

**République Algérienne Démocratique et Populaire**  
**Ministère de l'enseignement supérieur et de la recherche scientifique**

**UNIVERSITE 08 MAI 1945 GUELMA**

**Faculté des Mathématiques et de l'Informatique et des Sciences de la  
Matière Département des Sciences de la Matière**

**Mémoire du projet de fin d'étude**



**Master 2**

**Filière : Chimie**

**Spécialité : Chimie Physique**

*Présenté par :*

*Amirouche Rayane*

*Saidani Sarra*

---

---

**Inhibiteurs du SARS-CoV-2 : Relation Quantitative  
Structure-Activité (QSAR)**

---

---

**Sous la direction de : Pr Djemil Rayenne**

**Juin 2022**

## Remerciements

En premier lieu, nous remercions *Allah* de nous avoir donné la force et la patience nécessaire pour achever ce travail de mémoire.

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

Nous remercions sincèrement : *Dr Djadi Nabila, Dr Cheriet Mouna, Pr Nigri Soraya, Dr Kraim Khairedine et Mme Bencheikh Bouchra.*

Nous remercions pour leurs aides.

Nous tenons à exprimer nos profondes reconnaissances à tous les membres de jury pour avoir accepté d'examiner ce travail.

Enfin, à tous ceux qui nous ont aidés de près ou de loin, que ce soit par leurs amitiés, leurs conseils ou leurs soutiens moraux, ils trouveront dans ces quelques lignes l'expression de nos remerciements les plus vifs

Merci à tous.



## ***Dédicace***

### ***À ma mère***

Ma raison d'être, ma raison de vivre, la lanterne qui éclaire mon chemin et m'illumine de douceur et d'amour

### ***À mon père***

En signe d'amour, de reconnaissance et de gratitude pour tous les soutiens et les sacrifices dont il a fait preuve à mon égard

### ***À mes très chères sœurs***

*Meryem et Hadjer*

Aucun mot ne pourra décrire vos dévouements et vos sacrifices

### ***A ma très chère binôme***

*Rayane*

Pour son encouragement et pour tous les bons moments qu'on a vécus ensemble

### ***À mes très chères amies***

*Ibtissem et chayma*

En témoignage de l'amitié sincère qui nous a liées et des bons moments passés ensemble.

### ***À ma très chère amie***

*Manel*, Merci pour tous vos conseils et d'être à mes côtés

### ***À mon ami***

*Moncef*, Merci beaucoup de nous avoir aidés pendant les cinq années

### ***A tous mes autres amies***

À mes collègues de la promotion de chimie physique, À tous les gens qui me connaissent, qui m'aiment et ceux qui me souhaitent la réussite et le bonheur dans ma vie. À tous les gens que j'aime sans exception.

Je vous dis :

*Merci Beaucoup*

*Sarra*



## ***Dédicace***

Je dédie ce travail aux personnes les plus chères dans ma vie, ***mes parents***, qui m'ont fourni d'une telle éducation, je vous remercie pour tous vos sacrifices, votre amour, votre soutien, c'est grâce à vous que je suis là aujourd'hui, je vous aime énormément.

**A mes plus chères frères** *Mohamed El Arbi* et *Aymen*, merci d'être à mes côtés durant tous les moments de ma vie, je vous souhaite tout le bonheur et la réussite.

### ***À mes très chères sœurs***

*Amina et Roma*

Pour meilleur moments qui passé ensemble

**A ma géniale binôme** *Sarra* la plus proche de moi, qu'on a partagé les plus bons moments durant ces cinq ans.

**A mes adorables amies** *Chayma* et *Ibtissem* avec lesquelles j'ai partagé des souvenirs inoubliables.

### ***À mon ami***

*Moncef*, Merci beaucoup de nous avoir aidés pendant les cinq années

**A ma meilleur Promo de chimie physique**, à toute personne que j'aime et que j'ai oublié de citer, je vous dédie ce mémoire.

Je vous dis :

***Merci Beaucoup***

*Rayane*

## Résumé

---

Dans ce travail, nous avons réalisé une étude QSAR sur une série des Anti-HIV dérivés de Dragon, à partir de molécules optimisées par la fonctionnelle B3LYP/6-31G(d, p) intégrée au Gaussian. Un modèle basé sur la régression linéaire multiple (RLM), a été développé par Molegro pour la prédiction des pIC50 de l'ensemble de vingt inhibiteurs. Le traitement statistique réalisé par Minitab montre que les paramètres statistiques obtenus pour les ensembles de calibration et de validation (coefficients de corrélation  $R=0.92$ , de détermination  $R^2=0.84$  et de prédiction  $Q^2=0.84$ ) font ressortir la qualité et la pertinence du modèle calculé. La validation rigoureuse a été considérée pour juger la stabilité et la capacité prédictive de ce modèle. La qualité de l'ajustement du modèle développé a été vérifiée par représentation des valeurs calculées en fonction de celles observées ( $R_0^2 = 0.84$ ).

**Mots clés :** QSAR, SARS-CoV-2, RLM, IC50, Anti-HIV

---

## Résumé

---

In this work, we performed a QSAR study on a series of propiophenone-derived Anti-HIV used against SARS-CoV-2. Descriptors were obtained by Dragon software, from molecules optimized by the Gaussian-integrated B3LYP/6-31G(d, p) functional. A model based on multiple linear regression (MLR), was developed by Molegro for the prediction of pIC50s of the set of twenty inhibitors. The statistical treatment performed by Minitab shows that the statistical parameters obtained for the calibration and validation sets (correlation coefficients  $R=0.92$ , determination coefficients  $R^2=0.84$  and prediction coefficients  $Q^2=0.84$ ) highlight the quality and relevance of the calculated model. Rigorous validation was considered to judge the stability and predictive capacity of this model. The goodness of fit of the developed model was verified by plotting the calculated values against the observed ones ( $R_0^2 = 0.84$ ).

**Key words:** QSAR, SARS-CoV-2, RLM, IC50, Anti-HIV

---

## Résumé

---

في هذا العمل ، أجرينا دراسة QSAR على سلسلة من مضادات فيروس نقص المناعة البشرية المشتقة من البروبيوفينون المستخدمة ضد SARS-CoV-2. تم الحصول على الواصفات بواسطة برنامج Dragon ، من جزيئات محسنة بواسطة وظيفة B3LYP / 6-31G (d, p) المدمجة في Gaussian. تم تطوير نموذج يعتمد على الانحدار الخطي المتعدد (MLR) من قبل Molegro للتنبؤ ب pIC50s لمجموعة من عشرين مثبطا. تظهر المعالجة الإحصائية التي أجرتها Minitab أن المعلمات الإحصائية التي تم الحصول عليها لمجموعات المعايرة والتحقق من الصحة) معاملات الارتباط  $R = 0.92$  ، ومعاملات التحديد  $R^2 = 0.84$  ومعاملات التنبؤ  $Q^2 = 0.84$  تسلط الضوء على جودة وأهمية النموذج المحسوب. واعتبر التحقق الدقيق من الصحة للحكم على استقرار هذا النموذج وقدرته على التنبؤ. تم التحقق من ملاءمة النموذج المطور عن طريق رسم القيم المحسوبة مقابل القيم المرصودة. ( $R_0^2 = 0.84$ )

الكلمات المفتاحية: QSAR, SARS-CoV-2, RLM, IC50, Anti-HIV:

---

## Liste Des Figures

---

N	Intitulé des Figures	Page
	Chapitre I : SARS-CoV-2	
<b>Figure 01</b>	Classification des coronavirus	<b>5</b>
<b>Figure02</b>	Structure du SARS-CoV-2	<b>6</b>
<b>Figure 03</b>	Organisation génomique du SARS-CoV-2	<b>6</b>
<b>Figure 04</b>	Variants du SARS- CoV-2	<b>8</b>
<b>Figure 05</b>	Mécanisme d'activité de ACE2	<b>9</b>
<b>Figure 06</b>	Structure chimique de la chloroquine et de l'hydroxychloroquine	<b>11</b>
<b>Figure 07</b>	Principe d'action des inhibiteurs des protéases	<b>12</b>
<b>Figure 08.</b>	Structures des flavonoïdes	<b>13</b>
<b>Figure 09</b>	Structures chimiques de quelques Terpènes	<b>14</b>
<b>Chapitre II : Relation Quantitative Structure-Activité (QSAR)</b>		
<b>Figure 01</b>	Présentation de la méthodologie de QSAR	<b>22</b>
<b>Chapitre III : Résultats et discussion</b>		
<b>Figure 01</b>	Structure générale des dérivés de la propiophénone	<b>39</b>
<b>Figure 02</b>	Variation des valeurs expérimentales en fonction des valeurs prédites	<b>48</b>

---



## Liste Des Tableaux

---

<b>N</b>	<b>Intitulé des Tableaux</b>	<b>Page</b>
<b>Chapitre I : SARS-COV-2</b>		
<b>Tableau 01</b>	Comparaison entre différents types de coronavirus	<b>4</b>
<b>Chapitre II : Relation Quantitative Structure-Activité (QSAR)</b>		
<b>Tableau 01</b>	Les descripteurs électroniques calculés dans les études QSAR	<b>25</b>
<b>Tableau 02</b>	Différents descripteurs, employés dans les études QSAR, basés sur la dimension.	<b>26</b>
<b>Chapitre III : Résultats et discussion</b>		
<b>Tableau 01</b>	Structures des dérivés de la propiophénone et valeurs de pIC50	<b>40</b>
<b>Tableau 02</b>	Paramètres statistiques MLR de 3 descripteurs	<b>44</b>
<b>Tableau 03</b>	Valeurs de test de Fisher et test de Student.	<b>45</b>
<b>Tableau 04</b>	Définition et la classification des descripteurs	<b>46</b>
<b>Tableau05</b>	Paramètres statistiques de test	<b>47</b>

---

## Liste Des Abréviations

Symbole	Nomenclature
1D	Une Dimension
2D	Deux Dimensions
3D	Trois Dimensions.
3CLpro	3C-Like proteinase.
<b>A</b>	
ACE	Angiotensin Converting Enzyme.
ACE 2	Angiotensin Converting Enzyme 2
ARN	Acide RiboNucléique.
ADME/ Tox.	Absorption, Distribution, Métabolisme, Excrétion/ Toxicologie
<b>B</b>	
BzEt	BenzoylEthane
B3LYP	Becke 3-parameter, Lee-Yang-Parr
<b>C</b>	
COVID-19	COronaVirus Disease 2019
CoV	CoronaVirus
$\alpha$ -CoV	Alpha CoronaVirus
$\beta$ -CoV	Beta CoronaVirus
$\gamma$ -CoV	Gamma CoronaVirus
CQ	ChloroQuine
<i>CBD</i>	<i>CannaBiDiol</i>
<i>CYP2C8</i>	<i>Cytochrome P450 family subfamily C Member 8</i>
<b>D</b>	
DFT	Théorie de la Fonctionnelle de Densité
<b>G</b>	
GCG	GalloCatéchine Gallate

## Liste Des Abréviations

<b>H</b>	
HSAB	Hard and Soft Acides and Bases
HCQ	HydroxyChloroQuine
<i>HCoV</i>	<i>Coronavirus Humain</i>
HIV	Human Immunodeficiency Virus
HOMO	Highest Occupied Molecular Orbital
<b>I</b>	
IC50	Concentration Inhibitrice médiane
<b>L</b>	
LOO	Leave-one-out
LTO	Leave-Two-out
LMO	Leave-Many-out
LUMO	Lowest Unoccupied Molecular Orbital
<b>M</b>	
MERS-CoV	CoronaVirus du Syndrome Respiratoire du Moyen-Orient
MoRSE	Molecule Representation of Structures based on Electron diffraction
M-pro	Main protease
<b>N</b>	
NS3	Non-Structural Proteins 3
<b>O</b>	
OMS	Organisation Mondiale de la Santé
ORF	Open Reading Frame
ORF1a / b	Open Reading Frame a/b.
<b>P</b>	
PIC50	Cologarithme décimal de la Concentration inhibitrice médiane
PLpro	papain-like protease.
PLS	Régression des moindres carrés partiels
<b>Q</b>	
QSAR	Relation Quantitative Structure-Activité

## Liste Des Abréviations

RMN	Résonance Magnétique Nucléaire.
RdRp	RNA-dependent RNA Polymerase
RBD	Receptor Binding Domain.
RLM	Régression Linéaire Multiple
RNLM	Régression Non Linéaire Multiple
<b>S</b>	
SARS	Severe Acute Respiratory Syndrome related CoronaVirus
<b>V</b>	
VDW	Van Der Waals
<b>W</b>	
WHIM	Weighted Holistic Invariant Molecular descriptors

# Sommaire

## Table de matière

<b>Titre</b>	<b>Page</b>
<b>Sommaire</b>	<b>I</b>
<b>Liste des figures</b>	<b>Ii</b>
<b>Liste des tableaux</b>	<b>Iii</b>
<b>Liste des abréviations</b>	<b>vi</b>
<b>Introduction générale</b>	<b>1</b>
<b>Chapitre I : SARS-COV-2</b>	
<b>1. Historique :</b>	<b>3</b>
<b>2. Types et origine :</b>	<b>4</b>
<b>3. Classification :</b>	<b>4</b>
<b>4. Structure :</b>	<b>5</b>
<b>5. Génome :</b>	<b>6</b>
<b>6. Les variants du sars-cov2 :</b>	<b>7</b>
<b>6.1. Le variant Alpha, ou variante anglaise, ou VOC2020, ou B.1.1.7, ou 20I/501Y.V1</b>	<b>7</b>
<b>6.2. Le variant Beta, ou variante sud- africain, ou B.1.351, ou 20H/501Y.V2 :</b>	<b>7</b>
<b>6.3. Le variant Gamma, ou variante brésilienne, ou 20J/501Y.V3 :</b>	<b>7</b>
<b>6.4. Le variant Delta, ou variante indienne, ou B.1.617 :</b>	<b>8</b>
<b>6.5. Le variant Omicron ou B.1.1.529</b>	<b>8</b>
<b>7. Récepteurs :</b>	<b>9</b>
<b>a. Enzyme de Conversion de l'Angiotensine 2 (ACE2) :</b>	<b>9</b>
<b>1. Définition :</b>	<b>9</b>
<b>2. Mécanisme d'activité :</b>	<b>9</b>
<b>b. Domaine de liaison au récepteur RBD :</b>	<b>10</b>
<b>8. Les Protéases</b>	<b>10</b>
<b>9.1. Définition</b>	<b>10</b>
<b>9.2. Rôle :</b>	<b>10</b>
<b>9. Inhibiteurs :</b>	<b>11</b>
<b>a. Inhibiteurs synthétiques (Médicaments) :</b>	<b>11</b>
<b>1. La chloroquine (CQ) et l'hydroxychloroquine (HCQ) :</b>	<b>11</b>
<b>2. Inhibiteurs de la protéase du VIH :</b>	<b>12</b>
<b>b. Inhibiteurs d'origine naturelle :</b>	<b>12</b>

## Sommaire

<b>1. Les flavonoïdes :</b>	<b>13</b>
<b>2 .Les terpènes</b>	<b>13</b>
<b>3. Les vitamines :</b>	<b>14</b>
<b>Références bibliographiques</b>	<b>15</b>
<b>Chapitre II : Relation Quantitative Structure-Activité (QSAR)</b>	
<b>1.Historique :</b>	<b>20</b>
<b>2.Définition :</b>	<b>21</b>
<b>3.Principe :</b>	<b>21</b>
<b>4. Stratégie globale :</b>	<b>21</b>
<b>5.Bases de données</b>	<b>23</b>
<b>6.Descripteurs</b>	<b>23</b>
<b>6.1.Définition :</b>	<b>23</b>
<b>6.2.Types des descripteurs</b>	<b>23</b>
<b>6.2.2. Descripteurs constitutionnels :</b>	<b>23</b>
<b>6.2.3. Descripteurs topologiques :</b>	<b>24</b>
<b>6.2.4.Descripteurs géométriques :</b>	<b>24</b>
<b>6.2.5.Descripteurs électroniques :</b>	<b>24</b>
<b>6.2.6. Descripteurs spectroscopiques :</b>	<b>25</b>
<b>7. Méthodes Statistiques</b>	<b>26</b>
<b>7.1. La régression linéaire multiple (RLM)</b>	<b>26</b>
<b>7.2. La méthode de régression des moindres carrés partiels (PLS)</b>	<b>27</b>
<b>7.3. Régression non linéaire multiple</b>	<b>27</b>
<b>8. Paramètres statistiques de la validation</b>	<b>28</b>
<b>8.1.Coefficient de corrélation (R)</b>	<b>28</b>
<b>8.2.Coefficient de détermination (R<sup>2</sup>)</b>	<b>28</b>
<b>8.3.Test Fischer (F)</b>	<b>29</b>
<b>8.4.Test de Student</b>	<b>30</b>
<b>8.5. Ecart type (s)</b>	<b>30</b>
<b>8.6.R<sup>2</sup> ajusté</b>	<b>30</b>
<b>8.7.Coefficient de prédiction (q<sup>2</sup>)</b>	<b>31</b>
<b>8.8.Coefficient de corrélation prédit Q</b>	<b>31</b>
<b>9. Techniques de validation :</b>	<b>32</b>
<b>9.1. Validation interne :</b>	<b>32</b>

## Sommaire

<b>9.2. validation externe :</b>	<b>32</b>
<b>Références bibliographiques</b>	<b>34</b>
<b>Chapitre III</b> <b>Résultats et Discussion</b>	
<b>1.Introduction:</b>	<b>38</b>
<b>2.Méthodologie générale d'une étude QSAR</b>	<b>38</b>
<b>3.Collecte des données :</b>	<b>39</b>
<b>3.1.Données expérimentales :</b>	<b>39</b>
<b>4.Traitements des données :</b>	<b>43</b>
<b>4.1. Préparation de l'ensemble de données :</b>	<b>43</b>
<b>4.1.1. Chem3D Ultra Bio 18.0</b>	<b>43</b>
<b>4.1.2. Gaussian 09</b>	<b>43</b>
<b>4.2. Calcul des descripteurs :</b>	<b>43</b>
<b>4.3. Modélisation :</b>	<b>44</b>
<b>5. Résultats et discussion :</b>	<b>44</b>
<b>5.1.Modèle MLR :</b>	<b>44</b>
<b>5.2. Qualité d'ajustement :</b>	<b>45</b>
<b>5.3.Analyse des résultats :</b>	<b>46</b>
<b>6.Interprétation:</b>	<b>46</b>
<b>7.Validation du modèle :</b>	<b>47</b>
<b>8.Conclusion :</b>	<b>49</b>
<b>Références bibliographiques</b>	<b>50</b>
<b>Conclusion générale</b>	<b>51</b>

# **Introduction générale**



Le SARS-CoV-2 est le virus responsable du COVID-19, un syndrome respiratoire aigu sévère. Ce dernier est responsable de la crise sanitaire à laquelle le monde est confronté depuis la fin de l'année 2019. Par conséquent ces deux dernières années ont connu un grand essor pour le développement de nouvelles pistes thérapeutiques [1].

Le traitement de cette maladie a été basé principalement sur certains médicaments antiviraux déjà largement utilisés telles que les anti-malaria, les Anti-HIV. En effet, plusieurs classes d'agents antiviraux comme la famille des inhibiteurs de protéase HIV de diverses structures ont été proposés et étudiés en tant qu'inhibiteurs du SARS-CoV-2[2].

Le développement de nouveaux médicaments anti-SARS plus efficaces et moins chères constitue un défi pour l'industrie pharmaceutique car c'est un processus long et complexe. L'intégration de nouvelles approches in silico était l'une des meilleures solutions qui aide à élaborer des modèles fiables afin de prédire de façon rapide l'activité avant même que celles-ci ne soient synthétisées [3].

QSAR (Quantitative structure Activity Relationship) est parmi les nouvelles techniques de modélisation, mettant en jeu des relations entre les structures des composés et leurs activités utilisant des méthodes d'analyses multidimensionnelles par apprentissage statistique afin de modéliser l'activité biologique [4].

L'objectif de ce travail 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 propiophénone inhibiteurs du SARS-CoV-2.

Le manuscrit est divisé en trois chapitres :

**Le premier chapitre** : est dédié à une recherche bibliographique sur le SARS-CoV-2

**Le second chapitre** : est consacré à la modélisation de QSAR Quantitative Structure-Activité.

**Le dernier chapitre** : rassemble l'essentiel de nos résultats de modélisation de l'inhibition du SARS-CoV-2 par les dérivés du propiophénone.

### Références :

1. T. Gupta, and S. K. Gupta, Int. Immunopharmacol. 2020.
2. A. Sharma, S. Tiwari, M. K. Deb, J. L. Marty. Int. J. Antimicrob. 2020.
3. M. Ghamali, S. Chtita, M. Bouachrine, T. Lakhlifi. Revue. Université Moulay Ismail, Faculté des Sciences, Meknès, Maroc, 2016.
4. M. Karelson, S. Victor Lobanov, R. Alan Katritzky. chemical reviews, 1996.

# **Chapitre I : SARS CoV- 2**

### 1. Historique :

En fin d'année 2019, une alerte associée à l'apparition de cas d'infections respiratoires sévères d'étiologie inconnue a été provoquée par les autorités chinoises. Cette alerte rapportait que dans la ville de Wuhan, un nombre significatif de personnes ayant atteint le marché de Huanan avaient été hospitalisés pour des infections pulmonaires graves d'étiologie inconnue, avec un nombre significatif de décès. Des investigations microbiologiques ont permis l'identification de l'agent responsable de l'infection début janvier 2020[1] .Il s'agissait d'un coronavirus qui n'avait, jusqu'alors, jamais été détecté chez l'homme. L'information a été diffusée rapidement, afin de permettre à chacun de se préparer à un possible risque pandémique, ce risque étant toutefois tempéré par les expériences précédentes d'émergence d'autres coronavirus d'origine zoonotique ayant entraîné des cas d'importation sans développement d'un phénomène pandémique SARS en 2003 et MERS-CoV depuis 2012 [2-4]. Le 7 Mars 2020, l'Organisation mondiale de la santé (OMS) a déclaré la pandémie mondiale du coronavirus-19.Ce nouveau coronavirus, a été isolé de l'épithélium respiratoire des patients. Il appartient à une nouvelle branche évolutive au sein du CoV. Le 11 février 2020, le nouveau coronavirus a été officiellement renommé "SARS-CoV-2" à partir de "2019-nCoV" [5]. La maladie causée par le SRAS-CoV-2 a été appelée« Corona Virus Disease 2019 » (COVID-19)[6]. Sur la base de la récente mise à jour épidémiologique de l'OMS, en date du 11 décembre 2021, cinq COV du SARS-CoV-2 ont été identifiés depuis le début de la pandémie :

**Alpha (B.1.1.7) :** premier variant préoccupant décrit au Royaume-Uni (UK) fin décembre 2020

**Beta (B.1.351) :** signalé pour la première fois en Afrique du Sud en décembre 2020

**Gamma (P.1) :** signalé pour la première fois au Brésil début janvier 2021

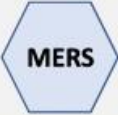


**Delta (B.1.617.2) :** signalé pour la première fois en Inde en décembre 2020

**Omicron (B.1.1.529) :** signalé pour la première fois en Afrique du Sud en novembre 2022[7]

### 2. Types et origine :

L'origine zoonotique (issu d'un hôte animal avec transmission à l'homme) des CoV est largement documentée. Les virus de cette famille infectent plus de 500 espèces de chiroptères (ordre de mammifères comprenant plus de 1 200 espèces de chauves-souris) qui représentent un réservoir important pour son évolution en permettant, entre autres, la recombinaison des génomes chez des animaux infectés simultanément par différentes souches virales [8-10]. Il est admis que la transmission zoonotique des CoV à l'homme passe par une espèce hôte intermédiaire, dans laquelle des virus mieux adaptés aux récepteurs humains peuvent être sélectionnés, favorisant ainsi le franchissement de la barrière d'espèce [11]. Les vecteurs de la transmission zoonotique peuvent être identifiés en examinant les relations phylogénétiques entre les nouveaux virus et ceux isolés à partir de virus d'espèces animales vivant dans les régions d'émergence.

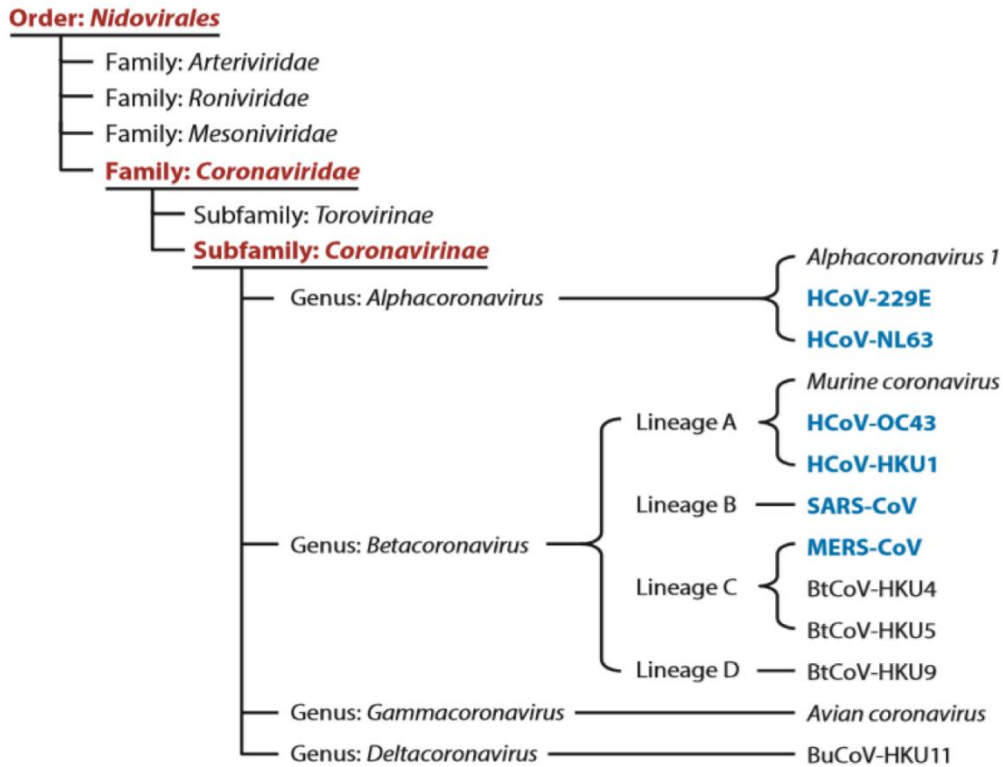
**Tableau 01** : Comparaison entre différents types de coronavirus

	 <b>MERS</b>	 <b>SARS</b>	 <b>Covid-19</b>
Virus species	<i>Middle East respiratory syndrome-related coronavirus</i>	<i>Severe acute respiratory syndrome-related coronavirus</i>	
Individuum	MERS-CoV	SARS-CoV	SARS-CoV-2
First reported	Saudi-Arabia (2012)	China (2002)	China (2019)
Current status	Active	No reports since 2004	Active
Infections	2521+	8098	3,500,000+
Fatality rate	34%	10%	~1%
Preventive/curative treatment	None	None	None

### 3. Classification :

l'ordre des Nidovirales, créé en 1996, regroupe actuellement trois familles, les Coronaviridae, les Arteriviridae, et les Roniviridae. Tous ces virus ont en commun l'organisation du génome ARN et la stratégie de réplication, mais ils diffèrent dans leur morphologie, la structure de leur capsid, et la taille de leur génome, qui va de 13 000 nucléotides pour les arterivirus à 31 000

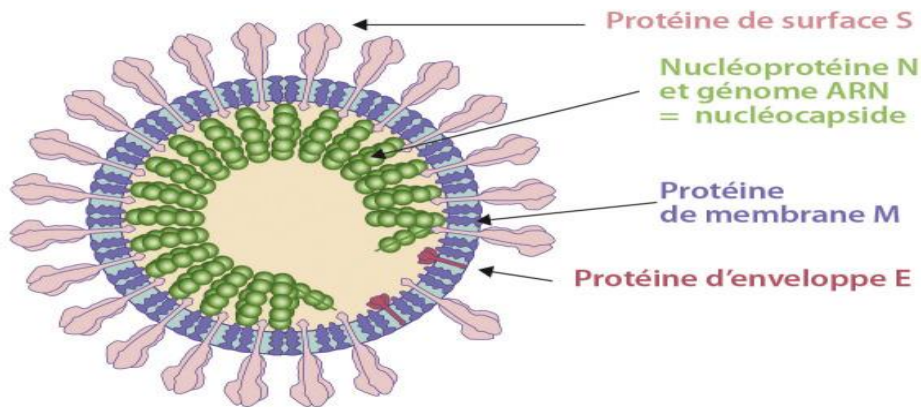
nucléotides pour les coronavirus. La famille des Coronaviridae est constituée de deux genres, les coronavirus et les torovirus. Parmi les Nidovirales, seul le genre coronavirus comprend des virus identifiés chez l'homme [12]. lesCoVs ont été organisés en quatre groupes : alpha coronavirus ( $\alpha$ -CoV), beta coronavirus ( $\beta$ -CoV), gamma coronavirus ( $\gamma$ -CoV) et delta coronavirus ( $\delta$ -CoV) [13,14].



**Figure 01** : Classification des coronavirus

#### 4. Structure :

Les coronavirus sont des virus enveloppés, plutôt sphériques, d'un diamètre compris entre 80 et 200 nm. Les protéines S (spike) forment une large couronne à leur surface, d'où le préfixe latin corona. Les protéines N, étroitement liées à l'acide ribonucléique (ARN) génomique, forment la nucléocapside. Les protéines M et E constituent la matrice et l'enveloppe [15-17].



**Figure 02 :** Structure du SARS-CoV-2.

### 5. Génome :

Il s'agit d'une molécule d'ARN monocaténaire linéaire non segmentée, de polarité positive. Le génome comprend deux régions non codantes en 5' et en 3'. La partie codante est divisée en plusieurs sections. Les deux premiers tiers du génome sont constitués de deux grandes régions chevauchantes, *open reading frame* (ORF) 1a et ORF1b, codant le complexe de réplication-transcription, dont le gène *RNA-dependent RNA Polymerase* (*RdRp*) qui code l'ARN polymérase ARN-dépendante. Le dernier tiers du génome code les protéines de structure (*S*, *E*, *M*, *N*) et des protéines non structurales variables selon les espèces de coronavirus [18].



**Figure 03 :** Organisation génomique du Sars-CoV-2. (ORF : open reading frame ; RdRp: gène codant l'ARN polymérase ARN-dépendante ; S, E, M, N : gènes codant les protéines de structure (S [surface], E [enveloppe], M [membrane], N [nucléoprotéine]).

### 6. Variants du SARS-CoV-2 :

#### 6.1. Variant Alpha, ou variante anglaise, ou VOC2020, ou B.1.1.7, ou 20I/501Y.V1

Fin décembre 2020, un nouveau variant préoccupant du SRAS-CoV-2, la lignée B.1.1.7 également appelé variant Alpha ou GRY (anciennement GR/501Y.V1), a été signalé au Royaume-Uni sur la base du génome entier séquençage d'échantillons de patients testés positifs pour le SRAS-CoV-2 [19]. Le variant B.1.1.7 comprend 17 mutations dans le génome viral. Parmi celles-ci, huit mutations se trouvent dans la protéine de pointe (S). La mutation N501Y montre une affinité accrue de la protéine de pointe pour les récepteurs ACE 2, améliorant la fixation virale et l'entrée ultérieure dans les cellules hôtes. [20] Cependant, des études ultérieures ont depuis rapporté que les personnes infectées par la variante de la lignée B.1.1.7 présentaient une gravité accrue de la maladie par rapport aux personnes infectées par d'autres formes circulantes de variants viraux [21-24].

#### 6.2. Variant Beta, ou variant sud- africain, ou B.1.351, ou 20H/501Y.V2 :

Un autre variant du SRAS-CoV-2, B.1.351 également appelée variant bêta ou GH501Y.V2 avec de multiples mutations de pointe, a entraîné la deuxième vague d'infections au COVID-19, a été détectée pour la première fois en Afrique du Sud en octobre 2020 [25]. Le variant B.1.351 comprend neuf mutations dans la protéine de pointe, dont trois mutations sont situées dans le RBD et augmenter l'affinité de liaison pour les récepteurs ACE. Ce variant aurait un risque accru de transmission et une neutralisation réduite par la thérapie par anticorps monoclonaux, les sérums de convalescence et les sérums post-vaccination [26].

#### 6.3. Variant Gamma, ou variant brésilien, ou 20J/501Y.V3 :

Le troisième variant préoccupant, le variant P1 également connue sous le nom de variant Gamma ou GR/501Y.V3 a été identifié en décembre 2020 au Brésil et a été détectée pour la première fois aux États-Unis en janvier 2021. Le variant B.1.1.28 héberge dix mutations dans la protéine de pointe et trois mutations sont localisées dans le RBD, similaire au variant B.1.351 [27].



### 6.4. Variant Delta, ou variant indien, ou B.1.617 :

À partir de début mars 2021, Inde a enregistré une recrudescence massive du nombre de cas [28]. Cette reprise épidémique a été constatée dans tout le pays, impliquant différents lignages viraux, dont le variant Alpha. Cependant, très rapidement, il a été observé des profils génétiques inconnus, qui ont émergé à occasion de cette reprise épidémique [29]. Parmi ces nouveaux lignages, trois profils génétiques différents ont été décrits (B.1.617.1, B.1.617.2 et B.1.617.3). Tous avaient en commun une mutation L452R sur la protéine de spicule pouvant être associée à une augmentation de la transmissibilité et un échappement immunitaire du virus [30,31]. Le variant Delta est hautement transmissible, estimée à environ 60% plus transmissible que la variant Alpha

### 6.5. Variant Omicron ou B.1.1.529

Le variant Omicron (B.1.1.529) a été détecté pour la première fois en novembre 2021, a été immédiatement déclaré COV par l'Organisation mondiale de la santé (OMS) et a rapidement augmenté en fréquence dans le monde entier. Le variant Omicron est considérablement muté par rapport à tous les isolats de SRAS-CoV-2 précédemment décrits, y compris 37 substitutions de résidus dans la protéine de pointe dans l'haplotype prédominant. Quinze des mutations d'Omicron sont regroupées dans le RBD, qui est la principale cible des anticorps neutralisants après une infection ou une vaccination [32, 33] suggérant qu'Omicron pourrait échapper aux anticorps induits par les infections et les vaccins et aux anticorps monoclonaux thérapeutiques

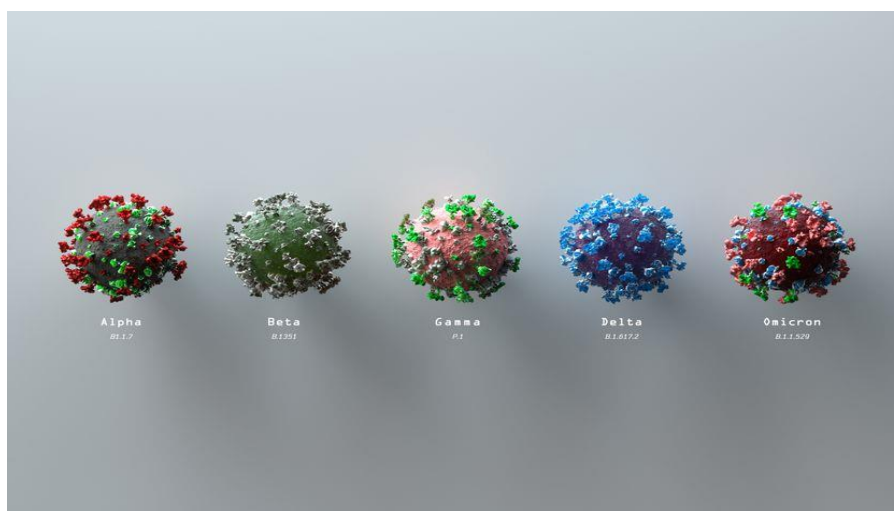


Figure 04 : Variants du SARS- CoV- 2

### 7. Récepteurs :

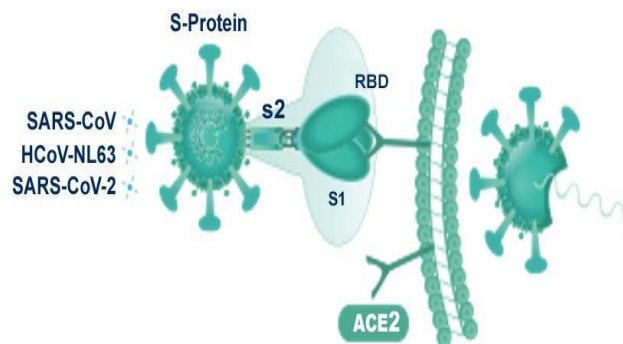
#### a. Enzyme de Conversion de l'Angiotensine 2 (ACE2) :

##### 1. Définition :

L'ACE2 appartient à la famille des enzymes de conversion de l'angiotensine ACE qui ont une distribution tissulaire plus large [34,35]. Ces enzymes (ECA) sont des métallo peptidases de zinc largement répandue de 170 kDa codée par un gène de 21 kb situé sur le chromosome 17 (17q23) [36]. L'ACE2 humaine [37,34] est une protéine transmembranaire de surface de type I exprimée dans les artères, le cœur, les reins et les épithéliums du poumon et de l'intestin grêle [37, 34].

##### 2. Mécanisme d'activité :

L'ACE2 joue un rôle essentiel dans la pathogénèse humaine de la maladie à coronavirus-19 (COVID-19) causée par le coronavirus-2 du syndrome respiratoire aigu sévère (SARS-CoV-2, ou 2019-nCoV). En effet, il a été établi que hACE2 est le récepteur de la protéine Spike (S) des coronavirus SARS-CoV et SARS-CoV-2, facilitant l'entrée du virus dans les cellules cibles [38-40]. Les pics de glycoprotéine présents sur l'enveloppe externe du virus se lient au récepteur membranaire ACE2, avec les protéases TMPRSS2, qui assurent sa liaison au domaine extracellulaire de ACE2, entraînant la pénétration du virus dans la cellule cible [41].



**Figure 05 :** Mécanisme d'activité de ACE2

### **b. Domaine de liaison au récepteur RBD :**

Receptor Binding Domain" (RBD) est un court fragment immunogène d'un virus qui se lie à une séquence de récepteur endogène spécifique pour entrer dans les cellules hôtes. Plus précisément, il s'agit d'une partie de la glycoprotéine "spike" (domaine S) qui est nécessaire pour interagir avec les récepteurs endogènes afin de faciliter la fusion membranaire et la libération dans le cytoplasme. Généralement, le domaine S est également le site des anticorps neutralisants. Le RBD est directement impliqué dans la fixation du virus au récepteur ACE2 de la cellule hôte [42-45].

### **8. Protéases :**

#### **8.1. Définition :**

Les coronavirus possèdent deux protéases [46]: la protéase de type papaïne (PL<sup>pro</sup> pour papain-like protease) et la protéase de type 3C (3CL<sup>pro</sup> pour 3C-like protease). Ces protéases sont des cibles antivirales intéressantes car elles sont essentielles à la réplication des coronavirus. La protéase majeure du SARS-CoV-2, 3CL<sup>pro</sup> (3C-like proteinase), elle appartient à la famille de la chymotrypsine. Sa séquence protéique est identique à 96 % à celle des protéases 3CL<sup>pro</sup> des autres coronavirus, qui ont été très étudiées [47-49]. Elle coupe la poly-protéine virale 11 fois au niveau de motifs Leu-Gln-(Ser/Ala/Gly). La deuxième protéase du SARS-CoV-2, PL<sup>pro</sup> (papain-like protease), a des propriétés différentes de celles de 3CL<sup>pro</sup>, notamment une activité déubiquitineuse, et elle pourrait intervenir pour altérer la réponse immunitaire innée de l'hôte infecté.

#### **8.2. Rôle :**

Les protéases virales sont essentielles à la production de virions par la cellule infectée car elles participent à une étape de maturation en coupant les poly-protéines virales synthétisées à des sites bien précis [50]. La PL<sup>pro</sup> est localisé dans la protéine non structurale 3 (NS3) du polypeptide viral. Elle permet l'autoclivage de nsp3 et la libération des protéines nsp1 et nsp2. Le ciblage de la protéase de type papaïne avec des médicaments antiviraux peut avoir un avantage non seulement en inhibant la réplication virale, mais aussi en inhibant le dérèglement des cascades de signalisation dans les cellules infectées qui peuvent conduire à la mort cellulaire dans les cellules environnantes non infectées. La protéase M-pro est responsable de la libération des protéines de nsp4 à nsp16. Elle joue un rôle essentiel dans le cycle de vie de virus et sa réplication [51,52]

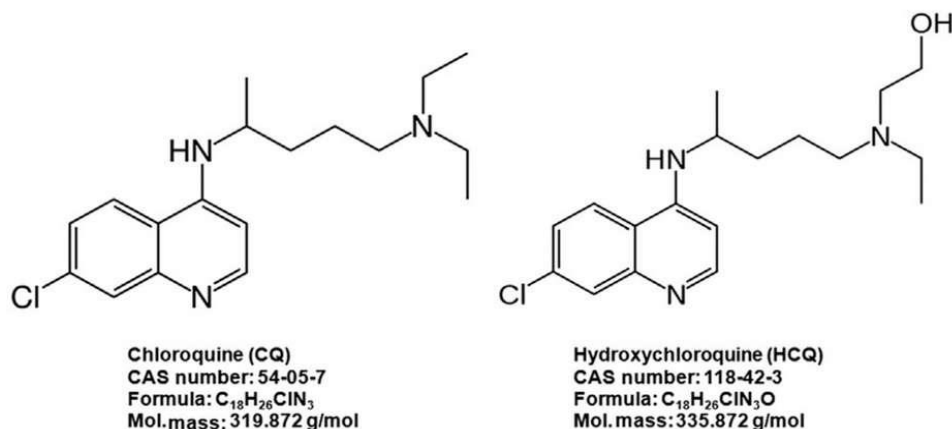
### 9. Inhibiteurs :

#### a. Inhibiteurs Synthétiques (Médicaments) :

##### 1. La chloroquine (CQ) et l'hydroxychloroquine (HCQ) :

La chloroquine (CQ) et l'hydroxychloroquine (HCQ) sont deux anti-malariques de la classe des amino-4-quinoléines. Ils ont un grand volume de distribution et sont fortement liés aux protéines plasmatiques. Leur métabolisme est hépatique et leur élimination principalement rénale. La demi-vie de la CQ est de 6 à 60 jours (moyenne 20 jours), et celle de son métabolite actif est de 59-67 jours. Celle de l'HCQ est d'environ 30 jours. Ces molécules s'accumulent donc dans les tissus et peuvent être retrouvées plusieurs semaines à mois dans les tissus. La CQ est substrat du CYP2C8 et de possibles interactions médicamenteuses pharmacocinétiques devraient être prises en compte [53].

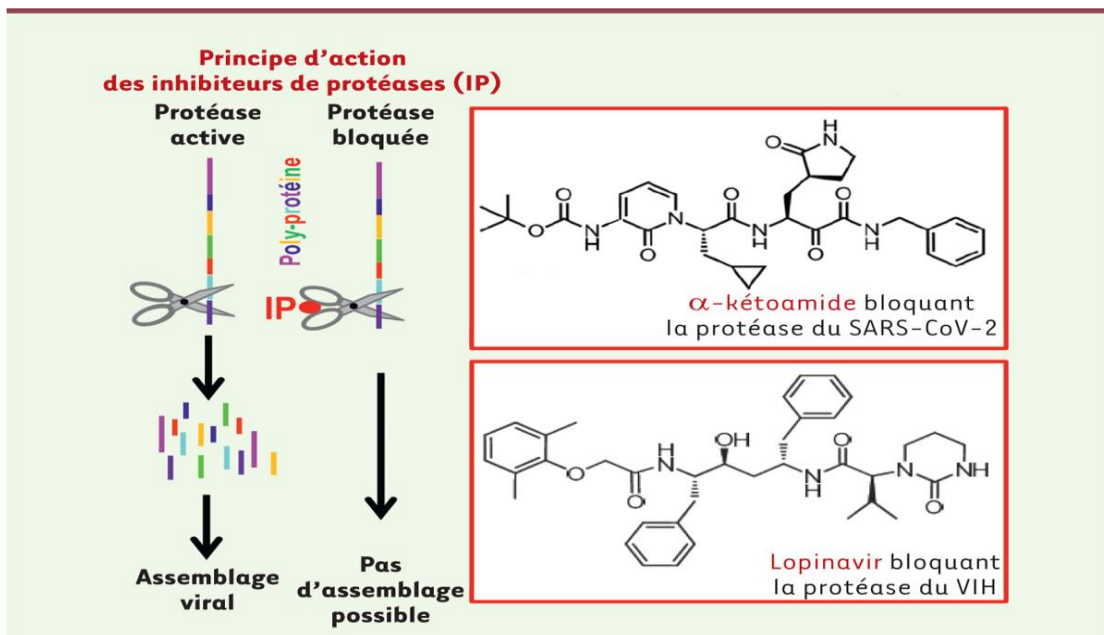
La chloroquine (C) et son dérivé l'hydroxy chloroquine (HC) sont des bases faibles qui s'accumulent au niveau des lysosomes, modifient leur pH, et interfèrent avec certaines enzymes. Elles ont ainsi la capacité d'inhiber la réplication des virus dont l'enveloppe fusionne avec celle de l'endosome acidifié. Ces médicaments empêchent en effet l'acidification des endosomes [54,55]. 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 [56].



**Figure 06 :** Structures chimiques de la chloroquine et de l'hydroxychloroquine.

### 2. Inhibiteurs de la protéase du VIH :

Le principe d'action de ces molécules est bloquer les protéases. Des inhibiteurs de la protéase 3CL<sup>pro</sup> des coronavirus déjà connus ont été testés pour leur capacité à inhiber celle du SARS-CoV-2. Par exemple Une molécule de la famille des  $\alpha$ -kétamides a été retenue pour son efficacité d'insertion dans le site catalytique de la protéase, ce qui bloque son activité et permet de réduire considérablement la production virale dans des cellules de poumon humain infectées avec le SARS-CoV-2 in vitro, ainsi que dans des modèles d'études précliniques [57]



**Figure 07 :** Principe d'action des inhibiteurs des protéases et comparaison des structures chimiques de deux inhibiteurs de protéases virales (le lopinavir et un  $\alpha$ -kétamide)

### b. Inhibiteurs d'origine naturelle :

Dans une situation d'urgence comme celle de la pandémie actuelle de COVID-19, et en absence de traitement conventionnel, la mise au point de médicaments de synthèse efficaces ou d'éventuels vaccins est soumise à la temporalité propre aux protocoles de recherches et de validation par des essais cliniques. Les médicaments à base de plantes et de produits naturels facilement disponibles et dont l'innocuité a déjà été prouvée peuvent faire gagner du temps en tant que première ligne de défense [58]

De tous les ingrédients naturels, les flavonoïdes et les terpènes pour avoir les mécanismes inhibiteurs les plus efficaces sur les molécules 3CLpro. La conclusion de ces ingrédients actifs peut aider à préparer le terrain pour développer les agents anti-COVID-19 nouveaux.

### 1. Les flavonoïdes :

Les flavonoïdes sont des produits largement distribués dans le règne végétal et sont couramment consommés quotidiennement sous forme de fruits, légumes et boissons telles que le thé. Ils sont capables de moduler l'activité de certaines enzymes et de modifier le comportement de plusieurs systèmes cellulaires, suggérant qu'ils pourraient exercer une multitude d'activités biologiques, notamment des propriétés antioxydantes, vasculoprotectrices, antihépatotoxiques, antiallergiques, anti-inflammatoires, antiulcéreuses et même antitumorales significatives[59]. Des flavonoïdes tels que la quercétine, l'épigallocatechine gallate, et la gallicatéchine gallate (GCG) ont montré de bonnes capacités d'inhibition sur l'enzyme 3CLpro qui est nécessaire à la réplication virale du coronavirus associé au syndrome respiratoire aigu sévère (SARS-CoV) [60].

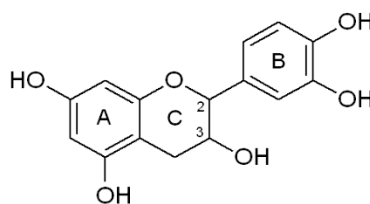


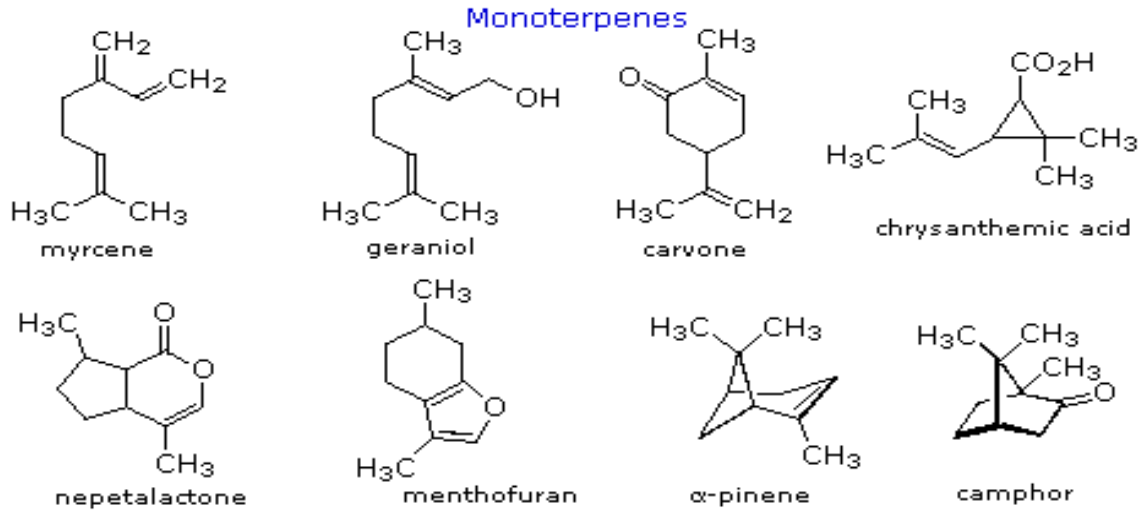
Figure 08 : Structures des flavonoïdes.

### 2. Les terpènes :

Les **terpènes** sont des hydrocarbures naturels **extraits de résines végétales et/ou d'huiles essentielles**. Ces composés organiques volatils sont produits par de nombreuses plantes et notamment par les conifères. Composants majeurs de la résine, de l'essence de térébenthine, du menthol, ils sont aujourd'hui fréquemment rencontrés dans nos environnements intérieurs.

L'activité d'une nouvelle formulation à base de terpènes, dont le nom de code est NT-VRL-1, contre la souche 229E du Coronavirus humain (HCoV) a été évaluée sur des fibroblastes pulmonaires humains (cellules MRC-5), avec et sans l'ajout de cannabidiol (CBD). Les principaux constituants de la formulation terpénique utilisée pour l'expérience étaient le bêta-caryophyllène, l'eucalyptol et le citral. La formulation testée a présenté un effet antiviral lorsqu'elle a été pré-incubée avec les cellules hôtes avant l'infection virale. La combinaison de

NT-VRL-1 et de CBD a potentialisé l'effet antiviral mieux que les contrôles positifs pyrazofurine et glycyrrhizine [61].



**Figure 09 :** Structures chimiques de quelques Terpènes

### 3. Les vitamines :

Avec la nouvelle épidémie de coronavirus 2019 (SARS-CoV-2) qui se propage maintenant à travers le monde, les gens cherchent des moyens de se protéger contre le virus ou d'atténuer ses effets une fois attrapé. Parmi les moyens étudiés on trouve les vitamines C et D. Plusieurs travaux ont montré que ces vitamines améliorent significativement l'état de santé chez les patients symptomatiques non graves infectés par COVID-19 [62].

### Références :

1. Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong et al. *N Engl. J. Med.* 2020, p 1199 – 1207.
2. A.J. Kayode, F.O. Banji-Onisile, A.O. Olaniran, A.I. Okoh *Agents pathogènes*, 2021, p 1108.
3. J.S. Peiris, K.Y. Yuen, A.D. Osterhaus, K. Stöhr, *N. Engl. J. Med.* 2003, p 2431 – 2441.
4. A. Sharif. Yakan, S.S. Kanj, *PLoS Pathog.* 2014. P 1004457.
5. A.E. Gorbalenya, S.C. Baker, R.S. Baric, R.J. Groot, C. Drosten, A. Gulyaeva, B.L. Haagmans, C. Lauber, A.M. Leontovich, B.W. Neuman et al. *bioRxiv*, 2020.
6. [http://www.xinhuanet.com/world/2020-02/12/c\\_1125561389.htm](http://www.xinhuanet.com/world/2020-02/12/c_1125561389.htm)
7. M. Cascella, M. Rajnik, A. Aleem, S. Duleboh, *National library of medicine*, 2022.
8. V.D. Menachery, B.L. Yount, K. Debbink, et al. *Nat Med*, 2015, p 1508–1513.
9. B. Hu, L. P. Zeng, X.L. Yang, et al. *PLoS Pathog*, 2017, p 1006698.
10. H.K.H. Luk, X. Li, J. Fung, et al. *Genet Evol*, 2019, p 21–30.
11. J. Cui, F. Li, Z.L. Shi, *Nat. Rev. Microbiol*, 2019, p 181–192.
12. D. Cavanagh *Arch. Virol.* 1997, p 629-633.
13. V.A.N. Regenmortel, M. H. V, C. M. Fauquet, D. H. L. Bishop, et al, *M.H.V v R*, Fauquet CM, DHL B, Carstens EB, Estes MK, Lemon SM, et al. San Diego: Academic Press, 2000, P 835-49.
14. P. C. Woo, S. K. Lau, C. S. Lam. et al. *J. Virol*, 2012, p 3995-4008.
15. A. Vabret, J. Brouard, J. Petitjean, G. Eugene. Ruellan, F. Freymuth, *Presse Med.* 1998, p 1813–7.
16. C. Risco, I. M. Anton, L. Enjuanes, J. L. Carrascosa, *J. Virol.* 1996, p 4773–7.
17. L. Enjuanes, I. Sola, F. Almazan, *J. Biotechnol*, 2001, p 183–204.
18. C. Hulo, E. Castro, P. Masson, *Viral zone*, 2011.



**19.**S.E. Galloway, P. Paul, D. R. MacCannell, M. A. Johansson, J. T. Brooks, A. MacNeil. R. B. Slayton, S. Tong, B. J. Silk, G. L. Armstrong, M. Biggerstaff, V. G. Dugan, 2021, p 95-99

**20.**E. Volz, S. Mishra, M. Chand, J. C. Barrett, R. Johnson, L. Geidelberg, W. R. Hinsley, D. J. aydon, G. Dabrera, Á. O'Toole, R. Amato, M. Ragonnet.Cronin, I. Harrison, B. Jackson, C. V. Ariani, O. Boyd, N. J. Loman, J. T. McCrone, S. Gonçalves, D. Jorgensen ,R. Myers, V. Hill, D. K. Jackson, K. Gaythorpe, N. Groves, J. Sillitoe, D. P. Kwiatkowski O. Ratmann, S. Bhatt, S. Hopkins, A. Gandy, A. Rambaut, N. M. Ferguson, Nature, 2021.

**21.**K. Wu, A. P. Werner, J. IMoliva, M. Koch, A. Choi, G. B. E. Stewart-Jones,H. Bennett ,S. Boyoglu-Barnum,W. Shi, B. S. Graham, A. Carfi, K. S. Corbett, R. A. Seder, D.K Edwards, bioRxiv. 2021.

**22.** N. G. Davies, S. Abbott, R. C. Barnard, C. I. Jarvis, A.J. Kucharski, J. D. Munday , C. A .B. Pearson, T.W. Russell, D. C. Tully, A. D. Washburne, T. Wenseleers, A. Gimma, W. Waites, K. L. M. Wong, K. vanZandvoort, J. D. Silverman, K. Diaz.Ordaz, R. Keogh , R. M. Eggo, S. Funk, M. Jit, K. E. Atkins, W. J. Edmunds, R. P. Walensky, H. T. Walke, A.S. Fauci, JAMA, 2021.

**23.**N.G. Davies, R. C. Barnard, C. I. Jarvis, T. W. Russell, M. G. Semple, M. Jit, W.J Edmunds, Lancet Infect Dis. 2021.

**24.**E.Volz, S. Mishra, M. Chand, J.C Barrett, R. Johnson, L. Geidelberg, W. RHinsley, D. J. Laydon, G. Dabrera, Á. O'Toole, R. Amato, M. Ragonnet.Cronin, I. Harrison, B. Jackson, C. V. Ariani, O. Boyd, N. J. Loman, J. T. McCrone, S. Gonçalves, D. Jorgensen, R. Myers, V .Hill, D. K. Jackson, K. Gaythorpe, N. Groves, J. Sillitoe, D. P. Kwiatkowski, S. Flaxman, O. Ratmann, S. Bhatt, S. Hopkins, A. Gandy, A. Rambaut, N.M. Ferguson, Nature, 2021.

**25.** H. Tegally, E. Wilkinson, M. Giovanetti, A. Iranzadeh, V. Fonseca, J. Giandhari, D. Doolabh, S. Pillay, E. J. San, N. Msomi, K. Mlisana, A. vonGottberg, S. Walaza, M.Allam, A.Ismail, T.Mohale, A.J.Verre, S.Engelbrecht, G.Van Zyl, W.L.Preiser, W. Petruccione, F.Sigal, A.Hardie, D.Marais, G.Hsiao, N.Y.Korsman, S.Davies, M.A.Tyers, L.Mudau, I.York,

D.Maslo, C.Goedhals, D.Abrahams, S.Laguda-Akingba, O. Alisoltani,Dehkordi, A.Godzik, A. Wibmer, C.K.Sewell, B.T. Lourenço, J.Alcantara, L.C.J. Kosakovsky, S.L.Pond, S. Weaver, D. Martin, R.J.Lessells, J.N.Bhiman, C.Williamson, Nature, 2021.

**26.**P. Wang, R. G. Casner, M. S. Nair, M. Wang, J. Yu, G. Cerutti, L. Liu, P. D. Kwong, Y. Huang, L. Shapiro, Ho.DD, bioRxiv, 2021.

**27.**N. R. Faria, T. A. Mellan, C. Whittaker, I.M. Claro, D. D. S. Candido, S. Mishra, M. A. E. Crispim, F.C. Sales, I. Hawryluk, J. T. McCrone, R. J. GHulswit, L.A.M. Franco, M. S. Ramundo, J. G. Jesus, P. S. Andrade, T. M. Coletti ,G. M. Ferreira, C. A. M. Silva, E. R. Manuli, R. H. M .Pereira, P. S. Peixoto, M. U. Kraemer, N. Gaburo, C. D. C. Camilo, H. Hoeltgebaum, W. M. Souza, E. C. Rocha, L. M. Souza, M. C. Pinho, L. J. T. Araujo, F. S. VMalte, A. BLima, J. D. PSilva, D. A. G .Zauli, S. Ferreira, A. C. R. P. Schnekenberg, D. JLayerdon, P. G. TWalker, H. M. Schlüter, A.L.P, M. S. Vidal, V. S, R.M.F. Filho, H.M, R. S.Aguiar, J. L. P. Modène, B. Nelson, J. A. Hay, M. Monod, X. Miscouridou, H. Coupland, R. Sonabend, M. Vollmer, A. Gandy, M. A. Suchard, T. A. Bowden, S. L. K. Pond, O. Ratmann, N. M. Ferguson, C. Dye, N. J. Loman, P. Lemey, A. Rambaut, N. A. Fraiji, O. G. Pybus, S. Flaxman, S. Bhatt, E. C. Sabino, l.medRxiv , 2021.

**28.**S. Cherian, V. Potdar, S. Jadhav, P. Yadav, N. Gupta, M. Das, BioRxiv, 2021, p 1542.

**29.**G. Vaidyanathan, Nature, 2021. p 321-322.

**30.**V. Tchesnokova, H. Kulakesara, L. Larson, V. Bowers, E. Rechkina, D. Kisiela, J. Clin Microbiol, 2021.

**31.**S.Cherian, V.Potdar, S.Jadhav, P.Yadav, N.Gupta, M.Das, et al. Microorganismes, 2021. p 1542.

**32.**L.Piccoli, J. candida medical association, 2020, p 1024-1042 .

**33.** A.J. Greaney, cell Host Microbe, 2021.p 463–476.

**34.**D. Harmer, FEBS Letters, 2002, p 107-110.

**35.**I. Hamming, J. Patho, 2004, p 631-637.

**36.**T.E. Howard, S.Y. Shai, K.G. Langford, B.M. Martin, K.E. Bernstein, gene Mol. Cell. Biol. 1990, P 4294-43.

**37.** M. Donoghue, Cir.Research, 2000, p 1-9.

**38.** W. Li, Nature, 2003, p 450-454.

39. M. Hoffmann, *Cell*. 2020, p 1-16.
40. P. Zhou, *Nature*, 2020, p 270-273.
41. T. Behl, I. Kaur, S. Bungau, A. Kumar, M.d .Sahab Uddin, C.h. Kumar, G.Pal, Sahil, K. Shrivastaa, G. Zengin, *Life Sciences*, 2020.
42. S. Chakraborti, *J. Virol.* 2005.
- 43 . Y.He, *Biochem Biophys Res Commun*. 2004, p 773–781.
- 44.H.Mou, *J.virol*, 2013, p 9379–9383.
- 45.W.Tai, *Cell Mol Immunol*, 2020, p 613–62.
46. R. Hilgenfeld, *J. Febs*. 2014, p 4085–4096.
- 47.K. Anand, J. Ziebuhr, P. Wadhwani, *Science*, 2003, p 1763–1767.
- 48.D.Juckel, J. Dubuisson, S. Belouzard, *Med. Sci. Paris*, p 633–41.
- 49.L.Zhang, D. Lin, X. Sun, 2020, *Science*, p 409–12.
50. S. Jager, P. Cimermancic, N. Gulbahce, J. R. Johnson, K. E. McGovern, S. C. Clarke, M. Shales, G. Mercenne, L. Pache, K. Li, H.Hernandez, G. M .Jang, S. L. Roth, E. Akiva, J. Marlett, M. Stephens, I. D’Orso, J. Fernandes, M. Fahey, C. Mahon, A.J. O’Donoghue, A. Todorovic, J. H. Morris, D. A. Maltby, T. Alber, G. Cagney, F. D. Bushman, A. John ,K. Young, Sumit , I. Chanda, Wesley ,T. Sundquist, R. Kortemme, D. Hernandez, Ch. S. Craik, A. Burlingame, A. Sali, A. D. Frankel, N. J. Krogan. *Nature*, 2011, p 365-370.
- 51.K.Anand, J. Ziebuhr, P. Wadhwani, J. R. Mesters, R. Hilgenfeld, *Science*, 2003.
- 52.D.Juckel, J. Dubuisson, S. Belouzard, *Médecine/sciences*, 2020, p 633-41.
- 53.C. Samer, F. Curtin, K. R Ing.Lorenzini, P. Vetter, *Vetter Date création*, 2020.
- 54.M. Wang, R. Cao, L. Zhang, *Vitro. Cell. Res*. 2020, p 269 – 271.
- 55.J. Liu, R. Cao, M. Xu, *vitro Cell Discov*. 2020, p 16.
- 56.J. Fuhier, F. Cova. *Society for Judgment and Decision Making*, 2020, p 889-908.
- 57.L. Zhang, D.Lin, X.Sun, *science*, 2020, p 409 - 12 .
- 58.R. A. Khamitov, S. I. Loginova, V. N. Shchukina, S. V. Borisevich, V.A. Maksimov, AM. Shuster, *Vopr Virusol*. 2008, p 9-13.
- 59.K. Ghedira, *Phytotherapy* ,2005, **3**, 162–169

**60.**T. T. H. Nguyen, H. J. Woo, H. K. Kang, V. D. Nguyen, Y. M. Kim, D. W. Kim, S. A. Ahn, Y. Xia, D. Kim, *Biotechnol. Lett.* 2012,

**61.**L. Chatow , A. Nudel , I. Neshher , D. HayoHemo <https://pubmed.ncbi.nlm.nih.gov/33805385/-affiliation-1>, P. Rozenberg , H. Voropaev , I. Winkler , R. Levy , Z. Kerem , Z. Yaniv , N. Eyal , *Life Basel.* 2021.

**62.**M. Castillo, *Clinicaltrials*, 2020.

**Chapitre II : Relation  
Quantitative  
Structure-Activité  
(QSAR)**

### 1. Historique :

La modélisation QSAR (La relation quantitative structure-activité) a été largement appliquée à la découverte de médicaments assistée par ordinateur. Les premiers travaux utilisant la méthode (QSAR) ont commencé au début des années 60 avec Hansch [1] d'une part qui a effectivement été le pionnier du domaine [2] et de Free et Wilson [3] d'autre part, qui ont proposé un modèle mathématique pour corréler l'activité biologique et la structure chimique. Les développements de cette étude (QSAR) sont très anciens, à partir de 1868, lorsqu'Alexander Crum-Brown et Thomas Fraser montrent l'existence d'une relation entre l'activité physiologiques et la structure chimique [4] suivis, avec Richet qui établit une relation entre la toxicité et les propriétés physicochimiques [5,6].

De manière indépendante, Meyer et Overton ont décrit une corrélation linéaire entre la lipophilie (coefficient de partage huile-eau) et les effets biologiques (narcotiques) . Pendant les dernières décennies, ce domaine a largement été étudié et les données bibliographiques disponibles sur cette approche sont maintenant importantes [7,8] Le domaine s'est développé et a considérablement évolué après l'apparition d'une revue complète de la modélisation QSAR [9], discutant l'évolution des méthodes et des meilleures pratiques de QSAR. De nombreuses publications ont fait progresser les domaines traditionnels de la modélisation QSAR [10] tels que la prédiction des activités biologiques et des propriétés ADME/Tox, en s'appuyant sur l'utilisation réussie de la modélisation QSAR en chimie, agrochimie, pharmaceutique, [11] et industries cosmétiques[12] Cependant, de nouvelles directions et domaines d'application intéressants ont également émergé, tels que la chimie des procédés [13,14] et la prédiction et l'optimisation des voies (rétro) synthétiques[15] Ainsi, les modèles sont devenus une partie intégrante du processus de découverte de médicaments, fournissant des conseils substantiels dans la planification des expériences [16,17].

### 2. Définition :

La relation quantitative structure-activité (QSAR) est un modèle mathématique qui associe un ou plusieurs paramètres quantitatifs dérivés de la structure chimique, à une mesure quantitative d'une activité.

Cette approche computationnelle relie l'activité et la structure d'un composé chimique par un certain algorithme mathématique, basée sur le postulat de base selon lequel les composés chimiques similaires ont des activités similaires.

L'application de ces modèles consiste à prédire ou concevoir de nouveaux produits chimiques avec l'activité souhaitée [18], La forme générale d'un tel modèle est la suivante :

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

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

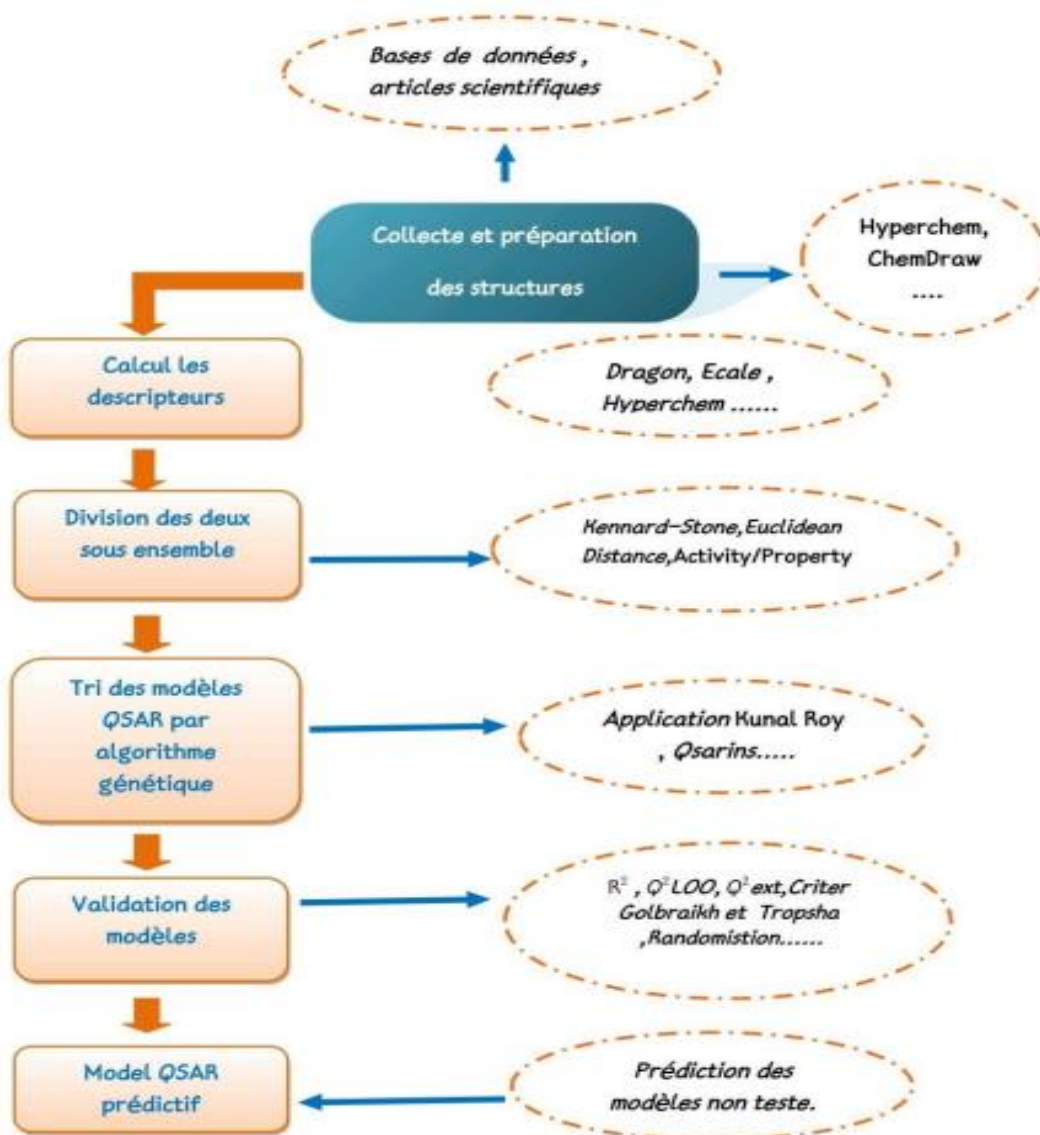
### 3. Principe :

Le principe d'une étude QSAR est de trouver une relation mathématique liant quantitativement une activité biologique, mesurée pour une série de composés ayant le même squelette structurel dans les mêmes conditions expérimentales, avec des descripteurs moléculaires (propriétés physico-chimiques) à l'aide de méthodes statistiques [19]. Un QSAR est conçu comme un modèle statistique qui relie un ensemble de descripteurs structuraux d'un composé chimique à ses activités biologiques.

### 4. 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. Une fois développé, le modèle doit être validé en termes de corrélation.

L'influence des composés du jeu d'entraînement sur le modèle est estimée par des méthodes de validation interne. Pour estimer le pouvoir prédictif du modèle, il est nécessaire de disposer de données expérimentales supplémentaires, c'est la validation externe [20].



**Figure 01** : Présentation de la méthodologie de QSAR



### 5. 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 [21,22].

Les données biologiques ou biochimiques des composés étudiés (ex : concentration inhibitrice médiane IC50), sont généralement exprimées sur une échelle logarithmique (PIC50), en raison de la relation linéaire entre la réponse et le logarithme de la concentration du composé. En outre, les logarithmes inverses de l'activité ( $\log 1/IC50$ ), sont également utilisés pour obtenir des valeurs mathématiques plus élevées, lorsque les structures sont fortement bioactives [23,24].

### 6. Descripteurs

#### 6.1. Définition :

Un descripteur est un paramètre numérique propre à une structure chimique donnée, également c'est 'un résultat d'un procédé mathématique ou d'une certaine expérience normalisée. Ces descripteurs représentent un moyen de corréler une structure chimique et une valeur de propriété physico-chimique ou biologique. Les descripteurs sont finalement des intermédiaires numériques se substituant à la molécule elle-même [25]. Ces descripteurs peuvent être utilisés pour évaluer l'influence de la structure moléculaire ou les relations activité-structure, aussi pour l'analyse de symétrie et la projection des bases de données des molécules [26].

#### 6.2. Types des descripteurs

##### 6.2.1. Descripteurs constitutionnels:

Ces descripteurs se basent sur la composition chimique du système moléculaire, ils sont très couramment utilisés fait de leur extrême simplicité non seulement d'un point de vue conceptuel mais surtout calculatoire. Ils sont notamment à la base des analyses par groupes de contribution,

comme la méthode de Benson [27], dans laquelle l'effet des substituants est introduit sous la forme d'une valeur tabulée.

L'inconvénient majeur de ce type de descripteurs pour obtention des modèles QSAR les plus simples, qu'ils ne permettent pas une bonne interprétation des mécanismes d'interaction mis en jeu pour la propriété étudiée. Par exemple, ils ne permettent pas de distinguer les isomères. Or, pour la grande majorité des propriétés, la position d'un substituant modifie la valeur expérimentale d'une propriété.

### 6.2.2. Descripteurs topologiques :

Ils utilisent des représentations graphiques bidimensionnelles des molécules, contenant des informations sur la connectivité des fragments moléculaires, et des estimations de propriétés physicochimiques. En effet, les indices topologiques sont généralement utilisés, pour l'interprétation des équations QSAR résultantes, afin de comprendre les liaisons et interactions physicochimiques.

Cependant, les descripteurs 2D sont insuffisants, pour expliquer convenablement certaines propriétés ou activités biologiques [28,29].

### 6.2.3. Descripteurs géométriques :

Les descripteurs géométriques sont dérivés des structures tridimensionnelles (3D) des molécules définies par les coordonnées des atomes, des noyaux atomiques et la taille des molécules représentées [30], ainsi que des rayons et des masses atomiques. En général, ces descripteurs sont calculés soit à partir de la géométrie moléculaire optimisée par modélisation moléculaire soit à partir des coordonnées cristallographiques (coordonnées cartésiennes x, y, z des atomes de la molécule) [31].

La plupart de ces descripteurs (descripteurs basés sur la grille) ont besoin de règles d'alignement pour réaliser la comparabilité des molécules.

### 6.2.4. Descripteurs électroniques :

Ces descripteurs sont d'une importance critique dans la détermination des types de forces intermoléculaires qui modulent les interactions des récepteurs médicamenteux, ils décrivent l'influence d'un certain groupe ou d'un substituant sur la distribution de la densité électronique (polarité des molécules) [32].

Les descripteurs les plus importants sont montrés dans le tableau 01.

**Tabelau 01:** Les descripteurs électroniques calculés dans les études QSAR [33].

Descripteur	Définition
<b>L'énergie HOMO</b> ( $E_{\text{HOMO}}$ )	Est le niveau d'énergie le plus élevé dans la Molécule qui contient des électrons
<b>L'énergie LUMO</b> ( $E_{\text{LUMO}}$ )	Est le niveau d'énergie le plus bas dans la molécule qui ne contient pas d'électrons.
<b>Le moment dipolaire</b> ( $\mu$ )	Mesure la polarité nette moléculaire, qui résulte S'il y a une différence d'électronégativité entre les atomes de molécule.
<b>L'électronégativité</b> ( $\chi$ )	$\chi = -\mu = -\left(\frac{\partial E}{\partial N}\right)_{v(r)} = -\frac{(E_{\text{LUMO}} + E_{\text{HOMO}})}{2}$

### 6.2.5. Descripteurs spectroscopiques :

Les molécules peuvent être caractérisées par des mesures spectroscopiques, par exemples par leurs fonctions d'onde vibrationnelles. En effet, les vibrations d'une molécule dépendent de la masse des atomes et des forces d'interaction entre ceux-ci ; ces vibrations fournissent donc des informations sur la structure de la molécule et sur sa conformation. Les spectres infrarouges peuvent être obtenus soit de manière expérimentale, soit par calcul théorique, après recherche de la géométrie optimale de la molécule. Ces spectres sont alors codés en vecteurs de descripteurs de taille fixe [34]. Les descripteurs de type MoRSE (Molecule Representation of Structures based on Electron diffraction) sont calculés à partir d'une simulation du spectre infrarouge ; ils font appel au calcul des intensités théoriques de diffraction d'électrons [35].

**Tableau 02** : Différents descripteurs, employés dans les études QSAR, basés sur la dimension.

<b>Dimension de descripteurs</b>	<b>Paramètres</b>
<b>Descripteurs 0D</b>	Indices constitutionnels, propriétés moléculaires dénombrement d'atomes et de liaisons.
<b>Descripteurs 1D</b>	Nombre de fragments, empreintes digitales.
<b>Descripteurs 2D</b>	Paramètres topologiques, paramètres structuraux paramètres physico-chimiques incluant des descripteurs thermodynamiques
<b>Descripteurs 3D</b>	Paramètres électroniques, paramètres spatiaux, paramètres d'analyse de forme moléculaire, paramètres d'analyse de champ moléculaire et paramètres d'analyse de surface de récepteur.

### 7. 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. Les variantes inclusives peuvent être PLS, RLM et RNLM par étapes [36].

#### 7.1. Régression linéaire multiple (RLM)

La plus simple méthode statistique de modélisation et la plus appliquée dans les études de la relation structure-activité. La méthode a été popularisée par Hansch en reliant l'activité biologique aux propriétés expérimentales lipophiliques, électroniques et stériques pour des séries de composés [37].

La méthode RLM repose sur l'hypothèse qu'il existe une relation linéaire entre une variable dépendante Y (activité biologique) et une série de n variables indépendantes Xi (descripteurs moléculaires).

$\epsilon_i$  (l'erreur du modèle, il résume les informations manquantes qui permettrait d'expliquer linéairement les valeurs de Y à l'aide des p variables Xn ). L'objectif est d'arriver à une équation mathématique de la forme :

$$Y = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_n X_n + \epsilon_i$$

La régression linéaire est facile à mettre en œuvre, et les coefficients an obtenus peuvent être interprétés : ils mesurent l'influence de chacune des variables sur la grandeur étudiée. Cependant, il est souvent nécessaire d'avoir recours à des modèles qui prennent en compte la corrélation non linéaire [38].

### 7.2. Régression des moindres carrés partiels (PLS)

La régression PLS, est une méthode rapide, efficace et optimale pour une parfaite minimisation des covariances. Cette méthode est très utilisée lorsqu'il y a un grand nombre de variables prédictives, ou lorsqu'il y a de fortes colinéarités entre ces dernières [39]. En premier temps, la PLS réduit les variables prédictives, à un plus petit ensemble de composantes non corrélées. Ensuite, elle effectue la régression par les moindres carrés, sur ces composantes plutôt que sur les données initiales [40]. Contrairement à la régression linéaire multiple (MLR), la PLS ne suppose pas, que les variables prédictives sont fixes. Ainsi, les mesures de ces variables tolèrent des erreurs, en incluant des incertitudes des mesures [41].

### 7.3. Régression non linéaire multiple

La régression non linéaire multiple (RNLM) est une méthode non linéaire (exponentielle, logarithmique, polynomiale, ...) qui permet de déterminer le modèle mathématique qui peut expliquer non-linéairement au mieux la variabilité d'une propriété ou d'une activité Y en fonction des descripteurs moléculaires [42] :

$$Y = a_0 + \sum_{i=1}^n a_i x_i + \dots + b_i x_i^n$$

Avec :

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

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

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

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

**a<sub>i</sub>** et **b<sub>i</sub>** : sont les coefficients de descripteurs dans l'équation du modèle [43].

### 8. Paramètres statistiques de la validation

#### 8.1. 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 [44].

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

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

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

#### 8.2. Coefficient de détermination (R<sup>2</sup>)

Le coefficient de détermination noté R<sup>2</sup> 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 R<sup>2</sup> proche de 1, indique un bon ajustement du modèle de régression [45-47]

## Chapitre II : Relation Quantitative Structure-Activité (QSAR)

---

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  $R^2$  est calculé selon la formule : Le coefficient  $R^2$ , 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 carrés totaux

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

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

Ainsi: **TSS = ESS + RSS**

### 8.3. 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 [48-50].

$F$  peut s'écrire comme suit :

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

**ESS** =  $\sum (\hat{y} - \bar{y})^2$ : Somme des carrés 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.

### 8.4. 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  $\hat{a}_i$  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) [51].

### 8.5. 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 par [52] :

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

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

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

### 8.6. $R^2$ 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),  $R^2_{\text{ajusté}}$  a été défini de la manière suivante :

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

**p**: nombre de variables utilisées

Le  $R^2_{\text{ajusté}}$  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  $R^2_{\text{ajusté}}$  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 [53]

### 8.7. 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 régression [54-56] . Ce paramètre est exprimé comme suit :

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

Où :

$Y_{\text{Prédite}}$ ,  $Y_{\text{réelle}}$  et  $\bar{y}$ : sont les valeurs prédites, réelles et moyennes de l'activité biologique ( $pIC_{50}$ ) de la molécule  $i$ , respectivement

### 8.8. Coefficient de corrélation prédit $Q$

La racine carrée du coefficient de prédiction  $Q$ , appelée également le coefficient de corrélation prédit, est calculé entre les valeurs prédites et celles expérimentales, lors de la validation croisée du modèle de régression. En effet, plus la valeur absolue du coefficient de corrélation prédit  $Q$  est proche de 1, plus la modèle de régression est fiable et valide [57-60].

### 9. Techniques de validation

La validation des modèles est nécessaire pour estimer leur fiabilité et détermine la reproductibilité des résultats et la pertinence d'un modèle développé pour une application donnée [61] Trois méthodes sont employées pour la validation d'un modèle de QSAR, à savoir les méthodes de validation interne, externe et le test de randomisation.

#### 9.1. Validation interne

Cette validation est appliquée sur le lot de test, afin de vérifier et de déterminer la stabilité du modèle QSAR et de tester l'impact et l'influence de chaque molécule sur le modèle final. Les procédures de validation croisée (en anglais cross-validation) sont souvent utilisées [62]

Les méthodes de validation croisée les plus populaires utilisent la technique Leave-n-out [63,64]: *Leave-one-out (LOO)*, *Leave-Two-out (LTO)* et *Leave-Many-out (LMO)*

Généralement, Ces techniques de validation permettent l'évaluation de la robustesse du modèle, autrement dit la stabilité des paramètres du modèle QSAR vis-à-vis des molécules du lot de test. Cela dit, qu'elles ne permettent en aucun cas de démontrer le pouvoir prédictif des modèles [65,66].

#### Leave-n-out

Cette méthode, consiste à extraire à chaque fois, un certain nombre « n » de molécules de lot d'essai initial à « k » molécules et à construire un nouveau modèle, avec les « k-n » molécules restantes, en utilisant les mêmes descripteurs moléculaires, choisis initialement pour construire le modèle d'origine

Ensuite, ce processus est répété « p » fois, pour retirer et prédire les valeurs de toutes les molécules de lot d'essai, par le modèle généré [67,68].

#### 9.2. validation externe

Cette méthode est utilisée afin de valider la fiabilité et la performance prédictive du modèle, dans la détermination de la corrélation existante, entre les activités prédites et celles expérimentales des molécules externes (lot de test). Elle consiste à prédire l'activité d'une série de molécules appelée généralement série de test qui ne sont pas dans la série de développement du modèle. En effet, plus cette corrélation est importante, plus le modèle est performant et

valide, pour prédire les activités biologiques des molécules, qui n'appartiennent pas au lot de test [70 -73]

Cette validation est caractérisée par le paramètre  $R^2_{(test)}$ . Récemment plusieurs études [74,75] ont montré l'insuffisance du paramètre  $R^2$ , pour vérifier le pouvoir prédictif des modèles QSAR. Par conséquent, d'autres paramètres doivent être vérifiés pour cet objectif. Ces paramètres sont connus sous le nom « critères de validation externe » ou souvent appelés « critères de Trophsa » [76]

Critères de validation Externe (série de test) :

- $R^2_{test} > 0.6$
- $\frac{R^2 - R_0^2}{R^2} < 0.1$      $0.85 \leq K \leq 1.15$  ou
- $\frac{R^2 - R_0'^2}{R^2} < 0.1$      $0.85 \leq K' \leq 1.15$
- $|R^2 - R_0^2| \leq 0.3$

$R_0^2$ : coefficient de corrélation entre les valeurs prédites et expérimentales pour la série de test.

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

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

### Références :

1. C. Hansch, P. Maloney, T. Fujita, R. Muir. *J. Nature*, 1962, p 178–180.
2. A. Cherkasov, E. Muratov, D. Fourches, A. Varnek, I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y. Martin, R. Todeschini, V. Consonni, V. Kuzmin, R. Cramer, R. Benigni, C. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard et A. Tropsha, *J. Med. Chim.* 2014, p 4977–5010.
3. H. Kubinyi. *J. Drug Discovery Today*, 1997, p 457–467.
4. F. Ban, K. Dalal, H. Li, E. LeBlanc, P. S. Rennie et A. Cherkasov. *J. Chem.*, 2017, p 1018–1028.
5. V. Alves, E. Muratov, A. Zakharov, N. Muratov, C. Andrade et A. Tropsha. *Food Chem. Toxicol.* 2018, p 526–534.
6. L. Simón-Vidal, O. García-Calvo, U. Oteo, S. Arrasate, E. Lete, N. Sotomayor et H. González-Díaz, *J. Chem. Inf. Model.* 2018, p 1384–1396.
7. R. Sheridan, W. Schafer, P. Piras, K. Zawatzky, E. C. Sherer, C. Roussel et C. J. Welch, *J. Chromatogr. A*, 2016, p 206–213.
8. B. Grzybowski, S. Szymkuć, E. Gajewska, K. Molga, P. Dittwald, A. Wołos et T. Klucznik. *J. Chem.*, 2018, p 390–398.
9. S. Capuzzi, W. Sun, E. N. Muratov, C. Martínez-Romero, S. He, W. Zhu, H. Li, G. Tawa, E. Fisher, M. Xu, P. Shinn, X. Qiu, A. García-Sastre, W. Zheng et A. Tropsha. *J. Med. Chim.* 2018, p 3582–3594.
10. C. Hansch, T. Fujita, *J. Am. Chem. Soc.* 1964, p 1616–1626.
11. S. M. Free, J. W. Wilson, *J. Med. Chem.* 1964, p 395–399.
12. A. Crum Brown, T. R. Fraser, *Trans. Roy. Soc. Edinburgh*, 1868, p 151–203.
13. A. Crum-Brown, T. R. Fraser, *Trans. Roy. Soc. Edinburgh*, 1868.
14. C. Richet, *C. R., Séances Soc. Biol. Ses. Fil.*, 1893.
15. H. Meyer, *Arch. Exp. Pathol. Pharmacol.*, 1899.

16. E. Overton, Jena: Gustav Fischer, Germany, 1901.
17. A. K. Debnath. Mini Reviews in Medicinal Chemistry, 2001, p 187-195.
18. EN. Muratov, J. Bajorath, RP Sheridan, Chemical Society, 2020.
19. J. Ferlay, M. Colombet, I. Soerjomataram, T. Dyba, G. Randi, M. Bettio, A. Gavin, O. Visser, F. Bray. J. Cancer, 2018, p 356–387.
20. S. Chtita. These, Université Moulay Ismail, Faculté des Sciences, Meknès, Maroc, 2017.
21. T. Mark, D. Cronin, T. Schultz, W. Chem. Res. Toxicol. 2001, p 1284- 1295.
22. S. Balaza, E. Sturdik, E Durkova, M .Antalik, P.Sulo. Biochimica et Biophysica Acta (BBA) – Bioenergetics, 1986, p 93-98.
23. S .Bouaziz-Terrachet, R. Terrachet, S. Tairi-Kellou. Med. Chem. Res, 2013, p 1529–1537.
24. S.Tairi-Kellou, S. Bouaziz Terrachet, B. Maouche, G. Moreau .J. Mol. Des, 2008, p 161-185.
25. O.R. Hansen. Acta Chemica scan dinavica, 1962, p 1593–1600.
26. C. Hansch and T. Fujita. J. Am. Chem. Soc. 1964, p 1616–1626.
27. S.W. Benson, J.H. Buss. J. Chem. Phys, 1958, p 546-572.
28. P. Geerlings, D. Proft. Chem. Rev. 2003, p 1793-1874.
29. H. Chermette. J. Comput. Chem, 1999, p 129-154.
30. M. Karelson. John Wiley & Sons, New York, 2000.
31. T. Puzyn, N. Suzuki, M. Haranczyk. J. chem. Inf. Model. 2008, p 1174-80.
32. S.M. Free, J.W. Wilson. J. Med. Chem. 1964, p 395–399.
33. Z. Zhou, RG. .Parr. J. Am. Chem. Soc. 1990, p 5720-4.
34. T.W. Heritage, et al. Drug Discovery and Design, 1998, p 381-398.
35. J.H. Schuur, P.Selzer, Gasteiger. J. of Chem. Inf. Comput. Sci. 1996, p 334-344.
36. T. Puzyn, J. Leszczynski, MT. Cronin. Springer Science & Business Media, 2010.
37. E. Besalu, J. V. De Julian-Ortiz, L. Pogliani. J. Chem. Inf. Model. 2007, p 751–760.
38. V. Consonni, D. Ballabio, R. Todeschini. J. Chemom. ,210, p 194–201.

39. S. Wold. Verlag-Chemie, Weinheim, Germany,1994.
40. M. Clark, R.D. Cramer III, Quant. Struct.- Act. Relat.1993, 12, p 137-145.
41. M. Sjöström, L. Eriksson. han van deWaterbeemd,1995, p 63- 90.
42. S. Niculescu, M. Lewis, J. Tigner. Environ. Res. 2008, p 735- 750.
43. I. Mitra, A. Saha, K. Roy. Mol Simult,2010, p 1067–1079.
44. K. Roy, I. Mitra, S. Kar, P. K. Ojha, R. N. Das, H. Kabir. J. Chem. Inf. Model. 2012, p 396–408.
45. T. Mark Cronin, T. Schultz. Chem. Res. Toxicol. 2001, p 1284-1295.
46. S. Balaza , E. Sturdik, E. Durkova ,M. Antalik , P. Sulo. Biochimica et Biophysica Acta (BBA) – Bioenergetics,1986, p 93-98.
47. J. M. Luco, F. H. Ferretti.J. Chem. Inf. Comput. Sci ,1997, p 392-401.
48. T.D. Mark Cronin, T. W. Schultz. Chem. Res. Toxicol. 2001, p 1284-1295.
49. S. Balaza, E. Sturdik, E. Durkova , M. Antalik , P. Sulo. Biochimica et Biophysica Acta (BBA) – Bioenergetics,1986, p 93-98.
50. J .M. Luco , F.H. Ferretti. J. Chem. Inf. Comput. Sci,1997, p 392-401.
51. S. L. Zabell. J. American Statistical Association,2008, p 1-7.
52. P.K. Ojha, I. Mitra, R.N. Das, K. Roy. Chemom. Intell. Lab. Sys,2011, p 194–205.
53. K. Roy, I. Mitra, P. Ojha, S. Kar, R.N. Das, H. Kabir. Chemom. Intell. Lab. Sys ,2012, p 200–210.
54. T.D. Mark. Cronin, T.W. Schultz. Chem. Res. Tox. 2001, p 1284-1295.
55. S. Balaza, E. Sturdik., E. Durkova , M. Antalik , P. Sulo. Biochimica et Biophysica Acta (BBA) – Bioenergetics,1986, p 93-98.
56. J . M. Luco, F. H. Ferretti. J. Chem. Inf. Comput. Sci ,1997, p 392-401.
57. T. D. Mark Cronin, T.W Schultz. Chem. Res. Tox.2001, p 1284-1295.

58. S. Balaza , E. Sturdik , E. Durkova ,M. Antalík., P. Sulo. *Biochimica et Biophysica Acta (BBA) – Bioenergetics*, 1986, p 93-98.
59. R.H. Azarboni, M. Keyanpour, M. Yaghouti. *Engineering Analysis with Boundary Elements*, 2019, p 204-210.
60. J.M. Luco, F.H. Ferretti. *J. Chem. Inf. Comput. Sci* ,1997, p 392-401.
61. A. Varnek, A. Tropsha. *Royal Society of Chemistry. Royaume- Uni*,2008.
62. A. Tropsha, P. Gramatica, V.K. Gombar. *Comb. Sci.* 2003, p 69–77.
63. Y. Yao, X.Z. Cheng, L. Wang, S. Wang, G. Ren. *Inter. J. Mol. Sci.* 2011, p 6445-6451.
64. A. Tropsha, P. Gramatica, V. Gombar. *QSAR & Comb. Sci.* 2003, p 69–77.
65. A. Tropsha, P. Gramatica ,V.K. Gombar. *QSAR & Comb. Sci* ,2003, p 69–77.
66. A. Golbraikh, A. Tropsha. *J. Mol. Graph. Model.* 2002, p 269–276.
67. Y.Yao , X .Z. Cheng ,L .X. Wang ,S .H .Wang. *J. Inter. J. Mol. Sci.* 2011, p 6445-6451.
68. A. Tropsha, P. Gramatica, V. K Gombar. *QSAR & Comb. Sci.* 2003, p 69–77.
69. P. Gramatica. Clifton, N.J, 2013, p.499-526.
70. R . Kunal, M. Indrani. *Mini Reviews in Medicinal Chemistry*, 2013, p 491-504.
71. R. H. Azarboni, M. Keyanpour, M.Yaghouti. *Engineering Analysis with Boundary Elements*, 2019, p 204-210.
72. T. K. A. Shameera, K. Vijisha, K . M. Rajan. *Food Science and Human Wellness* ,2019, p 53-62.
73. J .M. Luco , F. H. Ferretti. *J. Chem. Inf. Comput. Sci.* 1997, p 392-401.
74. A .T .A. Golbraikh .*J. Mol. Graph. Model.* 2002, p 269–276.
75. P. H. D. M. Y. E. M. A. G. H. Z. a. A. T. M. Martin. *J. Chem. Inf. Model.* 2012, p 2570–2578.
76. A. T. A. Golbraikh. *J. Mol. Graph. Model.* 2002, p 269 276.

# **Chapitre III : Résultats et discussion**



### 1. Introduction :

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

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

La validité d'un modèle QSAR dépendra donc des choix que l'on aura fait sur les paramètres. Il est ainsi judicieux d'identifier et d'évaluer quantitativement les paramètres les plus pertinents à utiliser en fonction de l'activité et de molécule sélectionnées.

Dans ce travail, nous intéressons à l'étude de la relation structure-activité d'une série des anti-HIV (dérivés de propiophénone) utilisée contre le SARS-CoV-2 en se basant sur la méthode QSAR. Notre objectif est de développer un modèle fiable entre l'activité inhibitrice mesurée (la concentration d'inhibition médiane: IC<sub>50</sub>) de cette famille d'inhibiteurs en fonction des paramètres structuraux moléculaires (appelés descripteurs), afin de prédire l'activité biologique de nouvelles molécules.

### 2. Méthodologie générale d'une étude QSAR

Pour aboutir à un modèle QSAR explicatif et prédictif on doit [3] :

- Constituer une base de données à partir des mesures expérimentales fiables de l'activité de chaque composé.
- Sélectionner les descripteurs en relation avec l'activité étudiée.
- Diviser cette base de données, aléatoirement, en une série d'apprentissage (Training set) qui contient généralement les 2/3 de la base de données et une série de test (Test set) constituée par le 1/3 restant.
- Etablir des modèles mathématiques en utilisant la série d'apprentissage.

- Valider les modèles élaborés en utilisant la série de test et calculer leurs paramètres statistiques de validation externe.

### 3. 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.

#### Données expérimentales :

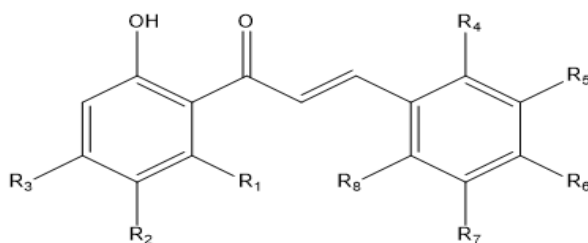
La propiophénone ou benzoyléthane (BzEt) est un composé aromatique de la famille des phénones, de formule brute  $C_9H_{10}O$ . C'est un intermédiaire dans la préparation d'autres composés, et un intermédiaire dans la synthèse de composés pharmaceutiques et autres composés organiques avec son odeur florale douce, elle est également le composant de certains parfums [4].

Un ensemble de données de 20 composés dérivés de propiophénone a été choisi à partir de la littérature avec les valeurs de l'activité inhibitrice médiane [5].

Les molécules utilisées dans cette étude possèdent des propriétés pharmacologiques. La nature des groupements qui se lient sur les noyaux benzéniques de base des molécules étudiées affecte sur leurs propriétés physicochimiques et par conséquent sur leurs propriétés pharmacologiques.

La concentration d'inhibition médiane (IC50) de cette famille d'inhibiteurs a été utilisée pour développer un modèle RLM

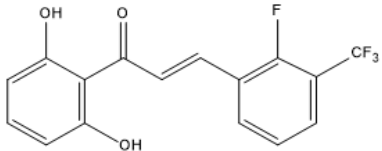
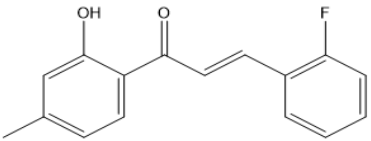
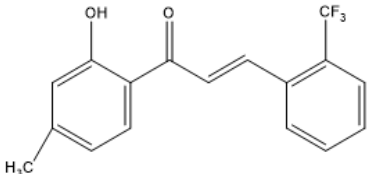
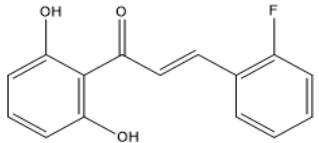
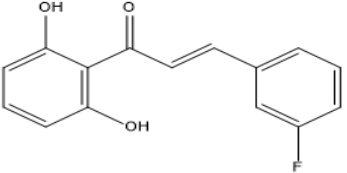
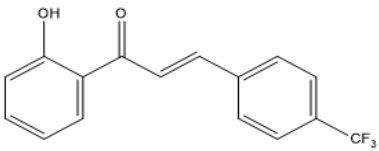
Nous avons étudié l'activité inhibitrice 20 dérivés de la propiophénone représenté par le cologarithme décimal d'IC50 (pIC50) par la modélisation QSAR.



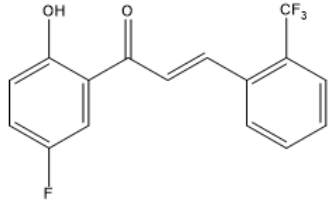
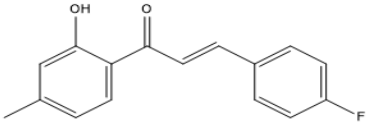
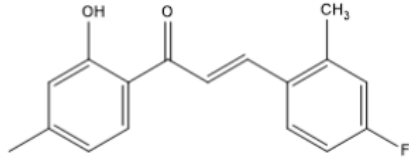
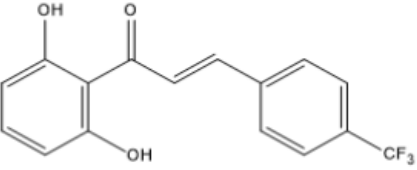
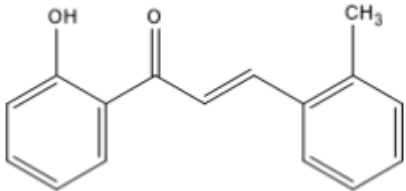
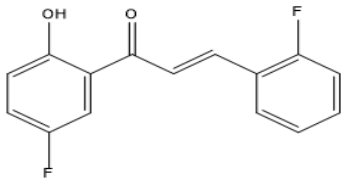
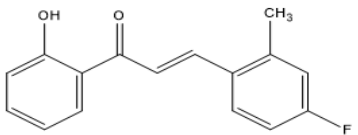
**Figure 01** : Structure générale des dérivés de la propiophénone

## Chapitre III : Résultats et discussion

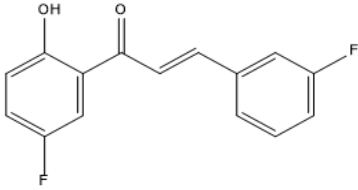
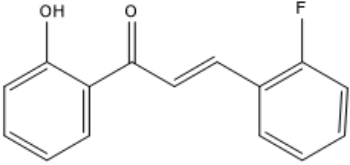
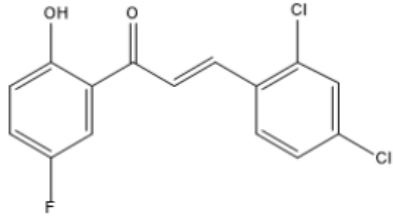
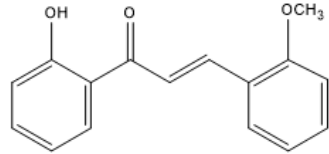
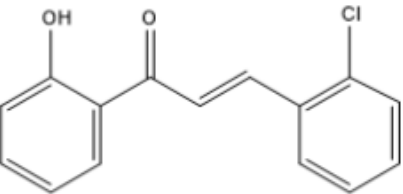
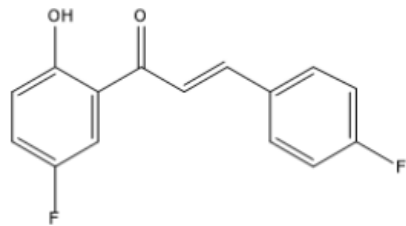
**Tableau 01** : Structures des dérivés de la propiophénone et valeurs de pIC<sub>50</sub> .

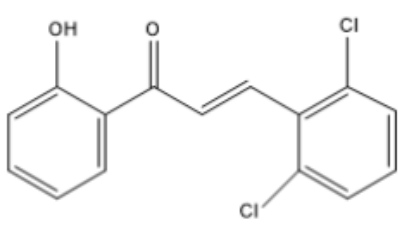
Composés	La structure	Le nom	PIC <sub>50exp</sub> (μmol)
C1		(E)-1-(2,6-dihydroxyphenyl)-3-(2-fluoro-3-(trifluoromethyl)phenyl)prop-2-en-1-one	8.921
C2		(E)-3-(2-fluorophenyl)-1-(2-hydroxy-4-methylphenyl)prop-2-en-1-one	6.983
C3		(E)-1-(2-hydroxy-4-methylphenyl)-3-(2-(trifluoromethyl)phenyl)prop-2-en-1-one	6.961
C4		(E)-1-(2,6-dihydroxyphenyl)-3-(2-fluorophenyl)prop-2-en-1-one	6.974
C5		(E)-1-(2,6-dihydroxyphenyl)-3-(3-fluorophenyl)prop-2-en-1-one	5.942
C6		(E)-1-(2-hydroxyphenyl)-3-(4-(trifluoromethyl)phenyl)prop-2-en-1-one	6.005

### Chapitre III : Résultats et discussion

C7		(E)-1-(5-fluoro-2-hydroxyphenyl)-3-(2-(trifluoromethyl)phenyl)prop-2-en-1-one	7.018
C8		(E)-3-(4-fluorophenyl)-1-(2-hydroxy-4-methylphenyl)prop-2-en-1-one	7.854
C9		(E)-3-(4-fluoro-2-methylphenyl)-1-(2-hydroxy-4-methylphenyl)prop-2-en-1-one	7.983
C10		(E)-1-(2,6-dihydroxyphenyl)-3-(4-(trifluoromethyl)phenyl)prop-2-en-1-one	7.983
C11		(E)-1-(2-hydroxyphenyl)-3-(o-tolyl)prop-2-en-1-one	4.995
C12		(E)-1-(5-fluoro-2-hydroxyphenyl)-3-(2-fluorophenyl)prop-2-en-1-one	7.004
C13		(E)-3-(4-fluoro-2-methylphenyl)-1-(2-hydroxyphenyl)prop-2-en-1-one	7.921

### Chapitre III : Résultats et discussion

C14		(E)-1-(5-fluoro-2-hydroxyphenyl)-3-(4-fluoro-2-methylphenyl)prop-2-en-1-one	7.983
C15		(E)-3-(2-fluorophenyl)-1-(2-hydroxyphenyl)prop-2-en-1-one	6.900
C16		(E)-3-(2,4-dichlorophenyl)-1-(5-fluoro-2-hydroxyphenyl)prop-2-en-1-one	5.853
C17		(E)-1-(2-hydroxyphenyl)-3-(2-methoxyphenyl)prop-2-en-1-one	5.772
C18		(E)-3-(2,6-dichlorophenyl)-1-(2-hydroxyphenyl)prop-2-en-1-one	5.767
C19		(E)-1-(5-fluoro-2-hydroxyphenyl)-3-(4-fluorophenyl)prop-2-en-1-one	7.699

C20		(E)-3-(2,6-dichlorophenyl)-1-(2-hydroxyphenyl)prop-2-en-1-one	6.006
-----	---	---	-------

### 4. Traitements des données :

Le traitement des données pour l'étude QSAR passe par trois étapes :

- Etape 1 : Préparation de l'ensemble de données.
- Etape 2 : Calcul des descripteurs.
- Etape 3 : Modélisation

#### 4.1. Préparation de l'ensemble de données :

Dans cette étape, nous avons utilisé deux logiciels :

**Chem3D Ultra Bio 18.0** [6] : Ce logiciel est utilisé pour la construction en trois dimensions des molécules étudiés (propiophénone) dans ce manuscrit.

**Gaussian 09** [7] : 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.

Après construction, chaque molécule a été optimisée en utilisant le logiciel Gaussian avec la fonctionnelle B3LYP associée à la base 6-31G (d, p).

#### 4.2. Calcul des descripteurs :

Nous avons utilisé le logiciel **DRAGON V5.4** [8] qui est un logiciel destiné au calcul des descripteurs moléculaires. Ces descripteurs 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.

Les géométries optimisées ont été transférées dans le logiciel Dragon 5.4 qui calcule plus de 1664 descripteurs.

### 4.3. Modélisation :

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 les méthode **RLM** au moyen du logiciel Molegro Data Modeller 03[9]

### 5. Résultats et discussion :

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

Les résultats ainsi obtenus ont été traité par le Mnitab 17 .

**MINITAB 17 [10]** : En utilisant ce programme pour obtenir une régression linéaire.

#### 5.1. Modèle RLM :

Le modèle obtenu par RLM est résumé par l'équation de régression suivante :

$$\text{pIC}_{50} = 2,59 - 8,55 \text{ PCR} + 3,470 \text{ MATS8m} + 102,7 \text{ G1m} \quad (\text{Eq.1})$$

Ce modèle reliant la variable dépendante  $\text{pIC}_{50}$  avec les trois descripteurs, est accompagné par les mesures statistiques suivantes : coefficient de corrélation multiple  $R$ , coefficient de détermination  $R^2$ , coefficient de détermination ajusté  $R^2_{\text{adj}}$ , coefficient de prédiction  $Q^2$ , test de Fisher  $F$  et l'écart type  $S$ .

Les paramètres statistiques obtenus par la régression sont rassemblés dans le tableau suivant :

**Tableau 02:** Paramètres statistiques RLM de 3 descripteurs

<b>R</b>	<b>R<sup>2</sup></b>	<b>Q<sup>2</sup></b>	<b>R<sup>2</sup><sub>ajust</sub></b>	<b>F</b>	<b>S</b>
0.920	0.8471	0.8471	0.8119	24.02	0.47

Le modèle QSAR doit considérer un  $R^2 > 0.6$  pour qu'il soit valide. La valeur du coefficient de corrélation multiple  $R = 0.920$  (modèle), indique qu'il existe une forte corrélation entre les différents descripteurs utilisés et l'activité biologique dans ( $\text{pIC}_{50}$ ). Les valeurs des coefficients de détermination  $R^2$  et  $R^2$  ajusté prouve une bonne variabilité du modèle et expliquent respectivement plus de 84% et de 81% de la variance des valeurs observées de l'activité biologique. La valeur  $F$  est jugé statistiquement significative au niveau de 95%, étant donné que la valeur  $F$  calculée est supérieure par rapport aux valeurs du tableau critiques [11,12]

La statistique de Fisher F nous permet de vérifier la présence ou l'absence de corrélation entre l'ensemble de descripteurs et l'activité biologique.

\*Notre modèle a de très bonnes capacités prédictives confirmées par la valeur de  $Q^2$  qui est supérieure à 90%. La statistique de Fisher montre que notre modèle est significatif

### 5.2. Qualité d'ajustement :

Nous avons ensuite examiné la table des coefficients pour vérifier la signification de chaque descripteur et sa contribution dans l'explication de l'activité biologique. L'utilisation des valeurs de la statistique test de Student, affichées dans la table ci-dessous pour chaque descripteur, nous permet de vérifier la présence ou l'absence de corrélation entre chaque descripteur et l'activité biologique

**Tableau 03** : Valeurs de test de Fisher et test de Student.

source	DL	Valeur de F	Valeur de P
<b>régression</b>	3	24.02	0.000
<b>PCR</b>	1	10.80	0.006
<b>MATS8m</b>	1	32.52	0.000
<b>G1m</b>	1	25.90	0.000

**D'après le tableau:**

#### Test de Fisher :

La valeur du test de Fisher  $F= 24.02$  et  $P=0.000$

$P < 0.05$        $\beta_j \neq 0$

$$F > F_{\text{seuil}}$$

F validé, La liaison globale est significatif

#### Test de student :

$$\left\{ \begin{array}{l} p = 0.000 \\ P < 0.05 \\ \beta_j \neq 0 \end{array} \right\}$$

donc : test de student est validé .

La liaison globale est significative

$\beta_j$  : coefficient correspond au descripteur j

\*Les descripteurs entrant dans le modèle, leurs classes et deux brèves définitions sont donnés dans le tableau



**Tableau 04:** Définition et Classification des descripteurs

Descripteurs	Classe	Définition
PCR	Comptes de parcours et de Chemin	Ratio de parcours multiple
MATS8m	Autocorrélations 2D	Autocorrélation de Moran du décalage 8 pondéré par la masse
G1m	Descripteur géométrique	Indice WHIM directionnel de symétrie du 1er composant/pondéré par les masses atomiques

### 5.3. Analyse des résultats :

$R^2=0.847$  et  $R^2$  ajusté=0.8119 Donc pourcentage **84,7 1%** de variation de la variable à modéliser. Cela nous indique que **84,7%** 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.

#### Interprétation:

Les trois descripteurs sont utilisés pour obtenir des informations pertinentes sur la molécule. Ils sont importants pour construire un modèle QSAR robuste et hautement prédictible.

Dans l'équation de modèle, le coefficient positif élevé de l'Indice de symétrie WHIM explique qu'il entraîne une augmentation de l'activité biologique de ces composés. Cela signifie que ce descripteur géométrique (**G1m**) basé sur des indices statistiques et des projections calculées des atomes le long des axes principaux, nous donne d'importantes informations sur **pIC50**.

Par conséquent, plus la valeur de ce descripteur est élevée, plus grande sera l'activité biologique.

Le coefficient de descripteur (**MATS8m**) a un signe positif dans le modèle, donc toute augmentation de cette autocorrélation entraîne une augmentation de l'activité biologique.

## Chapitre III : Résultats et discussion

Le coefficient de descripteur (**PCR**) a un signe négatif dans le modèle, ainsi une augmentation d'activité peut être obtenu en diminuant sa valeur. Pour modéliser un nouveau composé avec haute activité, le nouveau composé doit avoir la valeur la plus basse possible de ce descripteur.

### 4. Validation du modèle :

Le modèle QSAR doit être obligatoirement validé avant d'être utilisé à fins prédictives.

Les paramètres statistiques calculés  $R^2_{test}$  coefficient de corrélation et le coefficient de corrélation entre les valeurs prédites et expérimentales pour la série de test  $R_0^2$  sont rassemblés dans le tableau suivant :

**Tableau 05** : Paramètres statistiques de test

$R^2_{test}$	$R_0^2$	$ R^2 - R_0^2 $	$\frac{R^2 - R_0^2}{R^2}$
0.83	0.84	0.01	-0.01

Selon Tropsha et ses collaborateurs [13] un modèle de QSAR ne possède une puissance prédictive acceptable que si les conditions suivantes sont satisfaites :

$$R^2_{test} > 0.6$$

$$\frac{R^2 - R_0^2}{R^2} < 0.1 \quad \text{et} \quad 0.85 \leq K \leq 1.15$$

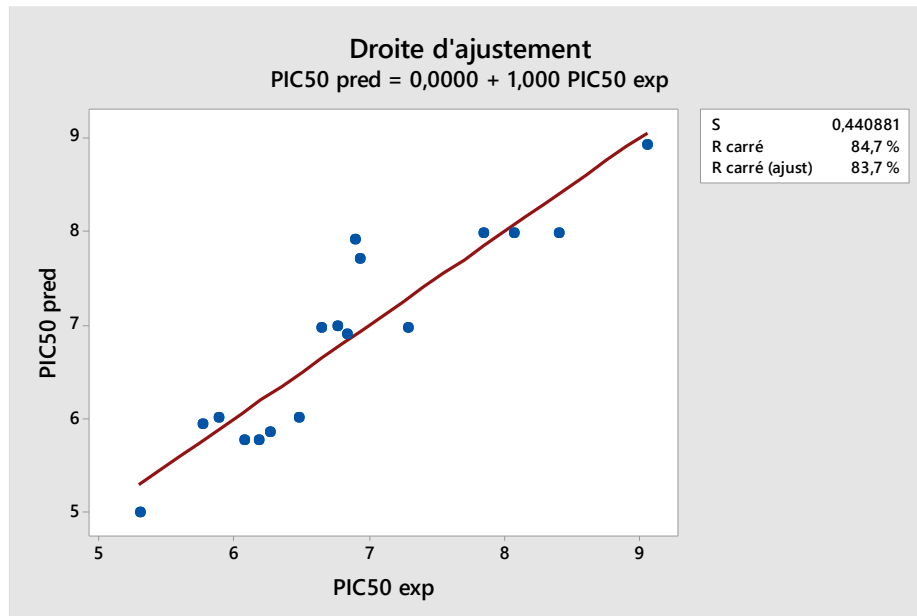
$$|R^2 - R_0^2| \leq 0.3$$

Dans notre modèle, les conditions suivantes sont satisfaites :

- $R^2_{test} = 0.83 > 0.6$
- $\frac{R^2 - R_0^2}{R^2} = -0.01 < 0.1$
- $|R^2 - R_0^2| = 0.01 \leq 0.3$
- $K = 1 \quad 0.85 \leq K \leq 1.15$

D'après les critères et les conditions, on déduit que notre modèle à trois Descripteurs est fiable.

Les résultats montrent des valeurs élevées du coefficient de détermination issus de la validation externe ( $R^2_{\text{test}} = 0.83$ ,  $R_0^2 = 0.84$ ), ceci prouve une bonne stabilité et une bonne prédiction du modèle.



**Figure 02 :** Variation des valeurs expérimentales en fonction des valeurs prédites

### Interprétation du graphe :

D'après la figure précédente, on remarque une faible dispersion autour de la droite 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 84%, pour les deux ensembles (Calibration et Validation) ainsi que  $R_{\text{ajust}}$  et la valeur de l'écart type ( $S = 0.44$ ). Cela prouve la performance du modèle établi.

### **Conclusion :**

Le modèle QSAR (Eq.1) ne permet pas d'extraire toute information pratique sur la relation entre activité biologique et structure moléculaire du composé, en raison de la nature complexe des trois descripteurs sélectionnés. L'utilité principale du modèle QSAR est qu'il peut être utilisé pour prédire l'activité biologique d'un nouveau composé, dont l'activité biologique n'est pas encore connue.

### Références :

1. V. N. Viswanadhan, M.R.Reddy, R. J. Bacquet et M.D. Erion.J. 1993, p 1019.
2. V. N. Viswanadhan, A.K. Ghoseet, J.N. Weinstein , 1990, p 1039.
3. K. A. Benazzouz Hicham, Université Abou Bekr Belkaïd Tlemcen,2016.
4. Merriam-Webster.com , 2 juin 2012.
5. M. Jovanovic, N. Turkovic, B. Ivkovic, Z. Vujic, K. Nikolic ,S. Grubisic, 2021.
6. [http: www.chemistrysoftware.com](http://www.chemistrysoftware.com) , 2021.
7. 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, HP .Hratchian , AF. Izmaylov, J. Bloino, G. Zheng, JL. Sonnenberg , M.Hada, M. Ehara , K .Toyota , R. Fukuda , J.Hasegawa, M, Ishida,T .Nakajima, Y. Honda , O .Kitao , H. Nakai , T. Vreven , J.Montgomery, J. Peralta, F. Ogliaro, M.Bearpark , J.Heyd , E.Brothers , K. Kudin, V.Staroverov , R. Kobayashi , J. Normand , K. Raghavachari , A. Rendell , J. Burant , S .Iyengar S, J .Tomasi , M .Cossi , N.Regga, J.Millam , M. Klene , J. Knox, J. Cross , V.Bakken , C.Adamo , J. Jaramillo , R.Gomperts, R.Stratmann , O.Yazyev , A.Austin , R.Cammi , C.Pomelli , J.Ochterski , R.Martin, K.Morokuma , V.Zakrzewski , G.Voth , P. Salvador ,J. Dannenberg , S.Dapprich ,A. Daniels , O.Farkas, J.Foresman, J.Ortiz , J.Cioslowski ,2009.
8. R. Todeschini. Milano Chemometrics and QSAR Research Group, 2006
9. A .Golbraikh , A .Tropsha. J. Mol. Graphics et Modelling, 2002, P 269–276.
10. J. Clautriaux , A. Delvaux, 1994, p 20.
11. Z.Almi, et al. International Letters of Chemistry, Physics and Astronomy,2014, p 113-124
12. K. Dermeche, N. Tchouar, S. Belaidi, T. Salah, Journal of Bionanoscience, 2015, 9, P 395-400.
13. A. Golbraikh ,2002 , p 269–276
14. H. HADDAG , mémoire de majister , Chimie et environnement, Université BADJI Mokhtar -Annaba, 2009.

# Conclusion générale

## Conclusion générale

---

Afin de lutter contre SARS\_CoV\_2 , les chercheurs ont mené de nombreuses expériences, Plus précisément en tentant d'inhiber l'enzyme de conversion de l'angiotensine-2 (ACE2), un récepteur qui facilite l'entrée virale dans les cellules hôte, et inhibe protéases essentielles au cycle de vie du virus, par des médicaments et des composés naturels.

Dans le but d'étudier l'activité inhibitrice d'une famille des Anti-HIV contre le SARS-CoV-2 par une des méthodes in silico fréquemment utilisée dans ce domaine, nous avons appliqué la méthode QSAR (Quantitative structure Activity Relationship) sur une série des dérivés de propiophénone.

Afin de trouver une corrélation entre l'activité inhibitrice mesurée des composés (Concentration Inhibitrice Médiane IC50) et certains descripteurs moléculaires, un grand nombre de descripteurs moléculaires a été calculé à partir des géométries préalablement optimisées avec la fonctionnelle B3LYP/6-31G (d, p) intégré au programme Gaussian.

Pour chaque composé plus de 300 descripteurs sont calculés par le logiciel Dragon. L'ensemble des molécules a été divisé en 2 parties « training » et « test » avec un choix aléatoire, puis les meilleurs descripteurs ont été sélectionnés.

Les données ont été modelées en utilisant l'analyse de la régression linéaire multiple (MLR = Multiple Linear Régression) implanté dans le logiciel Molegro Data Modeller qui nous a permis d'aboutir à une corrélation entre pIC50 qui détermine l'activité inhibitrice de cette famille d'inhibiteurs (dérivés de propiophénone) et trois descripteurs.

Le traitement statistique suivi par le Minitab montre que le modèle obtenu à une qualité statistique élevée et fiable et de faible erreur de prédiction ( $R=0.92$ ,  $R^2=0.84$ ,  $Q^2=0.84$ ). Une validation rigoureuse externe 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 du celles observées ( $R_0^2 = 0.84$ ).

Ce modèle QSAR qui indique l'existence d'une forte corrélation entre trois 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.