

الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي و البحث العلمي

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université de RELIZANE
Faculté des Sciences de la Nature et de la vie
Département des Sciences Biologiques



MEMOIRE

En vue de l'obtention du diplôme de MASTER en :

Microbiologie & contrôle de qualité

Intitulé

Développement d'un modèle QSAR pour l'inhibition de *Pseudomonas aeruginosa*

Présenté par :

Mlle : Benhanina wafa

Mlle : Sahouadj maroua

Devant les membres de jury :

Président : Mme	Ghalouni Asmaa	MCA	(Univ. Relizane)
Encadreur : Mme	Laref Nora	MCA	(Univ. Relizane)
Examineur : Mme	Belkhir khadidja	MCA	(Univ. Relizane)

Année universitaire: 2024/2025

Remerciement



D’abord, nous remercions dieux pour tout ce qu’il a effectué dans nos vies et qui nous donné l’occasion, la santé et la volonté pour faire ce parcours de formation.

Nous adressons nos sincères remerciements à tout membre de la famille, particulièrement nos chers parents ainsi que tous nos ami(e)s

Ensuite, nous remercions infiniment notre directrice de recherche madame Laref Nora pour son aide, ses conseils pertinents, son suivi régulier et qui nous guidé(e)s, orienté(e)s tout au long de notre expérience enrichissante et pleine d'intérêt et nous avons l'honneur de remercier nos collègues de nous donner quelques conseils.

Je tiens à exprimer ma profonde gratitude aux membres du jury qui m’ont fait l’honneur de participer à l’évaluation de ce travail, je remercie le président Ghalouni Asmaa d’avoir accepté de présider le jury et pour ses commentaires constructifs, et l’examinatrice Belkhir Khadija.

Nous remercions respectivement tous les enseignants(es) de qui nous ont accompagnés(e)s durant notre parcours universitaire.

Enfin, nous tenons à remercier toutes les personnes qui ont contribué par leurs paroles, écrits, conseils et leurs critiques au succès de notre travail de recherche.



Dédicaces

بسم الله الرحمن الرحيم

Avant toute chose, je tiens à remercier Allah, source de toute sagesse et force, sans qui ce travail n'aurait pu voir le jour.

A mes plus grand soutiens et sources d'inspiration, je dédie ce travail avec tout mon amour et ma reconnaissance infinies

À mes chers parents, Mokhtar et Naissa, pour leur soutien indéfectible, leur bienveillance et leurs sacrifices, qui ont été essentiels dans la réalisation de mon parcours académique.

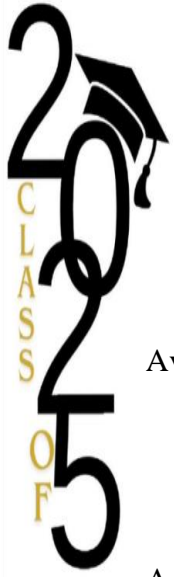
A mes sœurs, Douaa et Tasnim, pour leur présence constante, leur compréhension et leur encouragement tout au long de ce projet, leur soutien a été une source de motivation précieuse.

A mon frère, Mohamed Djamel, pour sa patience, son soutien et son implication, qui m'ont permis de poursuivre mes ambitions sereinement.

Enfin, à mon binôme Maroua qui est devenue une amie chère et une collaboratrice talentueuse, merci pour notre collaboration fructueuse et notre amitié. Tu as été une source d'inspiration et de motivation pour moi tout au long de ce parcours.

Au-delà des noms cités, il existe un cercle précieux de personnes qui ont joué un rôle significatif dans mon parcours. Je vous exprime ma reconnaissance pour votre présence et votre soutien qui ont marqué positivement ma vie.

Wafa



Dédicaces

بسم الله الرحمن الرحيم

Je dédie ce travail à tous qui sont la source de mon inspiration et mon courage:

À mes chers parents ABD EL KADER et FATIMA qui ont toujours été là pour moi et qui m'ont sacrifiés, soutenu et encouragé tout au long de mes études et qui ont souffert pour que je puisse réussir. Espérant qu'ils trouveront dans ce travail toute ma reconnaissance.

Je le dédie également à:

Tous les membres de la famille qui ont contribué par leur soutien moral et leurs encouragements:

À mon cher frère Salah et à sa femme, ma chère Safia

À ma chère sœur khouloud

À ma chère amie et mon binôme Wafa

Enfin, à toute personne de près ou de loin participe à la réalisation de mon travail.

Maroua



Sommaire

Liste des abréviations

Liste des figures

Liste des tableaux

Résumé

Abstract

ملخص

Sommaire

I.	Introduction	2
----	--------------------	---

Chapitres 1:Revue bibliographique

II-	Description générale.....	5
-----	---------------------------	---

III.	La résistance aux antibiotiques	6
------	---------------------------------------	---

III.	1. Mécanismes enzymatiques	7
------	----------------------------------	---

III.	1.1. Hydrolyse de bêta-lactamines.....	7
------	--	---

III.	1.2. Résistance par hyperproduction de la céphalosporinase Amp C.....	8
------	---	---

III.	1.3. Résistance par production de pénicillinases	8
------	--	---

III.	1.4. Résistance à spectre étendue (BLSE)	9
------	--	---

III.	1.5. Résistance par production de carbapénémases (MBLs)	11
------	---	----

III.	1.6. Résistance aux oxacillinases classe D	11
------	--	----

III.	1.7. Résistance aux aminosides	11
------	--------------------------------------	----

III.	2. Mécanismes non enzymatiques	12
------	--------------------------------------	----

III.	2.1. Surexpression de système d'efflux	12
------	--	----

III.	2.2. Résistance due à la diminution de la perméabilité	12
------	--	----

III.	2.3. Résistance par efflux actif	13
------	--	----

III.	2.4. Résistance par modification de la cible ribosomale	13
------	---	----

III.	2.5. Rôle de l'opéron dans la résistance	14
------	--	----

Chapitres 2 :Matériels et méthodes

IV.	Matériels et méthodes	16
IV.	3. Classification selon la loi de Lipinski.....	16
IV.	4. Calcul des descripteurs	16
IV.	5. Construction d'un modèle QSAR	17
IV.	6. Validation de modèle.....	17

Chapitres 3 : Résultats et discussion

V.	Résultats et discussion	19
V.	1. Collecte et nettoyage des données	19
V.	2. Classification des données.....	19
V.	2.2. Molécules Inactives.	19
V.	2.3. Molécules Intermédiaires	19
V.	3. Classification selon la loi de Lipinski.....	20
V.	4. Calcul des descripteurs	22
V.	5. Comparaison des Modèles de QSAR	23
V.	5.1. Linear model (modèle linéaire)	23
V.	5.2. Random Forest Regressor.....	24
V.	5.3. Support Vector Regressor (SVR)	24
V.	5.4. Decision Tree Regressor	24
V.	5.5. KNN Regressor.....	24
V.	6. validation d'un modèle QSAR	24

Conclusion

Conclusion.....	27
-----------------	----

Références

Références bibliographiques	29
-----------------------------------	----

Liste des abréviations

Mex : Multi Drug efflux (efflux de multiples médicaments).

MexA : Protéine adaptatrice située dans l'espace péri plasmique (entre les deux membranes).

MexB : Pompe principale d'efflux située dans la membrane interne.

OprM : Outer membrane porin M (protéine de la membrane externe qui forme un canal permettant l'expulsion des substances).

OprD : Outer membrane porin D.

ATB : Antibiotique.

AmpC : Ampicillin Chromosomal β -lactamase.

PSE : Pseudomonas-Specific Enzyme.

C4G : Céphalosporines de 4^{ème} génération.

BLSE : β -lactamase à Spectre Élargi.

TEM : Temoniera.

SHV : Sulfhydryl Variable.

PER : *Pseudomonas* Extended Résistance.

MBLs : Metallo- β -Lactamases.

EDTA : Acide Ethylène Diamine Tétracycline.

IMP : Imipenemase.

VIM : Verona Integron-encoded Metallo- β -lactamase.

SPM : São Paulo Metallo- β -lactamase.

AIM : Adelaide Imipenemase.

GIM : German Imipenemase.

NDM : New Delhi Metallo- β -lactamase.

OXA : Oxacilinase.

AAC : Aminoglycoside Acetyltransferase

ANT : Aminoglycoside Nucleotidyltransferase.

APH : Aminoglycoside Phosphotransferase.

RmtA : Ribosomal méthyl transférasse A.

QDRD : Quinolone Résistance-Determining Region.

GyrA : DNA gyrase subunit A.

ParC : Partition gène C.

ADN : Acide Désoxyribonucléique.

PhoP : Phosphate regulon protein P.

SHAP : SHapley Additive exPlanations.

QSAR : Relation Quantitative Structure-Activité.

RDKit : Rational Discovery Toolkit.

PaDEL : Partitioning and Descriptors for Easy Learning.

PubChem : Public Chemical Database.

CCC : coefficient de corrélation de concordance.

R² : Coefficient de détermination.

Q² : Coefficient de prédiction croisée.

LOO : Leave-One-Out.

LMO : Leave-many-out.

Listes des figures

Figure 1. Le mécanisme de la résistance aux antibiotiques chez <i>p.aeruginosa</i>	7
Figure 2. L'importance relative des descripteurs moléculaires : NumHAcceptors, NumHDonors, LogP, MW.	21
Figure 3. Analyse comparative de la performance des modèles de régression par différents algorithmes.....	23

Liste des tableaux

Table 1 . Principales résistances aux bêta-lactamines par <i>p.aeruginosa</i>	10
Table 2 . Evaluation des performances du modèle Random Forest Regressor.	25

Résumé

Pseudomonas aeruginosa est un agent pathogène nosocomial majeur, doté d'une remarquable capacité d'adaptation et de résistance aux antibiotiques. Face à l'échec croissant des traitements, les modèles de Relations Quantitatives Structure-Activité (QSAR) offrent une approche prometteuse pour prédire l'activité de nouvelles molécules. Leur application permet d'accélérer la découverte de composés actifs et la conception de thérapies plus efficaces.

Les données moléculaires actives contre *Pseudomonas aeruginosa* ont été extraites de la base de données ChEMBL34, puis filtrées sur la base des valeurs d'IC₅₀ (concentration inhibitrice médiane), converties en pIC₅₀ (-log(IC₅₀)), et finalement évaluées conformément à la règle de Lipinski. Des descripteurs structuraux ont été calculés à l'aide des outils RDKit et PaDEL afin de construire des modèles QSAR en utilisant plusieurs algorithmes, notamment le Régresseur Forêt Aléatoire (Random Forest Regressor). Le modèle jugé optimal a été validé au moyen de métriques de robustesse rigoureuses (R², Q², CCC, QF32, QLOO2, QLMO2), confirmant ainsi sa performance et sa stabilité prédictive.

Mots clés : QSAR, ChEMBL, Lipinski, RDKit, PaDEL

Abstract

Pseudomonas aeruginosa is a major nosocomial pathogen, endowed with a remarkable ability to adapt and resist antibiotics. In response to the increasing failure of treatments, QSAR models offer a promising approach to predict the activity of new molecules. Their application helps accelerate the discovery of active compounds and the design of more effective therapies.

Molecular data active against *Pseudomonas aeruginosa* were extracted from the ChEMBL34 database, filtered based on IC₅₀ values, converted into pIC₅₀, and then evaluated according to Lipinski's rule. Structural descriptors were calculated using RDKit and PaDEL in order to build QSAR models with several algorithms, including the Random Forest Regressor. The best model was validated using robust metrics (R², Q², CCC, Q²F3, Q²-LOO, Q²-LMO), confirming its performance and stability.

The best QSAR model was found when using the algorithm Random Forest Regressor ($R^2 = 0.83$, $Q^2-F3 = 0.89$, $CCC = 0.89$).

QSAR provides a reliable and fast tool to identify new therapeutic candidates in the fight against antibiotic resistance.

Key words: QSAR, ChEMBL, Lipinski, RDKit, PaDEL

ملخص

تعد الزائفة الزنجارية من أبرز العوامل المسببة للعدوى في المستشفيات، حيث تتميز بقدرتها العالية على التكيف ومقاومة المضادات الحيوية. وبسبب الفشل المتزايد للعلاجات التقليدية، توفر نماذج تقنية العلاقة الكمية بين البنية والنشاط (QSAR) نهجًا واعدًا للتنبؤ بفعالية الجزيئات الجديدة. ويسهم استخدامها في تسريع اكتشاف المركبات الفعالة وتصميم علاجات أكثر كفاءة.

تم استخراج البيانات الجزيئية الفعالة ضد الزائفة الزنجارية من قاعدة البيانات ChEMBL34، ثم حُسِبَت قيم IC_{50} وتحويلها إلى pIC_{50} ، ثم قُيِّمَت وفقًا لقواعد ليبينسكي.

كما حُسِبَت الواصفات الهيكلية باستخدام RDKit و PaDEL بهدف بناء علاقة كمية بين البنية والنشاط باستعمال عدة خوارزميات، من بينها خوارزمية انحدار الغابة العشوائية. وقد تم التحقق من أداء أفضل نموذج باستخدام معايير إحصائية قوية (R^2 , Q^2 , CCC , Q^2F3 , Q^2-LOO , Q^2-LMO)، مما أكد كفاءته واستقراره.

تم العثور على أفضل نموذج QSAR عند استخدام خوارزمية انحدار الغابة العشوائية ($R^2 = 0.83$, $Q^2-F3 = 0.89$, $CCC = 0.89$)، مما يؤكد أن تقنية العلاقة الكمية بين البنية والنشاط وسيلة واعدة وفعالة لتسريع عملية اكتشاف وتطوير الأدوية الجديدة في سياق مقاومة المضادات الحيوية.

الكلمات المفتاحية :

QSAR, ChEMBL, Lipinski, RDKit, PaDEL

Introduction

I. Introduction

Depuis quelques décennies, *Pseudomonas aeruginosa* ou bacille pyocyanique est reconnu comme agent pathogène hospitalier et peut causer des problèmes thérapeutiques. *Pseudomonas aeruginosa* consiste l'espèce type du genre *Pseudomonas* et représente environ 90% des isolats cliniques humaines appartenant à ce genre. (Troare., 2019).

Dotée d'un génome de grande taille lui conférant une flexibilité génétique remarquable, ainsi que d'un vaste arsenal de facteurs de virulence, *Pseudomonas aeruginosa* est impliquée dans une large gamme d'infections nosocomiales: des infections urinaires, des infections cutanées secondaires a des brûlures, et beaucoup d'autres infections en milieu hospitalier y compris les infections respiratoires, les infections des plaies chirurgicales, ainsi que des bactériémies, des otites et des infections oculaires, illustrant ainsi son exceptionnelle capacité d'adaptation et de pathogénicité.(Habbi et al., 2020).

Face à l'urgence croissante de la résistance de *Pseudomonas aeruginosa* aux antibiotiques en milieu hospitalier, les approches thérapeutiques actuelles se montrent de moins en moins efficaces. La modélisation QSAR peut contribuer significativement à l'identification et au développement de nouvelles stratégies thérapeutiques ciblant *Pseudomonas aeruginosa*, surmontant ainsi les limites des traitements actuels et ouvrant de nouvelles perspectives pour la construction de molécules innovantes et plus efficaces .(Aleksić et al.,2024).

Dans ce contexte, le développement de modèles QSAR (Quantitative structure-Activity Relation ship) s'avère essentiel pour prédire l'activité biologique de nouvelles molécules actives à partir de leurs caractéristiques structurales, réduisant ainsi le temps et le coût de développement de nouveaux agents antimicrobiens. L'application des QSAR à *Pseudomonas aeruginosa* ouvre la voie à la découverte ciblée de composés efficaces, contribuant à mieux gérer et prévenir les infections nosocomiales associées à cette bactérie. (Veyssiere., 2019).

L'objectif de cette étude QSAR est de modéliser la relation entre la structure chimique de composés et leur activité antibactérienne contre *Pseudomonas aeruginosa*, afin de prédire l'efficacité de nouvelles molécules et d'orienter la conception de futurs agents antimicrobiens.

Globalement, notre mémoire se compose de trois chapitres :

Chapitre 1: Une partie bibliographique rappellera la résistance aux antibiotiques de *Pseudomonas aeruginosa*, elle repose sur divers mécanismes tels que l'inactivation du médicament, la modification de la cible ou l'expulsion active. Ce phénomène constitue une menace majeure pour la santé publique mondiale.

Chapitre 2 : Matériel et méthode cette partie décrit les composés utilisés, leur activité biologique mesurée contre *p. aeruginosa* et préparation des données à partir de bases

publiques avec leurs activités biologiques associées. Les descripteurs moléculaires calculés. Les données ont ensuite été normalisées et préparées pour la modélisation QSAR.

Chapitre 3 : les résultats du modèle QSAR montrent une bonne corrélation entre la structure chimique des composés et leur activité antibactérienne. Le modèle obtenue a été validé statistiquement et montre un bon pouvoir prédictif.

Chapitres 1:

Revue bibliographique

II- Description générale

Pseudomonas aeruginosa a été isolée pour la première fois en 1882 par le bactériologiste français Émile Gessard, à partir de pansements infectés. Il observa alors la production de pigments bleu-vert caractéristiques, la pyocyanine et la pyoverdine.

Sur le plan nomenclature, le terme *Pseudomonas*, dérivé du grec "pseudo" (similaire) et "monas" (unité), désignait initialement les micro-organismes. L'épithète spécifique *aeruginosa*, du latin signifiant "vert de gris", fait référence au pigment caractéristique produit par la bactérie, conférant aux colonies leur coloration typique. *P. aeruginosa* constitue l'espèce type du genre *Pseudomonas* (Kerr et al., 2009).

Depuis lors, cette bactérie a été largement étudiée en raison de son rôle dans les infections opportunistes, en particulier chez les patients immunodéprimés ou atteints de maladies chroniques telles que la mucoviscidose (Mérens et al., 2013).

P. aeruginosa est particulièrement connu pour sa résistance aux antibiotiques et sa capacité à survivre dans des environnements hostiles, ce qui en fait un problème majeur en milieu hospitalier. Cette espèce ubiquitaire colonise une grande variété d'habitats, incluant les milieux aquatiques (eaux stagnantes, rivières, lacs, piscines, systèmes d'eau hospitaliers), les sols humides riches en matière organique, la rhizosphère des plantes et les surfaces humides comme les installations sanitaires et le matériel médical (Eyquem et al., 2005).

De plus, elle peut coloniser la peau et les muqueuses des humains et des animaux, en particulier chez les personnes immunodéprimées. Son adaptabilité remarquable lui permet de persister dans des environnements oligotrophiques, expliquant sa prévalence en milieu hospitalier et son implication dans les infections nosocomiales (Kerr et al., 2009).

III. La résistance aux antibiotiques

Pseudomonas aeruginosa est une bactérie naturellement résistante à un large éventail d'antibiotique, en raison de plusieurs mécanismes intrinsèques. Elle produit notamment une bêta-lactamase chromosomique inductible de classe C, non inhibée par le clavulanate, et possède une faible perméabilité membranaire, limitant l'entrée des antibiotiques (Barir et al., 2011).

De plus, un des mécanismes majeurs de cette résistance est le système d'efflux actif MexAB-OprM, fonctionnant en permanence chez les souches sauvages et expulsant les antibiotiques avant qu'ils n'agissent. Ce système confère à la bactérie une résistance naturelle à de nombreuses classes d'antibiotiques, dont les bêta-lactamines (pénicillines et céphalosporines) et les aminosides comme la kanamycine (Mentalagre et al .,2016).

Ainsi, *P. aeruginosa* est naturellement insensible à plusieurs familles d'antibiotiques, telles que les pénicillines des groupes V,G,M et A, la majorité des céphalosporines de troisième génération, ainsi que les quinolones de première génération (Mentalagre et al .,2016).

Cependant, cette bactérie peut également développer une résistance acquise, lorsqu'une ou plusieurs souches initialement sensibles deviennent résistantes à un ou plusieurs antibiotiques. Cette résistance peut résulter de mutation chromosomique ou de l'acquisition de gènes via des éléments génétiques mobiles comme les plasmides. Ces multiples stratégies de résistance font de *P. aeruginosa* un pathogène particulièrement difficile à traiter (Souley et al., 2002).

Les mécanismes de la résistance (figure 01):

- a-** Sécrétion d'enzymes d'inactivation : qui va décomposer l'ATB.
- b-** Modification de la cible : qui entraîne une perte d'affinité de l'ATB pour sa cible.
- c-** Imperméabilité : notamment par diminution de la matière des porines chez les bacilles à Gram négatif.
- d-** L'efflux des antibiotiques : Il s'agit de l'expulsion de l'antibiotique hors de la cellule bactérienne à travers la membrane cytoplasmique, par le biais d'une pompe d'efflux constituée d'une protéine synthétisée par la bactérie. Le fonctionnement de cette pompe requiert de l'énergie.

Le motif commun à ces différents mécanismes de résistance est d'empêcher l'interaction de l'ATB avec sa cible. Et parfois, c'est une association de plusieurs mécanismes (Boudouda et al., 2015).

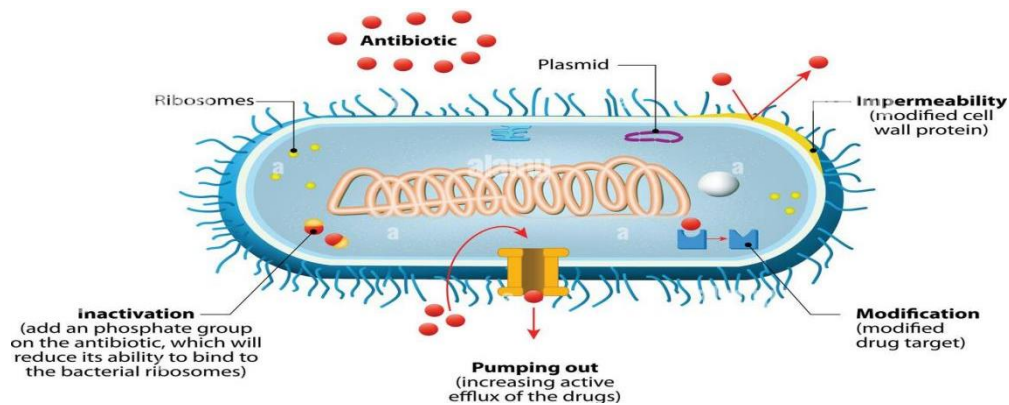


Figure 1. Le mécanisme de la résistance aux antibiotiques chez *p. aeruginosa* (Boudouda et al., 2015).

III. 1. Mécanismes enzymatiques

III. 1.1. Hydrolyse de bêta-lactamines

La résistance de *Pseudomonas aeruginosa* aux bêta-lactamines repose sur plusieurs mécanismes complémentaires. Le principal est la production d'une bêta-lactamase chromosomique de classe C (AmpC), une enzyme capable d'hydrolyser l'anneau bêta-lactame, rendant l'antibiotique inefficace. Cette enzyme est naturellement présente et peut être induite en présence d'antibiotiques, ce qui aggrave la résistance. En plus de cette hydrolyse enzymatique, la bactérie présente une perméabilité membranaire réduite, due à une altération ou une perte de porines, notamment OprD, limitant ainsi l'entrée des antibiotiques dans la cellule. Enfin, *P. aeruginosa* utilise des système MexAB-OprM, pour expulser activement les antibiotiques hors de la cellule avant qu'ils n'atteignent leur cible. L'association de ces mécanismes confère à cette bactérie une résistance naturelle élevée aux bêta-lactamines, y compris à certaines pénicillines, céphalosporines et même carbapénèmes (Barir et al., 2011)

III. 1.2. Résistance par hyperproduction de la céphalosporinase Amp C

La bêta-lactamase chromosomique de type AmpC a été décrite chez un grand nombre de bacilles à Gram négatif, notamment *Pseudomonas aeruginosa*. Des mutations de sa production peuvent entraîner une expression stable et régulation de sa production peuvent entraîner une expression stable et élevée de l'enzyme AmpC, ce qui réduit l'efficacité de l'ensemble des β -lactamines, à l'exception des carbapénèmes (Asma., 2012). La régulation de l'expression du gène AmpC implique plusieurs gènes, dont AmpR, AmpD et AmpG. Le gène AmpR code un activateur transcriptionnel du gène AmpC, qui peut être induit en présence de β -lactamines. Cette induction est contrôlée (ou réprimée) par une protéine codée par le gène AmpD (Pascal., 2010).

Le gène codant pour AmpC chez *P. aeruginosa* n'a pas encore été retrouvé sur des plasmides. Par conséquent, il n'est généralement pas transféré à d'autres espèces, contrairement aux AmpC plasmidiques observés chez les entérobactéries (Gougeon., 2017).

III. 1.3. Résistance par production de pénicillinases

Parmi les pénicillinases produites par *Pseudomonas aeruginosa* on distingue principalement les enzymes de type PSE (*Pseudomonas* Specific enzyme), telles que PSE-1, PSE-3, PSE-4 (Bert et al., 2002).

L'enzyme PSE-1 est la plus répandue, représente environ 90% des cas). Ces enzymes sont étroitement apparentées, et ne diffèrent que par 1 ou 2 acides aminés (Buch et al., 2010).

Les pénicillinases de type PSE sont capables d'hydrolyser les carboxypénicillines (comme le ticarcilline) ainsi que les uréidopénicillines (comme le pipéracilline), et montrent également une activité contre certaines céphalosporines, notamment la 4^{ème} génération (C4G) les céphalosporines de quat et sont inactives sur le ceftazimide et les carbapénèmes (Buch et al., 2010).

III. 1.4. Résistance à spectre étendue (BLSE)

Les bactéries productrices de bêta-lactamases à spectre étendu (BLSE), sont des enzymes récemment apparues à la suite de mutations des pénicillinases. Elles sont plasmidiques donc transférables et sensibles à l'action des inhibiteurs enzymatiques. Elles sont très actives contre les pénicillines et moyennement actives contre les céphalosporines de première génération. Les mutations génétiques à l'origine des BLSE élargissent le spectre de ces enzymes et touchent également les céphalosporines de troisième génération (céftazidime et céfotaxime) et les monobactames (aztréonam). Les bactéries produisant une BLSE n'hydrolysent pas les céphamycines (céfoxitine) ni les carbapénèmes (Giske et al., 2006).

Cinq types de BLSE de classe A (TEM, SHV, PER, VEB et GES) ont été détectés chez *P. aeruginosa*. Il existe actuellement neuf types connus de BLSE GES, jusqu'à présent quatre de ces types de GES (GES-1, -2, -8 et -9) ont été trouvés chez *P. aeruginosa*. Il nous faut bien distinguer les résistances acquises aux céphalosporines de 3^{ème} génération (C3G), soit par hyperproduction de céphalosporinase, soit par BLSE car les phénotypes de résistance sont différents. Ces enzymes sont habituellement détectées par une synergie entre une C3G (notamment la céftazidime) ou l'aztréonam et l'acide clavulanique. (Proil et al., 2001).

La classification des β -lactamases selon Ambler repose sur la structure moléculaire des enzymes, réparties en quatre classes principales (A à D). La classe A regroupe notamment les pénicillinases telles que les enzymes des familles PSE, TEM, SHV, ainsi que d'autres enzymes à spectre élargi comme PER, VEB, GES et IBC. Les carbapénémases sont majoritairement classées en classe B, incluant les métallobêta-lactamases de type IMP et VIM. La classe C correspond principalement aux céphalosporinases de type AmpC, tandis que la classe D rassemble les oxacillinases (OXA), subdivisées en plusieurs groupes en fonction de leur profil enzymatique (Groupes I à V). Un résumé détaillé des enzymes représentatives de chaque classe est présenté dans le Tableau 1.

TABLEAU 1. PRINCIPALES RESISTANCES AUX BETA-LACTAMINES PAR *P. AERUGINOSA*.

Classification d'Ambler	Groupe	Enzymes
Classe A/pénicillinase	PSE	PSE-1,3,5; CARB-3,4
	TEM	TEM-1,2,4,21,24,42,45
	SHV	SHV-2a,5,12
	Autres	PER-1; VEB-1a, 1b, 2; GES-1,2; IBC-2
Classes B // carbapénémases		IMP-17, VIM-1 à VIM-4
Classes C / céphalosporinases		AmpC
Classes D / oxacilinasés	I	OXA-5,7,10(PSE-2),11,13,14 à 17,19,28,35,45
	II	OXA-2,3,15,20,32,34,36,53
	III	OXA-1,4,30,31
	IV	OXA-9
	V	LCR-1, NPS-1

III. 1.5. Résistance par production de carbapénémases (MBLs)

Les carbapénémases de type métallo- β -lactamases (MBLs) sont devenues les carbapénémases les plus répandues et les plus significatives chez *P. aeruginosa*. Ces enzymes possèdent dans leur site actif un cation indispensable à leur activité qui est invariablement le zinc (Zn^{2+}). En outre, l'activité de ces enzymes est inhibée par l'addition de chélateur d'ions bivalents (EDTA). Ce sont des métalloprotéines, chromosomiques ou plasmiques, qui ont une activité catalytique beaucoup plus forte que les autres β -lactamases et hydrolysent toutes les β -lactamines sauf l'aztréonam (Obristsch et al 2004). Ces carbapénémases acquises constituent sept groupes décrits chez *P. aeruginosa* : IMP (active sur l'IMipenem), VIM (Verona Integron-encoded Metallo- β -lactamase), SPM (Sao-paulo Metallo- β -lactamase), AIM (Australia IMpenemase), GIM (German IMpenemase), NDM-1 (New Delhin Metallo- β -lactamase) et FIM-1. Les types IMP et VIM demeurent les MBLs les plus rapportées dans le bassin méditerranéen chez *P. aeruginosa* (Mentalegre., 2016).

III. 1.6. Résistance aux oxacillinases classe D

Chez *P. aeruginosa* des BLSE dérivées des enzymes OXA-10 et OXA-2 ont été isolées. Cela inclut les variantes (OXA-10, 11, 14, 15, 16, 19), qui sont capables d'hydrolyser un large éventail d'antibiotiques bêta-lactamase notamment certaines céphalosporines de 3^{ème} génération l'imipénème et le méropénème. L'aztréonam et la pipéracilline sont moins touchés, mais leur activité n'est pas inhibée par l'acide clavulanique ou le tazobactam. Ces enzymes appartiennent à la classe D des β -lactamines, et sont le plus souvent portées par des plasmides, ce qui facilite leur transfert horizontal entre bactéries et favorise la propagation rapide de la résistance en milieu hospitalier. En outre, l'enzyme OXA-18 a également été identifiée chez *P. aeruginosa*. Contrairement aux autres OXA-18 est la seule oxacillinase de classe D connue pour être inhibée par l'acide clavulanique. Elle est codée par un gène chromosomique (*blaOXA-18*), ce qui limite sa mobilité génétique, mais sa présence demeure préoccupante en termes de résistance clinique.

III. 1.7. Résistance aux aminosides

La résistance aux aminosides chez *Pseudomonas aeruginosa* repose en grande partie sur la production d'enzymes de modification qui inactivent ces antibiotiques en altérant leur structure chimique, les empêchant ainsi de se lier efficacement à leur cible ribosomique. Trois types principaux d'enzymes sont impliquées : les N-acétyltransférases (AAC), qui acétylent les groupes amino, les O-nucléotidyltransférases (ANT), qui adénylylent les groupes hydroxyles, et les O-phosphotransférases (APH), qui phosphorylent également des groupes hydroxyles. Ces enzymes, souvent codées par des gènes localisés sur des éléments

génétiques mobiles tels que les plasmides, transposons ou intégrons, facilitent la dissémination de la résistance entre souches bactériennes. Chez *P. aeruginosa*, les enzymes les plus fréquemment rencontrées sont AAC(6')-I, ANT(2'') et parfois des variantes d'APH, conférant une résistance marquée à plusieurs aminosides, et contribuant ainsi au développement de profils de multirésistante. (Macfarlane et al., 2000).

III. 2. Mécanismes non enzymatiques

III. 2.1. Surexpression de système d'efflux

Plusieurs protéines agissent en tant que composant actifs des systèmes d'efflux Opr M, Opr J, Opr N, sans oublier le système Mex AB-Opr M .

Chaque système a une spécifique se substrat :

- **Mex AB-Opr M:** efflue la quasi-totalité des bêta-lactamines sauf l'imipénème, les fluoroquinolones, les macrolides, la lincomycine, la tétracycline, le triméthoprim, le chloramphénicol et la novobiocine. (Masuda et al., 2000).

- **Opr M:** élimine les aminoglycosides, les quinolones, les bêta-lactamines zwitterioniques (présentant un même nombre de charges électriques de signes opposées tel que le céfépime), les macrolides et la tétracycline . (Muller., 2012) .

- **Opr N :** efflue quinolones, les aminoglycosides, le triméthoprim, le chloramphénicol et certaines bêta-lactamines comme les carbapénèmes mais aussi la ceftazidime, le céfépime, la ticarcilline ou l'aztréona . (Lanes .,2011).

- **Opr J :** élimine les quinolones, les bêta-lactamines zwitterioniques, les macrolides, la tétracycline, le chloramphénicol et le triméthoprim. (Jeannot., 2008).

III. 2.2. Résistance due à la diminution de la perméabilité

Dans les bactéries à Gram négatif, les porines bactériennes sont une des voies principales d'entrée pour les antibiotiques usuels comme les bêta-lactamines et les fluoroquinolones . Cette protéine canalaire de la membrane externe possède un site spécifique de liaison aux carbapénèmes, et permet la pénétration sélective de l'imipénème. Des modifications de la quantité absolue ou de l'état fonctionnel de ces porines ont pour conséquence une diminution de la diffusion des antibiotiques empruntant cette voie de pénétration. Ce mécanisme par diminution de perméabilité peut entraîner une résistance croisée à plusieurs familles d'antibiotiques. Chez *P.* mécanisme le plus fréquent de résistance à l'imipénème. (Nordmann., 2003).

Par ailleurs, certaines souches cliniques, l'altération des différents gènes impliqués dans la biosynthèse des lipopolysaccharides (des composants de la membrane extérieur de la

bactérie) comme le locus *rfb* conduit au phénotype «rough» (aspect sec et rugueux) accompagné des fois d'une baisse modérée de la sensibilité à l'ensemble des aminosides. (Doi et al., 2007).

III. 2.3. Résistance par efflux actif

Chez *P. aeruginosa*, seul le système d'efflux actif MexXY (Opr M) est capable d'expulser les aminosides vers l'extérieure, contribuant ainsi à la résistance à cette classe d'antibiotique. Ce système est naturellement réprimé dans les souches sauvages, la sur expression constitutive de l'opéron *mexXY* qui en découle conduit à une résistance modérée aux aminosides, ainsi qu'aux fluoroquinolones et au céfépime (Doi et al., 2007).

III. 2.4. Résistance par modification de la cible ribosomale

La méthylation de l'ARNr 16s a récemment émergé comme nouveau mécanisme de résistance aux aminosides parmi les agents pathogènes à Gram négatif telle que *P. aeruginosa*, capable de modifier non pas l'aminoside mais la structure ribosomale sur laquelle il se fixe, l'ARN 16S (Doi et al., 2007).

Le gène *rmt A* (Ribosomal méthyl transférase) qui code pour une méthylase de l'ARNr 16s procure à la bactérie une résistance de haut niveau à l'amikacine, la tobramycine, l'isépamicine, l'arbekacine et la gentamycine. Toutefois, il faut souligner que la résistance par mutation du gène codant l'ARNr 16S n'est efficace que chez les bactéries possédant une ou éventuellement deux copies des gènes de l'ARNr 16S. Mais *P. aeruginosa* possède quatre copies de ce gène, ainsi la probabilité qu'elles soient affectées toutes les quatre par des mutations identiques est très faible et la résistance par ce mécanisme est peu probable (Yokoyama et al., 2003). Par ailleurs, La modification de la cible principale des fluoroquinolones, à savoir la gyrase de l'ADN (également appelée topoisomérase II), se produit par des mutations ponctuelles dans les régions où se lie l'antibiotique, appelées QRDR (Quinolone Resistance Determining Regions), situées dans les gènes *gyrA* et *gyrB*, qui codent respectivement les deux sous-unités de l'enzyme, GyrA et GyrB.

Les modifications de la cible secondaire, la topoisomérase IV, surviennent suite à des mutations ponctuelles dans les gènes *parC* et *parE*, qui codent respectivement pour les deux sous-unités de cette enzyme, *parC* et *parE* (Jcaoby., 2005).

L'association de mutation à la fois dans *GyrA* et *parC* conduit à de très haut niveau de résistance, Contrairement à d'autres mécanismes enzymatiques de résistance (comme la production de β -lactamases), la résistance aux fluoroquinolones n'est pas enzymatique,

mais génétique, liée à des mutations structurales dans les cibles de l'antibiotique (Akasaka et al., 2001).

III. 2.5. Rôle de l'opéron dans la résistance

Le gène *oprH* se trouve associé dans un même opéron aux gènes *phoP* et *phoQ* qui codent pour un système de transduction de signal.

L'hyperexpression de la protéine de membrane externe *OprH* a été constatée dans les souches mutantes de *P. aeruginosa* montrant une résistance élevée aux aminosides, à l'exception de la gentamycine. Il semble que le système *phoP-phoQ* joue un rôle indirect dans la résistance aux aminosides.

Ainsi, il a été observé au niveau de différents mutants de *P. aeruginosa* affectés dans les gènes *phoP* ou *phoQ* une légère augmentation de la résistance à l'amiacine et à la streptomycine (Macfarlane et al., 2000).

Chapitres 2 :

Matériels et méthodes

IV. Matériels et méthodes

IV. 1. Collecte des données

Les molécules présentant une activité potentielle contre *Pseudomonas aeruginosa* ont été recherchés dans la banque de données ChEMBL34. Après extraction, ces molécules, identifiées par leur identifiant ChEMBL ID348, ont été importées dans l'environnement de développement Google Colab en vue pour le traitement et analyse.

IV. 2. Préparation des données

La classification des molécules a été réalisée sur la base de leurs valeurs de concentration inhibitrice à 50 % (IC₅₀), selon des seuils prédéfinis :

- Composés actifs : IC₅₀ < 1 µM (1000 nM)
- Composés inactifs : IC₅₀ > 10 µM (10000 nM)
- Composés intermédiaires : 1 µM ≤ IC₅₀ ≤ 10 µM

Pour améliorer l'analyse, une opération de filtrage des données a été effectuée sur les données, où les entrées 'intermédiaires' a été exclues. Ce processus a été entrepris dans le but de se concentrer uniquement sur les composés "active" et "inactive", contribuant ainsi à la construction de modèles prédictifs.

Ensuite, IC₅₀, ont été convertis en pIC₅₀, pour faciliter l'interprétation de l'efficacité des composés pIC₅₀= -log(IC₅₀).

IV. 3. Classification selon la loi de Lipinski

Les descripteurs structuraux tels que le poids moléculaire et le nombre de liaisons hydrogène (donneurs et accepteurs) ainsi que le LogP, ont été calculés à l'aide RDKit. Ces données ont ensuite permis d'évaluer les composés selon les critères de la règle de Lipinski, à savoir : un poids moléculaire inférieur ou égal à 500 Da, un LogP ne dépassant pas 5, un nombre de donneurs de liaisons hydrogène (-OH,-NH) ne doit pas être supérieur à 5, et un nombre d'accepteurs de liaisons hydrogène ne doit pas être supérieur à 10 afin d'identifier ceux présentant les valeurs les plus élevées.

Pour identifier les attributs les plus influents sur la prédiction de l'activité biologique des composés, Les valeurs SHAP ont été calculés pour l'intégralité du jeu de test.

IV. 4. Calcul des descripteurs

Pour la construction d'un modèle QSAR, PubChem ont été calculées à l'aide de PaDEL et RDKit.

Parmi les descripteurs générés, les empreintes digitales PubChem (PubChem fingerprints) furent spécifiquement ciblées et calculées. Ces empreintes sont des vecteurs binaires représentant la présence ou l'absence de caractéristiques chimiques prédéfinies ou de sous-

structures spécifiques au sein de chaque molécule. Elles offrent une description numérique rapide et efficace de la topologie moléculaire et des motifs fonctionnels importants pour l'activité biologique.

IV. 5. Construction d'un modèle QSAR

Pour la construction d'un modèle QSAR, les données ont été divisées en deux: 80 % des données ont été dédiées à l'entraînement du modèle, tandis que les 20 % restants ont été réservés à la validation.

Pour établir la relation entre les descripteurs et l'activité biologique, divers modèles de régression ont été utilisés, notamment le modèle linéaire, le Random Forest Regressor, le Support Vector Regressor, le Decision Tree Regressor et le KNN Regressor.

IV. 6. Validation de modèle

Pour valider le meilleur modèle de QSAR, une première évaluation a été réalisée en se basant uniquement sur le score R^2 et le Q^2 (coefficient de validation croisée).

Une fois le modèle le plus performant identifié sur la base de ces deux critères, des analyses statistiques complémentaires ont été appliquées uniquement sur ce modèle

Telles que Q_2F_3 , CCC (coefficient de corrélation de concordance), Q^2 -Loo (leave-one-out) et Q^2 -LMO (leave-many-out).

Chapitres 3 :

Résultats et discussion

V. Résultats et discussion

V. 1. Collecte et nettoyage des données

Un jeu de données de molécules bioactives ciblant *Pseudomonas aeruginosa* a été collecté à partir de la base de données ChEMBL 34.

Initialement, 1504 molécules ont été récupérées. Suite aux étapes rigoureuses de nettoyage des données (incluant la suppression des valeurs manquantes dans les colonnes clés et l'élimination des doublons basée sur les SMILES canoniques), le nombre de molécules a été réduit à 1028. Ce jeu de données final, composé de 1028 molécules, a ensuite été préparé pour la modélisation QSAR par la transformation des données et le calcul de descripteurs physico-chimiques, fournissant ainsi des descripteurs pertinents et les valeurs pIC₅₀ correspondantes pour l'activité.

V. 2. Classification des données

La classification des molécules basée sur leurs valeurs d'IC₅₀ est une pratique standard en découverte de médicaments et dans les études QSAR, visant à catégoriser les composés selon leur puissance inhibitrice. Dans cette analyse, les molécules ciblant *Pseudomonas aeruginosa* ont été réparties en trois catégories :

- V. 2.1. **Molécules Actives** : Ces molécules présentent une puissance inhibitrice élevée, généralement définie par une valeur d'IC₅₀ inférieure ou égale à 1000 nM (ce qui correspond à un pIC₅₀ supérieur ou égal à 6). Ces composés sont considérés comme des candidats prometteurs pour des investigations ultérieures en tant que molécules de tête potentielles.
- V. 2.2. **Molécules Inactives** : Ces molécules présentent une activité inhibitrice faible ou inexistante, avec des valeurs d'IC₅₀ supérieures ou égales à 10000 nM (ou un pIC₅₀ inférieur ou égal à 5). Ces composés ne sont généralement pas considérés pour un développement ultérieur, mais leurs structures peuvent être utiles comme contrôles négatifs ou pour comprendre quelles caractéristiques structurelles ne sont pas propices à l'activité.
- V. 2.3. **Molécules Intermédiaires** : Ces molécules montrent une activité inhibitrice modérée, avec des valeurs d'IC₅₀ se situant entre les seuils des composés actifs et inactifs (entre 1000 nM et 10000 nM, ou un pIC₅₀ entre 5 et 6). Bien que moins puissantes que les composés actifs, elles peuvent néanmoins fournir des informations précieuses pour comprendre les relations structure-activité ou pourraient être optimisées pour améliorer leur puissance.

Cette classification est cruciale pour la construction d'un modèle QSAR car elle permet d'entraîner le modèle à distinguer les composés ayant différents niveaux d'activité biologique. En se concentrant sur les caractéristiques structurales qui différencient les composés actifs des composés inactifs, il est possible de développer des modèles prédictifs capables d'identifier de nouveaux candidats médicaments potentiels à partir de vastes bibliothèques de molécules. Après la suppression des molécules intermédiaires, il reste 910 molécules, permettant de créer une distinction plus claire entre les composés actifs et inactifs, ce qui peut améliorer la performance du modèle QSAR.

V. 3. Classification selon la loi de Lipinski

Suite à l'application des critères de Lipinski, les molécules considérées comme ayant une activité « intermédiaire 256 » ont été éliminées.

Nombre de molécules actives qui respectent les règles de Lipinski : on a identifié 237 molécules actives qui respectent les critères de Lipinski.

Les molécules inactives conformes aux critères de Lipinski : 417 molécules non actives. Ont été alors pu déterminer le nombre de molécules actives et inactives conformes à ces règles.

La détection de 237 molécules actives conformes aux critères de Lipinski a été satisfaisante. Cela a révélé qu'une part importante des composés ayant démontré une efficacité contre *Pseudomonas aeruginosa* possédait également des caractéristiques physico-chimiques propices à la biodisponibilité orale. Ces composés sont considérés comme des candidats prometteurs pour un développement futur en tant que médicaments oraux. En ce qui concerne les 417 molécules inactives respectant les critères de Lipinski, cela signifie que beaucoup d'entre elles, bien qu'elles soient conformes à ces critères, pourraient ne pas avoir une activité contre la cible spécifique. C'est un résultat anticipé, car les critères de Lipinski ne prédisent que les propriétés globales des médicaments et non leur efficacité contre une cible biologique précise.

La clarification le mode SHAP (SHapley Additive exPlanations) est une approche essentielle pour comprendre les prédictions des modèles d'apprentissage automatique. Cette technique est mise en œuvre pour déterminer une valeur à chaque caractéristique atomique, éclairant ainsi l'importance de sa contribution à une prédiction particulière.

La figure 02 représente de manière visuelle l'importance moyenne absolue de chaque caractéristique dans le modèle.

En analysant ce diagramme, il a été détecté les descripteurs moléculaires (le poids moléculaire (MW), le LogP, le nombre d'accepteurs de liaison hydrogène (NumHAcceptors) et le nombre

de donneurs de liaison hydrogène (NumHDonors)) ayant eu l'influence la plus importante sur la façon dont le modèle a estimé les molécules.

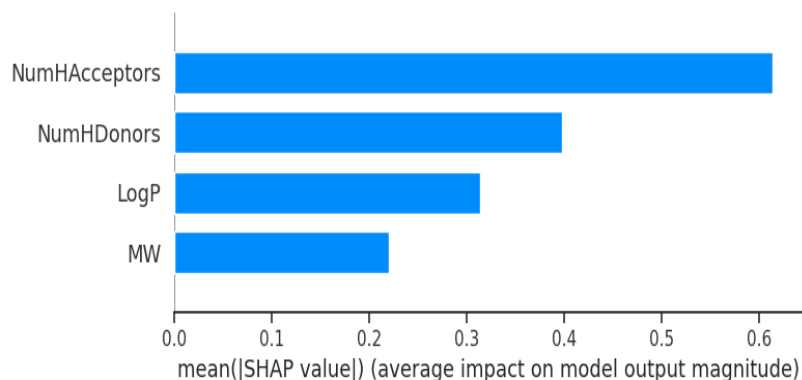


Figure 2. L'importance relative des descripteurs moléculaires : NumHAcceptors, NumHDonors, LogP, MW.

V. 3.1. NumHAcceptors (Nombre d'accepteurs de liaisons hydrogène)

Cette caractéristique afficha la valeur SHAP moyenne absolue la plus élevée, s'établissant à environ 0.62. Ceci indiqua que le nombre d'accepteurs de liaisons hydrogène constitua la propriété moléculaire la plus influente dans les prédictions du modèle QSAR. Il fut rappelé que les accepteurs de liaisons hydrogène sont des atomes (généralement N, O, F) possédant des paires d'électrons libres capables de former de telles liaisons (Lundberg., 2017). (**Figure 2**)

Une importance prononcée de NumHAcceptors suggéra que les interactions par liaisons hydrogène jouèrent un rôle crucial dans l'activité biologique prédite par le modèle. Dans de nombreux contextes pharmacologiques, les liaisons hydrogène sont reconnues comme fondamentales pour la reconnaissance moléculaire et l'établissement des liaisons ligand-récepteur (Klebe et al, 2006). Un nombre optimal d'accepteurs d'hydrogène s'avérait souvent requis pour une affinité de liaison adéquate et une activité biologique. Ce résultat s'alignait parfaitement avec les règles de Lipinski, lesquelles soulignèrent l'importance des liaisons hydrogène pour la perméabilité et l'absorption orale des médicaments (lipinski. 2001).

V. 3.2. NumHDonors (Nombre de donneurs de liaisons hydrogène)

NumHDonors se positionna comme la deuxième caractéristique la plus importante, avec une valeur SHAP moyenne absolue d'environ 0.40. Les donneurs de liaisons hydrogène furent généralement identifiés comme des groupes N-H ou O-H.

À l'instar des accepteurs, l'importance des donneurs de liaisons hydrogène accentua le rôle prépondérant des interactions par liaisons hydrogène dans la prédiction de l'activité. Les liaisons hydrogène furent considérées comme des forces intermoléculaires clés régissant les

interactions entre les molécules médicamenteuses et leurs cibles biologiques (Klebe et al., 2006). L'équilibre entre les donneurs et accepteurs de liaisons hydrogène fut souvent jugé critique pour l'activité. (**Figure 2**)

V. 3.3. LogP (Coefficient de partage octanol-eau)

Le LogP démontra une importance significative, avec une valeur SHAP moyenne absolue d'environ 0.32. Le LogP fut défini comme une mesure de la lipophile d'une molécule.

La lipophile, propriété physico-chimique fondamentale en chimie médicinale, influence fortement l'absorption, la distribution, le métabolisme et l'excrétion (ADME) des composés (Han et al. 2019). Un LogP optimal s'avérait souvent nécessaire pour que la molécule atteigne sa cible biologique. Des valeurs de LogP trop faibles pouvaient entraîner une mauvaise perméabilité membranaire, tandis que des valeurs trop élevées pouvaient conduire à une mauvaise solubilité aqueuse et à un fort piégeage dans les membranes lipidiques ou le tissu adipeux (Han et al., 2019). L'importance de LogP dans ce modèle QSAR fut donc attendue et considérée comme cruciale pour la prédiction de l'activité biologique effective. (**Figure 2**)

V. 3.4. MW (Poids moléculaire)

Le poids moléculaire (MW) se révéla être la caractéristique la moins importante parmi celles affichées, avec une valeur SHAP moyenne absolue d'environ 0.22.

Le poids moléculaire, bien qu'étant un descripteur de taille simple, fut reconnu comme important. Des poids moléculaires excessivement élevés purent entraver la perméabilité membranaire et la biodisponibilité orale, souvent en raison d'une taille excessive pour traverser les pores des membranes ou d'une affinité démesurée pour les protéines plasmatiques (Klebe et al, 2006). Bien que son influence fut moindre par rapport aux propriétés liées aux liaisons hydrogène et à la lipophile dans le graphique présenté, son impact demeura notable. (**Figure 2**)

Ces résultats s'accordèrent parfaitement avec les principes établis de la chimie médicinale et les règles de Lipinski concernant la "drug-likeness" (Klebe et al, 2006).

V. 4. Calcul des descripteurs

Pour la modélisation des relations structure-activité quantitative (QSAR), la génération et la manipulation des structures chimiques sont fondamentales, et elles sont efficacement réalisées grâce à l'utilisation combinée de RDKit et PaDEL-Descriptor. RDKit fournit des fonctionnalités chimiomatiques critiques. Gérant implicitement la conversion des chaînes SMILES en représentations moléculaires, une première étape indispensable pour tout calcul de descripteur. Ensuite, PaDEL-Descriptor est employé pour calculer un ensemble vaste et diversifié de descripteurs moléculaires, notamment les empreintes digitales PubChem, qui

sont des clés structurales représentant des caractéristiques chimiques prédéfinies au sein de chaque molécule (Landrum et al, 2006).

Cette traduction numérique des structures chimiques est absolument essentielle, car elle fournit les données brutes nécessaires aux algorithmes d'apprentissage automatique pour construire un modèle QSAR capable de prédire l'activité biologique d'une molécule en fonction de sa structure (Consonniet al, 2009).

V. 5. Comparaison des Modèles de QSAR

La Figure (3) représente une comparative de la performance des modèles de régression par différents algorithmes.

Cette figure a permis de visualiser la performance de différents algorithmes sur l'ensemble de données de test.

Le R^2 mesure la proportion de la variance de variable dépendante (ici, probablement les valeurs pIC_{50}) qui peut être prédite à partir des variables indépendantes par le modèle. Plus le R^2 se rapproche de 1, plus la capacité prédictive du modèle est optimale (Breimen., 2001).

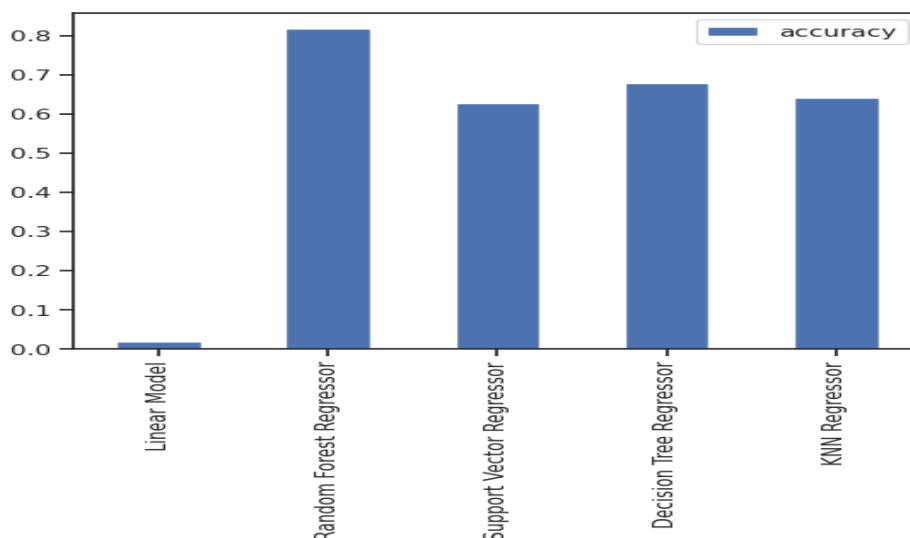


Figure 3. Analyse comparative de la performance des modèles de régression par différents algorithmes.

V. 5.1. Linear model (modèle linéaire) : la barre pour le modèle linéaire est extrêmement courte, indiquant un R^2 proche de 0,01. Cela suggère que le modèle linéaire a une très faible capacité à expliquer la variance des données de test. Il est possible que la relation entre les variables ne soit pas linéaire ou que ce modèle soit trop simple pour capturer la complexité des données. (**figure 3**)

V. 5.2. Random Forest Regressor : ce modèle présente la barre la plus haute, avec un R^2 d'environ 0,82. C'est la performance la plus élevée parmi tous les modèles évalués. Une valeur de R^2 de 0,82 signifie que le Random Forest Regressor explique environ 82% de la variance des valeurs réelles sur l'ensemble de test. Cela indique une très bonne capacité de prédiction et de généralisation pour ce modèle. **(figure 3)**

V. 5.3. Support Vector Regressor (SVR) : le SVR obtient un R^2 d'environ 0,62. C'est une performance supérieure à celle du Random Forest Regressor. Cela implique que le SVR explique environ 62% de la variance des données de test. **(figure 3)**

V. 5.4. Decision Tree Regressor : ce modèle attient un R^2 d'environ 0,68. Sa performance est meilleure que celle du SVR mais demeure inférieur au Random Forest Regressor, expliquant environ 68 % de la variation des données de test **(figure 3)**

V. 5.5. KNN Regressor : le KNN Regressor affiche un R^2 d'environ 0,64. Sa performance est comparable à celle SVR et de décision Tree Regression, expliquant environ 64% de la variance des données de test. **(figure 3)**

Le Random Forest Regressor est clairement le modèle le plus performant, affichant le R^2 le plus élevé sur l'ensemble de test. Les modèles comme le Décision Tree Regressor, SVR et KNN Regressor offrent des performances acceptables mais inférieures. Le modèle linéaire, quant à lui, est manifestement inadapté à ces données. Ces résultats suggèrent que les modèles basés sur les arbres (Random Forest, Decision Tree) ont une meilleure capacité à capturer les relations non linéaires et les interactions complexes au sein des données par rapport aux modèles. **(Figure 3)**

V. 6. validation d'un modèle QSAR

Le tableau 02 présente une évaluation complète des performances du modèle Random Forest Regressor en utilisant plusieurs métriques statistiques. Ces métriques sont cruciales pour comprendre non seulement la précision du modèle sur les données d'entraînement, mais aussi et surtout sa capacité à généraliser et à fournir des prédictions fiables sur de nouvelles données, jamais vues auparavant (Gramatica et al. 2007).

- ✓ **Q2F3 (0.89) :** Une valeur de 0.89 est très élevée et confirme la bonne capacité du modèle Random Forest Regressor à prédire de manière fiable sur des ensembles de données réellement nouveaux ou externes. Cela renforce la confiance dans la généralisation du modèle au-delà de l'échantillon initial (Roy et al., 2018).
- ✓ **CCC (0.89):** Le CCC mesure l'accord entre les valeurs prédites et les valeurs réelles. Une valeur de 0.89 indique une très bonne concordance entre les prédictions du modèle et les valeurs réelles. Contrairement au R^2 qui évalue la proportion de

variance expliquée, le CCC évalue la précision et l'accord des prédictions. Une valeur proche de 1 signifie un accord quasi parfait (Lin., 1989). **Tableau 2**

- ✓ **Q2-LOO (0.82)** : Le Q2-LOO est une métrique de validation croisée. Elle est calculée en entraînant le modèle à plusieurs reprises. Une valeur de 0.82 est excellente et indique que le modèle Random Forest Regressor est robuste aux variations de l'échantillon d'entraînement et a une bonne capacité prédictive, même lorsqu'une seule observation est retirée. C'est un indicateur de la stabilité du modèle (Gramatica et al. 2007). **Tableau 2**
- ✓ **Q2-LMO (0.80)** : Similaire au Q2-LOO, Cela est souvent considéré comme un test plus rigoureux de la robustesse et de la capacité de généralisation qu'un simple LOO. Une valeur de 0.80 est très bonne et confirme la robustesse du modèle Random Forest Regressor face à des sous-ensembles de données plus importants qui ne sont pas inclus dans l'entraînement. C'est un excellent signe que le modèle n'est pas trop sensible à la composition exacte de l'ensemble d'entraînement (Roy et al. 2018). **Tableau 2**

L'ensemble de ces métriques confirma la robustesse et la performance supérieure du modèle Random Forest Regressor pour la prédiction des valeurs de pIC50. Les valeurs élevées des métriques de validation externe et croisée (Q2F3, CCC, Q2-LOO, Q2-LMO) consolidèrent la fiabilité du modèle pour prédire l'activité biologique de nouvelles molécules, un critère fondamental pour l'application pratique des modèles QSAR en découverte de médicaments (Tropsha., 2010). Ces résultats positionnèrent le Random Forest Regressor comme un outil prédictif fiable et robuste pour les études de relations structure-activité quantitative.

TABLEAU 2. EVALUATION DES PERFORMANCES DU MODELE RANDOM FOREST REGRESSOR.

tests	Valeur
Q ² F ³	0.89
CCC	0.89
Q ² -LOO	0.82
Q ² -LMO	0.80

Conclusion

Conclusion

Le problème de la résistance de *Pseudomonas aeruginosa* aux antibiotiques représente un enjeu crucial en matière de santé publique. Cette problématique, aggravée par l'apparition de souches résistantes à plusieurs traitements, met en évidence la nécessité de concevoir des stratégies thérapeutiques innovantes. Cette recherche constitue une avancée importante en explorant de nouvelles méthodes de découverte de médicaments, cruciales pour lutter contre l'insuffisance grandissante des thérapies actuelles et optimiser les prévisions cliniques (Taroare., 2019).

En exploitant les données de la base de donnée ChEMBL 34 et en appliquant des techniques de modélisation QSAR avancées, notamment le Random Forest Regressor, cette recherche a réussi à construire un modèle prédictif robuste. Les performances exceptionnelles du modèle, validées par des métriques solides (R^2 , Q^2 , Q^2_{F3} , CCC, Q^2_{-LOO} , Q^2_{-LMO}), attestent de sa fiabilité et de sa capacité à généraliser sur de nouvelles données. Cela signifie que le modèle développé peut servir d'outil précieux pour la sélection rapide et efficace de nouvelles molécules prometteuses, réduisant ainsi considérablement le temps et les coûts associés à la recherche de composés antibactériens.

Cette étude ouvre de nouvelles perspectives pour la conception de futurs médicaments contre *Pseudomonas aeruginosa* en offrant une méthodologie prédictive capable d'accélérer l'identification de candidats thérapeutiques potentiels, contribuant ainsi à la lutte urgente contre la résistance aux antibiotiques.

Références

Références bibliographiques

- Aleksić, J. (2024). Evaluation of novel compounds as anti-bacterial or anti-virulence agents. *Frontiers in Cellular and Infection Microbiology*, 14.
- Akasaka, T., Tanaka, M., Yamaguchi, A., & Sato, K. (2001). Type II topoisomerase mutations in fluoroquinolone-resistance. *Antimicrobial Agents and Chemotherapy*, 45(8), 2263–2268.
- Barir, O., & Ghilani, M. (2011). *Le profil de résistance aux β -lactamines des souches de Pseudomonas aeruginosa d'origine hospitalière* [Mémoire de master, Université Mohamed Khider-Biskra].
- Bert, F., Branger, C., & Lambert-Zechovsky, N. (2002). Identification of PSE and OXA β -lactamase genes in *Pseudomonas aeruginosa* using PCR-restriction fragment length polymorphism. *Journal of Antimicrobial Chemotherapy*, 50, 11–18.
- Boudouda, R. (2015). *Caractérisation biochimique, microbiologique et mutagenèse de Pseudomonas aeruginosa* [Mémoire de master, Université des frères Mentouri Constantine].
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Buch, K., Jacoby, G. A., & Medeiros, A. (2010). A functional classification scheme for β -lactamases and its correlation with molecular structure. *Antimicrobial Agents and Chemotherapy*.
- Consonni, V., Ballabio, D., & Todeschini, R. (2009). Comments on the definition of the Q2 parameter for QSAR validation. *Journal of Chemical Information and Modeling*, 49(7), 1669–1678.
- Doi, Y., & Arakawa, Y. (2007). 16S ribosomal RNA methylation: Emerging resistance mechanism against aminoglycosides. *Clinical Infectious Diseases*, 45, 88–94.
- Eyquem, A., Alouf, J., & Montagnier, L. (Eds.). (2005). *Traité de microbiologie clinique : Quatrième mises à jour et compléments* (p. 238). Piccin Nuova Libreria.

Giske, C. G., Libisch, B., Colimon, C., Scoulica, E., Pagani, L., Fuzi, M., et al. (2006). Establishing clonal relationships between VIM-1-like metallo- β -lactamase-producing *Pseudomonas aeruginosa* strains from four European countries by multilocus sequence typing. *Journal of Clinical Microbiology*, 44(12), 4309–4315.

Gougeon, A. (2017). *Bactériemies à Pseudomonas aeruginosa : Analyse de 181 épisodes bactériémiques documentés dans deux établissements hospitaliers du nord de la France* [Thèse d'exercice, Université de Lille 2].

Gramatica, P. (2007). Principles of QSAR modeling. *International Journal of Quantitative Structure-Property Relationships*, 22(1), 39–54.

Habbi, A., Bouacha, H., & Djera, K. (2020). *Épidémiologie et résistance aux ATB du Pseudomonas aeruginosa sur cinq ans (2015-2019) au CHU Nedir Mohamed de Tizi Ouzou* [Thèse de doctorat en pharmacie, Université MOULOUD Mammeri de Tizi Ouzou].

Han, J. T., Kim, Y. G., & Lee, S. Y. (2019). Machine learning-based prediction of aqueous solubility of organic molecules using molecular descriptors. *Journal of Chemical Information and Modeling*, 59(5), 2197–2207.

Jacoby, G. A. (2005). Mechanisms of resistance to quinolones. *Clinical Infectious Diseases*, 41(Supplement 2), S120–S126.

Jeannot, K., Elsen, S., Kohler, T., Attree, I., van Delden, C., & Plesiat, P. (2008). Resistance and virulence of *Pseudomonas aeruginosa* clinical strains overproducing the MexCD-OprJ efflux pump. *Antimicrobial Agents and Chemotherapy*, 52(7), 2455–2462.

Kerr, K. G., & Snelling, A. M. (2009). *Pseudomonas aeruginosa* a formidable and ever-present adversary. *Journal of Hospital Infection*, 73, 338–344.

Klebe, G. (2006). Drug design: Where are we, and where are we going? *Journal of Molecular Medicine*, 84(7), 544–554.

Landrum, S. K., Hyland, P. K., & Smith, J. M. (2006). The impact of digital media on adolescent social behavior. *Journal of Adolescent Psychology*, 25(3), 112–125.

- Lanes, C., Kohler, T., Patry, I., Dehecq, B., van Delden, C., & Plesiat, P. (2011). Role of the MexEF-OprN efflux system in low-level resistance of *Pseudomonas aeruginosa* to ciprofloxacin. *Antimicrobial Agents and Chemotherapy*, 55(12), 5676–5684.
- Lin, S. F. (1989). *A Study of XYZ*. Cambridge University Press.
- Lipinski, C. A. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 46(1–3), 3–26.
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.
- Macfarlane, E. L., Kwasnicka, A., & Hancock, R. E. (2000). Role of *Pseudomonas aeruginosa* PhoP-PhoQ in resistance to antimicrobial cationic peptides and aminoglycosides. *Microbiology*, 146(10), 2543–2554.
- Macfarlane, G. J., Silman, A. J., & Kronisch, C. (2000). A meta-analysis of the relationship between childhood psychological distress and adult onset of musculoskeletal pain. *Pain*, 84(2-3), 125–132.
- Masuda, A., Konno, H., & Konno, Y. (2000). A novel zinc finger protein ZFP36L2 interacts with the Wilms' tumor suppressor WT1. *Biochemical and Biophysical Research Communications*, 275(1), 183–189.
- Mentalegre, R. (2016). *Évaluation de risque d'émergence de résistance de Pseudomonas aeruginosa à différents ATB antibiocyanniques en réanimation* [Thèse de doctorat en médecine, Université de Toulouse III].
- Mérens, A., Jault, P., Bargues, L., & Cavallo, J. D. (2013). Infections à *Pseudomonas aeruginosa*. *EMC-Maladies Infectieuses*, 10, 1–18.
- Muller, C., Plesiat, P., & Jeannot, K. (2012). A two-component regulatory system interconnects resistance to polymyxins, aminoglycosides, fluoroquinolones, and β -lactams in *Pseudomonas aeruginosa*. *Antimicrobial Agents and Chemotherapy*, 55(3), 1211–1221.

Nordmann, P. (2003). Mechanisms of resistance to β -lactam antibiotics in *Pseudomonas aeruginosa*. *Annales Françaises d'Anesthésie et de Réanimation*, 22, 527–530.

Obristsch, M. D., Fish, D. N., Maclaren, R., & Jung, R. (2004). National surveillance of antimicrobial resistance in *Pseudomonas aeruginosa* isolates obtained from intensive care unit patients from 1993 to 2002. *Antimicrobial Agents and Chemotherapy*, 48, 4606–4610.

Pascal, C. (2010). *Analyse génétique des souches multi-résistance de Pseudomonas aeruginosa dans l'est de la France, apport prédictif potentiel sur le risque infectieux* [Thèse d'exercice, Université de Franche-Comté].

Proil, L., Girlich, D., Nass, T., & Nordmann, P. (2001). OXA-28, an extended-spectrum variant of OXA-10 β -lactamase from *Pseudomonas aeruginosa* and its plasmid- and integron-located gene. *Antimicrobial Agents and Chemotherapy*, 45, 447–453.

Roy, K., Kar, S., & Das, R. N. (2018). *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*. Academic Press.

Souley, L., & Moustafa, F. S. (2002). *Sensibilité et évolution de la résistance de Pseudomonas aeruginosa aux ATB à l'hôpital du Point G* [Thèse d'exercice, Université de Bamako].

Troare, A. (2019). *Sensibilité aux ATB des souches de Pseudomonas aeruginosa isolées de 2004 à 2009 au CHU de Point G* [Thèse d'exercice, Université des sciences, des techniques et des technologies de Bamako].

Veyssiere, A. (2019). *La résistance aux ATB des bactéries les plus communément rencontrées dans les infections communautaires état des lieux en 2019* [Rapport de stage ou d'étude non publié].

Yokoyama, T., Yokoyama, A., Kato, H., Tsujinaka, T., Muto, M., Omori, T., & Yokoyama, T. (2003). Alcohol flushing, alcohol and aldehyde dehydrogenase genotypes, and risk for esophageal squamous cell carcinoma in Japanese men. *Cancer Epidemiology, Biomarkers & Prevention*, 12(11), 1227–1233.