f tleap_shortcuts' >> script_MD_mohamed
echo 'source /usr/local/amber14/dat/leap/cmd/leaprc.ff14SB' >> tleap_shortcuts
echo 'source leaprc.gaff' >> tleap_shortcuts
echo 'fmod1 = loadamberparams frcmod_'code'' >> tleap_shortcuts
echo 'fmod2 = loadamberparams /usr/local/amber14/dat/leap/parm/frcmod.ionsjc,ip3p' >>
tleap_shortcuts
echo 'prot = loadpdb /home/moham/nfs/AMBER/dynamic_simulation_2/rcptr_4qsm.pdb' >>
tleap_shortcuts
echo 'lig = loadmol2'code'_ac.mol2' >> tleap_shortcuts
echo 'complex = combine prot lig' >> tleap_shortcuts
echo 'solvatebox complex TIP3PBOX 20' >> tleap_shortcuts
echo 'charge complex' >> tleap_shortcuts
echo 'addions complex Cl- 0' >> tleap_shortcuts
echo 'saveamberparm complex 'code'_dyn.prmtop'code'_dyn.inpcrd' >> tleap_shortcuts
echo 'quit' >> tleap_shortcuts
echo 'ambpdb -p 'code'_dyn.prmtop <'code'_dyn.inpcrd> 'code'_dyn.pdb' >>
script_MD_mohamed
echo 'setenv mp "/usr/local/amber14/bin/mpirun -n4"' >>
script_MD_mohamed
echo 'setenv sn "/usr/local/amber14/bin/sander.MPI"' >> script_MD_mohamed
echo 'mp sn -O -i/home/moham/nfs/AMBER/param/fix_min.in -omin1.out -
c'code'_dyn.inpcrd -p 'code'_dyn.prmtop -r'code'_dyn_1.rst -ref 'code'_dyn.inpcrd <
/dev/null' >> script_MD_mohamed
echo 'cd /home/moham/nfs/AMBER/dynamic_simulation_2/'code'' >> script_MD_run

```

```
echo `pmemd.cuda -O -i /home/moham/nfs/AMBER/param/dyn1.in -o dyn1.out -c `code'_dyn_1.rst -p`code'_dyn.prmtop -r `code'_dyn_2.rst -ref`code'_dyn_1.rst -xmdcrd_1_`code' < /dev/null` » script_MD_run
echo `pmemd.cuda -O -i /home/moham/nfs/AMBER/param/dyn2.in -o dyn2.out -c `code'_dyn_2.rst -p`code'_dyn.prmtop -r `code'_dyn_3.rst -ref`code'_dyn_2.rst -xmdcrd_2_`code' < /dev/null' >> script_MD_run
echo`pmemd.cuda -O -i/home/moham/nfs/AMBER/param/dyn3.in -o dyn3.out -c`code'_dyn_3.rst -p `code'_dyn.prmtop -r`code'_dyn_4.rst -ref `code'_dyn_3.rst -xmdcrd_3_`code' < /dev/null` » script_MD_run
end
```

6.5.3 Analyse des résultats

Les résultats ont été validés en utilisant deux critères :

- Calcul du RMSD moyen entre la pose du ligand pendant toute la dynamique et sa position initiale, c'est-à-dire celle proposée par le docking. Les valeurs de $RMSD \leq 2\text{Å}$ indiquent comment le ligand ne s'est pas éloigné du site catalytique et a maintenu assez fortement la conformation de la position initiale ; des valeurs RMSD moyennes plus élevées ont également été prises en compte, si elles sont soutenues par une tendance particulièrement stable ;
- Calcul des interactions entre le ligand et l'enzyme lors de la simulation DM ; Cette fonction permet d'identifier les étapes (frame) dont les interactions sont créées, maintenues ou perdues. Pour chaque liaison hydrogène renvoie un pourcentage moyen de la durée, la distance et l'angle entre les atomes. Les interactions stables et durables qui reflètent celles du ligand d'origine ont été récompensées.

On a basé sur ces deux critères, on a obtenu 42 molécules qui présentent une stabilité avec des interactions remarquables.

Il est également rapporté, à titre de comparaison, le graphe d'une molécule qui a été rejetée :

La ligne rouge indique la tendance RMSD de la molécule, il est évident qu'aucune stabilisation n'est obtenue dans la conformation qui continue à changer par rapport à la valeur d'origine. Dans les 20 dernières étapes de DM, nous pouvons également observer une fluctuation des valeurs RMSD des carbones-alpha de la protéine. Cette molécule a ensuite été jetée.



FIGURE 6.16 – Graphe de RMSD pour la molécule VS_f17_Z1127125184 pendant la DM.

6.5.4 Analyse visuelle

De plus, chaque DM a été analysé manuellement à l'aide du programme "UCSF CHIMERA", qui vous permet de visualiser l'animation de la dynamique à l'aide de la commande MD-Movie. De cette manière, il était possible de donner une évaluation plus approfondie de la procédure de dynamique.

La Figure 6.17 montre une tendance typique d'une DM, visualisée avec le software CHIMERA : Voici les graphes des 9 molécules les plus prometteuses. L'évolution

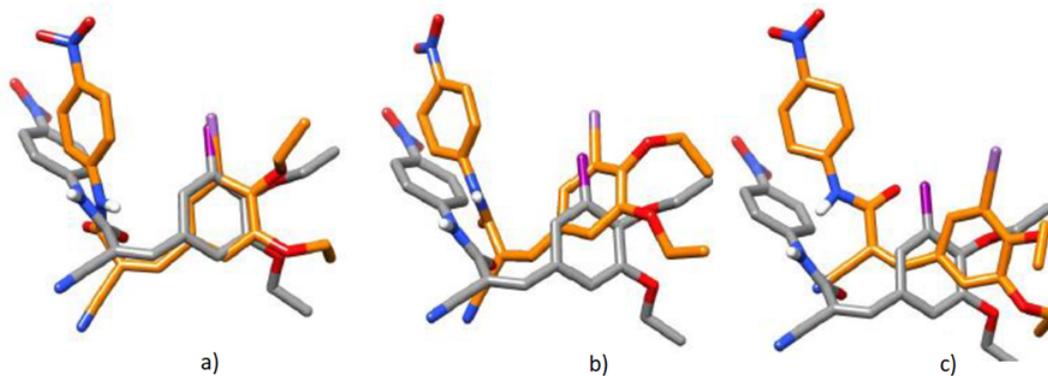


FIGURE 6.17 – Exemple d'une DM : en gris la conformation initiale, en orange les différentes étapes de la DM ; étape 1 (a), étape 28 (b), étape 100 (c).

des carbones-alpha de la structure protéique est similaire dans toutes les simulations de la dynamique. Dans les premiers 3,5 ns ils sont bloqués et stabilisés, suivis d'un pic qui correspond au début de la dynamique libre, puis reviennent à un état d'équilibre.

En général, la valeur RMSD moyenne des Carbones-alpha pour tous les complexes est d'environ 2 Å; ceci indique comment la protéine reste stable et ne s'éloigne pas de la position de départ malgré la liberté de mouvement.

Pour chaque graphe est également rapportée la valeur moyenne de RMSD pour les carbones-alpha et pour le ligand, les liaisons hydrogène les plus représentatives, le résidu contenant le donneur, le résidu contenant l'accepteur et le nombre d'étape dans lesquels *Cpptraj* a été calculée.

Les molécules ont été renommées avec le code suivant :

Molécules obtenues utilisant le virtuel screening à base du modèle pharmacophore numéro 17 :

VS_f17_(code Enamine database)

Le ligand VS_f17_Z1317967696 :

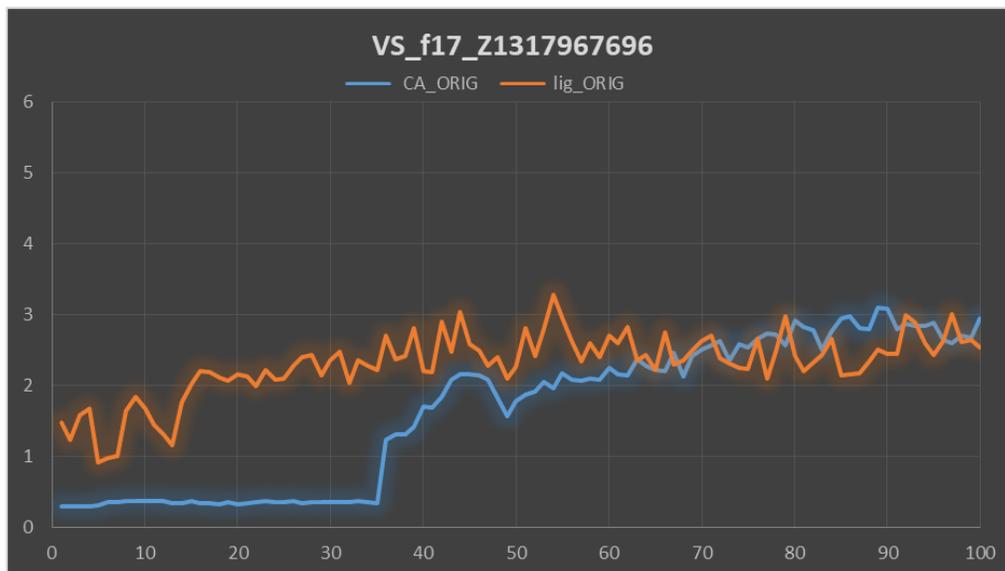


FIGURE 6.18 – Graphe de RMSD pour la molécule VS_f17_Z1317967696 pendant la DM.

TABLE 6.8 – Les interactions du ligand VS_f17_Z1317967696 pendant les différentes étapes de DM.

	CA_ORIG	Lig_ORIG
Moyenne du RMSD	1,643408	2,295383
Liaison hydrogènes		
Accepteur	Donneur	N° d'étape de DM (frame)
ASP_52	Lig_333	81
ASP_52	Lig_333	51
ASP_52	Lig_333	30
Lig_333	GLY_97	25
ASP_52	Lig_333	23

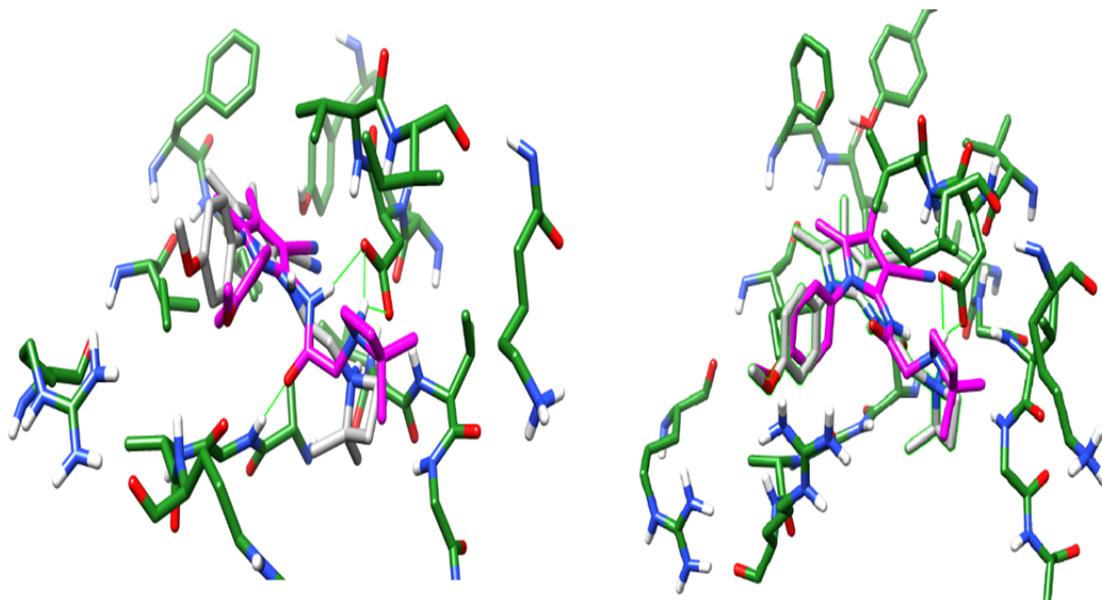


FIGURE 6.19 – DM VS_f17_Z1317967696; comparaison entre l'étape 1 (à gauche) et l'étape 100 (à droite). En gris la conformation initiale de référence; en violet la molécule pendant la dynamique.

Le ligand VS_f17_Z226451846 :

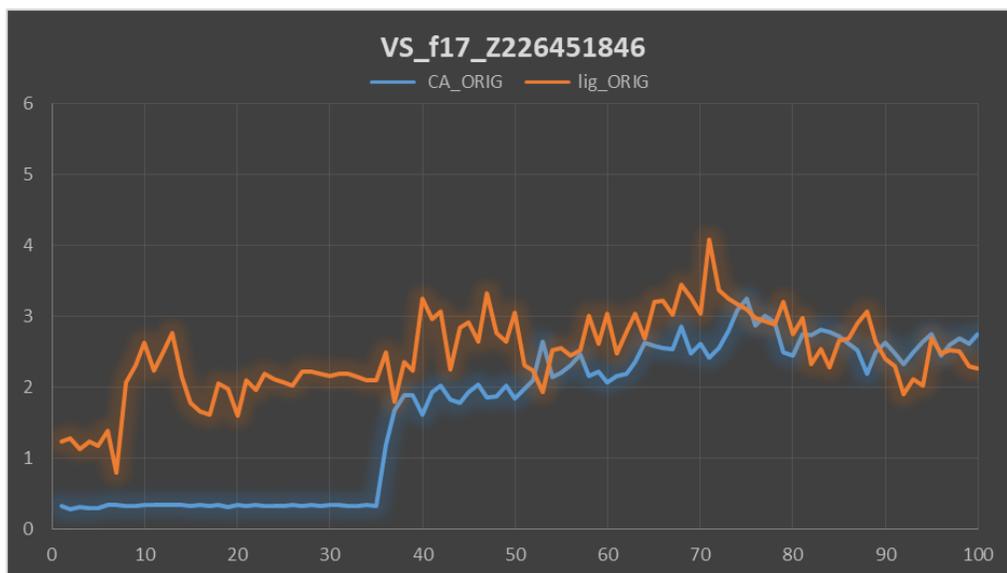


FIGURE 6.20 – Graphe de RMSD pour la molécule VS_f17_Z226451846 pendant la DM.

TABLE 6.9 – Les interactions du ligand VS_f17_Z226451846 pendant les différentes étapes de DM.

	CA_ORIG	Lig_ORIG
Moyenne du RMSD	1,653152	2,442208
Liaison hydrogènes		
Accepteur	Donneur	N° d'étape de DM (frame)
ASP_52	Lig_333	74
Lig_333	GLY_97	72
Lig_333	ARG_99	45
Lig_333	ARG_112	33
ASP_52	Lig_333	23

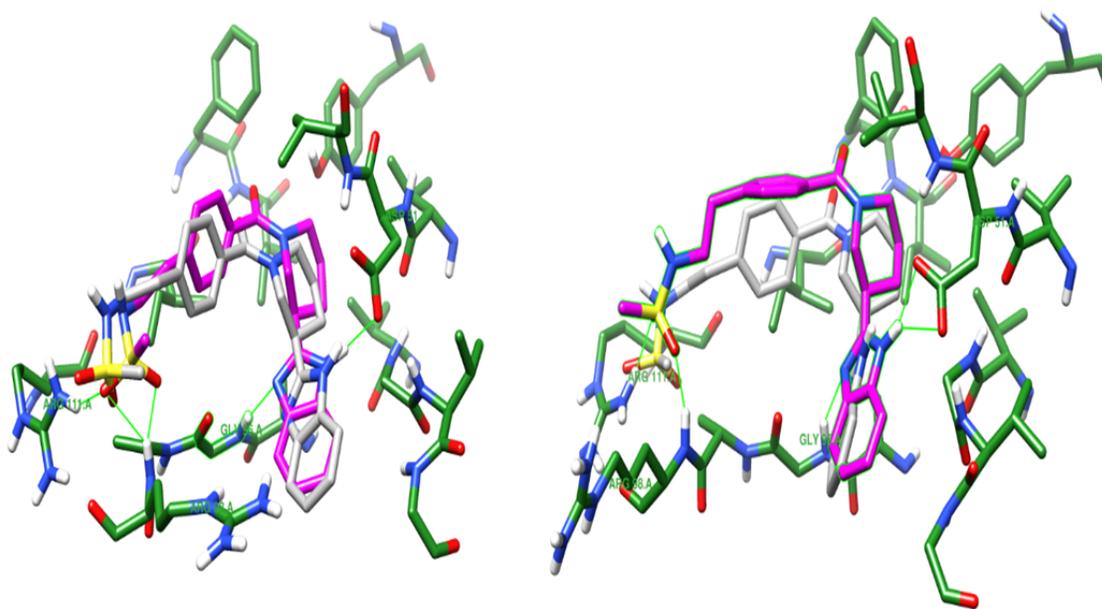


FIGURE 6.21 – DM VS_f17_Z226451846; comparaison entre l'étape 1 (à gauche) et l'étape 100 (à droite). En gris la conformation initiale de référence ; en violet la molécule pendant la dynamique.

Le ligand VS_f17_Z16507440 :



FIGURE 6.22 – Graphe de RMSD pour la molécule VS_f17_Z16507440 pendant la DM.

TABLE 6.10 – Les interactions du ligand VS_f17_Z16507440 pendant les différentes étapes de DM.

	CA_ORIG	Lig_ORIG
Moyenne du RMSD	1,62048	1,313699
Liaison hydrogènes		
Accepteur	Donneur	N° d'étape de DM (frame)
Lig33	TYR	72
Lig33	GLY7	70
Lig33	VAL1	56
ASP2	Li_333	28

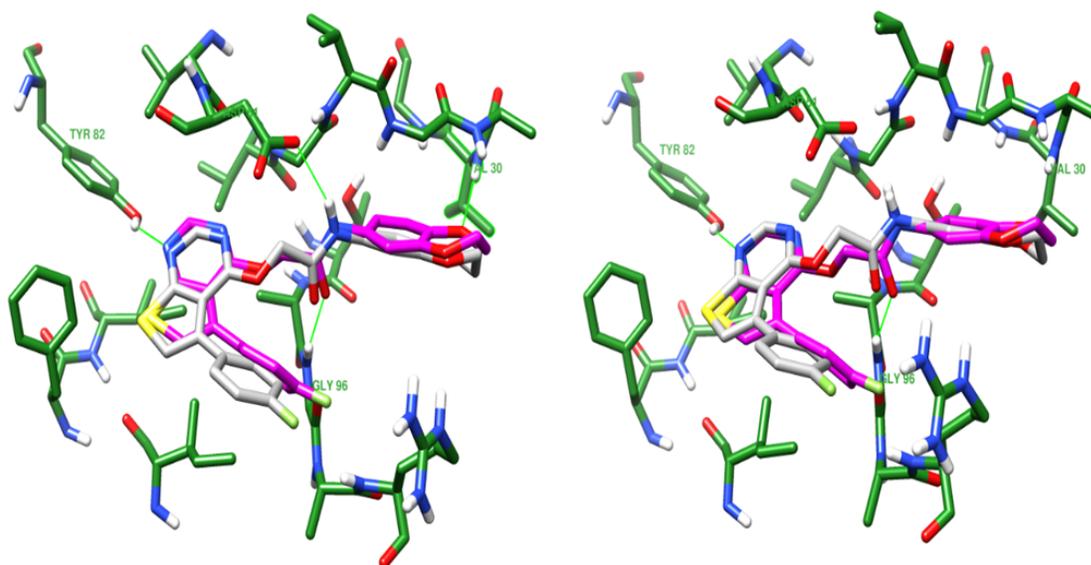


FIGURE 6.23 – DM VS_f17_Z16507440; comparaison entre l'étape 1 (à gauche) et l'étape 100 (à droite). En gris la conformation initiale de référence; en violet la molécule pendant la dynamique.

Le ligand VS_f17_Z16507442 :



FIGURE 6.24 – Graphe de RMSD pour la molécule VS_f17_Z16507442 pendant la DM.

TABLE 6.11 – Les interactions du ligand VS_f17_Z16507442 pendant les différentes étapes de DM.

	CA_ORIG	Lig_ORIG
Moyenne du RMSD	1,47871	1,294582
Liaison hydrogènes		
Accepteur	Donneur	N° d'étape de DM (frame)
Lig_333	GLY_97	85
Lig_333	TYR_83	80
ASP_52	Lig333	60

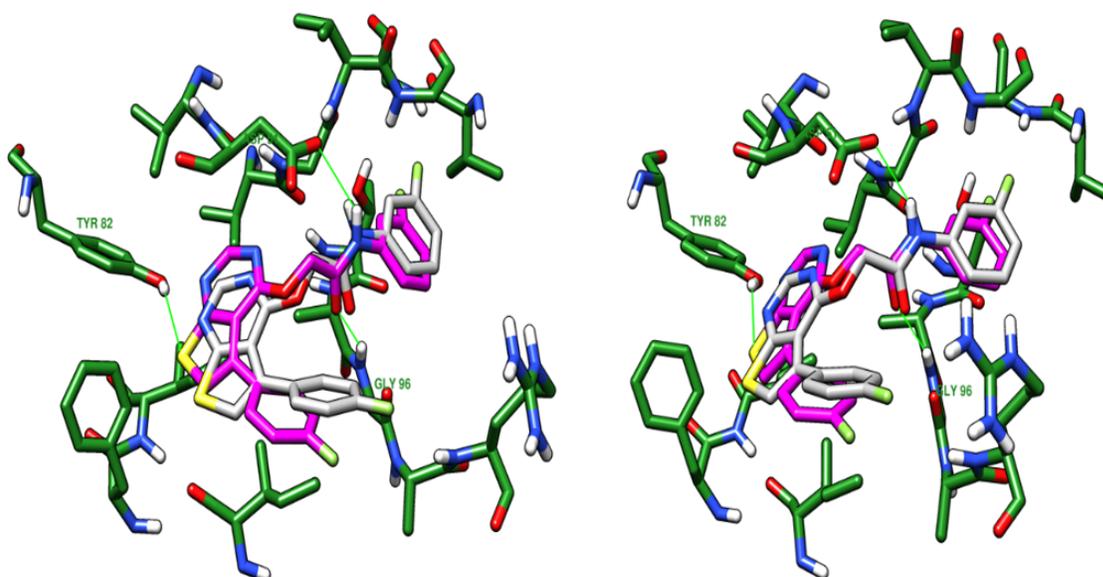


FIGURE 6.25 – DM VS_f17_Z16507442; comparaison entre l'étape 1 (à gauche) et l'étape 100 (à droite). En gris la conformation initiale de référence; en violet la molécule pendant la dynamique.

Le ligand VS_f17_Z16507998 :



FIGURE 6.26 – Graphe de RMSD pour la molécule VS_f17_Z16507998 pendant la DM.

TABLE 6.12 – Les interactions du ligand VS_f17_Z16507998 pendant les différentes étapes de DM.

	CA_ORIG	Lig_ORIG
Moyenne du RMSD	1,648489	1,974369
Liaison hydrogènes		
Accepteur	Donneur	N° d'étape de DM (frame)
Lig_333	GLY_97	73
Lig_333	TYR_83	72
ASP_52	Lig_333	50

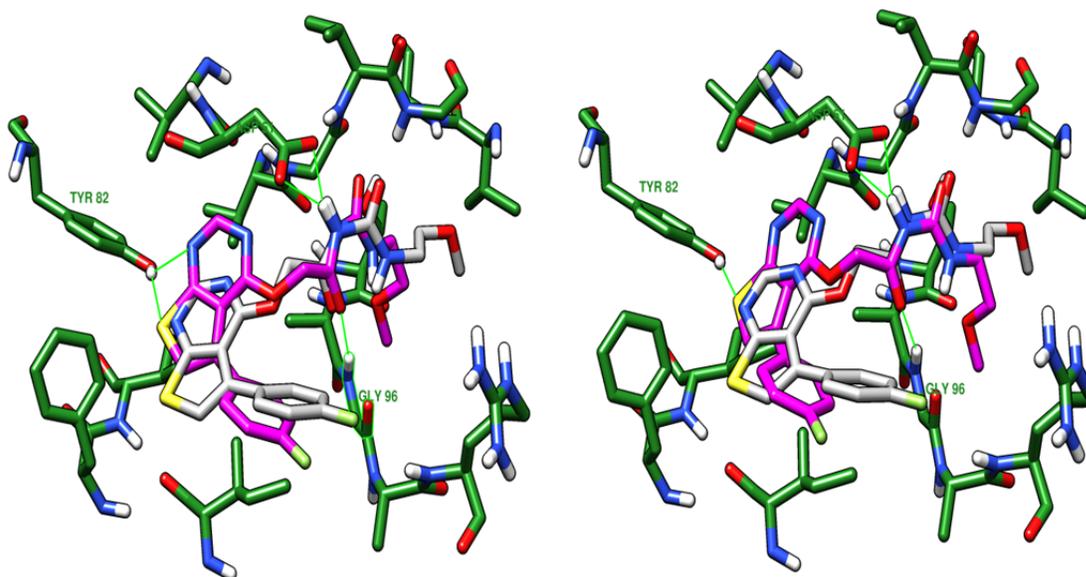


FIGURE 6.27 – DM VS_f17_Z16507998 ; comparaison entre l'étape 1 (à gauche) et l'étape 100 (à droite). En gris la conformation initiale de référence ; en violet la molécule pendant la dynamique.

Le ligand VS_f17_Z979261714 :



FIGURE 6.28 – Graphe de RMSD pour la molécule VS_f17_Z979261714 pendant la DM.

TABLE 6.13 – Les interactions du ligand VS_f17_Z979261714 pendant les différentes étapes de DM.

	CA_ORIG	Lig_ORIG
Moyenne du RMSD	1,764489	2.3624
Liaison hydrogènes		
Accepteur	Donneur	N° d'étape de DM (frame)
Lig_333	GLY_97	79
ASP_52	Lig_333	37

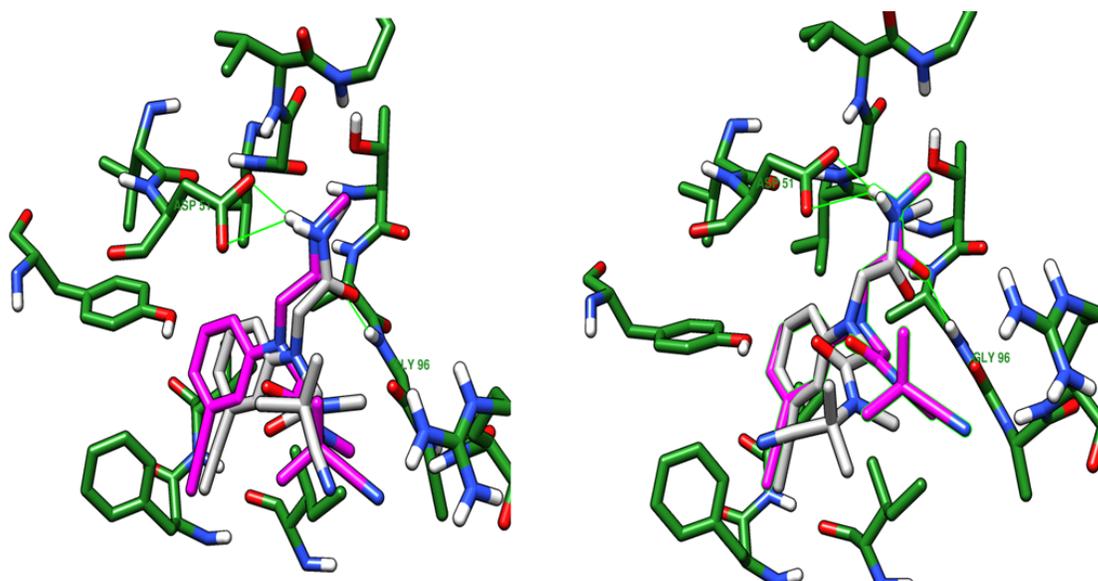


FIGURE 6.29 – DM VS_f17_Z979261714; comparaison entre l'étape 1 (à gauche) et l'étape 100 (à droite). En gris la conformation initiale de référence ; en violet la molécule pendant la dynamique.

Le ligand VS_f17_Z16508305 :

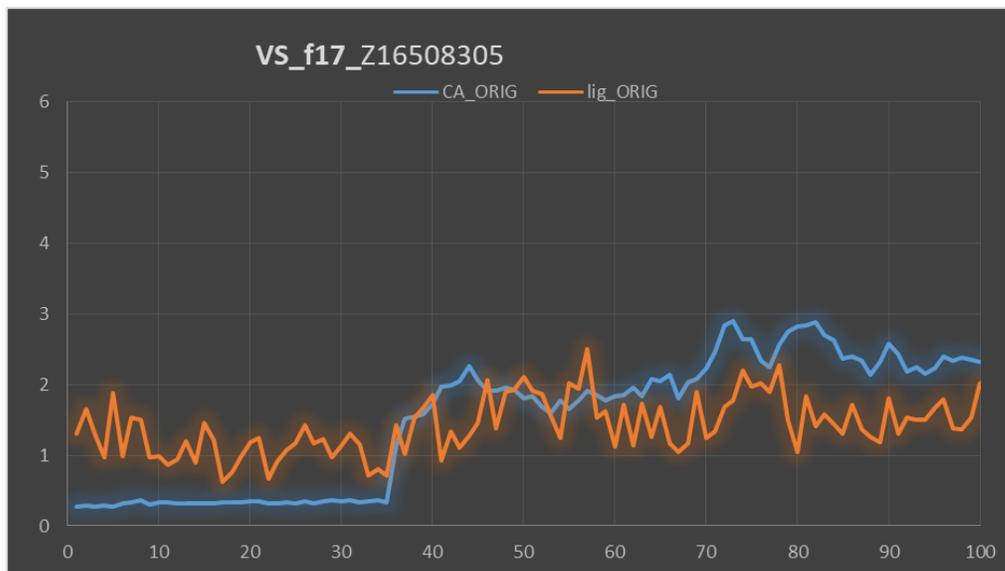


FIGURE 6.30 – Graphe de RMSD pour la molécule VS_f17_Z16508305 pendant la DM.

TABLE 6.14 – Les interactions du ligand VS_f17_Z16508305 pendant les différentes étapes de DM.

	CA_ORIG	Lig_ORIG
Moyenne du RMSD	1,508072	1,41251
Liaison hydrogènes		
Accepteur	Donneur	N° d'étape de DM (frame)
Lig_333	GLY_97	72
ASP_52	Lig_333	64

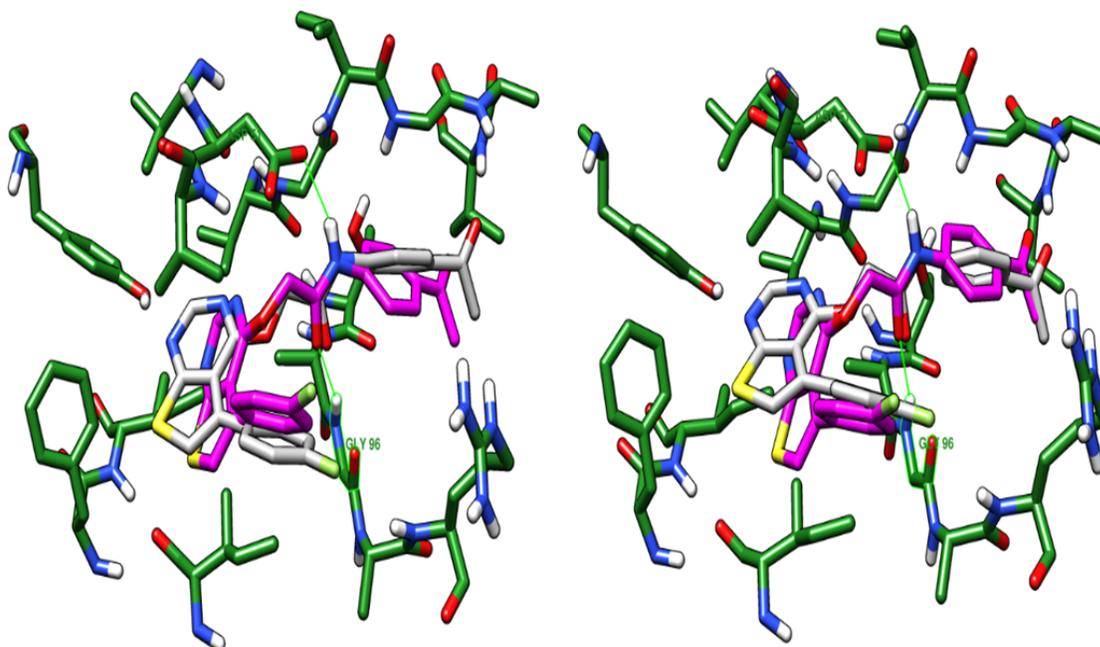


FIGURE 6.31 – DM VS_f17_Z16508305 ; comparaison entre l'étape 1 (à gauche) et l'étape 100 (à droite). En gris la conformation initiale de référence ; en violet la molécule pendant la dynamique.

Le ligand VS_f17_Z358317798 :



FIGURE 6.32 – Graphe de RMSD pour la molécule VS_f17_Z358317798 pendant la DM.

TABLE 6.15 – Les interactions du ligand VS_f17_Z358317798 pendant les différentes étapes de DM.

	CA_ORIG	Lig_ORIG
Moyenne du RMSD	1,51812	1,282947
Liaison hydrogènes		
Accepteur	Donneur	N° d'étape de DM (frame)
Lig_333	ARG_99	99
Lig_333	ARG_112	93
Lig_333	Gly_97	75
ASP_52	Lig_333	51

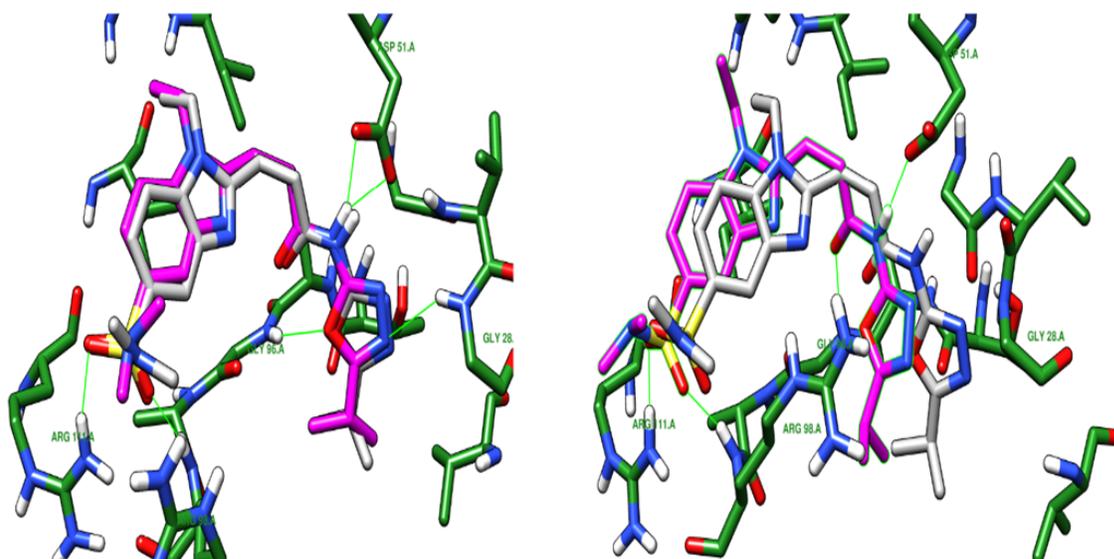


FIGURE 6.33 – DM VS_f17_Z358317798; comparaison entre l'étape 1 (à gauche) et l'étape 100 (à droite). En gris la conformation initiale de référence; en violet la molécule pendant la dynamique.

Le ligand VS_f17_Z17766169 :



FIGURE 6.34 – Graphe de RMSD pour la molécule VS_f17_Z17766169 pendant la DM.

TABLE 6.16 – Les interactions du ligand VS_f17_Z17766169 pendant les différentes étapes de DM.

	CA_ORIG	Lig_ORIG
Moyenne du RMSD	1,51812	1,282947
Liaison hydrogènes		
Accepteur	Donneur	N° d'étape de DM (frame)
Lig_333	Gly_97	75
Lig_333	ARG_99	68
Lig_333	ARG_112	63
ASP_52	Lig_333	51

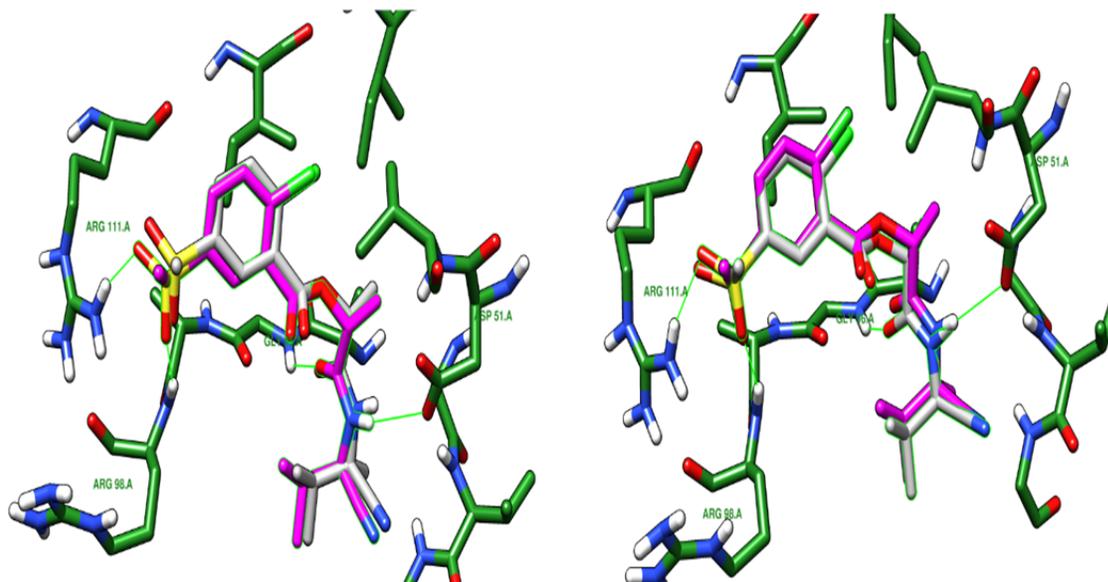


FIGURE 6.35 – DM VS_f17_Z17766169; comparaison entre l'étape 1 (à gauche) et l'étape 100 (à droite). En gris la conformation initiale de référence ; en violet la molécule pendant la dynamique.

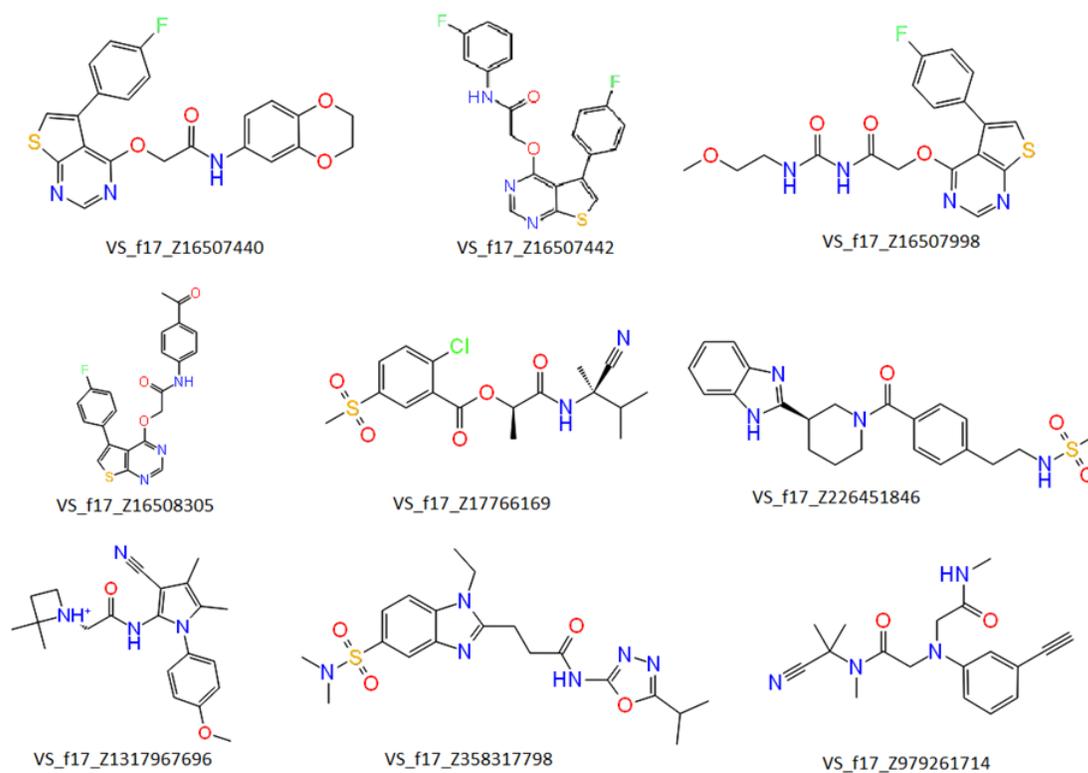


FIGURE 6.36 – La structure développée des 9 molécules stables après la DM.

6.6 Conclusion

Ce chapitre de thèse avait pour but le développement in silico de nouvelles structures hypothétiques agissant comme inhibiteurs très puissants de LDH dans le domaine de la chimiothérapie cancéreuse en basant sur la structure cristallographique obtenu en 2014 (code PDB : 4QSM) en interaction avec le ligand actif (ligand 9) obtenu en 2016. L'inhibiteur cristallisé 38K et le ligand 9, présentent une conformation intéressante au sein du site catalytique en interactions directe avec les résidus de récepteurs. Dans cette optique, nous avons cherché à développer une nouvelle procédure de VS en basant sur les interactions de ces deux ligands. La base de données Enamine, qui contient 2 000.000 composés, a été utilisée pour le VS. Dans cette partie de mon projet, j'ai utilisé le programme LigandScout pour construire un modèle pharmacophore qui prenait en compte les interactions les plus importantes des deux ligands. Sur les 73218 molécules obtenues à travers le filtre utilisant le modèle pharmacophore, 360 ont été sélectionnées pour des études de Docking utilisant le programme Glide_XP. Des études de DM sur 164 molécules ainsi obtenues ont permis d'éliminer les ligands non stables, conduisant éventuellement à l'obtention de 9 molécules comme inhibiteurs de lactate déshydrogénase. 5 molécules ont été considérées comme particulièrement prometteuses et seront achetées pour être soumises à des dosages enzymatiques. Si les résultats des dosages enzymatiques sont positifs, ces molécules peuvent être utilisées comme nouvelles pistes pour le développement d'inhibiteurs puissants de lactate déshydrogénase.

Bibliographie

- [1] Wei Cui, Wei Lv, Ying Qu, Rui Ma, Yi-Wei Wang, Yong-Jun Xu, Di Wu, and Xuanhuang Chen. Discovery of 2-((3-cyanopyridin-2-yl) thio) acetamides as human lactate dehydrogenase a inhibitors to reduce the growth of mg-63 osteosarcoma cells : Virtual screening and biological validation. *Bioorganic & medicinal chemistry letters*, 26(16) :3984–3987, 2016.
- [2] Subramaniapillai Kolappan, David L Shen, Renee Mosi, Jianyu Sun, Ernest J McEachern, David J Vocadlo, and Lisa Craig. Structures of lactate dehydrogenase a (ldha) in apo, ternary and inhibitor-bound forms. *Acta Crystallographica Section D : Biological Crystallography*, 71(2) :185–195, 2015.
- [3] Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, Talapady N Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. The protein data bank, 1999–. In *International Tables for Crystallography Volume F : Crystallography of biological macromolecules*, pages 675–684. Springer, 2006.
- [4] AN Shivanyuk, SV Ryabukhin, AV Bogolyubsky, DM Mykytenko, AA Chupryna, W Heilman, AN Kostyuk, and A Tolmachev. Enamine real database : making chemical diversity real. *Chemistry Today*, 25 :58–59, 2007.
- [5] Schrödinger Release. 3 : Ligprep. *Schrödinger, LLC, New York, NY*, 2016.
- [6] Thomas A Halgren, Robert B Murphy, Richard A Friesner, Hege S Beard, Leah L Frye, W Thomas Pollard, and Jay L Banks. Glide : a new approach for rapid, accurate docking and scoring. 2. enrichment factors in database screening. *Journal of medicinal chemistry*, 47(7) :1750–1759, 2004.
- [7] JM Haile. *Molecular dynamics simulation*, volume 18. Wiley, New York, 1992.
- [8] Jacob D Durrant and J Andrew McCammon. Molecular dynamics simulations and drug discovery. *BMC biology*, 9(1) :71, 2011.
- [9] Yinglong Miao, Victoria A Feher, and J Andrew McCammon. Gaussian accelerated molecular dynamics : Unconstrained enhanced sampling and free energy calculation. *Journal of chemical theory and computation*, 11(8) :3584–3595, 2015.

Conclusions générales et perspectives

Notre manuscrit s'articule autour de deux thèmes, le travail est présenté en deux parties consacrées à la modélisation moléculaire (QSAR et VS) à la base du modèle pharmacophore).

Au cours de la première partie, l'objectif a été de modéliser l'activité antibactérienne pour former un modèle QSAR robuste, stable, et précis capable de prédire l'activité biologique par rapport à la structure des dérivés des quinolones.

Parmi les agents antimicrobiens, la classe des quinolones suscite un intérêt considérable depuis sa découverte. Avec un spectre d'activité antibactérienne étendu et des propriétés pharmacocinétiques intéressantes, les quinolones représentent une des principales classes d'agents antimicrobiens actuels.

Pour cela nous avons, dans un premier temps, mené notre étude à partir d'une bibliothèque de 85 molécules dérivées de quinolones, qui ont montré une activité inhibitrice vis-à-vis la bactérie *E. coli* afin de modéliser l'activité bactérienne exprimée par log CMI. Un modèle a été construit en utilisant la méthode des algorithmes génétiques comme méthode de sélection. Une méthodologie basée sur la régression linéaire multiple (MLR), implémentée dans le logiciel XIStat a été utilisée pour la construction du modèle optimal. Le modèle est obtenu en utilisant la totalité des descripteurs issus du logiciel DRAGON 6.

Le modèle QSAR développé est simple, interprétable et transparent en utilisant un nombre réduit de descripteurs. En outre, il est caractérisé par la stabilité, la robustesse et le pouvoir prédictif élevé vérifié par la validation interne et externe. Ainsi, le modèle est considéré comme validé et applicable pour l'exploitation de la base de données.

Les perspectives de cette partie de travail nous semblent diverses. D'une part, nous avons l'intention de reprendre les mêmes bases de données et élaborer des modèles en utilisant d'autres méthodes d'analyse de données telle que PLS, PCA, NN, SVM,... D'autre part, nous projetons d'élaborer des modèles QSAR pour d'autres bases de données de molécules antibactériennes vis-à-vis d'autres espèces microbiennes et vis-à-vis de l'être humain (cytotoxicité, mutagénèse) en utilisant les descripteurs physico-chimique, et pourquoi pas les descripteurs 3D (CoMFA, CoMSIA).

Au cours de la deuxième partie présentée dans ce manuscrit, nous avons exploré plusieurs outils et stratégies de criblage virtuel (structure-based) pour leur application dans le cadre de la découverte de nouveaux anticancéreux comme des inhibiteurs du lactate déshydrogénase.

La lactate déshydrogénase (LDH) est une enzyme d'oxydoréductase capable de catalyser la conversion du pyruvate en lactate et vice-versa. Plusieurs études ont mis en évidence l'implication directe de cette enzyme dans la genèse et la prolifération de certains types de tumeurs (effet Warburg), et constitue donc une cible attractive dans le domaine de l'oncologie. L'effet Warburg a été très étudié mais sa nature précise reste floue, ce qui freine l'émergence des projets ayant pour but d'explorer son potentiel thérapeutique, c'est pourquoi nous adoptons sur le Virtuel Screening pour trouver des solutions.

Pour cela la chimiothèque commercial Enamine qui contient plus de 2 000.000 composés a été sélectionnée pour cette thématique de recherche. Sur la base des interactions des deux composés référence actifs (38K et composé 9), un modèle pharmacophore a été développé utilisant le software Ligandscout. Le modèle pharmacophore qui contient 8 interactions avec le récepteur a été proposé pour la prochaine étape de Virtuel Screening, 3 interactions considérée comme obligatoires, et les autres 5 interactions comme optionnelles.

73 218 molécules ont été obtenues après le filtrage à base du modèle pharmacophore, des molécules avec 5 interactions au minimum et 8 interactions au maximum.

La prochaine étape consistera à faire un docking dans le site catalytique, utilisant le software Glide_XP. Nous avons sélectionné 360 molécules "Top_Scoring" avec des valeurs du RMSD inférieur à 3 Å°. Avant la dernière étape qui consiste à faire une simulation avec la dynamique moléculaire, une analyse visuelle doit être réalisée pour les 360 molécules avec le modèle pharmacophore utilisant le software Chimera, dans le but de mieux explorer les interactions des molécules, ainsi le nombre est encore réduit jusqu'à 164 molécules.

Les études de dynamique moléculaire sur les 164 molécules obtenues ont permis d'éliminer les ligands non stables, conduisant éventuellement à l'obtention de 42 molécules comme inhibiteurs du lactate déshydrogénase. 9 molécules ont été considérées comme particulièrement prometteuses et seront achetées pour être soumises à des dosages enzymatiques. Ces molécules peuvent être utilisées comme nouvelles pistes pour le développement d'inhibiteurs puissants du lactate déshydrogénase si les résultats des essais biologiques sont intéressants.

Annexe

6.7 Les descripteurs constitutionnels

Les descripteurs constitutionnels sont les plus simples, d'un point de vue conceptuel et du coût calculatoire, et ils sont très couramment utilisés. Ils reflètent la composition chimique d'un composé sans aucune information sur sa géométrie moléculaire ni sur sa structure électronique. Dragon calcule 43 descripteurs constitutionnels (Tableau 6.17), beaucoup d'entre eux sont bien expliqués par leur définition tels que : le poids moléculaire (MW), le nombre d'atomes (nAT), le nombre de liaison (nBT), etc.

TABLE 6.17: Liste des descripteurs constitutionnels générés par le logiciel Dragon.

N°	Nom	Description
1	MW	molecularweight
2	AMW	avagemolecularweight
3	Sv	sum of atomic van der Waals volumes (scaled on Carbon atom)
4	Se	sum of atomic Sanderson electronegativities (scaled on Carbon atom)
5	Sp	sum of atomic polarizabilities (scaled on Carbon atom)
6	Si	sum of first ionization potentials (scaled on Carbon atom)
7	Mv	mean atomic van der Waals volume (scaled on Carbon atom)
8	Me	mean atomic Sanderson electronegativity (scaled on Carbon atom)
9	Mp	mean atomic polarizability (scaled on Carbon atom)
10	Mi	mean first ionization potential (scaled on Carbon atom)
11	nAT	number of atoms
12	nSK	number of non-H atoms
13	nBT	number of bonds
14	nBO	number of non-H bonds
15	nBM	number of multiple bonds
16	SCBO	sum of conventional bond orders (H-depleted)
17	RBN	number of rotatable bonds
18	RBF	rotatable bond fraction
19	Ndb	number of double bonds
20	nTB	number of triple bonds
21	nAB	number of aromatic bonds
22	nH	number of Hydrogenatoms
23	nC	number of Carbonatoms
24	nN	number of Nitrogenatoms

25	nO	number of Oxygenatoms
26	nP	number of Phosphorousatoms
27	nS	number of Sulfuratoms
28	nF	number of Fluorine atoms
29	nCL	number of Chlorineatoms
30	nBR	number of Bromineatoms
31	nI	number of Bromineatoms
32	nB	number of Boronatoms
33	nHM	number of heavyatoms
34	nHet	number of heteroatoms
35	nX	number of halogenatoms
36	H%	percentage of H atoms
37	C%	percentage of C atoms
38	N%	percentage of N atoms
39	O%	percentage of O atoms
40	X%	percentage of halogenatoms
41	nCsp3	number of sp3 hybridized Carbon atoms
42	nCsp2	number of sp2 hybridized Carbon atoms
43	nCsp	number of sp hybridized Carbon atoms

Ce type de descripteurs est employé pour obtenir les modèles QSAR les plus simples, mais dans certains cas ils sont défailants car ils ne permettent pas la proposition de mécanismes d'interactions mises en jeu pour l'activité étudiée. Par exemple, ils ne permettent pas de distinguer les isomères. Or, pour la majorité des cas, la position d'un substituant modifie la valeur expérimentale de l'activité.

6.8 Descripteurs de groupements fonctionnels

Ce sont des descripteurs moléculaires simples, définis comme étant le nombre de groupes fonctionnels spécifiques dans une molécule. Ils sont calculés sur la base de la composition moléculaire et des connectivités atomiques. Dragon calcule 154 groupes fonctionnels. Le Tableau 6.18 rassemble quelques descripteurs impliqués dans le bloc des groupements fonctionnels.

TABLE 6.18: Présentation de quelques descripteurs moléculaires du bloc de groupements fonctionnels.

N°	Symbole	Structure	Définition
1	nArOCN	$\text{Ar}-\text{O}-\text{C}\equiv\text{N}$	Nombre de cyanates (aromatique)
2	nArNCO	$\text{Ar}-\text{N}=\text{C}=\text{O}$	Nombre d'isocyanates (aromatique)
3	nArCONH ₂	$\begin{array}{c} \text{H}_2\text{N} \\ \\ \text{Ar}-\text{C}=\text{O} \end{array}$	Nombre d'amides primaires (aromatique)
4	nRCSOH	$\begin{array}{c} \text{HO} \quad \text{HS} \\ \quad \\ \text{Al}-\text{C}=\text{S} \quad \text{Al}-\text{C}=\text{O} \end{array}$	Nombre de thioacides (aliphatique)
5	nRCO	$\begin{array}{c} \text{Al} \\ \\ \text{Al}-\text{C}=\text{O} \end{array}$	Nombre de cétones (aliphatique)
6	nRNHO	$\begin{array}{c} \text{Al} \\ \\ \text{Al}-\text{N}-\text{O}-\text{Y} \end{array}$	Nombre d'hydroxylamines (aliphatique) Y = H, Al ou Ar
7	nRNO ₂	$\begin{array}{c} \text{O} \\ \\ \text{Al}-\text{N} \\ \\ \text{O} \end{array}$	Nombre de groupes nitro (aliphatique) Al= H ou groupe aliphatique lié avec un carbone
8	nO(C=O) ₂	$\begin{array}{c} \text{Y} \quad \text{Y} \\ \quad \\ -\text{C}-\text{O}-\text{C}- \end{array}$	Nombre d'anhydrides (thio-) Y = O ou S
9	nH ₂ O	H-O-H	Nombre de molécules d'eau
10	nSH	$\begin{array}{c} \\ \text{H}-\text{C}-\text{S}-\text{H} \\ \end{array}$	Nombre de thiols
11	nCHR ₂ X	$\begin{array}{c} \text{X} \\ \\ \text{C}-\text{C}-\text{H} \\ \\ \text{C} \end{array}$	Nombre de CHR ₂ X
12	nRCX	X-C≡C	Nombre de RCX
13	nHBonds	$\begin{array}{c} \text{H} \quad \text{H} \\ \quad \\ \text{Y}_1-\text{C}-\text{C}-\text{Y}_2 \\ \quad \\ \text{Y}_1 \quad \text{Y}_2 \end{array}$	Nombre de liaisons H intramoléculaires Y ₁ = B, N, O, Al, P, S. Y ₂ = N, O, F

Les lettres X, b, Ar, r, = et représentent respectivement, dans les symboles des descripteurs moléculaires, un halogène, un benzène, un groupe aromatique lié par un carbone du groupement, tout cycle non aromatique, des doubles et triples liaisons.

6.9 Les descripteurs topologiques

Les descripteurs topologiques sont inspirés de la théorie des graphes, et sont des indices obtenus à partir d'une structure 2D de la molécule, ils donnent des informations sur la taille globale du système, sa forme globale et ses ramifications.

Les descripteurs topologiques sont basés sur une représentation graphique de la molécule dont le système moléculaire est considéré comme un graphe $G[V,R]$ dans lequel les atomes sont des sommets V et les liaisons chimiques des arêtes R . Ils peuvent être sensibles à une ou plusieurs caractéristiques structurales de la molécule comme la taille, la forme, la symétrie, la ramification et la cyclicité, et peuvent également coder l'information chimique concernant les types d'atomes et la multiplicité des liaisons.

La distance topologique d entre deux atomes est définie comme le nombre de liaisons minimum reliant ces atomes. L'indice de Wiener par exemple, permet de caractériser le volume moléculaire et le taux de ramification des molécules. Il est calculé comme la somme de toutes les distances topologiques entre atomes au sein de la molécule.

$$\sum_{i,j} d_{ij} \quad (6.2)$$

Le Tableau 6.19 rassemble quelques descripteurs topologiques issus du logiciel Dragon.

TABLE 6.19: Quelques descripteurs topologiques fournis par logiciel Dragon.

N°	Nom	Description
1	Xt	total structure connectivity index
2	Ram	ramification index
3	Pol	polaritynumber
4	MSD	mean square distance index (Balaban)
5	MDDD	mean distance degreedeviation
6	UNIP	unipolarity
7	SMTIV	Schultz Molecular
8	GMTI	GutmanMolecular
9	MAXDN	maximal electro topological negative variation
10	MAXDP	maximal electro topological positive variation
11	DELS	molecular electro topological variation
12	TIE	E-state topologicalparameter

13	Psi_e_A	electro topological state pseudo connectivity index - type S average
14	Psi_e_0	electrotopological state pseudo connectivity index - type 0
15	Psi_e_1	electrotopological state pseudo connectivity index - type 1
16	Psi_e_tel	electro topological state pseudo connectivity index - type T
17	Psi_e_0d	electro topological state pseudo connectivity index - type 0d
18	Psi_e_1d	electro topological state pseudo connectivity index - type 1d
19	Psi_e_1s	electrotopological state pseudo connectivity index - type 1s
20	BAC	Balaban centric index
21	LOC	looping centric index

6.10 Les descripteurs géométriques

Les descripteurs géométriques sont issus de la structure tridimensionnelle de la molécule, et sont évalués à partir des positions relatives des atomes d'une molécule dans l'espace, ainsi que des rayons et des masses atomiques. En général, les descripteurs géométriques sont calculés soit à partir de la géométrie moléculaire optimisée par modélisation moléculaire soit à partir des coordonnées cristallographiques. Par conséquent, ils fournissent plus d'informations sur la structure moléculaire par rapport aux descripteurs topologiques. Cependant, ils sont plus coûteux en calcul. Les descripteurs géométriques calculés par Dragon sont divisés en sous-blocs tels que les indices de taille et les indices de forme. Parmi ces descripteurs on trouve le volume et la surface moléculaires, le moment d'inertie ou encore les distances et les angles dièdres entre atomes dans la molécule.

article

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Computational Study of Quinolone's Antibacterial Activity Using QSAR Approach.

Mohamed Aissaoui^{a*}, Sofiane Ali Rachedi^a, Ahmed Mostefa Mokhnane^b,
Habiba Louzim^b, and Abdelhafid Djerourou^a

^aOrganic Synthesis and Biocatalysis Laboratory, Faculty of Science, Chemistry Department, P.B.: 12, 23000, Annaba, Algeria.

^bChemistry Therapeutic Group, Organic Synthesis and Biocatalysis Laboratory, Faculty of Medicine, Pharmacy Department, P.B.: 12, 23000, Annaba, Algeria.

ABSTRACT

The urinary tract infection is a common cause of consultation and hospitalization in Urology. Quinolones are used in second line treatment of these infections. The resistance developed by the bacteria against these molecules is evolving. To evaluate this resistance, we have processed the files of 1506 patients hospitalized at Urology Service of the University Hospital center of Annaba, Algeria. In an attempt to achieve a better understanding of the parameters controlling the bacterial properties against gram negative *E. coli* and linking chemical activities with molecular structures and compositions, a QSAR study was undertaken. In the present study, 85 quinolone derivatives (Data set), were evaluated as antibacterial agents, expressed by the minimal inhibition concentration (MIC) of these compounds against *E. coli*. A linear QSAR model was developed using Multiple Linear Regression technique, while Genetic algorithm was adopted for selecting the most appropriate descriptors. The predictive ability of the Model proposed was experimentally validated using the external validation set and the Y-randomization technique. The QSAR model developed in this study will be helpful to design new potent quinolone derivatives.

Keywords: QSAR, Quinolone Derivatives, Minimum Inhibitory Concentration (MIC), *E. coli*, Genetic Algorithm, Multiple Linear Regression.

*Corresponding author

INTRODUCTION

It's been over seventy (70) years that antibiotics have been one of the greatest medical advances making them indispensable. [1, 2] Nowadays, bacterial resistance to antibiotics is both a present reality and a threat for the future.[3] Its involvement in morbidity and mortality makes the therapeutic choices more complex, affecting seriously the quality of medical care. It reduces our therapeutic range without giving the pharmaceutical researches the time to respond to the new needs with new more active products. Therefore, control of bacterial resistance to antibiotics is a major health issue for our country. Although actions have been taken for several years in both the monitoring of resistance and the prevention of the transmission of resistant bacteria in health facilities in order to promote better use of antibiotics. This work started with an observation in Urology Service of the University Hospital Center at Annaba (Algeria) of bacterial resistance to quinolones.

This study was carried out from 1 January 2011 to 31 December 2013, including 1506 patients. The microbiological exam of urine cultures showed that about 26% of patients are positive as shown in (Figure 1). *E. coli* is the most incriminated germ being responsible for about 3/4 of the positive results of urinalysis, other bacteria were also isolated, as shown in Figure 1.

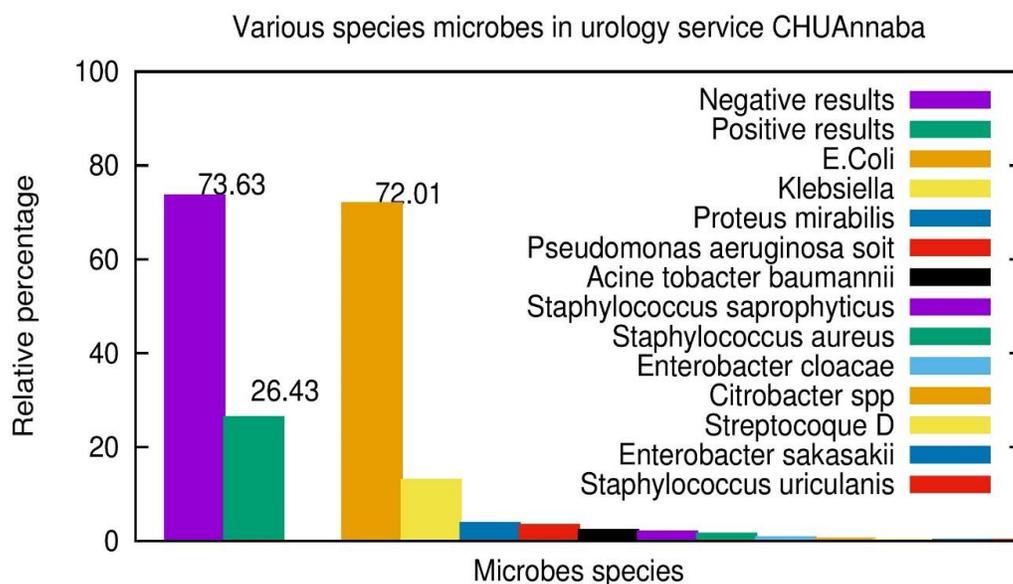


Figure 1 The percentage of microbiological exam of urine.

As for the resistance to quinolones a worrying development was noted for Ofloxacin®: passing from 23% in 2011 to 30% in 2013. Another quinolone: The Levofloxacin® seems more active with a resistance rate of 10 % in 2011, slightly increased to 13 % in 2013. [3, 4] Hence, conclusions were drawn; the bacterial resistance developed in this service is increasing, mainly for *E. coli*. [5, 6] Quinolones of different generations, however, remain effective in these urinary tract infections without confirming how long this effectiveness is affirmed, Ofloxacin® and Levofloxacin® are stereoisomers with different resistance levels, suggest that even small structural variations affect the antibiotic activity of quinolones.[7, 8].

Based on these data, and remaining in the same vein, the QSAR study we conducted is interested to quinolones following the establishment of a relationship between their structure and their activities.[9] The concept of structure-activity relationship (QSAR) used for several years in the designing new drugs and predicting their therapeutics activities even before their synthesis. This prediction although not total, has the advantage of saving a part of the initial testing; thus, the structure activity relationship has become an integral part in the modeling of biological activity. From this finding, recent works involving toxicologists, chemists and computer scientist have shown that compliance with a number of simple rules is possible to obtain comparable results to those obtained during the development of therapeutics molecules.

This article presents at first the attempts to set out the basic rules that must be respected in this approach. Then, an application example is given with the limits of the prediction obtained when using a QSAR. This example treats the minimum inhibitory concentration (MIC) of a set of quinolone molecules against the most isolated germ in the study previously mentioned *E. coli*.

MATERIALS AND METHODS

QSAR Study

Select data set and biological data

In an attempt to achieve a better understanding of the parameters controlling the antibacterial properties against the gram negative (*E. coli*), a QSAR study was realised [10–14].

The present QSAR study is limited to the properties exhibited against *E. coli* due to the incredible increase in the resistance of these bacteria.[15, 16] The data set contains 85 quinolone derivatives with an antibacterial activity against *E. coli*. Their molecular structure and activity in vitro are listed in figures 2 and table 2 respectively.

Training set

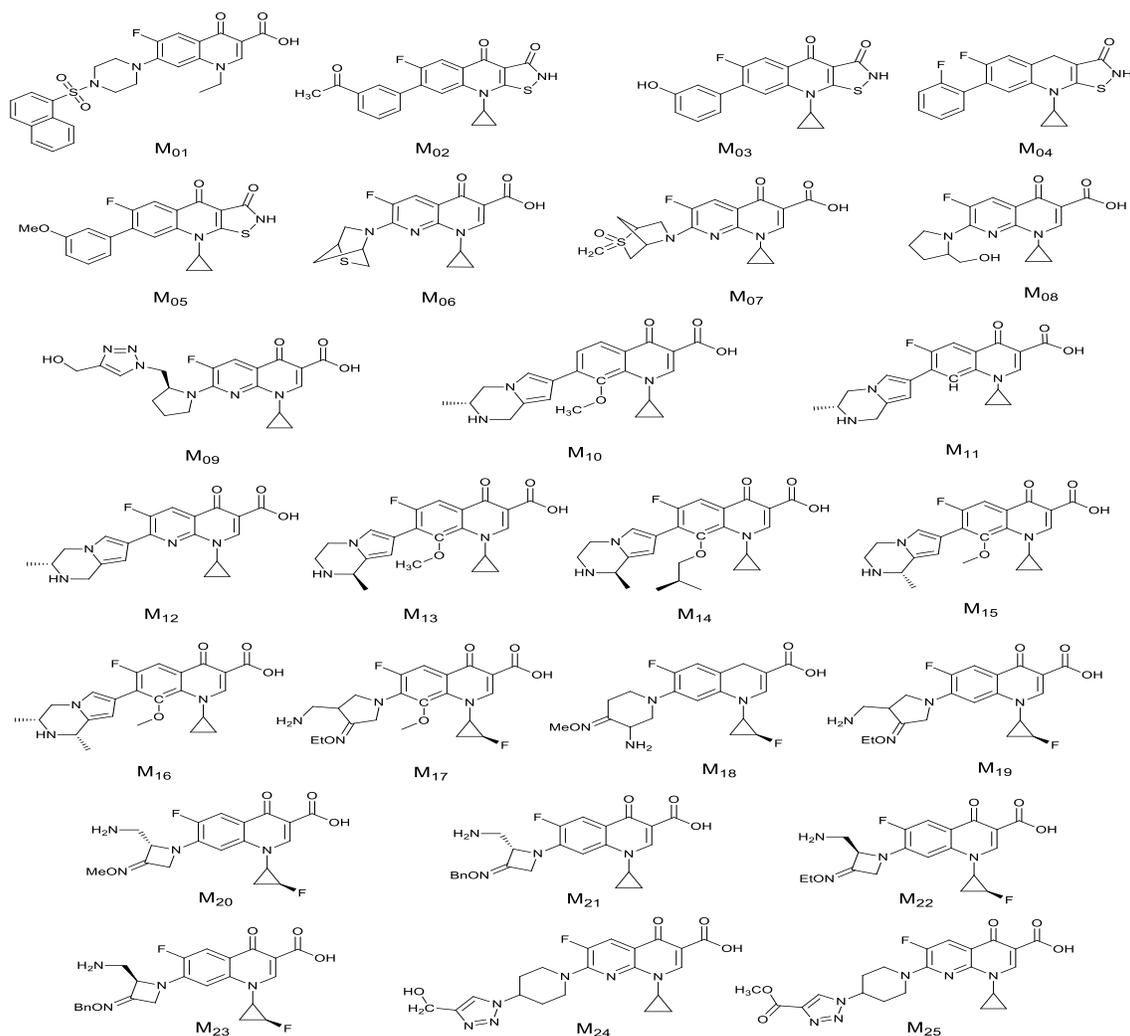
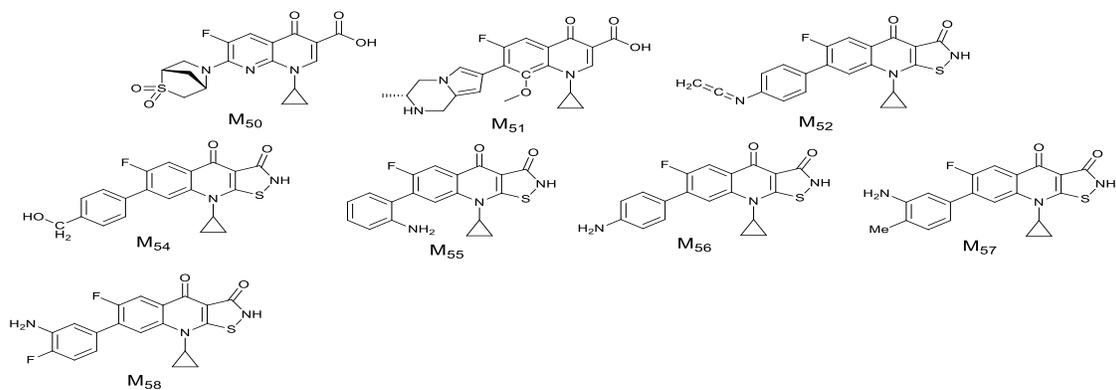


Figure 2: Training Set and Test Set Molecules



Test set:

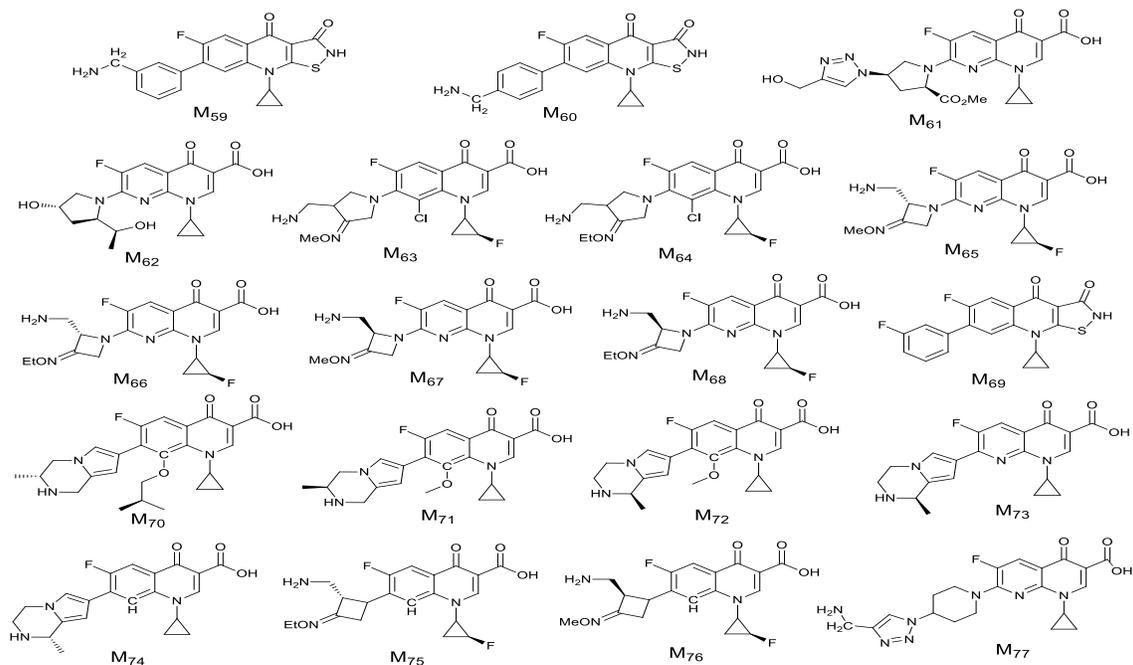


Figure 2: Training Set and Test Set Molecules continued.

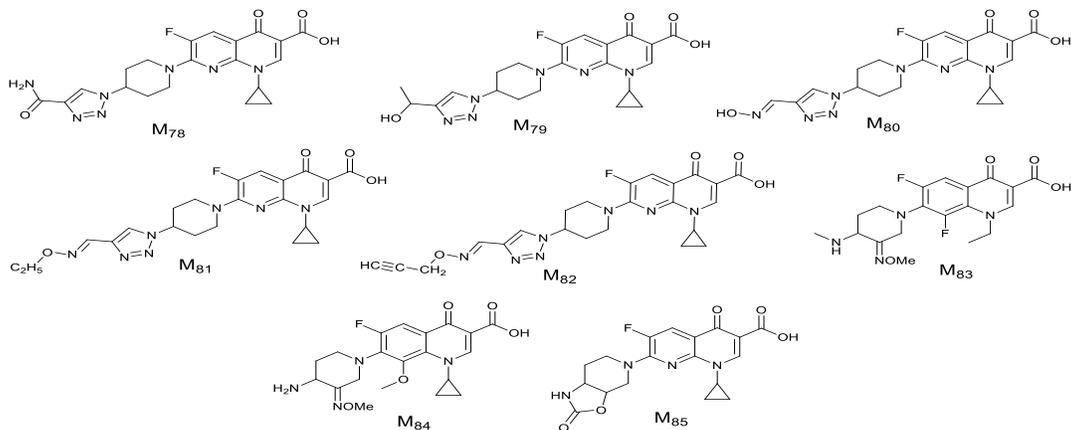


Figure 2: Training Set and Test Set Molecules continued.

All molecules were selected from different papers in literature,[17–22] The reason of this selection is to obtain a structural diversity for a better understanding of the observed bio-properties and to determine the most important structural parameters controlling the antibacterial activity. The biological activity modeled in this study was converted into -Log MIC where MIC is the minimal inhibition concentration against the organism shown.[23].

To reduce the number of descriptors. The objective selection is used to select a subset of descriptors that are best in encoding the property of quinolone derivatives using only the independent variables (descriptors). The reduction of variables was performed by the Dragon software using objective selection to remove descriptors which had identical or zero values for greater than 90% of the data set, one descriptors in any pair of descriptors whose pairwise correlation exceeding 0.9 was also eliminated.[32, 33]

Calculation

Geometric optimization

The structures of 85 quinolones derivatives was sketched using *ChemBioDraw* ultra 12.0 software [24] and was exported to *Gaussview* 5.0.9 and *Gaussian* 09 software [25]. The three dimensional structures of all molecules were generated, and their geometries were optimized preliminary with semi-empirical method AM1 [26], then using the quantum chemical DFT method(Density Functional Theory) included in Gaussian 09 software.[27, 28] After optimization the (x,y,z) atomic coordinates of the minimal energy conformation for each molecule can be determined.

Molecular descriptors

The optimized structures were transferred into *Dragon* 5.0 software[29] (developed by Milano chemometrics and QSPR group) to calculate 4485 descriptors in 29 different blocs.[30, 31]

QSAR model development

Objective feature selection of descriptors

Initially, over 4485 descriptors were calculated for each molecule in the data set, a feature selection was used to reduce the number of descriptors. The objective selection is used to select a subset of descriptors that are best in encoding the property of quinolone derivatives using only the independent variables (descriptors). The reduction of variables was performed by the Dragon software using objective selection to remove descriptors which had identical or zero values for greater than 90% of the data set, one descriptors in any pair of descriptors whose pairwise correlation exceeding 0.9 was also eliminated.[32, 33]

Subjective feature selection of descriptors

After elimination with objective selection, we obtained only 2255 descriptors, the number was reduced another time by subjective feature selection,[34–36] In this stage, descriptors selection is based on MLR analysis in combination with GA.[37, 38] The general aim behind the MLR procedure is to build a multiple linear regression model from a set of independent variables (descriptors) by entering and removing predictor in a stepwise manner until there is no justifiable reason to enter or remove any more (until no significant variables varies). All molecular descriptors are used to build QSAR model by MLR analysis, as implemented in XLStat software. [39] In this study, we used the stepping criteria: $\alpha=0.15$ to enter and $\alpha=0.15$ to remove in XLStat software. [39]. GA is a search heuristic method that belongs to the larger class of the *Evolution Algorithms* (EA) which generate solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, crossover, mutation, and selection. [40] The GA simulation conditions were: 10000 iterations, equation length change from 3 to 8 descriptors per model for map the set of quinolone derivatives to activity, number of crossovers was 5000, smoothness factor was 1, mutation probability was 0.5, and initial number of equations generated was 500. GA was executed multiple times (10-15 times) till an optimized solution is found.

A major decision to develop successive QSAR model is knowing when to stop adding descriptors to the model during GA-MLR procedure. A simple technique to control the model expansion is called "Breaking Point" in the improvement of the statistical quality of the model, by analyzing the plot of the number of descriptors involved in the models obtained (in our study from 3 to 8 descriptors) versus the squared correlation of coefficient R^2 and of cross validation correlation coefficient Q_{cv}^2 . The model corresponding to the breaking point is considered as the optimum model.[41].

Model validation and evaluation

Cross validation test

The first technique applied for the validation of the proposed model was based on leave-one-out algorithm. The "Leave one out" of the proposed model was based on leave-one-out algorithm. The "Leave one out" (Loo) cross validation was used to evaluate the predictivity of the final QSAR equation. This step is necessary, because a high value of the square of correlation coefficient R^2 indicates the best fit of the data, but does not contain information about the ability to predict the dependent variable or no included data in the training set. From the Loo cross-validation procedure, the square of cross-validation coefficient Q^2 is obtained, which is used as a criterion to evaluate both the robustness and the predictive ability of the generated model. According to Tropsha et al, a QSAR model is considered predictive if the following conditions are satisfied.[42, 43]

$$Q_{cv-Loo}^2 > 0.5 \quad (1)$$

$$R^2 > 0.6 \quad (2)$$

$$R^2 - \frac{R_0^2}{R^2} < 0.1 \quad (3)$$

$$0.85 < K < 1.15 \quad (4)$$

The Mathematical definition of R^2 , R_0^2 , $R_0'^2$, K and K' are based on regression for the test set of the observed activity against predicted activity and vice versa. According to Roy et al[44], the difference between values of R_0^2 and $R_0'^2$ must be studied and given importance. They suggested the following modified R^2 from:

$$R_m = R^2 \left(1 - \left| \sqrt{R^2 - R_0^2} \right| \right) \quad (5)$$

Y-Randomization test for the MLR model

The model was further validated by applying the Y-randomization test,[45] in which random MLR models are generated by randomly shuffling. The dependent variables while keeping the independent variables as they are. The new QSAR models are expected to have significant low R^2 and Q^2 values for several trials, which confirm that the developed QSAR models are robust. Another parameter, CR^2 is also calculated, which should be more than 0.5 for passing this test with:

$$CR^2 = R \sqrt{(R^2 - (\text{average } R_r)^2)} \quad (6)$$

Where: average R_r = average R of random models.

Euclidean based applicability domain (AD)

Applicability domain (AD) is the physicochemical structure or biological space, knowledge or information on which the training set of the model has been developed. The resulting model can be reliably applicable for only those compounds which are inside this domain.[46, 47] It is based on mean distance scores calculated by the Euclidean distance norms. A simple way of defining the range of a QSAR model is according

to the leverage of a compound. The leverage h_i of a compound can be used to measure outlyingness in the X-space and measures its influence on the model. Observations that are outliers in the X-space are known as high leverage points to distinguish them from observations that are outliers in the response variable (those with large standardized residuals). In matrix terms, the leverage of a compound in the original variable space is defined as:

$$h_i = X_i^T (X^T X)^{-1} X_i \dots (i = 1, \dots, n) \quad (7)$$

Where X_i is the descriptor row-vector of the query compound, and X the matrix of k model descriptor values for n training set compounds. The superscript T refers to the transpose of the matrix/vector. The warning leverage h^* is defined as follows:

$$h^* = 3h = \frac{3(\sum_i h_i)}{n} = \frac{3p'}{n} \dots (i = 1, N, n) \quad (8)$$

Where n is the number of training compounds and p' is the number of model adjustable parameters.[40].

The applicability domain (AD) of QSAR model is defined from the Williams plot. The plot of leverage values versus standardized residuals was used to give a graphical detection of both the response outliers (Y outliers) and the structurally influential compounds (X outliers). In this plot, the two horizontal lines indicate the limit of normal values for Y outliers (i.e., samples with standardized residuals greater than 3.0 standard deviation units, 3.0 s); the vertical straight lines indicate the limits of normal values for X outliers (i.e., samples with leverage values greater than the threshold value, $h > h^*$). For a sample in the external test set, whose leverage value is greater than h^* . It's prediction is the result of a substantial extrapolation of the model.[48] Conversely, when the leverage value of a compound is lower than the critical value, the probability of accordance between predicted and experimental values is as high as that for the compounds in the training set.

RESULT AND DISCUSSION

During model selection or formulation process, the task is to design, or more often to select an algorithm to mathematically describe the relations of descriptors and biological activity.

Selection of relevant descriptors

In order to select the predominant descriptors that will affect the MIC of these compounds, correlation analysis was performed with statistical software XLStat software. [39], taking every calculated descriptor as independent variable and $-\log$ MIC as a dependent variable, using the GA technique as method of selection. To select the most important descriptors and the optimal number in the model, we performed consecutively several equations with different number of descriptors from 3 to 8 variables. To define the optimum model, we must pass through an important step, that ensures the over parameterization of the model and prevents to some extent the chance correlations between descriptors. As mentioned above, this procedure is based on the break point rule, the change in the slope as shown in the plot of R^2 and Q_{CV-LOO}^2 versus number of descriptors added [49]. It was indicated that the maximum improvement of R^2 and Q_{CV-LOO}^2 was at five descriptors as shown in figure3.

Table 1 Correlation coefficients and VIF values among the five variables.

VARIABLES (DESCRIPTORS)	DP	RDF055U	RCI	MOR17I	CATS2D 04 AL	VIF
DP	1.000	0.384	0.213	-0.302	0.588	1.72
RDF055U		1.000	0.341	-0.354	0.256	1.36
RCI			1.000	-0.396	0.157	1.28
MOR17I				1.000	-0.273	1.37
CATS2D 04 AL					1.000	1.58

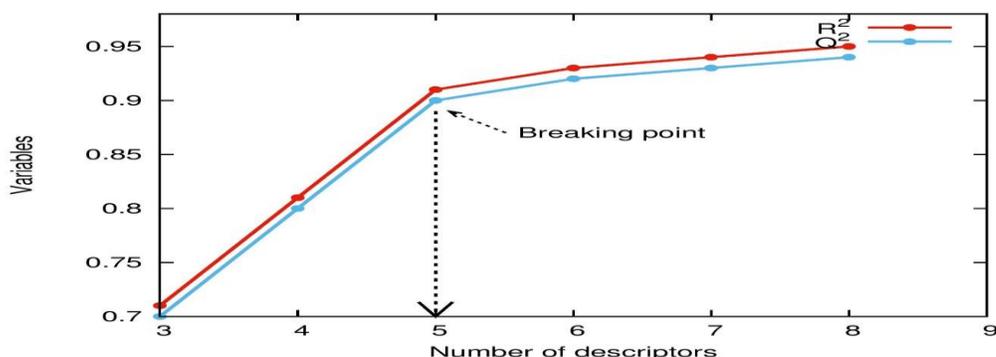


Figure 3 The optimum number of descriptors for the MLR-GA model.

The multi-collinearity test between the 5 descriptors in the equation was investigated by the VIF parameter, where VIF is the variable inflation factor. The values for these parameters are given in table1.

All descriptors have variance of VIF values between 1.28 and 1.72, indicating that the collinearity is not a problem for these data, so the obtained model have statistical significance, and the descriptors were found to be reasonably orthogonal. The same picture emerges if we examine the correlation coefficient matrix using XLStat software. As a result of the collinearity, low correlation coefficients (table 1) were detected among the value of the independent variables of the model represented by equation 9.

QSAR modeling

The aforementioned stepwise MLR technique coupled with GA, was used to establish the QSAR model. The structures of 85 quinolone derivatives was drawn (figure7, 8) and their antibacterial activity against *E. coli* were listed in table 2. Before starting the actual calculation for the model, we divided the data set into training and test set, using the Euclidean Based Kennard-Stone algorithm. The training set and the test set consisted 58 and 17 compounds respectively. On the MLR-GA method, we perform several equations from 3 to 8 variables. As noted above, the powerful QSAR model contains 5 descriptors shown with their high statistical parameters:

$$MIC = 5.723 - 8.60Dp - 0.03385RDF055u - 6.808RCI - 1.022Mor17i - 0.2049CATS2D\ 04\ AL \quad (9)$$

$$N = 58; R^2 = 0.91; S = 0.2743; F = 120.41; P < 10^{-4}; Q_{cv-Lo0}^2 = 0.9031; S_{cv-Lo0} = 0.2886$$

Where

- **N** is the number of compounds (training set),
- **S** is the standard deviation of the regression,
- **R²** is the squared correlation coefficient,
- **F** is the Fischer ratio,
- **S_{cv-Lo0}** is the cross validation standard deviation,
- **Q_{cv-Lo0}²** is deviation and square of the correlation coefficient.

Table 2 The experimental and predicted value of -log MIC with the selected descriptors.

No.	-Log MIC Exp (mol/ml)	-Log MIC Pred (mol/ml)	Dp	RDF055u	RCI	Mor17i	CATS2D 04 AL	Leverage (hi)
Training Set								
1	0.60	0.74	0.380	28.397	-0.091	-1.066	12.000	0.25067135
2	3.50	3.55	0.345	10.716	-0.086	-1.566	5.000	0.25962742
3	3.79	4.10	0.314	9.572	-0.113	-1.228	3.000	0.06509144
4	3.47	3.58	0.295	8.198	-0.019	-1.134	3.000	0.09549577



5	3.48	3.43	0.322	9.096	-0.093	-1.157	5.000	0.05317436
6	3.46	3.34	0.325	11.623	-0.014	-1.901	6.000	0.27592132
7	2.60	2.71	0.327	17.098	-0.017	-1.863	8.000	0.27394688
8	2.24	1.88	0.310	18.826	0.003	-0.895	7.000	0.16156481
9	1.13	1.24	0.329	28.471	0.023	-1.284	9.000	0.17257736
10	2.63	2.93	0.320	24.247	-0.028	-1.979	7.000	0.07082532
11	3.50	3.24	0.309	12.863	0.058	-1.790	4.000	0.14068250
12	3.18	3.07	0.309	14.710	0.047	-1.811	5.000	0.11444626
13	2.92	2.83	0.342	22.960	0.018	-1.731	4.000	0.05801818
14	2.92	2.98	0.348	38.413	0.036	-2.561	4.000	0.21743082
15	2.32	2.58	0.332	24.797	0.018	-1.469	4.000	0.06486185
16	2.63	2.67	0.344	30.443	-0.016	-1.615	4.000	0.06958903
17	2.05	2.10	0.361	31.690	-0.034	-0.917	3.000	0.20257029
18	3.21	3.54	0.289	22.881	-0.095	-1.025	3.000	0.10438099
19	3.85	3.32	0.321	21.233	-0.066	-1.216	3.000	0.05038511
20	4.69	4.48	0.285	20.024	-0.148	-1.460	3.000	0.13707945
21	2.64	2.98	0.330	31.095	-0.115	-1.768	7.000	0.14241815
22	4.70	4.22	0.284	19.092	-0.135	-1.253	3.000	0.10910278
23	3.27	2.73	0.348	21.720	-0.115	-1.360	7.000	0.06447378
24	2.93	2.83	0.325	15.535	-0.026	-1.256	5.000	0.03874262
25	2.66	2.89	0.368	15.313	-0.054	-1.482	5.000	0.06484982
26	3.23	3.02	0.339	14.457	-0.035	-1.464	5.000	0.03153311
27	2.66	2.40	0.361	16.897	-0.006	-1.911	8.000	0.06316960
28	2.36	2.70	0.377	15.577	-0.031	-1.533	5.000	0.28750275
29	2.08	2.01	0.405	17.715	-0.014	-1.474	6.000	0.15183708
30	1.82	1.66	0.414	19.303	-0.017	-1.846	9.000	0.13406368
31	3.54	3.16	0.365	30.427	-0.070	-1.906	4.000	0.11804362
32	4.15	3.87	0.360	18.795	-0.113	-1.689	3.000	0.13063692
33	2.92	2.97	0.342	19.481	-0.108	-1.923	9.000	0.12224275
34	2.01	2.40	0.383	21.149	-0.055	-1.716	7.000	0.25903428
35	1.47	1.48	0.369	38.565	0.031	-1.847	7.000	0.12157641
36	1.75	1.61	0.348	28.655	0.065	-1.692	7.000	0.10470795
37	0.91	0.87	0.425	27.258	-0.052	-1.190	9.000	0.20343004
38	3.50	3.73	0.296	9.319	-0.100	-1.186	5.000	0.06889501
39	2.05	2.26	0.402	28.763	-0.013	-2.673	9.000	0.18387913
40	3.22	3.20	0.307	21.240	-0.036	-1.990	7.000	0.22526157
41	3.85	3.25	0.325	23.785	-0.061	-1.300	3.000	0.05582371
42	1.17	1.35	0.376	18.490	-0.025	-1.541	11.000	0.12495147
43	3.51	3.14	0.357	19.955	-0.071	-2.075	7.000	0.07228965
44	1.21	1.04	0.375	27.245	-0.067	-1.439	12.000	0.18091071
45	3.51	3.75	0.349	11.730	-0.122	-1.589	5.000	0.08634628
46	2.92	2.92	0.335	16.078	-0.017	-2.103	8.000	0.09257504
47	2.64	2.60	0.328	36.111	0.067	-2.152	4.000	0.17339414
48	3.23	3.77	0.300	30.370	-0.106	-1.519	3.000	0.14764938
49	4.08	4.31	0.306	8.642	-0.101	-1.414	3.000	0.06714957
50	3.50	3.50	0.323	16.047	-0.027	-2.104	6.000	0.07766224
51	2.21	2.58	0.334	22.232	-0.051	-1.613	4.000	0.09129070
52	4.29	3.87	0.295	9.758	-0.083	-1.446	5.000	0.06670504
53	3.82	3.61	0.332	10.493	-0.091	-1.476	5.000	0.04814234
54	3.82	4.05	0.307	9.992	-0.075	-1.383	3.000	0.05245198
55	3.46	3.57	0.310	10.480	-0.091	-0.852	3.000	0.07675033
56	3.78	3.98	0.284	8.658	-0.094	-0.946	3.000	0.08288749

57	3.48	3.64	0.311	9.843	-0.058	-1.122	3.000	0.05990887
58	3.48	3.50	0.304	9.503	-0.101	-0.635	3.000	0.10936951
Test Set								
59	4.4	4.04	0.324	11.910	-0.102	-1.404	3.000	0.16794792
60	3.82	4.08	0.284	9.576	-0.068	-1.250	3.000	0.22267215
61	0.97	1.17	0.347	28.622	-0.054	-1.181	7.000	0.27527671
62	1.67	2.04	0.357	18.471	-0.064	-0.796	6.000	0.27688470
63	3.54	3.05	0.345	19.473	-0.074	-0.841	2.000	0.27251481
64	2.96	2.85	0.344	22.528	-0.062	-0.820	2.000	0.27840659
65	3.81	3.38	0.287	19.446	-0.036	-1.338	4.000	0.13920913
66	3.51	3.29	0.288	19.427	-0.019	-1.362	4.000	0.15194296
67	3.19	3.17	0.297	21.033	-0.036	-1.269	4.000	0.11544441
68	2.92	2.60	0.325	25.584	-0.019	-1.202	4.000	0.12669077
69	3.77	4.06	0.309	9.101	-0.101	-1.208	3.000	0.1753176
70	2.96	2.56	0.361	38.159	0.069	-2.476	4.000	0.25828249
71	2.95	2.63	0.335	21.783	0.051	-1.661	4.000	0.19939862
72	2.89	3.31	0.342	14.502	0.021	-1.948	4.000	0.24221746
73	2.88	3.29	0.316	16.515	0.008	-1.883	5.000	0.13524025
74	3.18	3.38	0.314	16.040	0.021	-1.825	4.000	0.17348944
75	4.13	4.38	0.286	22.151	-0.135	-1.534	3.000	0.30026026
76	3.89	4.26	0.282	18.356	-0.148	-1.166	3.000	0.30520997
77	2.43	2.69	0.332	20.012	-0.018	-1.375	5.000	0.04948037
78	2.24	2.51	0.342	16.272	-0.014	-1.190	5.000	0.08904159
79	2.64	2.24	0.350	15.798	0.019	-1.598	7.000	0.20680778
80	2.34	2.88	0.330	15.452	-0.024	-1.357	5.000	0.06155180
81	2.06	2.68	0.395	17.214	-0.036	-1.685	5.000	0.28856971
82	2.07	2.16	0.343	17.514	-0.015	-1.090	6.000	0.13723289
83	2.91	3.32	0.356	18.897	-0.124	-1.051	3.000	0.24274444
84	2.52	2.81	0.349	24.323	-0.032	-1.485	4.000	0.09370465
85	2.59	2.55	0.324	13.671	0.026	-1.659	7.000	0.23920959

Table 3 The definition of descriptors based on the model with their bloc kind.

DESCRIPTORS SYMBOL	DESCRIPTION	BLOC
DP	D total accessibility index / weighted by polarizability	WHIM descriptors
RDF055U	Radial Distribution Function - 055 / unweight	RDF descriptors
RCI	ring complexity index	Ring descriptors
MOR17I	signal 17 / weighted by ionization potential	3D-MoRSE descriptors
CATS2D 04 AL	CATS2D ACCEPTOR-LIPOPHILIC AT LAG 04	CATS 2D

The definition of descriptors based on the model used in the present study was designed as follow in the table 3.

The values of descriptors for all the training set as well as the test set are summarized in table 2. As can be seen from the statistical parameters of the above equation, a considerable improvement was achieved by combining five descriptors, equation 9 can explain the 91% of the experimental variance of the dependent variable MIC. The MLR equation is given by value of regression ($R^2 = 0.91$) which explain around 91% of the variance of data, the model presents the greatest Fisher value ($F = 120.41$) and the lowest standard deviation for the data ($S = 0.2743$) which confirm the former selection and indicate the ability of predictive power of this QSAR model. The present QSAR model developed with MLR coupled with GA as method of selection for variable, confirms that the prediction of the MIC against *E. coli* bacteria is statistically very significant. The validity and predictability of the QSAR model for anti-bacterial activity were cross validated by correlation coefficient value ($R^2 = 0.903$) and standard deviation of cross validation value ($S_{cv-Loo} = 0.2886$), obtained by the leave on out (Loo) method. The high value of the determination

coefficient of leave one out cross-validation for the obtained model, and small standard deviation cross validation are proving the predictive power of this approach and the stability of the model. The model was further validated by applying the Y-randomization test, several random shuffles of the Y vector were performed and the low R^2 and Q_{cv-LoO}^2 values that were obtained show that the good results in our model are not due to a chance correlation or structural dependency of the training set. The results of the Y-randomization test are presented in table 4.

Table 4 R^2 and Q_{cv-LoO}^2 values after several Y-randomizations.

Model	R	R^2	Q_{cv-LoO}^2
Original model	0.959	0.920	0.903
Random 1	0.386	0.149	-0.075
Random 2	0.227	0.052	-0.236
Random 3	0.296	0.088	-0.106
Random 4	0.278	0.077	-0.117
Random 5	0.222	0.049	-0.225
Random 6	0.168	0.028	-0.205
Random 7	0.260	0.067	-0.131
Random 8	0.318	0.101	-0.103
Random 9	0.233	0.054	-0.190
Random 10	0.255	0.065	-0.155

As mentioned in section 1.4. We verified the predictive ability using Golbraikh et al criteria (equations 1, 2, 3, 4).

$$Q_{cv-LoO}^2 = 0.9031, \text{ passed (threshold value } Q^2 > 0.5)$$

$$R^2 = 0.719, \text{ passed (threshold value } R^2 > 0.6)$$

$$\frac{R^2 - R_0^2}{R^2} < 0.00051, \text{ passed (threshold value } \frac{R^2 - R_0^2}{R^2} < 0.1)$$

$$K = 1.035, \text{ passed (threshold value } 0.85 < K < 1.15)$$

Equation 9 represents our best performing QSAR model, Figure 4 shows the corresponding scatter plot of the estimated versus experimental activity values for quinolone derivatives, this figure indicates that there is a significant correlation between experimental and predicted values of log MIC for training set and test set.

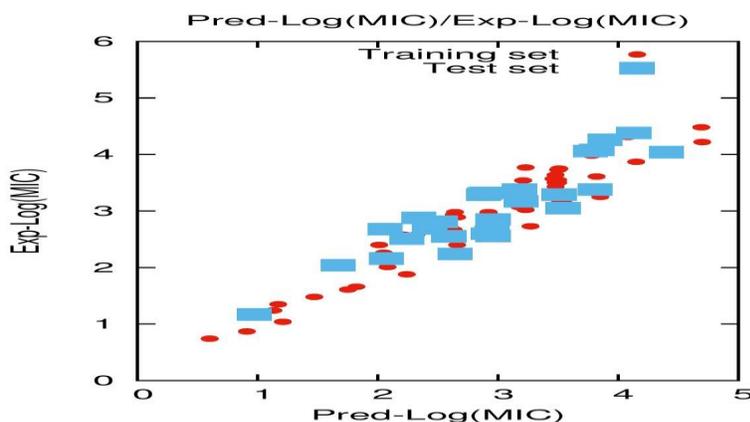


Figure 4 Predicted versus the experimental values of $-\log$ MIC For the training set and validation set.

The results illustrate that the MLR technique combined with GA as variable selection procedure are adequate to generate an efficient QSAR model. The residual of the predicted values of MIC against the experimental values for the present model is shown in the Figure 5.

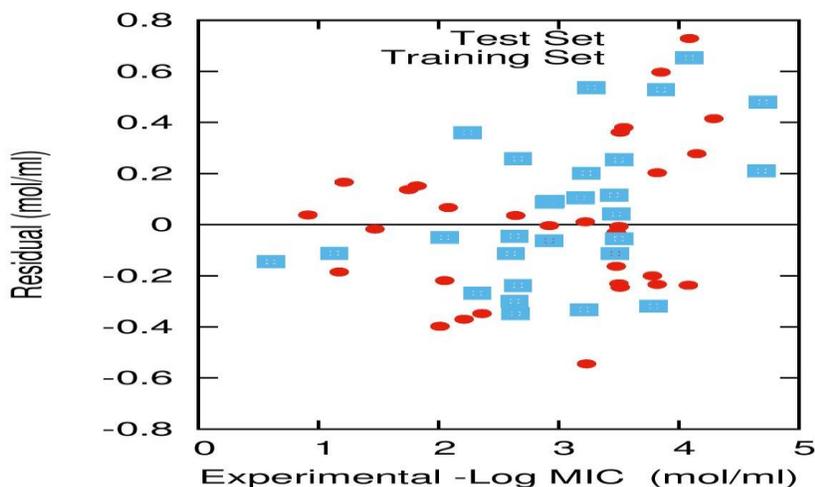


Figure 5 Plot of the residuals for calculated values of $-\log$ MIC from the GA-MLR model versus their experimental values for the training and validation sets.

As most of the calculated residuals are distributed on two sides of zero line, a conclusion maybe drawn that there is no systematic error in the development of the present model. To see the importance of each descriptor for the prediction of antibacterial activity, the relative contributions of 5 descriptors to the MLR-GA model were determined and are plotted in Figure 6. Interpreting a QSAR model in terms of the specific contribution of substituents and other molecular features to the modeled activity is always a difficult task. Dp and RCI are the most important variables in this equation (figure 6), because they present the highest contribution in our model. This two descriptors have negative influence on the studied property, the reason why the inhibitory activity of the quinolone derivatives increases.

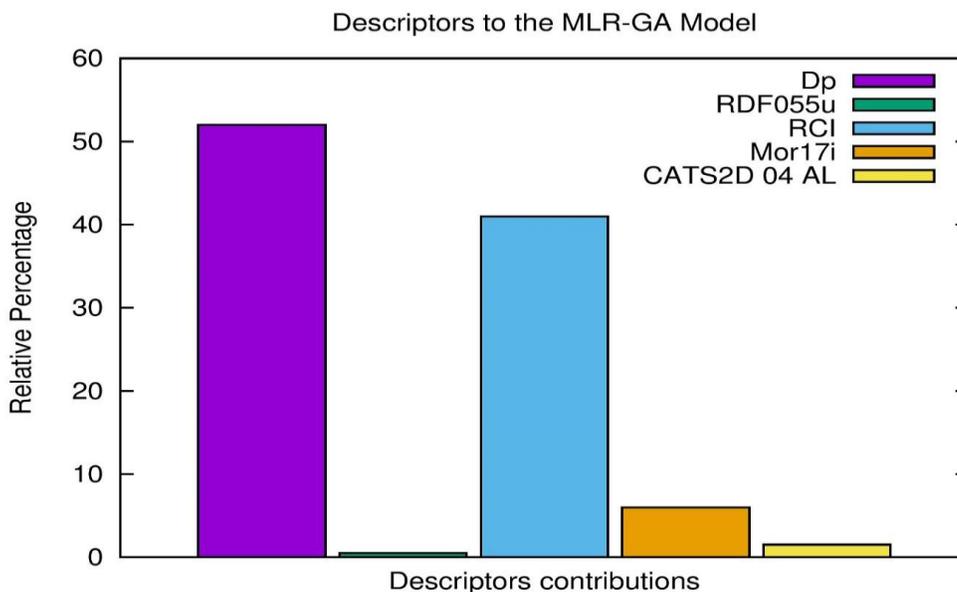


Figure 6 Histogram depicting the relative contributions of the five descriptors to the MLR-GA model.

We made a comparison between compounds N: 26 and 30, 20 and 48. As we can see, the difference between compounds 30 and 26 is in the radical at position 7 of quinolone derivatives, whereas the radical in molecule 26 contains 4 atoms of nitrogen and one of oxygen, and for molecule 30, the radical contains 5 atoms of nitrogen and one of oxygen (increase of polarizability), this difference leads to increase Dp values from 0.339 to 0.414 and consequently decrease the MIC value as shown in figure 7.

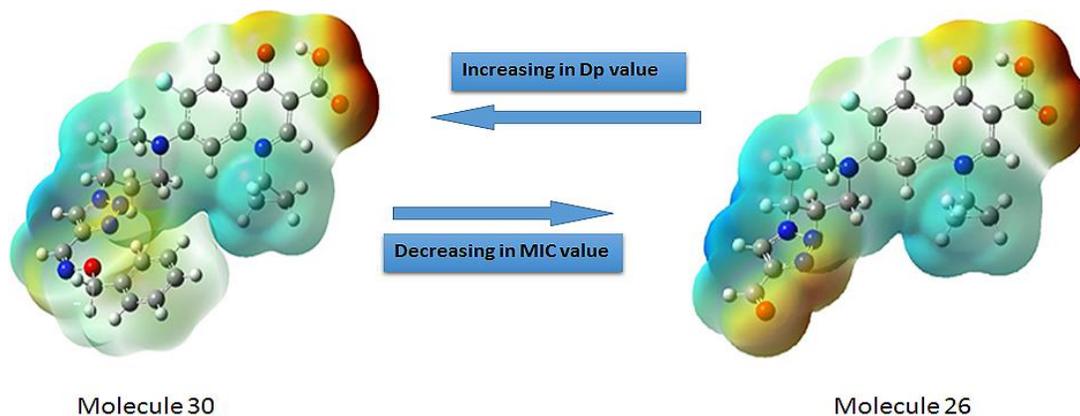


Figure 7 Molecular Electrostatic potentials (MEP) comparison mapped on the electron density surface calculated by the DFT/B3LYP method between compounds N30 and N26.

On the other hand, as shown in equation 9, the inhibition of quinolone derivatives was dependent on the steric parameter (ring complexity index). Indeed, structure analog of 2-benzoylacrylic in quinolone molecule consists a big conjugated system contained 12 of π electrons. Carbon atoms and Fluorine, Nitrogen and Oxygen heteroatoms provide lone pair electrons. Previous study stating that the electron transfer may occur when quinolones interact with DNA or protein.[50] It is clear that which substitution position and what kind of substituent may affect the mechanism of action of pharmacophore (2-benzoylacrylic) with DNA. That's why another comparison was made between compounds 48 and 20; the only difference between these two compounds is the kind of substituent in the pharmacophore of quinolone derivatives at position 8. In spite of this, apparently small difference leads to increase of RCI values from -0.148 to -0.106 and consequently decreases the value of MIC as shown in figure 8.

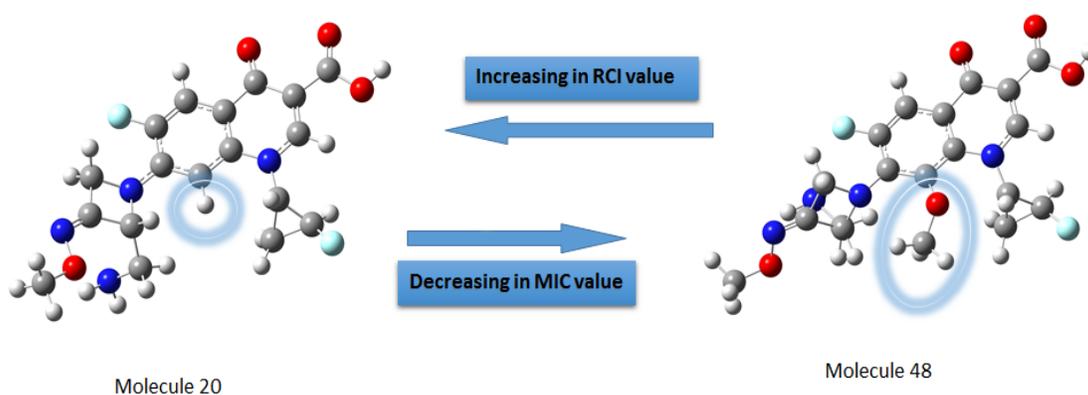


Figure 8 Structure comparison between compounds N20 and N48.

Applicability domain

The prime overall goal of QSAR research is to develop models that provide accurate predictions for as many chemical structures as possible in the universe, particularly for those that have not been tested or for which reliable experimental data is still not available. To this end, however, QSAR model must always be

verified for their applicability with regard to chemical domain, in order to produce predicted data that can be considered reliable only for too structurally dissimilar chemicals. The applicability domain of the model was analyzed using a Williams plot (Figure 9), where the vertical line is the critical leverage value ($h^*=0.31$), and the horizontal lines are the cut off value for Y space. From this plot, the applicability domain is established inside a squared area within 3 standard deviations and a leverage threshold h^* leverage= 0.31. For making predictions: training data must be considered reliable only for those compounds that fall within this AD on which the model was constructed. It can be seen from Figure 9 that the majority of compounds in the data set are inside this area. Apparently, no influential or outlier data was detected. Therefore, the model presented by Equation 9 displays the best statistical parameters, good prediction, and applicability.

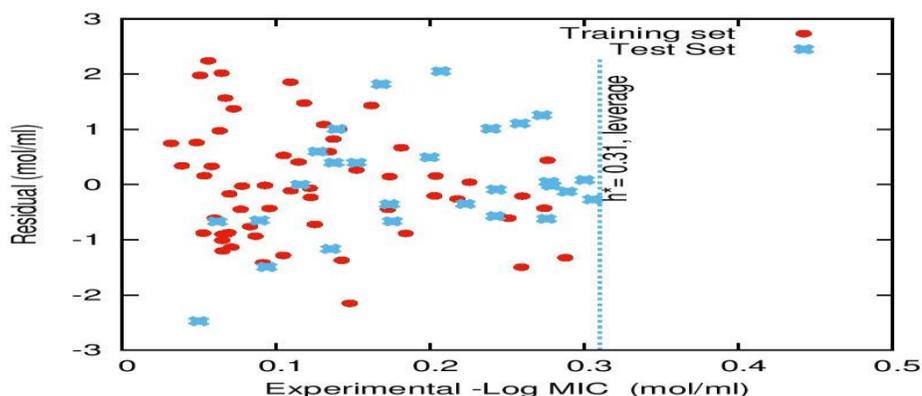


Figure 9 Projection of the training set and the validation set in the Williams plot.

1. CONCLUSION

The aim of the present work was to develop a QSAR study and to predict the minimal inhibition concentration of quinolone derivatives to gram negative (*E. coli*). Minimal inhibition concentration for a set of 85 quinolone derivatives was modeled with success by Multiple Linear Regression analysis, using Genetic Algorithms as variable selection method. The five selected descriptors showed that the polarizability properties and the structure of the molecule play a main role in the inhibition of quinolone derivatives. The proposed model has good stability, robustness and predictivity when verified by internal validation (cross validation by Leave One Out and Y randomization) and external validation. The chemical applicability domain of the studied model served as a valuable tool to filter out dissimilar and outlier compounds.

ACKNOWLEDGEMENTS

This work was supported financially by The General Directorate for Scientific Research and Technological Development (DG-RSDT), Algerian Ministry of Scientific Research, Organic Synthesis and Biocatalysis Laboratory (FNR 2000).

REFERENCES

- [1] Alexander Fleming. The discovery of penicillin. *British Medical Bulletin*, 2(1):4–5, 1944.
- [2] AW Bauer, WMM Kirby, J C&turck Sherris, and Marvin Turck. Antibiotic susceptibility testing by a standardized single disk method. *American journal of clinical pathology*, 45(4):493, 1966.
- [3] JACKSON E Fowler Jr and THOMAS A Stamey. Studies of in- troital colonization in women with recurrent urinary infections. vii. the role of bacterial adherence. *The Journal of urology*, 117 (4):472–476, 1977.
- [4] G Ka“llenius, SB Svenson, H Hultberg, R Mo“llby, I Helin, B Cedergren, and J Winberg. Occurrence of p-fimbriated es- cherichia coli in urinary tract infections. *The Lancet*, 318(8260): 1369–1372, 1981.
- [5] Jochen Kuhlmann, Axel Dalhoff, and H-J Zeiler. *Quinolone an- tibacterials*, volume 127. Springer Science & Business Media, 2012.

- [6] Hubert Ph Endtz, Gijs J Ruijs, Bert van Klingeren, Wim H Jansen, Tanny van der Reyden, and R Peter Mouton. Quinolone resistance in campylobacter isolated from man and poultry following the introduction of fluoroquinolones in veterinary medicine. *Journal of antimicrobial chemotherapy*, 27(2):199–208, 1991.
- [7] Lester A Mitscher. Bacterial topoisomerase inhibitors: quinolone and pyridone antibacterial agents. *Chemical reviews*, 105(2):559–592, 2005.
- [8] Munekazu Yamakuchi, Masanori Nakata, Ko-ichi Kawahara, Isao Kitajima, and Ikuro Maruyama. New quinolones, ofloxacin and levofloxacin, inhibit telomerase activity in transitional cell carcinoma cell lines. *Cancer letters*, 119(2):213–219, 1997.
- [9] Corwin Hansch. Quantitative structure-activity relationships in drug design. *Drug design*, 1:271–342, 1971.
- [10] Mohammed I El-Gamal, Said M Bayomi, Saadia M El-Ashry, Shehta A Said, A-M Alaa, and Naglaa I Abdel-Aziz. Synthesis and anti-inflammatory activity of novel (substituted) benzylidene acetone oxime ether derivatives: molecular modeling study. *European journal of medicinal chemistry*, 45(4):1403–1414, 2010.
- [11] [A-M Alaa, Kamal EH ElTahir, and Yousif A Asiri. Synthesis, anti-inflammatory activity and cox-1/cox-2 inhibition of novel substituted cyclic imides. part 1: Molecular docking study. *European journal of medicinal chemistry*, 46(5):1648–1655, 2011.
- [12] A-A Magda, Naglaa I Abdel-Aziz, A-M Alaa, Adel S El-Azab, Yousif A Asiri, and Kamal EH ElTahir. Design, synthesis, and biological evaluation of substituted hydrazone and pyrazole derivatives as selective cox-2 inhibitors: molecular docking study. *Bioorganic & medicinal chemistry*, 19(11):3416–3424, 2011.
- [13] Adel S El-Azab, Mohamed A Al-Omar, A-M Alaa, Naglaa I Abdel-Aziz, A-A Magda, Abdulaziz M Aleisa, Mohamed M Sayed-Ahmed, and Sami G Abdel-Hamide. Design, synthesis and biological evaluation of novel quinazoline derivatives as potential antitumor agents: molecular docking study. *European journal of medicinal chemistry*, 45(9):4188–4198, 2010.
- [14] Fatmah AM Al-Omary, Laila A Abou-Zeid, Mahmoud N Nagi, El-Sayed E Habib, A-M Alaa, Adel S El-Azab, Sami G Abdel-Hamide, Mohamed A Al-Omar, Abdulrahman M Al-Obaid, and Hussein I El-Subbagh. Non-classical antifolates. part 2: synthesis, biological evaluation, and molecular modeling study of some new 2, 6-substituted-quinazolin-4-ones. *Bioorganic & medicinal chemistry*, 18(8):2849–2863, 2010.
- [15] W Goettsch, W Van Pelt, N Nagelkerke, MGR Hendrix, AGM Buiting, PL Petit, LJM Sabbe, AJA Van Griethuysen, and AJ De Neeling. Increasing resistance to fluoroquinolones in *Escherichia coli* from urinary tract infections in the Netherlands. *Journal of Antimicrobial Chemotherapy*, 46(2):223–228, 2000.
- [16] TW Maddox, PD Clegg, NJ Williams, and GL Pinchbeck. Antimicrobial resistance in bacteria from horses: Epidemiology of antimicrobial resistance. *Equine veterinary journal*, 47(6):756–765, 2015.
- [17] Jason A Wiles, Qiuping Wang, Edlaine Lucien, Akihiro Hashimoto, Yongsheng Song, Jijun Cheng, Christopher W Marlor, Yangsi Ou, Steven D Podos, Jane A Thanassi, et al. Isothiazoloquinolones containing functionalized aromatic hydrocarbons at the 7-position: synthesis and in vitro activity of a series of potent antibacterial agents with diminished cytotoxicity in human cells. *Bioorganic & medicinal chemistry letters*, 16(5):1272–1276, 2006.
- [18] Xiaoguang Huang, Dongliang Chen, Ning Wu, Ai Qin Zhang, Zhenhua Jia, and Xingshu Li. The synthesis and biological evaluation of a novel series of c7 non-basic substituted fluoroquinolones as antibacterial agents. *Bioorganic & medicinal chemistry letters*, 19(15):4130–4133, 2009.
- [19] Bin Zhu, Brett A Marinelli, Raul Goldschmidt, Barbara D Foleno, James J Hilliard, Karen Bush, and Mark J Macielag. Synthesis and antibacterial activity of 7-(1, 2, 3, 4-tetrahydropyridin-1(2H)-yl) quinolones. *Bioorganic & medicinal chemistry letters*, 19(17):4933–4936, 2009.

- [20] Xiaoguang Huang, Aiqin Zhang, Dongliang Chen, Zhenhua Jia, and Xingshu Li. 4-substituted 4-(1h-1, 2, 3-triazol-1-yl) piperidine: Novel c7 moieties of fluoroquinolones as antibacterial agents. *Bioorganic & medicinal chemistry letters*, 20(9):2859–2863, 2010.
- [21] Kai Lv, Jinwei Wu, Jian Wang, Mingliang Liu, Zengquan Wei, Jue Cao, Yexin Sun, and Huiyuan Guo. Synthesis and in vitro antibacterial activity of quinolone/naphthyridone derivatives containing 3-alkoxyimino-4-(methyl) aminopiperidine scaffolds. *Bioorganic & medicinal chemistry letters*, 23 (6):1754–1759, 2013.
- [22] Hongmin Liu, Ju Huang, Jiayang Wang, Minghua Wang, Mingliang Liu, Bin Wang, Huiyuan Guo, and Yu Lu. Synthesis, antimycobacterial and antibacterial evaluation of l-[(1r, 2s)-2-fluorocyclopropyl] fluoroquinolone derivatives containing an oxime functional moiety. *European journal of medicinal chemistry*, 86:628–638, 2014.
- [23] Jennifer M Andrews. Determination of minimum inhibitory concentrations. *Journal of antimicrobial Chemotherapy*, 48 (suppl 1):5–16, 2001.
- [24] ChemBioDraw Ultra. 12.0. Cambridge, USA: Cambridge Soft Corporation, 1986.
- [25] MJ Frisch, GW Trucks, HB Schlegel, GE Scuseria, MA Robb, JR Cheesemann, VG Zakrzewski, JA Montgomery Jr, RE Stratmann, JC Burant, et al. Gaussian software, 2003.
- [26] Michael JS Dewar, Eve G Zebisch, Eamonn F Healy, and James JP Stewart. Development and use of quantum mechanical molecular models. 76. am1: a new general purpose quantum mechanical molecular model. *Journal of the American Chemical Society*, 107(13):3902–3909, 1985.
- [27] Chengteh Lee, Weitao Yang, and Robert G Parr. Development of the collesalvetti correlation-energy formula into a functional of the electron density. *Physical review B*, 37(2):785, 1988.
- [28] Warren J. Hehre. *Ab initio molecular orbital theory*. Wiley-Interscience, 1986.
- [29] Andrea Mauri, Viviana Consonni, Manuela Pavan, and Roberto Todeschini. Dragon software: An easy approach to molecular descriptor calculations. *Match*, 56(2):237–248, 2006.
- [30] R Todeschini, V Consonni, A Mauri, and M Pavan. Dragon software for the calculation of molecular descriptors. *Web version*, 3, 2003.
- [31] R Todeschini, V Consonni, and R Mannhold. Methods and principles in medicinal chemistry. *Kubinyi H, Timmerman H (Series eds) Handbook of molecular descriptors*. Wiley-VCH, Weinheim, 2000.
- [32] María Jimena Martínez, Ignacio Ponzoni, Mónica F D'Íaz, Gustavo E Vázquez, and Axel J Soto. Visual analytics in cheminformatics: user-supervised descriptor selection for qsar methods. *Journal of cheminformatics*, 7(1):1–17, 2015.
- [33] Brian E Mattioni and Peter C Jurs. Prediction of dihydrofolate reductase inhibition and selectivity using computational neural networks and linear discriminant analysis. *Journal of Molecular Graphics and Modelling*, 21(5):391–419, 2003.
- [34] Subhash C Basak. Mathematical descriptors for the prediction of property, bioactivity, and toxicity of chemicals from their structure: A chemical-cum-biochemical approach. *Current computer-aided drug design*, 9(4):449–462, 2013.
- [35] Linnan He and Peter C Jurs. Assessing the reliability of a qsar model's predictions. *Journal of Molecular Graphics and Modelling*, 23(6):503–523, 2005.
- [36] Rajarshi Guha, Jon R Serra, and Peter C Jurs. Generation of qsar sets with a self-organizing map. *Journal of Molecular Graphics and Modelling*, 23(1):1–14, 2004.
- [37] Leona S Aiken, Stephen G West, and Steven C Pitts. Multiple linear regression. *Handbook of psychology*, 2003.
- [38] Raimund Mannhold, Hugo Kubinyi, Hendrik Timmerman, and David E Clark. *Evolutionary algorithms in molecular design*, volume 8. John Wiley & Sons, 2008.
- [39] Addinsoft. Xlstat 2015, data analysis and statistics software for microsoft excel, 2010.
- [40] Peter J Rousseeuw and Annick M Leroy. *Robust regression and outlier detection*, volume 589. John Wiley & Sons, 2005.

- [42] Alan R Katritzky, Liliana M Pacureanu, Svetoslav Slavov, Dimitar A Dobchev, and Mati Karelson. Qsar study of antiplatelet agents. *Bioorganic & medicinal chemistry*, 14(22):7490–7500, 2006.
- [43] Alexander Golbraikh and Alexander Tropsha. Predictive qsar modeling based on diversity sampling of experimental datasets for the training and test set selection. *Molecular diversity*, 5(4): 231–243, 2000.
- [44] Alexander Golbraikh and Alexander Tropsha. Beware of q^2 ! *Journal of Molecular Graphics and Modelling*, 20(4):269–276, 2002.
- [45] Kunal Roy, Pratim Chakraborty, Indrani Mitra, Probir Kumar Ojha, Supratik Kar, and Rudra Narayan Das. Some case studies on application of “ rm^2 ” metrics for judging quality of quantitative structure–activity relationship predictions: emphasis on scaling of response data. *Journal of computational chemistry*, 34(12):1071–1082, 2013.
- [46] Liying Zhang, Hao Zhu, Tudor I Oprea, Alexander Golbraikh, and Alexander Tropsha. Qsar modeling of the blood–brain barrier permeability for diverse organic compounds. *Pharmaceutical research*, 25(8):1902–1914, 2008.
- [47] E Papa, JC Dearden, and P Gramatica. Linear qsar regression models for the prediction of bioconcentration factors by physicochemical properties and structural theoretical molecular descriptors. *Chemosphere*, 67(2):351–358, 2007.
- [48] Paola Gramatica, Elisa Giani, and Ester Papa. Statistical external validation and consensus modeling: A qspr case study for k_{oc} prediction. *Journal of Molecular Graphics and Modelling*, 25(6):755–766, 2007.
- [49] Lennart Eriksson, Joanna Jaworska, Andrew P Worth, Mark TD Cronin, Robert M McDowell, and Paola Gramatica. Methods for reliability and uncertainty assessment and for applicability evaluations of classification-and regression-based qsars. *Environmental health perspectives*, 111(10):1361, 2003.
- [50] Lee Phillips. *gnuplot Cookbook*. Packt Publishing Ltd, 2012.
- [51] Jianying Hu, Wanfeng Wang, Zhou Zhu, Hong Chang, Feng Pan, and Binle Lin. Quantitative structure-activity relationship model for prediction of genotoxic potential for quinolone antibacterials. *Environmental science & technology*, 41(13):4806–4812, 2007.

