

الجمهورية الجزائرية الديمقراطية الشعبية

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET
POPULAIRE

وزارة التعليم العالي والبحث العلمي

MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA
RECHERCHE SCIENTIFIQUE

جامعة صالح بونبندر قسنطينة 3

UNIVERSITE SALAH BOUBNIDER CONSTANTINE 3

كلية الطب - قسم الصيدلة

FACULTE DE MEDECINE - DEPARTEMENT DE PHARMACIE



Mémoire de fin d'études en vue de l'obtention du diplôme de

DOCTEUR EN PHARMACIE

**Développement de modèles QSPR/QSAR pour la prédiction de
propriétés physicochimiques et biologiques des substances
organiques**

Encadré par :

Pr. LALAOUNA Abd El Djalil

Présenté par :

- Baghi Mohammed El Hadi
- Kaidouchi Safwan
- Mansouri Houssef Eddine

ANNEE UNIVERSITAIRE : 2022 - 2023

Sommaire

Remerciements

Dédicaces

Sommaire

Liste des figures	I
Liste des tableaux	VII
Liste des annexes	XIII
Liste des abréviations et acronymes	XIV
Liste des définitions	XVII
Introduction générale	1
Contexte de l'étude	3

Partie 1 . Etude bibliographique

Chapitre I .Médicaments : définition, développement et classification

1.Introduction	8
2. Définitions	8
3. Différentes phases de recherche et développement d'un médicament	10
3.1. Recherche d'un médicament	10
3.2. Essais précliniques	11
3.2.A. Objectif des essais précliniques	11
3.2.B. Etapes des essais précliniques	11
3.2.C. Réglementations des essais précliniques	12
3.3. Essais cliniques	12
3.3.A. Etapes des essais cliniques	13
3.3.B. Réglementations des essais cliniques	14
3.4. Autorisation de la mise en marche	14
3.4.A. Etapes d'AMM	14
3.5. Pharmacovigilance	15
3.5.A. Importance de la pharmacovigilance	15

3.5.B. Imputabilité intrinsèque	16
3.5.C. Imputabilité extrinsèque	16
4. Principales classes médicamenteuses	21
4.1. Médicaments analgésiques	21
4.2. Médicaments anti-infectieux	21
4.3. Médicaments du système cardiovasculaire	22
4.4. Médicaments neurologiques et neuropsychiatriques	23
4.5. Médicaments pour les troubles sanguins	23
4.6. Médicaments pour les troubles broncho-pulmonaires	24
4.7. Médicaments anti-inflammatoires	24
4.8. Médicaments pour les troubles gastro-intestinaux	25
4.9. Médicaments pour les troubles endocriniens	25
4.10. Médicaments pour les troubles métaboliques	26
4.11. Médicaments ophtalmiques	27
4.12. Médicaments ORL (Oto-Rhino-Laryngologie)	28
4.13. Médicaments pour les troubles génito-urinaires	28
4.14. Médicaments pour les troubles génico-obstétriques	29
Chapitre II . Propriétés physico-chimiques, biologiques et caractéristiques structurales des molécules organiques	
1. Introduction	31
2. Propriétés physico-chimiques des molécules organiques	31
2.1. Descripteurs 2D	32
2.1.A. Notion et terminologie	32
2.1.B. Propriétés physiques	33
2.1.C. Description de la théorie Huckel	35
2.1.D. Surfaces subdivisées	37
2.1.E. Nombre d'atomes et nombre de liaisons	39
2.1.F. Indices de connectivité et de forme Kier & Hall	41
2.1.G. Descripteurs des matrices d'adjacence et de distance	43

2.1.H. Descripteurs des caractéristiques pharmacophore	45
2.1.J. Descripteurs de charges partielles	47
2.2. Descripteurs 3D	51
2.2.A. Descripteurs d'énergie partielle	51
2.2.B. Descripteurs de la MOPAC	54
2.2.C. Descripteurs de forme/surface/volume	56
2.2.D. Descripteurs de charges dépendant de la conformation	58
3. Propriétés biologiques des molécules organiques	59
3.1. Absorption	60
3.1.A. Perméabilité Caco-2	60
3.1.B. Absorption intestinale (humaine)	61
3.1.C. Solubilité dans l'eau	61
3.1.D. Substrats de glycoprotéine P	61
3.1.E. Inhibiteurs de la P-glycoprotéine I et II	61
3.1.F. Perméabilité cutanée	62
3.2. Distribution	62
3.2.A. Volume de distribution (humain)	62
3.2.B. Perméabilité de barrière hémato-encéphalique	63
3.2.C. Fraction non liée	64
3.2.D. Perméabilité du système nerveux central	64
3.3. Métabolisme	64
3.3.A. Inhibiteurs de cytochrome P450	65
3.3.B. Substrats de CYP 2D/CYP3A4	65
3.4. Excrétion	65
3.4.A. Clearance totale	66
3.4.B. Substrats rénaux pour OCT2 (Organic Cation Transporter 2)	66
3.5. Toxicité	67
3.5.A. Dose maximale tolérée	67

3.5.B. DL50 chez le rat	68
3.5.C. Test d'Ames	68
3.5.D. Toxicité chez T. pyriformis	68
3.5.E. Toxicité du vairon	69
3.5.F. Toxicité hépatique	69
3.5.G. Toxicité orale chronique chez le rat	69
3.5.H. Sensibilisation cutanée	69
3.5.I. Inhibiteurs HERG I et inhibiteurs HERG II	70
CHAPITRE III. Modèles QSPR/QSAR	
1. Introduction	72
2. Historique	72
3. Développement des modèles QSPR/QSAR	74
4. Etapes de développement des modèles QSPR/QSAR	75
4.1 Collecte et prétraitement des données	75
4.2 Partitionnement de l'ensemble de données	78
4.3. Sélection des descripteurs	79
4.3.A. Méthode ascendante	79
4.3.B. Méthode descendante	80
4.3.C. Méthode pas à pas	80
4.3.D. Algorithme génétique	80
5. Méthodes chimiométriques	81
5.1. Régression linéaire multiple	82
5.2. Réseaux de neurones artificiels	83
6. Evaluation et validation des modèles chimiométriques	85
6.1. Validation des modèles chimiométriques	85
6.4.A. Validation interne	85
6.4.B. Validation externe	86
6.2. Evaluation des performances des modèles chimiométriques	87

6.2.A. Racine de l'Erreur Quadratique Moyenne (REQM) (ou RMSE pour Root Mean Square Error)	87
6.2.B. Coefficient de détermination (r^2) et coefficient de détermination ajusté (r^2 ajusté)	87
Partie 2 . Partie pratique	
1. Logiciels utilisés	91
2. Collecte des données	92
3. Présentation des individus	92
4. Prétraitement des données	93
4.1.Méthode	93
4.2.Résultats	94
4.3.Discussion et conclusion	97
5. Selection des propriétés physiochimique et biologique à étudier	97
6. Sélection des individus	98
6.1.Méthode	98
6.2.Résultats	98
6.3.Discussion et conclusion	100
7. Modélisation prédictive des données par régression linéaire multiple : une analyse préliminaire	101
7.1.Méthode	101
7.2.Résultats	101
7.3.Discussion et conclusion	103
8. Conception et optimisation des modèles prédictifs par régression linéaire multiple	
8.1. Modèle MLR pour le calcul de "LogP"	105
8.1.1. Méthode	105
8.1.2. Résultats	105
8.1.3. Discussion et conclusion	113
8.2. Modèle MLR pour le calcul de "h_LogP"	114
8.2.1. Méthode	114
8.2.2. Résultats	114

8.2.3. Discussion et conclusion	119
8.3. Modèle MLR pour le calcul de "h_LogS"	120
8.3.1. Méthode	120
8.3.2. Résultats	120
8.3.3. Discussion et conclusion	127
8.4. Modèle MLR pour le calcul de "LogS"	128
8.4.1. Méthode	128
8.4.2. Résultats	128
8.4.3. Discussion et conclusion	134
8.5. Modèle MLR pour le calcul de "h_mr"	134
8.5.1. Méthode	134
8.5.2. Résultats	134
8.5.3. Discussion et conclusion	142
8.6. Modèle MLR pour le calcul de "mr"	142
8.6.1. Méthode	142
8.6.2. Résultats	142
8.6.3. Discussion et conclusion	151
8.7. Modèle MLR pour le calcul de "TPSA"	151
8.7.1. Méthode	151
8.7.2. Résultats	151
8.7.3. Discussion et conclusion	156
8.8. Modèle MLR pour le calcul du "BBB permiability"	156
8.8.1. Méthode	156
8.8.2. Résultats	156
8.8.3. Discussion et conclusion	166
8.9. Modèle MLR pour le calcul de "CNS permiability"	166
8.9.1. Méthode	166
8.9.2. Résultats	167

8.9.3. Discussion et conclusion	171
8.10. Modèle MLR pour le calcul "d'Intestinal absorption"	172
8.10.1. Méthode	172
8.10.2. Résultats	172
8.10.3. Discussion et conclusion	176
8.11. Modèle MLR pour le calcul de "Caco2 permeability"	176
8.11.1. Méthode	176
8.11.2. Résultats	176
8.11.3. Discussion et conclusion	179
8.12. Modèle MLR pour le calcul de " Oral rat chronic toxicity (LOAEL) "	180
8.12.1. Méthode	180
8.12.2. Résultats	180
8.12.3. Discussion et conclusion	183
8.13. Modèle MLR pour le calcul du "Minnow Toxicity"	183
8.13.1. Méthode	183
8.13.2. Résultats	183
8.13.3. Discussion et conclusion	187
9. Conception et optimisation des modèles prédictifs par réseaux de neurones artificiels	
9.1. Modèle ANN pour le calcul de "LogP"	188
9.1.1. Méthode	188
9.1.2. Résultats	188
9.1.3. Discussion et conclusion	191
9.2. Modèle MLR pour le calcul du "H_LogP"	192
9.2.1. Méthode	192
9.2.2. Résultats	192
9.2.3. Discussion et conclusion	195
9.3. Modèle MLR pour le calcul du "H_LogS"	196
9.3.1. Méthode	196

9.3.2. Résultats	196
9.3.3. Discussion et conclusion	199
9.4. Modèle MLR pour le calcul de "LogS"	200
9.4.1. Méthode	200
9.4.2. Résultats	200
9.4.3. Discussion et conclusion	203
9.5. Modèle MLR pour le calcul du "H_Mr"	204
9.5.1. Méthode	204
9.5.2. Résultats	204
9.5.3. Discussion et conclusion	207
9.6. Modèle MLR pour le calcul de "Mr"	208
9.6.1. Méthode	208
9.6.2. Résultats	208
9.6.3. Discussion et conclusion	211
9.7. Modèle MLR pour le calcul du "TPSA"	212
9.7.1. Méthode	212
9.7.2. Résultats	212
9.7.3. Discussion et conclusion	215
9.8. Modèle MLR pour le calcul de "BBB perméabilité"	216
9.8.1. Méthode	216
9.8.2. Résultats	216
9.8.3. Discussion et conclusion	219
9.9. Modèle MLR pour le calcul de "CNS perméabilité"	220
9.9.1. Méthode	220
9.9.2. Résultats	220
9.9.3. Discussion et conclusion	223
9.10. Modèle MLR pour le calcul "d'Intestinal absorption"	224
9.10.1. Méthode	224

9.10.2.	Résultats	224
9.10.3.	Discussion et conclusion	227
9.11.	Modèle MLR pour le calcul de "Caco2 permeability"	228
9.11.1.	Méthode	228
9.11.2.	Résultats	228
9.11.3.	Discussion et conclusion	231
9.12.	Modèle MLR pour le calcul de " Oral rat chronic toxicity (LOAEL) "	232
9.12.1.	Méthode	232
9.12.2.	Résultats	232
9.12.3.	Discussion et conclusion	235
9.13.	Modèle MLR pour le calcul de "Minnow Toxicity"	236
9.13.1.	Méthode	236
9.13.2.	Résultats	236
9.13.3.	Discussion et conclusion	239
10.	Discussion générale	240
11.	Conclusion générale	244
	Références bibliographiques	
	Annexes	
	Résumé	

ABSTRACT

The physicochemical and biological properties of organic compounds play a crucial role in various fields such as drug manufacturing, development of analytical methods, and toxicity assessment. Prediction models for structure-property relationships (QSPR) and structure-activity relationships (QSAR) offer a promising solution for rapidly and cost-effectively estimating these properties.

In this study, we developed QSPR and QSAR models using relevant chemical descriptors and chemometric techniques. The dataset included the structures of 472 organic molecules sourced from the PubChem and Mol-Instincts databases. 383 descriptors were calculated for each structure, with 354 obtained using the MOE software and 29 via the pkCSM server.

After preprocessing the data, we selected 264 descriptors, of which 222 were used as independent variables and 26 as responses. The dataset was divided using the Kennard Stone algorithm.

An initial modeling was performed using multiple linear regression (MLR) with simple descriptors that are easily accessible and straightforward to use. Various variable selection methods (forward, backward, stepwise, and genetic algorithm) were employed.

The chosen responses for the chemometric models were: $\log P(o/w)$, h_logP , $\log S$, h_logS , mr , h_mr , TPSA, Caco2 permeability, Intestinal absorption, BBB permeability, CNS permeability, Oral Rat Chronic Toxicity (LOAEL), and Minnow toxicity.

MLR yielded models with R^2 values ranging from 0.252 to 0.987. Among the 14 studied responses, 12 models achieved $R^2 \geq 0.6$.

To improve predictions, we also explored models based on artificial neural networks (ANN). The ANN models outperformed MLR significantly, with R^2 values ranging from 0.394 to 0.999. Among the 14 studied responses, 12 ANN models achieved $R^2 \geq 0.7$, and 9 models achieved $R^2 \geq 0.9$.

In conclusion, the results confirm the effectiveness of MLR and ANN models for accurate modeling and prediction of the studied organic molecule properties. These approaches offer promising prospects for rapid and cost-effective estimation of physicochemical and biological properties of organic compounds.

Keywords: QSPR, QSAR, chemometrics, organic molecules, Kennard Stone, multiple linear regression, artificial neural networks.