

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique Université 8 Mai
1945 Guelma
Faculté de mathématique informatique et sciences de la matière



Département des sciences de la matière
Laboratoire de chimie computationnelle et nanoparticules
THÈSE

**EN VUE DE L'OBTENTION DU DIPLOME DE DOCTORAT EN 3^{ème}
CYCLE**

Domaine : Sciences de la matière Filière : Chimie
Spécialité : Chimie théorique et computationnelle

Présentée par

BENCHEIKH Bochra

Intitulée

**Criblage in silico des inhibiteurs du SARS-CoV-2 par
Docking moléculaire et QSAR**

Devant le Jury composé de :

	Nom et Prénom	Grade	
Présidente :	Mme NOUAR Leila	Prof	Univ. De Guelma
Directrice de thèse :	Mme DJEMIL Rayenne	Prof	Univ. De Guelma
Examineurs :	Mr MERDES Rachid	Prof	Univ. De Guelma
	Mme MADI Fatiha	Prof	Univ. De Guelma
	Mme BERRDJEM Yamina	Prof	Univ. D'Annaba
	Mme ATTOUI- YAHIA Ouassila	Prof	Univ. D'Annaba

Année Universitaire :2022/2023

Remerciements

Tout d'abord, Je tiens à remercier très sincèrement ma directrice de thèse Madame **DJEMIL Rayenne**, Professeur à l'Université 8mai 1945-GUELMA, pour avoir dirigé mes travaux et pour toute l'aide qu'elle m'a apportée au cours de ce travail. Son intérêt, sa disponibilité, son suivi quotidien de mes travaux et ses précieux conseils m'ont été d'un grand profit. Je la remercie infiniment.

Ce travail de recherche qui fait l'objet de cette thèse a été réalisé au sein du laboratoire de Chimie computationnelle et nanoparticule (LCCN), Université 8mai 1945-GUELMA.

Nous remercions sincèrement : *Pr NIGRI Soraya* , *Dr DJADI Nabila*, *Dr CHERIET Mouna*, *Dr KRAIM Khairedine* pour leurs aides.

Mes meilleurs remerciements s'adressent à **Mme NOUAR Leila**, pour m'avoir fait l'honneur de présider le jury de cette thèse.

Je remercie vivement **Mr MERDES Rachid** qui m'a fait l'honneur d'examiner ce travail et de faire partie du jury de cette thèse.

J'exprime ma gratitude à **Mme MADI Fatiha**, pour avoir accepté de juger ce travail. Je suis honoré de la compter parmi les membres de ce jury.

Je remercie tout particulièrement **Mme BERRDJEM Yamina** et **Mme ATTOUI- YAHIA Ouassila** pour avoir accepté de faire partie du jury de cette thèse.

Merci à tous.

Dédicace

Je dédie cette thèse ...□



A mon Dieu, le tout puissant ! Au seigneur le tout généreux !

A ma chère mère

*Tes prières ont été pour moi d'un grand soutien au cours de ce long parcours.
Tu m'as toujours donné de ton temps, de ton énergie, de la liberté, de ton
cœur et de ton amour.*

*En ce jour j'espère réaliser chère mère un de tes rêves (doctorat,,,,) ,
sachant que tout ce que je pourrais faire ou dire ne pourrait égaler ce que tu
m'as donné et fait pour moi.*

*Qu'Allah, te garde, te procure santé, bonheur et longue vie pour que tu
demeures le flambeau illuminant mon chemin...*

A mon cher père

*A celui qui m'a tout donné sans compter, à celui qui m'a soutenu toute ma
vie, à celui à qui je dois ce que je suis et ce que je serai.*

*Puisse Allah, te protéger de tout mal, te combler de santé, de bonheur et
t'accorder une longue et heureuse vie, afin que je puisse te rendre un
minimum de ce que je te dois.*

*A mon cher mari,
Je veux juste vous remercier pour votre aide précieuse, Je vous souhaite une
vie commune pleine de bonheur et de réussite.*

*A mon rayon de soleil fille NAYA,
T'es la bénédiction de ma vie. Dieu m'a béni avec une fille merveilleuse qui ne
cesse de m'étonner. Sois toujours forte et courageuse ...*

*A Mon fils mon bonheur MOHAMED TAMIME,
Tu étais et es toujours, un de mes souhaits les plus chers. Tu es mon
essentiel, Je te souhaite d'être heureux le plus souvent possible, Je te
souhaite de t'épanouir autant dans la vie.*

*A mon cher frère ABDOU,
Je te souhaite beaucoup de bonheur et de santé, de réussite.*

*A ma chère sœur RAYANE,
Je t'ai toujours considéré comme une seconde mère. Merci pour ta serviabilité.
Je te souhaite une vie pleine de bonheur et de réussite.*

*A ma petite sœur DOUAA,
Tu es brillante. Tu ne mérites que le meilleur. Je te souhaite une vie pleine d
bonheur et de réussite.*

*À mes collègues de la promotion de chimie théorique et computationnelle, À
ma famille.*

Je vous dis : Merci Beaucoup

المخلص:

فيروس SARS-CoV-2 هو العامل المسبب لـ COVID-19 والذي يمثل تهديدًا كبيرًا للبشرية. إذ لا يزال اكتشاف نظام دوائي فعال ضد هذه العدوى قيد التطوير من أجل المساهمة في مكافحة COVID-19 والذي يمثل حتى الآن جدلا واسعا في الوسط العلمي، درسنا تثبيط SARS-CoV-2 من خلال الأساليب النظرية للفحص الافتراضي.

تم فحص سلسلة من المركبات والأدوية الطبيعية باستخدام الالتحام الجزيئي. في أنظمة المستقبلات التي تمت دراستها، فإن مستقبلات SARS-CoV-2 المستخدمة كأهداف علاجية هي 6LU7 (M^{Pro} protease) والإنزيم المحول للأنجيوتنسين 2 (ACE2) (6LZG). في حين أن الروابط المستخدمة كمثبطات هي سلسلة من مركبات الفلافونويد والكومارين مع بعض الأدوية، مثل CQ و HCQ المستخدمة كمراجع.

تظهر نتائج الإرساء أن القوة التثبيطية للمركبات الطبيعية التي تمت دراستها باستخدام M^{Pro} أكبر منها مع ACE2. غالبية الروابط الطبيعية لديها تقارب أفضل مع M^{Pro} ، مقارنة بالأدوية المستخدمة ضد SARS-CoV-2. تتوافق الخصائص البيولوجية والدوائية للروابط التي تحدها اختبارات ADMET مع نتائج الإرساء والديناميكيات الجزيئية.

سمح لنا منهج QSAR (العلاقة الكمية بين البنية والنشاط) بالحصول على نموذج موثوق يشرح ويتنبأ بالنشاط المضاد للفيروسات المعبر عنه بتركيز التثبيط المتوسط (IC₅₀) لسلسلة من مشتقات amino quinazolin-4 (3H)-one-2 من مثبطات SARS-CoV-2، من أجل التنبؤ بالنشاط البيولوجي لمثبطات جديدة لهذا الفيروس.

الكلمات المفتاحية: COVID-19، SARS-CoV-2، الالتحام الجزيئي، QSAR، ADMET، الديناميكية الجزيئية.

Résumé

Le virus SARS-CoV-2 est l'agent causal du COVID-19 et constitue toujours une grande menace pour l'humanité. La découverte d'un régime médicamenteux efficace contre cette infection est encore en développement. Dans le but de contribuer à la lutte contre le COVID-19 qui représente jusqu'à ce jour une polémique, nous avons étudié l'inhibition du SARS-CoV-2 par des approches théoriques du criblage virtuel.

Une série de composés naturels et de médicaments ont été criblés utilisant le Docking moléculaire. Dans les systèmes récepteurs-ligands étudiés, les deux récepteurs du SARS-CoV-2 utilisés comme cibles thérapeutiques sont la protéase M^{Pro} (6LU7) et l'enzyme de conversion de l'angiotensine 2 ACE2 (6LZG). Tandis que les ligands utilisés comme inhibiteurs sont une série de flavonoïdes et des coumarines avec quelques médicaments, tels que la CQ et l'HCQ utilisés comme références.

Les résultats de Docking montrent que le pouvoir inhibiteur des composés naturels étudiés avec le M^{Pro} est plus important qu'avec l'ACE2. La majorité des ligands naturels ont une meilleure affinité pour M^{Pro}, par rapport aux médicaments utilisés contre le SARS-CoV-2. Les propriétés biologiques et pharmacologiques des ligands déterminées par les tests ADMET sont en accord avec les résultats du Docking et de dynamique moléculaire.

L'approche QSAR (Relation Quantitative Structure-Activité) nous a permis d'avoir un modèle fiable qui explique et prédit l'activité antivirale exprimée par la concentration d'inhibition médiane (IC₅₀) d'une série de dérivés de 2-amino quinazolin-4(3H)-one inhibiteurs du SARS-CoV-2, afin de prédire l'activité biologique de nouveaux inhibiteurs de ce virus.

Mot clés : COVID-19, SARS-CoV-2, Docking moléculaire, QSAR, ADMET, dynamique moléculaire .

Abstract

The SARS-CoV-2 virus is the causative agent of COVID-19 and still poses a great threat to humanity. The discovery of an effective drug regimen against this infection is still in development. In order to contribute to the fight against COVID-19, which to date represents a controversy; we studied the inhibition of SARS-CoV-2 by theoretical approaches of virtual screening.

A series of natural compounds and drugs have been screened using molecular Docking. In the receptor-ligand systems studied, the two SARS-CoV-2 receptors used as therapeutic targets are the M^{Pro} protease (6LU7) and the angiotensin 2 converting enzyme ACE2 (6LZG). While the ligands used as inhibitors are a series of flavonoids and coumarins with some drugs, such as CQ and HCQ used as references.

Docking's results show that the inhibitory power of the natural compounds studied with MPro is greater than with ACE2. The majority of natural ligands have a better affinity for MPro, compared to drugs used against SARS-CoV-2. The biological and pharmacological properties of the ligands determined by the ADMET tests are in agreement with the results of Docking and molecular dynamics.

The QSAR (Quantitative Structure-Activity Relationship) approach has allowed us to have a reliable model that explains and predicts the antiviral activity expressed by the median inhibition concentration (IC₅₀) of a series of 2-amino quinazolin derivatives -4(3H)-one inhibitors of SARS-CoV-2, in order to predict the biological activity of new inhibitors of this virus.

Key words: COVID-19 , SARS-CoV-2, Docking molecular, QSAR, ADMET, molecular dynamics.



LISTE DES TABLEAUX



Liste des tableaux

N°	Titre	Page
CHAPITRE 01		
Tableau 1-1	Bases de données des petites molécules les plus connues et les plus importantes.	7
Tableau 1-2	Valeurs utilisées dans les différents filtres permettant la sélection de molécules ou fragments aux propriétés pharmacocinétiques « drug-like » ou « lead-like ».	10
Tableau 1-3	Classification des méthodes QSAR selon la dimension des descripteurs utilisés.	14
CHAPITRE 02		
Tableau 2-1	Comparaison épidémiologique entre différents types de coronavirus.	34
CHAPITRE 03		
Tableau 3-1	la structure 2D des inhibiteurs.	61
Tableau 3-2	La boîte GRID BOX des récepteurs.	69
Tableau 3-3	L'activité anti-SARS-CoV-2 des dérivés de 2-amino quinazolin-4(3H)-one.	72
CHAPITRE 04		
Tableau 4-1	RMSD des deux complexes protéine-ligand.	77
Tableau 4-2	La boîte GRID BOX des protéines.	77
Tableau 4-3	Résultats du docking 6LZG-Ligand.	78
Tableau 4-4	Résultats du Docking 6LU7-Ligand (tous les inhibiteurs).	81
Tableau 4-5	Résultats du Docking 6LU7-Ligand.	85
Tableau 4-6	Résultats des propriétés de Druglikeness des ligands.	88
Tableau 4-7	Résultats du test ADME/T des ligands (diverses propriétés pharmacocinétiques et pharmacodynamiques).	90
Tableau 4-8	Paramètres statistiques RLM de 2 descripteurs.	95
Tableau 4-9	Résultats de l'analyse de la variance.	96

Tableau 4-10	Définition des descripteurs.	97
Tableau 4-11	Paramètres de validation croisée LOO.	98
Tableau 4-12	Paramètres statistiques de test.	98



LISTE DES FIGURES



Liste des figures

N°	Titre	Page
CHAPITRE 01		
Figure 1.1	Classification des méthodes de criblage virtuel.	5
Figure 1.2	Principales étapes de l'élucidation d'un pharmacophore.	11
Figure 1.3	Illustration du processus de conception de modèles QSAR.	12
Figure 1.4	Méthodes pour obtenir la structure 3D d'une cible biologique d'intérêt.	20
Figure 1.5	Représentation schématique du Docking moléculaire.	22
CHAPITRE 02		
Figure 2.1	Classification et taxonomie des coronavirus humains (HCoV) Classification des HCoV selon l'ICTV en 2017.	33
Figure 2.2	Origines animales des coronavirus humains.	35
Figure 2.3	Structure schématisée du SARS-CoV-2.	37
Figure 2.4	Le domaine RBD (Receptor Binding Domain).	37
Figure 2.5	(a) Structure du domaine de liaison au récepteur (RBD) de la protéine S du SARS-CoV-2 (rubans bleu) complexé avec l'ACE2 humaine (rubans rouge).	38
Figure 2.6	Présentation schématique de l'organisation du génome du SRAS-CoV-2, des ARNm sous génomiques canoniques et de la structure du virion.	39
Figure 2.7	Cycle de réplication du SARS-CoV-2 dans les cellules hôtes.	40
Figure 2.8	Structures chimiques de la chloroquine et de l'hydroxychloroquine.	44
Figure 2.9	Structures de Lopinavir et Ritonavir.	45
Figure 2.10	Structures de Remdesivir.	46
Figure 2.11	Structures de Favipiravir.	47

Figure 2.12	Structures de Umifenovir.	47
Figure 2.13	Structures de L'azithromycine.	48
Figure 2.14	Structure de Vitamine D.	50
Figure 2.15	Structure de ZINC.	51
CHAPITRE 03		
Figure 3.1	Stratégie globale d'une étude QSAR.	71
CHAPITRE 04		
Figure 4.1	Comparaison de la conformation expérimentale du ligand avec sa conformation optimale calculée par Docking moléculaire avec AutoDock Vina.	77
Figure 4.2	Mode d'interaction de 6LU7-bergamotine 2D, 3D.	87
Figure 4.3	Mode d'interaction de 6LU7-naringénine 2D, 3D.	87
Figure 4.4	Mode d'interaction de 6LU7-hydroxychloroquine 2D, 3D.	87
Figure 4.5	Mode d'interaction de 6LU7-Chloroquine 2D, 3D.	88
Figure 4.6	Comparaison et représentation détaillée de (A) RMSD : écarts quadratiques moyens, (B) RMSF : fluctuations quadratiques moyennes, (C) Rg : rayon de giration et (D) SASA : surface accessible au solvant, sur le temps de 50 ns de référence MPro et tous complexes.	93
Figure 4.7	Valeurs résiduelles ajustées en fonction de la valeur ajustée.	96
Figure 4.8	Variation des valeurs expérimentales en fonction des valeurs prédites.	99



LISTE DES ABBREVIATIONS



Liste des abbreviations:

Symbole	Nomenclature
1D	Une Dimension.
2D	Deux Dimensions.
3D	Trois Dimensions.
3CLpro	3C-Like proteinase.
A	
ACE	Angiotensin Converting Enzyme.
ACE 2	Angiotensin Converting Enzyme 2.
ARN	Acide RiboNucléique.
ADME/ Tox.	Absorption, Distribution, Métabolisme, Excrétion/ Toxicologie.
.AM1	Austin Model 1.
AMBER	Assisted Model Building with Energy Refinement.
B	
B3LYP	Becke 3-parameter, Lee- Yang-Parr.
BBB :	Blood-Brain Barrie.
C	
COVID-19	COronaVirus Disease 2019.
CoV	CoronaVirus.
α -CoV	Alpha CoronaVirus.
β -CoV	Beta CoronaVirus.
γ -CoV	Gamma CoronaVirus.
CQ	Chloroquine.
CHARMM	Chemistry at HARvard Macromolecular Mechanics.
CNS	Central Nervons System.
D	

DFT	Théorie de la Fonctionnelle de Densité.
F	
FDA	Food and Drug Administration.
G	
GCG	GalloCatéchine Gallate.
H	
HSAB	Hard and Soft Acides and Bases.
HCQ	HydroxyChloroQuine.
HCoV	Coronavirus Humain.
HIV	Human Immunodeficiency Virus.
HOMO	Highest Occupied Molecular Orbital.
HIV	Human immunodeficiency virus.
HE	Hémagglutinine –esterase.
HIA	Human Intestinal Absorption.
HERG	Human Ether à go-go- Related Gene.
I	
IC50	Concentration Inhibitrice médiane.
L	
LOO	Leave-one-out.
LTO	Leave-Two-out.
LMO	Leave-Many-out.
LUMO	Lowest Unoccupied Molecular Orbital.
M	
MERS-CoV	CoronaVirus du Syndrome Respiratoire du Moyen-Orient.
MoRSE	Molecule Representation of Structures based on Electron diffraction.
M ^{pro}	Main protease.
N	

NS3	Non-Structural Proteins 3.
O	
OMS	Organisation Mondiale de la Santé.
ORF	Open Reading Frame.
ORF1a / b	Open Reading Frame a/b.
P	
PIC50	Cologarithme décimal de la Concentration inhibitrice médiane.
PLpro	Papain-like protease.
PLS	Régression des moindres carrés partiels.
Q	
QSAR	Relation Quantitative Structure-Activité.
R	
RMN	Résonance Magnétique Nucléaire.
RdRp	RNA-dependent RNA Polymerase.
RBD	Receptor Binding Domain.
RLM	Régression Linéaire Multiple.
RNLM	Régression Non Linéaire Multiple.
S	
SARS	Severe Acute Respiratory Syndrome related CoronaVirus.
V	
VDW	Van Der Waals.
W	
WHIM	Weighted Holistic Invariant Molecular descriptors.

Les acides aminés

- GLU: Acide Glutamique.
- ASP: Acide Aspartique.
- ALA: Alanine.
- ASN: Asparagine.
- CYS: Cystéine.
- GLN: Glutamine.
- GLY: Glycine.
- HIS: Histidine.
- LEU: Leucine.
- LYS: Lysine.
- MET: Methionine.
- PRO: Proline.
- PHE: Phenylalanine.
- SER: Serine.
- THR: Threonine.
- TYR: Tyrosine.
- VAL: Valine.
- ARG: Arginine.

Sommaire :

Titre	Page
- Remerciements	
- Dédicace	
- ملخص	
- Résumé	
- Abstract	
- Liste des tableaux	
- Liste des figures	
- Liste des abréviations	
- Les acides aminées	
- Sommaire	
- Introduction générale	1
Chapitre 01 : Criblage virtuel in silico	
1. Généralités sur le criblage virtuel.	5
2. Chimiothèques.	6
2.1 Différentes chimiothèques.	6
2.2 Filtrage ADME-Tox.	7
2.3 Règles de Lipinski.	7
2.4 Autres critères de sélection ADME-Tox.	8
2.5 Toxicologie.	9
3. Méthodes de criblage.	10
3.1 Criblage virtuel « ligand-based ».	10
3.1.1. Recherche de similarité.	10
3.1.2. Modèles pharmacophoriques « ligand-based ».	10
3.1.3. Quantitative Structure-Activity Relationship (QSAR) « ligand-based ».	11
A. Principe.	12
B. Stratégie globale.	12

C. Bases de données.	13
D. Descripteurs.	13
E. Méthodes statistiques.	15
F. Paramètres statistiques de la validation.	16
G. Techniques de validation.	19
3.2 Criblage virtuel « structure based»...	20
3.2.1. Modèles pharmacophoriques « structure-based ».	21
3.2.2. Quantitative Structure-Activity Relationship (QSAR) « structure-based » RD-QSAR (Receptor Dependent-Quantitative Structure-ActivityRelationship).	21
3.2.3. Approches de Novo.	21
3.2.4. Docking moléculaire.	22
A. Principe générale de Docking moléculaire.	23
B. Types de Docking moléculaires.	23
C. Algorithmes de recherche.	24
D. Le processus de scoring.	26
Chapitre 02 : COVID-19 et ses inhibiteurs	
1. Historique du covid.	33
2. Caractéristique virologique du SARS CoV- 2.	33
2.1. Classification et taxonomie.	33
2.2. Type et origine.	34
2.3. Structure de virion et son génome.	35
2.3.1. Le virion.	36
2.3.2. Le Génome.	38
3. Pathogénie du SARS CoV-2.	39
3.1. Transmissions.	39
3.2. Pénétration du virus dans la cellule hôte.	39
3.3. Cycle de réplication.	40
3.4. Les variants du SARS-CoV-2.	41
3.4.1. Signification de « variant ».	41
3.4.2. La dénomination des différents variants.	41

4. Epidémiologie.	42
4.1. Evolution de l'épidémie.	42
5. Principales thérapies et prophylaxie.	43
5.1. Les thérapies antivirales.	43
5.1.1. La chloroquine - L'hydroxychloroquine.	43
5.1.2. Inhibiteurs de la protéase du VIH.	45
5.2. Les thérapies adjuvantes.	44
5.2.1. La thérapie antibactérienne.	48
5.2.2. La vitaminothérapie.	49
Chapitre 03 : Matériel et méthode	
1. Docking moléculaire .	57
1.1. Matériels utilisés.	57
1.1.1. Programmes (logiciels) .	57
1.1.2. Banques de données (chimiothèques) .	58
1.2. Méthodes.	58
1.2.1. Docking moléculaire.	58
1.2.2. Reproduction de la structure cristalline.	59
1.2.3. Le choix du complexe.	59
1.2.4. Choix du logiciel de docking moléculaire.	60
1.2.5. Préparation des molécules au docking moléculaire.	60
1.3. Criblage Par Docking.	68
1.3.1. Protocole du Docking moléculaire.	68
1.3.2. Paramètres de la grille de calculs.	69
1.3.3. L'algorithme de recherche.	69
2. QSAR (Quantitative structure Activity Relationship) .	69
2.1. Stratégie globale d'une étude QSAR .	70
2.2. Collecte des données.	71
2.3. Traitements des données.	73
2.4. Matériels utilisés.	73
2.4.1. Programmes (logiciels) .	73

2.5. Méthodologie du QSAR.	73
2.6. Répartition des échantillons.	74
2.7. Méthodes statistique utilisées.	74
Chapitre 04 : Résultat et discussion	
1. DOCKING MOLECULAIRE.	76
1.1. RMSD.	76
1.2. La boîte GRID BOX des protéines.	77
1.3. Interactions ACE2.	78
1.4. Interactions 6LU7 M ^{Pro} .	81
1.5. Comparaison entre les résultats de Docking obtenus avec les deux récepteurs ACE2 et M ^{Pro} .	84
1.6. Inhibition du récepteur SARS-CoV-2 (M ^{Pro}) par les ligands d'origine naturelle.	84
1.7. Énergies d'interaction : affinité des ligands pour le récepteur M ^{Pro} .	85
1.8. Etude des interactions du M ^{Pro} (6LU7) .	86
1.9. Propriétés pharmacologiques des ligands.	88
1.10. Résultats ADMET (absorption, distribution, métabolisme, l'excrétion, et de toxicité) .	90
1.11. Simulation de dynamique moléculaire.	92
2. Relation Quantitative Structure-Activité (QSAR) .	94
1.1. Régression linéaire multiple.	94
1.1.1. Equation et analyse de régression.	95
1.1.2. Analyse de variance.	95
A. Analyse des résultats.	97
B. Interprétation.	97
1.1.3. Validation du modèle.	98
A. Validation Leave one out.	98
B. Vérification des critères de Tropsha.	99
Conclusion générale	102
Annexe	104



INTRODUCTION GÉNÉRALE



Introduction générale

Depuis sa première apparition en Chine en décembre 2019, la pandémie de coronavirus COVID-19 s'est propagée à travers le monde à une vitesse ahurissante. La propagation a été sans précédent en termes de vitesse et d'impact, constituant une urgence de santé publique à portée internationale, et contribuant à une perturbation socio-économique généralisée. ¹ Cette maladie infectieuse a conduit de nombreux pays à travers le monde à adopter diverses mesures: « distanciation physique, le confinement, couvre-feu...etc. ».

L'agent causal étant un nouveau Beta coronavirus, similaire au coronavirus du syndrome respiratoire aigu sévère (SRAS-CoV-1), et celui du syndrome respiratoire du Moyen Orient (MERS-CoV). Le génome du SRAS-CoV-2 a été rapidement séquencé afin de permettre le diagnostic, le suivi épidémiologique et le développement de thérapeutique préventive ou curative.

Comprendre le mécanisme de ce virus pourrait fournir plus d'informations sur la transmission virale et révéler des cibles thérapeutiques. Commenant par la porte d'entrée de virus, l'enzyme de conversion de l'angiotensine-2 (ACE2), un récepteur fonctionnel du SARS-CoV-2 qui peut faciliter l'entrée du virus dans les cellules de l'hôte en se liant à la protéine de pointe (S) du virus. Par conséquent, bloquer cette liaison, semble être une piste prometteuse pour traiter une infection virale, comme le SARS-CoV-2. Une autre possibilité d'inhiber ce virus est d'empêcher la protéase virale principale (M^{Pro}) responsable de sa réplication. Beaucoup de composés qui empêchent ces enzymes ont été décrits et avancent par la voie de développement de médicament.

Cependant, la découverte des nouveaux médicaments est un processus extrêmement long et fastidieux qui dure en moyenne une dizaine, voire une quinzaine d'années en dépensant près d'un milliard de dollar. La méthode pour concevoir un nouveau médicament basé sur le développement de nouvelles approches telles que le criblage virtuel par Docking moléculaire était nécessaire².

Afin de contribuer à la recherche des nouveaux inhibiteurs antiviraux efficaces pour prévenir et traiter l'infection du SARS-CoV-2, nous avons fait appel aux approches par criblage virtuel in silico. Ces méthodes consistent à prédire l'affinité d'un ensemble de molécules pour le site actif ciblé, ce qui peut accélérer le processus de découverte de médicaments et réduire le cout et les ressources nécessaires. Par conséquent, ils ont devenu un outil précieux dans la recherche

de produits biologiquement actifs. En effet, les industries pharmaceutiques consacrent aujourd'hui près de 10% de leur budget à la recherche de médicaments assistée par ordinateur. Pour cette raison, plusieurs programmes de criblage virtuel ont été développés tels que : Gold³, Glide⁴, FlexX^{5,6} et AutoDock Vina⁷.

Le criblage virtuel a été largement utilisé dans la recherche des médicaments pour la maladie COVID-19. Il est utilisé pour identifier des molécules potentielles qui peuvent inhiber la réplication du SARS-CoV-2 et pour évaluer leur potentiel d'effets secondaires.

Dans cet optique, la première partie de notre travail est une contribution à la recherche des nouveaux inhibiteurs du SARS-CoV-2 par criblage virtuel. Nous avons donc étudié l'inhibition du SARS-CoV-2 par une série des molécules d'origine naturelles et des médicaments via le Docking moléculaire au moyen du logiciel AutoDock Vina. Dont l'objectif est, dans un premier temps, d'identifier le point d'intervention de ces options thérapeutiques, la stratégie suivie consiste à inhiber les récepteurs du SARS-CoV-2 se trouvant à l'entrée du virus (ACE2) et la protéase (M^{Pro}) responsable de la réplication. Ensuite, étudier les interactions mis en jeu dans l'inhibition du SARS-CoV-2 par diverses molécules tout en évaluant leur énergie d'interaction. Ce nous permettra de déterminer l'inhibiteur le plus puissant envers du SARS-CoV-2. Enfin, l'application de la règle de Lipinski et des paramètres ADME/TOX, nous permettra de nous renseigner sur les propriétés pharmacocinétiques des composés proposés. Dans un deuxième temps, nous avons jugé utile de réaliser une simulation de dynamique moléculaire (DM) afin d'évaluer les micro- interactions entre la structure du ligand et celle de la protéine.

Dans la deuxième partie, nous avons fait appel à la méthode QSAR (Quantitative structure Activity Relationship), qui permettent d'établir des relations entre les structures des composés et leurs activités utilisant des méthodes d'analyses multidimensionnelles par apprentissage statistique afin de modéliser l'activité biologique⁸. Notre objectif vise à utiliser la méthodologie QSAR afin d'élaborer un modèle robuste et fiable qui explique et prédit l'activité antivirale exprimée par la concentration d'inhibition médiane (IC₅₀) d'une série de dérivés de 2-Amino quinazolin-4(3H)-one inhibiteurs du SARS-CoV-2.

L'approche QSAR la plus populaire est basée sur la régression linéaire multiple (MLR), permettant de prédire l'activité biologique des nouveaux composés chimiques qui entrent dans le domaine d'applicabilité du modèle développé.

Notre thèse est présentée sous forme de deux parties. Une revue bibliographique sera rapportée dans deux chapitres. Nous abordons dans le premier chapitre le principe du criblage virtuel et ses différentes méthodes basées sur le ligand ou sur la structure, tout en détaillant les deux méthodes utilisées dans la réalisation de ce travail, le Docking moléculaire et la Relation Quantitative Structure- Activité dite QSAR. Dans le deuxième chapitre, nous présenterons des rappels généraux sur le SARS-CoV-2 et ses inhibiteurs.

Dans la partie pratique, nous décrirons le matériel utilisé pour l'obtention des structures des cibles étudiées et celles des ligands nécessaires au processus du Docking moléculaire effectué par le logiciel Auto Dock Vina tout en détaillant le protocole de Docking ainsi que la description des différentes étapes de l'approche QSAR réalisée par plusieurs programmes en l'occurrence le Molegro Data Molleder (MDM). Les résultats obtenus de chaque méthode ont été présenté et discutés.

Enfin une conclusion générale résume l'ensemble du travail réalisé et présente les perspectives d'études que nous envisageons pour l'avenir.

Références :

1. Coronaviridae Study Group of the International Committee on Taxonomy of V. **The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2.** *Nat Microbiol.* (2020) DOI : 10.1038/s41564-020-0695-z.
2. Criblage virtuel par docking moléculaire bioinfopharma. ustrasbg.fr/labwebsite/publications/paper99.pdf. Consulté le (20-05.2015).
3. Jones G., Willett P., Glen R. C. **Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation,**(1995). *J Mol Biol*, Jan, 1995 ;245(1) :43–53.
4. Friesner Richard A., Banks Jay L., Murphy Robert B., Halgren Thomas A., Klicic Jasna J., Mainz Daniel T., Repasky Matthew P., Knoll Eric H., Shelley Mee, Perry Jason K., Shaw David E., Francis Perry, Shenkin Peter S. **Glide : A new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy.** *Journal of Medicinal Chemistry* (2004) ;47(7) :1739–1749. PMID : 15027865.
5. Kramer B, Rarey M, Lengauer T. **Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking.** *Proteins : Structure, Function, and Bioinformatics* (1999) ;37(2) :228–241.
6. Rarey M, Wefing S, Lengauer T. **Placement of medium-sized molecular fragments into active sites of proteins.** *Journal of Computer-Aided Molecular Design* (1996) ;10(1) :41–54
7. Trott O, Olson A J. **Autodock vina : Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading.** *Journal of Computational Chemistry* (2010) ;31(2) :455–461.
8. M. Karelson, S.Victor Lobanov, R. Alan Katritzky. **chemical reviews**, (1996).

PARTIE

BIBLIOGRAPHIQUE



CHAPITRE : 01
CRIBLAGE VIRTUEL IN
SILICO



1. Généralités sur le criblage virtuel

Le terme criblage regroupe un ensemble de techniques computationnelles ayant pour objectif l'exploration des bases de composés à la recherche des molécules d'intérêt thérapeutique ce qui est d'une importance croissante dans la conception des médicaments, principalement basée sur les progrès en informatique et en bio-informatique ¹. Le criblage virtuel peut être assimilé de manière schématique à un entonnoir dans lequel un grand nombre (généralement entre cent mille et dix millions) des composés sont versés, constituant la chimiothèque à cribler, afin d'obtenir, au moyen d'un algorithme de criblage, un nombre plus faible (quelques dizaines à quelques milliers) des composés qui seront ensuite testés expérimentalement ².

Ce processus vise à réduire considérablement la chimiothèque de départ à une liste limitée de composés jugés les plus prometteurs. Par conséquent, le temps aussi bien que les coûts peuvent être réduits de façon significative. Les protocoles établis permettent d'éliminer les composés considérés inactifs ou indésirables en raison de leurs propriétés pharmacologiques et sélectionneront les composés présumés actifs en leur attribuant des scores d'affinité élevés à la cible ³. Initié au début des années 1980, le criblage virtuel s'est développé pour devenir, de nos jours, un outil incontournable dans la recherche de produits biologiquement actifs.



Figure 1.1: Classification des méthodes de criblage virtuel

2. Chimiothèques

2.1. Différentes chimiothèques

Pour procéder aux criblages, il faut d'abord sélectionner une chimiothèque de composés. Le succès des différents types de criblages dépend du choix de la chimiothèque utilisée, sa taille, composition, diversité, les caractéristiques des molécules qu'elle regroupe ⁴.

Les chimiothèques utilisées pour les criblages virtuels sont principalement commerciales. (Tableau 1-1).

Tableau 1-1: Bases de données des petites molécules les plus connues et les plus importantes.

Chimiothèque	License Type	Publisher
NCI Open database	Publicly available	National Cancer Institute
MDDR	Commercial	Accelrys Inc
Binding DB	Publicly available	University of Maryland USA
Relibase	Commercial	Cambridge cristalographic data center
Chem spider	Publicly available	Royal society of chemistry ,UK
DrugBank	Publicly available	University of alberia CANADA
PubChem	Publicly available	National center for biotechnology information
Therapeutic Traget Database	Publicly available	National University of Singapore, Singapore
ZINC	Publicly available	University of California san Francisco, USA
ChEMBL	Publicly available	European bioinformatics Institute, KU
World Drug Index	Commercial	Thomson Reuters

Il existe plusieurs types de chimiothèques tels que les chimiothèques focalisées dédiées à la mise en œuvre de criblages sur une famille de cibles spécifique. Plusieurs chimiothèques commerciales sont disponibles : « **ChemBridge**, (ChemBridge screening libraries 2016), **Asinex**, (ASINEX screening libraries 2016), **Life Chemicals**, (Life Chemicals screening libraries 2016), **Timtec**, (TimTec screening libraries 2016), **Maybridge** (Maybridge screening libraries.2016) ».

Ainsi, il existe des chimiothèques de produits naturels, ces produits ont généralement une sélectivité significative par rapport à leurs cibles, présentent de bonnes caractéristiques ADME-Tox ^{5 et 6}.

Après avoir sélectionné une chimiothèque et avant le criblage, les molécules doivent être préparées dans le format approprié aux méthodes utilisées, à partir des données brutes informatiques. Différents types de filtres peuvent également être utilisés pour réduire la taille de la chimiothèque et d'optimiser les résultats.

2.2. Filtrage ADME-Tox

Dans les années 90, les échecs dans la conception de médicaments étaient principalement liés aux mauvaises performances pharmacocinétiques. L'utilisation d'un filtrage précédant des chimiothèques a permis de résoudre ce problème et réduisant ainsi les taux d'échec dans les phases de développement ⁷.

Les filtres de type ADME-Tox (Absorption, Distribution, Métabolisme, Elimination et Toxicité) sont rapidement devenus populaires et largement utilisés. Ils se basent sur plusieurs critères déterminant les propriétés pharmacocinétiques potentielles des molécules. Avant tout processus de criblage et pour réduire le nombre de composés d'une chimiothèque, les plus appropriés pour devenir des candidats médicaments seront sélectionnés. L'utilisation de ces filtres a donné des très bons réduisant de taux d'échec imputable à une cause pharmacocinétique de 40% à moins de 10% en 10 ans ⁷.

Désormais, la toxicité et le manque d'efficacité des candidats médicaments sont les deux plus grandes causes d'échecs dans le développement d'un médicament ⁷.

2.3. Règles de Lipinski

Ainsi, Lipinski et al. ont défini dès 1997 des règles simples permettant d'identifier rapidement et à grande échelle des molécules à caractère « drug-like », plus susceptibles d'avoir les caractéristiques de biodisponibilité nécessaires à l'élaboration d'un candidat médicament ⁸.

Ces règles, communément appelées « règles de Lipinski » ou « la règle de 5 », comportent quatre critères physico-chimiques qui décrivent la molécule : poids moléculaire ≤ 500 Da, $\log P \leq 5$, nombre d'accepteurs de liaisons hydrogènes ≤ 10 et nombre de donneurs de liaisons hydrogènes ≤ 5 (**En ci-dessous le Tableau 1.2**). Particulièrement, la mesure du $\log P$ caractérise la polarité du

composé (estimée par le coefficient de partition octanol/eau), ce qui permet d'estimer la distribution du composé dans l'organisme. Les molécules hydrophobes (valeurs du logP élevées) sont principalement réparties dans les régions hydrophobes, comme la bicouche lipidique des cellules. En revanche, les molécules hydrophiles se trouvent principalement dans des régions aqueuses, telles que le sérum sanguin. La 5^{ème} règle de Lipinski indique que les adjuvants et les produits assimilés font exception aux quatre autres règles.

Selon ces règles, les composés ne validant pas au moins deux critères suivants qui sont très susceptibles d'avoir des problèmes d'absorption ou de perméabilité intestinale ⁸. Une mauvaise biodisponibilité orale serait également détectée en cas de violation de l'une de ces règles.

2.4. Autres critères de sélection ADME-Tox

Par la suite, d'autres critères ont été mis en place pour compléter et ajuster les règles de Lipinski dans la sélection de composés « drug-like ». Ainsi, Veber et al. choisissent d'utiliser les quatre critères suivants : nombre de liaisons rotatives, aire de la surface polaire et nombre de donneurset d'accepteurs de liaisons hydrogènes (**Tableau 1.2**) ⁹. D'autres approches ont été envisagées pour la sélection ADME-Tox, notamment basées sur les types d'atomes et de liaisons, ainsi que sur le comptage de points pharmacophoriques spécifiques ¹⁰.

Les méthodes d'apprentissage ont également été explorées pour identifier de composés « drug-like », notamment les machines à vecteurs desupport ^{11 et 12}. (Support Vector Machines ou SVM) et les réseaux de neurones, ^{11 et 13} avec des résultats comparables aux méthodes de Lipinski et Veber.

Tableau 1-2: Valeurs utilisées dans les différents filtres permettant la sélection de molécules ou fragments aux propriétés pharmacocinétiques « drug-like » ou « lead-like ». (*) Pour les critères de Veber (**). Dans l'utilisation de la « règle de 3 »

Propriétés physico-chimiques	« Drug-likeness »		« Lead-likeness »	
	Règles de Lipinski	Veber et al.	Hann & Oprea	« Règle de 3 »
Poids moléculaire (en Da)	≤ 500	-	≤ 460	< 300
Lipophilie (logP)	≤ 5	-	[-4, 4.2]	≤ 3
Nombre de donneurs de liaisons hydrogènes (DLH)	≤ 5	DLH + ALH ≤ 12 *	≤ 5	≤ 3
Nombre d'accepteurs de liaisons hydrogènes (ALH)	≤ 10		≤ 9	≤ 3
Aire de la surface polaire	-	$\leq 140 \text{ \AA}^2$ *	-	$\leq 60 \text{ \AA}^2$ **
Nombre de liaisons rotatives	-	≤ 10	≤ 10	≤ 3 **
Nombre de cycles aromatiques	-	-	≤ 4	-
Solubilité dans l'eau (logS _w)	-	-	≥ -5	-

2.5. Toxicologie :

En plus des caractéristiques ADME, il est également possible de filtrer les chimiothèques en fonction de la toxicité potentielle des composés. Parmi les caractéristiques de toxicité les plus importantes, la carcino-génicité, muta-génicité, cardio-toxicité et hépato-toxicité doivent être étudiées ¹⁴. L'émergence des méthodes de toxicologie computationnelle a permis de prédire de telles caractéristiques, en s'appuyant principalement sur deux techniques : l'établissement de modèles QSAR et les méthodes de Docking ^{15 et 16}.

L'application de ces techniques est rendue possible grâce à la disponibilité de données toxicologiques prévenant de diverses sources : industrie pharmaceutique et recherche académique, mais aussi industrie cosmétique, alimentaire et des produits de consommation courants.

3. Méthodes de criblage

Les méthodes de criblage permettent d'éliminer les composés présumés inactifs ou les molécules indésirables tout en privilégiant les composés les plus susceptibles d'être actifs, ce qui est plus facile à mettre en œuvre pour un coût bien moindre que les criblages expérimentaux. Les méthodes de criblage sont séparées en deux grandes familles : « ligand-based » et « structure-based », et le choix de leur utilisation est basé sur la disponibilité des données.

3.1. Criblage virtuel « ligand-based »

Les méthodes de criblage « ligand-based » basées sur la connaissance préalable de ligands avec une activité sur la cible thérapeutique. Il sera ainsi possible d'utiliser ces ligands comme une première base de « hits » afin d'identifier d'autres composés similaires, ayant des caractéristiques d'activité communes aux ligands connus de la cible.^{17 et 18}

Différents types de descripteurs moléculaires peuvent être calculés afin de quantifier la similitude entre composés. Selon le nombre de ligands connus de la cible thérapeutique, plusieurs méthodes peuvent être utilisées : la recherche de similarité, le criblage de pharmacophores, ou les approches QSAR.¹⁹

3.1.1. Recherche de similarité

La recherche de similarité est utilisée lorsqu'un ligand connu de la cible thérapeutique est identifié. Des métriques de similarité permettront ensuite de comparer les ligands de référence aux molécules criblées, à l'aide de descripteurs moléculaires appropriés, afin de prédire leurs profils d'activité.

3.1.2. Modèles pharmacophoriques « ligand-based »

Le terme « pharmacophore » utilisé pour la première fois par *Lemont.Kier* en 1971, publié plus tard en 1978²⁰. Le pharmacophore est défini comme étant l'agencement tridimensionnel des caractéristiques permettant à une molécule d'avoir une activité biologique spécifique. Il représente l'ensemble des propriétés moléculaires stériques et électrostatiques dites pointes

pharmacophoriques (les donneurs et accepteurs de liaisons hydrogènes, les interactions électrostatiques, les groupes aromatiques et hydrophobes) nécessaires à établir des interactions supramoléculaires optimales avec une cible biologique spécifique pour déclencher ou bloquer sa réponse biologique. Un pharmacophore peut donc être utilisé comme référence pour le criblage de chimiothèques, en recherchant les molécules qui s'inscrivent dans celui-ci.

Un pharmacophore est appelé « ligand-based » lorsqu'il est déterminé à partir de composés actifs utilisés comme référence sans tenir compte de la structure de leur récepteur. Inversement, un pharmacophore est appelé « structure-based » lorsqu'il est construit à partir de la structure du site de liaison de la cible étudiée.

La majorité des méthodes de construction de pharmacophore fonctionnent sur le même principe (**Figure 1. 2**)

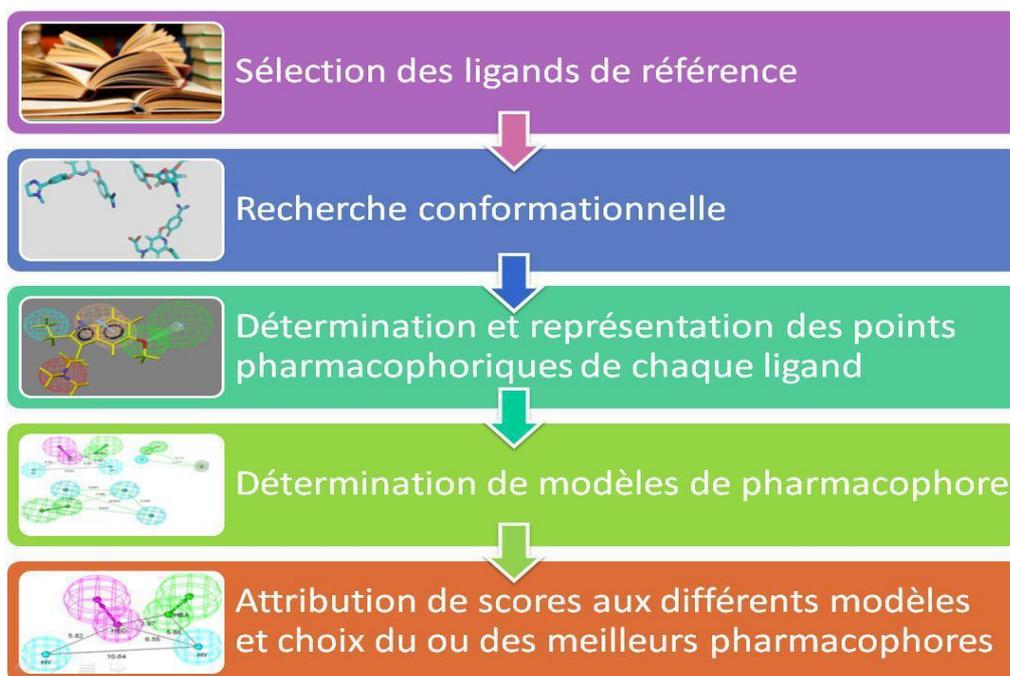


Figure 1.2 : Principales étapes de l'élucidation d'un pharmacophore.

3.1.3. Quantitative Structure-Activity Relationship (QSAR) « ligand-based »

Les méthodes QSAR « ligand-based » ont pour objectif d'établir des relations entre les propriétés physico-chimiques de ligands de référence avec leurs propriétés biologiques (activité expérimentale, toxicité, solubilité, etc.) au moyen de modèles statistiques (les régressions linéaires multiples (Multiple Linear Regression ou MLR) ou de moindres carrés (Partial Least Squares ou PLS), l'analyse en composantes principales (Principal Component Analysis ou PCA).

A. Principe :

Le principe des méthodes QSAR est comme leur nom l'indique, de mettre en place une relation mathématique reliant de manière quantitative la structure moléculaire, codée par des propriétés moléculaires appelées descripteurs, avec une activité en utilisant des méthodes d'analyse de données, l'objectif de ces méthodes est d'analyser les données structurales afin de déterminer les facteurs influençant l'activité mesurée.

La forme générale d'un tel modèle est la suivante :

$$\text{Activité} = f(D_1, D_2, \dots, D_n)$$

- D_n : sont des descripteurs des structures moléculaires

B. Stratégie globale :

Pour développer un modèle, il faut débiter par la recherche du maximum possible des données expérimentales fiables. Ensuite, le développement d'une série de descripteurs qui caractérisent les structures moléculaires des composés de la base de données en vue de les relier à l'activité expérimentale étudiée.

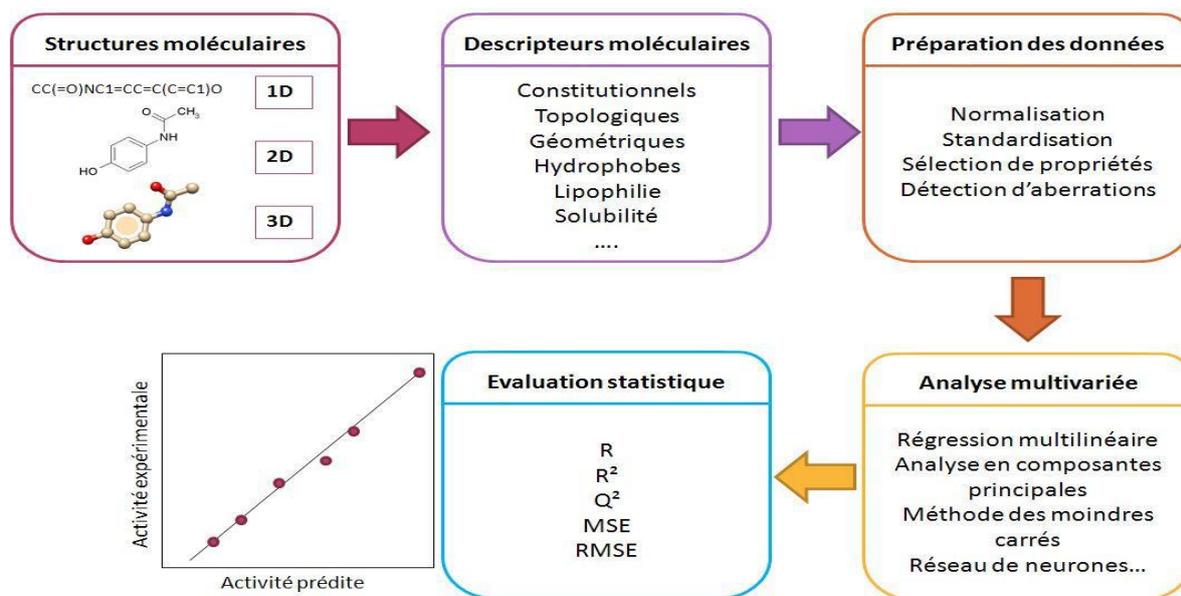


Figure 1.3 : Illustration du processus de conception de modèles QSAR ²¹.

C. Bases de données :

Une base de données QSAR, doit contenir des résultats fiables et homogènes, d'un ensemble de composés chimiques testés expérimentalement. Cette base est répartie approximativement, en deux ensembles : un lot d'essai ou d'entraînement (training set), contenant $\frac{3}{4}$ de l'ensemble de molécules et un lot de test (test set), contenant $\frac{1}{4}$ des molécules restantes ²¹.

Parmi les données biologiques ou biochimiques des composés étudiés la concentration inhibitrice médiane IC₅₀ et la constante d'affinité K_I sont généralement exprimées sur une échelle logarithmique (pIC₅₀, pK_I).

D. Descripteurs :

Un descripteur moléculaire est un paramètre (une valeur numérique) propre à une structure chimique donnée. Ces valeurs peuvent être obtenues expérimentalement ou calculées à partir de la structure de la molécule. Les descripteurs calculés, permettent d'effectuer des prédictions sans avoir à synthétiser les molécules, ce qui est l'un des objectifs de la modélisation moléculaire.

Les descripteurs moléculaires jouent un rôle fondamental dans les études de la relation quantitative structure activité/propriété. Ils sont utilisés en tant que variables indépendantes pour prédire une variable dépendante (activité ou propriété).

Il est possible d'établir des relations entre ceux-ci et une activité (ou propriété) à l'aide des outils de la modélisation .²²

De nombreux logiciels ont été développés pour calculer les différents descripteurs moléculaires tels que : OASIS ²³ , ADAPT ²⁴ ,DRAGON ²⁵ et MOPAC ²⁶ etc.

Dans ce travail de thèse nous avons utilisé le logiciel DRAGON, qui est capable de générer plus de 5000 descripteurs moléculaires.

Tableau 1-3: Classification des méthodes QSAR selon la dimension des descripteurs utilisés.

20,27

Dimension	Structure durécepteur	Méthode de prédiction
1D-QSAR	Ignorée	Corrèle l'activité à des descripteurs moléculaires 1D constitutionnels(poids moléculaire, pKa, logP, nombres absolus et relatifs d'atomes, d'hétéroatomes, de liaisons, de cycles et de groupes fonctionnels d'un type donné, etc.)
2D-QSAR	Ignorée	Corrèle l'activité à des descripteurs topologiques sans prendre en compte la géométrie (indices de connectivité, de complexité, de ramification, quantiques, etc.)
3D-QSAR	Optionnelle	Corrèle l'activité à l'arrangement spatial des champs d'interactions non-liées
4D-QSAR	Optionnelle	QSAR-3D avec inclusion de plusieurs conformations des ligands
5D-QSAR	Ignorée	QSAR-4D avec représentation de différents modèles « induced-fit » adaptation structurale de la cible au ligand
6D-QSAR	Ignorée	QSAR-5D avec inclusion de différents modèles de solvation
7D-QSAR	Nécessaire	QSAR-6D avec prise en compte explicite de la structure du récepteur

Lors d'une étude QSAR (**Figure1.3**), tous les descripteurs moléculaires possibles doivent être calculés à partir de la structure moléculaire de plusieurs actifs. Après élimination des descripteurs dont la valeur ne varie pas ou peu sur l'ensemble des molécules, une analyse multivariée suivie d'une évaluation statistique sont menées pour obtenir le modèle QSAR. Une dernière étape de validation du modèle est nécessaire pour s'assurer de sa fiabilité.

E. Méthodes statistiques :

Les techniques statistiques ou chimio-métriques constituent la base mathématique de la construction d'un modèle QSAR. La méthode la plus facile à interpréter est l'analyse de régression linéaire parmi diverses méthodes statistiques pour QSAR. Ces régressions représentent la corrélation directe des variables indépendantes (x) avec une variable dépendante (y). Ce modèle peut être considéré pour la prédiction de y à partir des données de variables x. Cela peut appartenir à un ensemble de systèmes qualitatifs ou quantitatifs.

a) Régression linéaire multiple (MLR)

Cette méthode est l'extension de SLR à plus d'une dimension. Dans cette méthode, des calculs de régression multi variables standard sont effectués. L'identification d'une propriété de médicament est effectuée sur tous les descripteurs sous enquête.

La possibilité de corrélation est vérifiée par la valeur du coefficient de corrélation multiple (r), valeur t à l'aide de la méthode « leave-one-out ». La corrélation est vérifiée par des valeurs R² ou Q² qui sont généralement connues sous le nom de coefficient de corrélation croisé. La relation est exprimée en équation linéaire unique à plusieurs termes comme suit ²⁸:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_mx_m + e. \quad (1)$$

b) Partiel des moindres carrés (PLS)

Le PLS donne une solution statistiquement robuste même lorsque les variables indépendantes sont fortement interreliées entre elles, ou lorsque les variables indépendantes dépassent le nombre d'observations. PLS est une méthode de régression itérative qui produit ses solutions basées sur la transformation linéaire d'un grand nombre de descripteurs originaux en un petit nombre de nouveaux termes orthogonaux appelés variables latentes. Ainsi, cette méthode est comptée comme statistique standard. ²⁹

c) *Analyse en composantes principales (ACP)*

Cette méthode est connue pour créer un nouvel ensemble de descripteurs orthogonaux appelés composants principaux (PC) qui décrivent la plupart des informations contenues dans les variables indépendantes dans l'ordre de la variance décroissante. Dans cette méthode, le modèle QSAR n'est pas généré mais il témoigne encore de la relation entre des variables différentes. PCA réduit la dimensionnalité de l'ensemble de données des descripteurs à la quantité réelle de données. Pour générer une équation linéaire à plusieurs termes, la PCR applique les scores de la décomposition de l'ACP en tant que régresseurs dans le modèle QSAR³⁰.

F. Paramètres statistiques de la validation

- *Coefficient de corrélation (R)*

C'est le coefficient de corrélation de Bravais Pearson entre \hat{y} et y , c'est à dire entre valeurs observées et prédites par le modèle de régression, il est noté R, sa valeur varie entre 0 et 1.³¹

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

Avec : y_i et \hat{y}_i sont respectivement les valeurs observées et calculées de la variable dépendante.

- $\bar{y} = y_{\text{obs}} = \sum \frac{y_i}{n}$
- $\hat{y} = y_{\text{cal}}$

- *Coefficient de détermination (R²)*

Le coefficient de détermination noté R² est la mesure de degré de liaison entre la variante dépendante (activité biologique) et la variante indépendante (Descripteur moléculaire), compris entre 0 et 1 ($0 \leq R^2 \leq 1$). En effet, une valeur de R² proche de 1, indique un bon ajustement du modèle de régression.³²

Il sert à déterminer à quel point l'équation de régression est adaptée pour décrire la distribution des points. Il représente la variabilité de y par la corrélation de y avec x .

Le R² est calculé selon la formule : Le coefficient R², est calculé selon l'équation suivante :

$$R^2 = \frac{ESS}{TSS} = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}$$

A noter que :

TSS = $\sum (y_i - \bar{y})^2$: Somme des carrées totaux

ESS = $\sum (\hat{y} - \bar{y})^2$: Somme des carrées exprimés par le modèle de régression

RSS = $\sum (y_i - \hat{y})^2$: Somme des carrés résiduels

Ainsi: **TSS = ESS + RSS**

- **Test Fischer (F)**

Le test Fischer, permet de justifier la liaison entre l'activité biologique et le descripteur moléculaire, en mesurant le rapport entre la variance de l'activité biologique expliquée et non expliquée par le modèle de régression. En effet, le test de Fischer permet de tester l'hypothèse nulle, lorsque la somme des carrés expliquée par la régression est grande. En d'autres termes, si la valeur de F observée dépasse le seuil ($F > F_{\text{seuil}}$) l'hypothèse sera rejetée, et dans le cas contraire ($F < F_{\text{seuil}}$) l'hypothèse sera conservée.³³

F peut s'écrire comme suit :

$$F = \frac{ESS}{p} \frac{n - p - 1}{RSS}$$

ESS = $\sum (\hat{y} - \bar{y})^2$: Somme des carrées exprimés par le modèle de régression

RSS = $\sum (y_i - \hat{y})^2$: Somme des carrés résiduels

Où : **ESS** est associé avec p degrés de liberté et **RSS** associé avec n-p-1 degrés de liberté.

- **Test de Student**

Pour voir la contribution de chaque paramètre dans l'explication de la variable dépendante Y. On utilise la statistique « t » définie auparavant en régression simple.

On calcule t-test pour chaque paramètre en utilisant la formule ci-dessous

$$t_{\text{Observé}} = \frac{\hat{a}_i}{s(\hat{a}_i)}$$

Avec $s(\hat{a}_i)$ est l'erreur sur le paramètre \hat{a}_i

En pratique si on choisit le risque $\alpha=5\%$ et si n est grand ($n>30$) pour approcher la loi de Student par la loi normale la valeur critique de t-test oscille le plus souvent autour de 2 (1,96).³⁴

- **Ecart type (s)**

L'écart type (s) est un autre paramètre habituellement rapporté ; il indique dans quelle mesure la fonction de régression prédit les données observées, ce paramètre est donné par³⁵ :

$$S = \sqrt{\frac{\text{RSS}}{n-p-1}}$$

Où :

$$\text{RSS} = \sum (y_i - \hat{y})^2$$

P : est le nombre de variables indépendants (descripteurs moléculaires).

- **R^2 ajusté**

Si l'on continue à augmenter le nombre de descripteurs dans un modèle pour un fixe nombre d'observations, les valeurs R^2 augmenteront toujours, mais cela conduira à une diminution du degré de liberté et à une faible fiabilité statistique. Ainsi, une valeur élevée de R^2 n'est pas nécessairement une indication d'un bon modèle statistique qui convient bien les données disponibles. Pour mieux refléter la variance expliquée (la fraction de la variance des données expliquée par le modèle), R^2 ajusté a été défini de la manière suivante :

$$R_{\text{ajust}}^2 = \frac{(R^2)(n-1)-p}{n-p-1}$$

p: nombre de variables utilisées

Le R^2 ajusté quantifie la part du modèle expliquée par les variables explicatives en tenant compte du nombre de variables utilisées, privilégiant les modèles contenant peu de variables. On choisit le modèle dont le R^2 ajusté est le plus

élevé. Ce critère est beaucoup plus judicieux que le R^2 classique qui lui privilégiera toujours le modèle contenant toutes les variables.³⁶

- **Coefficient de prédiction (Q^2)**

Le coefficient de prédiction q^2 , est une mesure de la justesse de la prédiction, puisqu'il mesure la capacité prédictive d'un modèle de régression³⁷.

Ce paramètre est exprimé comme suit :

$$Q_{cv}^2 = 1 - \frac{\sum_{i=1}^N (Y_{prédite} - Y_{réelle})^2}{\sum_{i=1}^N (Y_{réelle} - \bar{y})^2}$$

Où: \bar{y} : Y Prédite, réelle sont les valeurs prédites, réelles et moyennes de l'activité biologique (pIC50) de la molécule i, respectivement

G. Techniques de validation

Après l'obtention de l'équation du modèle, outre la stabilité et la qualité de l'ajustement du modèle, il est également important d'estimer la puissance et la validité du modèle avant de l'utiliser pour prédire l'activité biologique. La validité consiste à établir la fiabilité et la signification de la méthode pour un usage particulier. Par conséquent, la validation d'un modèle QSAR doit être effectuée.³⁸

Il existe deux types de validation : validation interne / validation externe.

➤ Validation interne

La validation interne utilise l'ensemble de données à partir duquel le modèle est construit et vérifie la stabilité interne. La technique de validation croisée (CV) est largement utilisée comme méthode de validation interne des modèles statistiques. Habituellement, un composé de l'ensemble est extrait à chaque fois, puis le modèle est recalculé en utilisant comme ensemble d'entraînement le n-1 (où n est le nombre de composés) des composés restants, de sorte que la valeur d'activité biologique pour le composé extrait est prédite une fois pour tous les composés. Ce processus est répété n fois pour tous les composés de l'ensemble initial, obtenant ainsi une prédiction pour chaque objet³⁹. Ce processus fait référence à la méthode LOO (leave-one-out).

➤ Validation externe

Un modèle QSAR avec une excellente qualité d'ajustement et des prédictions acceptables peut être déficient dans la relation réelle entre les descripteurs structurels et l'activité biologique. La validité parfaite du modèle est examinée par validation externe, qui évalue la généralisation du modèle. L'ensemble des données originales est divisé en deux sous-groupes, l'ensemble d'apprentissage et l'ensemble d'essai. L'ensemble d'apprentissage ou d'étalonnage est utilisé pour dériver un modèle d'étalonnage qui sera utilisé plus tard pour prédire les activités des composés d'essai ou de validation. D'autre part, un ensemble de test externe qui n'a été inclus dans aucune étape de la construction du modèle peut être utilisé comme ensemble de test ⁴⁰

3.2. Criblage virtuel « structure-based »

Comme leur nom l'indique, ces méthodes utilisent la structure de la cible afin de découvrir des nouveaux composés actifs. Lorsque la structure 3D de la cible biologique d'intérêt est disponible, des méthodes dites basées sur la structure peuvent être utilisées pour le criblage virtuel. Il existe deux types de méthodes expérimentales pour obtenir la structure 3D d'une cible: la cristallographie aux rayons X et la résonance magnétique nucléaire (RMN). ⁴¹ Lorsque la structure 3D expérimentale n'a pas encore été résolue, il est également possible de recourir à la modélisation par homologie. (**Figure1.4**)

Après avoir obtenu une structure de la cible thérapeutique, plusieurs approches « structure-based » peuvent être employées pour l'identification de composés actifs : la construction de modèles pharmacophoriques, la conception de ligands de novo, ou les méthodes RD-QSAR les méthodes de Docking qui sont les plus populaires. Chacune de ces méthodes nécessite l'identification préalable d'un site de liaison sur la cible thérapeutique.

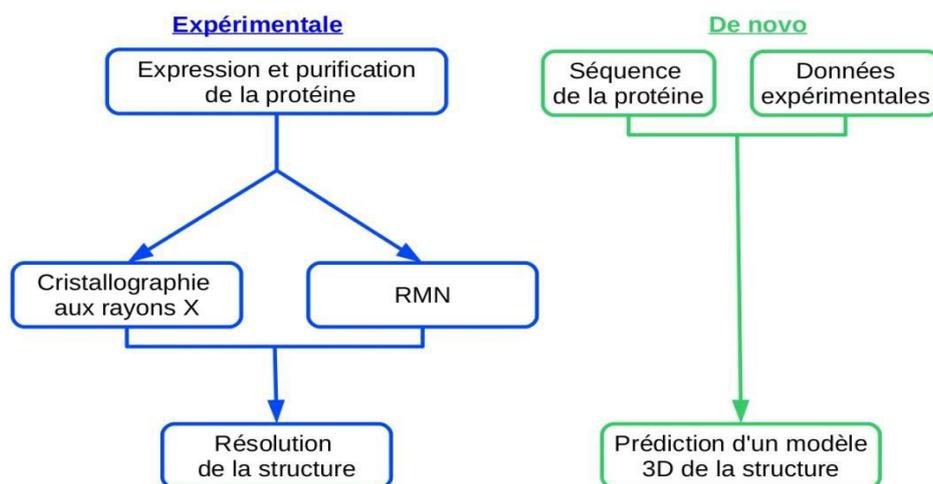


Figure 1. 4: Méthodes pour obtenir la structure 3D d'une cible biologique d'intérêt

3.2.1. Modèles pharmacophoriques « structure-based »

Comme décrit précédemment, les modèles pharmacophoriques « ligand-based » permettent d'identifier de nouveaux composés actifs à partir de la connaissance de ligands de la cible étudiée. Cependant, lorsque la structure de la cible a été résolue, l'utilisation des informations de forme, de volume ou physico-chimiques du récepteur peut rendre les méthodes pharmacophoriques plus puissantes.^{42,43} Les modèles pharmacophoriques « structure-based » peuvent être construits soit à partir de la structure du récepteur, soit à partir d'un complexe récepteur-ligand.^{42,43}

3.2.2. Quantitative Structure-Activity Relationship (QSAR) « structure-based » RD-QSAR (Receptor Dependent-Quantitative Structure-Activity Relationship)

Les méthodes QSAR « ligand-based » ont initialement été proposées et développées à une période pendant laquelle un faible nombre de structures de protéines étaient disponibles. Le logiciel CoMSIA237 a été proposé en 1999 alors que la PDB disposait d'environ 9000 structures cristallographiques. Par la suite, les méthodes QSAR ont été améliorées par utilisation des données structurales de la cible pour déduire les modes possibles de liaison des ligands et affiner leurs modèles, ce qui permet de mieux prédire les propriétés biologiques des composés. Les méthodes QSAR « structure-based », également dites RD QSAR, sont généralement classées selon trois catégories : 3D, 4D ou 7D-QSAR. Ce dernier (7D-QSAR désigne l'utilisation de méthodes RD 4D-QSAR tenant compte spécifiquement l'adaptation du

récepteur lors de l'accommodation d'un ligand (« induced-fit ») et différents modèles de solvation des complexes récepteur-ligand simulés.

3.2.3. Approches de Novo

La conception de *Novo* est une approche qui consiste à obtenir des composés bioactifs par une construction incrémentale ou combinatoire de ligands dans le site actif d'un récepteur (ou d'un enzyme).⁴⁴ Les méthodes de *novo* ont été développées afin de réaliser des assemblages d'atomes uniques ou de fragments moléculaires (approches « fragment-based »). La première étape est donc d'extraire toutes les informations concernant la structure du site actif de la cible étudiée et les interactions ligand-récepteur de type liaisons hydrogène, électro-statistique et hydrophobe pour déterminer les sites d'interaction au placement des divers blocs dans le site de liaison ligand-récepteur. Ces blocs peuvent être soit des atomes uniques, soit des fragments.

La majorité de ces méthodes sont basées sur l'utilisation de chimiothèques de fragments et d'algorithmes stochastiques permettant une optimisation locale à chaque étape de la construction des molécules dans le site de liaison de la cible.⁴⁴

Une solution de conception de *novo* doit être efficace en trois points principaux :

- (i) L'assemblage des différents fragments moléculaires,
L'échantillonnage de l'espace chimique accessible
- (ii) L'évaluation de la qualité des assemblages.

3.2.4. Docking moléculaire

Le Docking in silico (ancrage, amarrage ou arrimage moléculaire), est une technique informatique qui vise à prédire la structure d'un complexe formé par deux molécules (cible et ligand) figure 5 ou plusieurs molécules. La formation de ces complexes dépend de la reconnaissance de la structure tridimensionnelle d'un ligand par un site récepteur et contrôle de l'activité de nombreuses molécules. Le plus souvent, le récepteur est une protéine qui a un ou plusieurs sites actifs spécifiques, plus ou moins accessibles selon les cas. Le ligand, est généralement une molécule flexible étrangère de petite taille.⁴⁵ (voir figure 5)

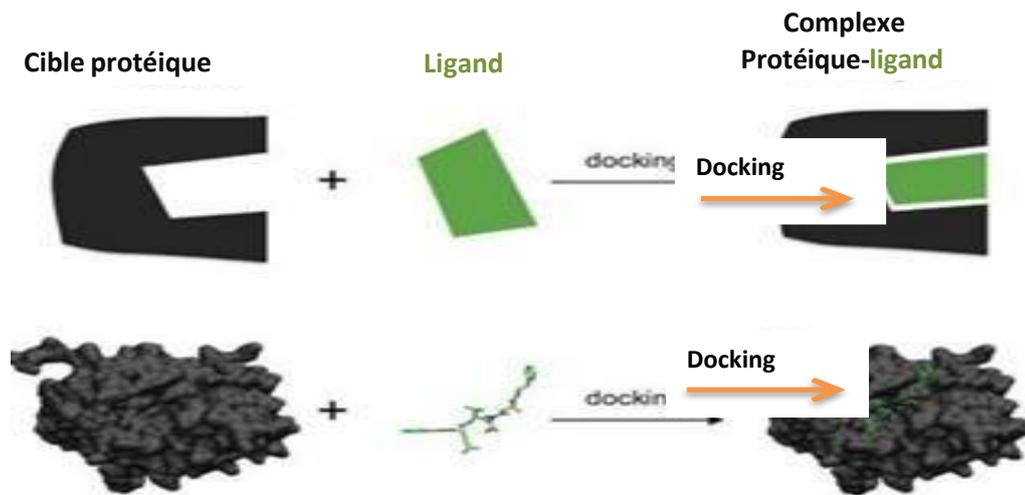


Figure 1.5: Représentation schématique du Docking moléculaire

L'utilisation des méthodes de Docking pour la conception de médicaments a commencé il y a plus de 30 ans. Ces méthodes ont deux objectifs principaux : (i) l'identification de ligands de la cible étudiée à partir de larges chimiothèques, grâce à la prédiction de scores d'affinité ou de valeurs d'affinité réelles (K_i , IC_{50} , etc.) et (ii) déterminer les modes de liaison adoptés par les ligands dans le récepteur.

A. Principe générale de Docking moléculaire

Le Docking se divise en deux étapes distinctes. La première section est constituée d'algorithmes de recherche, ces algorithmes sont capables de générer un grand nombre de structures possibles et de déterminer le mode de liaison. La deuxième section est consacrée au scoring, qui sont des méthodes mathématiques, utilisées pour estimer la puissance d'interaction et l'affinité de liaison entre deux molécules après la phase de Docking. Le meilleur résultat pour le Docking est le complexe protéine-ligand ayant la plus basse énergie. ⁴⁶

B. Types du Docking moléculaire

Il existe trois types de Docking moléculaire

a) *Docking rigide*

Le mécanisme de liaison d'un ligand à son récepteur a été considéré comme un processus statique dans lequel le ligand constituait une clé de forme complémentaire à celle de la serrure qu'il était capable « d'ouvrir », le récepteur (modèle « clé-serrure » ou « lock-and-key » model). Dans le cas des méthodes de Docking rigide, la recherche de la pose optimale se limite au positionnement par

l'énumération de toutes les rotations et translations possibles pour un ligand à l'intérieur du site de liaison. L'intérêt des méthodes de Docking rigide réside dans leur rapidité d'exécution qui permet de réaliser un premier filtrage de très larges chimiothèques avant de procéder, par exemple, à la mise en œuvre de méthodes de Docking flexible.⁴⁷

b) Docking flexible

Le système biologique ligand-récepteur a été présenté comme un système dynamique, dont la liaison du ligand à un site d'interaction se fait par un processus de sélection de conformation suivie par une étape d'adaptation du ligand et du récepteur l'un à l'autre.

Il est donc important pour les méthodes de Docking de considérer cette flexibilité du ligand et du récepteur. Lorsque les méthodes de Docking prennent en compte la flexibilité du ligand, deux étapes sont effectuées successivement pendant toute la durée du Docking.

Il existe plusieurs types d'algorithmes pour le traitement de la flexibilité du ligand : les méthodes systématiques (fragmentation/reconstruction), les méthodes aléatoires, et les méthodes de simulation (dynamique moléculaire).⁴⁸

c) Docking semi-flexible

Lorsque l'espace conformationnel des ligands est exploré, le nombre de degrés de liberté de l'espace de recherche peut être significatif pour les molécules très flexibles. Par conséquent, l'utilisation de méthodes de recherche exhaustives semble souvent inappropriée car elles nécessitent des simplifications importantes au niveau de l'échantillonnage. Les algorithmes de fragmentation sont donc employés pour construire de façon incrémentielle le ligand au sein du site actif de la protéine. L'espace des conformations du ligand est alors limité au voisinage d'un ensemble initial d'états simplifiés. Cette stratégie de recherche par construction est adoptée par les programmes DOCK, FLEXX, et Hammerhead.⁴⁷

C. Algorithmes de recherche

Il existe différentes Algorithmes qui permettent de prendre en compte la flexibilité du ligand notamment des Algorithmes systématiques et des Algorithmes aléatoires ou stochastiques.⁴⁹

a) Algorithme systématique

Les algorithmes de recherche systématique ont pour objectif d'explorer tous les degrés de liberté des ligands par rotation de 0 à 360° de toutes les liaisons rotatives à l'aide d'un pas incrémental choisi. Il en résulte que le nombre de conformations ainsi générées peut-être très élevé, c'est ce qu'on appelle l'explosion combinatoire. Pour pallier à ce problème, les méthodes de recherche systématique utilisent donc souvent un algorithme de fragmentation de reconstruction. Le principe général est de couper le ligand en fragments rigides et flexibles. Une ou plusieurs ancres peuvent être définies entre les points où des rotations sont possibles. Tout d'abord, un ou plusieurs fragments qui doivent être rigides sont placés au sein du site actif et donc mis en interaction avec la cible, puis le ligand est reconstruit en plaçant les fragments flexibles d'une manière successive tout en exploitant les angles de torsion. Cette méthode a été intégrée dans plusieurs programmes tels que Dock, FlexX et Surflex.⁵⁰

b) Algorithme stochastique

L'approche stochastique consiste à effectuer des changements aléatoires dans la structure tri-dimensionnelle du ligand. L'algorithme génétique constitue l'un des algorithmes stochastiques les plus importants. Cette méthode a été implémentée dans plusieurs programmes y compris AutoDock et GOLD. Concernant la recherche stochastique, on peut citer : La méthode Monté Carlo (MC), algorithme génétique.⁵¹

Dans la méthode Monté Carlo, le ligand est placé au sein du site d'interaction d'une manière aléatoire, où l'interaction sera évaluée par une fonction objective (fonction de score), et ainsi cette conformation est dite scorée. Une nouvelle conformation est générée selon des changements aléatoires au niveau des liaisons rotatables du ligand et sa position spatiale (translation et rotation). Après chaque changement, le ligand est minimisé et son énergie est calculée (scoré). Alors, si la nouvelle configuration est meilleure que la précédente selon une fonction de probabilité dérivée de Boltzmann, appelée critère de Metropolis, elle sera immédiatement acceptée.⁵²

D. Le processus de scoring

a) Principe

Le score est une donnée numérique utile pour quantifier le degré avec lequel un ligand se complexe à un récepteur. Il s'agit généralement d'une approximation de l'énergie libre résultant du passage de la forme libre de la protéine et du ligand à l'association sous forme de complexe. Le principe thermodynamique est le suivant⁵⁴ ;

$$\Delta G = \Delta G_{\text{complexe}} - \Delta G_{\text{ligand}} - \Delta G_{\text{protéine}} \quad (2)$$

L'utilisation des fonctions de scoring pour estimer mathématiquement l'affinité de liaison entre un récepteur et chacune des poses générées pendant le Docking est double. Tout d'abord, elles permettent de déterminer la meilleure conformation du ligand concerné. Cette conformation est appelée « première pose ». L'autre utilisation des scores est de pouvoir classer les premières poses de chaque ligand afin d'établir un classement final des molécules les plus prometteuses.

b) Familles de fonctions de score.

Les fonctions de scoring peuvent être classées dans plusieurs catégories: fonctions de score basées sur les champs de force et fonctions de scoring empiriques et fonctions de type knowledge-based (encore appelées « knowledge-based »). Ces classes de fonctions sont basées sur un ensemble de complexes ligand-protéine.⁵⁴ Elles ont donc la faiblesse de ne pouvoir prédire que des interactions qui ne dévient pas trop de celles ayant été répertoriées dans les complexes étudiés.⁵⁵

➤ **Fonctions de score basées sur les champs de force :**

Dans ce type de fonctions, le score est la somme de l'énergie d'interaction ligand-récepteur, de l'énergie interne du ligand et de l'énergie interne de la protéine, l'énergie interne de la protéine étant négligeable lorsque la conformation de celle-ci ne varie pas. Les énergies sont calculées à partir d'un terme de type Van Der Waals souvent (représentée par un potentiel de Lennard-Jones) et d'un terme électrostatique (potentiel de Coulomb) assorti d'une fonction diélectrique dépendante de la distance qui permet d'atténuer les contributions des interactions charge-charge. De nombreux logiciels de Docking utilisent des fonctions de score basées sur les champs de force. Cependant, les valeurs des champs de force ne sont

pas toujours les mêmes. Ainsi, les fonctions G-Score (basée sur le champ de force de Tripos (et celle implémentée dans AutoDock (basée sur le champ de force AMBER) sont des exemples de ce type de fonction de score.

➤ **Fonctions de score empiriques :**

Ce type de fonction de score approxime l'énergie libre de liaison en sommant de façon pondérée différents termes d'interaction dérivés de paramètres structuraux. Les différents poids de la fonction de score sont ajustés pour reproduire en priorité des données expérimentales, telles que les constantes de liaison tirées d'un jeu d'entraînement de complexes protéine-ligand. La plupart des programmes de Docking implémentent ce type de fonction de score témoignant de leur efficacité (en terme de rapport précision/rapidité). Cependant, le principal inconvénient de ces fonctions empiriques est leur forte dépendance aux données utilisées pour les calibrer qui, en cas de mauvaise paramétrisation, peut limiter leur transférabilité sur des systèmes différents. Parmi les principales fonctions de score empiriques, on peut citer : ChemScore, PLP, et LigScore. ⁵⁶

➤ **Fonctions de type knowledge-based :**

Ces fonctions proviennent de l'analyse des structures tridimensionnelles de complexes ligand- protéine déterminés de manière expérimentale. Des règles définissant la géométrie préférentielle des interactions sont déduites de ces structures grâce à des moyens statistiques.

- Cette alternative aux fonctions empiriques est plus tolérante quant aux interactions présentes au sein du complexe. Leurs expressions sont moins strictes que dans le cas des fonctions empiriques. La fonction PMF fait partie de cette classe de fonction.

Références:

1. Leach, A.R., Shoichet, B.K. and Peishoff, C.E. **Docking and Scoring-Perspective. Journal of Medicinal Chemistry**, (2006), 49, 5851-5855.
<http://dx.doi.org/10.1021/jm060999m>.
2. Enyedy, I. J. and Egan, W. J., J. **Comput.-Aided Mol**,(2008), Des. 22, 161
<https://doi.org/10.1007/s10822-007-9165-4>.
3. Reddy NP, Reddy CM, Aparoy P, Achari C, Ramu Sridhar P, Reddanna P. **Structure Based Drug Design, Synthesis and Evaluation of 4-(benzyloxy)-1-phenylbut-2-yn-1-ol derivatives as 5-Lipoxygenase inhibitors. Eur. J. Med. Chem**,(2012),47(1):351–359.
4. Lavecchia, A. & Giovanni, C. D. **Virtual screening strategies in drug discovery: A critical review. Curr. Med. Chem**, (2013),20, 2839–2860.
5. Lagunin, A. D. Filimonov, and Poroikov, V., **Curr. Pharm. Des**, (2010), 16, 1703–1717.
6. Cragg, G. M. Paclitaxel (Taxol®): **A success story with valuable lessons for natural product drug discovery and development. Med. Res**, (1998), Rev. 18, 315–331 .
7. Knox AJS, Meegan MJ, Carta G, Lloyd DG. **Considerations in compound database preparation—“hidden” impact on virtual screening results. J Chem Inf Model**,(2005),;45:1908–1919.
8. Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J. **Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. Advanced Drug Delivery Reviews**, (1997), 23, 3-25.
9. VEBER, D. F.; JOHNSON, S. R.; CHENG, H-Y.; SMITH, B.R.;WARD, K. W.; KOPPLE, K. D. **Molecular properties that influence the oral bioavailability of drug candidates. J. Med. Chem**, (2002), v.45, p.2615-2623,
10. Muegge I, Heald SL, Brittelli D. **Simple selection criteria for drug-like chemical matter. J Med Chem** (2001); 44:1841.
11. Zernov VV, Balakin KV, Ivaschenko AA, Savchuk NP, Pletnev IV: **Drug discovery using support vector machines. The case studies of drug-likeness, agrochemical-likeness, and enzyme inhibition predictions. J Chem Inf Comput Sci** ,(2003), 43:2048-2056.
12. Cortes, C. and Vapnik, V. **Support-Vector Networks. Machine Learning**, (1995) , 20, 273-297.
13. Ajay, Walters, W.P. and Murcko, M.A., **J. Med. Chem**, (1998) ,41 3314.
14. Toropov, A. A., Toropova, A. P., Raska Jr, I., Leszczynska, D. & Leszczynski, J. 181 **Comprehension of drug toxicity: Software and databases. Comput. Biol. Med**, (2014), 45, 20–25 .
15. Toropov, A.A , Toropova, A.P, Raska, I, Jr, D. Leszczynska, J. Leszczynski, **Comput. Biol. Med.**,(2014), 45, 20–25.
16. Valerio Jr., L.G. **In Silico Toxicology for the Pharmaceutical Sciences. Toxicology and Applied Pharmacology**, (2009) , 241, 356-370.
17. Bawden, D. **Applications of two-dimensional chemical similarity measures to**

- database analysis and querying.** In, Johnson, M.A. and Maggiora, G.M. (Eds.) **Concepts and Applications of Molecular Similarity**, Wiley, New York, (1990), pp. 65–76.
18. Hawkins PC, Skillman AG, Nicholls A ,**Comparison of shape-matching Chemometrics and QSPR Group, Talete srl, Milan, Italy and docking as virtual screening tools.** *J Med Chem*, (2007) , 50(1):74–82.
 19. Mason, JS, Good, AC, Martin ,EJ, **3-D pharmacophores in drug discovery.** *Curr Pharm Des*, (2001) , 7:567–597.
 20. KIER, LB. **Molecular orbital theory in drug design.** New York, American Press, (1971), 258 p
 21. Puzyn, T , Leszczynski, J,Cronin, M. T. D. , « **Recent advances in QSAR studies-Methods and applications** » Springer Dordrecht Heidelberg London New York, (2010) / . M. Tichý, *Science.Tot. Env*,(1991) 109, 407-410.
 22. Wu, J, Wang, Y, Shen, Y , **Chemomet. Intel. Lab. Syst**, (2014) ,13, 22-30.
 23. Mekenyan, O, Bonchev, D, *Acta Pharm. Jugosl* ,(1986),36,225-237.
[/http://oasis-lmc.org](http://oasis-lmc.org).
 24. Guha, R, Serra, J. R, Jurs, P. C, *J. Mol. Graph. Mod*,(2004),23, , 1–14./
<http://www.netsci.org/Resources/Software/Modeling/CADD/adapt.html>.
 25. Todeschini, R, Milano (http://www.taletemi.it/products/dragon_description.htm).
 26. <http://openmopac.net/>.
 27. Verma, J, Khedkar ,VM, Coutinho ,EC, **3D-QSAR in drug design - A review.** *Curr Top Med Chem*, (2010) , 10: 95-115.
 28. Heritage, T.Wet al, **Drug Discovery and Design**,(1998), p 381-398.
 29. Cherkasov, A, Muratov, E, Fourches, D, Varnek, A, Baskin, II, Cronin. M , Dearden . J , Gramatica. P , Martin. Y, Todeschini. R , Consonni .V , Kuzmin . V , Cramer. R , Benigni. R , Yang .C , Rathman. J , Terfloth . L , Gasteiger. J, Richard .A et A. Tropsha, *J. Med. Chim*, (2014), p4977–5010 .
 30. Schuur , J.H. , Selzer .P, Gasteiger. J. *of Chem. Inf. Comput*. (1996), p 334-344.
 31. Roy. K , Mitra . I , Kar .S, Ojha, P. K, Das , R. N , H. Kabir. J. *Chem. Inf. Model*, (2012), p(396–408.).
 32. Mark Cronin .T , Schultz .T, *Chem. Res. Toxicol*, (2001), p 1284-1295. / Balaza .S, Sturdik. E , Durkova . E, Antalik. M, Sulo. P, **Biochimica et Biophysica Acta (BBA) – Bioenergetics**,(1986), p 93-98. / Luco, J. M , F. H Ferretti.*J. Chem. Inf. Comput. Sci* ,(1997), p 392-401.
 33. Mark Cronin , T.D , Schultz ,T. W, *Chem. Res. Toxicol*, (2001) , p 1284-1295 / Balaza. S, Sturdik. E , Durkova . E, Antalik .M , Sulo .P, **Biochimica et Biophysica Acta (BBA) – Bioenergetics**,(1986), p 93-98./ Luco, J .M , F.H. Ferretti. *J. Chem. Inf. Comput. Sci*,(1997), p 392-401.
 34. Zabell, S. L, *J. American Statistical Association*, (2008), p 1-7.
 35. (Ojha, P.K , Mitra. I , Das. R.N , Roy. K , **Chemom. Intell. LabSyst**,(2011), p 194–205.
 36. Roy. K , Mitra .I, Ojha. P , Kar. S, Das .R.N , Kabir .H, **Chemom. Intell. Lab. Sys** ,(2012), p 200–210.
 37. Mark.T.D, Cronin, T.W. Schultz. *Chem. Res. Tox*. (2001), p 1284-1295. / S. Balaza, E.

- Sturdik., E. Durkova , M. Antalík , P. Sulo. **Biochimica et Biophysica Acta (BBA) – Bioenergetics**,(1986), p 93-98/ J . M. Luco, F. H. Ferretti. **J. Chem. Inf. Comput. Sci** ,(1997), p 392-401.
38. Shao J. **Linear model selection by cross-validation. Journal of the American statistical Association**,(1993),88(422):486-94.
 39. Golbraikh A, Shen M, Xiao Z, Xiao Y-D, Lee K-H, Tropsha A. **Rational selection of training and test sets for the development of validated QSAR models. Journal of computer-aided molecular design**,(2003),17(2-4):241-53.
 40. Shao J. **Linear model selection by cross-validation. Journal of the American statistical Association**,(1993), 88(422):486-94.
 41. Belbin A., Schneider,B., Clowney,L., Hsieh,S.-H., Olson,W.K. and Berman,H.M. **Geometric parameters in nucleic acids: sugar and phosphate constituents**, (1996), 118, 519–528.
 42. Jones G, Willett P, Glen RC, Leach AR, Taylor R. **Development and validation of a genetic algorithm for flexible docking. J Mol Biol**,(1997), Apr 4;267(3):727-48.
 43. Shulman-Peleg,A., Mintz,S., Nussinov,R. and Wolfson,H.J**Protein–protein interfaces: recognition of similar spatial and chemical organizations**, . (2004) , pp.194–205.
 44. Olivecrona et al., (2017); Schneider and Fechner, (2005)
 45. van Gunsteren, Wilfred F. ; Berendsen, Herman J. C **Angewandte Chemie**,(September 1990), vol. 102, issue 9, pp. 1020-1055
 46. Louacheni, F.**Docking protéine-ligand avec AutoDock, Mémoire de Master, Département d’informatique, Francophone Institute for computer sciences, Hanoi**, (2014).
 47. Alexandre ,B. **Développement et validation de la plateforme de criblage virtuel. Thèse de Doctorat : chimie informatique et théorique. Nancy : Université de Nancy. France**,(2008),. 122p.
 48. Ayaz, M.D, and Shafia M . **Molecular Docking: Approaches, Types, Applications and Basic Challenges. J Anal Bioanal Tech**,(2017).
 49. Brooijmans, N.; Kuntz, I.D. **Molecular recognition and docking algorithms. Annu Rev Biophys Biomol Struct**,(2003), 32: p. 335-73
 50. KRID ,A, **thèse de magister en chimie théorique .Modélisation de nouvelles molécules biologiques actives liées au stress oxydant. Université Mentouri de Constantine** , (2008),.
 51. Guillemain, H. **Thèse de doctorat, Evaluation et application de méthodes de criblage in silico, Université de Cnam**,(2012) .
 52. Zauhar, R.J..Moyna, G. L.Tian, et al, **Shape signatures: a new approach to computeraided ligand- and receptor-based drug design. J Med Chem.**,((2003), 46, 5674-90.
 53. ALMI ,I .**Mémoire en Master 2. Contribution à l’étude du Pharmacophore d’une série de dérivés de nitrobenzoxadiazole, sont des inhibiteurs de glutathione-S-transferase P1-1(GSTP) .Université Med Khider Biskra**,(2016),
 54. Stahl, M.; Rarey, M. **Detailed analysis of scoring functions for virtual screening J. Med. Chem.** (2001), 44, 1035-1042.
 55. Kollman, P.A. Massova, I. C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y.

- Duan, W. Wang, O. Domini, P. Cieplak, J. Srinivasan, D.A. Case, T.E. Cheatham, (2000).
56. **ASSES, Y. thèse de doctorat en Chimie Informatique et Théorique. France.**
Conception par modélisation et criblage in silico d'inhibiteurs du récepteur c-Met .
Université Henri Poincaré – Nancy I ,(2011),



CHAPITRE 02 :
COVID-19 ET SES INHIBITEURS



1. Historique du Covid.

Les coronavirus (CoV) sont des virus qui infectent l'homme et de nombreuses espèces animales (mammifères et oiseaux). Les premiers CoV ont été décrits chez les animaux depuis les années 1937. Tandis que, chez l'homme les 1er CoV ont été isolés en culture cellulaire dans les années 1960, à partir des sécrétions respiratoires d'individus présentant une infection aiguë.

Ces CoV n'ont d'abord pas reçu l'appellation « coronavirus », apparue plus tardivement dans le 1er rapport du Comité International de Taxonomie Virale (ICTV) en 1971.

En 1968, le terme « coronavirus » fait officiellement son apparition dans la revue Nature. Ce nouveau groupe de virus se définit alors, à partir de critères essentiellement morphologiques.

Les CoV sont des virus enveloppés grossièrement sphériques, avec un diamètre allant de 80 à 200 nm. Leur enveloppe porte à la surface de hautes projections formées de la protéine de surface S (spike) et disposées en couronne, d'où le préfixe « corona ».

L'intérêt accru porté au genre coronavirus a été exacerbé en 2003 avec la découverte du coronavirus humain impliqué dans le syndrome respiratoire aigu sévère (SARS-CoV) qui a brutalement émergé. L'homme se fait contaminer par la civette qui est un hôte intermédiaire. Par contre la chauve-souris semble être le réservoir naturel. ¹

Le 20 Septembre 2012, le premier cas de syndrome respiratoire sévère lié à un nouveau coronavirus a été rapporté ². Il s'agissait du Syndrome Respiratoire du Moyen-Orient lié au coronavirus (MERS-CoV) et une étude rétrospective faite sur des échantillons a révélé que ce nouveau coronavirus était responsable de maladie respiratoire grave impliquant des patients et des professionnels de santé en Jordanie. ³

Ce nouveau sous-type de coronavirus, contrairement celui responsable du syndrome respiratoire aigu sévère qui touchait l'homme par l'intermédiaire de la civette, utilise les camélidés comme vecteur. ⁴

Le 9 janvier 2020 le SARS-CoV-2 a été identifié. D'abord nommé « coronavirus de Wuhan ou 2019-nCoV », l'appellation « SARS-CoV-2 » a été proposé par International Committee on Taxonomy of Viruses, l'organisme en charge de la classification des virus.

Il est l'agent étiologique incriminé lors de l'épidémie infectieuse qui s'est répandue tout d'abord en Chine avant d'atteindre différents pays du globe terrestre à partir de fin décembre 2019. Cette maladie a été appelée « COVID- 19 » par l'OMS, le 11 février 2020.⁴

2. Caractéristique virologique du SARS CoV- 2

2.1. Classification et taxonomie

Depuis les années 2000, la taxonomie des CoV a été régulièrement revue par l'ICTV. Actuellement, les CoV appartiennent à l'ordre des *Nidovirales* et à la famille des *Coronaviridae*, elle-même subdivisée en 2 sous- familles, les *Coronavirinae* et les *Torovirinae*. En 2009, les *Coronavirinae* ont été divisés en 4 genres appelés *Alpha-*, *Beta-*, *Gamma-*, et *Deltacoronavirus*. Les *Alpha-*, *Beta-* et *Gammacoronavirus* remplacent les anciens CoV de types 1, 2 et 3. Le genre *Deltacoronavirus* a été défini en 2011 et regroupe essentiellement des virus aviaires. Le genre *Betacoronavirus* est subdivisé en 4 clades, nommés « a, b, c et d » ou « A, B, C, et D » selon les auteurs ». (Al hosani f. et jones TI,2014) (Figure 2.1).

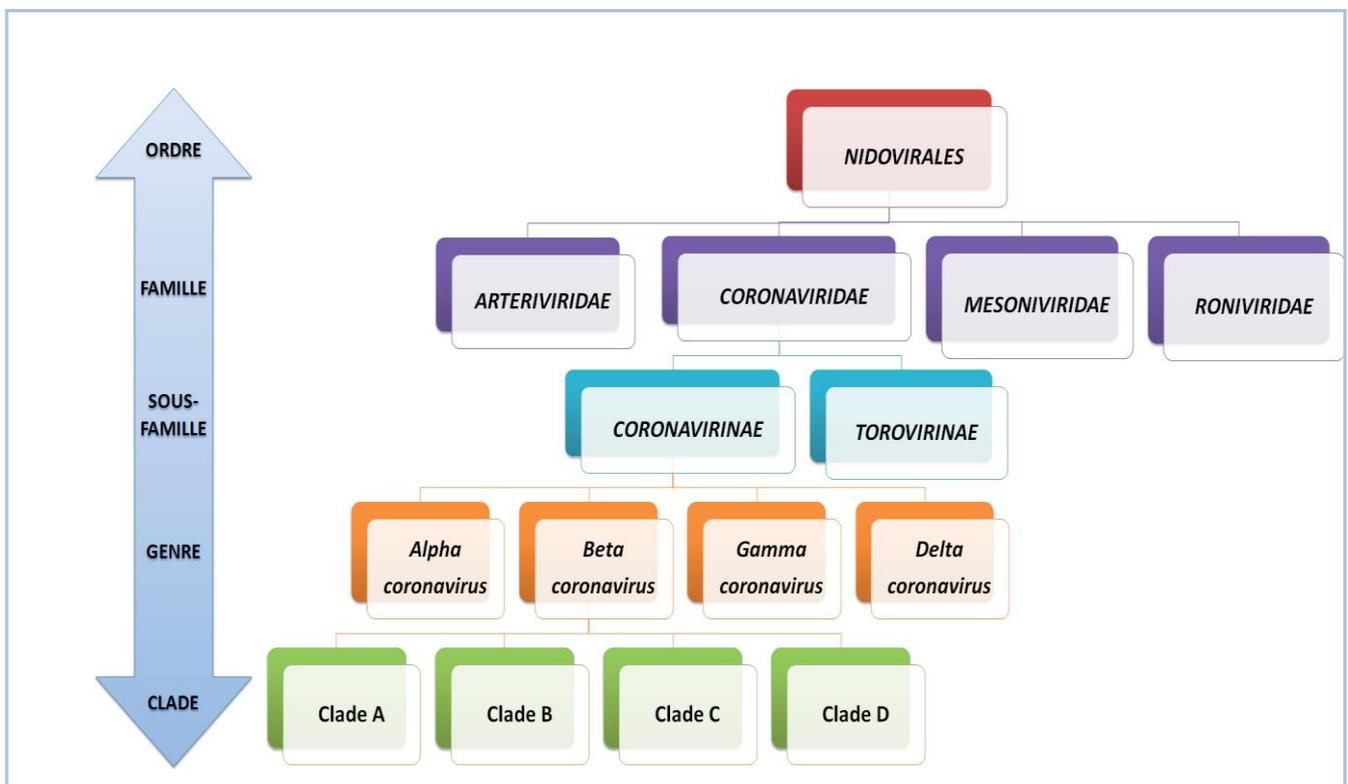


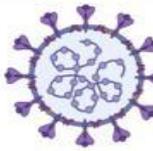
Figure 2.1: Classification et taxonomie des coronavirus humains (HCoV) Classification des HCoV selon l'ICTV en 2017

2.2. Types et origine

Les CoV sont des virus zoonotique, ce qui signifie qu'il se transmet de l'animal à l'être humain, ils infectent plus de 500 espèces de chiroptères, les oiseaux, le bétail, les civettes de palmiers masqués, les souris, les chiens et les chats ⁵.

Les alphacoronavirus et les betacoronavirus infectent seulement les mammifères. Les gammacoronavirus et deltacoronavirus infectent les oiseaux, mais certains d'entre eux peuvent également infecter les mammifères. Les Alphacoronavirus et les betacoronavirus causent habituellement une maladie respiratoire chez les humains et une entérite gastro-intestinale chez les animaux. Les deux virus hautement pathogènes, le SARS-CoV et le MERS-CoV causent des problèmes respiratoires graves le syndrome chez les humains tandis que les coronavirus (HCoV-NL63, HCoV-229E, HCoV-OC43 et HKU1) n'induisent que de légères maladies des voies respiratoires supérieures hôtes immunocompétents, bien que certains d'entre eux peut causer des infections graves chez les nourrissons, les jeunes enfants et les personnes âgées. ⁶

Tableau 2-1: Comparaison épidémiologique entre différents types de coronavirus

Maladie	Maladie à coronavirus 2019 (COVID-19)	Syndrome respiratoire aigu sévère (SRAS)	Syndrome respiratoire du Moyen-Orient (MERS)
Virus causant la maladie	 SRAS-CoV-2	 SRAS-CoV	 MERS-CoV
Taux de reproduction (R_0)	2 à 5,7 ^a	3	0,3 à 0,8
Taux de létalité	Autour de 1 % ^b	9,6 à 11 %	34,4 %
Temps d'incubation	2 à 14 jours ^c	2 à 7 jours	6 jours
Taux d'hospitalisation	6 à 19 % ^d	La plupart des cas	La plupart des cas
Taux potentiel d'infection communautaire	35 à 80 % ^e	10 à 60 %	4 à 13 %

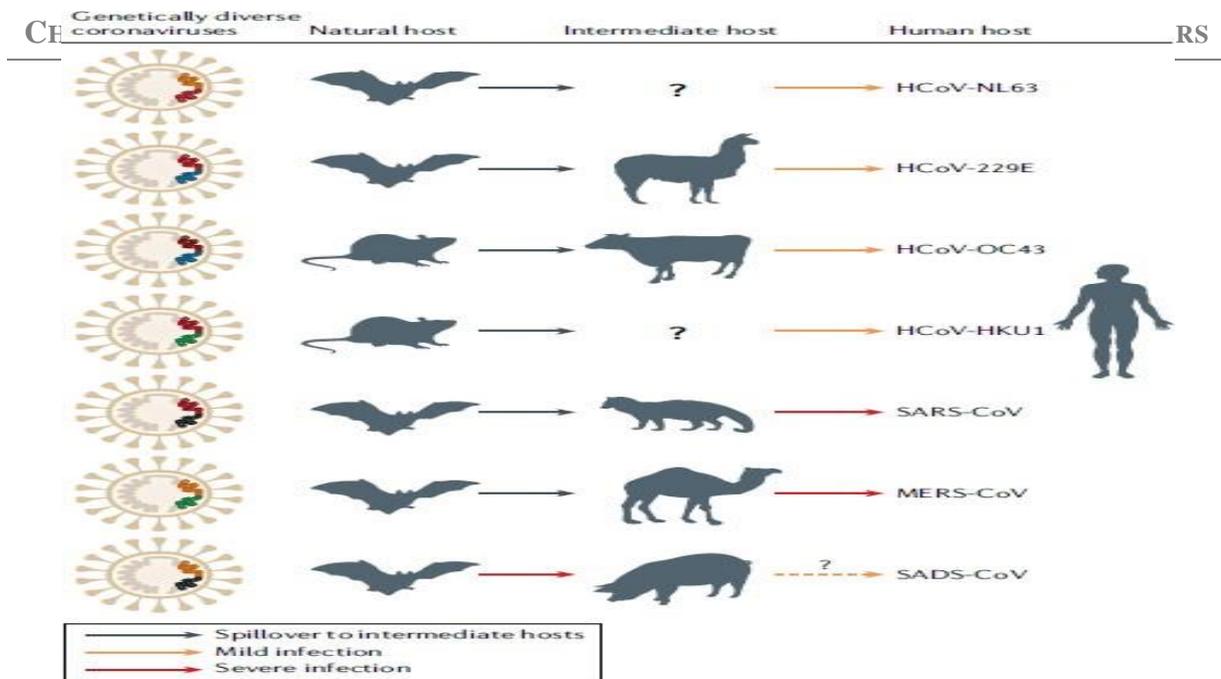


Figure 2.2 : Origines animales des coronavirus humains ⁶

2.3. Structure de virion et son génome

2.3.1. Le virion

Le virus du SRAS-CoV-2 possède une enveloppe de forme ronde ou elliptique et souvent pléomorphe, avec un diamètre de 60 à 140 nm. Les pléomorphes à la surface du virion sont des particules sphériques de 150 à 160 nm, donnant au virus sa morphologie « corona » ou en forme de couronne au microscope. Le virion du SARS-CoV-2 comporte de l'intérieur de la particule vers l'extérieur : le génome associé à un ARN monocaténaire positif non segmentée, a des nucléoprotéines, entouré d'une capsid de type hélicoïdale formée de protéine N, une matrice formée de protéine M puis une enveloppe lipidique dans laquelle sont enchâssées la glycoprotéine S (spike), la petite protéine d'enveloppe (E) et l'hémagglutinine-estérase (HE) (Figure 2.3) ⁷

A. La Glycoprotéine S

La protéine S est la protéine qui lie le répéteur cellulaire du SARS-CoV-2 (ACE2) et permet l'entrée dans la cellule. Elle est formée de deux sous-unités S1 et S2. ⁸

Dans la sous-unité S1, se trouve le domaine de liaison au récepteur (RBD, *receptor binding domain*) contenant le motif de liaison au récepteur (RBM, *receptor-binding motif*) (Figure 2.4). La sous-unité S2 contient notamment le peptide de fusion. Cette protéine S induit la production

d'anticorps neutralisants par le sujet infecté. Le domaine RBD (receptor binding domain) est la région de la protéine spike entrant en contact avec le récepteur cellulaire ACE2 qui sert de porte d'entrée au SARS-CoV-2 dans les cellules qu'il infecte ⁹.

B. La protéine M

La protéine M de matrice (ou de membrane) est la protéine virale la plus abondante présente dans la particule du virion donnant une forme définie à l'enveloppe virale, ¹⁰ elle se lie à la nucléocapside et agit comme un organisateur central de l'assemblage des coronavirus. ¹¹

Les protéines M des coronavirus sont très diverses en termes de contenu en acides aminés, mais conservent une similarité structurelle globale au sein des différents genres, elle possède trois domaines transmembranaires, flanqués d'une courte extrémité aminée à l'extérieur du virion et d'une longue extrémité carboxy terminale. ¹²

C. La protéine E

La protéine E du coronavirus est la plus petite des principales protéines structurelles. ¹³ Elle joue un rôle dans l'assemblage, la pathogenèse et la libération du virus. ¹⁴ Ce petit polypeptide membranaire intégral agit comme une viroporine (canal ionique). ¹⁵ L'absence de cette protéine est liée à l'altération de la virulence des coronavirus en raison de changements dans la morphologie et le tropisme. ¹⁶

D. La protéine N

La protéine N des coronavirus est polyvalente. Parmi plusieurs fonctions, elle joue un rôle dans la formation de complexes avec le génome viral, facilite l'interaction avec la protéine M nécessaire lors de l'assemblage du virion, et améliore l'efficacité de la transcription du virus. ^{17,18}

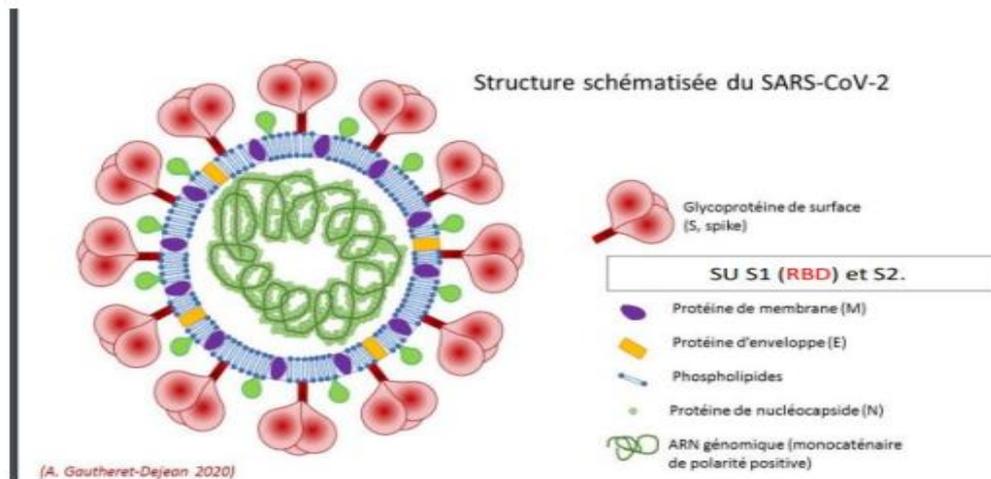


Figure 2.3: Structure schématisée du SARS-CoV-2.

Le génome est constitué d'un brin d'ARN de polarité positive pour interagir avec la nucléoprotéine de capsid (N) pour former une nucléocapside en forme d'hélice. ¹⁹

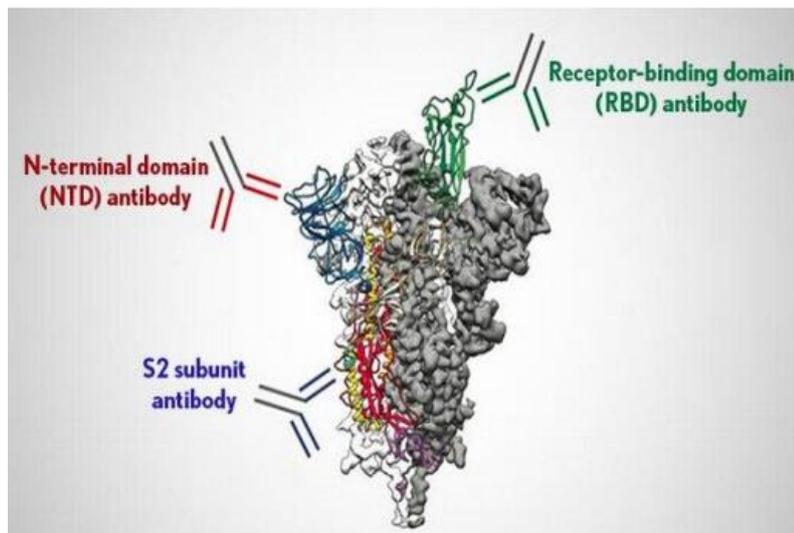


Figure 2.4 : Le domaine RBD (Receptor Binding Domain)

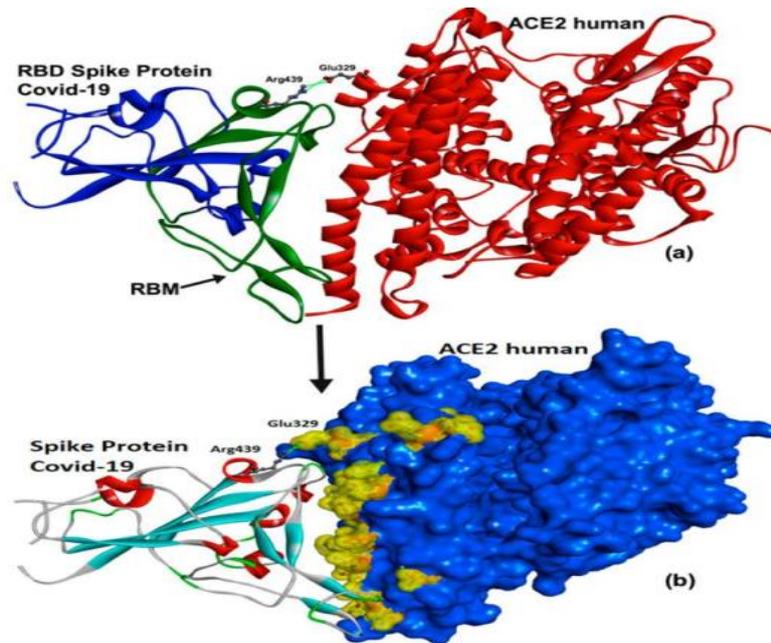


Figure 2.5 : (a) Structure du domaine de liaison au récepteur (RBD) de la protéine S du SARS-CoV-2 (rubans bleu) complexé avec l'ACE2 humaine (rubans rouge).

*Le ruban vert indique le motif de liaison au récepteur (RBM) dans les résidus d'acides aminés. (b) Le site actif de l'ACE2 (couleur jaune) qui interagit directement avec le RBD de la protéine S du SARS-CoV-2. L'interaction entre la protéine S de SARS-CoV-2 et hACE2 est stabilisée par une liaison hydrogène (lignes vertes).*²⁰

2.3.3. Le Génome

Le génome d'une longueur d'environ 29,8 à 29,9 kbases comporte 10 cadres de lecture ouverts (*open reading frames*, ORF) et code environ 30 protéines.²¹ Les deux tiers du génome codent pour un vaste gène réplicase ORF1ab (composé de orf1a et orf1b) qui sera traduit en deux polyprotéines, par la suite clivées en seize protéines non structurales indispensables à la réplication virale incluant deux protéases virales et une ARN-polymerase ARN-dépendante.²² Le reste du génome code essentiellement pour les protéines de structures du virus dont quatre glycoprotéines membranaires - la protéine Spike (S), l'Hémagglutinine-Esterase (HE) et les protéines de membrane (M) et d'enveloppe (E)-ainsi que la protéine de capsid (N). L'ORF1a code les protéines 1 à 11 et l'ORF1b celles de 12 à 16.^{23,24}

Six gènes codant des protéines accessoires sont présents entre les gènes des protéines de structure (**Figures 2. 6**)

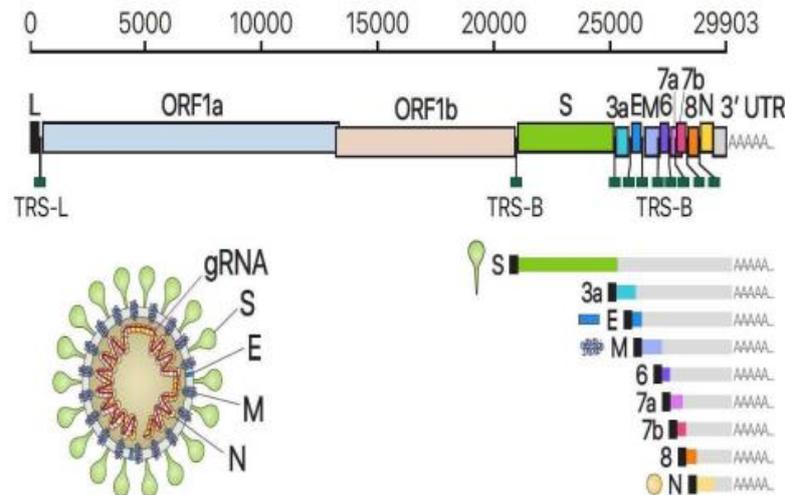


Figure 2. 6 : Présentation schématique de l'organisation du génome du SRAS-CoV-2, des ARNm sous génomiques canoniques et de la structure du virion ²⁵

3. Pathogénie du SARS CoV-2

3.1. Transmissions

La transmission de ce virus de l'animal à l'homme était l'hypothèse de base, car plus de la moitié des sujets atteints avaient fréquenté le marché de fruits de mer, cependant après un certain temps, elle était écartée. De nos jours, il est admis que la transmission interhumaine est la principale voie de dissémination ²⁶. Le virus peut pénétrer dans l'organisme :

- Par contact avec les yeux, nez .
- Par inhalation de gouttelettes/sécrétions d'un malade
- Cas de contact avec des surfaces infectées.

3.2. Pénétration du virus dans la cellule hôte

La protéine S du SARS-CoV-2 utilise le récepteur cellulaire ACE2, une métalloprotéase dont la fonction première est la dégradation de l'angiotensine II en angiotensine 1-7 - pour rentrer dans la cellule hôte .²⁷ Bien étudiée chez le SARS-CoV-1, la liaison de la sous unité S1 à ACE2 entraîne une modification conformationnelle de la protéine S, exposant S2 et permettant l'endocytose puis la fusion membranaire .²⁸ Cette fusion nécessite l'activation de S par le clivage au niveau de la jonction S1/S2 et d'un autre site de S2, notamment réalisée par la protéase membranaire TMPRSS2 (transmembrane protease serine 2). ²⁹

Dans le cas du SARS-CoV-2, l'ajout d'un site de clivage furine ³⁰ permet un clivage des sousunités S1/S2 dès la biosynthèse virale ³¹ et pourrait majorer le potentiel infectant du virus ³². De fac, on intéressante, en dehors d'ACE2, le SARS-CoV-2 pourrait également utiliser d'autres récepteurs cellulaires de la protéine S pour infecter les cellules n'exprimant pas ACE2, ainsi que démontrée sur des lymphocytes T in vitro .²⁶

3.3. Cycle de réplication

Le cycle de réplication des coronavirus a été largement étudié. Après la fusion et le largage de la nucléocapside dans le cytosol de la cellule hôte, la machinerie cellulaire traduit le gène de la réplicase

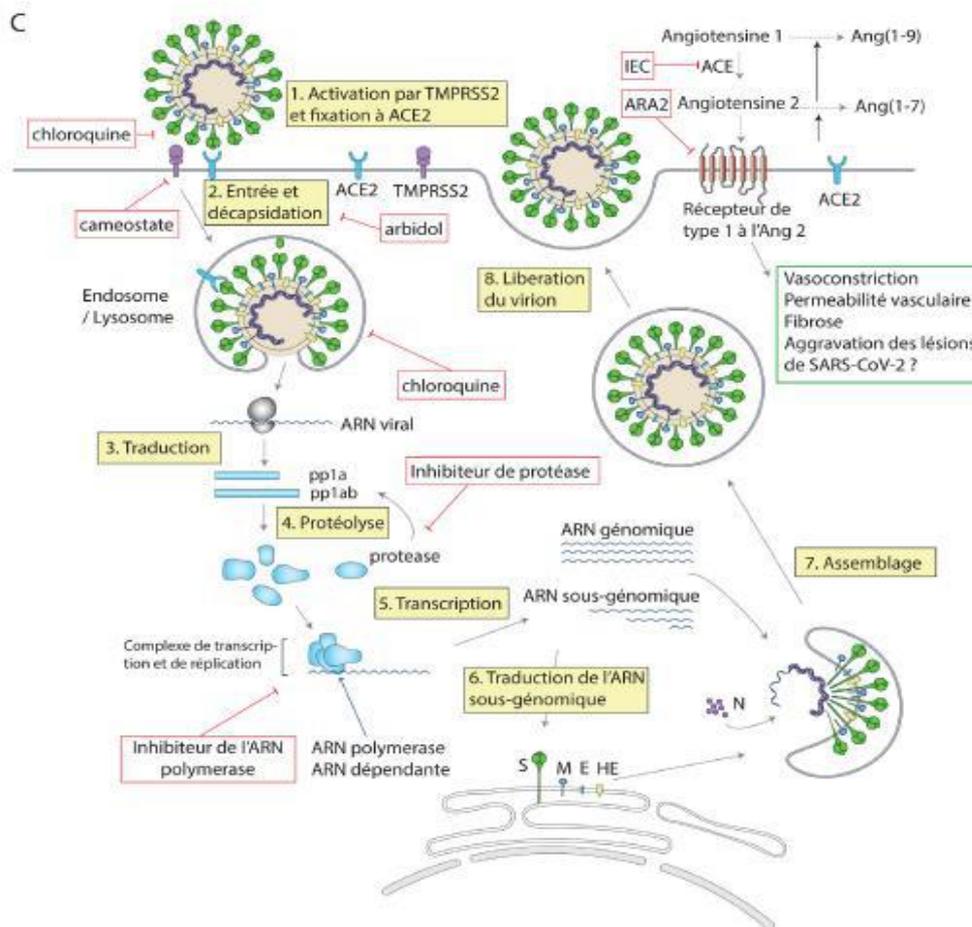


Figure 2.7 : Cycle de réplication du SARS-CoV-2 dans les cellules hôtes (Zhou et *al.*, 2020).

La nucléocapside formée de l'ARN viral complexé à la protéine N est enchâssée à l'intérieur de l'enveloppe. Le génome viral doit être répliqué et traduit afin de produire de nouveaux virions.

- La première étape de la réplication est la traduction du gène de la réplicase en deux polyprotéines (pp1a et pp1ab) clivées en nombreuses protéines indispensables au cycle viral (notamment deux protéases virales et une ARN-polymérase ARN-dépendant) (RdRp) s'assemblant en un large complexe de transcription et de réplication (RTC) .²⁹
- Ce complexe permet d'une part de reproduire l'ARN viral et d'autre part, par le biais de la formation de petits brins d'ARN anti-sens appelés ARN sous-génomiques, la production de protéines de structure des nouveaux virions.
- Finalement les brins d'ARN synthétisés sont combinés avec la protéine N pour former la nucléocapside et l'assemblage avec les glycoprotéines d'enveloppe permet le bourgeonnement de nouvelles particules virales. ³³ La connaissance du cycle viral permet de déterminer les cibles thérapeutiques inhibant sa réplication.

3.4. Les variants du SARS-CoV-2

3.4.1. Signification de « variant »

Au cours du processus de réplication du virus (la production de nouvelles unités), de petits changements dans le virus appelés mutations se produisent. Un virus qui porte une ou plusieurs nouvelles mutations est appelé un variant du virus d'origine. La propagation du virus augmente la fréquence des mutations. Ces mutations produisent parfois des variants plus adaptés à leur environnement que le virus d'origine. Le terme « évolution virale » fait référence à ce processus de mutation et de sélection de variants plus appropriés. Certaines mutations peuvent provoquer des changements dans les caractéristiques du virus, comme une modification de la transmission (par exemple, il peut se propager plus facilement) ou de sa gravité (par exemple, il peut provoquer des formes de maladie plus graves.).³⁴

3.4.2. La dénomination des différents variants

A. Les variants préoccupants ou VOC (variants of concern) :

Les premiers variants préoccupants (VOC) sont apparus à partir de fin 2020. Ils font l'objet d'une dénomination internationale, définie par l'OMS et basée sur l'alphabet grec.

a) Le variant Alpha s'est rapidement propagé en France après son introduction fin 2020 et est devenu majoritaire en mars 2021.

- b) *Les variants Beta et Gamma* ont également circulé au premier semestre 2021, de manière toutefois moindre.
- c) *Le variant Delta* est apparu en mai 2021 et a rapidement remplacé les précédents variants : il est devenu majoritaire en France en juillet 2021 et représentait plus de 99% des variants circulants à partir du mois d'août 2021.
- d) *Le variant Omicron* est le dernier VOC apparu, fin novembre 2021, et sa diffusion a été croissante. Il a rapidement augmenté en fréquence dans le monde entier.

B. Les variants à suivre ou VOI (variants of interest)

D'autres variants, porteurs de mutations qui les distinguent des souches virales de référence du SARS-CoV-2, sont également régulièrement identifiés, en partie du fait de ce renforcement des capacités de séquençage.

Pour certains de ces variants nouvellement détectés, leur impact en santé publique n'est pas formellement démontré, mais leurs caractéristiques virologiques, cliniques et/ou épidémiologiques justifient le classement en « **variants à suivre** » (« variants of interest », ou VOI en anglais).

C. Les variants sous surveillance ou VUM (variant under monitoring)

Les variants du SARS-CoV-2 sous surveillance sont le résultat de plusieurs mutations génétiques susceptibles d'affecter les caractéristiques du virus. C'est un potentiel risque futur sans pour autant que les répercussions phénotypiques ou épidémiologiques ne soient claires à ce jour. Il faut accorder une surveillance particulière à ces variants et répéter les évaluations en attendant de nouvelles preuves ³⁵

4. Épidémiologie

4.1. Évolution de l'épidémie

Actuellement, les patients COVID-19 sont la principale source d'infection et les patients sévères sont considérés comme plus contagieux que les patients bénins. Les personnes infectées de manière asymptomatique ou les patients en incubation qui ne présentent aucun signe ou

symptôme d'infection respiratoire dont la dissémination de virus infectieux est prouvée peuvent également être des sources potentielles d'infection.

De plus, les échantillons prélevés sur des patients récupérés du COVID-19 montrent en continu un test RT-PCR positif, ce qui n'a jamais été vu dans l'histoire des maladies infectieuses humaines. En d'autres termes, les personnes infectées de manière asymptomatique et les patients en incubation ou récupérés du COVID-19 peuvent poser de sérieux défis pour la prévention et le contrôle de la maladie .³⁶

- **Au Monde :** A partir de 29-10-2022 il y'a plus de 630 millions de cas de Covid-19 confirmés dans le monde plus de 6,59 millions de décès liés au coronavirus.
- **En Algérie :** En mars 2020, le premier cas de maladie à coronavirus confirmée (COVID-19) a été enregistré à l'institut Pasteur, le patient qui a présenté une pneumonie aiguë était un cas importé d'Europe. Depuis lors et jusqu'à et jusqu'à 29-10-2022, un total de 271 k cas confirmés en laboratoire d'infection par le coronavirus 2 du syndrome respiratoire aigu sévère (SARS-CoV-2) ont été signalés dans les différentes régions du pays dont 6881 décès.

5. Principale thérapies et prophylaxie

Pendant l'épidémie de COVID-19, certains médicaments antiviraux potentiels sont administrés en urgence aux patients atteints de COVID-19 .

Parmi ces médicaments :

5.1. Les thérapies antivirales

5.1.1. La chloroquine - L'hydroxychloroquine

La chloroquine (CQ) et l'hydroxychloroquine (HCQ) sont des anti-malariques de la classe des amino-4-quinoléines ont un grand volume de distribution et sont fortement liés aux protéines plasmatiques. Leur métabolisme est hépatique et leur élimination principalement rénale. La CQ est un substrat du CYP2C8 capable de conduire des interactions médicamenteuses qui doivent être prises en compte. ³⁷

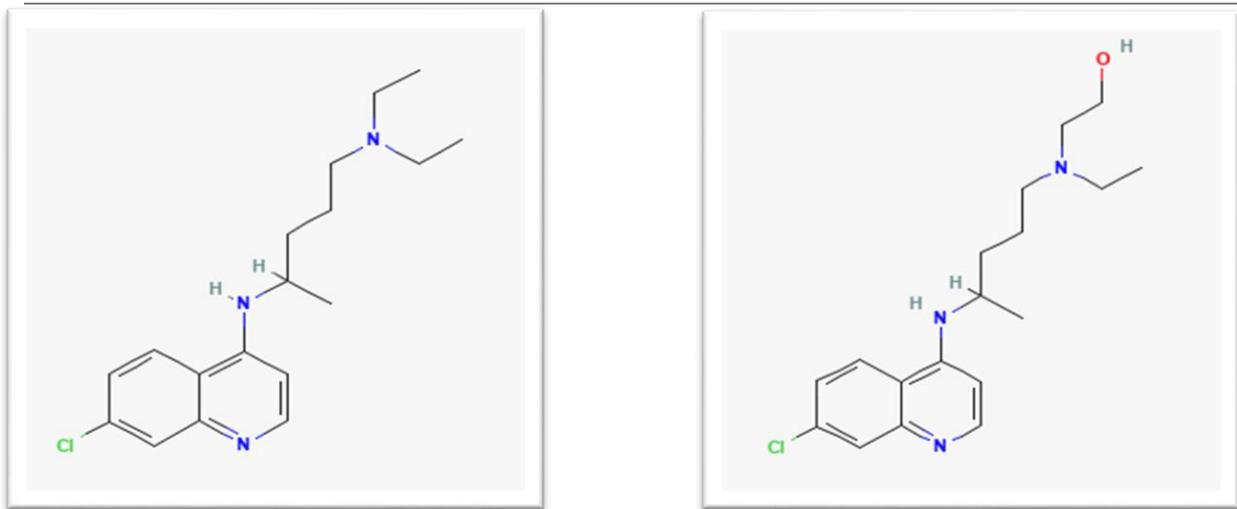


Figure 2.8 : Structures chimiques de la chloroquine et de l'hydroxychloroquine

En plus de leurs effets antipaludiques, la chloroquine et l'hydroxychloroquine ont également des effets antiviraux à large spectre. Les effets antiviraux de la chloroquine ont été suggérés il y a au moins 50 ans. Dès lors plusieurs études ont montrées l'activité antivirale de la chloroquine et de l'hydroxychloroquine contre plusieurs souches virales CoV et non-CoV. La majorité de ces études ont révélé une capacité substantielle de la CQ et de l'HCQ ainsi que de certains de leurs dérivés à inhiber la réplication virale avec une toxicité nulle à faible. Il a été démontré que la CQ inhibe la réplication de différents CoV, notamment le SARS-CoV, le MERS-CoV et le SARS-CoV-2. Pour l'HCQ qui ne semble pas avoir d'effet significatif sur la réduction de la réplication du SRAS-CoV, elle a démontré de puissants effets inhibiteurs in vitro contre la réplication du SRAS-CoV-2.

Outre l'activité antivirale, la chloroquine a une activité de modulation immunitaire, qui peut renforcer synergiquement son effet antiviral in vivo ³⁸. La CQ et l'HCQ sont des bases faibles connues pour élever le pH des organelles intracellulaires acides, telles que les endosomes/lysosomes, essentiels à la fusion des membranes. La cavité interne des endosomes et des lysosomes est maintenue à un pH de 4,5 à 6,5. La CQ et l'HCQ pourraient perturber le processus d'acidification en empêchant la maturation de l'endosome et en empêchant le transport et la libération ultérieurs de virions. Ce mécanisme semble important pour la pathogénie du SARS-CoV.³⁸

En effet, le professeur Didier Raoult, à Marseille, qui fait partie de l'essai clinique européen Discovery, a empêché ces modifications de pH et enrayé ainsi l'entrée du virus dans la cellule. ³⁹

5.1.2. Inhibiteurs de la protéase du VIH :

Le principe d'action de ces molécules est bloquer les protéases. Des inhibiteurs de la protéase 3CLpro des coronavirus déjà connus ont été testés pour leur capacité à inhiber celle du SARS-CoV-2.

Dans l'urgence, les firmes du monde entier tentent d'identifier des agents antiviraux efficaces pour lutter contre le COVID-19, et s'intéressent, entre autres, à plusieurs molécules déjà utilisées contre le HIV. Ces inhibiteurs de protéase du HIV, sont tous des peptidomimétiques. Ils se fixent sur le site actif de la protéase virale, et l'empêchent de catalyser le clivage des polypeptides du virus. La maturation protéique étant altérée, les virions formés sont inactifs et non infectieux ⁴. Parmi ces médicaments qui ont montré une efficacité contre le coronavirus en inhibant la réplication du virus, l'association lopinavir / ritonavir, remdesivir, favipiravir, umifenovir. ⁴¹

A. Lopinavir-ritonavir

Le lopinavir-ritonavir est une association de lopinavir, qui inhibe la protéase chymotrypsine 3 virale, et de ritonavir, qui inhibe l'isoenzyme 3A4 du cytochrome P450, soit l'enzyme qui métabolise le lopinavir, ce qui en augmente la biodisponibilité. Cette association utilisée pour le traitement du VIH a montré être efficace pour le traitement des virus SRAS-CoV en inhibant la réplication du virus. Divers rapports ont également revendiqué avoir démontré l'efficacité du lopinavir-ritonavir contre le virus SRAS-CoV-2.

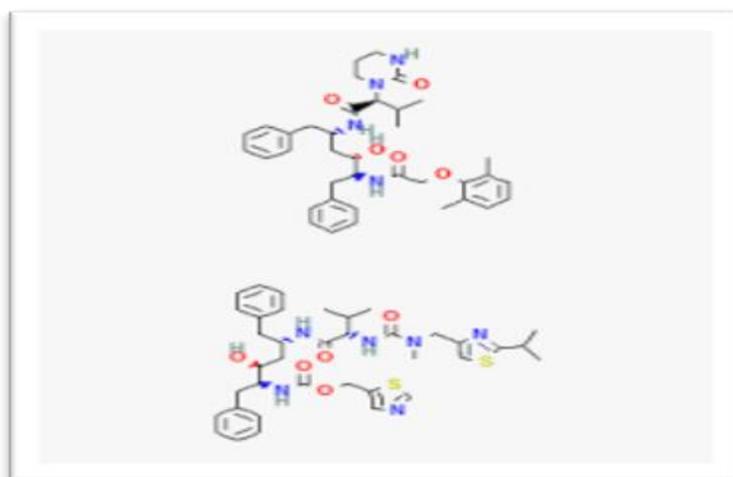


Figure 2.9 : Structures de Lopinavir et Ritonavir

Il a été mis en évidence que l'association lopinavir / ritonavir avec ribavirine comparé à la ribavirine seule, pourrait être utilisé en prophylaxie contre le SRAS.

B. Remdesivir

L'agence américaine des produits alimentaires et médicamenteux (FDA) a accordé une autorisation d'utilisation urgente du médicament antiviral REMDESIVIR afin de traiter COVID-19 (FDA, 2020b). Selon les tests cliniques effectués pour le SARS-CoV-2, ce médicament améliore les chances de survie des patients sous oxygénothérapie et réduit leur temps d'hospitalisation (ECDC, 2020c). Le REMDESIVIR est très efficace pour arrêter le mécanisme de réplication du SARS-CoV-2 (Microbe Notes, 2020c). Ses effets indésirables sont l'élévation des taux des enzymes hépatiques, les complications gastro-intestinales, les éruptions cutanées, l'insuffisance rénale et l'hypotension (OMS, 2020i).

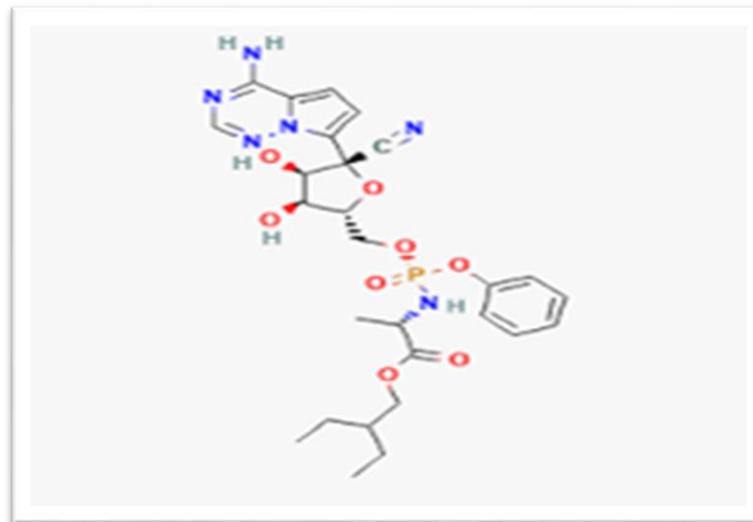


Figure 2.10 : Structures de REMDESIVIR

C. Favipiravir

Était l'un des médicaments expérimentaux à l'essai pour COVID-19 (Microbe Notes, 2020d). Des doses élevées de ce médicament peuvent empêcher les cellules humaines d'être infectées par le SARS-CoV-2 (GoodRx, 2020). Le mécanisme sous-jacent du FAVIPIRAVIR est la métabolisation de sa forme active qui inhibe l'action de RdRp en arrêtant sa transcription et sa réplication (Microbe Notes, 2020d).⁴¹

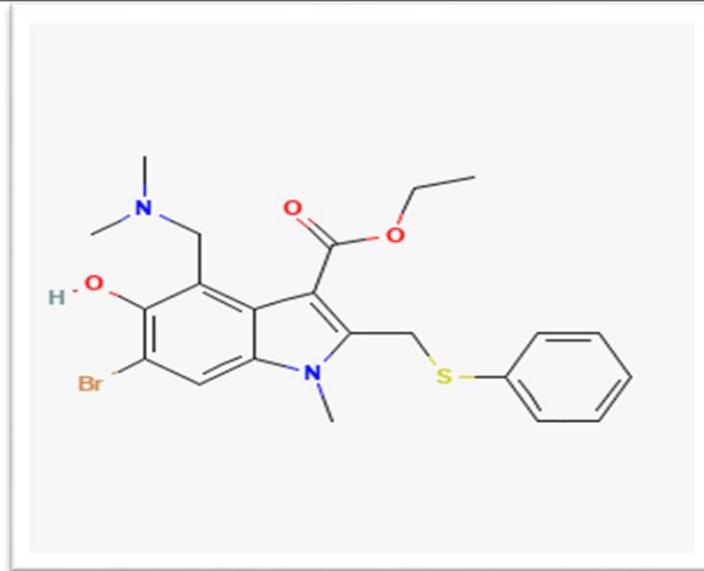


Figure 2.11 : Structures de FAVIPIRAVIR

D. Umifenovir

Approuvé pour la prophylaxie et le traitement de la grippe et d'autres infections virales respiratoire. ⁴² L'UMIFENOVIR a également montré une activité inhibitrice contre les virus des hépatites B et C. In vitro, il semble que l'UMIFENOVIR a une activité antivirale efficace contre le pathogène du SRAS, et que le Mésylate D'arbidol (dérivé de l'UMIFENOVIR) avait un effet presque cinq fois supérieur à celui de l'UMIFENOVIR pour réduire la reproduction du SRAS dans les cellules. ⁴³

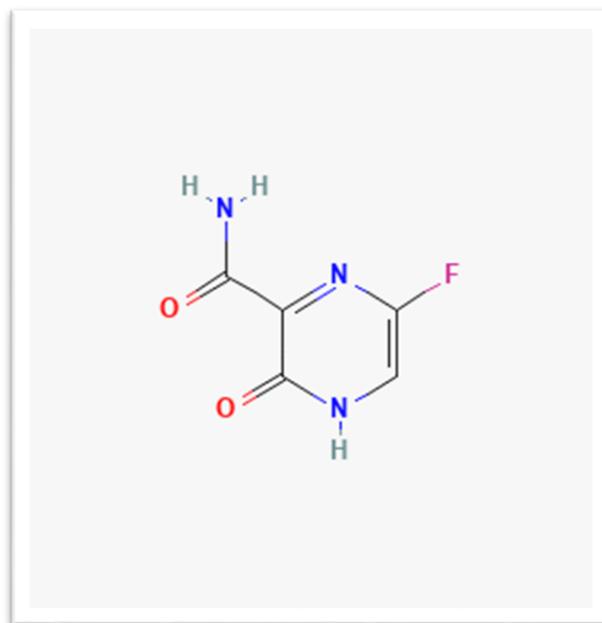


Figure 2.12 : Structures de l'UMIFENOVIR

5.2. Les thérapies adjuvantes

5.2.1. La thérapie antibactérienne

Une co-infection ou une infection croisée d'agents pathogènes bactériens comme le staphylococcus aureus (*S. aureus*), peut être observée chez les patients atteints de pneumonie virale grave, dans ces cas la Société Américaine des Maladies infectieuses (IDSA) en 2018 a recommandé l'administration d'antibiotique en plus du traitement antiviral. Alors un traitement antibiotique est recommandé dans le traitement des patients COVID-19.⁴⁴

A. L'azithromycine

L'azithromycine (AZM) est un antibiotique de la famille des macrolides connu pour ses propriétés anti-inflammatoires et immunomodulatrices. Les macrolides sont connus pour leur capacité à réguler la baisse des réponses inflammatoires et réduire la production excessive de cytokines pro-inflammatoires associée aux infections virales respiratoires.^{45,46} et inhibant l'activation des neutrophiles.

C'est encore une nouvelle fois l'équipe marseillaise de Didier Raoult qui est à l'origine de données suggérant une efficacité clinique et virologique (diminution de la charge virale) grâce à l'association de l'hydroxychloroquine et de l'azithromycine.⁴⁷

Par ailleurs, il s'agit d'une molécule qui allonge également l'espace QT chez des patients à risque (âge, antécédents cardiaques, dyslipidémies...) nécessitant une surveillance accrue lors de son administration

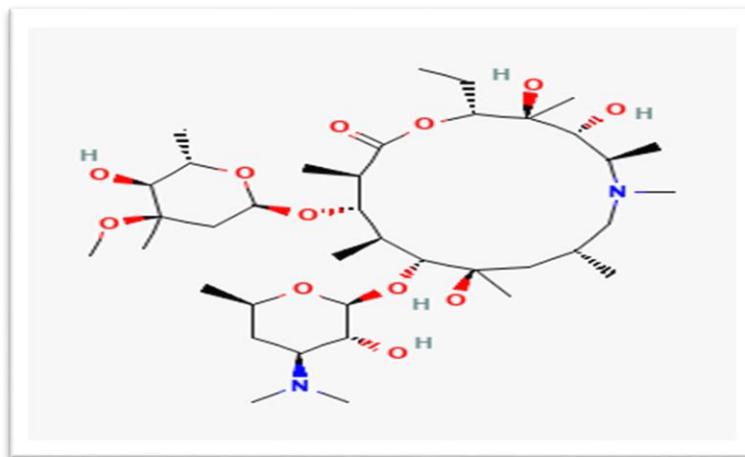


Figure 2.13 : Structures de l'azithromycine.

B. Les immunomodulateurs

Diverses molécules ont été proposées afin de moduler les réponses immunitaires des patients atteints de la Covid-19. On note que des troubles immunitaires ont été rapportés chez les patients atteints de Covid-19. Et ceci vu l'activation des cellules immunitaires suite à l'infection virale, entraînant ainsi une augmentation des cytokines et une aggravation de l'état du patient.
48,49

Les agents immunomodulateurs peuvent être aussi importants que la réplication virale pour prévenir la progression vers un dysfonctionnement multiorganique (**Figure 2.14**). Il existe plusieurs agents immunomodulateurs spécifiques et non spécifiques, y compris les anticorps monoclonaux neutralisants, les corticostéroïdes et d'autres molécules dans la prise en charge du COVID-19 sévère, et leur impact sur la survie et les symptômes cliniques.

5.2.2. La vitaminothérapie

L'administration complémentaire de nombreuses vitamines (dont la vitamine A, C, D, E) et plusieurs oligo-éléments (Zinc, Sélénium, Cuivre et Fer) s'est avérée pouvoir être un allié efficace dans la prévention et/ou le traitement de la Covid-19. En effet la vitamine C, la vitamine D et le zinc peuvent jouer un rôle essentiel dans le renforcement du système immunitaire des patients COVID-19. La vitamine C joue un rôle dans plusieurs aspects de l'immunité, y compris la croissance, la fonction des cellules immunitaires et la production d'anticorps.

A. Vitamine D

La vitamine D est un précurseur hormonal produit par notre propre corps à l'aide de la lumière solaire qui a un rôle important dans l'immunité adaptative et la différenciation cellulaire, la maturation et la prolifération de plusieurs cellules immunitaires.

Plusieurs auteurs suggèrent que le déficit en vitamine D pourrait être associé à un plus haut taux d'atteintes par le SARS-CoV-2⁵⁰. Ainsi, pour les patients à fort risque de développer la Covid-19 et les patients positifs mais asymptomatiques, il est conseillé de contrôler le statut en vitamine D. Dans le cas où la valeur est inférieure à 20 µg/L, un traitement médicamenteux sera nécessaire avec une dose de charge (25 000 à 50 000 UI/semaine) pendant quelques semaines.

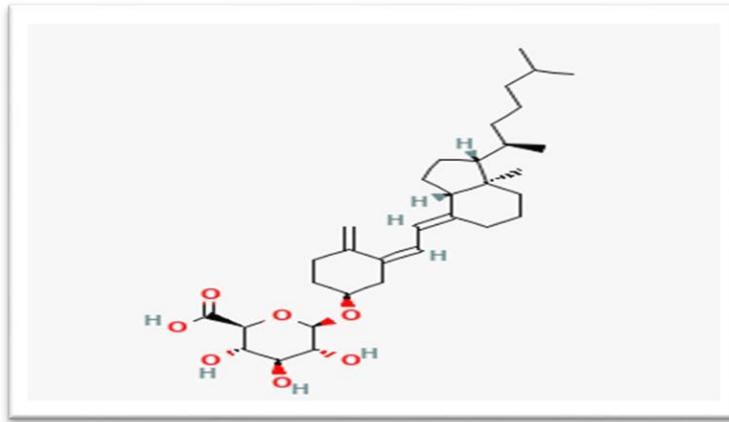


Figure 2.14 : Structure de Vitamine D.

B. Zinc

Le zinc (Zn) est un micronutriment essentiel, avec des concentrations systémiques et intracellulaires strictement réglementées. Cet oligoélément est indispensable dans la transcription du génome grâce à son rôle structural fondamental dans la constitution des protéines.⁵¹ Le Zinc est mêlé à la protection contre la rupture de la barrière muqueuse de voies aériennes supérieures mais aussi dans le maintien de l'intégrité des cellules endothéliales pulmonaires. Il possède une activité antivirale efficace. En effet, il a été démontré que le zinc inhibe l'activité de l'ARN polymérase ARN dépendante (RdRp) du virus de l'hépatite E. Dans le cadre de la Covid-19, les ionophores de zinc bloquent la réplication du coronavirus, en déplaçant les ions magnésium du centre actif. Des similitudes avec d'autres virus comme le SARSCov1 ou le MERS-CoV laissent penser que le Zinc aurait les mêmes propriétés sur le SARS-Cov2 in vivo chez l'homme.⁵²

Des auteurs avaient émis l'hypothèse qu'une supplémentation efficace en zinc pendant le traitement de COVID-19 en association avec la CQ ou l'HCQ pourrait entraîner une inhibition plus efficace de la réplication intracellulaire du SRAS-CoV-2, améliorant ainsi les résultats cliniques des patients COVID-19 traités par CQ ou HCQ .^{45,46}

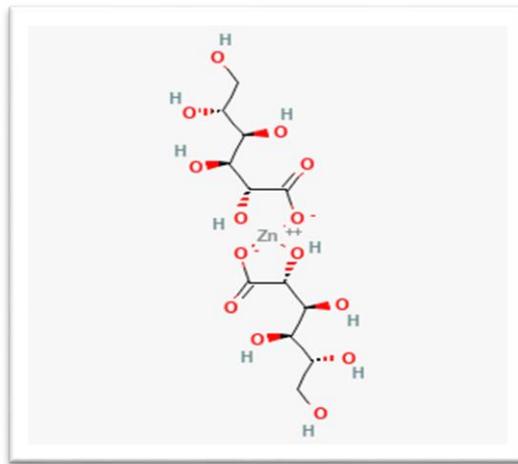


Figure 2. 15 : Complexe de ZINC.

Références :

1. Blancou, J. **compte rendu sur les maladies émergentes.**
2. Zaki,A.M, vanBoheemen,S, Bestebroer,T.M, Osterhaus,A.D, Fouchier,R.A. **Isolation of a novel coronavirus from a man with pneumonia in Saudi – Arabia.** *NEnglJ Med*,(2012);367:1814–2.
3. DeGroot .RJ, Baker. SC, Baric .RS, Brown. CS, Drosten .C, Enjuanes . L, et all. **Middle East Respiratory Syndrome Coronavirus (MERS-CoV): announcement of the Coronavirus Study Group.** *JVirol*(2013); 87:7790–2.
4. J.-M.M. **MERS-CoV: l’hypothèse du chameau est confirmé mais...** *Revue francophone des laboratoires* (Février 2014), N°459.
5. Sallard .E , Halloy .J , Casane . D, Helden, J. v,E.Decroly.**Médecine/sciences,EDPSciences**,(2020),p783-796.
6. Cui. J, Li. F and Shi. Z, **Nature Reviews Microbiology**,(2019).
7. Mousavizadeh L, Ghasemi S. **Genotype and phenotype of COVID-19: their roles in pathogenesis.** *J Microbiol Immunol Infect* 2020. [http://dx.doi.org/10.1016/j.jmii.\(2020\).03.022](http://dx.doi.org/10.1016/j.jmii.(2020).03.022)
8. Mousavizadeh L, Ghasemi S. **Genotype and phenotype of COVID-19: their roles in pathogenesis.** *J Microbiol Immunol Infect*(2020). <http://dx.doi.org/10.1016/j.jmii.2020.03.022>
9. V. Bonnya,A. Maillardb, Mousseaux, L. Placais, Q. **COVID-19 : physiopathologie d’une maladie à plusieurs visages**, (2020.), Richier doi.org/10.1016/j.
10. Grant, O.C.; Montgomery, D.; Ito, K.;Woods, R.J. **Analysis of the SARS-CoV-2 spike protein glycan shield: Implications for immune recognition.** *BioRxiv Prepr. Serv. Biol.*(2020).
11. Neuman. BW, Kiss .G, Kunding .AH, Bhella. D, Baksh .MF, Connelly .S,Droese. B, Klaus .JP, Makino .S, Sawicki .SG, Siddell .SG, Stamou .DG, Wilson. IA, Kuhn. P, Buchmeier .MJ. (2011). **A structural analysis of M protein in coronavirus assembly and morphology.** 174:11–22. <https://doi.org/10.1016/j.jsb.2010.11.021>.
12. Nal. B, Chan. C, Kien .F, Siu. L, Tse .J, Chu .K, Kam. J, Staropoli. I, Crescenzo-Chaigne. B, Escriou. N, van der Werf S, Yuen K-Y, Altmeyer R. **Differential maturation and subcellular localization of severe acute respiratory syndrome coronavirus surface proteins** , (2005),S, M and E. *J Gen Virol* 86:1423–1434. <https://doi.org/10.1099/vir.0.80671-0>
13. Arndt .AL, Larson .BJ, Hogue. BG. **A conserved domain in the coronavirus membrane protein tail is important for virus assembly** (2010).. *J Virol* 84:11418 – 11428. <https://doi.org/10.1128/JVI.01131-10>
14. Schoeman .D, Fielding. **BCCoronavirus envelope protein: current knowledge** , (2019), *Virol J* 16:69.
15. Nieto-Torres. JL, DeDiego. ML, Verdiá-Báguena. C, Jimenez-Guardeño .JM, ReglaNava. JA, Fernandez-Delgado .R, Castaño-Rodriguez. C, Alcaraz. A,Torres. J, Aguilera. VM, Enjuanes .L**Severe acute respiratory syndrome coronavirus envelope protein ion channel activity promotes virus fitness and pathogenesis** , (2014).. *PLoS Pathog* 10:e1004077. <https://doi.org/10.1371/journal.ppat.1004077>

16. Pervushin. K, Tan. E, Parthasarathy. K, Lin .X, Jiang. FL, Yu. D, Vararattanavech .A, Soong .TW, Liu. DX, Torres .J **Structure and inhibition of the SARS coronavirus envelope protein ion channel.** ,(2009). <https://doi.org/10.1371/journal.ppat.1000511>
17. DeDiego. ML, Alvarez .E, Almazán .F, Rejas. MT, Lamirand .E, Roberts. A, Shieh. WJ, Zaki. SR, Subbarao .K, Enjuanes. LA **A severe acute respiratory syndrome corona virus that lacks the E gene is attenuated in vitro and in vivo.** ,(2007). *J Virol* 81:1701–1713. <https://doi.org/10.1128/JVI.01467-06>).
18. Chang. C-K, Sue .S-C, Yu T-H, Hsieh. C-M, Tsai. C-K, Chiang. Y-C, Lee. S-J, Hsiao .HH, Wu. W-J, Chang .W-L, Lin. C-H, Huang .T-H. **Modular organization of SARS coronavirus nucleocapsid protein** ,(2006).. *J Biomed Sci* 13:59 –72. <https://doi.org/10.1007/s11373-005-9035-9>,
19. Sheikh. A, Al-Taher .A, Al-Nazawi .M, Al-Mubarak .AI, Kandeel. M **Analysis of preferred codon usage in the coronavirus N genes and their implications for genome evolution and vaccine design.** ,(2020), *J Virol Methods* 277:113806. <https://doi.org/10.1016/j.jviromet.2019.113806>
20. Bonny. V, Maillard. A, Mousseaux. C, Plaçais. L, & Richier. Q , **COVID- 19: Pathogenesis of a multi-faceted disease. La Revue de Médecine Interne.** (2020), doi:10.1016/j.revmed.(2020).05.003
21. Mohd Gazzali. A, et al. **Molecules**,(2020),1;25(17):3980. doi:10.3390/molecules25173980 .
22. Khailany. RA, Safdar. M, Ozaslan; M. **Genomic characterization of a novel SARSCoV-2.** (2020);19:100682.
23. Mousavizadeh. L, Ghasemi .S. **Genotype and phenotype of COVID-19: their roles in pathogenesis. J Microbiol Immunol Infect** , (2020). <http://dx.doi.org/10.1016/j.jmii.2020.03.022>
24. Jungreis. I, Sealfon .R, Kellis. M. ,(2020.) <http://dx.doi.org/https://doi.org/10.1101/2020.06.02.130955>
25. Chen. WH, Hotez. PJ, Bottazzi. ME. **Potential for developing a SARS-CoV receptor-binding domain (RBD) recombinant protein as a heterologous human vaccine against coronavirus infectious disease (COVID)-19. Hum Vaccin Immunother**,(2020);16(6):1239-42. <http://dx.doi.org/10.1080/21645515.2020.1740560>.
26. Kim. D, Lee. JY, Yang .JS, Kim. JW, Kim .VN, Chang. H. **The architecture of SARS-CoV 2 transcriptome. bioRxiv** 2 ,(2020).
27. Wölfel .R., Corman V.M., Guggemos. W., Seilmaier .M., Zange. S., Müller M.A. **Virological assessment of hospitalized patients with COVID-2019. Nature**,(2020), DOI : 10.1038/s41586-020-2196-x.
28. Dong. L., Tian .J., He. S., Zhu. C., Wang .J., Liu .C. **Possible vertical transmission of SARS-CoV-2 from an infected mother to her newborn. JAMA**,(2020) :e204621. DOI : 10.1001/jama.(2020).4621.
29. Zhou. P., Yang. X.L, Wang .X.G, Hu. B, Zhang. L, Zhang .W, **A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature**,(2020);579(7798):270–273.
30. Wilde. A.H., Snijder .E.J, Kikkert. M, van Hemert M.J. **Host factors in coronavirus**

- replication. *Curr Top Microbiol Immunol*, (2018);419:1–42.
31. Hoffmann. M, Kleine-Weber .H, Schroeder .S, Krüger .N, Herrler .T, Erichsen. S, **SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor.** *Cell*,(2020) .
 32. Coutard. B., Valle. C, de Lamballerie. X, Canard. B., Seidah .N.G., Decroly .E, **The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade.** *Antiviral Res*,(2020);176:104742.
 33. Walls .A.C, Park .Y.-J, Tortorici. M.A, Wall. A, McGuire .A., Veesler D. **Structure, function, and antigenicity of the SARS-CoV-2 Spike glycoprotein.** *Cell*,(2020);181(2)
 34. Wit. E, van Doremalen. N, Falzarano .D, Munster V.J. **SARS and MERS: recent insights into emerging coronaviruses.** *Nat Rev Microbiol*, (2016);14:523–534.
 35. <https://www.who.int/fr/news-room/q-a-detail/sars-cov-2-evolution>. (Consulté le 18 Juin 2021). n.d.
 36. Santé publique France. Covid-19 : une enquête pour suivre l'évolution des comportements et de la santé mentale pendant l'épidémie.
 37. Yuefei J ., Haiyan .Y, Wangquan. J , Weidong . W, Shuaiyin. Ch , Weiguo. Z and Guangcai. D. **Viruses**,(2020).
 38. Samer. C, Curtin .F, Ing-Lorenzini. KR, Vetter .P, **Département de Médecine Aiguë, Service de Pharmacologie et Toxicologie Cliniques**, (2020).
 39. Traore. B, Tsoumbou Bakana. G, Nani. S, Hassoune. S , **Revue Marocaine de santé publique**, (2020).
 40. Fuhier. J, Cova. F, **Judgment and decision Making**. Vol. 15, (6 November 2020), p 889-908.
 41. Bursaux. É, **Médecine/sciences**, (1996).
 42. Benchikh. A, Benzaaza .H, **Mémoire de master. Université Mohamed El bachir El Ibrahimi B.B.A**, (2020), p 13
 43. Blaising.J, Polyak, S. J. & Pecheur, E. **I. Arbidol as a broad-spectrum antiviral: an update.** *Antivir.* ,(2014), Res. 107, 84–94.
 44. Uyeki TM, Bernstein HH, Bradley JS, Englund JA, File TM, Fry AM, et al. **Clinical Practice Guidelines by the Infectious Diseases Society of America:**(2018) Update on Diagnosis, Treatment, Chemoprophylaxis, and Institutional Outbreak Management of Seasonal Influenza. *Clin Infect Dis Off Publ Infect Dis Soc Am.* (2019) 05;68(6):e1–47 .
 45. Kanoh S, Rubin BK. **Mechanisms of action and clinical application of macrolides as immunomodulatory medications.** *Clin Microbiol Rev.* 2010 Jul;23(3):590– 615.
 46. Zarogoulidis P, Papanas N, Kioumis I, Chatzaki E, Maltezos E, Zarogoulidis K. **Macrolides: from in vitro anti-inflammatory and immunomodulatory properties to clinical practice in respiratory diseases.** *Eur J Clin Pharmacol.* (2012)May;68(5):479–503.
 47. Gao. J, Tian. Z, Yang .X. **Breakthrough: Chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies.** *BioScience Trends*,(2020);14(1):72-73.
 48. Beigel JH, Nam HH, Adams PL, Krafft A, Ince WL, ElKamary SS, et al. **Advances in**

- respiratory virus therapeutics - A meeting report from the 6th isirv Antiviral Group conference. Antiviral Res.**(2019)Apr 8;167:45–67.
49. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. **Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study.** The Lancet. (2020) Feb 15;395(10223):507–13
50. D’Avolio et al, 2020 ; Hastie et al, 2020 ; Ilie, Stefanescu & Smith, 2020; Meltzer et al, (2020).
51. conseil Supérieur de la Santé. Vitamine D, **Zinc et Covid-19**,(2021), Avis n° 9620. **Service public Fédéral de la Santé publique de la Sécurité de la Chaîne alimentaire et de l’Environnement.**
52. Wessels, Rolles & Rink, (2020) ; Arentz et al, 2020 ; Skalny et. al, (2020) ; Razaque, 2020 ; Pormohammad et al, (2020) ; Mayor-Ibarguren et al,(2020) ; Kumar et al, 2020

PARTIE

COMPUTATIONNELLE



CHAPITRE 03 :
MATÉRIELS ET MÉTHODES



1. Docking moléculaire

L'interaction entre une protéine et son substrat est la première étape de la plupart des réactions biologiques. Comprendre son mode de fonctionnement et définir quels sont les résidus mis en jeu, est donc primordial pour pouvoir expliquer les mécanismes qui influent sur l'affinité entre les molécules. De même, la découverte de nouvelles molécules activant ou inhibant l'activité biologique d'une protéine ne peut se faire qu'en prédisant leur affinité respective. C'est dans ce but que des techniques de modélisations moléculaires, regroupées sous le nom "Docking", ont été développées.

Dans le cadre de la recherche de nouvelles molécules naturelles efficaces contre le SARS-CoV-2, et afin de trouver de nouveaux inhibiteurs de ce virus, nous avons étudié par Docking moléculaire l'inhibition d'une série de molécules d'origine naturelle et médicaments avec deux récepteurs du SARS-CoV-2, : ACE2 et M^{Pro}. L'enzyme de conversion de l'angiotensine-2 (ACE2) est un récepteur fonctionnel du SARSCoV-2 qui peut faciliter l'entrée du virus dans les cellules de l'hôte en se liant à la protéine de pointe (S) du virus. La protéase virale principale (M^{Pro}) est essentielle à la réplication des coronavirus.

1.1. Matériels utilisés

1.1.1. Programmes (logiciels)

Plusieurs logiciels ont été utilisés pour la réalisation de ce travail :

- ❖ **Logiciel Gaussian¹** : est un logiciel de chimie numérique, créé et sorti en 1970 par John Pople et collaborateurs (Gaussian 70). Le nom vient de l'utilisation des fonctions gaussiennes pour représenter les orbitales atomiques (OA) pour accélérer le calcul par rapport aux logiciels utilisant des orbitales de Slater. Ceci a facilité le développement de la chimie numérique, en particulier les méthodes ab initio comme Hartree-Fock. Ce programme mondialement reconnu par ses capacités de premier plan pour la modélisation électronique de structures chimiques permet de pronostiquer des énergies, structures moléculaires et fréquence de vibration de système moléculaire complexes et d'anticiper leurs autres propriétés chimiques.
- ❖ **Logiciel de Docking moléculaire : AutoDock Vina²** : est un programme de Docking dérivé d'AutoDock comme son nom l'indique, qui a également été développé par le groupe de

Arthur J. Olson au “Scripps Research Institute” (<http://autodock.scripps.edu/>). Ce logiciel d’amarrage moléculaire très répandu et connu par sa fiabilité et son pouvoir de reproduire les ligands déjà complexés et utilisé dans le docking moléculaire. Autodock Vina utilise un algorithme de recherche d’optimisation globale stochastique de type Iterated Local.³ La fonction de scoring utilisée par Autodock Vina permet de faire la somme des contributions intramoléculaires et intermoléculaires du système d’une manière inspirée par le programme Xscore⁴ qui utilise une fonction de score empirique. Elle combine ainsi les avantages des fonctions de score empiriques et knowledge-based.

- ❖ **Logiciel Discovery : Le visualiseur Discovery Studio**⁵ est un visualiseur gratuit qui peut être utilisé pour ouvrir des données générées par d’autres logiciels de la gamme de produits Discovery Studio. Il est conçu pour offrir un environnement interactif permettant de visualiser et d’éditer des structures moléculaires, des séquences, des données de réflexion des rayons X, des scripts et d’autres données.

1.1.2. Banques de données (chimiothèques)

A. PDB (Protein Data Bank)

La banque de données protéiques (Protein Data Bank ou PDB) est une collection mondiale de données sur la structure tridimensionnelle (Structure 3D) de macromolécules biologiques : protéines essentielles et acides nucléiques. Ces structures sont déterminées par cristallographie aux rayons X ou par spectroscopie RMN. Cette base de données est accessible gratuitement à l’adresse (<http://www.rcsb.org/pdb/>).

B. PUBCHEM (<https://pubchem.ncbi.nlm.nih.gov/>).

C’est une banque de données américaine de molécules chimiques gérée par le National Center for Biotechnology Information (NCBI), branche de la Bibliothèque nationale de médecine des États-Unis, sous l’autorité de la National Institutes of Health (NCI).

1.2. Méthodes

1.2.1. Docking moléculaire

Le Docking moléculaire est un procédé très utile dans la découverte de médicaments. Il permet de prédire la conformation (position et orientation relatives d’une petite molécule par rapport à une autre comme cible macromoléculaire) la plus favorable de deux molécules en

interaction pour avoir le complexe le plus stable, dans lequel différentes approches sont combinées pour étudier les modes d'interaction entre ces deux molécules. La connaissance de cette conformation préférentielle permet par la suite l'estimation de la force d'association (affinité de liaison) entre ces deux molécules.

Ce travail consiste à étudier les interactions entre série de molécules avec l'enzyme ou la cible choisis comme inhibiteurs du SARS-CoV-2 par le Docking moléculaire en utilisant le logiciel AutoDock Vina.

1.2.2. Reproduction de la structure cristalline

Souvent, les performances d'un programme de Docking moléculaire sont évaluées en termes de capacité à reproduire le mieux possible des complexes expérimentaux. Autrement dit, il faut que l'évaluation quantitative de la déformation spatiale entre la pose du ligand calculée par AutoDock Vina et la conformation dans le complexe expérimental soit la plus petite possible. Pour cela, on utilise une grandeur scalaire : RMSD (Root-Mean-Square-Deviation).

Le RMSD correspond à la moyenne de la déviation de chacun des atomes du ligand (ligand docké) par rapport à ceux de la molécule d'origine (le ligand de référence).

Le positionnement, c'est-à-dire l'identification correcte du site de liaison sur la protéine, l'orientation et la conformation du ligand influent sur la valeur du RMSD. La prédiction est acceptable si sa valeur ne dépasse pas 2 angströms.

1.2.3. Choix du complexe

Parmi plusieurs protéines proposées par la PDB, nous avons choisi les codes 6LU7 et 6LZG définit comme suit :

- **6LU7** : Ce code représente la structure 3D de M^{-pro} constituée d'une seule chaîne (A), dans laquelle se trouve un seul ligand, avec une résolution de 2.16 Å.
- **6LZG** : ce code représentant la structure 3D d'ACE2 constituée de deux chaînes (A et B) le ligand naturel se trouve dans la chaîne (B) avec une résolution de 2.5 Å.

1.2.4. Choix du logiciel de Docking moléculaire

Le logiciel « AutoDock Vina » a été sélectionné pour les calculs de Docking moléculaire de par sa disponibilité au laboratoire, sa grande modularité des paramètres (Docking, scoring) et sa précision de Docking.

1.2.5. Préparation des molécules au Docking moléculaire

Dans ce travail nous avons suivi les étapes suivantes :

A. Préparation des récepteurs

Cette étape préliminaire du Docking moléculaire, consiste au téléchargement des structures chimiques. La plupart des structures sont disponibles via la «Protein Data Bank » et téléchargés sous forme PDB (<http://www.pdb.org>).

Le récepteur est ensuite préparé en définissant les charges partielles de ses atomes par l'intermédiaire du sous-programme AutoTools qui attribue des charges de type "Kollman" calculées avec le champ de forces empirique MMFF.

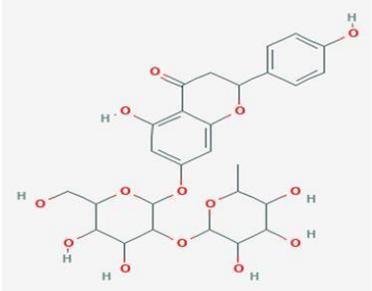
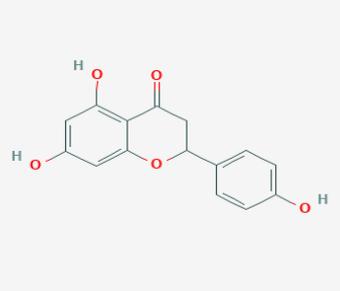
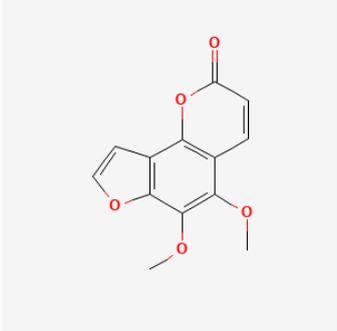
Avant le Docking, chaque enzyme est séparée de son ligand co-cristallisé pour obtenir une cible avec un site actif libre. Les atomes d'hydrogène sont rajoutés et, les structures sont enregistrées en format PDBQT.

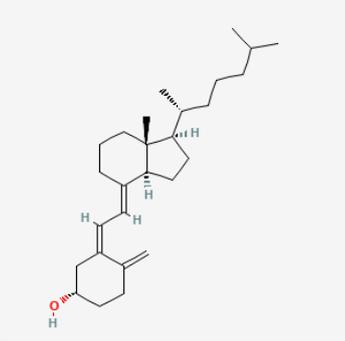
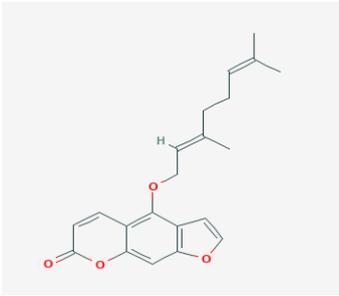
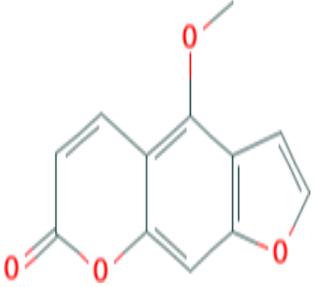
B. Préparations des inhibiteurs

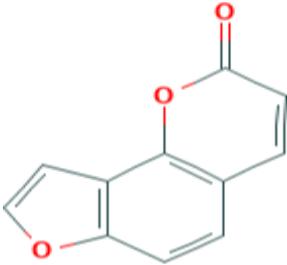
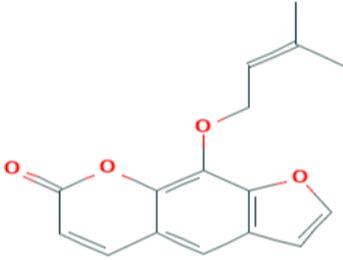
Les ligands étudiés ont été téléchargés via la chimiothèque PUBCHEM (<https://pubchem.ncbi.nlm.nih.gov/>). Dans ce travail nous avons étudié 28 inhibiteurs d'origine naturelle et des médicaments.

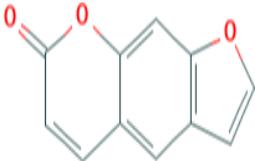
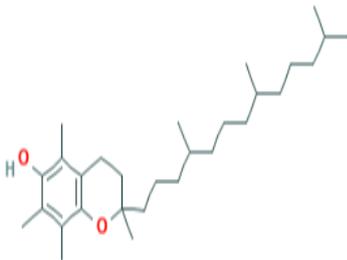
Puis nous avons utilisés le logiciel GUASSIAN 09W, pour l'optimisation des ligands par la méthode semi empirique (PM6) afin de trouver les conformations les plus stables. Les molécules ainsi obtenues sont enregistrées en format pdb. La phase finale de préparation du ligand a été réalisée par le sous-programme vina. Ce dernier a pour rôle de contrôler les mouvements du ligand, en définissant les atomes aromatiques, les liaisons flexibles et les angles de torsion. La molécule finale est enregistrée en format pdbqt avant son utilisation par AutoDock Vina.

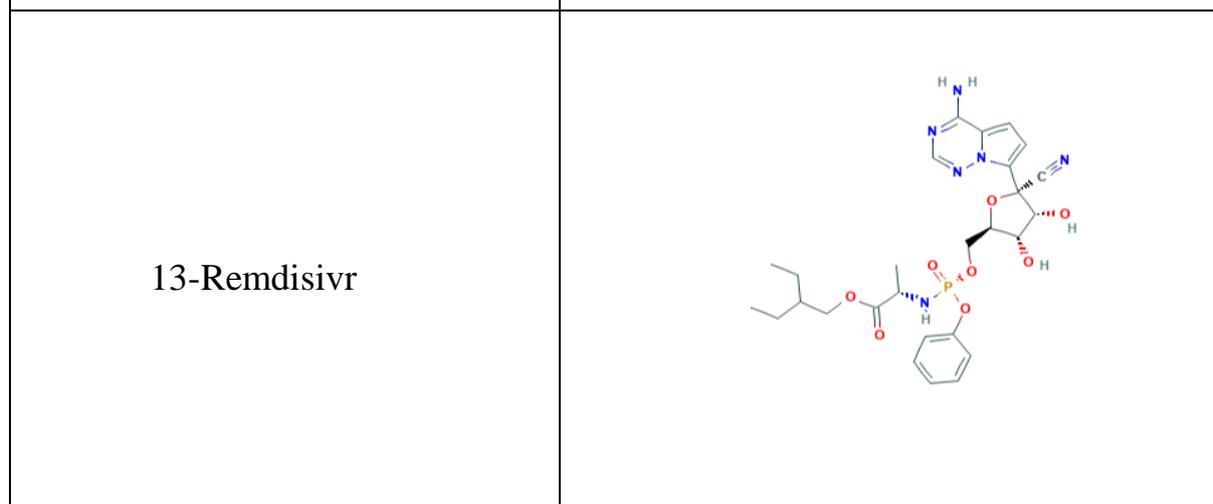
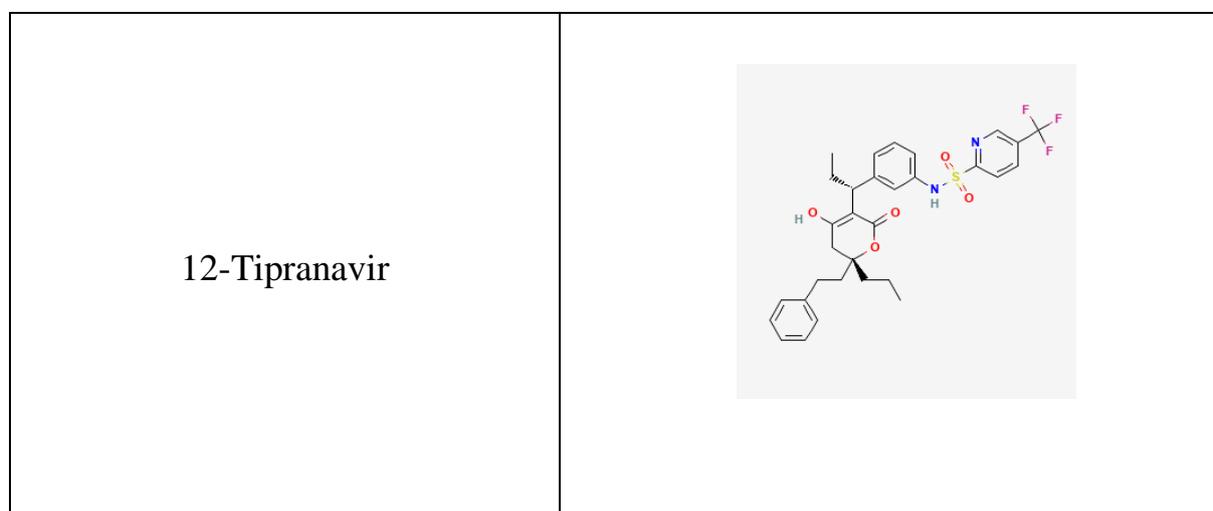
Tableau 3-1 : la structure 2D des inhibiteurs

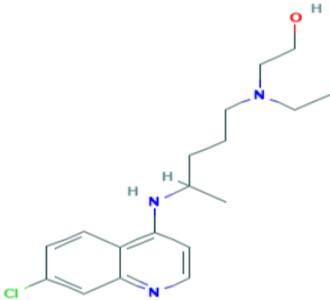
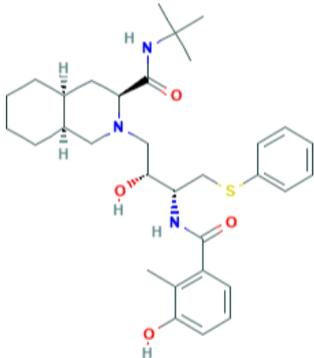
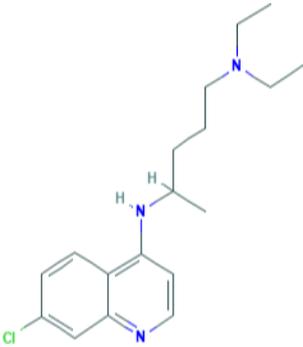
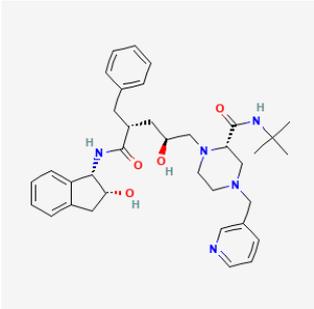
Ligand	2D
1-Naringine	 <p>The chemical structure of 1-Naringine is a complex polyphenolic compound. It features a central flavanone core (naringenin) with a 4-hydroxyphenyl group at the C-7 position. This core is linked via an ether bridge to a quaternary ammonium cation, specifically a trimethylammonium group attached to a 1,2,3,4-tetrahydroquinoline ring system. The structure is highly polar due to the presence of multiple hydroxyl groups and the charged nitrogen atom.</p>
2-Naringinine	 <p>The chemical structure of 2-Naringinine is a flavanone derivative. It consists of a flavanone core with a 4-hydroxyphenyl group at the C-7 position and a 4-hydroxyphenyl group at the C-2 position. The structure is highly polar due to the presence of multiple hydroxyl groups.</p>
3-Pimpinelline	 <p>The chemical structure of 3-Pimpinelline is a complex polyphenolic compound. It features a central flavanone core with a furfuryl group at the C-7 position and a methoxy group at the C-3 position. The structure is highly polar due to the presence of multiple hydroxyl groups and the furfuryl group.</p>

<p>4-Cholecalciferol</p>	 <p>The image shows the chemical structure of 4-Cholecalciferol, a complex steroid molecule. It features a four-ring steroid nucleus with a hydroxyl group at C-3, a double bond at C-5, and a long side chain at C-13 that includes a double bond and a methyl group. Stereochemistry is indicated with wedges and dashes.</p>
<p>5-Bergamottine</p>	 <p>The image shows the chemical structure of 5-Bergamottine, a furanocoumarin. It consists of a coumarin core with a furan ring fused at the 5-position and a prenyl side chain at the 7-position. The oxygen atoms in the furan and coumarin rings are highlighted in red.</p>
<p>6-Bergaptene</p>	 <p>The image shows the chemical structure of 6-Bergaptene, a furanocoumarin. It features a coumarin core with a furan ring fused at the 6-position and a methoxy group at the 7-position. The oxygen atoms in the furan and coumarin rings are highlighted in red.</p>

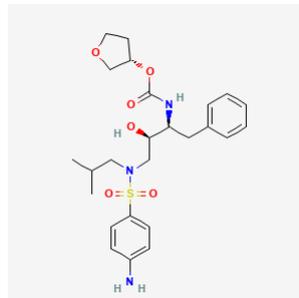
7-Angélicine	 <p>The chemical structure of 7-Angelicin is a coumarin derivative. It features a benzopyrone core with a furfuryl group at the 7-position and an angeloyl group at the 2-position. The furfuryl group is a five-membered furan ring fused to a benzene ring. The angeloyl group is a propenoic acid derivative with a double bond between the alpha and beta carbons and a methyl group at the gamma position.</p>
8-Impératorine	 <p>The chemical structure of 8-Impératorin is a coumarin derivative. It features a benzopyrone core with a furfuryl group at the 8-position and an isopentenyl group at the 3-position. The furfuryl group is a five-membered furan ring fused to a benzene ring. The isopentenyl group is a prenyl chain with a double bond between the alpha and beta carbons and a methyl group at the gamma position.</p>

9-psoralene	 <p>The chemical structure of Psoralen is a coumarin derivative. It features a benzopyrone core with furfuryl groups at the 6 and 7 positions. The furfuryl groups are five-membered furan rings fused to benzene rings.</p>
10- α -Tocophérol (vitamine E)	 <p>The chemical structure of α-Tocopherol is a vitamin E derivative. It features a chromanol ring with a hydroxyl group at the 6-position and a phytyl side chain at the 2-position. The phytyl side chain is a long, branched hydrocarbon chain with methyl groups at the 2, 6, and 10 positions.</p>

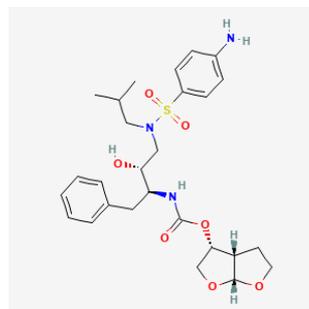


14-HydroxyChloroquine	 <p>The structure shows a quinoline ring system with a chlorine atom at the 7-position. At the 4-position, there is a nitrogen atom bonded to a hydrogen atom and a side chain. The side chain consists of a methylene group, a methine group with a methyl group, and a four-carbon chain ending in a secondary amine with an ethyl group and a hydroxyl group.</p>
15-Nelfinavir	 <p>The structure features a piperidine ring with a nitrogen atom bonded to a hydrogen atom and a side chain. The side chain includes a methylene group, a methine group with a methyl group, and a chiral center with a hydroxyl group. This is followed by a methylene group, a sulfur atom bonded to a phenyl ring, and a carbonyl group attached to a 2,4-dihydroxyphenyl ring.</p>
16-Chloroquine	 <p>The structure shows a quinoline ring system with a chlorine atom at the 7-position. At the 4-position, there is a nitrogen atom bonded to a hydrogen atom and a side chain. The side chain consists of a methylene group, a methine group with a methyl group, and a four-carbon chain ending in a secondary amine with two ethyl groups.</p>
17-Indinavir	 <p>The structure is a complex molecule with a central piperidine ring. It features a benzimidazole ring system, a phenyl ring, a hydroxyl group, a carbonyl group, and a side chain with a methylene group, a methine group with a methyl group, and a nitrogen atom bonded to a hydrogen atom and a piperidine ring. The piperidine ring is further substituted with a methyl group and a pyridine ring.</p>

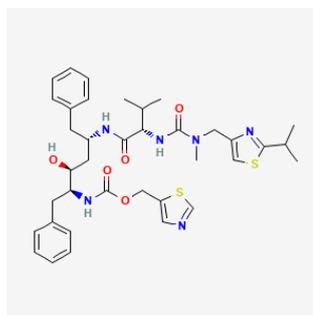
18-Amprenavir



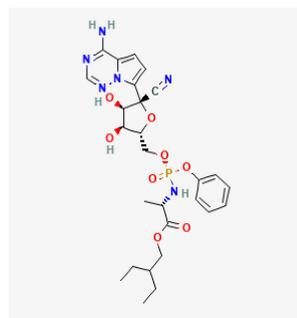
19-Darunavir

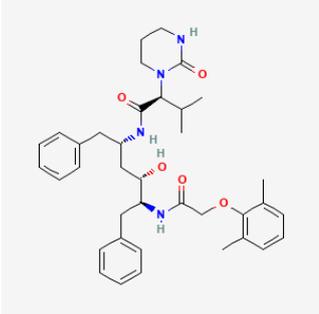
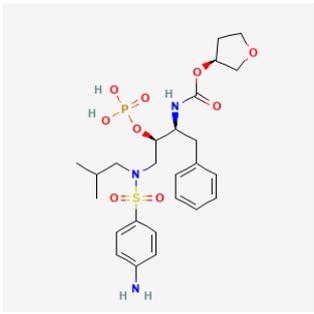
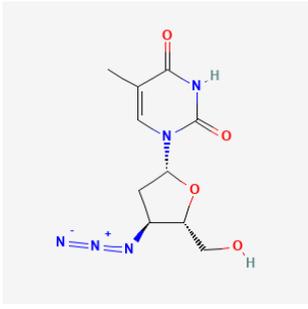
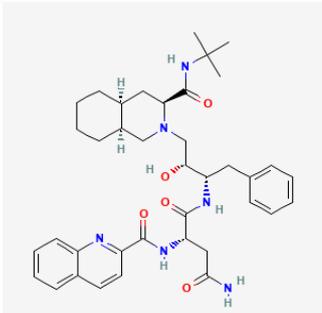


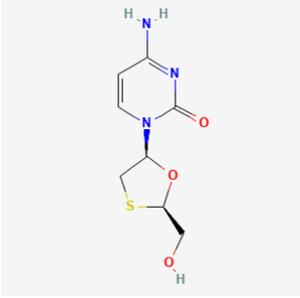
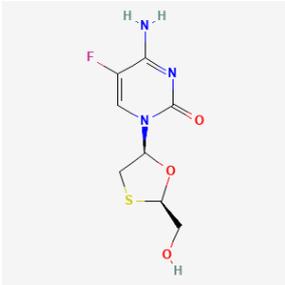
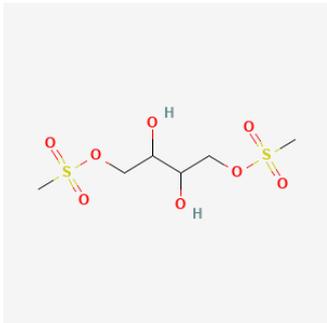
20-Ritonavir



21-Remdesivir



22-Lopinavir	 <p>The chemical structure of Lopinavir is a complex molecule. It features a central piperidine ring with a carbonyl group and a methyl group. This is connected to a chiral center that is further linked to a benzyl group and a side chain containing another chiral center with a hydroxyl group and a benzyl group. The side chain also includes a carbonyl group and a methoxy group attached to a benzene ring.</p>
23-Fosamprenavir	 <p>The chemical structure of Fosamprenavir consists of a central piperidine ring substituted with a methyl group and a sulfonamide group. The sulfonamide group is linked to a side chain that includes a benzyl group, a hydroxyl group, and a phosphate group. The phosphate group is further connected to a cyclopentane ring with an oxygen atom, representing the fosfate ester.</p>
24- zidovudine	 <p>The chemical structure of zidovudine is a nucleoside. It features a pyrimidine ring with a methyl group at the 5-position and a carbonyl group at the 2-position. The nitrogen at the 1-position is linked to a ribose sugar ring. The ribose ring has a hydroxyl group at the 3-position and a diazotrimethylammonium group at the 4-position.</p>
25-Saquinavir	 <p>The chemical structure of Saquinavir is a complex molecule. It features a central piperidine ring with a carbonyl group and a methyl group. This is connected to a side chain that includes a benzyl group, a hydroxyl group, and a carbonyl group. The carbonyl group is further linked to a quinoline ring system.</p>

26-Lamivudine	 <p>The image shows the chemical structure of Lamivudine. It consists of a pyrimidine ring with a carbonyl group at the 2-position and an amino group at the 4-position. The pyrimidine ring is attached to a thiazolidine ring, which is further substituted with a hydroxymethyl group.</p>
27- Emtricitabine	 <p>The image shows the chemical structure of Emtricitabine. It is similar to Lamivudine but has a fluorine atom at the 5-position of the pyrimidine ring.</p>
28- DHB	 <p>The image shows the chemical structure of DHB (Disodium Hydrogen Phosphate). It consists of a central carbon atom bonded to two hydroxyl groups and two phosphate groups.</p>

1.3. Criblage Par Docking

Tous les ligands de notre série ont été criblé par Docking moléculaire. Le « score » des poses obtenues pour chaque ligand mène à un classement des meilleures poses basées sur leurs énergies d'interaction.

1.3.1. Protocole du Docking moléculaire

Le Docking moléculaire a été réalisé à l'aide du logiciel AutoDock Vina

1.3.2. Paramètres de la grille de calculs

Le programme AutoDock Vina utilise dans les calculs de Docking moléculaire une grille tridimensionnelle, permettant d'augmenter la vitesse d'évaluation de l'énergie d'interaction ligand/récepteur. Cette grille englobe largement le site actif de la protéine étudiée et permet la rotation libre du ligand dans ce site. Le centre de cette boîte est déterminé par les coordonnées X et Y et Z. La boîte est alors centrée sur le ligand et ses dimensions sont proportionnelles à la taille de tous les ligands étudiés. Les paramètres par défaut ont été utilisés pour tous les calculs.

Tableau 3-2: La boîte GRID BOX des récepteurs

Récepteurs	X	Y	Z	Dimension
6LU7	13.069	9.740	68.490	60 x 60 x 60
6LZG	-32.201	28.937	6.225	54x54x54

1.3.3. Algorithme de recherche

Nous avons choisi, algorithme de recherche d'optimisation globale stochastique de type Iterated Local.

A. Fonctions de score

Lors de la simulation du Docking plusieurs poses candidates peuvent être obtenues pour chaque ligand, nous avons sélectionné les meilleures poses avec les énergies les plus basses selon la fonction de score basée sur une grille « AutoDock Score ».

2. QSAR (Quantitative structure Activity Relationship) :

Les relations structure-activité quantitatives (QSAR) visent la recherche d'une relation entre les molécules de grandes bases de données et leurs paramètres biologiques pour développer un modèle statistique. L'étude QSAR se repose sur un principe qui affirme que la différence de

propriétés structurelles est responsable de variations observées dans la réponse biologique des molécules.

La qualité d'un modèle est déterminée par son aptitude à l'ajustement et à la prédiction. L'un des objectifs importants de la modélisation QSAR est la prédiction de l'activité biologique de nouveaux composés chimiques qui entrent dans le domaine d'applicabilité du modèle développé.

Dans ce travail, nous intéressons à l'étude de la Relation Structure-Activité d'une série des (dérivés de 2-amino quinazolin-4(3H)-ones.) utilisée contre le SARS-CoV-2 en se basant sur la méthode QSAR. L'objectif principal de notre étude est de développer un modèle QSAR sur le plan statistique pour prédire de manière précise et fiable l'activité biologique.

2.1. Stratégie globale d'une étude QSAR :

La stratégie de développement d'un QSAR et plus générale qu'un QSPR, elle s'articule autour de 6 points :

- Constituer une base de données Structure - Propriété à partir de mesures quantitatives, fiables et normalisées de la propriété cible, pour chaque composé et Sélectionner des descripteurs chimiques en relation avec la propriété cible ;
- Diviser cet ensemble de données en un ensemble jeu d'apprentissage et un ensemble de test;
- Construire des modèles à partir d'un ensemble d'apprentissage avec les outils statistiques souhaités. Caractériser ces modèles par leurs indices de validation internes et vérifier leur robustesse par un test d'hasardisation ;
- Valider les modèles avec l'ensemble de test et calculer leur indice de corrélation externe.
- Répéter l'opération de division pour obtenir d'autres ensembles d'apprentissage et de test. La division optimale donne le plus petit ensemble d'apprentissage capable de bonnes prévisions pour le plus grand ensemble de test ;
- Explorer et exploiter les modèles validés pour comprendre les mécanismes possibles et faire des prévisions.

Enfin, pour tout modèle, il est important de savoir pour quel type de molécules il est utilisable ou non, c'est-à-dire connaître son domaine d'applicabilité.

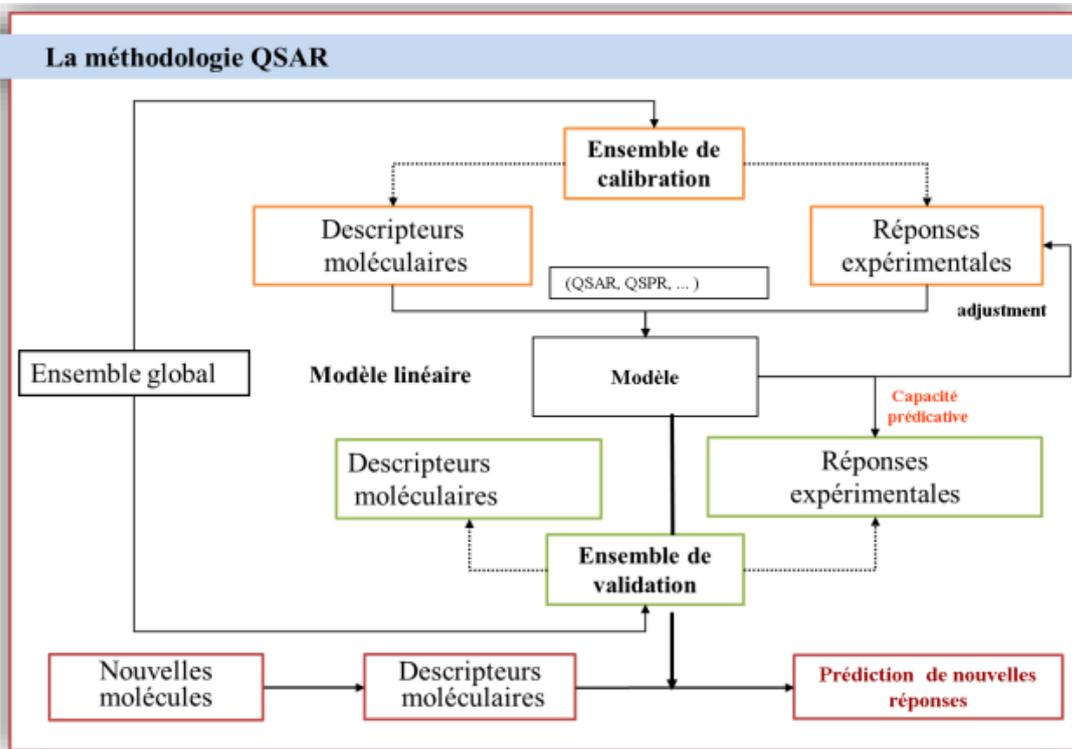
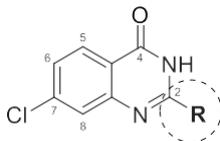


Figure 3.1 : Stratégie globale d’une étude QSAR.

2.2. Collecte des données

L’homogénéité des données est essentielle pour l’activité, elle doit être mesurée par un seul et même protocole et dans les mêmes conditions expérimentales.

- ❖ **Bases de données :** Plusieurs dérivés de 2-amino quinazolin-4(3H)-one présentent des propriétés pharmacologiques importantes. Ils sont utilisés dans le traitement de SARS-CoV-2. Le choix de la base de données expérimentale initiale est une étape critique pour le développement des modèles QSAR. Elle doit être composée de données expérimentales fiables. Pour cette étude nous avons utilisé l’activité inhibitrice IC₅₀ des dérivés de 2-amino quinazolin-4(3H)-one. Les structures et l’activité biologique expérimentale des dérivés de 2-amino quinazolin-4(3H) - one sont rapportées dans le tableau suivant :

Tableau 3-3 : L'activité anti-SARS-CoV-2 des dérivés de 2-amino quinazolin-4(3H)-one.

Composes	IC50 ^a (μM)	R
1-2	2.6	2,3,4-F3-Ph-NH-
2-1A	0.23	H
3-9B	1.4	3-F-Ph-NH-
4-2A	0.33	-COCH ₃
5-2B	0.29	-COCH ₃
6-2C	0.11	-COCH ₃
7-5C	0.15	H
8-9G	0.23	3,5-Cl ₂ -Ph-NH-
9-9H	11	3-CH ₃ O-Ph-NH-
10-9E	0.24	3,5-F ₂ -Ph-NH-
11-6A	0.21	-COCH ₂ CH ₃
12-9K	1.7	3-CN-Ph-NH-
13-9L	0.68	3-CF ₃ -Ph-NH-
14-9M	8.8	
15-9N	14	Cyclohexylmethyl-NH-
16-6B	7.05	-CO ₂ CH ₃
17-6C	5.66	-CH ₃
18-8	2.57	-COCH ₂ OH
19-10A	0.20	7-CF ₃
20-10B	0.41	7-NO ₂
21-1B	0.24	H
22-10E	0.47	5-OH
23-10F	1.6	8-OH
24-10I	0.33	5,8-Cl ₂
25-10J	0.25	7,8-Cl ₂
26-11A	0.21	7-CF ₃
27-11B	0.51	7-NO ₂
28-11E	0.15	5-OH
29-11F	0.37	8-OH
30-11I	0.35	5,8-Cl ₂
31-11J	0.24	7,8-Cl ₂
32-6D	0.57	<i>p</i> -Ts

33-Remdesivir	3.47	
---------------	------	--

2.3. Traitements des données :

Le développement du modèle QSAR nécessite le passage de notre base de composés chimiques par trois étapes :

- Etape 1 : Préparation de l'ensemble de données ;
- Etape 2 : Calcul des descripteurs ;
- Etape 3 : La modélisation. En utilisant plusieurs logiciels de traitement ci-dessous.

2.4. Matériels utilisés

2.4.1. Programmes (logiciels)

- A. **ChemDraw Ultra Bio** : Est un outil complet destiné aux chimistes et biologistes, intégrant toute une gamme d'outils intelligents. Il est imposé comme une référence pour le dessin de structures des composés chimiques. Simple à utiliser, puissant et qui permet de dessiner de manière intuitive et efficace en deux et trois dimensions.
- B. **Dragon** : est une application de calcul de descripteurs moléculaires, Sachant qu'il n'a pas été conçu comme un logiciel QSAR, il ne fournit que des descripteurs moléculaires et ne peut pas effectuer une analyse QSAR. Cependant, DRAGON permet la fusion des descripteurs moléculaires calculés (1664 descripteurs appartenant à différentes classes) et des propriétés définies par l'utilisateur pour un ensemble de molécules, fournissant un fichier de sortie complet qui est facilement chargé.
- C. **Molegro Data Modeller (MDM)** est un environnement intégré d'analyse et d'exploration de données. Il peut être utilisé pour créer des modèles de régression et de classification à l'aide de descripteurs numériques importés.
- D. **Minitab 18** : En utilisant ce programme nous appliquons la méthode QSAR pour obtenir une instruction de régression linéaire.

2.5. Méthodologie du QSAR

L'étude Relations Quantitatives Structure-Activité (QSAR) passe par trois étapes :

- ❖ **Construction des molécules** : Pour la construction des structures des molécules on a utilisé le logiciel « Chem Draw Ultra Bio» .⁶
- ❖ **Optimisation** : tous les calculs d'optimisation de la structure de la molécule ont été réalisés avec le logiciel Gaussian 9¹ par la fonctionnelle B3LYP associée à la base 6-31G (d, p).
- ❖ **Modélisation** : La modélisation passe par deux phases :
 - ◆ Calcul du descripteur : une géométrie optimisée a donc été adoptée Plus de 1600 calculs avec le logiciel DragonV5.4 qui est un logiciel peuvent être utilisés pour évaluer l'influence de la structure moléculaire ou les relations propriétés-structure, aussi pour l'analyse de symétrie et la projection des bases de données des molécules.
 - ◆ Conception du modèle : La modélisation consiste à construire des modèles mathématiques, afin de prévoir et de mieux en comprendre la relation entre la structure moléculaire et l'activité biologique. Nous avons utilisé la méthode RLM au moyen du logiciel Molegro Data Modeller 03.

2.6. Répartition des échantillons :

Pour la sélection des composés de calibrage et de validation, le bon modèle a été choisi avec selon les paramètres statistiques notamment avec une meilleur valeur du coefficient de corrélation et le test du Fisher.

2.7. Méthodes statistique utilisées :

❖ La Régression Linéaire multiple MLR :

Est l'une des méthodes de modélisation les plus populaires grâce à sa simplicité d'utilisation et de sa facilité d'interprétation. L'avantage important de la régression linéaire multiple est qu'elle est très transparente, puisque l'algorithme est disponible, et que les prédictions peuvent être réalisées facilement.

Références :

1. F.M.T.G.S.H.S.G.R.M.C.J.S.G.B.V.M.B.P.G.N.H.C.M.L. X. H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R.Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery, J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N.Kudin, V.N. Staroverov, R. Kobayashi, J Normand, K. Raghavachari, A. Rendell, J.C.Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E, Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G.Zakrzewski, G.A. Voth, P.S. alvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, O.Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox. Gaussian 09, Gaussian Inc., Wallingford CT, (2009) .
2. Trott. O, Olson. A. J, **AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading.** *Journal of Computational Chemistry*, (2010), p 455-461.
3. Chang. M.W, Ayeni .C, Breuer .S, Torbett B.E. ‘**Virtual screening for HIV protease inhibitors: a comparison of AutoDock 4 and Vina**’. *PLoS ONE*, (2010),5. p11955.
4. Nocedal. J, Wright. S.J, **Quasi-Newton Methods.** ‘*In Numerical Optimization.* T. V. Mikosch, S. I. Resnick, S. M. Robinson (Ed.). (2006)
5. Kitchen. D. B, Decornez. H., Furr. J. R, Bajorath. J. **Docking and scoring in virtual screening for drug discovery: methods and applications.** *Drug discovery*, (2004), p 935.



**CHAPITRE 04 :
RÉSULTAT ET DISCUSSION**



1- Docking moléculaire : Résultats et discussion

Dans le cadre de la recherche de nouvelles molécules naturelles efficaces contre le COVID-19, et afin de trouver de nouveaux inhibiteurs de ce virus, nous avons étudié par Docking moléculaire l'inhibition d'une série de molécules d'origine naturelle et quelques médicaments avec deux récepteurs du COVID-19 : ACE2 et M^{Pro}.

L'enzyme de conversion de l'angiotensine-2 (ACE2) est un récepteur fonctionnel du SARS-CoV-2 qui peut faciliter l'entrée du virus dans les cellules de l'hôte en se liant à la protéine de pointe (S) du virus.

La protéase virale principale (M^{Pro}) est essentielle à la réplication des coronavirus.

Les ligands naturels utilisés dans notre étude appartenant à la famille des flavonoïdes, des furanocoumarines on les trouve dans plusieurs sources naturelles telles que : les agrumes (citron, orange pamplemousse), le poivron, le maïs, les épinards ainsi que plusieurs fruits et légumes. Tandis que les médicaments sont des anti HIV et des vitamines.

Nous avons criblé l'ensemble des inhibiteurs étudiés par Docking moléculaire après avoir choisi le récepteur qui présente la meilleure affinité pour cette série de composés.

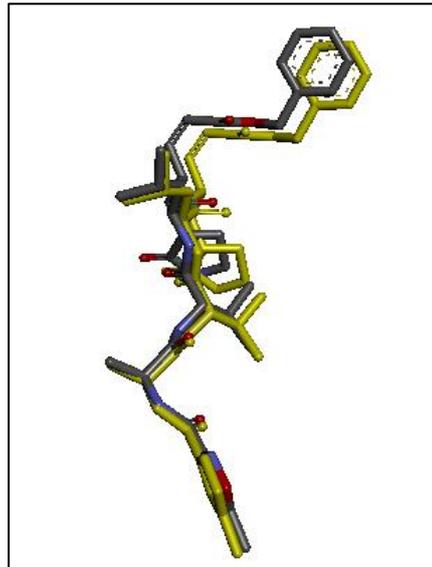
1.1. Le RMSD

La prédiction du mode d'interaction consiste à déterminer le positionnement correct du ligand par rapport à son récepteur. La valeur admise de la déviation quadratique moyenne ou RMSD (Root-Mean-Square-Derivation) de la pose du ligand calculée par le logiciel par rapport au ligand de référence est de 2 angströms, au-delà de laquelle la prédiction est considérée non adéquate.

Les valeurs du RMSD des complexes de M^{Pro} et du ACE2 sont inférieure à 2 Å pour les deux récepteurs, indiquant que les paramètres pour la simulation d'amarrage est bonne pour reproduire les structures cristallines. La structure cristalline (6LU7) a une petite valeur de RMSD (0.733 Å) ; par conséquent, son ligand co-cristallisé a été sélectionné et validé pour le reste de cette étude.

Tableau 4.1: RMSD des deux complexe protéine-ligand

CODE PDB	RMSD (Å) ⁰
6LU7	0.7333

**Figure 4.1 :** Comparaison de la conformation expérimentale du ligand avec sa conformation optimale calculée par Docking moléculaire avec AutoDock Vina (coloré en).

1.2. La boîte GRID BOX des protéines

Pour augmenter la vitesse d'évaluation de l'énergie du système, les récepteurs 6LU7 (M^{Pro}) et 6LZG (ACE2) sont plongés dans une grille tridimensionnelle englobant largement le site actif de la protéine étudié et permettant la libre rotation du ligand dans ce site. La boîte est alors centrée sur le ligand et ses dimensions sont proportionnelles à la taille de tous les ligands étudiés.

Tableau 4-2 : La boîte GRID BOX des protéines.

Récepteurs	X	Y	Z	Dimension A ³	Resolution
6LU7	-9.732	11.403	68.925	60x60x60	2.16
6LZG	-32.201	28.937	6.225	54x54x54	2.5

1.3. Interactions ACE2

1.3.1. Présentation du récepteur : Interface S-RBD

Le domaine RBD se lie à la protéine du récepteur ACE2 via ces résidus importants de l'interface : Phe486, Asn487, Ala475, Glu484, Tyr489, Leu455, Gln493, Tyr453, Ser494, Tyr449, Gln498, Gly496, Gly446 et Asn501, qui contribuent à la stabilité du complexe entre S-RBD et ACE2.

Tableau 4-3 : Résultats du Docking 6LZG-Ligand.

Ligands	Affinité ΔG (Kcal/mol)	Liaison hydrogène	Interactions Hydrophobes	Interactions électrostatiques
saquinavir	-12.5	GLN 98 GLU 208 GLY205 LYS 562	LEU 391 LYS 562 ALA 99	
Nelfinavir	-11.1	HIS 345	HIS 378	
Tipranavir	-10.5	TYR 127 ASN 149 HIS 505 ARG 273	PHE 504	GLU 145
Indinavir	-10.4	ALA 348 ARG 393 ASP 382	HIS 401 ARG 393 PHE 390 TRP 349	
Amprenavir	-10	ALA 348 ASP 382	HIS 401	HIS401
Darunavir	-9.7	ASP 350	PHE 40 TRP 349 TRP 69 HIS 401	
Ritonavir	-9.7	TRP 369 ASP 350 ALA 348 HIS 401 TYR 385	TRP 349 PHE 40 TRP 69	ASP 382
Naringine	-8.9	ASN 198 LEU 85 GLN 98	LEU 85	

		LYS 94 PRO 84 ASN 194		
DHB	-8.6	ASP206 TYR 202 TYR 196	TYR 202	
Bergaptene	-8.3	ILE 291 PRO 415 ALA 413	ILE 291 PRO 415 PHE 438	
Bergamotine	-7.9	SER 47	TRP 349 PHE 390 ARG 393 PHE40	
Remdesivir	-7.9	ASP 350 ASN 394 ASP 382	TRP 349 ARG 393	ASP 350 ASP 382
Methoxane	-7.8	ASN 290 ILE 291	ILE 291 PRO 415 PHE 438 ALA 413	
Imperatorine	-7.8	GLN98	VAL 209 PRO 565 VAL 212 LYS 562 LEU 95	LYS 562
Naringénine	-7.8	TRP 566 GLU 208	LEU 95 PRO 565 VAL 209	LYS 562 ASP 206
lopinavir	-7.8	ASN 394 ALA 348 ASP 350	TRP 349 ALA 348 HIS 378 PHE 40	
Fosampenavir	-7.6	ALA 348 ASP 350 ASN 394 TYR 385 ASP 382	PHE 40 PHE 390 TRP 349 HIS 378 HIS 401	ASP 350 ASP 382
Angelicine	-7.4	HIS 540	ILE 291 PHE 438 PRO 415	
Pimpinelline	-7.3	LYS 562 ASP 206 ASN 210	LYS 562 LEU 95 VAL 209	
Hydroxychloroquine	-6.8	ASN 290 ILE 291	ILE 291 PRO 415 LYS 541 HIS 540 PHE 436	
Psoralène	-6.7	TRP 349	TRP 349	

		SER 44 PHE 40		
Tocophérols	-6.7	PHE 390	TRP 69 PHE 40 LEU 73 LEU 100 PHE 390 LEU 391	
Zidovudine	-6.4	TYR 196 GLN 98 GLU 208		
Chloroquine	-6.3	GLU 208 ALA 396	LYS 562 LEU 95 VAL 209 VAL 212 PRO 565	
Emtricitabine	-5.8	ASP 509 SER 511 TRP 203	TRP203	
Lamivudine	-5.8	ASP 509 SER 511 TRP 203 ARG 514	TRP 203	

D'après les résultats de Docking des inhibiteurs avec l'ACE2, on remarque que le meilleur score égale à -12 kcal/mol, a été observé avec le médicament saquinavir. Pour Chacun des inhibiteurs médicamenteux ou naturels, les interactions mis en jeu avec le récepteur sont de type faible en particulier des liaisons hydrogène et hydrophobe avec des résidus non impliqués à l'interface S-RBD.

De point de vue énergétique, les anti HIV (saquinavir, nelfinavir, tipranavir, inidinavir, amprenavir, darunavir et ritonavir) ont les meilleures affinités pour le récepteur ACE2. Tandis que les deux anti HIV (lopinavir, fosampénavir) ont été classés après les ligands naturels (naringine, DHB, bargptene, bergamotine, imperatorine, naringénine).

Par ailleurs, tous les ligands naturels ont présenté des valeurs supérieures à celles de la CQ et l'HCQ.

Donc, on constate que l'anti- HIV saquinavir présente le meilleur pouvoir inhibiteur du récepteur ACE2 par rapport aux autres ligands étudiés. Malgré les valeurs d'affinité de la plupart des inhibiteurs y compris les inhibiteurs naturels pour ce récepteur, les interactions ligand- récepteur ont été moins importantes.

1.4. Interactions 6LU7 M^{Pro} :

1.4.1. Présentation du récepteur :

Site actif : Le site actif de l'enzyme est constitué des résidus suivants : His41, Met 49, Phe140, Lue141, Asn142, Gly143, His163, His164, Met165, Glu166, Leu167, Pro168, His172, Gln189, Thr190 et Ala191. Les résidus d'acides aminés His41 et Cys145 représentent la dyade catalytique dans le site actif, situé entre les domaines I et II ¹. His164 est essentiel pour l'activité enzymatique. Selon Yang et al. His163, His172 et Glu166 fournissent la porte d'ouverture pour les substrats actifs ².

Tableau 4-4 : Résultats du Docking 6LU7-Ligand (tous les inhibiteurs).

Inhibiteurs	Affinité	Liaison hydrogène	Interactions Hydrophobes	Interactions électrostatiques
saquinavir	-10.1	TYR154 ARG298 ILE152	ARG298 PHE8	
indinavir	-10	ASP153	PHE294 ARG298 VAL104	
Nelfinavir	-9.5	SER158	PHE294 VAL104	ASP153
tipranavir	-9.5	TYR154	PHE294 ARG298 PHE8	
Darunavir	-8.9	THR199 LEU287 GLY275 LEU272	LEU287 LEU286	
Naringine	-8.6	SER158 ARG298 ASP153 GLN110	VAL104 ILE106	
amprenavir	-8.5	LEU287 TYR237 LEU272	LEU286 LEU287	
Ritonavir	-8	SER158 SER158 ILE152	PHE294 VAL104	
lopinavir	-7.8	THR45 SER46	HIS41 PRO168	GLU166

		ASN142 THR24 GLN189	MET165 MET49	
Bergamotine	-7.8	CYS145 GLY143	MET165 HIS163	HIS41
Remdesivir	-7.7	GLY143 GLU166	GLU166 MET49	GLU166 HIS41
Fosampénavir	-7.6	HIS41 HIS41 GLY143 CYS145 HIS163 GLU166 LEU141 HIS164	MET165 MET49 HIS41	GLU166
DHB	-7.5	THR111 ARG298	TYR154 ARG298	
Naringénine	-7.5	ASP187 GLU166	MET49 CYS145	
imperatorine	-6.8	GLY143 HIS164 CYS145	MET49 MET165	
Pimpinelline	-6.4	HIS41 GLU16	HIS41 MET165 CYS145	
Methoxane	-6.3	ARG131 LYS137 THR199 TYR239	LEU286	LYS137 ASP289
Acide ascorbique	-6.3	ARG279 LEU220 TRP218 SER267		
Hydroxychloroquine	-6.2	HIS164 CYS145 GLY143 GLU166	CYS145	HIS163
bergaptene	-6.2	HIS 41, TYR54	MET165 HIS41	
Chloroquine	-5.8	GLY143 GLU166 HIS164 CYS145	CYS145	HIS163
Angelicine	-5.7	GLN189 HIS164	MET165 HIS41 MET165	
Psoralène	-5.6		ARG298 PHE8	

Tocophérols	-5.6	GLN110	ILE106 VAL104 PHE8	
Emtricitabine	-5.6	CYS145 HIS163 MET165 PHE140 GLU166 PHE140 GLU166	CYS145	
lamivudine	-5.2	HIS41 SER144 MET165 CYS145 ASN142		

D'après les résultats de Docking des inhibiteurs avec le M^{Pro} on remarque que le meilleur score égale à -10 kcal/mol, a été observé avec le médicament saquinavir. Les médicaments anti HIV présentant les meilleurs scores ont établies des interactions avec des résidus non impliqués au site actif du récepteur M^{Pro}, tels que saquinavir, indinavir, nelfinavir, tipranavir et darunavir. Par ailleurs, les anti HIV lopinavir, remdesivir et fosampénavir moyennement classés ainsi que les deux derniers emtricitabine et lamivudine ont établies des liaisons avec les acides aminés du site actif (Glu166, Met 165, His 163, Gln 189, Leu 141...) et avec la dyade catalytique His41 et Cys145.

Les inhibiteurs d'origine naturelle présentent des interactions de type hydrogène ou hydrophobe ou électrostatique avec des résidus impliqués au site actif et en particulier la dyade catalytique sauf la naringine malgré sa valeur d'énergie (-8.6 kcal/mol).

De point de vue énergétique, les ligands naturels ont présenté des valeurs supérieures ou à celles de la CQ et l'HCQ sauf l'angélicine, le psoralène et les tocophérols.

Donc, on constate que l'anti- HIV saquinavir présente le meilleur pouvoir inhibiteur du récepteur M^{Pro} par rapport aux autres ligands étudiés. Malgré les valeurs significatives de l'énergie trouvée avec les médicaments, les interactions mis en jeu avec le récepteur ont été moins importantes contrairement aux inhibiteurs naturels.

1.5. Comparaison entre les résultats de Docking obtenus avec les deux récepteurs ACE2 et M^{Pro} :

Par comparaison des résultats de Docking de la même série d'inhibiteurs d'origine naturelle et des médicaments avec le M^{Pro} et l'ACE2 (S-RBD), on constate que les meilleurs scores (-12 et -10 kcal/mol) ont été observés avec le médicament saquinavir pour l'ACE2 et le M^{Pro} respectivement. Tandis que pour les composés d'origine naturelle, les meilleures affinités (-8.9 et -8.6 kcal/mol) ont été obtenues avec la naringine pour l'ACE2 et le M^{Pro} respectivement. De manière générale, les inhibiteurs d'origine naturelle présentent une affinité vis-à-vis à M^{Pro} voisine à celle avec l'ACE2. Ces deux inhibiteurs de différente nature (saquinavir et naringine) présentent avec les deux récepteurs les plus faibles énergies d'interaction.

La stabilité de la plupart des complexes est attribuée aux liaisons hydrogène et hydrophobes. Tandis que les interactions électrostatiques contribuent aussi à la stabilité de quelques complexes avec le M^{Pro} tels que les anti HIV (lopinavir, remdesivir, fosampénavir, nelfénavir), les anti malaria (CQ et HCQ) et le ligand naturel bergamotine ainsi que d'autres complexes avec l'ACE2 (tipranavir, ritonavir, remdesivir, fosampénavir) et les ligands naturels (impératorine et naringénine).

Donc, on peut conclure que le pouvoir inhibiteur des composés naturels étudiés avec le M^{Pro} est plus important qu'avec l'ACE2 en raison des interactions mis en jeu avec le site actif du récepteur. Ces interactions ligand- récepteur sont de type faible en l'occurrence des liaisons hydrogène et hydrophobes.

1.6. Inhibition du récepteur SARS-CoV-2 (M^{Pro}) par les ligands d'origine naturelle:

Dans le but de trouver de nouveaux composés naturels efficaces contre la pneumonie COVID-19, nous avons criblés par Docking moléculaire l'ensemble des inhibiteurs étudiés. Donc, dans le reste de cette partie nous nous sommes concentrés sur l'inhibition de SARS-CoV-2 (M^{Pro}) par une série de molécules naturelles et deux médicaments (Chloroquine et Hydroxychloroquine).

Tableau 4-5 : Résultats du Docking moléculaire 6LU7-LIGAND

Ligands	Affinité	Liaison hydrogène	Interactions Hydrophobes	Interactions électrostatiques
Bergamotine	-7.8	CYS145, GLY143	MET165, HIS163	HIS41
Naringénine	-7.5	ASP187, GLU166	MET49, CYS145	
Impératorine	-6.8	GLY143, HIS164 CYS145	MET49, MET165	
Hydroxychloroquine	-6.2	HIS164, CYS145 GLY143, GLU166	CYS145	HIS163
Bergaptene	-6.2	HIS 41, TYR54	MET165, HIS41	
Chloroquine	-5.8	GLY143, GLU166 HIS164, CYS145	CYS145	HIS163

1.7. Énergies d'interaction : affinité des ligands pour le récepteur M^{Pro}

La Bergamotine, la naringénine, la chloroquine, l'hydroxychloroquine, le bergaptène et l'impératorine ont été étudiés comme inhibiteurs potentiels du SARS-CoV-2 M^{Pro} (**Tableau 4-2**). L'énergie d'interaction (ΔG) a été estimée comme suit :

$$\Delta G = \Delta G_{\text{complexe}} - \Delta G_{\text{ligand}} - \Delta G_{\text{protéine}}$$

Les énergies d'interaction obtenues à partir du Docking moléculaire de 6LU7 avec les ligands naturels : Bergamotine, Naringénine, Impératorine, Hydroxychloroquine, Bergaptene et Chloroquine sont :- 7,8, -7,5,-6,8, -6,2, -6,2, -5,8 kcal/mol, respectivement.

L'analyse du Docking a révélé le potentiel inhibiteur des ligands étudiés, classés par affinité : Bergamotine > Naringénine > Impératorine > Hydroxychloroquine Bergaptene > Chloroquine.

La stabilité de ces complexes (ligand- récepteur) est assurée par des interactions faibles établies entre le ligand (inhibiteur) et le récepteur (protéine) telles que les liaisons VDW, les liaisons électrostatiques et les liaisons H. Pour élucider le type d'interaction entre le ligand et M^{Pro}, nous avons analysé les interactions du complexe obtenu.

1.8. Etude des interactions du M^{Pro} (6LU7)

Les résultats obtenus montrent que tous les ligands naturels interagissent avec les résidus du site actif et en particulier avec la dyade catalytique de HIS41 et CYS145. La bergamotine présente la plus grande valeur d'affinité pour le récepteur M^{Pro} (PDB ID 6LU7), ce ligand est stabilisé par deux interactions hydrogène l'une avec CYS145 et l'autre avec GLY143 et deux liaisons hydrophobes formées avec les résidus MET165 et HIS163 ainsi qu'une interaction électrostatique avec HIS41 (**Figure 4. 2**).

Deux interaction hydrophobes réalisés avec les résidus MET49 et CYS145, ainsi que deux liaisons hydrogène avec ASP187 et GLU166, stabilisent le ligand naringénine (**Figure 4. 3**).

Le ligand impératrice a montré trois interactions hydrogène avec GLY143, HIS164 et CYS145, ainsi que deux interactions hydrophobes avec MET49 et MET165.

De plus, la molécule d'hydroxychloroquine est stabilisée par des liaisons hydrogène générées par les résidus GLY143, HIS164, CYS145 et GLU166, ainsi que par des interactions hydrophobes et électrostatiques formés par CYS145 et HIS163 (**Figure 4.4**). Le bergaptène forme deux interactions hydrogène avec HIS41 et TYR54 et deux interactions hydrophobes avec les résidus MET165 et HIS41, ce dernier étant l'un des résidus de la dyade catalytique.

Le ligand chloroquine a révélé quatre liaisons hydrogène avec les résidus GLY143, GLU166, HIS164 et CYS145, une CYS145 hydrophobe ainsi qu'une liaison hydrophobe CYS 145 et une liaison électrostatique avec HIS163 (**Figure 4.5**).

Deux résidus de la dyade catalytique, CYS145 et HIS41, participent à différentes interactions (liaison H et liaison électrostatique) pour le ligand bergamontine, qui a la meilleure affinité pour le récepteur M^{Pro} (énergie de liaison la plus faible). Dans les autres ligands, un seul résidu de la dyade catalytique est impliqué.

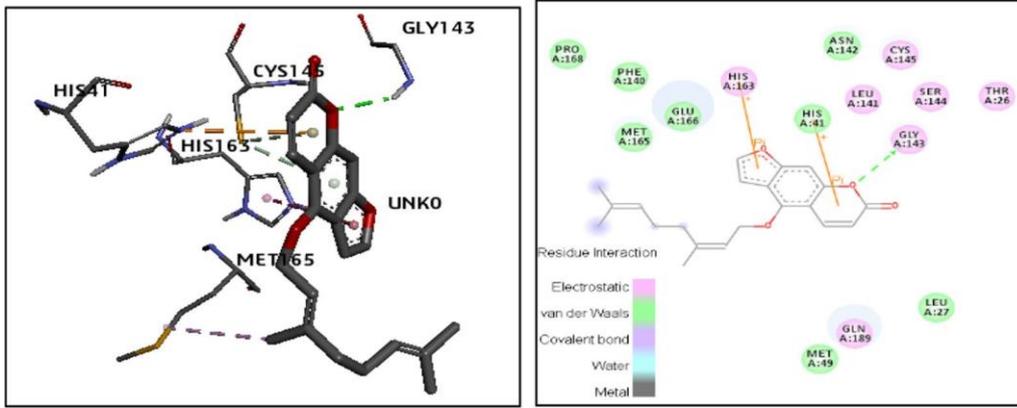


Figure 4.2 : Mode d'interaction de 6LU7-bergamotone 2D, 3D.

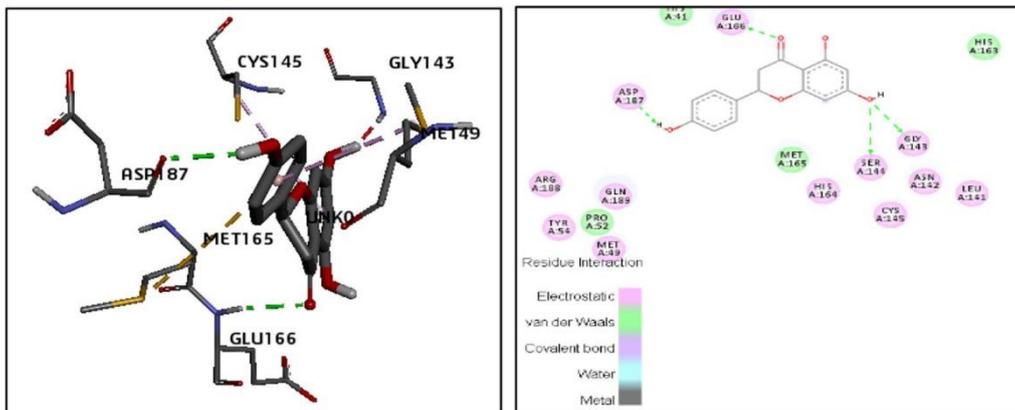


Figure 4.3 : Mode d'interaction de 6LU7-naringénine 2D, 3D.

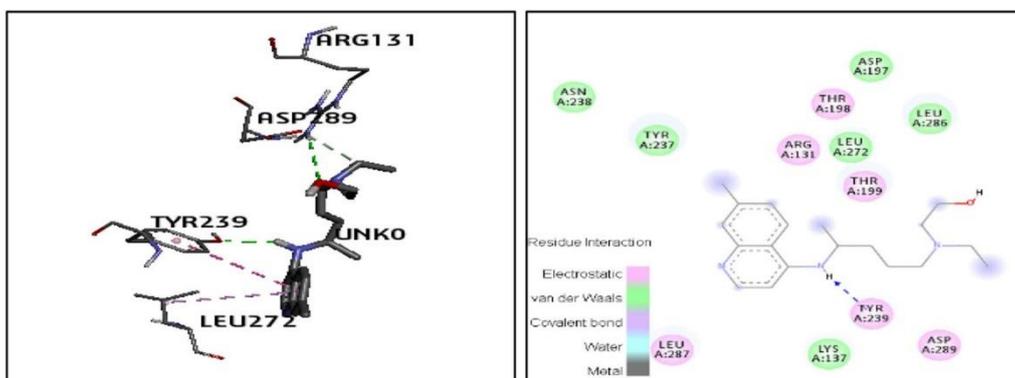


Figure 4.4 : Mode d'interaction de 6LU7-hydroxychloroquine 2D, 3D.

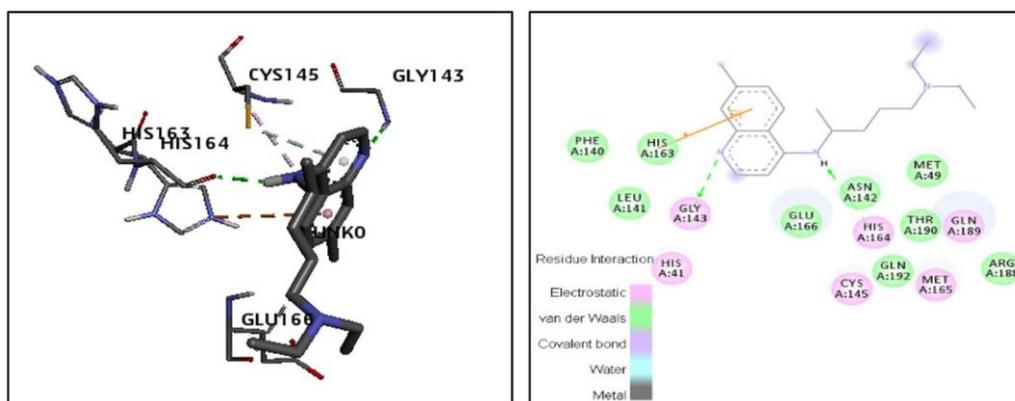


Figure 4.5 : Mode d'interaction de 6LU7-Chloroquine 2D, 3D.

1.9. Propriétés pharmacologiques des ligands

Ce test est réalisé pour faciliter la création de nouvelles molécules médicamenteuses dont les molécules doivent répondre aux conditions des cinq règles de Lipinski suivantes : poids moléculaire : ≤ 500 , nombre de donneurs de liaisons hydrogène : ≤ 5 , nombre d'accepteurs de liaisons hydrogène : ≤ 10 , lipophilie (exprimé en LogP) : 5 et réfraction molaire de 40 à 130.32. Les résultats de cette étude sont présentés dans le tableau.

La naringénine, l'impératorine, la bergamottine, l'hydroxychloroquine, le bergaptène et la chloroquine ont les poids moléculaires suivants : (293, 273, 340, 335, 234 et 226) g/mol et ont un poids moléculaire de 500 g/mol et présentent également les caractéristiques polaires topologiques suivantes. Valeurs de surface (TPSA) (97, 53, 44, 48, 48, 44 et 64) Å². Des valeurs TPSA plus basses donnent de bons résultats.

Tableau 4-6 : Résultats des propriétés de Druglikeness des ligands.

Propriété médicamenteuse	Naringenine	Imperatorine	Bergamottine	Hydroxy chloroquine	bergaptene	chloroquine
Poids moléculaires g/mol	293	273	340	335	234	226
Consensus Log Po/w	0.44	2.47	4.76	3.29	1.06	0.6
Log S	-1.67	-3.52	-5.2	-3.91	-1.77	-0.6
Num. H-bond accepteurs	5	4	4	3	4	4
Num. Donneurs de liaisons	4	0	0	2	0	4

Réactivité moléculaires	74	74	99	98	56	58
Lipinski	YES	YES	YES	YES	YES	YES
Ghose	YES	YES	YES	YES	YES	YES
Veber	YES	YES	YES	YES	YES	YES
Egan	YES	YES	YES	YES	YES	YES
Muegge	YES	YES	NO	YES	YES	YES
Score de biodisponibilité	0.55	0.55	0.55	0.55	0.55	0.55
TPSA	97	44	48	48	44	64
Num. liaisons rotatives	4	3	6	9	1	9
Solubilité	-1.67	-3.52	-5.2	-3.91	-1.77	-0.6

En comparant les valeurs de lipophilie (LogP) de notre série de ligands, nous observons qu'ils ont des valeurs inférieures à cinq, indiquant qu'ils ont produit d'excellents résultats et peuvent être facilement absorbés par l'organisme.

Cependant, les valeurs de solubilité LogS des ligands testés sont comprises entre 4,06 et 0,25. La bergamotine a la valeur la plus élevée ; les valeurs inférieures sont toujours visibles.

Tous les ligands ont un nombre de donneurs de liaisons hydrogène : 5, un nombre d'accepteurs de liaisons hydrogène : 10 et des valeurs de réactivité comprises entre 35,75 et 130. On peut donc dire que la règle des cinq de Lipinski a été vérifiée. La règle de Weber, qui représente la biodisponibilité orale d'une molécule médicamenteuse potentielle ; est vérifiée pour tous les ligands étudiés.

De même, la règle d'Egan qui définit l'absorption d'une molécule médicamenteuse est vérifiée pour tous les ligands ; en revanche, les deux règles de Veber, Egan et Muegge sont vérifiées.

Nous avons également remarqué que certains ligands ont des effets sur la reproduction, des propriétés irritantes, mutagènes et tous les ligands ont montré un score de biodisponibilité de 0,55 et un bon score de médicament, solubilité et score de médicament.

1.10. Résultats ADMET (absorption, distribution, métabolisme, l'excrétion, et de toxicité) :

Les tests ADME permettent de déterminer les propriétés pharmacologiques et pharmacodynamiques d'un médicament dans un système biologique.

Les propriétés ADME/Tox (absorption, distribution, métabolisme, excrétion/toxicologie) de tous les ligands ont été calculées à l'aide du serveur en ligne PreADMET (<https://preadmet.bmdrc.kr/>). Ce serveur permet de calculer les propriétés pharmacocinétiques telles que : l'absorption intestinale humaine (HIA), la perméabilité cellulaire, le caco-2 in vitro (PCaCO2), la perméabilité des cellules Kidney Maden Darby (PMDCK), la perméabilité (PSkin), la liaison aux protéines plasmatiques (PPB) et le sang, pénétration cérébrale.

Les résultats sont répertoriés dans le **tableau 4-7**.

Tableau 4-7 : Résultats du test ADME/T des ligands (diverses propriétés pharmacocinétiques et pharmacodynamiques).

	Propriétés	Naringénine	Imperatorine	Bergamotine	Hydroxy chloroquine	bergaptene	chloroquine
Absorption	HIA (%)	87.31	97.72	97.39	94.66	98.16	98.05
	Caco-2 cellule (nm sec-1)	10.52	56.10	49.82	46.08	43.47	56.61
	MDCK	44.63	228.71	0.12	45.10	52.91	0.29
Distribution	Peau_ Perméabilité	-4.18	-3.16	-2.49	-3.08	-3.70	-2.53
	BBB	0.59	2.70	0.36	2.28	2.07	7.73
Métabolisme	Cyp-2C19-inhibition	Inhibiteur	Inhibiteur	Inhibiteur	Non	Inhibiteur	Non
	Cyp-2C9-inhibition	Inhibiteur	Inhibiteur	Inhibiteur	Non	Inhibiteur	Non
	Cyp-2D6-inhibition	Non	Non	Non	Inhibiteur	Non	Inhibiteur
	Cyp-2D6-Substrat	Non	Non	Non	Substrat	Non	Substrat

	Cyp-3A4- Inhibition	Inhibiteur	Inhibiteur	Inhibiteur	Non	Inhibiteur	Non
	Cyp-3A4- Substrat	Non	Faible	Substrat	Substrat	Non	Substrat
toxicité	Mutagénicité	mutagène	mutagène	mutagène	mutagène	mutagène	mutagène
	Cancérogénicité Souris	négative	négative	Positive	négative	négative	négative
	Cancérogénicité Rat	Positive	négative	négative	négative	Positive	Positive
	HERG inhibition	Risque moyen					

Le paramètre cellulaire Caco-2 est utilisé pour prédire les paramètres d'absorption et de distribution afin de déterminer la perméabilité du composé. Les paramètres des cellules Caco-2 variaient de 4 à 70 nm/s, y compris la perméabilité moyenne. L'absorption intestinale humaine (HIA) peut prédire le pourcentage d'absorption de l'intestin humain (%HIA) et la liaison aux protéines plasmatiques (%PPB) est utilisée pour déterminer le degré auquel les médicaments se lient aux protéines dans le sang. Plus la valeur du médicament lié est inférieure à 90 %, plus il peut traverser efficacement les membranes cellulaires ou se diffuser.

En raison de ces données et valeurs, les ligands ont un score d'absorption intestinale humaine (HIA) plus élevé. Une valeur HIA plus élevée indique que le composé pourrait être mieux absorbé par le tractus intestinal par administration orale, alors les valeurs de Caco-2 concordent avec les valeurs obtenues (valeur entre 4 et 70 nm/s). Ainsi, tous les ligands se sont avérés avoir une perméabilité moyenne. Valeurs BBB (CBrain/CBlood) et composés de débogage qui ont des valeurs supérieures à 1.

(CBrain/CBblood > 1) sont considérés comme actifs dans le Système Nerveux Central (SNC) et peuvent provoquer des effets indésirables, tandis que les composés avec des valeurs inférieures à 1 (CBrain/CBblood <1) sont classés comme inactifs dans le SNC.

Par conséquent, les ligands Bergaptene, impérotine, hydroxychloroquine et chloroquine avaient des valeurs supérieures à 1 ($C_{Brain}/C_{Blood} > 1$), indiquant qu'ils sont actifs dans le SNC et peuvent provoquer des effets indésirables.

Effets, tandis que la naringénine et la bergamotine avaient des valeurs inférieures à 1 ($C_{Brain}/C_{Blood} < 1$), qui sont considérées comme inactives dans le SNC.

Le test de toxicité AMES est utilisé pour déterminer si une substance est mutagène ou non ; la plupart des ligands testés sont mutagènes. Une prédiction de cancérogénicité négative indique qu'il existe des preuves d'activité cancérogène, tandis qu'une cancérogénicité positive indique que la substance d'essai n'est pas cancérogène. De plus, l'inhibition de HERG pour tous les ligands présente un risque modéré.

Sur la base des résultats d'ADMET, on peut conclure que les propriétés biologiques et pharmacologiques des ligands naturels sont plausibles, leur permettant d'avoir une biodisponibilité acceptable et donc des options intéressantes contre le SRAS-CoV-2.

1.11. Simulation de Dynamique moléculaire

La simulation de dynamique moléculaire (DM) est la méthode la plus largement utilisée pour évaluer les micro- interactions entre la structure du ligand et celle de la protéine.

Pour une meilleure compréhension de la stabilité et de la dynamique de tous les complexes, des simulations MD de 50 ns ont été utilisées. RMSD, RMSF, Rg et SASA ont été évalués après des simulations MD.

Les simulations MD de 50 ns ont été réalisées dans des conditions standard. Le champ de force CHARMM était appliquées pour obtenir les attributs topologiques des protéines et des ligands³. Le thermostat Berendsen contrôlait la température des simulations. Les interactions électrostatiques ont été traitées par le maillage particulière Méthode d'Ewald (PME) avec une distance de coupure de 1,2 nm ; et les interactions de van der Waals étaient modélisé à l'aide d'un seuil de 1,4 nm⁴. L'ensemble du système a été assoupli à l'aide de l'algorithme de descente la plus raide. Les équations classiques du mouvement de Newton ont été intégrées à l'aide de l'algorithme Leap-Frog et la méthode de couplage faible de Berendsen⁵. Les coordonnées de la trajectoire ont été enregistrées tous les 500 pas avec un pas de temps de 1 fs. Toutes les simulations MD ont été réalisées à l'aide de la version du package GROMACS 2020. Des

calculs de l'écart quadratique moyen (RMSD) ont été effectués pour identifier les ensembles partageant mêmes caractéristiques conformationnelles. Alors que les calculs de la fluctuation quadratique moyenne (RMSF) trajectoires ont été utilisées pour déterminer les parties les plus mobiles du système, en particulier celles impliquées dans l'interaction entre la protéine et les ligands. Le rayon de giration (R_g) a été utilisé pour évaluer la compacité de la protéine ainsi que sa stabilité.

Enfin, la surface accessible aux solvants (SASA) a été utilisée afin de conserver le repliement et la stabilité des protéines.

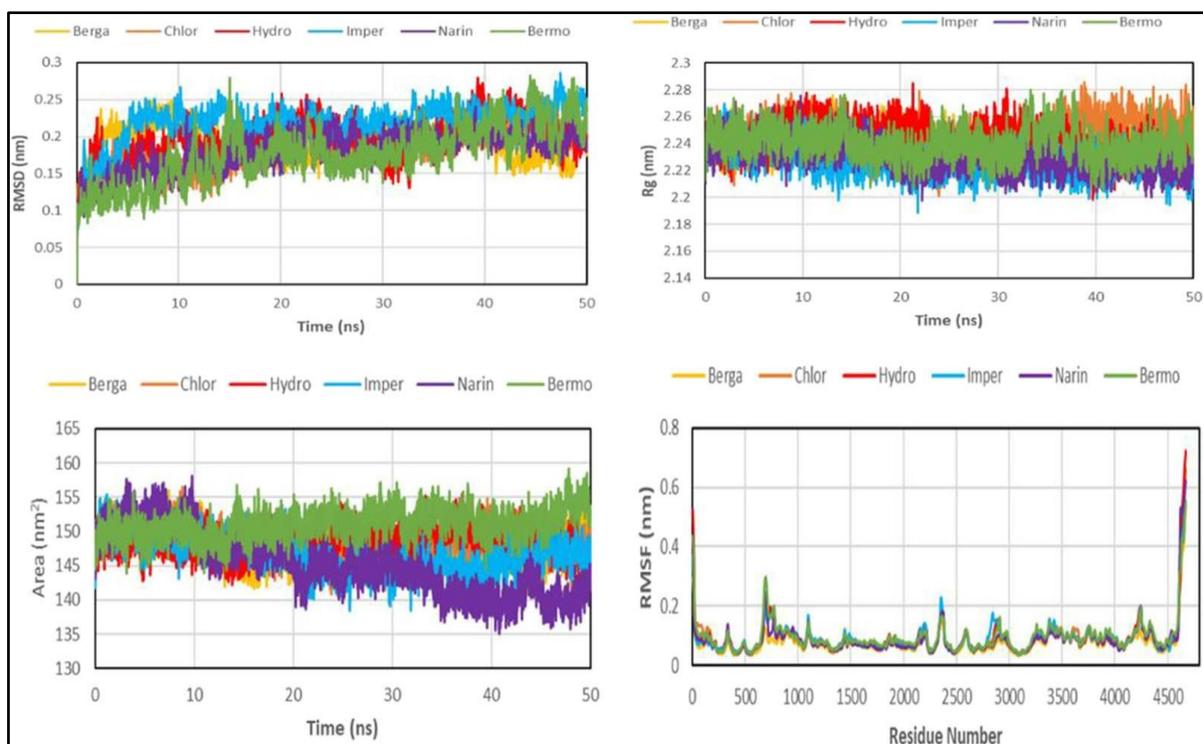


Figure 4.6 : Comparaison et représentation détaillée de (A) RMSD : écarts quadratiques moyens, (B) RMSF : fluctuations quadratiques moyennes, (C) R_g : rayon de giration et (D) SASA : surface accessible au solvant, sur le temps de 50 ns de référence M^{Pro} et tous complexes.

Les tracés de l'écart quadratique moyen (RMSD) montrent que tous les complexes sont restés stables sur 15 à 40 ns de simulation avec des valeurs moyennes allant de 0,18 à 0,22 nm. Les résultats RMSD indiquent que la liaison entre M^{Pro} et tous ces ligands est stable. Le rayon de giration (R_g) a été utilisé pour évaluer la compacité des complexes et R_g de tous les systèmes s'est avéré stable avec la RMSD du système. Cela indique que la protéine était stable et

comprimée pendant 50 ns. Les valeurs moyennes de R_g vont de 2,23 à 2,27 nm. Pour calculer davantage la flexibilité résiduelle sur un temps de 50 ns, les fluctuations de RMSF ont été mesurées. Les fluctuations totales de tous les complexes sont approximativement les mêmes. Les valeurs moyennes de RMSF sont comprises entre 0,079 et 0,098 nm. La surface accessible au solvant (SASA) est une méthode alternative pour maintenir le repliement et la stabilité des protéines. Les valeurs SASA calculées pour le type sauvage et les mutants sont présentés à la (**Figure 4.6**). Les valeurs SASA moyennes pour les complexes se sont avérées être de 147 nm² (Berga), 148 nm² (Chlor), 148 nm² (Hydro), 146 nm² (Imper), 145 nm² (Narin) et 151 nm² (Bermo). Cela signifie qu'il n'y avait pas de différences notables dans la zone disponible de tous les systèmes pendant le processus de simulation.

2.Relation Quantitative Structure-Activité (QSAR)

L'étude QSAR a été focalisée sur une série de 33 substituants aux dérivés de 2-aminoquinazolin-4(3H)-one., afin de déterminer une relation quantitative entre la structure et les activités inhibitrices du SARS-CoV-2.

2.1 Régression linéaire multiple

Afin de sélectionner les descripteurs prédominants qui affecteront les activités anti-SARS-CoV-2 des composés étudiés, l'analyse de corrélation a été effectuée avec le logiciel Molegro Data Modeller 03, prenant chaque descripteur calculé en tant que variable indépendante et pIC50 comme variable dépendante. Sur la base de l'analyse de corrélation, la méthode de régression linéaire multiple a été utilisée pour établir le modèle QSAR.

Nous avons divisé l'ensemble de données en deux sous-ensembles :

- **Training _set** : destiné à l'apprentissage d'un modèle.
- **Test set** : destiné à l'évaluation du modèle

Afin de sélectionner la meilleure performance de régression, on utilise les paramètres statistiques suivants : le coefficient de corrélation multiple R, le coefficient de détermination R^2 , le coefficient de détermination ajusté R^2 ajusté, le coefficient de prédiction Q^2 , la statistique de Fisher F et l'écart type S.

2.1.1 Equation et analyse de régression

Le modèle QSAR obtenu en utilisant la méthode de régression linéaire multiple (RLM) est représenté par l'équation suivante :

$$\text{pIC50} = 1.643 + 62.1 \text{ JGI3} + 0.657 \text{ G3u}$$

Ce modèle reliant la variable dépendante pIC50 avec les deux descripteurs, est accompagné par les mesures statistiques R, R², R²_{ajust}, Q² et les paramètres statistiques obtenus par la régression sont rassemblés dans le tableau suivant :

Tableau 4-8 : Paramètres statistiques RLM de 2 descripteurs

Les paramètres statistiques de ce modèle sont :

S	R	Q ²	R ²	R ² _{ajust}	F
0,405701	79,10%	40.50%	59,25%	56,54%	21.81

Le coefficient de corrélation multiple (R = 0.791) nous indique qu'il existe une corrélation entre les valeurs d'activité biologique observées et prédites par le modèle. En termes de variabilité, les résultats d'analyse ajustés ont montré une variance de 79,10% et 59,25%, respectivement.

Les observations sont regroupées selon deux descripteurs (JGI3 et G3u).

2.1.2 Analyse de variance

La variance est une mesure de la distance entre les données et la moyenne, la variance de l'échantillon est égale au carré de l'écart-type de l'échantillon.

Cette corrélation est vérifiée en examinant, la présence ou l'absence de corrélation entre l'ensemble de descripteurs et l'activité biologique par la statistique de Fisher et l'utilisation des valeurs de la statistique test de Student.

Tableau4-9: Résultats de l'analyse de la variance

Source	DDL	Somme des carrés	CM	F	P
Régression	2	7,181	3,590	21.81	0.000
Erreur résiduelle	30	4.938	0,164		
Total	32	12,119			

La Probabilité pour le test de Fisher $p < \alpha=0.05$ ce qui montre que le modèle est significatif.

Les statistiques DW sont utiles pour évaluer la présence ou l'absence d'une autocorrélation en série des résidus des modèles de régression. Si la statistique DW est sensiblement inférieure à 2, il existe des preuves d'autocorrélation en série positive et une valeur vers 4 indique une autocorrélation négative.

Dans notre étude, la valeur DW de modèle 1.65258 est plus proches de 2, ce qui montre que la valeur est au-dessus de l'autocorrélation positive.

Nous avons reconstruit le modèle optimal établi précédemment, en utilisant l'analyse de régression linéaire multiple par le logiciel MINITAB 18 .⁶

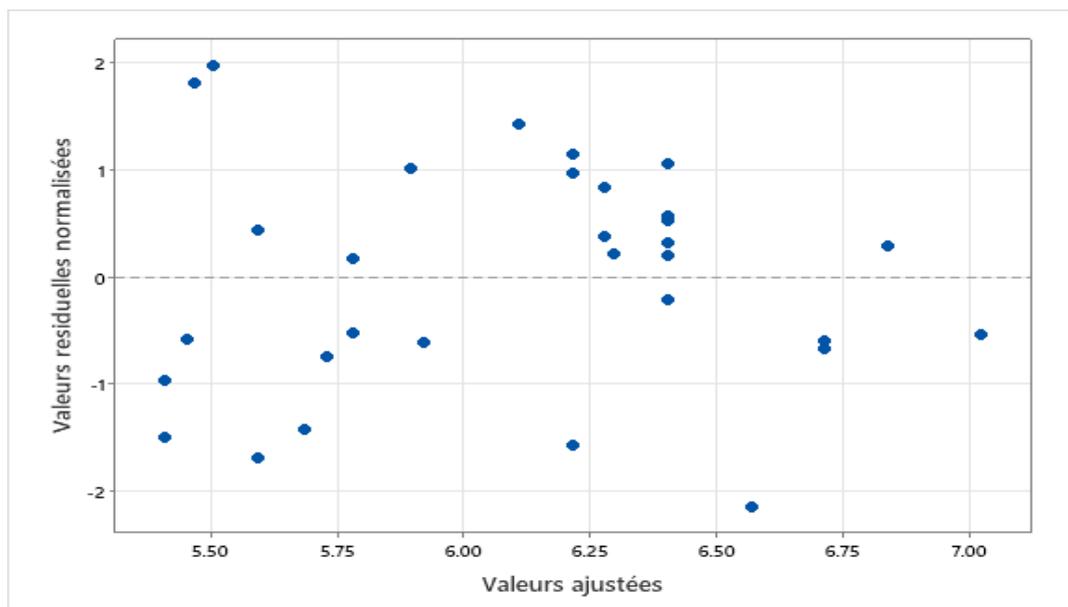
**Figure 4.7 :** Valeurs résiduelles ajustées en fonction des valeurs ajustées

Tableau 4-10 : Définition des descripteurs

Descripteur	Groupe	Définition
JGI3	Autocorrélations 2D	Indice de charge topologique moyen d'autocorrélation d'ordre 3-2D
G3u	Descripteurs WHIM	3ème composante symétrie directionnelle WHIM indice/non pondéré.

A. Analyse des résultats :

La valeur de coefficient de détermination $R^2=0.5925$ montre que 59,25% de la variabilité du pIC50 est expliquée par les descripteurs. Le reste de la variabilité est dû à des effets (autres variables explicatives) qui ne sont pas pris en compte ou bien les descripteurs sélectionnés ne sont pas totalement descriptifs.

B. Interprétation :

Les deux descripteurs sont utilisés pour obtenir des informations pertinentes sur la molécule. Ils sont importants pour construire un modèle QSAR

Dans l'équation de modèle, le coefficient positif élevé JGI3 (Indice de charge topologique moyen d'ordre 3) explique qu'il entraîne une augmentation de l'activité biologique de ces composés. Cela signifie que ce descripteur topologique qui renseigne sur l'échange de charges entre atomes situés à une distance topologique d'ordre 3, nous donne d'importantes informations sur pIC50. Par conséquent, plus la valeur de cet autocorrélation est élevée, plus grande sera l'activité biologique.

Le coefficient de descripteur de l'Indice de symétrie WHIM basé sur des indices statistiques et des projections calculées des atomes le long des axes principaux, a un signe positif dans le modèle. Le descripteur géométrique (G3u) donne des informations 3D moléculaires pertinentes concernant la taille, la forme, la symétrie et la distribution des atomes moléculaires, donc toute augmentation de ce descripteur entraîne une augmentation de l'activité biologique.

2.1.3 Validation du modèle

Un modèle QSAR peut être obtenu par le hasard, pour cette raison, il faut toujours s'assurer de sa stabilité. Pour la validation, nous avons utilisé 33 des composés étudiés.

A. Validation Leave one out (LOO)

La procédure LOO (mettre une molécule à parts) a donné une valeur moyenne

$$R^2_{\text{ajust}} = 56,54\%$$

Cette valeur est loin de la valeur de $R^2 = 59,25\%$.

Tableau 4-11 : Paramètres de validation croisée LOO

PRESS	SSY	PRESS/SSY	SPRESS	R ² ajust
0.16459	3.59074	0.045837	0.09094	56,54

PRESS (carrés de la somme résiduelle prédite) sa valeur étant inférieure à SSY (somme des carrés de la valeur de réponse) pour un modèle raisonnable de QSAR le rapport PRESS/SSY doit être inférieur à 0,4.

Les résultats présentés dans le tableau montre que le modèle QSAR développé présente une faible valeur de SPRESS est de 0.09094 (<0.200) ce qui indique que le modèle présente une petite valeur résiduelle entre l'activité biologique observée et prédite.

B. Vérification des critères de Tropsha

Tableau 4.12 : Paramètres statistiques de test

R ² _{test}	R ² ₀	R ² -R ² ₀	$\frac{ R^2-R^2_0 }{R^2}$
0.79	0.59	0.03	-0.05

$$R^2_0 = 0.59$$

$$|R^2 - R^2_0| = 0.03$$

$$\frac{R^2 - R_0^2}{R^2} = -0.05$$

$$R^2$$

Un modèle de QSAR ne possède une puissance prédictive acceptable que si les conditions suivantes sont satisfaites :

$$R^2 > 0.6, \quad \frac{|R^2 - R_0^2|}{R^2} < 0.1, \quad |R^2 - R_0^2| < 0.3$$

Tous les critères de Tropsha sont vérifiés, alors notre modèle est fiable.

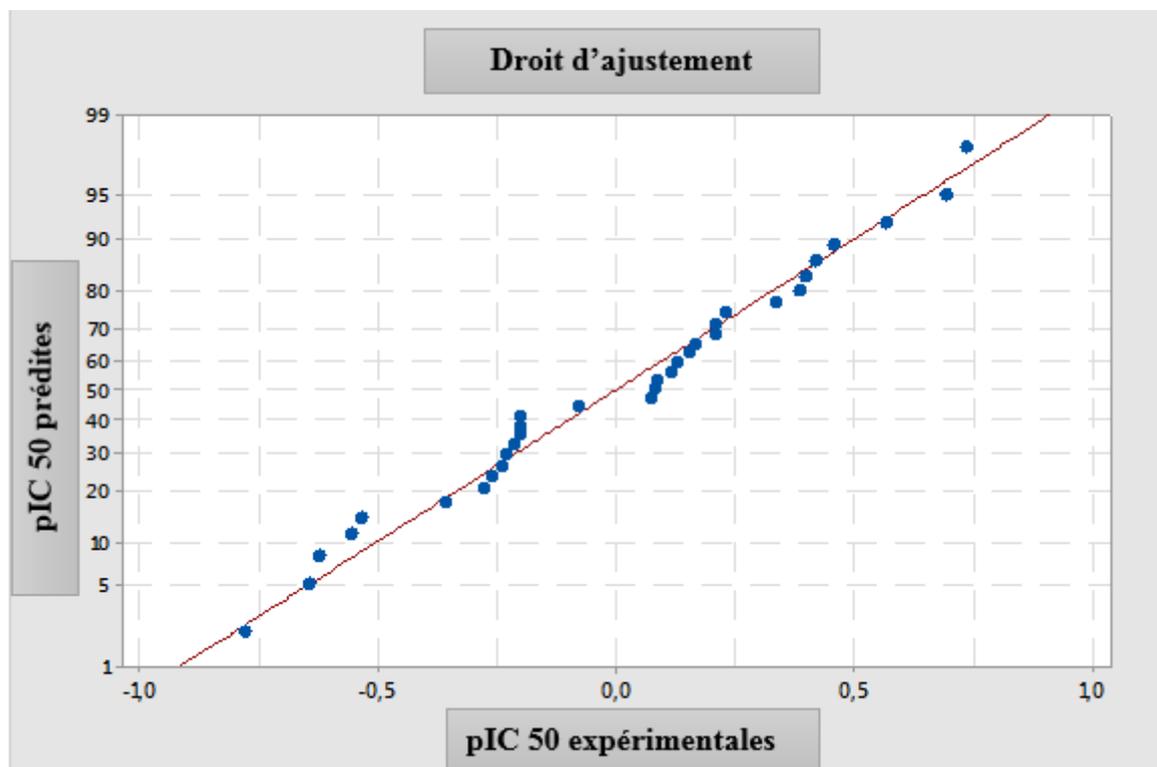


Figure 4.8 : Variation des valeurs expérimentales en fonction des valeurs prédites

❖ Interprétation du graphe :

La qualité de l'ajustement a été vérifiée en représentant les valeurs pIC50 expérimentales en fonction des valeurs prédites de pIC50 (**Figure 4.8**) montre une faible dispersion autour de la droite d'ajustement ce qui traduit la faiblesse des erreurs lors du calcul. On déduit donc, qu'on a un bon ajustement confirmé par les valeurs du coefficient de corrélation issus de la régression R^2 supérieures à 59%, pour les deux ensembles (Calibration et Validation) ainsi que R ajust et la valeur de l'écart type ($S= 0.40$). Cela prouve la performance du modèle établi.

Références :

1. MT. Ul. Qamar, S. M. Alqahtani, M.A. Alamri, LL. Chen. **Journal of Pharmaceutical Analysis**, (2020), p 313-319.
2. A. Speciale, C. Muscarà, M.S. Molonia, F. Cimino, A. Saija, S.V.Giofrè. Wileyonline **library.com/journal/ptr**, (2021).
3. B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, et al, “CHARMM: The Biomolecular Simulation Program,” *Journal of Computational Chemistry* 30, no. 10 (2009): 1545–614., doi:10.1002/jcc.21287.
4. H. G. Petersen, “Accuracy and Efficiency of the Particle Mesh Ewald Method,” *Journal of Chemical Physics*. 103, no. 9 (1995): 3668–79. doi:10.1063/1.470043.
5. W. F. Van Gunsteren, and H. J. Berendsen, “A Leap-Frog Algorithm for Stochastic Dynamics,” *Molecular Simulation* 1, no. 3 (1988): 173–85. doi:10.1080/08927028808080941.
6. Assistance Minitab 18 .<https://support.minitab.com/fr-fr/minitab/18/>.



CONCLUSION GÉNÉRALE



Conclusion générale

Dans la lutte contre le SRAS-CoV-2, les scientifiques ont adopté différentes approches, notamment en essayant d'inhiber les protéases nécessaires au cycle de vie du virus à l'aide de médicaments et de composés naturels.

Dans ce travail, nous avons donc étudié l'inhibition du SARS-CoV-2 en utilisant l'amarrage moléculaire et la dynamique moléculaire avec plusieurs composés naturels, dont l'objectif principal étant de trouver de nouvelles molécules capables d'inhiber ce virus. Nous avons réalisé le Docking deux récepteurs du SRAS-CoV-2 la protéase principale (6LU7) et l'enzyme de conversion de l'angiotensine 2 ACE2 (6LZG) avec une série de flavonoïdes et de coumarines et de médicaments (CQ et HCQ).

L'étude de leurs interactions avec ces enzyme (ACE2 et M^{Pro}), suivie d'une utilisation de la règle de Lipinski et les propriétés ADMET a été cruciale.

Le Docking moléculaire a été réalisée par Auto Dock Vina, un logiciel d'amarrage moléculaire très répandu et connu par sa fiabilité et sa performance. Le logiciel Discovery Studio nous a permis la visualisation de ces interactions.

Les ligands ont été classés sur la base de leurs affinités. Les résultats de Docking ont montré que le pouvoir inhibiteur des composés naturels étudiés avec le M^{Pro} est plus important qu'avec l'ACE2 en raison des interactions mis en jeu avec le site actif du récepteur. Ces interactions ligand- récepteur sont de type faible en l'occurrence des liaisons hydrogène et hydrophobes.

Donc, la protéase M^{Pro} du SARS-CoV-2 s'avère une cible thérapeutique prometteuse pour la découverte de nouveaux médicaments contre le COVID-19.

Les résultats d'ADMET ont révélés que la majorité des ligands naturels étudiés ont des propriétés biologiques et pharmacologiques intéressantes, leur permettant d'avoir une meilleure biodisponibilité.

Une simulation de dynamique moléculaire (MD) a été réalisée dans notre étude afin de mieux comprendre la stabilité et la dynamique de tous les complexes. Comme résultat préliminaire,

ces ligands naturels semblent être des candidats potentiels prometteurs pour l'inhibition du SRAS-CoV-2.

Dans le but de prédire l'activité biologique de nouveaux inhibiteurs du SARS-CoV-2, l'approche QSAR a été employée pour trouver une relation quantitative Structure-Activité d'une série des anti-SARS-CoV-2.

Ainsi, une corrélation a été obtenue entre l'activité inhibitrice mesurée des dérivés de 2-amino quinazolin-4(3H)-one (Concentration Inhibitrice Médiane IC50) et les descripteurs moléculaires calculés par le logiciel Dragon, à partir des géométries préalablement optimisées avec la fonctionnelle B3LYP/6-31G (d, p) intégré au programme Gaussian.

Afin de sélectionner les descripteurs prédominants qui affecteront les activités anti-SARS-CoV-2 des composés étudiés, l'analyse de corrélation a été effectuée avec le logiciel Molegro Data Modeller 03, la méthode de régression linéaire multiple a été utilisée pour établir le modèle QSAR.

Le traitement statistique suivi par le Minitab montre que le modèle obtenu à une qualité statistique acceptable et fiable ($R=0.79$ et $R^2=0.59$). Une validation rigoureuse a été utilisée pour juger la stabilité, la justesse et la capacité prédictive du modèle obtenu. La qualité de l'ajustement du modèle développé a été vérifiée en procédant à la représentation des valeurs calculées en fonction de celles observées ($R_0^2 = 0.56$).

Ce modèle QSAR qui indique l'existence d'une corrélation entre deux descripteurs issus de la structure moléculaire du composé et son activité biologique, peut être utilisé pour prédire l'activité biologique inconnue d'un nouveau composé.

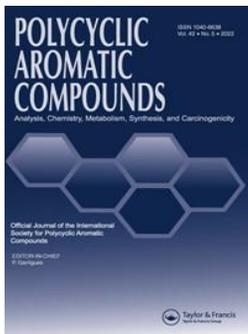
Sur la base de ces résultats, nous envisageons d'approfondir l'étude de l'effet de nouveaux composés sur l'inhibition du SARS-CoV-2, en faisant appel à d'autres programmes de Docking moléculaire et du QSAR avec l'utilisation des bases de données plus larges.

En termes de perspectives, une étude expérimentale *in vitro* et/ou *in vivo* s'avère nécessaire pour valider les molécules étudiées en tant qu'inhibiteurs du SARS-CoV-2.



ANNEXE





In Silico Study of Selected Natural Products as SARS-CoV-2 M^{Pro} Binder: Molecular Docking and Molecular Dynamics Simulation

Bochra Bencheikh, Mouna Cheriet, Rayenne Djemil & Djameleddine Khatmi

To cite this article: Bochra Bencheikh, Mouna Cheriet, Rayenne Djemil & Djameleddine Khatmi (2022): In Silico Study of Selected Natural Products as SARS-CoV-2 M^{Pro} Binder: Molecular Docking and Molecular Dynamics Simulation, Polycyclic Aromatic Compounds, DOI: [10.1080/10406638.2022.2097279](https://doi.org/10.1080/10406638.2022.2097279)

To link to this article: <https://doi.org/10.1080/10406638.2022.2097279>



Published online: 31 Jul 2022.



Submit your article to this journal [↗](#)



Article views: 6



View related articles [↗](#)



View Crossmark data [↗](#)



In Silico Study of Selected Natural Products as SARS-CoV-2 M^{Pro} Binder: Molecular Docking and Molecular Dynamics Simulation

Bochra Bencheikh^a, Mouna Cheriet^b , Rayenne Djemil^a, and Djameleddine Khatmi^c

^aLaboratory of Computational Chemistry and Nanostructures, University May 8, 1945, Guelma, Algeria;

^bMedicinal Plants Research Unit (URPM.3000, Laghouat) Attached to the Biotechnology Research Center (CRBt 25000, Constantine), Laghouat, Algeria; ^cInstitute of Molecular Sciences ISM2, Marseille, France

ABSTRACT

The SARS-CoV-2 virus is the causative agent of COVID-19 still pose significant threat to humanity. Discovery of an effective medication regime against this infection still under development. In the present study the potentiality of selected naturally occurring compounds as a SARS-CoV-2 M^{Pro} binder were investigated employing molecular docking and dynamics simulation. The Quantum chemical calculations were used to optimize the 3 D geometry of hits. Known inhibitor profiles were used for comparison purposes. The values of Binding Energy showed that the majority of natural ligands have better affinity to M^{Pro}, compared to drugs used to cure SARS-CoV-2. The biological and pharmacological properties of these ligands were on the same side as the docking and molecular dynamic results.

ARTICLE HISTORY

Received 3 January 2022
Accepted 24 June 2022

KEYWORDS

SARS-CoV-2 M^{Pro}; inhibition;

molecular docking;
molecular dynamic
simulations

1. Introduction

Coronaviruses (CoVs) are etiologic agents of severe infections in both animals and people, such that they cause illness in the respiratory and digestive systems. Earlier Coronaviruses inspections revealed that Coronaviruses affect various animals including reptiles, bird species and mammals.¹ The unique species of Coronaviruses was discovered at the end of 2019 and it is initially dubbed 2019-nCoV after an outbreak in Wuhan, China.² On January 30, 2020, the World Health Organization's Emergency Committee recognized an eruption in China as a Public Health Emergencies of International Concern (PHEIC).³ WHO officially designated this Coronaviruses as COVID-19 on February 11, 2020.⁴ Currently no specific treatments are available, and the researchers are in progress.⁵ However, the fact that the developed vaccinations are now being supplied, a mass vaccination campaign will take a long time to be established. As a result, efforts to develop safe and efficient anti-COVID-19 drugs must be continued.

Biologically active natural molecules are part of the defense system against infectious diseases, e.g., flavonoids, alkaloids and terpenes.⁶ Natural products pose a significant role for identifying bioactive molecules.⁷ Several biomolecules have been tested against COVID-19 with its different receptors.^{8–14} SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) is the virus responsible for Covid-19. There are proteins in SARS-CoV-2 that regulate how the virus multiplies and avoids the host's immune system. Initially, these nonstructural proteins appeared as two major polyproteins. One of the most attractive targets for inhibiting viral replication is SARS-CoV-2

M^{Pro} (also known as 3CL^{Pro}).¹⁵ The coronavirus's main protease M^{Pro} is a key CoV enzyme, which plays a pivotal role in mediating viral replication and transcription.^{16,17}

The M^{Pro} of 2019-nCoV shares 96% similarity with the M^{Pro} of the SARS-CoV.^{18–20} The M^{Pro} in CoV is essential for the proteolytic maturation of the virus and has been examined as a potential target protein to prevent the spread of infection by inhibiting the cleavage of the viral poly-protein.^{21,22} The discovery of the M^{Pro} protease structure in SARS-CoV-2 provides an excellent opportunity to identify potential drug candidates for treatment, which will be our target in this study.

Theoretical techniques play a critical role in searching for treatments for this disease.²³ Among these techniques, molecular docking and molecular dynamics simulations used in this study.

We investigate the inhibition of a series of compounds of natural origin and two medications (Antimalarial) with M^{Pro} as a receptor for SARS-CoV-2 treatment. Since the main protease of SARS-CoV-2 (M^{Pro}) plays a crucial role in the viral replication process, researchers will be able to use the findings of this investigation to discover an effective SARS-CoV-2 medication.

2. Computational methodology

2.1. Ligand preparation

In this paper, we looked at natural inhibitors such flavonoids and coumarins, as well as two anti-malarial medications as references. The Molecular Mechanics Force Field (MM) was used to optimize the three-dimensional (3D) structures of molecules (Table 1). The results were then optimized using the semi-empirical AM1 method to determine the most stable conformations. Using Discovery Studio software,²⁴ the molecules are saved in pdb format. Before being used by AutoDock Vina,²⁵ the final molecule is saved in pdbqt format.

2.2. Protein preparation

The present study focused on the main proteases in Coronaviruses (3CL^{Pro}/M^{Pro}), especially PDB ID 6LU7,²⁶ as potential target proteins for SARS-CoV-2 treatment. 6LU7 is a SARS-CoV-2 M^{Pro} that has been restructured and repositioned in PDB and has been publicly.

Accessible since early February 2020. This code, with a resolution of 2.16, illustrates the 3D structure of M^{Pro}, which comprises of a single chain A and a single ligand. This structure has been well defined and resolved. The protein is converted to the pdbqt format (Figure 1).

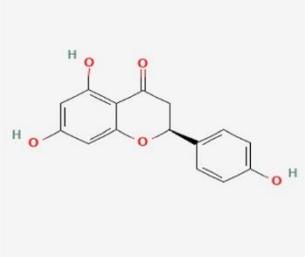
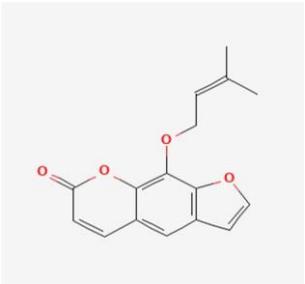
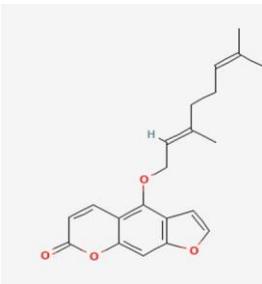
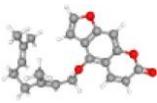
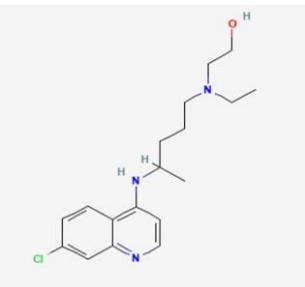
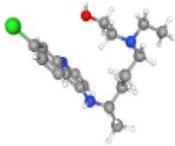
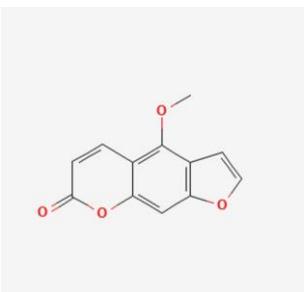
2.3. Molecular docking

All molecular docking calculations are performed using AutoDock Vina^{25,27}

The pdbqt file of SARS-CoV-2 main protease (M^{Pro}) was prepared using the AutoDock tools. All other docking parameters were maintained at their default settings. Based on the availability of resolved structures of M^{Pro} in association with inhibitors, the binding site of SARS-CoV-2 M^{Pro} was accurately determined. The box's dimensions (60 60 60) were chosen to adequately enclose the SARS-CoV-2 M^{Pro} active site. The grid centre's coordinates were 13.069, 9.740 and 68.490 (XYZ assignments, respectively). The Gasteiger method was used to calculate the atomic partial charges of the compounds.

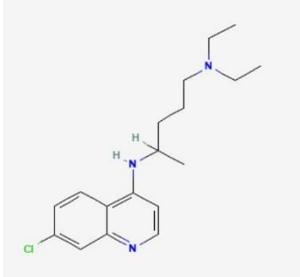
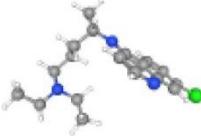
Re-docking was used to validate the protocol, and the bound standard ligand was re-docked in the originally attached active site of SARS-CoV-2 (6LU7). According to Sen et al. (2020),²⁸ the root mean square deviation (RMSD) limit for the ligand to binding site must not be greater than 2 Å. After completion of this process, the best pose of the docked co-ligand on its original crystallographic bound conformation was compared via superimposition, and a new RMSD value was

Table 1. 2D and 3D representations of the all ligands, the ligands structures were taken from PubChem server (www.pubchem.ncbi.nlm.nih.gov).

Compound name	2D Structure	3D Structure
Naringenin		
Imperatorin		
Bergamotin		
Hydroxychloroquine		
Bergapten		

(continued)

Table 1. Continued.

Compound name	2D Structure	3D Structure
Chloroquine		

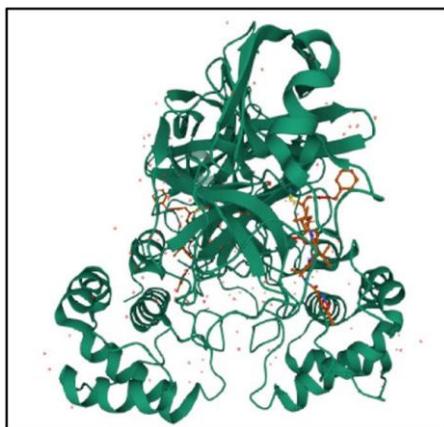


Figure 1. 3D structure of 2019-nCoV chimeric receptor-binding domain complexed with its receptor M^{Pro} (PDB ID: 6vw1). The structure was taken from Protein Data Bank online server (<https://www.rcsb.org/>).

calculated. M^{Pro} (6LU7) has an RMSD of 0.7333 Å, which is within acceptable limits (see Figure 2).

2.4. Druglikeness, ADME/toxicity qualities

The integration of ADMET (absorption, distribution, metabolism, excretion and toxicity) studies into the early phases of drug discovery has been universally implemented within the pharmaceutical industry.²⁹ The ultimate goal of computational research into ADMET prediction is to be able to identify compounds liable to later stage failure before they are even synthesized, bringing even greater efficiency benefits.³⁰ Many potential therapeutic agents fail to reach the clinic trials because of their unfavorable (ADMET) parameters;³¹ also it is not checking the druglikeness.

Drug likeness and ADME/toxicity analysis were performed for the selected ligand. The structure of the selected ligand was screened to predict whether the ligands obey Lipinski's rule of five.³² In order to confirm whether the physicochemical properties like molecular weight, hydrogen bond donor, hydrogen bond acceptor, molar refractivity, etc. of ligands follow Lipinski's rule of five or not.³³ SWISS-ADME server (<http://www.swissadme.ch/>) was used to evaluate the Ghose, Veber,³⁴ Egan,³⁵ Muegge rules, Polar surface area (TPSA) and number of rotatable.³³ The ADME/T provides absorption, distribution, metabolism, excretion, and toxicity of the given compounds. The pharmacodynamics properties such as, Caco-2 permeability, mutagenicity, carcinogenicity, Cytochrome P (CYP) inhibitory capability, human intestinal absorption, and blood brain

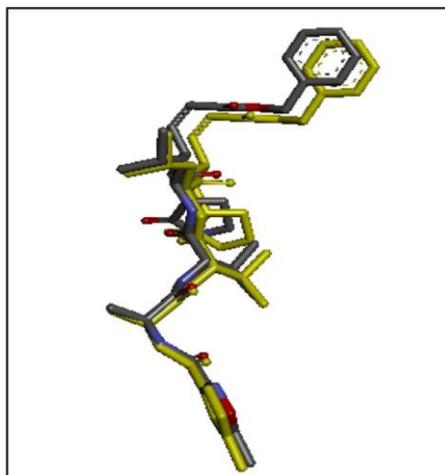


Figure 2. RMSD of natural and complexed ligand.

barrier permeability were analyzed. *In silico* toxicology aims to complement existing toxicity tests to predict toxicity, prioritize chemicals, guide toxicity tests, and minimize late-stage failures in drug design.

2.5. Molecular dynamics simulation

The MD simulations of 50 ns were achieved under standard conditions. CHARMM36 force field was applied to achieve topology attributes of protein and ligands.³⁶ The Berendsen thermostat³⁷ controlled the temperature of the simulations. The electrostatic interactions were treated by the particle mesh Ewald (PME) method with a cutoff distance of 1.2 nm; and the van der Waals interactions were modeled using a cutoff of 1.4 nm.³⁸ The whole system was relaxed using the steepest descent algorithm. The classical Newton's equations of motion were integrated using the Leap-Frog algorithm and the Berendsen weak-coupling method.³⁹ The trajectory coordinates were recorded every 500 steps with a time step of 1 fs. All MD simulations were carried out using the GROMACS package version 2020. Root-mean-square-deviation (RMSD) calculations were made to identify the sets sharing the same conformational characteristics. While, calculations of the root-mean-square fluctuation (RMSF) trajectories were used to determine the more mobile parts of the system, especially those involved in the interaction between protein and ligands. The radius of gyration (Rg) was used to assess the compactness of the protein as well as its stability. At last, solvent accessible surface area (SASA) was used in order to retain the protein folding and stability.

3. Results and discussion

3.1. Molecular docking

3.1.1. Binding energies: Ligands' affinity for M^{Pro} receptor

Bergamotin, naringenin, chloroquine, hydroxychloroquine, Bergapten, and Imperatorin have been investigated as potential SARS-CoV-2 M^{Pro} inhibitors (Table 2).

The binding energy (DG_{binding}) was estimated as follows:

DG_{binding} energy calculated as follows equation (1):

$$DG_{\text{binding}} = DG_{\text{complex}} - DG_{\text{ligand}} - DG_{\text{protein}} \quad (1)$$

Table 2. Docking results of 6LU7-Ligand.

Inhibitor	Affinity (kcal/mol)	Hydrogen Bond	Hydrophobic Bond	Electrostatic Bond
Bergamotin	-7.8	CYS145, GLY143	MET165, HIS163	HIS41
Naringenin	-7.5	ASP187, GLU166	MET49, CYS145	//////////
Imperatorin	-6.8	GLY143, HIS164 CYS145	MET49, MET165	//////////
Hydroxychloroquine	-6.2	HIS164, CYS145 GLY143, GLU166	CYS145	HIS163
Bergapten	-6.2	HIS 41, TYR54	MET165, HIS41	//////////
Chloroquine	-5.8	GLY143, GLU166 HIS164, CYS145	CYS145	HIS163

The binding energies obtained from docking 6LU7 with the all ligands Bergamotin, Naringenin, Imperatorin, Hydroxychloroquine, Bergapten and Chloroquine were -7.8, -7.5, -6.8, -6.2, -6.2, -5.8 Kcal/mol, respectively.

The docking analysis revealed the potential for inhibition of the studied ligands, classified by affinity (DG); Bergamotin > Naringenin > Imperatorin > Hydroxychloroquine $\frac{1}{4}$ Bergapten > Chloroquine.

In the general case, the interactions established between ligand and receptor (protein) are weak such as VDW bonds, electrostatic and H bonds. In order to understand the type of interaction between ligand and M^{Pro} , we analyzed the interactions of obtained complex.

3.1.2. Interaction analysis between M^{Pro} and ligands

The results obtained show that all the ligands interact with the catalytic dyad HIS41 and CYS145.

Bergamotin had the highest affinity for M^{Pro} receptor (PDB ID 6LU7), this ligand is stabilized by two hydrogen interactions one with CYS145 and the other with GLY143 and two hydrophobic bonds formed with residues MET165 and HIS163, as well as an electrostatic interaction with HIS 41 (Figure 3).

Two hydrophobic contacts generated with the residues MET49 and CYS145, as well as two hydrogen bonds with ASP187 and GLU166, stabilize the naringenin ligand (Figure 4).

The imperatorin ligand showed three hydrogen interactions with GLY143, HIS164 and CYS145, as well as two hydrophobic interactions with MET49 and MET165.

Furthermore, the hydroxychloroquine molecule is maintained by hydrogen bonds generated by residues GLY143, HIS164, CYS145 and GLU166, as well as hydrophobic and electrostatic contacts formed by CYS145 and HIS163 (Figure 5).

Bergapten forms two hydrogen interactions with HIS41 and TYR54, and two hydrophobic contacts with the residues MET165 and HIS41, the latter of which is one of the catalytic dyad's residues.

The chloroquine ligand has revealed four hydrogen bonds with the residues GLY143, GLU166, HIS164 and CYS145, one hydrophobic CYS145, as well as one hydrophobic CYS 145 bond and one electrostatic bond with HIS163 (Figure 6).

The two residues of the catalytic dyad, CYS145 and HIS41, are involved in the various interactions (H bond and an electrostatic bond) for the Bergamotin ligand that has the best affinity toward the M^{Pro} receptor (the lowest binding energy). In the other ligands, only one residue of the catalytic dyad is implicated.

3.2. Druglikeness properties of ligands

This test is performed to facilitate the creation of new drug molecules, the molecules of which must comply with the conditions of the following five Lipinski's rule: molecular weight: ≤ 500 , number of hydrogen bond donors: ≤ 5 , number of hydrogen bond acceptors: ≤ 10 , lipophilicity

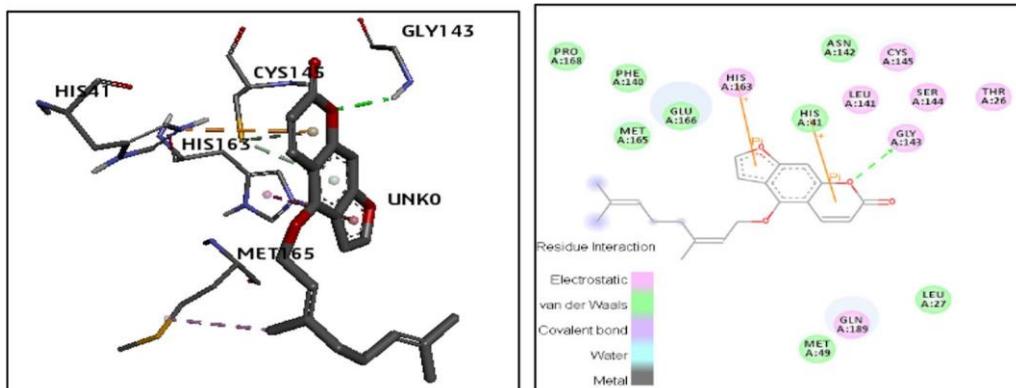


Figure 3. Interaction Mode of 6LU7-Bergamotin 2D, 3D.

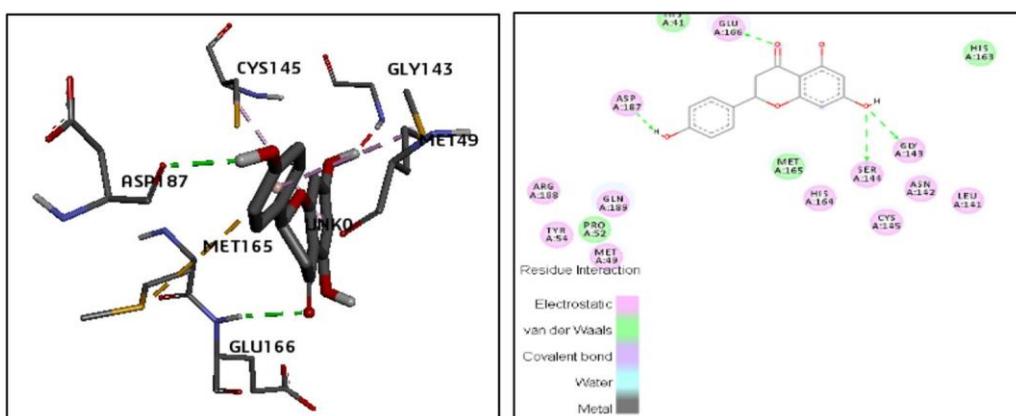


Figure 4. Interaction Mode of 6LU7-Naringenin 2D, 3D.

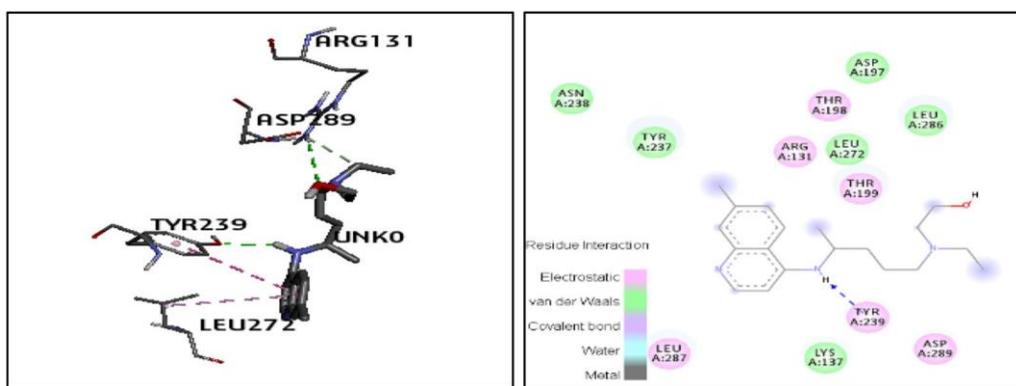


Figure 5. Interaction Mode of 6LU7-hydroxychloroquine 2D, 3D.

(expressed in LogP): $\bar{5}$ and molar refractivity from 40 to 130.³² The results of this study are shown in the Table 3.

Naringinin, imperatorin, bergamotin, hydroxychloroquine, bergapten and chloroquine have the following molecular weights: (293, 273, 340, 335, 234 and 226) g/mol, and they have a molecular weight 500 g/mol, and also respectively present the following values of the topological polar surface (TPSA)(97, 53, 44, 48, 48, 44, and 64) Å². The lowest TPSA values give good results.

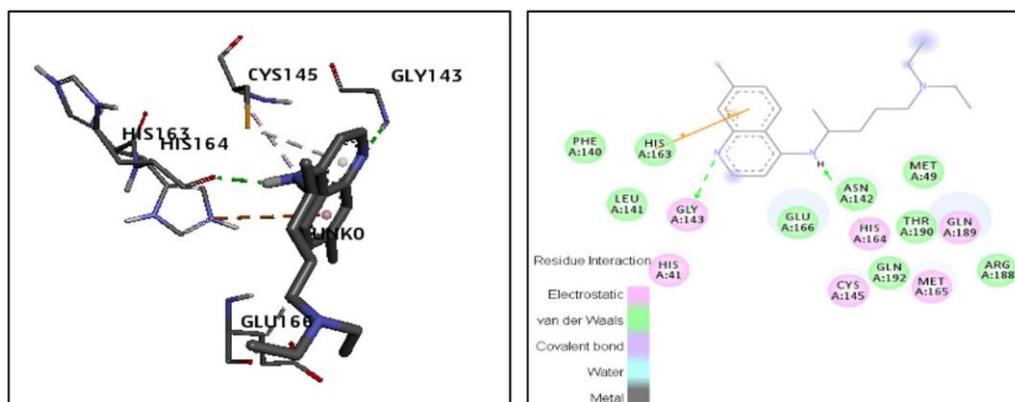


Figure 6. Interaction Mode of 6LU7- Chloroquine 2D, 3D.

Table 3. List of the results of the ligands druglikeness properties.

Drug likeness propriete	Narinnngin	Imperatorin	Bergamotin	Hydroxychloroquine	Bergapten	Chloroquine
Molecular weight g/mol	293	273	340	335	234	226
Consensus Log Po/w	0.44	2.47	4.76	3.29	1.06	0.6
Log S (ESOL)	-1.67	-3.52	-5.2	-3.91	-1.77	-0.6
Num. H-bond acceptors	5	4	4	3	4	4
Num. H-bond donors	4	0	0	2	0	4
Molar refractivity	74	74	99	98	56	58
Lipinski	YES	YES	YES	YES	YES	YES
Ghose	YES	YES	YES	YES	YES	YES
Veber	YES	YES	YES	YES	YES	YES
Egan	YES	YES	YES	YES	YES	YES
Muegge	YES	YES	NO	YES	YES	YES
Bioavailability score	0.55	0.55	0.55	0.55	0.55	0.55
TPSA	97	44	48	48	44	64
Num. rotatable bonds	4	3	6	9	1	9
Solubility	-1.67	-3.52	-5.2	-3.91	-1.77	-0.6

By comparing the lipophilicity (LogP) values of our ligands series, we observe that they have values less than five, indicating that they have produced excellent results and can be easily absorbed in the body.

However, the LogS solubility values of the ligands tested are between -4.06 and 0.25. Bergamotin has the highest value; the lowest values are always appreciable.

All the ligands have a number of hydrogen bonding donors: 2 and a number of hydrogen bond- ing acceptors: 10 and reactivity values between 35.75 and 130. Therefore, we can say that the five Lipinski's rule was verified. Veber's rule, which represents the oral bioavailability of a possible drug molecule; it is verified for all ligands studied.

We also note that another Egan's rule that defines the absorption of the drug molecule is verified for all ligands; on the other hand, the two rules of Ghose's, Muegge's,⁴⁰ Veber,³⁴ Egan³⁵ and Muegge⁴¹ are verified.

We also notice that some ligands give the reproductive effects, the irritant, Tumorigenic, Mutagenic properties and all the ligands showed a bioavailability score of 0.55, and good values of druglikeness score, Solubility values and drug score.

3.3. ADMET/drug results

The ADME tests were performed to determine the pharmacological and pharmacodynamic properties of a medication in a biological system, and the results are listed in Table 4.

Table 4. The ADME/T test results of ligands (various pharmacokinetic and pharmacodynamic properties).

Properties		Naringin	Imperatorin	Bergamotin	Hydroxychloroquine	Bergapten	Chloroquine
Absorption	HIA (%)	87.31	97.72	97.39	94.66	98.16	98.05
	Caco-2 cell (nm sec ⁻¹)	10.52	56.10	49.82	46.08	43.47	56.61
	MDCK	44.63	228.71	0.12	45.10	52.91	0.29
Distribution	Skin Permeability	−4.18	−3.16	−2.49	−3.08	−3.70	−2.53
	BBB	0.59	2.70	0.36	2.28	2.07	7.73
Metabolism	Cyp-2C19-inhibition	Inhibitor	Inhibitor	Inhibitor	Non	Inhibitor	Non
	Cyp-2C9-inhibition	Inhibitor	Inhibitor	Inhibitor	Non	Inhibitor	Non
	Cyp-2D6-inhibition	Non	Non	Non	Inhibitor	Non	Inhibitor
	Cyp-2D6-Substrate	Non	Non	Non	Substrate	Non	Substrate
	Cyp-3A4-inhibition	Inhibitor	Inhibitor	Inhibitor	Non	Inhibitor	Non
	Cyp-3A4-substrate	Non	Weakly	Substrate	Substrate	Non	Substrate
toxicity	Mutagenicity (Ames test)	Mutagen	mutagen	mutagen	mutagen	Mutagen	Mutagen
	Carcinogenicity mouse	Negative	Negative	Positive	Negative	Negative	Negative
	Carcinogenicity rat	Positive	Negative	Negative	Negative	Positive	Positive
	HERG inhibition	Medium risk	Medium risk	Medium risk	Medium risk	Medium risk	Medium risk

The properties of ADME/Tox (absorption, distribution, metabolism, excretion/toxicology) of all ligands were calculated using the online server PreADMET.

This server calculates pharmacokinetic properties like: Human Intestinal Absorption (HIA), Cell Permeability, Caco-2 in Vitro (PCaCO₂), Kidney Maden Darby Cell Permeability (PMDCK)), Permeability (PSkin), Protein Binding plasma (PPB) and hematomenigeal penetration are shown in Table 4.

To predict absorption and distribution parameters, Caco-2 cell parameter employed to determine compound permeability. Caco-2 cell parameters were ranged from 4 to 70 nm/sec including mean permeability. Human intestinal absorption (HIA) can predict percent of absorption of human intestine (% HIA) and plasma protein binding (% PPB) is used to determine the degree to which drugs bind to proteins in the blood. The bound drug value less than 90% is, the more efficiently it can traverse cell membranes or diffuse.

As a result of these data and values, the ligands have higher human intestinal absorption (HIA) scores. A higher HIA value indicates the compound might be better absorbed from the intestinal tract on oral administration, and then caco-2 values agree with the values obtained (value between 4 and 70 nm/sec). Thus, it is noted that all of the ligands had a medium permeability. BBB values (CBrain/CBlood) and tuning compounds that have values larger than 1 (CBrain/CBlood > 1) are considered active in the CNS may cause side effects while compounds with values less than 1 (CBrain/CBlood < 1) are classified as inactive in the CNS.

Therefore, ligands Bergapten, imperatorin, hydroxychloroquine and chloroquine had values higher than 1 (CBrain/CBlood > 1), indicating that they are active in the CNS and can cause side effects, while naringin and bergamotin had values less than 1 (CBrain/CBlood < 1), being considered inactive in the CNS.

The AMES toxicity test is used to determine whether a substance is mutagenic or not; the majority of the test ligands are mutagenic. Prediction of negative carcinogenicity indicates that there is evidence of carcinogenic activity while positive carcinogenicity shows that the substance under investigation is not carcinogenic. In addition, inhibition of HERG for all ligands present medium risk.

From the results of ADMET, it can be concluded that the biological and pharmacological properties of natural ligands are plausible, allowing them to have acceptable bioavailability and therefore to be attractive options against SARS-CoV-2.

3.4. Molecular dynamics simulations

Molecular dynamics (MD) simulation is the most important used method for evaluating the micro interaction between ligand structure and protein. To gain a better understanding of the

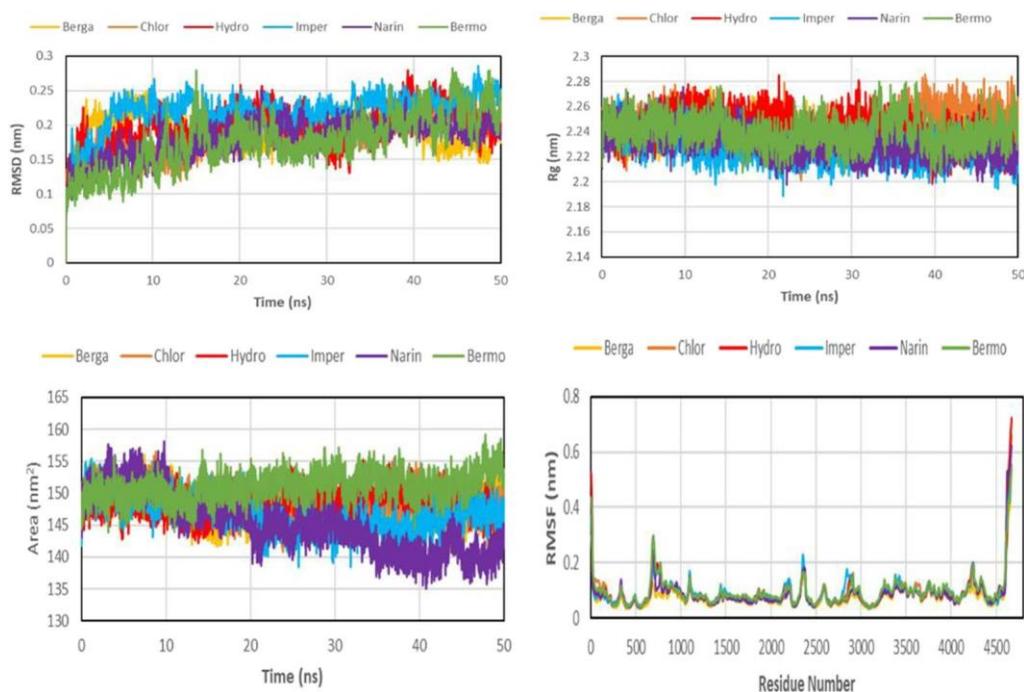


Figure 7. Comparison and detail representation of (A) RMSD: Root mean square deviations, (B) RMSF: Root mean square fluctuations, (C) Rg: Radius of gyration and (D) SASA: solvent accessible surface area, over the time of 50 ns of reference M^{Pro} and all complexes.

stability and dynamics of all complexes, 50 ns MD simulations were used. RMSD, RMSF, Rg, and SASA were evaluated after MD simulations.

The Root mean square deviations (RMSD) curves that all complexes remained stable in^{15–40} ns period time of the simulation with average values ranged from 0.18 to 0.22 nm. The results of RMSD signify that the binding between M^{Pro} and all of these ligands are steady. The Rayon of gyration (Rg) was used to assess the complexes compactness and it was revealed that Rg off all systems was steady with the RMSD of system. This depicts that protein was stable and compressed during throughout the 50 ns of time. The average Rg values are ranged from 2.23 to 2.27 nm. The Root-mean-square fluctuation RMSF were measured in order to further compute the residual flexibility over 50 ns of time. The total fluctuations of all the complexes are roughly the same. The average RMSF values are ranged from 0.079 to 0.098 nm. The Solvent accessible surface area (SASA) is alternative method in order to retain the protein folding and stability. The calculated SASA values for wild type and mutants are displayed in [Figure 7](#). The average SASA values for complexes were found to be 147 nm² (Berga), 148 nm² (Chlor), 148 nm² (Hydro), 146 nm² (Imper), 145 nm² (Narin) and 151 nm² (Bermo). Signifying that there were no notable differences in the available area of all the systems throughout the simulation process.

4. Conclusion

In order to fight against SARS-CoV-2, researchers have carried out variety of approaches, especially by trying to inhibit proteases essential to the virus's life cycle, by using drugs and natural compounds. In this work, we therefore studied the inhibition of SARS-CoV-2 by molecular docking and molecular dynamics with several natural compounds, the main goal was to find novel molecules that can inhibit this virus. We have docked the Main-protease (6LU7) as SARS-CoV receptor with a series of flavonoids and coumarins compared to two drugs (CQ and HCQ). The

study of their interactions with this enzyme (M^{Pro}), followed by filtration in using Lipinski's rule and ADMET properties was crucial. We classified the ligands based on their affinities. The results of docking revealed that the inhibitory power of the natural compounds studied with the M^{Pro} was significant. From the results of ADMET, it can be concluded that the majority of the natural ligands studied have interesting biological and pharmacological properties, allowing them to have better bioavailability. Molecular dynamics (MD) simulation was performed in our study in order to better understanding of the stability and dynamics of all complexes. As a preliminary result, these natural ligands appear to be promising potential candidates for SARS-CoV-2 inhibition.

References

1. Yashpal Singh Malik, Shubhankar Sircar, Sudipta Bhat, Khan Sharun, Kuldeep Dhama, Maryam Dadar, Ruchi Tiwari, and Wanpen Chaicumpa, "Emerging Novel Coronavirus (2019-nCoV) – Current Scenario, Evolutionary Perspective Based on Genome Analysis and Recent Developments," *The Veterinary Quarterly* 40, no. 1 (2020): 68–12. doi:10.1080/01652176.2020.1727993.
2. P. R. Lee, and P. I. Hsueh, "Emerging Threats from Zoonotic Coronaviruses-from SARS and MERS to 2019-nCoV," *Journal of Microbiology, Immunology and Infection* 53, no. 3 (2020): 1–3.
3. A. J. Rodr'iguez-Morales, K. MacGregor, S. Kanagarajah, D. Patel, and P. Schlagenhauf, "Going Global Travel and the 2019 Novel Coronavirus," *Travel Medicine and Infectious Disease* 33 (2020): 101578. doi:10.1016/j.tmaid.2020.101578.
4. World Health Organization (WHO). *Novel Coronavirus (2019-nCoV) WHO Bullno*, 2020. 1–7.
5. D. Sen, S. Bhaumik, P. Debnath, and S. Debnath, "Potentiality of Moringa Oleifera against SARS-CoV-2: Identified by a Rational Computer Aided Drug Design Method," *Journal of Biomolecular Structure and Dynamics* (2021): 1–18. doi:10.1080/07391102.2021.1898475.
6. Ziyad Tariq Muhseen, Alaa R. Hameed, Halah M. H. Al-Hasani, Muhammad Tahir Ul Qamar, and Guanglin Li, "Promising Terpenes as SARS-CoV-2 Spike Receptor-Binding Domain (RBD) Attachment Inhibitors to the Human ACE2 Receptor: Integrated Computational Approach," *Journal of Molecular Liquids* 320 (2020): 114493. doi:10.1016/j.molliq.2020.114493.
7. A. Kundu, D. Sen, and T. K. Chatterjee, "In-Vivo Antimalarial Study of PITC2 of *Pluchea indica* (L.) Less against *Plasmodium berghei* and Plasmodium Yolli Model," *Pharmacologyonline* 3 (2010): 817–23.
8. D. Juckel, J. Dubuisson, and S. Belouzard, "Les Coronavirus, Ennemis Incertains," *Médecine/Sciences* 36, no. 6–7 (2020): 633–41. p doi:10.1051/medsci/2020113.
9. M. T. Ul Qamar, S. M. Alqahtani, M. A. Alamri, and L.-L. Chen, "Structural Basis of SARS-CoV-2 3CLpro and anti-COVID-19 Drug Discovery from Medicinal Plants," *Journal of Pharmaceutical Analysis* 10, no. 4 (2020): 313–9. doi:10.1016/j.jpah.2020.03.009.
10. D. S. N. B. K. Prasanth, Manikanta Murahari, Vivek Chandramohan, Gangadharappa Bhavya, Atmakuri Lakshmana Rao, Siva Prasad Panda, G. S. N. Koteswara Rao, Guntupalli Chakravarthi, Nayudu Teja, Peddireddy Suguna Rani, et al, "In-Silico Strategies of Some Selected Phytoconstituents from *Melissa Officinalis* as SARS CoV-2 Main Protease and Spike Protein (COVID-19) Inhibitors," *Molecular Simulation* 47, no. 6 (2021): 457–70. doi:10.1080/08927022.2021.1880576.
11. M. K. Tripathi, P. Singh, S. Sharma, Tej P. Singh, A. S. Ethayathulla, and Punit. Kaur, "Identification of Bioactive Molecule from *Withania Somnifera* (Ashwagandha) as SARS-CoV-2 Main Protease Inhibitor," *Journal of Biomolecular Structure and Dynamics* 39, no. 15(2021): 5668–5681.
12. M. A. A. Ibrahim, E. A. R. Mohamed, a, A. H. M. Abdelrahman, a, K. S. Allemailem, M. F. Moustafa, A. M. Shawky, A. Mahzari, A. Refdan Hakami, K. A. A. Abdeljawaad, and M. A. M. Atia, "Rutin and Flavone Analogs as Prospective SARS-CoV-2 Main Protease Inhibitors: In Silico Drug Discovery Study," *Journal of Molecular Graphics & Modelling* 105 (2021): 107904. doi:10.1016/j.jmgs.2021.107904.
13. M. A. A. Ibrahim, A. H. M. Abdelrahman, T. A. Mohamed, M. A. M. Atia, M. A. M. Al-Hammady, K. A. A. Abdeljawaad, E. M. Elkady, M. F. Moustafa, F. Alrumaihi, K. S. Allemailem, et al, "In Silico Mining of Terpenes from Red-Sea Invertebrates for SARS-CoV-2 Main Protease (Mpro) Inhibitors," *Molecules* 26, no.7 (2021): 2082. doi:10.3390/molecules26072082.
14. M. A. A. Ibrahim, A.H.M. Abdelrahman, T. A. Hussien, E. A.A. Badr, T. A. Mohamed, H. R. El-Seedi, P. W. Pare, T. Efferth, M.-E. F. Hegazy *In Silico* Drug Discovery of Major Metabolites from Spices as SARS-CoV-2 Main Protease Inhibitors. *Computers in Biology and Medicine* 2020; 126: 104046[PMC][33065388]
15. L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, and R. Hilgenfeld, "Crystal Structure of SARS-CoV-2 Main Protease Provides a Basis for Design of Improved Alpha-Ketoamide Inhibitors," *Science (New York, N.Y.)* 368, no. 6489 (2020): 409–12. doi:10.1126/science.abb3405.

16. H. Hofmann, and S. Pohlmann, "Cellular Entry of the SARS Coronavirus," *Trends in Microbiology* 12, no. 10 (2004): 466–72. doi:10.1016/j.tim.2004.08.008.
17. R. Hilgenfeld, "From SARS to MERS: crystallographic Studies on Coronaviral Proteases Enable Antiviral Drug Design," *The FEBS Journal* 281, no. 18 (2014): 4085–96. Sepdoi:10.1111/febs.12936.
18. Jiabao Xu, Shizhe Zhao, Tieshan Teng, Abualgasim Elgaili Abdalla, Wan Zhu, Longxiang Xie, Yunlong Wang, and Xiangqian Guo, "Systematic Comparison of Two Animal-to-Human Transmitted Human Coronaviruses: SARS-CoV-2 and SARS-CoV," *Viruses* 12, no. 2 (2020): 244. doi:10.3390/v12020244.
19. Z. Xu, C. Peng, Y. Shi, Z. Zhu, K. Mu, and X. Wang, "Nelfinavir Was Predicted to Be a Potential Inhibitor of 2019-nCoV Main Protease by an Integrative Approach Combining Homology Modelling," *Molecular Docking and Binding Free Energy Calculation* 1201 (2020): 0–2.
20. A. Zhavoronkov, V. Aladinskiy, A. Zhebrak, B. Zagribelnyy, V. Terentiev, D. S. Bezrukov, D. Polykovskiy, R. Shayakhmetov, A. Filimonov, P. Orekhov, et al., "Potential 2019-nCoV 3C-like Protease Inhibitors Designed Using Generative Deep Learning Approaches," *Preprint* 10 (2020): 26434.
21. R. X. Liu, B. Zhang, Z. Jin, and H. Yang, "The Crytal Structure of 2019-nCoV Main Protease in Complex with an Inhibitor N3," 2020. Available from: <https://www.rcsb.org/structure/6lu7>.
22. M. A. A. Ibrahim, K. A. A. Abdeljawaad, A. H. M. Abdelrahman, and M.-E F. Hegazy, () "Natural-like Products as Potential SARS-CoV-2 Mpro Inhibitors: *In-Silico* Drug Discovery," *Journal of Biomolecular Structure and Dynamics* (2021),39(15):5722–5734.
23. D. Douche, Y. Sert, S. A. Brandán, A. A. Kawther, B. Bilmez, N. Dege, A. El Louzi, Khalid. Bougrin, K. Karrouchi, and B. Himmi, "5-((1H-Imidazol-1-yl) Methyl) Quinolin-8-ol as Potential Antiviral SARS-CoV-2 Candidate: Synthesis, Crystal Structure, Hirshfeld Surface Analysis, DFT and Molecular Docking Studies," *Journal of Molecular Structure* 1232 (2021): 130005. doi:10.1016/j.molstruc.2021.130005.
24. D. B. Kitchen, H. Decornez, J. R. Furr, and J. Bajorath, "Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications," *Nature Reviews. Drug Discovery* 3, no. 11 (2004): 935–49. doi:10.1038/nrd1549.
25. O. Trott, and A. 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, no. 2 (2010): 455–61.
26. R. X. Liu, B. Zhang, Z. Jin, and H. Yang, "The Crytal Structure of 2019-nCoV Main Protease in Complex with an Inhibitor N3," 2020.
27. M. Gümüş, Ş. N. Babacan, Y. Demir, Y. Sert, I. Koca, and I. Gölçin, "Discovery of Sulfadrag-Pyrrole Conjugates as Carbonic Anhydrase and Acetylcholinesterase Inhibitors," *Archiv Der Pharmazie* 355, no. 1 (2022): e2100242. doi:10.1002/ardp.202100242.
28. D. Sen, P. Debnath, B. Debnath, S. Bhaumik, and S. Debnath, "Identification of Potential Inhibitors of SARS-CoV-2 Main Protease and Spike Receptor from 10 Important Spices through Structure-Based Virtual Screening and Molecular Dynamic Study," *Journal of Biomolecular Structure and Dynamics* 40, no. 2(2022): 941-962. doi:10.1080/07391102.2020.1819883).
29. H. Yu, and A. Adedoyin, "ADME-Tox in Drug Discovery: Integration of Experimental and Computational Technologies," *Drug Discovery Today* 8, no. 18 (2003): 852–61. doi:10.1016/S1359-6446(03)02828-9.
30. A. P. Beresford, H. E. Selick, and M. H. Tarbit, "The Emerging Importance of Predictive ADME Simulation in Drug Discovery," *Drug Discovery Today* 7, no. 2 (2002): 109–16. doi:10.1016/S1359-6446(01)02100-6.
31. M. A. Miteva, S. Violas, M. Montes, D. Gomez, and P. Tuffery, "FAF-Drugs: Free ADME/Tox Filtering of Compound Collections," *Nucleic Acids Research* 34 (2006): 738–44.
32. C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, "Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings," *Advanced Drug Delivery Reviews* 46, no. 1-3 (2001): 3–26. doi:10.1016/S0169-409X(00)00129-0.
33. A. Daina, O. Michielin, and V. Zoete, "SwissADME: A Free Web Tool to Evaluate Pharmacokinetics, Drug-Likeness and Medicinal Chemistry Friendliness of Small Molecules," *Scientific Reports* 7 (2017): 42717.
34. D. F. Veber, S. R. Johnson, H. Y. Cheng, B. R. Smith, K. W. Ward, and K. D. Kopple, "Molecular Properties That Influence the Oral Bioavailability of Drug Candidates," *Journal of Medicinal Chemistry* 45, no. 12 (2002): 2615–23.
35. W. J. Egan, K. M. Merz, and J. J. Baldwin, "Prediction of Drug Absorption Using Multivariate Statistics," *Journal of Medicinal Chemistry* 43, no. 21 (2000): 3867–77. doi:10.1021/jm000292e.
36. B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, et al., "CHARMM: The Biomolecular Simulation Program," *Journal of Computational Chemistry* 30, no. 10 (2009): 1545–614. doi:10.1002/jcc.21287.
37. H. J. Berendsen, J. V. Postma, W. F. van Gunsteren, A. R. H. J. DiNola, and J. R. Haak, "Molecular Dynamics with Coupling to an External Bath," *Journal of Chemical Physics* 81, no. 8 (1984): 3684–90. doi:10.1063/1.448118.

38. H. G. Petersen, "Accuracy and Efficiency of the Particle Mesh Ewald Method," *Journal of Chemical Physics* 103, no. 9 (1995): 3668–79. doi:[10.1063/1.470043](https://doi.org/10.1063/1.470043).
39. W. F. Van Gunsteren, and H. J. Berendsen, "A Leap-Frog Algorithm for Stochastic Dynamics," *Molecular Simulation* 1, no. 3 (1988): 173–85. doi:[10.1080/08927028808080941](https://doi.org/10.1080/08927028808080941).
40. A. K. Ghose, V. N. Viswanadhan, and J. J. Wendoloski, "A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. "A Qualitative and Quantitative Characterization of Known Drug Databases," *Journal of Combinatorial Chemistry* 1, no. 1 (1999): 55–68. doi:[10.1021/cc9800071](https://doi.org/10.1021/cc9800071).
41. I. Muegge, Sarah L. Heald, and David. Brittelli, "Simple Selection Criteria for Drug-like Chemical Matter," *Journal of Medicinal Chemistry* 44, no. 12 (2001): 1841–6. doi:[10.1021/jm015507e](https://doi.org/10.1021/jm015507e).