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

UNIVERSITE YAHIA FARES DE MEDEA
Faculté des Sciences et de la Technologie
Département de Génie de Procédés et Environnement

MEMOIRE DE MASTER

En vue de l'obtention du Diplôme de Master en Génie des Procédés
Option : *Procédés Pharmaceutiques*

Thème

MODELISATION DE LA BIODISPONIBILITE DES MEDICAMENTS A PARTIR DE LEURS STRUCTURES CHIMIQUES

Etudié par :
M^elle^ Amina FATOUMI
M^me^ Chafika TERRAS

Dirigé par :
M S. BITAM

Soutenu devant le jury :

D^r^. Mohamed TIR	**Maitre de conférences, UYFM**	**Président**
D^r^. Mahdi YOUCEF	**Maitre de conférences, UYFM**	**Examinateur**

Promotion : 2015-2016

Remerciements

Tout d'abord nous rendons grâce à Allah, le tout puissant qui nous a donné la santé et la volonté pour arriver à ce stade.

Nous exprimons nos profonds remerciements à toutes les personnes qui ont contribué de prés ou de loin à la réalisation de ce travail.

Nos plus vifs remerciements iront également à messieurs les membres de jury de mémoire, pour avoir bien voulu consacrer une partie de leurs temps à examiner et rapporter ce travail.

Ce travail a été réalisé sous la diréction de monsieur **S.Bitam,** qu'il trouve ainsi l'expression de notre profonde reconnaissance pour sa compétence, ses encouragements et ses précieux conseils durant le déroulement de ce travail.

En fin nous exprimons nos profonds remerciements à tous les enseignants qui on contribué à notre formation.

Dédicace :

Je dédie ce modeste travail :

A ma mère qui a fait de sa vie le chemin de ma réussite et de sa jeunesse une bougie qui brule pour m'éclairer pendant ma traversée.

A mon père, école de mon enfance, qui a été mon ombre durant toutes les années des études, et qui a veillé tout au long de ma vie
à m'encourager, à me donner l'aide et à me protéger. Que dieu les garde et les protège.

A mon cher mari Hasni et ma petite belle fille Manar

A mon frère Abdelkader

A mes sœurs Soumia et son fils Abdelmoumen, Naziha et Rania

A toute ma famille

A tous les professeurs et enseignants qui ont participé à ma bonne formation depuis mes débus jusqu'à la fin de mon cursus universitaire.

A mes amies de la promotion de génie de procédés pharmaceutique 2015 -2016.

A tous ceux qui me sont chères

A tous ceux qui m'aiment et à tous ceux que j'aime

Dédicace

Je dédie ce modeste travail à :

Mon père, mon premier encadrant, depuis ma naissance ;

Ma très chère mère : qu'elle trouve ici l'hommage de ma gratitude qui, si grande qu'elle puisse être, ne sera à la hauteur de ses sacrifices et ses prière pour moi ;

Mes frères et ma sœur qui je souhaite beaucoup de réussite et de bonheur ;

Toute ma famille ;

Toutes mes amies qui me sont chers ;

Tous ceux que j'aime et qui m'aiment : qu'ils trouvent ici l'expression de mes sentiments les plus dévoués et mes vœux les plus sincères ;

Que dieu le tout puissant vous préserve tous et vous procure sagesse et bonheur.



Amina

الملخص

تعد الجرعات المتناولة عن طريق الفم الأسلوب الأكثر شيوعا في استعمال الأدوية، وتركيزات البلازما النهائية من الأدوية تتعلق بالتوافر البيولوجي.في هذه الدراسة ، تم تطوير علاقة بين بنية الدواء الكمية ووفرته في الدورة الدموية (QSAR) لمجموعة متنوعة من المركبات للسماح بالتنبؤ بالتوافر البيولوجي للدواء. وقد اتخذت بيانات التوافر البيولوجي ل 130 مركب من المراجع، ومن الهياكل الجزيئية تم إنشاء 950 واصف نظري كما تم استخدام الانحدار الخطي المتعدد لتطوير معادلة الانحدار على أساس هذه المركبات. تضمنت معادلة الانحدار النهائية ثمانية واصفات جزيئية . وتحقق ارتباط 0.916 لمجموعة اختبار التنبؤات التوافر البيولوجي مقارنة مع قيم المراجع.

الكلمات المفتاحية: العلاقة الكمية بين البنية والفعالية ,الانحدار الخطي المتعدد,الواصفات الجزيئية,التوافر البيولوجي.

Résumé

L'administration orale est la méthode la plus courante de l'administration du médicament, et les concentrations plasmatiques finales du médicament dépendent de sa biodisponibilité. Dans la présente étude, une relation structure-activité quantitative (QSAR) a été développée pour un large éventail de composés pour permettre la prédiction de la biodisponibilité du médicament. Les données de la biodisponibilité de 130 composés ont été prises à partir de la littérature, et les structures moléculaires de 950 descripteurs théoriques ont été générées. Une régression linéaire multiple a été utilisée pour développer une équation de régression basée sur ces composés. L'équation de régression finale comprenait huit descripteurs moléculaires.Une corrélation de 0,916 a été atteinte pour les tests de prédictions de la biodisponibilité comparativement aux valeurs de la littérature.

Mots-clés: QSAR; Régression linéaire multiple ; Descripteurs moléculaires; biodisponibilité.

Abstract

Oral dosing is the most common method of drug administration, and final plasma concentrations of the drug depend upon itsbioavailability. In the current study, a quantitative structure–pharmacokinetic relationship (QSAR) was developed for a diverserange of compounds to allow prediction of drug bioavailability. Bioavailability data for 130 compounds was taken from theliterature, and from the molecular structures 950 theoretical descriptors were generated. Stepwise regression was employed todevelop a regression equation based on 130 compounds. The final regression equation included eight descriptors molecular. Predicted bioavailability for the training set agreed more closely for drugs exhibiting mid-range literature bioavailabilityvalues. A correlation of 0.916 was achieved for test set bioavailability predictions when compared with literature values.

Keywords: QSAR; Multiple linear regression; Molecular descriptors; Bioavailability.

Table des matières

Abréviations III
Nomenclatures V
Liste des tableaux VI
Liste des figures VII
Introduction générale **1**
Chapitre I. Généralité sur la biodisponibilité **3**
I.1. Introduction 3
I.2. La pharmacocinétique 4
I.2.1. Définition 4
I.2.2. Absorption 4
I.2.2.1. Evaluation de l'absorption : la biodisponibilité (F) 6
a. La biodisponibilité absolue 6
b. La biodisponibilité relative 6
1) Surface Sous la Courbe (SSC) 7
2) Savoir interpréter la biodisponibilité absolue 7
3) Les facteurs influant sur la biodisponibilité 8
4) Le facteur vitesse 8
5) L'intérêt de l'étude de la biodisponibilité 10
6) Pourquoi recherche-t-on des biodisponibilités élevées ? 10
I.2.3. Distribution 10
I.2.4. Métabolisme 11
I.2.5. Elimination 12
I.3 La pharmacodynamique 12
I.4. Conclusion 12
Chapitre II. Modélisation moléculaire **13**
II.1. Introduction 13
II.2. La méthode QSAR : relation quantitative structure à activité 14
II.2.1. Historique 14
II.2.2. Principe 15
II.2.3. Importance des bases de données 17
II.2.4. Descripteurs moléculaires 18
II.2.4.1 Les descripteurs moléculaires théoriques 19
II.2.4.1 Les descripteurs moléculaires empiriques 21

Table des matières

II.2.5. Sélection et application 22
II.2.5.1 Carte d'identité d'une molécule 22
II.2.5.2Utilisation des descripteurs moléculaire 22
II.2.5.3Sélection à l'aide du modèle QSAR 23
II.2.6. Méthodes d'analyse de données 23
II.2.6.1Régression linéaire multiple (MLR) 25
1) Description de la méthode 26
2) Confiance accordée aux résultats 28
3) Qualité de la régression linéaire 28
a. Coefficient de corrélation (R) 28
b. Coefficient de détermination (R^2) 28
c. Coefficient de détermination ajusté (R^2ajusté) 30
d. Ecart type (s) 31
e. Coefficient de prédiction (Q^2) 31
f. Colinéarité des variables explicatives 31
4) Validations des modèles 32
II.2.7. Applications des méthodes QSAR 32
II.3. Conclusion 34
Chapitre III. Partie expérimentale 35
III.1. Introduction 35
III.2. Méthodologie de calcul 35
III.2.1. Calcul des descripteurs moléculaires 35
III.2.2. Elaboration de la base de données 42
III.2.3. Prétraitement de la base de données 43
III.2.4. L'étape de la régression linéaire multiple (MLR) 43
III.3 Résultats et Discussion 45
III.3.1. Equation de la régression linéaire multiple 45
III.3.2. Interprétation 47
III.4. Conclusion 53
Conclusion générale 54
Références bibliographiques 55
Annexes 58

Liste des Abréviations

Abréviation	Signification
1D	1 Dimension
2D	2Dimension
3D	3 Dimensions
4D	4Dimensions
C	la Concentration du soluté non ionisé dans la phase aqueuse
C′	la Concentration du soluté dans la phase huileuse
Cmax	Concentration maximale
D	Descripteur des structures moléculaires
EHOMO	Energie de l'orbitale la plus haute occupée
ELUMO	Energie de l'orbitale la plus basse vacante
ESS	Somme des Carrés Expliqués
Ka	vitesse d'absorption
MLR	Régression Multi-Linéaire
PA	Principe Actif
PCR	la régression en composante principale (*Principale Componant Regression)*
PLS	régressions aux moindres carrés partiels (*Partial Least Squares)*
QSAR	Quantitative Structure-Activity Relationship
RNA	les Réseaux de Neurones Artificiels
RSS	Somme des Carrée Résiduels

Abréviations

Sig	Significativité
SLR	Régression Linéaire Simple
SPSS	Statistical Package for the Social Sciences
SSC	Surface Sous la Courbe
Tmax	Temps maximal
TSS	Somme des Carrés Totaux
X	variables indépendantes
Y	variable dépendante
α	la polarisabilité
β	coefficients de régression
ε	Erreur
μ	Potentiel chimique électronique

Nomenclatures ET Symboles

Lecture	Désignation	Unité
1/F	L'inverse de la biodisponibilité	(-)
F	la biodisponibilité	(-)
K	constantes respectives de la réaction	(-)
K_0	constantes respectives d'une référence	(-)
log P	coefficient de partage eau/octanol	(-)
PE	l'erreur de prédiction	(-)
PRESS	La somme résiduelle prédictive des carrés (*Predictive Residual Sum of Squares*)	(-)
Q^2	Ecart type	(-)
R	Coefficient de corrélation	(-)
R^2	Coefficient de détermination	(-)
R^2ajusté	Coefficient de détermination ajusté	(-)
R^2_{CV}	coefficient de corrélation de validation croisée	(-)
R^2_j	coefficient de détermination de la régression de la variable Xj sur les autres variables	(-)
S	Ecart type	(-)
S_{PRESS}	la validation standard des erreurs de prédiction	(-)
VIF	Facteur d'inflation de la variance	(-)
ρ	une constante de réaction	(-)

Liste des tableaux

Numéro	Titre du tableau	Page
	Chapitre 3	
Tableau III.1	Liste des médicaments utilisés dans cette étude avec leurs valeurs de biodisponibilité expérimentales.	37-38-39-40
Tableau III.2	Les descripteurs moléculaires de cette étude	45
Tableau III.3	Récapitulatif du modèle	46
Tableau III.4	Tableau ANOVA	46
Tableau III.5	Comparaison entre les valeurs expérimentales et les valeurs prédites	48-49-50-51

Liste des figures

Figure	Titre de la figure	Page
Chapitre 1.		
Figure I.1	Comparaison entre deux voies d'administration (voie intra-vineuse et voie orale)	7
Figure I.2	La concentration du médicament en fonction du temps (voie orale)	8
Figure I.3	Comparaison des profils pharmacocinétiques de 2 médicaments	9
Chapitre 2		
Figure II.1	Modèle d'étude quantitative des relations structure/activité (QSAR)	16
Figure II.2	Influence d'une valeur atypique (A) sur une corrélation	17
Chapitre 3		
Figure III.1	Image de ChemSpider	36
Figure III.2	Image de ChemBio3D Ultra 14.0	37
Figure III.3	Image de PaDel-Descriptor	41
Figure III.4	Image représente l'ouverture d'un fichier (.csv) sur CSV Viewer	41
Figure III.5	la base de données sur Excel	42
Figure III.6	La première étape sur SPSS	43
Figure III.7	La deuxième étape sur SPSS	44
Figure III.8	Corrélation entre les valeurs de l'inverse de la biodisponibilité prédites et expérimentales	51

Introduction générale

La notion de la biodisponibilité a été introduite suite à l'observation d'accidents toxiques dus à des surdosages ou thérapeutiques attribués à des sous dosages. La biodisponibilité correspond à la fois à la quantité relative du principe actif absorbé à partir d'une forme pharmaceutique qui atteint la circulation générale et la vitesse à laquelle se produit ce phénomène.

On étudie la biodisponibilité pour plusieurs raisons, comme la formulation d'un nouveau médicament, la modification de la quantité du principe actif dans le médicament, le changement de la voie d'administration, la fabrication d'une nouvelle forme pharmaceutique…

La découverte des nouveaux médicaments de la manière la plus efficace et la moins coûteuse possible constitue un enjeu majeur pour les années à venir. Il est admis que, en moyenne, pour une molécule qui arrive sur le marché en tant que médicament innovant, 10 000 molécules sont synthétisées et testées. De plus, le développement d'un médicament demande généralement entre 10 et 15 ans de recherches. Il s'agit en effet de trouver une molécule qui doit à la fois présenter des propriétés thérapeutiques particulières, et posséder le minimum d'effets secondaires indésirables. Le prix de revient d'un médicament est essentiellement dû à ces synthèses longues et coûteuses.

Pour cette raison, l'industrie pharmaceutique s'oriente vers de nouvelles méthodes de recherche, qui consistent à prédire les activités de molécules avant même que celles-ci ne soient synthétisées. Les relations structure-activité ou QSAR (Quantitative Structure-Activity Relationships) sont développées en réponse à ce besoin.

La découverte d'une telle relation permet de prédire les propriétés physiques et chimiques et l'activité biologique de composés, de développer de nouvelles théories ou de comprendre les phénomènes observés.

Elle permet également de guider les synthèses de nouvelles molécules, sans avoir à les réaliser, ou à analyser des familles entières de composés.

Les relations entre les structures des molécules et leurs activités sont généralement établies à l'aide de méthodes de modélisation par apprentissage statistique. Les techniques usuelles reposent sur la caractérisation des molécules par un ensemble de descripteurs, nombres réels mesurés ou calculés à partir des structures moléculaires. Il est alors possible d'établir une relation entre ces descripteurs et la grandeur modélisée.

Ce manuscrit de thèse est divisé en trois chapitres suivant une introduction générale.

La première partie de ce mémoire présente une généralité sur la biodisponibilité, les types de la biodisponibilité, les facteurs influençant sur la biodisponibilité. On va définir la pharmacocinétique et ses phases : absorption, distribution, métabolisme et élimination de la substance active (ADME) et la pharmacodynamique.

Dans le deuxième chapitre on va définir la méthode QSAR, les descripteurs moléculaires et la régression linéaire multiple (MLR).

Le dernier chapitre présente l'élaboration de la base de données et résume les résultats de la régression linéaire multiple : l'équation de la régression, la validation du modèle obtenu et discussion.

Enfin, on va conclure par une conclusion générale qui résume notre travail : les conclusions tirées à partir de ce travail et les études qui peuvent être envisagées dans le futur.

I.1. Introduction

Le médicament n'est pas un produit commercial comme les autres, c'est toute substance ou composition présentée comme possédant des propriétés curatives ou préventives à l'égard des maladies humaines ou animales, ainsi que tout produit pouvant être administré en vue d'établir un diagnostique médical ou de restaurer, corriger ou modifier leurs fonctions organiques. [1]

Le but essentiel d'une thérapie médicamenteuse est d'atteindre un effet bénéfique avec un minimum d'effets secondaires. Lorsqu'un médicament a été sélectionné pour un patient, le médecin doit déterminer la dose la plus propice à atteindre ce but. [2]

La dose administrée (dose externe) n'est pas forcement égale à la dose qui va réellement agir encore appelée dose interne. Le thérapeute veut connaître l'exposition réelle de son patient qui est liée à la dose interne; la relation entre la dose administrée et la dose interne est donnée par le facteur de biodisponibilité. La mesure de la biodisponibilité est également informative pour les médicaments administrés localement afin d'anticiper la présence ou non d'effet systémiques ou de résidus de médicament pour la médecine vétérinaire. [3]

Une approche rationnelle afin d'atteindre l'objectif du médicament combine les principes de la pharmacologie qui regroupe la pharmacocinétique (devenir du médicament dans l'organisme) et la pharmacodynamie (action du médicament sur l'organisme). Elle doit être parfaitement connue pour mieux en parler dans le cadre du bon usage du médicament. **[**4]

La biodisponibilité c'est un outil essentiel en pharmacocinétique, car la biodisponibilité doit être prise en considération lors du calcul des doses pour des voies d'administration autres qu'intraveineuse. [5]

I.2. La pharmacocinétique

I.2.1. Définition

C'est la partie de la pharmacologie qui étudie, en fonction du temps, le devenir d'une substance après son introduction dans un organisme vivant. Cette destinée comporte essentiellement quatre phases : [6]

- ✓ L'absorption ou résorption,
- ✓ La distribution,
- ✓ Les biotransformations ou métabolismes,
- ✓ L'élimination.

La pharmacocinétique a pour but d'étudier le devenir d'un médicament dans l'organisme. La détermination des paramètres pharmacocinétiques d'un médicament apporte les informations qui permettent de choisir les voies d'administration et d'adapter les posologies pour son utilisation future. [7]

- **Devenir du médicament dans l'organisme**

Après avoir pénétrer dans l'organisme, le principe actif (PA) traverse un nombre variable de cellules pour aboutir dans la circulation sanguine. Au niveau du tube digestif, et en particulier de l'intestin, la résorption est facilitée par la très grande surface de contact avec le contenu intestinal. La voie orale est donc la plus utilisée, même si elle n'est pas adaptée à toutes les situations. [6]

I.2.2. Absorption

L'absorption est le processus par lequel le médicament inchangé passe de son site d'administration à la circulation générale. La voie d'administration du médicament influence cette dernière : la voie intraveineuse est la voie de référence puisque par définition, à la différence des autres voies (orale par exemple) toute la dose administrée atteint la circulation générale.

Des facteurs chimiques et physiologiques vont influencer la résorption des médicaments.

1) Nature chimique

Les molécules de charge neutre traverseront plus facilement les membranes par diffusion.

2) Poids moléculaire

On utilise souvent de petites molécules de poids inférieur à 1 k Da.

3) Solubilité

La solubilité des médicaments dans les lipides joue un rôle important dans leur absorption et est exprimée en coefficient de partage.

$$\textbf{Coefficient de partage} = \frac{\textbf{solubilité dans les lipides}}{\textbf{solubilité dans l'eau}}$$

Plus ce coefficient est élevé, plus rapidement la molécule va se diffuser à travers la membrane lipidique.

4) Variables physiologiques

a) Motilité gastrique : si on augmente ce paramètre, on va limiter l'absorption intestinale du médicament.

b) pH au site d'absorption : les médicaments avec des fonctions acides seront sous une forme non ionisée et traverseront facilement les membranes. À l'inverse, les médicaments à fonctions basiques qui seront complètement ionisés ne traverseront pas la membrane.

c) Aire des surfaces absorbantes : plus elle sera importante, plus le médicament sera absorbé.

d) Élimination pré-systématique : la voie d'administration pourra limiter l'accès des médicaments à la circulation générale.

Par exemple les gouttes oculaires (effet local, très peu entrent dans le sang) ou l'antibiotique pénicilline G (instable à pH acide et donc dégradé dans l'estomac).

e) Ingestion avec ou sans aliments : un médicament pris avec un verre d'eau froide va accélérer la vidange gastrique et donc le temps de contact entre le médicament et les zones absorbantes sera moins important (la dose sera diminuée).

À l'inverse, un médicament pris avec des lipides et une boisson acide retarde la vidange gastrique, le temps de contact est moins important et la densité de la substance dans le sang est plus importante (surdosage).

5) Aspects propres à la résorption digestive :

- Motilité du tube digestif ;
- Interférence avec d'autres substances digestives (lipides, etc.) ;

- Décomposition de la forme galénique par des enzymes digestives, la flore intestinale ou le foie ;
- Constitution même du médicament selon l'agent d'enrobage, le colorant, la taille des particules.

I.2.2.1. Evaluation de l'absorption : la biodisponibilité (F)

La biodisponibilité se définit comme étant la fraction de la dose de médicament administré qui atteint la circulation générale et la vitesse à laquelle elle l'atteint. Elle s'exprime en pourcentage. En pratique, elle correspond à la fraction du principe actif atteignant le compartiment central (volume virtuel dans lequel la concentration serait partout identique et que l'on pourrait mesurer en prélevant un échantillon de volume négligeable, généralement le sang dans la réalité) en tenant compte de la vitesse à laquelle ce compartiment est atteint. On parle de fraction car l'utilisation de certaines voies d'administration fait qu'une seule partie du principe actif peut atteindre le sang.

Le facteur quantitatif (F) de la biodisponibilité ne peut être apprécié que par rapport à une forme de référence. On distingue ainsi :

a. La **biodisponibilité absolue** : la forme étudiée est comparée à la forme de référence qui est le médicament administré par voie intraveineuse puisque par définition toute la dose atteint la circulation générale.
b. La **biodisponibilité relative** où la forme de référence est administrée par une autre voie que la voie intraveineuse. Cette forme de référence peut être administrée par la même voie que la forme à tester, mais il s'agit soit d'une autre forme galénique (solution aqueuse, suspension..) soit d'une autre formulation d'une forme commercialisée depuis longtemps (cas des génériques).

En général la quantification de la biodisponibilité (F) s'effectue par comparaison des surfaces sous la courbe des concentrations en fonction du temps (SSC) après administration de chaque forme séparément. [6]

1) Surface Sous la Courbe (SSC)

La surface sous la courbe(SSC) trace la concentration par rapport au temps. Elle correspond à l'intégral de la concentration plasmatique sur un intervalle de temps défini. Elle correspond à la biodisponibilité. Dans le cas de la voie intraveineuse le médicament est directement dans le sang, et donc la surface sous la courbe représente 100% de biodisponibilité. [4]

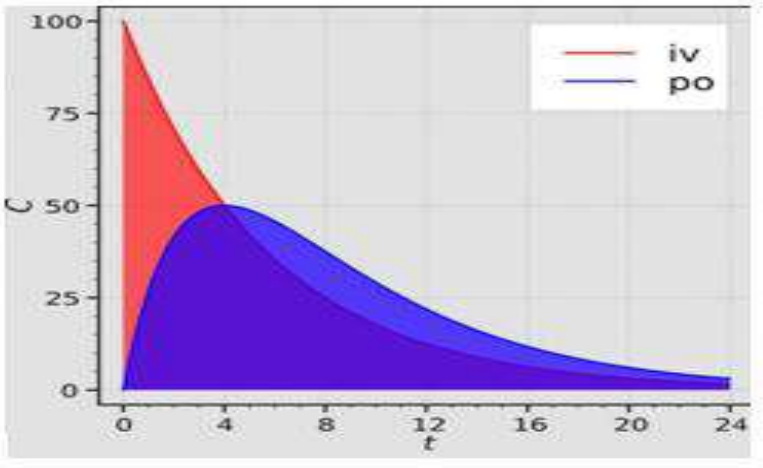


Figure I.1 Comparaison entre deux voies d'administration (voie intra-vineuse (iv) et voie orale (po))

2) Savoir interpréter la biodisponibilité absolue

$$\mathbf{F=}\frac{\mathbf{SSC\ (orale)}}{\mathbf{SSC\ (iv)}}\ \text{(biodisponibilité absolue)}$$

Avec :

SSCpo : surface sous la courbe de la voie orale

SSCiv : la surface sous la courbe de la voie intra-vineuse

On voit selon cette équation que si toute la dose administrée par voie orale est absorbée (comme en intraveineux) la biodisponibilité absolue de ce produit sera 1(100%).

Une biodisponibilité absolue de 0,5 (50%) pour un produit signifie que seule la moitié de la quantité administrée est retrouvée dans la circulation générale.

Ainsi, la dose contenue dans le comprimé ou la gélule ne reflète pas toujours la dose biodisponible. F est donc par définition compris entre 0 et 1(entre 0 et 100%).

3) Les facteurs influant sur la biodisponibilité

La biodisponibilité est relative à une présentation, c'est-à-dire à un principe actif, sous une forme pharmaceutique, d'un fabricant donné ; elle dépend de :

- la substance elle-même ;
- l'anatomie, c'est-à-dire de la voie d'administration (cf. voies orale ou rectale…) ;
- la forme galénique et du processus de fabrication (très important en pratique) ;
- les facteurs physiologiques (âge, alimentation…) ;
- les facteurs pathologiques (insuffisances hépatiques…). [7]

D'autres facteurs peuvent s'inclure, mais ne sont pas limités :

- Les propriétés physiques de la drogue (hydrophobie, pKa, solubilité);
- La formulation du médicament (à libération immédiate, excipients utilisés, les méthodes de fabrication, à libération modifiée - à libération retardée, à libération prolongée, etc.);
- si la formulation est administrée dans un repas ou à jeun;
- la vitesse de vidange gastrique;
- les différences circadiennes;
- les interactions avec d'autres médicaments/aliments. [8]

4) Le facteur vitesse

Le facteur vitesse est apprécié par la constante de vitesse d'absorption Ka ou plus facilement par la concentration maximale (Cmax) et le temps pour atteindre cette concentration (Tmax).

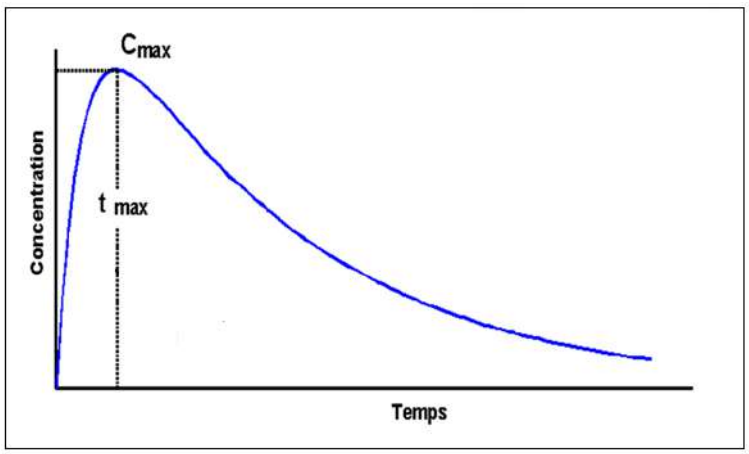
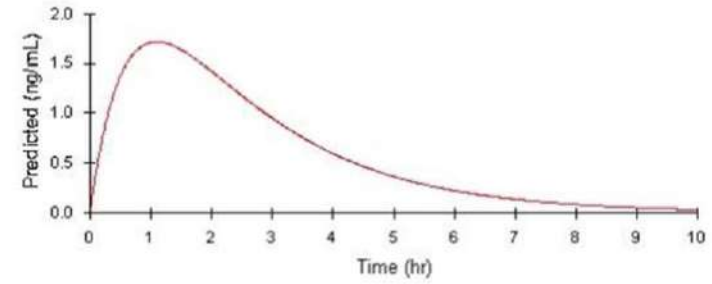


Figure I. La concentration du médicament en fonction du temps (voie orale)

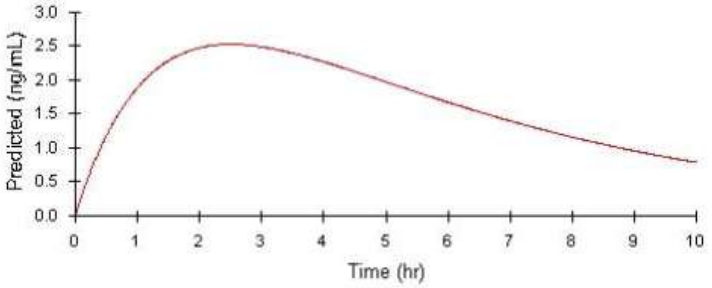
Au même titre que la quantité absorbée, la vitesse d'absorption d'un médicament est un paramètre significatif pour le délai d'action d'un principe actif.

La vitesse de passage est un paramètre prépondérant pour les médicaments destinés à une action rapide (antalgique par exemple) en prise unique ou de courte durée. Pour les traitements chroniques, où une imprégnation constante est recherchée, la notion de Tmax est moins déterminante.

Exemple : Comparaison des profils pharmacocinétiques de 2 médicaments



Médicament A



Médicament B

Figure I. Comparaison des profils pharmacocinétiques de 2 médicaments

T max : plus rapide pour A que B (1 heure contre 2.5 heures)
C max : plus élevée pour B que A (2.5 ng/ml contre 1.7 ng/ml)

5) L'intérêt de l'étude de la biodisponibilité

- La biodisponibilité est intéressante à connaitre, car elle permet, en comparant un ou plusieurs médicament contenant le même PA de vérifier s'ils sont bioéquivalents, c'est-à-dire si leur biodisponibilité est identique, ce qui les rend inchangeable pour le malade.
- La biodisponibilité absolue est déterminée lors de l'étude d'un nouveau médicament.
- La détermination de la biodisponibilité relative est utilisée pour comparer des formes galéniques ; elle est obligatoire pour tout changement de formulation (changement d'excipient...) et avant commercialisation d'un médicament.
- Il ne faut pas assimiler obligatoirement mauvaise biodisponibilité et faible efficacité. En effet, la mauvaise biodisponibilité peut provenir d'un captage hépatique au 1er passage. Il est possible que ce captage aboutisse à la transformation du médicament en métabolite pharmacologiquement actif. Dans ces conditions, malgré une faible biodisponibilité, le médicament administré par voie orale pourrait être aussi actif que par voie intraveineuse. [6]

6) Pourquoi recherche-t-on des biodisponibilités élevées ?

On recherche des biodisponibilités élevées (proche de 100%) pour quatre raisons majeures :

- ne pas "gaspiller" de la substance active c'est-à-dire en recherchant la sélectivité de la cinétique du médicament pour sa biophase,
- pour garantir une exposition systémique inter-sujet reproductible,
- pour ne pas avoir à augmenter inutilement les doses pour exposer correctement les sujets ayant les biodisponibilités les plus faibles
- pour ne pas offrir des opportunités d'interactions conduisant à des effets secondaires par des augmentations inattendues de la biodisponibilité.

I.2.3. Distribution

La distribution concerne la répartition du médicament dans l'organisme par l'intermédiaire de la circulation sanguine. Le médicament se dissout dans le plasma (partie liquide du sang), ou bien il se fixe sur les protéines du sang qui lui servent de transporteurs jusqu'à sa cible pharmacologique.

En même temps qu'il agit sur sa cible, le médicament est capté par des organes qui le transforment (ou le métabolisent par les professionnels). [9]

Les médicaments ont une taille moléculaire faible, ils peuvent quitter facilement la circulation sanguine par filtration capillaire pour atteindre les tissus cibles. Certaines protéines plasmatiques comme l'albumine vont pouvoir fixer certains médicaments sans pouvoir empêcher leur filtration.

Le volume de distribution d'un médicament est défini en termes de concentration totale du médicament sous sa forme libre qui va dépendre de son pKa. Les acides se concentrent dans le sang alors que les médicaments basiques se concentrent dans l'estomac. Les médicaments chargés électriquement ne pourront pas passer la barrière hémato-encéphalique sauf si les méninges sont inflammatoires. [10]

I.2.4. Métabolisme

Le métabolisme est la phase des transformations chimiques des principes actifs du médicament. Cette métabolisation s'effectue principalement au niveau hépatique mais aussi au niveau des poumons, de l'intestin, ou d'autres organes (tube digestif, reins…) ou tissus (peau…). Le foie, en raison de sa vascularisation et de sa richesse en enzymes, joue un rôle primordial. Il est le grand nettoyeur ou (métaboliseurs) de médicaments. Cette biotransformation forme des résidus appelés des métabolites qui peuvent être actifs, inactifs voire toxiques. [9]

La plupart des médicaments sont métabolisés avant d'être rejeté de l'organisme. Ces réactions métaboliques se font en deux phases :

- **Phase I** : oxydation, réduction et hydrolyse pour augmenter la polarité des molécules et faciliter l'action des enzymes de la phase II.
- **Phase II** : réactions de conjugaison et de synthèse qui vont ajouter un groupement chimique à la molécule, ce qui va augmenter la solubilité aqueuse de cette molécule et favorisera son élimination.

L'expression des enzymes est différente selon les tissus, et l'ingestion simultanée de deux ou plusieurs médicaments va pouvoir affecter la vitesse du métabolisme de l'un ou de plusieurs des médicaments. [10]

I.2.5. Elimination

L'élimination est la dernière phase. Elle correspond à l'excrétion des médicaments et de leurs métabolites. Les principales voies d'élimination sont la voie urinaire, la voie biliaire (les métabolites partent dans les selles) et la voie pulmonaire (les métabolites partent dans l'air expiré). [9]

1) Excrétion rénale

Le pH urinaire va déterminer la fraction ionisée des molécules non-réabsorbées au niveau des tubules. Une augmentation du débit urinaire va augmenter l'élimination rénale car le temps de contact entre les molécules et la surface du tissu va être davantage limité.

2) Excrétion gastro-intestinale

Si on augmente le péristaltisme intestinal, on va augmenter l'excrétion du médicament. Si le médicament est toxique, on peut faire un gavage au polyéthylène glycol pour accélérer la vidange gastrique. [10]

I.3. La pharmacodynamique

La pharmacodynamique, ou pharmacodynamie, décrit les effets qu'un principe actif produit sur l'organisme : c'est l'étude détaillée de l'interaction entre récepteur et substance active.
Cette réponse est une composante de l'effet thérapeutique recherché.

Lors de cette étape, la substance active quitte le système sanguin pour diffuser jusqu'au site d'action dans l'organe cible et se combine avec un récepteur, une enzyme ou une structure cellulaire quelconque pour provoquer la réponse. [11]

I.4. Conclusion

La biodisponibilité est une propriété fondamentale d'une formulation pharmaceutique qui va dépendre de facteurs propres à la substance pour une voie d'administration donnée (et donc non modifiables) et de facteurs pharmaceutiques sous contrôle des galénistes. La biodisponibilité n'est pas un paramètre pharmacocinétique sensu stricto mais une variable car elle ne dépend pas que de facteurs biologiques. Elle est la seule variable pharmacocinétique que l'on puisse modifier et tenter d'améliorer. [3]

II.1. Introduction

La modélisation moléculaire peut être considérée comme un ensemble de techniques informatiques basées sur des méthodes de chimie théorique et les données expérimentales qui peuvent être utilisés soit pour analyser les molécules et les systèmes moléculaires ou à prédire les propriétés moléculaires, chimiques et biochimiques [12]. Il sert de pont entre la théorie et expérience pour:

- Extrait résultats pour un modèle particulier.
- Comparer les résultats expérimentaux du système.
- Comparer les prédictions théoriques du modèle.
- Aide comprendre et interpréter les observations expérimentales.
- Corrélation entre détails microscopiques au niveau atomique et moléculaire et les propriétés macroscopiques.
- Fournir de l'information non disponible à partir d'expériences réelles. [13]

La modélisation moléculaire permet de justifier les données expérimentales disponibles et prédire les activités pour des nouveaux composés ou des composés pour lesquels les données expérimentales ne sont pas disponibles. Parmi ces méthodes de modélisation les plus utilisées, on peut citer la méthode QSAR (Quantitative Structure-Activity Relationship). Cette méthode s'appuie sur le principe que les activités biologiques des molécules dépendent fortement de leurs structures chimiques.

L'élaboration d'un modèle QSAR n'est pas une chose facile. La première difficulté réside dans la différence d'échelles existant entre les données à corréler. La structure étant à une échelle moléculaire alors que les activités à prédire sont à une échelle macroscopique. Un des problèmes importants réside également dans le traitement de données. En fait, de nombreux outils existent et il s'agit de trouver le moyen le plus adapté pour obtenir un modèle fiable à partir des données disponibles.

L'analyse de la régression est l'outil le plus utilisé en modélisation QSAR. L'idée est de décrire et d'évaluer la relation entre une variable (dite variable à expliquer ou variable dépendante) souvent notée Y, et une (ou plusieurs) variable(s), dite(s) variable(s) explicative(s) ou indépendante (s) notée (s) X. Dans ce chapitre, nous présentons la régression linéaire multiple (MLR) utilisées pour élaborer les modèles QSAR. [14]

II.2. La méthode QSAR : relation quantitative structure - activité (en anglais : *Quantitative structure-activity relationship*)

Les méthodes QSAR sont de plus en plus utilisées, du fait de la croissance des moyens de calculs (machines, logiciels,...). Récemment, on assiste à la mise en place d'un nouveau règlement REACH (*Registration, Evaluation, Authorisation and Restriction of Chemicals)* qui recommande l'utilisation des méthodes alternatives pour limiter le recours à l'expérimentation. On entend par étude de la relation quantitative structure-activité l'analyse de la structure moléculaire pour en tirer des informations prédictives sur ses propriétés essentielles telles que sa stabilité, sa distribution, sa captation, son absorption et sa toxicité. L'étude de la relation quantitative structure-activité est l'une des méthodes d'identification très importantes.

II.2.1.Historique

En fait, les premiers développements dans le sens de telles méthodologies sont plutôt anciens. Dès 1868, Crum-Brown et Fraser ont postulé l'existence de relations entre les activités physiologiques et les structures chimiques [15] en reliant les changements d'activité biologique à des modifications structurales simples, ne disposant alors pas de moyen pour caractériser les structures chimiques en termes quantitatifs.

Une avancée importante vers les modèles QSAR proprement dits a été réalisée grâce au développement des équations de Hammett [16-17] dans lesquelles les constantes du même nom σ caractérisent de manière quantitative les vitesses de réactions pour les composés organiques.

$$\mathrm{Log}(\,K/K_0) = \sigma\,\rho$$

Où K et K_0 sont les constantes respectives de la réaction étudiée et de celle d'une référence et ρ une constante de réaction dépendant du type de réaction.

Cela dit, les premiers travaux utilisant la méthodologie telle qu'employée à l'heure actuelle sont à attribuer aux contributions de Hansch [18] d'une part et de Free et Wilson [19] d'autre part.

D'un côté, Hansch a proposé des modèles reliant directement l'activité biologique des composés avec les propriétés hydrophobes, électroniques et stériques à l'échelle moléculaire. De l'autre, Free et Wilson ont développé des modèles dits de contributions de groupes pour

l'activité biologique dans lesquels l'identification des groupements particuliers renvoie à une valeur tabulée (empirique). La valeur finale est donc une combinaison de ces différentes valeurs correspondant aux différents groupes identifiés au sein de la molécule.

Dans les dernières années, l'utilisation des méthodes QSAR n'a cessé de progresser.
Elle est même devenue indispensable en chimie pharmaceutique et pour la conception de médicaments [20-21]. Leur développement dans une gamme plus large d'applications leur ouvre d'ailleurs de grandes perspectives (ex : points d'ébullition, températures critiques, densité) [22].

Il s'agit de présenter ici le principe des modèles QSAR ainsi que ceux des différents outils employés pour leur mise en place et leur évaluation : bases de données expérimentales, descripteurs, outils d'analyse de données.

II.2.2. Principe

Le principe des méthodes QSAR (Quantitative Structure Activity Relationship) consiste comme leurs noms l'indiquent à mettre en place une relation mathématique à l'aide de méthodes d'analyse de données, reliant des propriétés moléculaires microscopiques appelées descripteurs, à un effet expérimental (biodisponibilité des médicaments dans l'organisme, activité biologique, toxicité), pour une série de composés chimiques similaires. Le point de départ de telles méthodes se construit sur la définition des descripteurs moléculaires empiriques ou théoriques. Ces dernières prennent en compte des informations sur la structure et les caractéristiques physico-chimiques des molécules. La forme générale d'un tel modèle est la suivante :

$$\textbf{Activité} = f(D_1, D_2, \ldots D_n, \ldots)$$

Avec :

D_1, D_2, …D_n sont des descripteurs des structures moléculaires.

Le point de départ de telles méthodes se construit sur la définition des descripteurs moléculaires empiriques ou théoriques. Ces dernières prennent en compte des informations sur la structure et les caractéristiques physico-chimiques des molécules.

Une fois cette relation est établie et validée, elle peut alors être employée pour la prédiction de l'activité des nouvelles molécules, pour lesquelles les valeurs expérimentales ne sont pas disponibles.

Le choix de la base des données expérimentales de référence est décisif dans une étude QSAR. Elle doit être composée de données expérimentales fiables obtenues en suivant un protocole expérimental unique. En effet, la robustesse du modèle dépend fortement de la base sur lequel il se fonde.

Enfin, le lien entre les descripteurs et la base de données est déterminé grâce à des outils d'analyse comme les régressions multi-linéaires (MLR), les régressions aux moindres carrés partiels (PLS), les arbres de décisions, les réseaux de neurones, et les algorithmes génétiques.

Les informations extraites à partir des résultats d'étude de QSAR peuvent être utilisées pour obtenir une meilleure connaissance des structures moléculaires et probablement le mode d'action au niveau moléculaire. Ces informations peuvent alors être utilisées pour prévoir les activités biologiques de nouveaux composés ainsi que pour concevoir de nouvelles structures (figure II.1) [23].

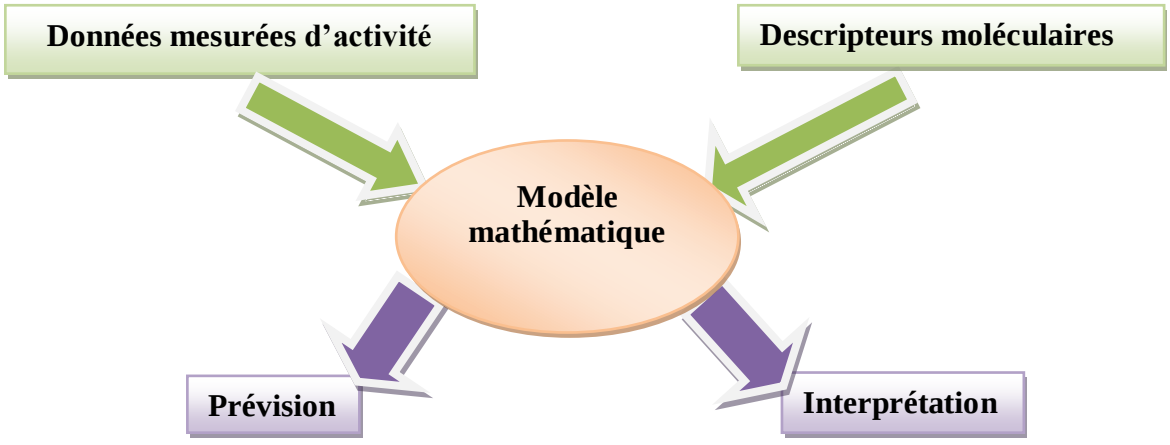


Figure II.1 Modèle d'étude quantitative des relations structure/activité (QSAR)

II.2.3. Importance des bases de données

Un modèle QSAR, de par sa construction, est très dépendant des données expérimentales de référence. Le choix de la base de données est donc un point critique des on développement. Dans la plupart des cas, les données expérimentales sont issues de la littérature.

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

De plus, les données doivent être obtenues suivant un protocole expérimental unique.
En effet, les conditions expérimentales ont, en général, une forte influence sur les valeurs obtenues.

Il faut d'ailleurs également que cette distribution soit la plus homogène (et normale) possible et ce pour deux raisons. La première vient du fait que la plupart des méthodes statistiques sont basées sur des distributions normales et sont donc plus performantes sur ce type de distributions. Ensuite, la présence de valeurs atypiques influence beaucoup les modèles et fausse les résultats. Sur la Figure II.2, par exemple, la présence du composé extrême A donne l'impression d'une corrélation. De tels composés sont donc à éliminer. Sur l'illustration proposée, l'exclusion de la valeur atypique réduit la corrélation à R^2=0,04.

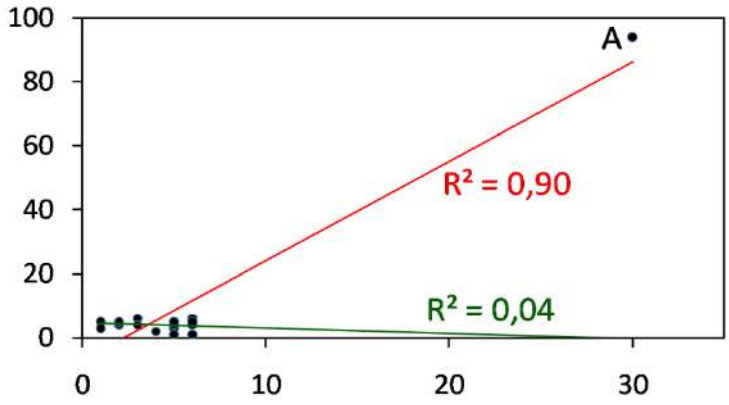


Figure II.2 Influence d'une valeur atypique (A) sur une corrélation

L'obtention d'un modèle efficace sera également dépendante du type de molécules qui y sont incluses. Un jeu convenable devrait contenir des composés pour lesquels le mécanisme mis en jeu est unique. De plus, plus le modèle présentera des molécules des structures proches, plus ce dernier aura de chances d'être performant au prix d'une réduction de son domaine d'applicabilité. Enfin, la présence de molécules au comportement déviant au sein de la base de données peut influencer de manière importante le modèle final. De même, la mise en évidence de sous-groupes de molécules dans un modèle a tendance à le dégrader. Il est préférable, dans de tels cas, de considérer des modèles multiples permettant de traiter les différents sous-groupes de manière indépendante. [24]

II.2.4. Descripteurs moléculaires

Les descripteurs moléculaires sont des propriétés numériques qui peuvent être calculées à partir de la table de connectivité d'une molécule ou d'une représentation planaire (2D) ou tridimensionnelle (3D) de la structure. Ils sont basés sur les éléments présents, les charges partielles, la nature des liaisons. On peut calculer ainsi:

- des propriétés physiques : polarisabilité, charge totale, réfractivité moléculaire, masse, densité, coefficient de partage eau/octanol (log P) …
- des approximations d'aires de surfaces : surface de Van der Waals, surface accessible au solvant…
- le dénombrement des atomes, liaisons et pivots ;
- des indices de formes et de connectivité : indices de Kier et Hall ;
- des descripteurs topologiques : matrices de distance et de connectivité ;
- les dénombrements de fonctions pharmacophoriques : donneurs, accepteurs de liaison hydrogène, polaire (donneur et accepteur), positif, négatif, hydrophobe et autres … ;
- les descripteurs de charges partielles : charge partielle positive (négative) totale, aire de la surface de Van der Waals positive (négative), aire de la surface de Van der Waals polaire, … ;
- les indices de Hansch : r, s et p ; … [13]

De nombreuses recherches ont été menées, au cours de ces dernières décennies, pour trouver la meilleure façon de représenter l'information contenue dans la structure des molécules et ces structures elles-mêmes en un ensemble de nombres réels appelés descripteurs qui sont soit théoriques ou empiriques :

II.2.4.1.Les descripteurs moléculaires théoriques

- **Les descripteurs 1D** : sont accessibles à partir de la formule brute de la molécule et décrivent des propriétés globales du composé comme le nombre d'atomes et la masse moléculaire,…etc.

  Ces descripteurs sont couramment utilisés du fait de leur extrême simplicité. Cependant, ils peuvent poser problème pour une bonne interprétation des mécanismes d'interaction du fait qu'ils ne permettent pas de tenir en compte des effets stériques et d'isomérie. [14]

- **Les descripteurs 2D** : sont calculés à partir de la formule développée de la molécule. On distingue :

a. **Descripteurs constitutionnels** : La manière la plus simple de représenter un système moléculaire est de considérer sa composition chimique, sans prise en compte de quelconques considérations géométriques ou électroniques. Des exemples caractéristiques de descripteurs constitutionnels sont :

   - les nombres absolus et relatifs d'atomes (C, H, N…) ;
   - les nombres absolus et relatifs de groupes fonctionnels (NO2, COOH, OH…) ;
   - les nombres absolus et relatifs de liaisons (simples, doubles, aromatiques…) ;
   - les nombres absolus et relatifs de cycles (aromatiques ou non) ;
   - la masse moléculaire.

Ces descripteurs sont très couramment utilisés du 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 [25], dans laquelle l'effet des substituants est introduit sous la forme d'une valeur tabulée.

Si ce type de descripteurs est employé pour obtenir les modèles QSAR les plus simples, ces derniers peuvent poser problème pour 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. Les

paramètres constitutionnels sont, dans de tels cas, défaillants. Il faut alors recourir à d'autres classes de descripteurs.

b. **Descripteurs topologiques** : Les descripteurs topologiques sont des indices obtenus à partir d'une structure 2D de la molécule, à savoir une simple table de connectivité des atomes dans la molécule. Ils contiennent en leur sein des informations sur la taille globale du système, sa forme globale et ses ramifications [26].

Ces descripteurs, ou indices topologiques, s'inspirent de la théorie des graphes appliquée à la table de connectivité qui n'est autre qu'une représentation compacte de la connectivité interatomique au sein de la molécule. [24]

Exemples:

Indice de Wiener qui permet de caractériser le volume moléculaire et le taux de ramification des molécules. Il est calculé comme la somme de toutes les distances topologiques entre atomes au sein de la molécule. [27]

Indice de Randic qui peut être interprété comme une mesure de l'aire de la molécule accessible au solvant. [28]

Parmi les autres indices topologiques courants, on retrouve les indices de Kier-Hall [29] ou encore l'indice de Balaban. [30]

En général, ce type de descripteurs simplifie grandement la représentation de la connectivité chimique au sein de la molécule puisqu'ils ne prennent pas en compte les différences de distances, d'angles et d'ordres de liaison ni même la nature des atomes dans la molécule. Si certains indices ont été développés pour intégrer de manière très approximative ce genre d'informations, ils restent souvent insuffisants pour caractériser l'intégralité des propriétés moléculaires. Finalement, les indices topologiques sont souvent considérés comme des descripteurs convenables d'un point de vue numérique. Cela dit, dans la plupart des cas, l'interprétation des équations QSAR qui en résultent n'est pas aisée, puisqu'il est difficile de les relier aux mécanismes sous-jacents. [24]

Ces descripteurs 2D permettent de prédire les propriétés physiques mais sont insuffisantes pour expliquer certaines propriétés et activités biologiques comme la toxicité.

- **Les descripteurs 3D :** sont évalués à partir des positions relatives des atomes dans l'espace, et décrivent des caractéristiques plus complexes. Leurs calculs nécessitent donc de connaître la géométrie 3D de la molécule.

a. **Les descripteurs 3D géométriques :** Il s'agit de descripteurs évalués à partir des positions relatives des atomes d'une molécule dans l'espace, ainsi que des rayons et masses atomiques. Ils nécessitent donc de connaître la structure 3D de la molécule. Celle-ci peut être obtenue expérimentalement bien entendu mais le plus souvent par modélisation moléculaire, empirique ou *ab initio*. Par conséquent, contrairement aux descripteurs précédents, ils nécessitent un certain temps de calcul.

Parmi ces descripteurs, on retrouve le volume et la surface moléculaire, le moment d'inertie ou encore des distances, angles ou angles dièdres particuliers entre atomes dans la molécule.

b. **Les descripteurs 3D électroniques** : permettent de quantifier les différents types d'interactions inter- et intramoléculaires, de grande influence sur l'activité biologique des molécules. Le calcul de la plupart de ces descripteurs nécessite la recherche de la géométrie pour laquelle l'énergie est minimale, et fait souvent appel à la chimie quantique (**ELUMO :** Energie de l'orbitale la plus basse vacante, **EHOMO :** Energie de l'orbitale la plus haute occupée, **μ** : Potentiel chimique électronique).

c. **Les descripteurs 3D spectroscopiques** : les molécules peuvent être caractérisées par des mesures spectroscopiques, par exemples par leurs fonctions d'onde vibrationnelles.

II.2.4.2. Les descripteurs moléculaires empiriques

Il existe plusieurs descripteurs qui peuvent être mesuré expérimentalement tels que le coefficient de partage (octanol-eau) (log P), la polarisabilité (α).

Exemple : **Coefficient de partage (octanol-water) Kow, log P**

Le coefficient de partition est défini comme le rapport de la concentration du soluté dans la phase huileuse C' à la concentration du soluté non ionisé dans la phase aqueuse C, à l'équilibre.

$C' > C \Rightarrow P > 1 \Rightarrow \log P > 0$: le soluté est dit lipophile (hydrophobe).

$C' < C \Rightarrow P < 1 \Rightarrow \log P < 0$: le soluté est dit hydrophile. [14]

II.2.5. Sélection et application

II.2.5.1. Carte d'identité d'une molécule

Afin d'établir des modèles de prédiction du comportement biopharmaceutique des molécules (absorption, métabolisme, élimination, biodisponibilité,…etc.).Des outils de chémo-informatique permettent une meilleure manipulation de l'ensemble des données bien collectés.

Cette modélisation repose sur l'utilisation des descripteurs moléculaires (de topologie, de géométrie,…), véritable carte d'identité d'une molécule.

Les descripteurs 1D sont déduits de la formule brute (nombre d'atomes, nombres de C, O, N…), les descripteurs 2D de la structure plane de la molécule (donneurs/accepteurs d'hydrogène, nombre de cycle…). Les descripteurs 3D sont déduits de la structure 3D de la molécule (volume, surface accessible au solvant…). La notion de 4D peut être employée quand on explore l'espace autours de la molécule (surfaces moléculaires).

II.2.5.2. Utilisation des descripteurs moléculaire

Ces descripteurs sont utilisés pour filtrer des bases de structures chimiques permettent d'éliminer des composés dont les caractéristiques sont trop éloignées de celles d'un médicament.

Les règles dites de Lipinsky utilisent, par exemple, une combinaison de propriétés physicochimiques essentielles dans l'absorption orale. Et on peut également vouloir s'afficher de certains fragments chimiques comme :

- les fonctions réactives susceptibles de former des liaisons covalentes [31] ;
- les fonctions connues comme générant des faux positifs pour des raisons physico-chimiques du type agents chélatants, inhibiteurs, suicides, composés polyanioniques, etc. [32];
- les fonctions chimiques susceptibles de réaliser des agrégats (promiscuous aggregating inhibitors) et faussant le résultat biologique [33].

Ces approches sont souvent utilisées dans des processus de focalisation de base. Elles permettent, dans le travail de Monge [34] de réduire une base chimique de 173 803 composés à 10 000 molécules ayant des propriétés compatibles avec celles d'un futur médicament.

Mais elles ont également été utilisées avec succès pour enrichir des chémiothèques avec des apports externes tant sur la base du potentiel médicamenteux que d'un point de vue diversité chimique.

II.2.5.3. Sélection à l'aide du modèle QSAR

Ensuite on s'engage la construction du modèle, par des approches QSAR (quantitative structure activity relationship) le plus souvent [35]. Ces techniques permettent de relier par une relation mathématique les descripteurs chimiques à l'activité biologique dans une approche de relation structure-activité quantitative. De tels modèles QSAR, basés sur des descripteurs 2D ou 3D sont le point de départ de nombreux processus de sélection de molécules. Ces modèles sont tous construits sur le même mode mettant en œuvre un jeu de référence appelé jeux d'apprentissage permettant de sélectionner le(s) descripteur(s) le(s) plus adapté(s) dans la construction d'un modèle. Les modes de construction de ces modèles peuvent être simples (régression linéaire ou multilinéaire) ou plus sophistiqués permettent de rechercher la meilleure combinaison de descripteurs (algorithmes génétique, réseaux de neurones, forets aléatoires, séparateurs à vaste marge).

Elle sera suivie d'une étape de validation appelée validation externe et réalisée avec un nouveau jeu de données, souvent généré postérieurement, et qui va permettre de bien mesurer la pertinence du modèle et de définir le domaine d'application.

Le pouvoir prédictif de ces modèles et leur domaine d'application pour de futures séries chimiques dépendent de la composition du jeu d'apprentissage utilisé pour construire le modèle. Il faut se rappeler qu'un modèle QSAR parfait ne pourra donc pas permettre une prédiction pour toute sorte de substances (molécules trop différents de celles du jeu d'apprentissage) et avoir un niveau de précision supérieur à celui de la méthode expérimentale qui a permis de générer les données du jeu d'apprentissage. [36]

II.2.6. Méthodes d'analyse de données

L'objectif de l'analyse statistique est justement de « démêler » ces descripteurs et d'identifier ceux qui sont corrélés à la variable cible, qui produisent du signal, de ceux qui ne le sont pas, qui produisent du bruit. L'analyse statistique permet également d'identifier les descripteurs qui sont corrélés entre eux pour ne garder que les principaux et réduire ainsi la redondance d'informations [37].

Dans la terminologie des statistiques, la propriété étudiée est appelée variable cible ou variable dépendante ou encore variable Y. C'est la caractéristique que l'on cherche à expliquer et à optimiser.

Les descripteurs, en l'occurrence chimiques, sont appelés variables explicatives, variables indépendantes ou encore variables X.

L'analyse statistique détermine et quantifie les corrélations entre les descripteurs et la variable cible. Elle indique également la contribution relative de chaque descripteur dans l'explication globale de l'activité. Le modèle statistique est une équation donnant la valeur de la variable cible en fonction de la somme des valeurs pondérées des descripteurs.

Les principaux outils statistiques pour obtenir un modèle sont :

- la régression linéaire multivariée (ou multiple) (*Multivariate Linear Regression* - MLR) ;
- la régression en composante principale (*Principale Componant Regression* - PCR) ;
- la régression des moindre carrés partiels (*Partial Least Squares* - PLS);
- les réseaux de neurones artificiels (RNA).

Le choix de la méthode dépend principalement de la question qui est posée et de la nature des données à traiter.

La MLR est d'un usage assez restrictif et peu adaptée au QSAR [38] : elle requiert un jeu de données très complet où toutes les combinaisons de substituants ont été testées. En d'autres termes, il faut faire autant d'expériences que de variations possibles, ce qui est rarement le cas en pratique.

On s'intéresse uniquement à la régression linéaire multiple MLR.

II.2.6.1. Régression linéaire multiple (MLR)

Nous présentons la régression linéaire ceci dit la régression linéaire simple (SLR) et la régression linéaire multiple (MLR) utilisées pour élaborer les modèles QSAR.

En statistiques, en économétrie et en apprentissage automatique, un modèle de régression linéaire est un modèle de régression d'une variable expliquée sur une ou plusieurs variables explicatives dans lequel on fait l'hypothèse que la fonction qui relie les variables explicatives à la variable expliquée est linéaire dans ses paramètres.

Régression linéaire simple

Définition

Un modèle de régression linéaire simple est de la forme :

$$\mathbf{Y = \beta 0 + \beta 1 X + \varepsilon}$$

Avec :

- Y est la variable dépendante ;
- β_0 et β_1 sont les coefficients (ordonnée à l'origine et pente) ;
- X est la variable indépendante (variable explicative) ;
- ε est une erreur aléatoire.

Régression linéaire multiple

La régression linéaire multiple cherche à approximer une relation trop complexe en général, par une fonction mathématique simple.

L'idée est de décrire et d'évaluer la relation entre une variable (dite variable à expliquer ou variable dépendante) souvent notée Y (ici, la biodisponibilité des médicaments dans l'organisme), et une (ou plusieurs) variable(s), dite(s) variable(s)explicative(s) ou indépendante (s) notée (s) X (ici, les descripteurs).). L'objectif est d'obtenir une équation de la forme suivante :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \varepsilon$$

Cette équation est linéaire par rapport aux paramètres (coefficients de régression) β_0 , β_1...,β_p. ε est une erreur.

1) Description de la méthode

La régression multilinéaire est la méthode la plus simple et la plus employée pour le développement de modèles prédictifs [39]. Elle repose sur l'hypothèse qu'il existe une relation linéaire entre une variable dépendante Y et une série de p variables indépendantes Xi L'objectif est d'obtenir une équation de la forme suivante :

$$\mathbf{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_p X_p + \varepsilon}$$

Où : $\boldsymbol{\beta_i}$ sont les coefficients de la régression.

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

$$Y_1 = \beta_0 + \beta_1 X_{1,1} + \beta_2 X_{2,1} + \ldots + \beta_p X_{n,1} + \varepsilon_1$$
$$Y_2 = \beta_0 + \beta_1 X_{1,2} + \beta_2 X_{2,2} + \ldots + \beta_p X_{n,2} + \varepsilon_2$$
$$\vdots$$
$$Y_n = \beta_0 + \beta_1 X_{1,p} + \beta_2 X_{2,p} + \ldots + \beta p X_{n,p} + \varepsilon_n$$

Où les résidus $\boldsymbol{\varepsilon_i}$ représentent l'erreur du modèle, constituée par l'incertitude sur la variable dépendante $\mathbf{Y_i}$ d'une part, sur les variables indépendantes $\mathbf{X_i}$ d'autre part, mais aussi par les informations contenues dans les variables indépendantes mais non exprimées via les variables dépendantes.

Ce système d'équations peut être écrit sous la forme matricielle suivante :

$$\begin{Bmatrix} Y1 \\ Y2 \\ . \\ . \\ . \\ . \\ . \\ Yn \end{Bmatrix} = \begin{Bmatrix} 1 & X1,1 & X2,2 & . & . & . & . & X1,p \\ 1 & . & . & & & & & \\ . & . & . & & & & & \\ . & . & . & & & & & \\ . & . & . & & & & & \\ . & . & . & & & & & \\ . & . & . & & & & & \\ . & . & . & & & & & \\ 1 & Xn,1 & Xn,2 & . & . & . & . & Xn,p \end{Bmatrix} \begin{Bmatrix} \beta 1 \\ \beta 2 \\ . \\ . \\ . \\ . \\ . \\ \beta n \end{Bmatrix} + \begin{Bmatrix} \varepsilon 1 \\ \varepsilon 2 \\ . \\ . \\ . \\ . \\ . \\ \varepsilon n \end{Bmatrix}$$

La matrice X *(n,p)* comporte dans sa première colonne un vecteur constitue uniquement de 1.Ce vecteur correspond à la constante $\mathbf{X_0}$. La matrice X est alors de dimension *(n, p+1)*.
Le modèle s'écrit alors sous forme $\mathbf{Y= X\ b + \varepsilon}$.

La méthode consiste alors à choisir les coefficients du vecteur **β** en faisant en sorte de minimiser la somme des carrés des écarts entre les valeurs prédites et les valeurs réelles sur l'intégralité de la base de données et ceci sous couvert de certaines hypothèses de départ.

En premier lieu, les variables indépendantes $\mathbf{X_i}$, comme leur nom l'indique, sont supposées indépendantes entre elles et leur incertitude est négligeable. Ensuite, les différents échantillons $\mathbf{Y_i}$ sont supposés indépendants entre eux. Enfin, par nature, la dépendance de **Y** vis-à-vis des $\mathbf{X_i}$ est supposée linéaire.

La valeur prédite de la variable dépendante est alors :

$$\widehat{\boldsymbol{Y}}=\boldsymbol{\beta_0}+\widehat{\boldsymbol{\beta}}_1\mathbf{X_{1,i}}+ \ldots +\widehat{\boldsymbol{\beta}}_p\mathbf{X_{n,i}}$$

Les résidus peuvent donc être définis comme la différence entre les valeurs prédites et observées de Y.

$$\boldsymbol{\varepsilon}=\mathbf{Y_i}-\widehat{\boldsymbol{Y}}_\mathbf{i}$$

Il s'agit alors de trouver les coefficients $\widehat{\boldsymbol{\beta}}_\mathbf{i}$ afin de minimiser la somme des carrés de ces résidus (*RSS*) pour l'intégralité de la base de données.

2) Confiance accordée aux résultats

La logique statistique nous conduit à effectuer des tests d'hypothèses, et à déterminer des intervalles de confiance autour des valeurs estimées. Successivement on va chercher a:

- tester la signification globale de la régression,
- tester l'hypothèse nulle $\beta_1 = 0$ et à calculer l'intervalle de confiance de β_1,
- tester l'hypothèse nulle $\beta_0 = 0$ et à calculer l'intervalle de confiance de β_0,
- calculer la précision de l'estimation de Y pour la moyenne et pour une observation individuelle. [14]

3) Qualité de la régression linéaire

Les objectifs d'une modélisation statistique peuvent être de natures différentes, les objectifs explicatifs et les objectifs prédictifs. Ces objectifs sont déterminés par plusieurs coefficients.

a. Coefficient de corrélation (R)

C'est le coefficient de corrélation de Bravais Pearson entre $\widehat{Y}_i$ et Y, c'est-à-dire entre valeurs observées et prédites par le modèle de régression, il est noté R, sa valeur variée entre 0 et 1.

b. Coefficient de détermination (R^2)

Pour évaluer la précision avec laquelle la dépendance trouvée décrit la variance de la variable dépendante (c'est à dire la qualité de l'ajustement statistique), le carré du coefficient de détermination ($\mathbf{R^2}$) est utilisé.

Définition :

Le coefficient de détermination (R^2) est un indicateur qui permet de juger la qualité d'une régression linéaire, simple ou multiple. D'une valeur comprise entre 0 et 1, il mesure l'adéquation entre le modèle et les données observées. Certes, le R^2 a ses imperfections, mais son utilité n'a d'égale que sa simplicité.

Deux façons de le définir (entre autres) :

Dans le cadre d'une régression linéaire simple, c'est le carré du coefficient de corrélation.

La deuxième façon est beaucoup plus riche en implications car elle s'applique aussi bien à une régression simple qu'à une régression multiple.

On sait que la valeur *yi* d'une observation peut être décomposée en deux parties : une part expliquée par le modèle et une part résiduelle. La dispersion de l'ensemble des observations se décompose donc en variance expliquée par la régression et en variance résiduelle (MSE), inexpliquée. Le R^2 se définit alors comme la part de variance expliquée par rapport à la variance totale, autrement dit de [(1 – somme des carrés des résidus) / somme des carrés totale] la somme des carrés totale étant la somme des carrés des distances entre les points du nuage et une droite horizontale qui passerait par son centre de gravité.

On définit alors le coefficient de détermination, qui mesure la part de la variance expliquée par le modèle dans la variance totale.

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

Avec :

Somme des carrés totaux : *TSS*

Somme des carrés expliqués : *ESS*

Somme des carrée résiduels : *RSS*

Ainsi, *TSS* = *ESS* + *RSS*

Il est compris entre 0 et 1. Plus il est proche de 1 et plus la régression permet d'expliquer une grande partie de la variance totale de la variable à expliquer.

Un bon ajustement correspondra à un R^2 proche de l'unité.

R^2 qui varie entre 0 et 1, mesure la proportion de variation totale de Y autour de la moyenne expliquée par la régression. Plus la valeur de R^2 sera proche de 1 (cas idéal) et plus

les valeurs prédites et observées sont corrélées. Un R^2 faible signifie que le modèle a un faible pouvoir explicatif et les descripteurs (certains d'eux) sont sans effet sur la réponse.

Le jugement sur la valeur de R^2 est très subjectif. Bien que ce coefficient soit très facile à comprendre, il faut se garder d'y attacher trop d'importance car il est loin de fournir un critère suffisant pour juger de la qualité d'une régression. Il n'est pas recommandé d'utiliser R^2 pour comparer des modèles avec un nombre différent de descripteurs, le coefficient R^2 nous dira toujours de choisir le modèle avec le plus grand nombre de descripteurs car son R^2 sera plus important (on projette sur un espace plus grand), même si les variables sont sans effets sur la réponse Y.

La valeur de R^2 dépend de la taille de l'échantillon et le nombre de variables prédictives dans l'équation. Il garde la même valeur ou augmente lors d'une nouvelle variable de prédiction est ajoutée à l'équation de régression, même si la variable ajoutée ne contribue pas à la réduction de la variance inexpliquée. Par conséquent, un autre paramètre statistique peut être utilisé, appelé R^2ajusté.

c. Coefficient de détermination ajusté (R^2ajusté)

Le principal défaut du R^2 est de croître avec le nombre de variables explicatives. Or, on sait qu'un excès de variables produit des modèles peu robustes. C'est pourquoi on s'intéresse davantage à cet indicateur qu'au R^2. Mais ce n'est pas un véritable « carré » et il peut même être négatif.

R^2ajusté est obtenu par une expression semblable à celle de R^2, mais la SSR et SST sont divisés par leurs degrés de liberté correspondants. Ce coefficient est utilisé en régression multiple par ce qu'il tient compte du nombre de paramètres du modèle. Voici deux présentations du R^2 ajusté, sachant que certains auteurs donnent une définition légèrement différente :

$$\mathbf{R^2ajusté} = \mathbf{R^2} - \frac{k(1-R^2)}{n-k-1} = 1 - \frac{(1-R^2)(n-1)}{n-k-1}$$

- n = La taille de l'échantillon
- k = Le nombre de variables indépendantes dans l'équation de régression

Cette formule indique notamment que R^2ajusté est toujours inferieur à R^2, et ceci d'autant plus que le modèle contient un grand nombre de descripteurs.

d. Ecart type (s)

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

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

Dont p est le nombre de variables indépendants.

e. Coefficient de prédiction (Q^2)

Si R^2 est une mesure de la justesse de la corrélation, Q^2 est une mesure de la justesse de la prédiction. Il mesure la capacité prédictive d'un modèle et donné par :

$$Q^2 = 1-\frac{PRESS}{SSY}$$

4) Colinéarité des variables explicatives

La détection d'une colinéarité se fait à l'aide de :

- Matrice de corrélation : elle permet de détecter des fortes corrélations entre deux variables.
- Facteur d'inflation de la variance VIF : On définit le facteur d'inflation de la variance (VIF) par

$$VIF = 1/ (1-R^2_j)$$

Où R^2_j est le coefficient de détermination de la régression de la variable Xj sur les autres variables. Plus Xj est linéairement proche des autres variables, plus R^2_j est proche de 1 et le VIF grand. L'avantage du VIF par rapport à la matrice de corrélation est qu'il prend en compte des corrélations multiples.

5) Validations des modèles

Quel que soit le but ultime du modèle QSAR, il doit être obligatoirement validé avant d'être interprété ou utilisé à fins prédictives. Il existe différentes solutions pour s'assurer de validité d'un modèle. La validation croisée *LOO* (*leave-one-out*) est un processus qui permet de tester la précision prédictive d'un modèle, cette méthode basée sur le calcul de quelques paramètres statistiques tels que : la somme des carrée résiduels RSS (PRESS), Somme des carrés totaux TSS, qualité d'ajustement R^2ajusté, coefficient de corrélation de validation croisée R^2_{CV}, la validation standard des erreurs de prédiction (SPRESS) et l'erreur de prédiction (PE). Ces paramètres statistiques sont calculés à partir des relations suivantes :

$PRESS = \sum (Y_{obs}-Y_{cal})$ La somme résiduelle prédictive des carrés (PRESS, *Predictive Residual Sum of Squares*) est une mesure des capacités prédictives d'un modèle.

$TSS = \sum (Y_{obs}-\bar{Y})^2$ La somme des carrés totaux

R^2ajusté$= (1-r^2)\left(\frac{n-1}{n-p-1}\right)$ Qualité d'ajustement

$R^2_{CV} = 1-\frac{PRESS}{TSS}$ coefficient de corrélation de validation croisée

$S_{PRESS} = \sqrt{\frac{PRESS}{n}}$ la validation standard des erreurs de prédiction

$PE = 0.6745\ (1-r^2)/\sqrt{n}$ l'erreur de prédiction [13]

II.2.7. Applications des méthodes QSAR

Les applications des méthodes QSAR sont très nombreuses, elles touchent tous les domaines ou la structure chimique intervient, entre autre on peut citer :

- ***<u>En chimie (Propriétés physico-chimiques</u>)***: point d'ébullition : Une des premières applications de la QSAR concerne la prédiction des points d'ébullition **[40].**Il est bien connu par exemple que pour une famille de composés chimiques, particulièrement en chimie organique, il existe une corrélation forte entre la structure et les propriétés observées. On peut citer comme exemple simple, la relation entre le nombre de carbones dans les alcanes et leur point d'ébullition. Il existe une tendance nette à l'augmentation de la température d'ébullition avec le nombre d'atomes de carbone, ce qui sert de moyen prédictif pour les points d'ébullition des alcanes les plus lourds. Les méthodes de Hammett, de Taft et prédiction de pKa sont des applications particulièrement intéressantes, ainsi le Point de fusion, densité, indice de réfraction,

température critique, viscosité, solubilité, pression de vapeur, tension superficielle, Coefficients de partition : eau/octanol, air/eau, huile/air, lait/plasma…

- ***En biologie*** : L'activité biologique des molécules est mesurée habituellement au moyen d'essais afin d'établir le niveau d'inhibition d'une transduction de signal ou d'une voie métabolique particulière. Les produits chimiques peuvent être biologiquement actifs par leur toxicité. La recherche de médicament implique parfois l'utilisation de la QSAR afin d'identifier les structures chimiques pouvant présenter de bons effets inhibiteurs sur des cibles spécifiques et possèdent une faible toxicité (activité non-spécifique). Un intérêt particulier est la prédiction du coefficient de partition log P, qui est une mesure importante pour l'identification de la «similarité médicamenteuse » selon la règle des cinq de Lipinski. Alors que de nombreuses analyses QSAR traitent des interactions d'une famille de molécules avec une enzyme ou un site récepteur, la QSAR peut aussi être utilisée pour étudier les interactions entre les domaines structuraux des protéines
- ***Autres propriétés/activités*** :

- Prédiction de la toxicité aquatique des composés chimiques vis-à-vis des espèces environnementales.

- Toxicité des nanoparticules
- Toxicité des pesticides et des colorants
- Propriétés inhibitrices de corrosion
- Concentration micellaire critique
- Prédiction de plusieurs propriétés dangereuses telle que l'explosibilité et l'inflammabilité de certaines familles de molécules chimiques.

- Conception des médicaments et de nombreux autres produits tels que les agents tensio-actifs, parfums, les colorants et les produits de la chimie fine… [41]

II.3. Conclusion

La modélisation moléculaire et l'analyse QSAR ont été effectuées pour déterminer quantitativement l'effet de la structure moléculaire des composés étudiés sur leur activité biologique. Des modèles mathématiques ont été développés pour prédire cette activité.

La validité des modèles obtenus ont été établie par la détermination des paramètres statistiques appropriés. Le développement d'un modèle QSAR exige un ensemble de diverses données, pour cet effet, un nombre considérable de descripteurs doivent être considérés dans l'étude QSAR. Les QSAR constituent un outil puissant d'information sur les caractéristiques probables d'une substance à condition d'être maniés et interprétés par un utilisateur averti. Ils peuvent permettre d'éviter des tests expérimentaux ou de ne les effectuer que lorsque l'interprétation des résultats laisse planer une suspicion. Ils peuvent permettre aussi de compléter l'information expérimentale. Les QSAR sont faciles à mettre en œuvre, ils fournissent un résultat immédiat et leurs coûts sont sans commune mesure avec ceux engendrés par la réalisation de tests expérimentaux. Ils peuvent être réalisés par des laboratoires équipés des logiciels adaptés ou par le demandeur lui-même s'il dispose des outils informatiques et des personnes compétentes pour valider les résultats car on ne saurait trop insister sur les risques de conclusions hâtives.

La relation quantitative structure activité(QSAR) est depuis quelques années utilisée de façon systématique dans la conception de nouveaux médicaments et permet de prédire l'activité thérapeutique ou la biodisponibilité des molécules avant même de les avoir synthétisées.

III.1. Introduction

Les outils classiques de l'analyse de données en sciences humaines sont connus. Ils appartiennent à une même famille : analyse de variance, analyse typologique, analyse en composante principales, etc. l'analyse révèle les causalités, des typologies, des caractéristiques qui facilitent la spécification de modèle explicatifs. L'approche classique consiste à formuler explicitement un modèle et à l'estimer. L'explication d'une réalité complexe peut toutefois buter sur les difficultés d'estimation des modèles complexe ou sur les limites des modèles plus facilement estimable mais exagérément simplifiés par rapport à la réalité.

Une bonne prédiction sur les données résulte généralement d'un bon modèle explicatif mais on sait qu'il est impossible de prévoir avec succès sans pour autant expliquer. Le choix de la forme fonctionnelle se fait alors sans référence à un cadre conceptuel à priori cohérent avec le phénomène analysé.

Dans ce chapitre nous présentons les résultats préliminaires de la modélisation de la biodisponibilité des médicaments à partir de leurs structures chimiques par la régression linéaire multiple.

Ce chapitre est divisé en deux parties : la première partie concerne la préparation du modèle (l'élaboration de la base de données). Dans la deuxième partie on s'intéressera aux résultats et discussion.

Enfin on conclura ce chapitre en analysant et interprétant les différents résultats obtenus.

III.2. Méthodologie de calcul

III.2.1.Calcul des descripteurs moléculaires

Un descripteur moléculaire peut être considéré comme la conséquence d'un processus logique et mathématique, appliqué à l'information chimique codifiée à travers la représentation d'une molécule.

Le choix des descripteurs dépend des outils dont on dispose, de la nature des composés décrits et de la propriété ciblée.

L'information codée d'un descripteur moléculaire dépend du type de représentation moléculaire employée et de l'algorithme défini pour son calcul. [13]

Avant calculer les descripteurs moléculaires, il faut obtenir les structures chimiques des médicaments.

Toutes les structures chimiques employées pour le développement des modèles QSAR ont été obtenus à l'aide du« ChemSpider »qui est une base de données de produits chimiques à partir du site : **www.chemspider.com**

Ces structures ont ensuite été employées pour le calcul des descripteurs moléculaires.



Figure III.1 Image de ChemSpider

Les molécules sont enregistrées sur format mol, et pour lire ce format en utilise le logiciel : « ChemBioOffice 2014(ChemBio3D Ultra 14.0) ». Ce logiciel nous donne encore la structure 2D de la molécule qu'on peut utiliser à la base de données.

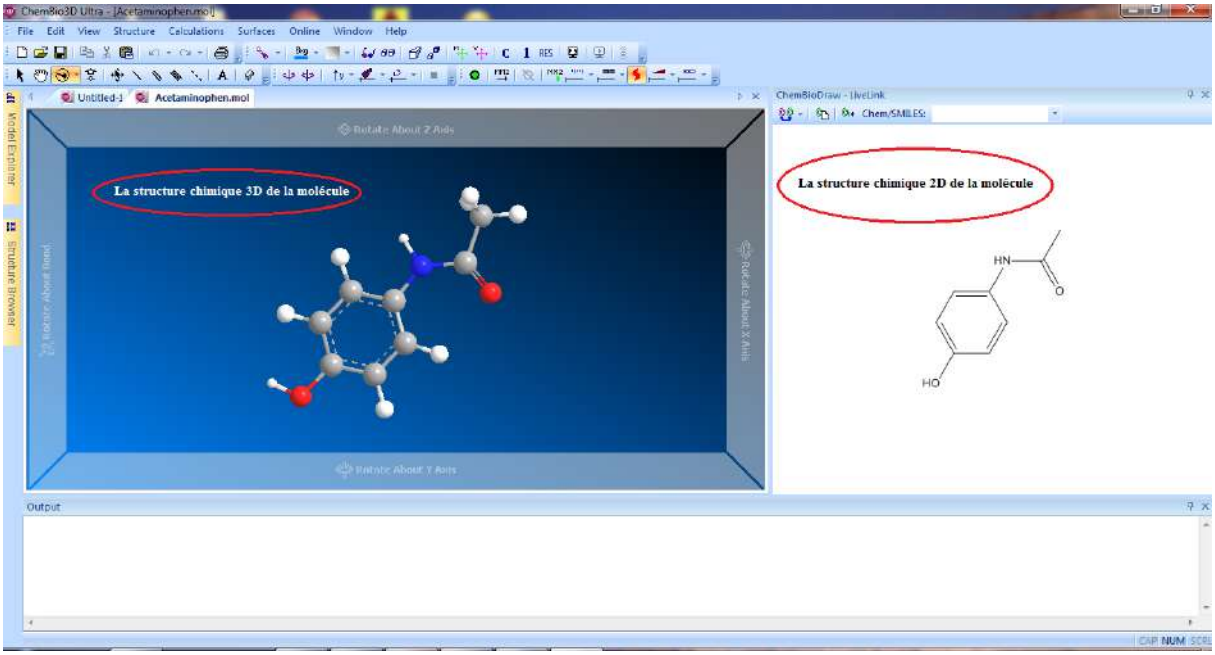


Figure III.2 Image de ChemBio3D Ultra 14.0

Dans notre travail nous avons utilisé un paramètre biologique : la biodisponibilité.Et dans le tableau suivant, il y a une liste des médicaments utilisés dans cette étude avec leurs biodisponibilités. [42]

Tableau III. Liste des médicaments utilisés dans cette étude avec leurs valeurs de biodisponibilité expérimentales.

N^0	Molécules	Structures chimiques	Biodisponibilité	Inverse de la biodisponibilité
1	Acetaminophen	$C_8H_9NO_2$	63	0,015873017
2	Acyclovir	$C_8H_{11}N_5O_3$	23	0,043478260
3	Alendronate	$C_4H_{13}NO_7P_2$	1	1,000000000
4	Allopurinol	$C_5H_4N_4O$	52	0,019230770
5	Amiloride	$C_6H_9Cl_2N_7O$	50	0,020000000
6	Aminoglutethimide	$C_{13}H_{16}N_2O_2$	90	0,011111111
7	Amiodarone	$C_{25}H_{29}I_2NO_3$	38	0,026315790
8	Amitriptiline	$C_{20}H_{23}N$	50	0,020000000
9	Amlodipine	$C_{20}H_{25}ClN_2O_5$	60	0,016666668
10	Amoxicillin	$C_{16}H_{19}N_3O_5S$	93	0,010752688
11	Ampicillin	$C_{16}H_{19}N_3O_4S$	60	0,016666668
12	Aspirin	$C_9H_8O_4$	68	0,014705882
13	Atenolol	$C_{14}H_{22}N_2O_3$	50	0,020000000
14	Atorvastatin	$C_{33}H_{35}FN_2O_5$	14	0,071428575
15	Atropine	$C_{17}H_{23}NO_3$	50	0,020000000
16	Baclofen	$C_{10}H_{12}ClNO_2$	70	0,014285714
17	Bendrofluazide	$C_{15}H_{14}F_3N_3O_4S_2$	90	0,011111111

18	Betamethasone	$C_{22}H_{29}FO_5$	70	0,014285714
19	Bromocriptine	$C_{32}H_{40}BrN_5O_5$	10	0,100000000
20	Bumetanide	$C_{17}H_{20}N_2O_5S$	80	0,012500000
21	Captopril	$C_9H_{15}NO_3S$	70	0,014285714
22	Carbamazepine	$C_{15}H_{12}N_2O$	70	0,014285714
23	Cephalexin	$C_{16}H_{17}N_3O_4S$	90	0,011111111
24	Chlortalidone	$C_{14}H_{11}ClN_2O_4S$	70	0,014285714
25	Cimetidine	$C_{10}H_{16}N_6S$	68	0,014705882
26	Ciprofloxacin	$C_{17}H_{18}FN_3O_3$	60	0,016666668
27	Cladribine	$C_{10}H_{12}ClN_5O_3$	50	0,020000000
28	Clonidine	$C_9H_9Cl_2N_3$	83	0,012048192
29	Cloxacillin	$C_{19}H_{18}ClN_3O_5S$	40	0,025000000
30	Dexamethasone	$C_{22}H_{29}FO_5$	90	0,011111111
31	Dicloxacillin	$C_{19}H_{17}Cl_2N_3O_5S$	60	0,016666668
32	Didanosine	$C_{10}H_{12}N_4O_3$	50	0,020000000
33	Disopyramide	$C_{21}H_{29}N_3O$	72	0,013888889
34	Domperidone	$C_{22}H_{24}ClN_5O_2$	15	0,066666670
35	Doxapram	$C_{24}H_{30}N_2O_2$	60	0,016666668
36	Doxepin	$C_{19}H_{21}NO$	30	0,033333335
37	Doxorubicin	$C_{27}H_{29}NO_{11}$	5	0,200000000
38	Enalapril	$C_{27}H_{40}$	60	0,016666668
39	Ethinyloestradiol	$C_{20}H_{24}O_2$	50	0,020000000
40	Etidronate	$C_2H_8O_7P_2$	5	0,200000000
41	Etoposide	$C_{29}H_{32}O_{13}$	45	0,022222223
42	Famotidine	$C_8H_{15}N_7O_2S_3$	50	0,020000000
43	Felodipine	$C_{18}H_{19}Cl_2NO_4$	15	0,066666670
44	Fexofenadine	$C_{32}H_{39}NO_4$	30	0,033333335
45	Finasteride	$C_{17}H_{20}F_6N_2O_3$	80	0,012500000
46	Flecainide	$C_{24}H_{26}FNO_4$	90	0,011111111
47	Fluconazole	$C_{13}H_{12}F_2N_6O$	90	0,011111111
48	Flucytozine	$C_4H_4FN_3O$	80	0,012500000
49	Fluorouracil	$C_4H_3FN_2O_2$	30	0,033333335
50	Fluvastatin	$C_{24}H_{26}FNO4$	22	0,045454547
51	Furosemide	$C_{12}H_{11}ClN_2O_5S$	60	0,016666668
52	Gabapentin	$C_9H_{17}NO_2$	70	0,014285714
53	Gemfibrozil	$C_{15}H_{22}O_3$	90	0,011111111
54	Glibenclamide	$C_{23}H_{28}ClN_3O_5S$	80	0,012500000
55	Glipizide	$C_{21}H_{27}N_5O_4S$	90	0,011111111
56	Haloperidol	$C_{21}H_{23}ClFNO_2$	60	0,016666668
57	Hydralazine	$C_8H_8N_4$	40	0,025000000
58	Hydrochlorothiazide	$C_7H_8ClN_3O_4S_2$	70	0,014285714
59	Hyoscine	$C_{17}H_{21}NO_4$	20	0,050000000
60	Ibuprofen	$C_{13}H_{18}O_2$	90	0,011111111
61	Idarubicin	$C_{26}H_{27}NO_9$	30	0,033333335
62	Imipramine	$C_{19}H_{24}N_2$	47	0,021276595
63	Indapamide	$C_{16}H_{16}ClN_3O_3S$	90	0,011111111
64	Indometacin	$C_{19}H_{16}ClNO_4$	87	0,011494253
65	Irbesartan	$C_{25}H_{28}N_6O$	70	0,014285714

66	Ketoprofen	$C_{16}H_{14}O_3$	90	0,011111111
67	Labetalol	$C_{19}H_{24}N_2O_3$	30	0,033333335
68	Lamivudine	$C_8H_{11}N_3O_3S$	80	0,012500000
69	Lamotrigine	$C_9H_7Cl_2N_5$	95	0,010526316
70	Lansoprazole	$C_{16}H_{14}F_3N_3O_2S$	80	0,012500000
71	Levodopa	$C_{14}H_{22}N_2O$	44	0,022727273
72	Lidocaine	$C_{14}H_{22}N_2O$	35	0,028571429
73	Lisinopril	$C_{21}H_{31}N_3O_5$	40	0,025000000
74	Lorazepam	$C_{15}H_{10}Cl_2N_2O_2$	95	0,010526316
75	Losartan	$C_{22}H_{23}ClN_6O$	33	0,030303031
76	Mercaptopurine	$C_5H_4N_4S$	46	0,021739130
77	Metformin	$C_4H_{11}N_5$	50	0,020000000
78	Methotrexate	$C_{20}H_{22}N_8O_5$	50	0,020000000
79	Methylprednisolone	$C_{22}H_{30}O_5$	70	0,014285714
80	Metoclopramide	$C_{14}H_{22}ClN_3O_2$	60	0,016666668
81	Metoprolol	$C_{15}H_{25}NO3$	50	0,020000000
82	Mianserin	$C_{18}H_{20}N_2$	30	0,033333335
83	Minocycline	$C_{23}H_{27}N_3O_7$	100	0,010000000
84	Misoprostol	$C_{22}H_{38}O_5$	80	0,012500000
85	Morphine	$C_{17}H_{19}NO_3$	27	0,037037037
86	Naloxone	$C_{19}H_{21}NO_4$	2	0,500000000
87	Naltrexone	$C_{20}H_{23}NO_4$	23	0,043478260
88	Norfloxacin	$C_{16}H_{18}FN_3O_3$	45	0,022222223
89	Omeprazole	$C_{17}H_{19}N_3O_3S$	50	0,020000000
90	Ondansetron	$C_{18}H_{19}N_3O$	60	0,016666668
91	Oxprenolol	$C_{15}H_{23}NO_3$	50	0,020000000
92	Oxycodone	$C_{18}H_{21}NO_4$	50	0,020000000
93	Pantoprazole	$C_{16}H_{15}F_2N_3O_4S$	80	0,012500000
94	Pethidine	$C_{15}H_{21}NO_2$	60	0,016666668
95	Phenobarbitone	$C_{12}H_{12}N_2O_3$	80	0,012500000
96	Phenytoin	$C_{15}H_{12}N_2O_2$	98	0,010204081
97	Pindolol	$C_{14}H_{20}N_2O_2$	90	0,011111111
98	Pravastatin	$C_{23}H_{36}O_7$	20	0,050000000
99	Prazosin	$C_{19}H_{21}N_5O_4$	57	0,017543860
100	Primidone	$C_{12}H_{14}N_2O_2$	70	0,014285714
101	Procainamide	$C_{13}H_{21}N_3O$	83	0,012048192
102	Prochlorperazine	$C_{20}H_{24}ClN_3S$	20	0,050000000
103	Promethazine	$C_{17}H_{20}N_2S$	25	0,040000000
104	Propranolol	$C_{16}H_{21}NO_2$	36	0,027777778
105	Propylthiouracil	$C_7H_{10}N_2OS$	80	0,012500000
106	Quinapril	$C_{25}H_{30}N_2O_5$	60	0,016666668
107	Quinine	$C_{20}H_{24}N_2O_2$	70	0,014285714
108	Ramipril	$C_{23}H_{32}N_2O_5$	30	0,033333335
109	Ribavirin	$C_8H_{12}N_4O_5$	60	0,016666668
110	Salicylic acid	$C_7H_6O_3$	60	0,016666668
111	Selegiline	$C_{13}H_{17}N$	20	0,050000000
112	Simvastatin	$C_{25}H_{38}O_5$	26	0,038461540
113	Sotalol	$C_{12}H_{20}N_2O_3S$	90	0,011111111

114	Spironolactone	$C_{24}H_{32}O_4S$	90	0,011111111
115	Sulfamethoxazole	$C_{10}H_{11}N_3O_3S$	80	0,012500000
116	Sulfisoxazole	$C_{11}H_{13}N_3O_3S$	90	0,011111111
117	Sumatriptan	$C_{14}H_{21}N_3O_2S$	20	0,050000000
118	Temazepam	$C_{16}H_{13}ClN_2O_2$	40	0,025000000
119	Terbutaline	$C_{12}H_{19}NO_3$	15	0,066666670
120	Testosterone	$C_{19}H_{28}O_2$	66	0,015151516
121	Tetracycline	$C_{22}H_{24}N_2O_8$	77	0,012987013
122	Theophyline	$C_7H_8N_4O_2$	90	0,011111111
123	Timolol	$C_{13}H_{24}N_4O_3S$	90	0,011111111
124	Tolbutamide	$C_{12}H_{18}N_2O_3S$	93	0,010752688
125	Triamterene	$C_{12}H_{11}N_7$	50	0,020000000
126	Trimethoprim	$C_{14}H_{18}N_4O_3$	88	0,011363637
127	Trimipramine	$C_{20}H_{26}N_2$	40	0,025000000
128	Valproic acid	$C_8H_{16}O_2$	100	0,010000000
129	Verapamil	$C_{27}H_{38}N_2O_4$	28	0,035714287
130	Zalcitabine	$C_9H_{13}N_3O_3$	70	0,014285714

Pour calculer les descripteurs moléculaires on utilise le logiciel : « PaDEL-Descriptor ».

PaDEL-Descriptor est un logiciel de calcul de descripteurs moléculaires et les empreintes digitales. PaDEL-Descriptor a été développé en utilisant le langage Java et se compose d'un élément de bibliothèque et un composant d'interface.

Le logiciel présente plusieurs avantages par rapport aux logiciels existants de calcul de descripteur moléculaire autonome. Il est gratuit, peut travailler sur toutes les grandes plates-formes (Windows, Linux, MacOS), prend en charge plus de 90 formats de fichiers moléculaires différents.

PaDEL -Descriptor est un ajout utile au logiciel de calcul des descripteurs moléculaires qui sont actuellement disponibles. Le logiciel peut être téléchargé à l'adresse :

http://padel.nus.edu.sg/softwar / padeldescriptor. [43]

Pour utiliser PaDEL–Descriptor :

1. Il faut sélectionner le dossier qui contient les structures chimiques des molécules (médicaments) ;
2. Il faut sélectionner la sortie du calcul ;
3. IL faut sélectionner les descripteurs (dans notre cas on calcule seulement les descripteurs 1D et 2D) ;
4. Enfin, cliquer sur start pour lancer le calcul.

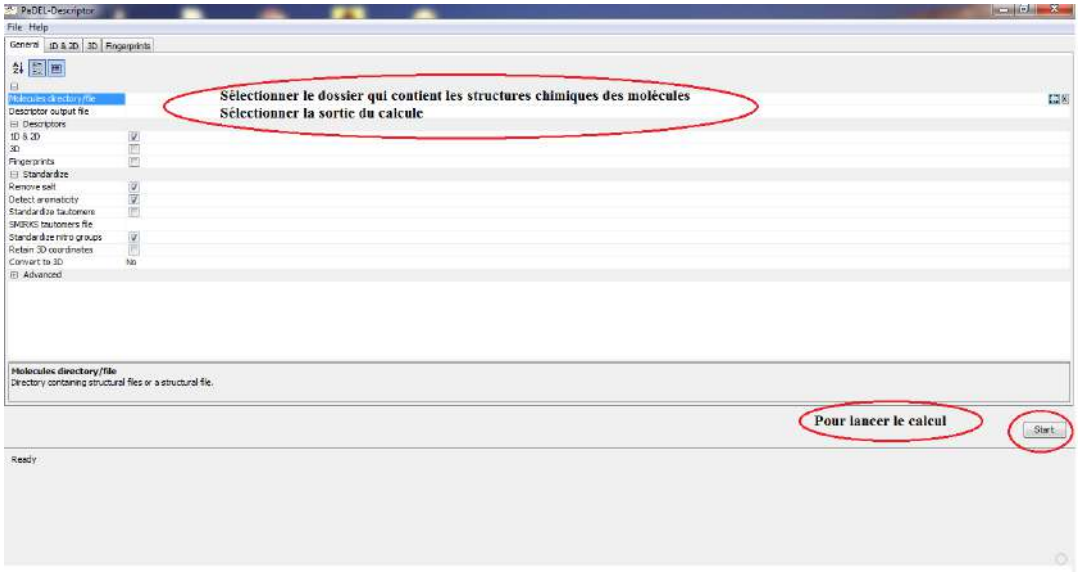


Figure III.3 Image de PaDel-Descriptor

Lorsque PaDEL –Descriptor termine le calcul on a besoin d'un logiciel qui lit les descripteurs moléculaires calculés, ce logiciel est : « CSV Viewer ».

CSV Viewer est une visionneuse de fichiers gratuite et simple pour les fichiers CSV, il peut ouvrir rapidement et formater les données dans les (* .csv), pour aider à voir facilement, il est très facile à utiliser et totalement gratuit.

Un fichier CSV est un fichier de valeurs séparées par des virgules (Comma Separated Value), il est un format de fichier texte brut qui est utilisé pour stocker des données, généralement utilisés dans un logiciel de tableur ou base de données. En utilisant un fichier CSV, les utilisateurs peuvent facilement échanger des données entre les différentes applications. [44]



Figure III.4Image représente l'ouverture d'un fichier (.csv) sur CSV Viewer

III.2.2. Elaboration de la base de données

Coller les descripteurs calculés par PaDEL–Descriptor sur Excel. Dans cette étape nous avons sélectionné 130 molécules de médicaments. La base de données de la **Figure III.5** contient 130 lignes qui représentent les molécules, et 1445 colonnes qui représentent leurs propriétés (descripteurs) et deux colonnes sont réservées pour décrire la biodisponibilité et son inverse de chaque molécule.

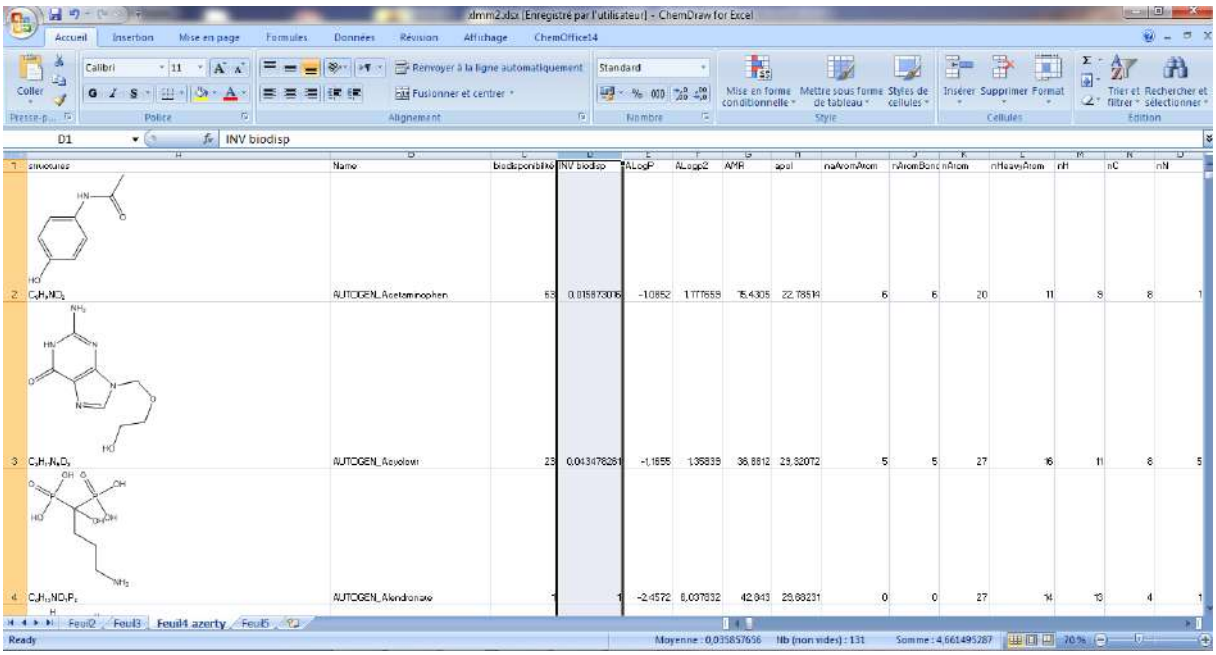


Figure III.5 la base de données sur Excel

III.2.3. Prétraitement de la base de données

1ère Étape : enlever les descripteurs avec des valeurs pratiquement constantes ou nulles.

2ème Étape : éliminer les descripteurs qui ont une valeur identique pour plus de 80% des échantillons.

La base de données qui résulte contient 130 lignes qui représentent les molécules, et 950 colonnes qui représentent leurs propriétés (descripteurs) et deux colonnes sont réservées pour décrire la biodisponibilité et son inverse de chaque molécule.

III.2.4. L'étape de la régression linéaire multiple (MLR)

Après l'élaboration de la base de données on a besoin d'un logiciel qui fait la régression linéaire qui est « SPSS 19 » (Statistical Package for the Social Sciences) est un logiciel utilisé pour l'analyse statistique. C'est aussi le nom de la société qui le revend (SPSS Inc). [45]

On va ouvrir la base de données de l'Excel sur SPSS et faire la régression linéaire.

Les étapes suivis sur SPSS

1. Ouvrir la base de données à partir de l'Excel ;
2. Sur SPSS, faire : Analyse ⟶ Régression ⟶ Linéaire ;

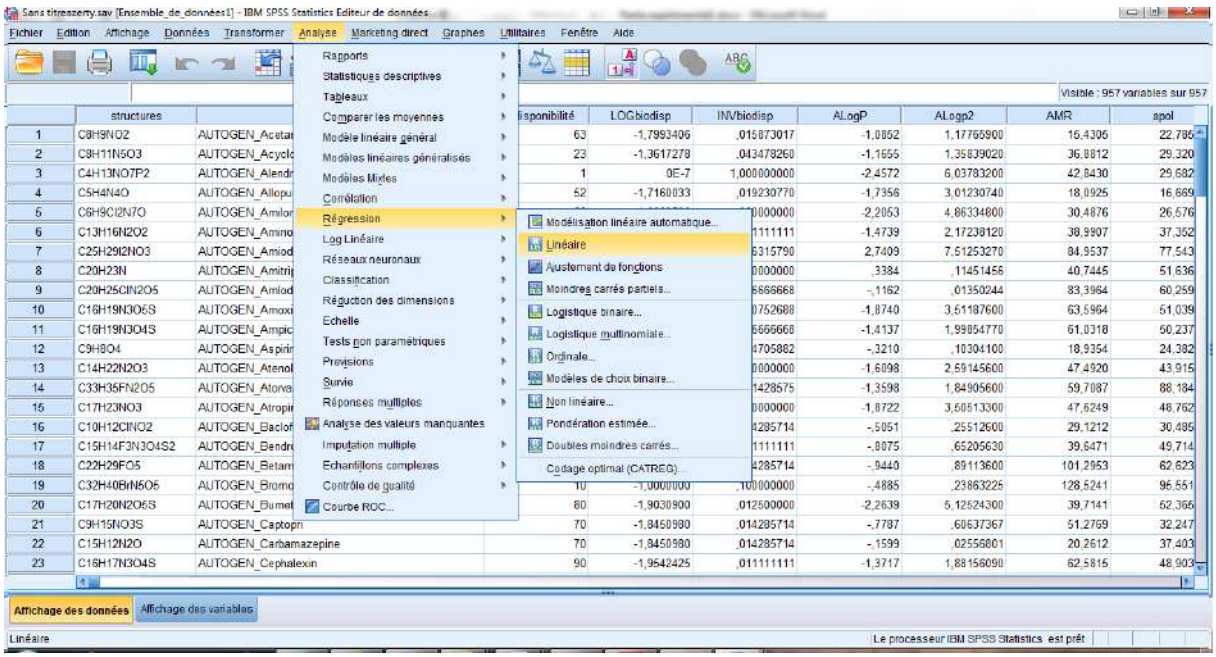


Figure III.6 La première étape sur SPSS

3. Sélectionner la variable indépendante ;
4. Sélectionner les variables indépendantes ;
5. Choisir la méthode pas à pas ;
6. Cliquer sur statistiques et choisir les coefficients de régression ;
7. Cliquer sur poursuivre, ensuite sur OK.

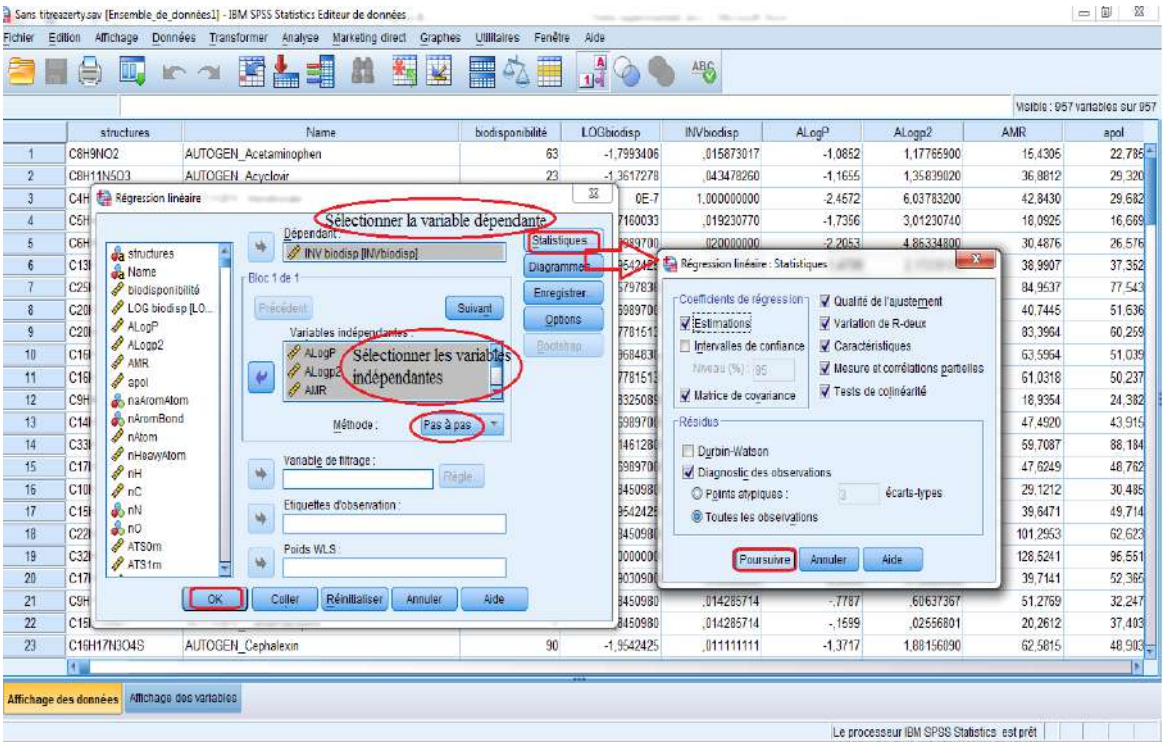


Figure III.7 La deuxième étape sur SPSS

III.3. Résultats et Discussion

III.3.1. Equation de la régression linéaire multiple

L'équation de la régression multiple est sous la forme :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots + \beta_p X_p + \varepsilon$$

On nomme la biodisponibilité par : «F » et à son inverse par : « 1/F»

Remplaçons maintenant lesβ par les coefficients fournis dans le tableau [Annexe 1].

La corrélation entre l'activité biologique (la biodisponibilité) et les descripteurs exprimés par la relation suivante:

$$1/F = -7.255 .10^{-3} + 3.996 .10^{-4} MPC8 + 1.005 .10^{-2} nAtomLAC + 1.419 .10^{-7} VR1_Dzs - 1.412 .10^{-7} VR1_Dt + 3.473 AATSC6c - 6.018 .10^{-5} ATS6e + 2.711 AATSC7c + 1.081 .10^{-2} maxwHBa$$

D'après cette équation on remarque que l'inverse de la biodisponibilité a une relation avec 8 descripteurs qui sont représentés dans le tableau III.2. [Annexe1]

Tableau III. Les descripteurs moléculaires de cette étude

Nom du descripteur	Description	La classe du descripteur	La classe prolongée
MPC8	Molecular path count of order 8 (compte de l'ordre du trajet moléculaire)	Path counts (compte de trajets) [46]	Topological descriptors (descripteurs topologiques)
nAtomLAC	Number of atoms in the longest aliphatic chain (nombre d'atomes dans la plus longue chaine aliphatique)	Longest aliphatic chain (La plus longue chaine aliphatique)	Constitutional descriptors (descripteurs constitutionnels)
VR1_Dzs	Randic-like eigenvector-based index from Barysz matrix / weighted by I-state (Indice de base-Randic vecteur propre comme de la matrice Barysz / pondérée par I-état)	Barysz matrix [47] (Barysz matrice)	Topological descriptors (descripteurs topologiques)
VR1_Dt	Randic-like eigenvector-based index from detour matrix ((Indice de base-Randic vecteur propre comme de la matrice de déviation)	Detour matrix [47] (matrice de déviation)	Topological descriptors (descripteurs topologiques)
AATSC6c	Average centered Broto-Moreau autocorrelation - lag 6 / weighted by charges (Moyenne centrée Broto-Moreau autocorrélation - lag 6 / pondérée par des charges)	Autocorrelation (autocorrélation) [48]	Autocorrelation descriptors (descripteurs d'autocorrélation)
ATS6e	Moran autocorrelation - lag 6 / weighted by Sanderson electronegativities (Moran autocorrélation - décalage 6 / pondéré par électronégativité Sanderson)	Autocorrelation (autocorrélation) [48]	Autocorrelation descriptors (descripteurs d'autocorrélation)
AATSC7c	Average centered Broto-Moreau autocorrelation - lag 7 / weighted by charges (Moyenne centrée Broto-Moreau autocorrélation - décalage 7 / pondérés par des charges)	Autocorrelation (autocorrélation) [48]	Autocorrelation descriptors (descripteurs d'autocorrélation)
maxwHBa	Maximum E-States for weak Hydrogen Bond acceptors (Maximum E-Etats pour accepteurs obligataires faibles d'hydrogène)	Atom type electrotopological state (type Atom état electrotopological) [49] (Type d'atome état electrotopological)	E-state descriptors (descriptions immobilières)

On a remarqué que l'équation de régression résultante contient 4 descripteurs de classe descripteurs topologiques, 4 descripteurs de classe descripteurs d'autocorrélation, 1descripteurs de classe descripteurs constitutionnels et 1descripteurs de classe immobilière.

Après la régression avec SPSS on a trouvé que le meilleur modèle a un coefficient de corrélation R=0.916, R^2=0.839 et R^2 ajusté = 0.829, il est représenté dans le tableau suivant : [annexe 2]

Tableau III.3Récapitulatif du modèle

R	R-deux	R-deux ajusté
0,916	0,839	0,829

Valeurs prédites : (constantes), MPC8, nAtomLAC, VR1_Dzs, VR1_Dt, AATSC6c, ATS6e, AATSC7c, maxwHBa.
Variable dépendante : Inverse de la biodisponibilité

On a besoin du tableau ANOVA pour tester la significativité [annexe 3]

Tableau III.4 Tableau ANOVA

Modèle	Somme des carrées	ddl	Moyen des carrées	D =F	Sig
Régression Résidu Total	1.050 0.201 1.251	8 121 129	0.131 0.002	78.937	0.000

Sig (significativité)= 0<0.05 .C'est à dire le modèle est significatif.

1. **Le coefficient de corrélation R** est la corrélation entre y et sa prédiction par le modèle : plus il est élevé plus ça signifie que le modèle donne de bonnes « prédictions » pour y. R est significatif, il donne deux informations importantes :

    a. Le sens de la relation linéaire entre les variables : Le coefficient de corrélation varie entre - 1 et 1. **R = 0.916** indique une corrélation positive parfaite entre les variables. lorsque les variables x augmentent, la variable y augmente dans la même proportion.

    b. La force de la relation linéaire entre les variables : Plus la valeur du coefficient est proche de + 1, plus les variables sont associées fortement.

2. **Le coefficient de détermination (R^2)** est un indicateur qui permet de juger la qualité d'une régression linéaire, simple ou multiple. D'une valeur comprise entre 0 et 1, il mesure l'adéquation entre le modèle et les données observées.

    Le R^2 est une mesure de la précision de l'ajustement de la droite de régression. Il s'agit du rapport entre la variation de la variable dépendante expliquée par le modèle de régression et sa variation totale.

    Le modèle QSAR doit considérer un coefficient de détermination $R^2 > 0.6$ pour qu'il soit valide [13].

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

$$ESS = Régression = 1.050$$
$$RSS = Résidu = 0.201$$
$$TSS = Total = 1.251$$

$R^2 = 1.050/1.251 = 0.839$

III.3.2.Interprétation

$R^2 = 0.839 = 83.9\%$ signifie que 83.9% des variations de la variable dépendante sont expliqués par le modèle de régression et que 16.1% restent par conséquent inexpliqués : la corrélation est bonne.

Les valeurs **$R = 0.916$ et $R^2 = 0.839$** nous permet d'indiquer fermement la corrélation entre les différents descripteurs utilisés et la biodisponibilité.

3. **Le coefficient de détermination ajusté** (tient compte du nombre de variables) **R^2 ajusté**. En effet, le principal défaut du R^2 est de croître avec le nombre de variables explicatives. Or, on sait qu'un excès de variables produit des modèles peu robustes. C'est pourquoi on

s'intéresse davantage à cet indicateur qu'au R^2. Mais ce n'est pas un véritable « carré » et il peut même être négatif.

$$R^2 \text{ ajusté} = 1-(1-R^2)\left[\frac{n-1}{n-(k+1)}\right]$$

$$R^2 \text{ ajusté} = 1-(1-0.839)\left[\frac{952-1}{952-(8+1)}\right]$$

- n = La taille de l'échantillon
- k = Le nombre de variables indépendantes dans l'équation de régression

R^2ajusté=0.829 Cette valeur est adopté pour la validité du modèle.

Dans le tableau suivant, on fait une comparaison entre les valeurs expérimentales (obtenus par les expériences) et les valeurs prédites (obtenus par le modèle de MLR)

Tableau III.5 Comparaison entre les valeurs expérimentales et les valeurs prédites

Numéro de l'observation	Erreur Résidu	INV biodisponibilité (valeur expérimentale)	Prévision (valeur prédite)	Résidu
1	,245	,015873017	,00587980567	,009993211327
2	,172	,043478260	,03645068070	,007027579304
3	,000	1,000000000	1,00002000954	-,000020009539
4	-,294	,019230770	,03122333826	-,011992568261
5	,676	,020000000	-,00758086779	,027580867792
6	,176	,011111111	,00391931906	,007191791945
7	-,482	,026315790	,04596048441	-,019644694415
8	-,509	,020000000	,04075007306	-,020750073062
9	,109	,016666668	,01220883142	,004457836581
10	-,414	,010752688	,02762871770	-,016876029700
11	-,175	,016666668	,02380328393	-,007136615926
12	,121	,014705882	,00978663608	,004919245918
13	,058	,020000000	,01763662922	,002363370780
14	,198	,071428575	,06334232144	,008086253563
15	-,267	,020000000	,03089179558	-,010891795578
16	-,137	,014285714	,01986049852	-,005574784520
17	-,321	,011111111	,02419751305	-,013086402053
18	-,480	,014285714	,03384185240	-,019556138400
19	-,004	,100000000	,10016543186	-,000165431861
20	-,637	,012500000	,03847918528	-,025979185277

21	,325	,014285714	,00104426015	,013241453855
22	,055	,014285714	,01203040542	,002255308577
23	-,341	,011111111	,02501252075	-,013901409755
24	-,137	,014285714	,01988399623	-,005598282227
25	,080	,014705882	,01144899975	,003256882246
26	-,073	,016666668	,01965165908	-,002984991081
27	-,283	,020000000	,03152164826	-,011521648262
28	,277	,012048192	,00073387859	,011314313409
29	,073	,025000000	,02201581402	,002984185980
30	-,557	,011111111	,03384185240	-,022730741400
31	-,289	,016666668	,02843747043	-,011770802433
32	-,171	,020000000	,02695871382	-,006958713818
33	,571	,013888889	-,00940721826	,023296107262
34	,470	,066666670	,04750279641	,019163873592
35	,408	,016666668	,00004173474	,016624933256
36	-,270	,033333335	,04434645081	-,011013115812
37	2,881	,200000000	,08252249567	,117477504334
38	,073	,016666668	,01367743866	,002989229340
39	-1,759	,020000000	,09173347391	-,071733473908
40	,865	,200000000	,16472825999	,035271740009
41	-1,251	,022222223	,07321609684	-,050993873837
42	-,587	,020000000	,04395119126	-,023951191264
43	1,608	,066666670	,00108055085	,065586119154
44	,911	,033333335	-,00383110350	,037164438495
45	,017	,012500000	,01182089828	,000679101722
46	,991	,011111111	-,02928506631	,040396177307
47	-,352	,011111111	,02548477625	-,014373665250
48	,265	,012500000	,00168126989	,010818730110
49	,807	,033333335	,00041378032	,032919554683
50	-,703	,045454547	,07411343702	-,028658890023
51	-,430	,016666668	,03420874787	-,017542079867
52	,708	,014285714	-,01456566431	,028851378311
53	-,779	,011111111	,04288914249	-,031778031486
54	,300	,012500000	,00025952172	,012240478281
55	,061	,011111111	,00861990837	,002491202630
56	-,343	,016666668	,03066585181	-,013999183810
57	,207	,025000000	,01655418926	,008445810740
58	-,768	,014285714	,04560303792	-,031317323920
59	,185	,050000000	,04246918271	,007530817289
60	-,413	,011111111	,02793916615	-,016828055151

61	-,565	,033333335	,05637902341	-,023045688411
62	-,251	,021276595	,03149249343	-,010215898431
63	-,382	,011111111	,02668358807	-,015572477073
64	-,585	,011494253	,03535979534	-,023865542341
65	-,759	,014285714	,04524115149	-,030955437487
66	-,496	,011111111	,03132655929	-,020215448289
67	,105	,033333335	,02906744793	,004265887071
68	-,045	,012500000	,01434222921	-,001842229212
69	-,231	,010526316	,01994943921	-,009423123214
70	-,418	,012500000	,02954481695	-,017044816948
71	-,270	,022727273	,03373798464	-,011010711636
72	,654	,028571429	,00190894302	,026662485984
73	-,116	,025000000	,02971897171	-,004718971712
74	-,142	,010526316	,01631992538	-,005793609377
75	-,342	,030303031	,04423406158	-,013931030576
76	-,413	,021739130	,03858516484	-,016846034841
77	,708	,020000000	-,00887069718	,028870697182
78	-,852	,020000000	,05475201370	-,034752013696
79	-,495	,014285714	,03447875727	-,020193043265
80	,383	,016666668	,00103211915	,015634548851
81	,104	,020000000	,01574549530	,004254504696
82	,228	,033333335	,02403910525	,009294229749
83	-,456	,010000000	,02860729466	-,018607294660
84	-,592	,012500000	,03665728530	-,024157285300
85	-1,514	,037037037	,09876465317	-,061727616169
86	8,381	,500000000	,15822016115	,341779838851
87	-1,711	,043478260	,11326754885	-,069789288850
88	,350	,022222223	,00795670286	,014265520141
89	,300	,020000000	,00776134973	,012238650269
90	-,248	,016666668	,02678588063	-,010119212630
91	-,365	,020000000	,03489628287	-,014896282866
92	-2,300	,020000000	,11379258109	-,093792581092
93	-,262	,012500000	,02316843999	-,010668439988
94	,832	,016666668	-,01727259517	,033939263166
95	,390	,012500000	-,00339450708	,015894507079
96	,059	,010204081	,00778884638	,002415234622
97	-,418	,011111111	,02814424908	-,017033138083
98	,192	,050000000	,04218943014	,007810569864
99	-,050	,017543860	,01959653789	-,002052677895
100	,512	,014285714	-,00657869691	,020864410908

101	-,009	,012048192	,01242046204	-,000372270043
102	,039	,050000000	,04839434236	,001605657643
103	-,241	,040000000	,04982584799	-,009825847988
104	,047	,027777778	,02585702614	,001920751865
105	-,419	,012500000	,02957257577	-,017072575766
106	-,108	,016666668	,02107563602	-,004408968022
107	-,579	,014285714	,03789003767	-,023604323672
108	,605	,033333335	,00866657832	,024666756685
109	,239	,016666668	,00690134706	,009765320938
110	-,747	,016666668	,04713606623	-,030469398232
111	-,252	,050000000	,06028693532	-,010286935315
112	,520	,038461540	,01726177257	,021199767435
113	,109	,011111111	,00666754116	,004443569842
114	-,561	,011111111	,03397592882	-,022864817815
115	-,084	,012500000	,01593804131	-,003438041313
116	,144	,011111111	,00522314365	,005887967349
117	,520	,050000000	,02878543264	,021214567359
118	,294	,025000000	,01302433064	,011975669356
119	1,831	,066666670	-,00801330999	,074679979986
120	-,009	,015151516	,01552504925	-,000373533246
121	-,351	,012987013	,02730561119	-,014318598187
122	,106	,011111111	,00677567663	,004335434367
123	-,120	,011111111	,01599815245	-,004887041453
124	-,551	,010752688	,03322343021	-,022470742213
125	-,220	,020000000	,02898630262	-,008986302622
126	,380	,011363637	-,00412749190	,015491128901
127	,091	,025000000	,02130925122	,003690748783
128	-,798	,010000000	,04252450829	-,032524508295
129	,386	,035714287	,01999157912	,015722707880
130	,149	,014285714	,00819258850	,006093125505

a. Variable dépendante : INV biodisponibilité

On trace le graphe des valeurs prédites en fonction des valeurs expérimentales de 1/F.

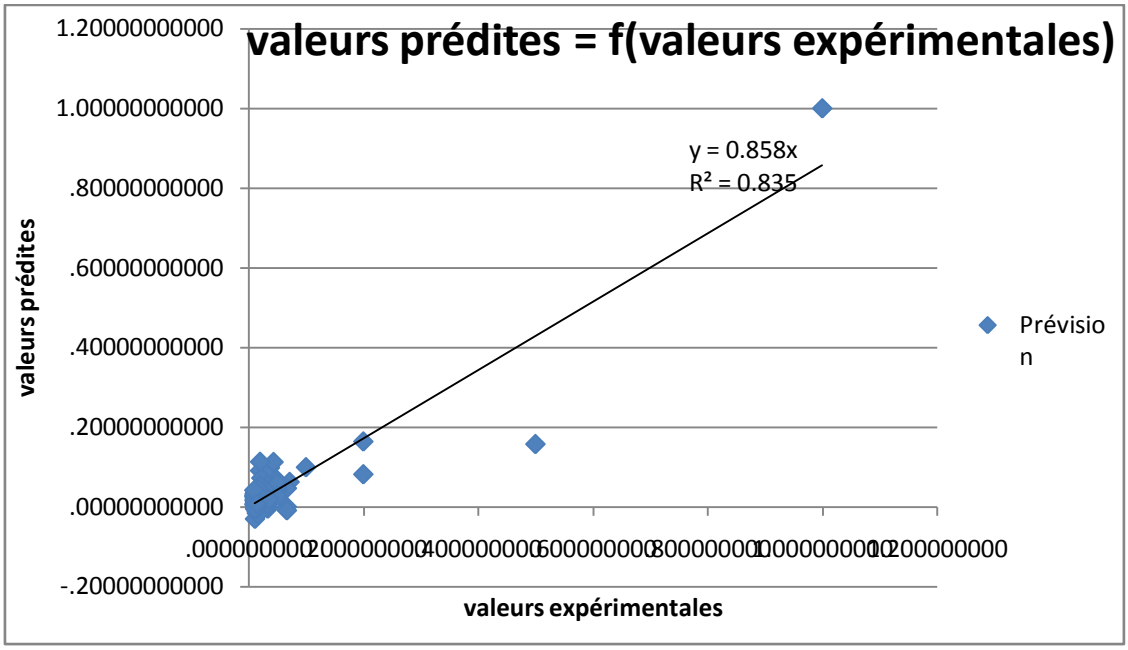


Figure III.8 Corrélation entre les valeurs de l'inverse de la biodisponibilité prédites et expérimentales

Interprétation du graphe :

La courbe est une droite qui passe par l'origine, son équation sous forme : Y=a X

Tel que Y=0.858X

Avec : Y=prévision

X= inverse de la biodisponibilité (1/F)

On remarque que la courbe passe par la majorité des points, et il ya une accumulation très importante des point, comme il ya quelques des points atypiques.

R^2=0.835. Il y a une forte corrélation entre les valeurs prédites et les valeurs expérimentales c'est-à-dire les molécules sont proches de la courbe mais il y a d'autres qui sont loin.

Plus la valeur de R^2 est proche de 1 (cas idéal) et plus les valeurs prédites et observées sont corrélées. R^2= 0.835=83.5% signifie que 83.5% des variations de la variable dépendante sont expliqués par le modèle de régression. Dons il indique que ce modèle peut être appliqué avec succès pour prédire la biodisponibilité des médicaments.

III.4. CONCLUSION

Dans ce chapitre nous avons abordé le problème de prédiction de la biodisponibilité des médicaments à partir de leurs structures chimiques, en utilisant des programmes PaDel-Descriptor et SPSS élaborés dans le cadre de ce travail.

Nous avons montré dans ce travail qu'il est possible de relier la biodisponibilité aux descripteurs moléculaires par une relation de régression. Les valeurs de **R = 0.916≈ 1 et R^2=0.839** nous permet d'indiquer fermement la corrélation entre les différents descripteurs utilisés et la biodisponibilité.

On a trouvé un coefficient de corrélation ajusté **R^2 ajusté=0.829**, il est proche de 1, donc l'équation de régression estimée correspond ou explique mieux la relation entre *X* et *Y*.

Les valeurs prédites et les valeurs expérimentales sont proches et corrélés ce qui montre la validité de notre travail.

Conclusion générale

Ce travail est une étude de la relation entre les structure et l'activité d'une série des médicaments basé sur la méthode QSAR, consiste à trouver une corrélation entre une activité biologique : la biodisponibilité mesurée pour un panel des composés et certains descripteurs moléculaires.

On a développé un modèle de régression multilinéaire avec huit descripteurs pour l'estimation de la biodisponibilité, ces descripteurs sont de quatre classes : topologiques, d'autocorrélation, constitutionnel et immobilière.

Les valeurs de **R = 0.916 et R^2==0.839** nous permet d'indiquer fermement la corrélation entre les différents descripteurs utilisés et la biodisponibilité.

On a trouvé un coefficient de corrélation ajusté **R^2 ajusté=0.829**, il est proche de 1, donc l'équation de régression estimée correspond ou explique mieux la relation entre *X* et *Y*.

Enfin, le modèle proposé pourrait fournir des informations utiles sur la biodisponibilité d'un médicament candidat potentiel avant sa synthèse.

Les perspectives de ce travail nous semblent diverses. D'une part, nous avons l'intention de reprendre les mêmes bases de données et élaborer des modèles en utilisant d'autres méthodes d'analyse de données telle que PLS, GA, RNA,...D'autre part, nous projetons d'élaborer des modèles QSAR pour d'autres bases de données de molécules.

Références bibliographiques

[1] Journal officiel de la santé public, Article L. 511-1 du Code de la Santé Publique (C.S.P) (ordonnance n^0 67-827 du 23 septembre 1967).

[2] D.Stora, Pharmacologie et thérapeutiques, éditeur Lamarre collection Étudiants IFSI, *(2013).*

[3] P. L .Toutain, A. Bousquet-Melou, La biodisponibilité et son évaluation, (2012).

[4] F. Pebret, Dictionnaire de pharmacologie générale. Heures de France, ISBN 10: 2853852598 (2004).

[5] Y.Nafti, « Contribution à l'étude de la cinétique de libération d'un principe actif », Mémoire De Fin D'Étude En Vue De l'Obtention Du Diplôme D'Ingénieur D'État En Biologie**,** sous la direction de M^{elle} Boucherit N, Djelfa, Université ZIANE Achour, (2008).

[6] P. Lechat. Pharmacologie - Service de pharmacologie Clinique, Université Pierre et Marie Curie, (2006).

[7] J. Dangoumau, Pharmacologie générale, Université Victor Segalen-Bordeaux 2, (2006).

[8] http://www.aquaportail.com/definition-4174-biodisponibilite.html. (Vue le 05-02-2016)

[9] http://www.santepratique.fr/medicament.php (Vue le 06-02-2016)

[10] L.Patrick, Chimie pharmaceutique, Traduction de la 2^e édition anglaise par P.Depovere, Université catholique de Louvain, ISBN 2-7445-0154-9.

[11] fr.wikipedia.org/wiki/Pharmacodynamique(Vue le 10-02-2016)

[12] H.-D. Höltje, G. Folkeis, Molecular Modeling: Basic Principles and Applications. VCH, New-York(1997).

[13] N .Melkami, « Etude des corrélations entre les paramètres structuraux et les propriétés bioactives de quelques composés organiques », Mémoire De Fin D'Étude En vue de l'obtention du diplôme de Doctorat En Sciences, Biskra, Université Mohamed Khider, (2011).

[14] K. Bellifa, « Etude des relations quantitatives structure–toxicité des composés chimiques à l'aide des descripteurs moléculaires », Mémoire De Fin D'Étude En Vue De l'Obtention Du Diplôme de Doctorat en Chimie Physique, Université Abou Bekr Belkaid de Tlemcen, (2010).

[15] A. Crum Brown, T.R. Fraser, on the connection between chemical constitution And physiological action. Part 1. On the physiological action of salts of the Ammonium bases, derived from strychnia, brucia, thebia, codeia, morphia and nicotia.Trans. Roy.Soc. Edinburgh, (1868).

[16] L.P. Hammett, Journal of the American Chemical Society, (1937)

[17] C. Hansch, A. Leo, R.W. Taft, the Journal of Physical Chemistry *C*, (1991).

[18] C. Hansch, T. Fujita, a Method for the Correlation of Biological Activity and Chemical Structure. J. Am. Chem. Soc, (1964).

[19] S.M. Free, J.W. Wilson, J. Med. A Mathematical Contribution to Structure-Activity Studies; 7:395–399. Doi: 10.1021/jm00334a001, (1964).

[20] M. Grover, B. Singh, M. Bakshi, S. Singh, Pharmaceutical science & technology today, (2000)

[21] M. Grover, B. Singh, M. Bakshi, S. Singh, Pharmaceutical science & technology today, (2000)

[22] A.R. Katritzky, D.C. Fara, R.O. Petrukhin, D.B. Tatham, U.Maran, A. Lomaka, M. Karelson, Curr. Top. Med. Che., (2002)

[23] H.T. Phuong, « Synthèse et étude des relations structure/activité quantitatives (QSAR/2D) d'analogues Benzo[c]phénanthridiniques », Mémoire De Fin D'Étude En Vue De l'Obtention Du Diplôme de Doctorat en Chimie organique et thérapeutique, Université d'Angers, (2007).

[24] G. Fayet, « Développement de modèles QSPR Pour la prédiction des propriétés d'explosibilité des composés nitroaramatiques » Mémoire De Fin D'Étude En Vue De l'Obtention Du Diplôme de Doctorat en Ecole Doctorale Chimie Physique et Analytique de Paris VI, Université pierre et marie curie, 2010.

[25] S.W. Benson, J.H. Buss, Journal of ChemicalPhysics, (1958).

[26] M. Karelson, Molecular Descriptors in *QSAR/QSPR*, Wiley, New York, (2000).

[27] H. Wiener, Structural determination of paraffin boiling points. J. Am. Chem. Soc., (1947).

[28] M. Randic, Characterization of molecular branching. J. Am. Chem. Soc. 97, 6609–6615, J. Am. Chem. Soc., (1975).

[29] L.B. Kier, L.H. Hall, W.J. Murray, Molecular connectivity I: Relationship tononspecific local anesthesia. J. Pharm. Sci., (1975).

[30] A.T. Balaban, Highly Discriminating Distance-Based Topological Index, Chem. Phys. Lett., (1982).

[31] T.Opera, Property distribution of drug-related chemical databases. J Comput Id Mol Des, (2000).

[32] G.Richton, Nonleadlikness in leadlikness in biochemical screening Drug Discov Today, (2003).

[33] J.Serdler, S.Mc Govern, T.Danan, B.Shaichet, Identification and prediction of promiscuous aggregating inhibitors among known drug.J Med Chem, (2003).

[34] O. Monge, Création et utilisation de chémiothèques optimisées pour la recherche in silico de nouveaux composés bioactifs, réalisé à Orléans, (2006).

[35] O.Sperandio, MA.Miveta, BO.Villoutriex, Combining ligand-and structure based methods in drug design projects. Curr Comput Aid Drug Design, (2008).

[36] S.Rezazi, « Modélisation de la prédiction de l'effet thérapeutique à partir de la structure moléculaire », Mémoire De Fin D'Étude En Vue De l'Obtention Du Diplôme de Magister en génie des procédés, Université Yahia Farès de Médèa, (2011).

[37] A. Fortuné, « Techniques de Modélisation Moléculaire Appliquées à l'Etude et à l'Optimisation de Molécules Immunogènes et de Modulateurs de la Chimiorésistance », Mémoire De Fin D'Étude En Vue De l'Obtention Du Diplôme de doctorat en Sciences du Médicament, Université Joseph Fourier – Grenoble I, (2006).

[38] W.J. Dunn, III, Quantitative structure-activity relationship (QSAR). Chemometrics and Intelligent Laboratory Systems, (1989).

[39] I. Charif, « Elaboration des corrélations quantitatives structure-activité des acides carbonés. Etude théorique des effets de solvants sur les équilibres libres cétoénoliques des composés b-dicarbonyles cycliques », Mémoire De Fin D'Étude En Vue De l'Obtention Du Diplôme de doctorat en Chimie Physique, Université de Tlemcen, (2012).

[40] D.Bonchev, D.H. Rouvray: Chemical Graph Theory: Introduction and Fundamentals. Gordon and Breach Science Publishers, (1990), ISBN 0-85626-454-7.

[41] fr.wikipedia.org/wiki/Relation_quantitative_structure_à_activit (Vue le 12-03-2016)

[42] Joseph V. Turner, Beverly D. Glass, Snezana Agatonovic-Kustrin, Prediction of drug bioavailability based on molecular structure, (2003).

[43] J Comput Chem. (2011 May); 32(7):1466-74. Doi : 10.1002/jcc.21707. Epub 2010

[44] http://www.csvviewer.com/, (Vue le 02-04 2016).

[45] https://fr.wikipedia.org/wiki/SPSS, (Vue le 04-04-2016).

[46] Todeschini, R. and Consonni, V. (2009). Molecular descriptors for chemoinformatics, (Weinheim: Wiley VCH) pg 574-579, pg 395-402.

[47] Todeschini, R. and Consonni, V. (2009). Molecular descriptors for chemoinformatics, (Weinheim: Wiley VCH) pg 714-726.

[48] Todeschini, R. and Consonni, V. (2009). Molecular descriptors for chemoinformatics, (Weinheim: Wiley VCH) pg 27-37.

[49] Hall, L. H., and Kier, L. B. Electrotopological state indices for atom types: A novel combination of electronic, topological, and valence state information. (1995)

Annexe 1

Coefficients [a]

Modèle	Coefficients non standardisés		Coefficients standardisés	T	Sig.	Corrélations			Statistiques de colinéarité	
	A	Erreur standard	Bêta			Corrélation simple	Partielle	Partie	Tolérance	VIF
(Constante)	-,007	,011		-,638	,525					
MPC8	,000	,000	,430	6,335	,000	,094	,499	,231	,289	3,465
nAtomLAC	,010	,002	,200	4,240	,000	,133	,360	,155	,597	1,676
VR1_Dzs	1,419E-7	,000	1,635	21,078	,000	,489	,887	,768	,221	4,527
VR1_Dt	-1,412E-7	,000	-1,380	-18,350	,000	,024	-,858	-,669	,235	4,253
AATSC6c	3,473	,790	,209	4,397	,000	,303	,371	,160	,586	1,706
ATS6e	-6,018E-5	,000	-,309	-4,297	,000	,001	-,364	-,157	,258	3,882
AATSC7c	2,711	1,098	,122	2,469	,015	-,390	,219	,090	,542	1,844
maxwHBa	,011	,005	,082	2,065	,041	-,105	,185	,075	,845	1,184

a. Variable dépendante : Inverse de la biodisponibilité (1/F)

Annexe 2

Récapitulatif des modèles [ae]

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation	Changement dans les statistiques				
					Variation de R-deux	Variation de F	ddl1	ddl2	Sig. Variation de F
1	,537[a]	,289	,283	,083383161941	,289	51,973	1	128	,000
2	,672[b]	,451	,443	,073525801231	,163	37,622	1	127	,000
3	,724[c]	,525	,513	,068696636187	,073	19,483	1	126	,000
4	,752[d]	,565	,551	,065998371640	,040	11,513	1	125	,001
5	,778[e]	,605	,589	,063165937012	,040	12,462	1	124	,001
6	,792[f]	,627	,609	,061578999982	,023	7,473	1	123	,007
7	,801[g]	,642	,622	,060571110212	,015	5,127	1	122	,025
8	,812[h]	,660	,637	,059332586477	,017	6,146	1	121	,015
9	,827[i]	,684	,660	,057437943916	,024	9,114	1	120	,003
10	,835[j]	,697	,672	,056406574004	,014	5,428	1	119	,021
11	,854[k]	,729	,704	,053604171047	,032	13,768	1	118	,000
12	,861[l]	,741	,714	,052642723054	,012	5,350	1	117	,022
13	,868[m]	,753	,725	,051640088795	,012	5,587	1	116	,020
14	,874[n]	,764	,735	,050711899115	,011	5,285	1	115	,023
15	,872[o]	,761	,734	,050827953192	-,003	1,532	1	115	,218
16	,870[p]	,757	,732	,050944488541	-,003	1,537	1	116	,218
17	,906[q]	,821	,801	,043981543211	,063	40,978	1	116	,000
18	,906[r]	,821	,802	,043796887117	,000	,020	1	116	,889
19	,906[s]	,820	,804	,043637464002	,000	,143	1	117	,706
20	,905[t]	,820	,805	,043528971193	-,001	,409	1	118	,524
21	,905[u]	,818	,805	,043515968221	-,001	,928	1	119	,337
22	,904[v]	,816	,804	,043568310129	-,002	1,291	1	120	,258
23	,901[w]	,812	,802	,043871309794	-,004	2,703	1	121	,103
24	,900[x]	,809	,800	,044049313172	-,003	2,000	1	122	,160
25	,907[y]	,822	,812	,042675699095	,013	9,045	1	122	,003
26	,905[z]	,820	,811	,042846356912	-,003	1,986	1	122	,161
27	,910[aa]	,828	,818	,042057150502	,008	5,660	1	122	,019
28	,909[ab]	,826	,817	,042097095299	-,002	1,234	1	122	,269
29	,913[ac]	,834	,824	,041320563694	,008	5,666	1	122	,019
30	,916[ad]	,839	,829	,040778383155	,006	4,266	1	121	,041

a. Valeurs prédites : (constantes), SHBint3

b. Valeurs prédites : (constantes), SHBint3, VE2_Dzm

c. Valeurs prédites : (constantes), SHBint3, VE2_Dzm, VE2_Dze

d. Valeurs prédites : (constantes), SHBint3, VE2_Dzm, VE2_Dze, AATS8s

e. Valeurs prédites : (constantes), SHBint3, VE2_Dzm, VE2_Dze, AATS8s, MPC8

f. Valeurs prédites : (constantes), SHBint3, VE2_Dzm, VE2_Dze, AATS8s, MPC8, AATSC4s

g. Valeurs prédites : (constantes), SHBint3, VE2_Dzm, VE2_Dze, AATS8s, MPC8, AATSC4s, GATS4p

h. Valeurs prédites : (constantes), SHBint3, VE2_Dzm, VE2_Dze, AATS8s, MPC8, AATSC4s, GATS4p, nAtomLAC

i. Valeurs prédites : (constantes), SHBint3, VE2_Dzm, VE2_Dze, AATS8s, MPC8, AATSC4s, GATS4p, nAtomLAC, maxsCH3

j. Valeurs prédites : (constantes), SHBint3, VE2_Dzm, VE2_Dze, AATS8s, MPC8, AATSC4s, GATS4p, nAtomLAC, maxsCH3, VR3_Dt

k. Valeurs prédites : (constantes), SHBint3, VE2_Dzm, VE2_Dze, AATS8s, MPC8, AATSC4s, GATS4p, nAtomLAC, maxsCH3, VR3_Dt, VR3_Dzs

l. Valeurs prédites : (constantes), SHBint3, VE2_Dzm, VE2_Dze, AATS8s, MPC8, AATSC4s, GATS4p, nAtomLAC, maxsCH3, VR3_Dt, VR3_Dzs, SsssCH

m. Valeurs prédites : (constantes), SHBint3, VE2_Dzm, VE2_Dze, AATS8s, MPC8, AATSC4s, GATS4p, nAtomLAC, maxsCH3, VR3_Dt, VR3_Dzs, SsssCH, MLFER_BH

n. Valeurs prédites : (constantes), SHBint3, VE2_Dzm, VE2_Dze, AATS8s, MPC8, AATSC4s, GATS4p, nAtomLAC, maxsCH3, VR3_Dt, VR3_Dzs, SsssCH, MLFER_BH, VR1_Dzs

o. Valeurs prédites : (constantes), SHBint3, VE2_Dzm, VE2_Dze, AATS8s, MPC8, GATS4p, nAtomLAC, maxsCH3, VR3_Dt, VR3_Dzs, SsssCH, MLFER_BH, VR1_Dzs

p. Valeurs prédites : (constantes), SHBint3, VE2_Dzm, VE2_Dze, AATS8s, MPC8, GATS4p, nAtomLAC, VR3_Dt, VR3_Dzs, SsssCH, MLFER_BH, VR1_Dzs

q. Valeurs prédites : (constantes), SHBint3, VE2_Dzm, VE2_Dze, AATS8s, MPC8, GATS4p, nAtomLAC, VR3_Dt, VR3_Dzs, SsssCH, MLFER_BH, VR1_Dzs, VR1_Dt

r. Valeurs prédites : (constantes), SHBint3, VE2_Dzm, VE2_Dze, AATS8s, MPC8, nAtomLAC, VR3_Dt, VR3_Dzs, SsssCH, MLFER_BH, VR1_Dzs, VR1_Dt

s. Valeurs prédites : (constantes), SHBint3, VE2_Dzm, VE2_Dze, AATS8s, MPC8, nAtomLAC, VR3_Dt, SsssCH, MLFER_BH, VR1_Dzs, VR1_Dt

t. Valeurs prédites : (constantes), SHBint3, VE2_Dzm, AATS8s, MPC8, nAtomLAC, VR3_Dt, SsssCH, MLFER_BH, VR1_Dzs, VR1_Dt

u. Valeurs prédites : (constantes), SHBint3, AATS8s, MPC8, nAtomLAC, VR3_Dt, SsssCH, MLFER_BH, VR1_Dzs, VR1_Dt

v. Valeurs prédites : (constantes), SHBint3, MPC8, nAtomLAC, VR3_Dt, SsssCH, MLFER_BH, VR1_Dzs, VR1_Dt

w. Valeurs prédites : (constantes), SHBint3, MPC8, nAtomLAC, VR3_Dt, MLFER_BH, VR1_Dzs, VR1_Dt

x. Valeurs prédites : (constantes), SHBint3, MPC8, nAtomLAC, VR3_Dt, VR1_Dzs, VR1_Dt

y. Valeurs prédites : (constantes), SHBint3, MPC8, nAtomLAC, VR3_Dt, VR1_Dzs, VR1_Dt, AATSC6c

z. Valeurs prédites : (constantes), MPC8, nAtomLAC, VR3_Dt, VR1_Dzs, VR1_Dt, AATSC6c

aa. Valeurs prédites : (constantes), MPC8, nAtomLAC, VR3_Dt, VR1_Dzs, VR1_Dt, AATSC6c, ATS6e

ab. Valeurs prédites : (constantes), MPC8, nAtomLAC, VR1_Dzs, VR1_Dt, AATSC6c, ATS6e

ac. Valeurs prédites : (constantes), MPC8, nAtomLAC, VR1_Dzs, VR1_Dt, AATSC6c, ATS6e, AATSC7c

ad. Valeurs prédites : (constantes), MPC8, nAtomLAC, VR1_Dzs, VR1_Dt, AATSC6c, ATS6e, AATSC7c, maxwHBa

ae. Variable dépendante : Inverse de la biodisponibilité (1/F)

Annexe 3

ANOVA [ae]

Modèle		Somme des carrés	ddl	Moyenne des carrés	D	Sig.
1	Régression	,361	1	,361	51,973	,000[a]
	Résidu	,890	128	,007		
	Total	1,251	129			
2	Régression	,565	2	,282	52,232	,000[b]
	Résidu	,687	127	,005		
	Total	1,251	129			
3	Régression	,657	3	,219	46,384	,000[c]
	Résidu	,595	126	,005		
	Total	1,251	129			
4	Régression	,707	4	,177	40,569	,000[d]
	Résidu	,544	125	,004		
	Total	1,251	129			
5	Régression	,757	5	,151	37,923	,000[e]
	Résidu	,495	124	,004		
	Total	1,251	129			
6	Régression	,785	6	,131	34,498	,000[f]
	Résidu	,466	123	,004		
	Total	1,251	129			
7	Régression	,804	7	,115	31,294	,000[g]
	Résidu	,448	122	,004		
	Total	1,251	129			
8	Régression	,825	8	,103	29,306	,000[h]
	Résidu	,426	121	,004		
	Total	1,251	129			
9	Régression	,855	9	,095	28,809	,000[i]
	Résidu	,396	120	,003		
	Total	1,251	129			
10	Régression	,873	10	,087	27,428	,000[j]
	Résidu	,379	119	,003		
	Total	1,251	129			
11	Régression	,912	11	,083	28,862	,000[k]
	Résidu	,339	118	,003		
	Total	1,251	129			
12	Régression	,927	12	,077	27,878	,000[l]
	Résidu	,324	117	,003		
	Total	1,251	129			

13	Régression	,942	13	,072	27,172	,000[m]
	Résidu	,309	116	,003		
	Total	1,251	129			
14	Régression	,956	14	,068	26,541	,000[n]
	Résidu	,296	115	,003		
	Total	1,251	129			
15	Régression	,952	13	,073	28,335	,000[o]
	Résidu	,300	116	,003		
	Total	1,251	129			
16	Régression	,948	12	,079	30,428	,000[p]
	Résidu	,304	117	,003		
	Total	1,251	129			
17	Régression	1,027	13	,079	40,837	,000[q]
	Résidu	,224	116	,002		
	Total	1,251	129			
18	Régression	1,027	12	,086	44,612	,000[r]
	Résidu	,224	117	,002		
	Total	1,251	129			
19	Régression	1,027	11	,093	49,011	,000[s]
	Résidu	,225	118	,002		
	Total	1,251	129			
20	Régression	1,026	10	,103	54,140	,000[t]
	Résidu	,225	119	,002		
	Total	1,251	129			
21	Régression	1,024	9	,114	60,088	,000[u]
	Résidu	,227	120	,002		
	Total	1,251	129			
22	Régression	1,022	8	,128	67,276	,000[v]
	Résidu	,230	121	,002		
	Total	1,251	129			
23	Régression	1,016	7	,145	75,448	,000[w]
	Résidu	,235	122	,002		
	Total	1,251	129			
24	Régression	1,013	6	,169	86,982	,000[x]
	Résidu	,239	123	,002		
	Total	1,251	129			
25	Régression	1,029	7	,147	80,725	,000[y]
	Résidu	,222	122	,002		
	Total	1,251	129			
26	Régression	1,026	6	,171	93,102	,000[z]

	Résidu	,226	123	,002		
	Total	1,251	129			
27	Régression	1,036	7	,148	83,633	,000[aa]
	Résidu	,216	122	,002		
	Total	1,251	129			
28	Régression	1,033	6	,172	97,182	,000[ab]
	Résidu	,218	123	,002		
	Total	1,251	129			
29	Régression	1,043	7	,149	87,268	,000[ac]
	Résidu	,208	122	,002		
	Total	1,251	129			
30	Régression	1,050	8	,131	78,937	,000[ad]
	Résidu	,201	121	,002		
	Total	1,251	129			