



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE



MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

Université Frères Mentouri Constantine 1
Faculté des Sciences de la Nature et de la Vie

1

Département de Biochimie et Biologie Cellulaire et Moléculaire
Mémoire présenté en vue de l'obtention du Diplôme de Master
Domaine : Science de la Nature et de la Vie
Filière : Science Biologique
Spécialité : Biochimie Appliquée

Thème

Etude des relations Quantitatives Structure-Activité
(QSAR) de plusieurs
familles de composés sur des données biologiques

Présenté et soutenu publiquement Par :

Le 09/07/2019

- **KHAINNAR Mehdi**
- **AIMEUR Oussama**

Jury d'évaluation :

Président du jury : BENSEGUENI A. (Professeur à UFM Constantine 1),
Examineur : MOKRANI E.H. (Maitre-assistant A à UFM Constantine 1),
Rapporteur : DEMS M.A.E. (Maître de recherche classe A au CRBt).
CO-ENCADREUR : BOUHEDJAR K (Chercheur permanent au CRBt)

Année universitaire

2018-2019

Remerciement

Notre premier remerciement à " ALLAH " le tout puissant de nous avoir offert la force pour accomplir ce travail "Dieu merci".

Nous tenons à remercier vivement notre encadreur Mr AbdEsselem DEMS maitre de recherche au centre de recherche en biotechnologie (C.R.Bt) et notre co-encadreur Khalid Bouhedjar, pour sa patience, sa disponibilité, sa gentillesse et surtout ses bons conseils afin d'élaborer ce mémoire dans de bonnes conditions. Nous lui exprimons toute notre reconnaissance ;

A toute l'équipe pédagogique de la faculté des sciences de la nature et de la vie ;

Aux membres du jury qui s'apprêtent à évaluer notre travail ;

A nos familles et amis qui nous ont tellement soutenus ;

A toutes les personnes qui ont contribué de près ou de loin à l'élaboration de ce mémoire ;

Nous exprimons nos forts remerciements

_ MERCI _

Résumé

Le travail présenté dans ce mémoire a pour objectif d'élaborer des modèles QSAR fiables, stables et prédictifs pour la prédiction des propriétés pIC₅₀ (concentration inhibitrice médiane) d'inhibiteurs de deux familles de molécules (AChE et COX2)

Le logiciel de modélisation moléculaire CHemDraw a été utilisé pour représenter les molécules, puis à l'aide de la méthode semi-empirique PM3 dans le logiciel HyperChem les géométries finales ont été obtenues. Différents descripteurs moléculaires sont calculés à l'aide du logiciel Dragon (V.5.5).

Des modèles QSAR ont été développés pour la prédiction des pIC₅₀ de divers ensembles de d'inhibiteurs. Des approches basées sur la régression linéaire multiple (RLM), conduisent à des modèles de qualités différentes. Les algorithmes génétiques (GA), ont été associés pour sélectionner les descripteurs les plus importants,

L'approche hybride algorithme génétique/ régression multilinéaire a été utilisée pour modéliser les pIC₅₀.

Les modèles établis, ont été validés selon les cinq principes avancés par l'Organisation de Coopération et de Développement Économiques (OCDE). Le domaine d'application des modèles est étudié à l'aide du diagramme de Williams pour détecter les composés aberrants en X et/ou en Y.

Des validations rigoureuses internes et externes ont été considérées pour juger la stabilité et la capacité prédictive de ces modèles afin de combler les lacunes dans les données des pIC₅₀ des molécules.

Mots clés : Acétylcholinestérase, Cyclooxygénase-2, QSAR/QSPR, MobyDigs, Descripteurs.

Abstract

The work presented in this thesis aims to develop reliable, stable and predictive QSAR models for the prediction of pIC₅₀ properties (median inhibitory concentration) of inhibitors of two families of molecules (AChE and COX2)

The ChemDraw molecular modeling software was used to represent the molecules, then using the semi-empirical PM3 method in the HyperChem software the final geometries were obtained. Different molecular descriptors are calculated using the Dragon software (V.5.5).

QSAR models have been developed for the prediction of pIC₅₀ of various sets of inhibitors. Multiple linear regression (MLR) approaches lead to different quality models. Genetic algorithms (GA), have been associated to select the most important descriptors,

The hybrid genetic algorithm / multilinear regression approach was used to model the pIC₅₀s.

The established models have been validated according to the five principles put forward by the Organization for Economic Cooperation and Development (OECD). The field of application of the models is studied using the Williams diagram to detect aberrant compounds in X and / or in Y.

Rigorous internal and external validations were considered to judge the stability and predictive ability of these models in order to fill the gaps in the pIC₅₀ data of the molecules.

Mots clés : Acétylcholinestérase, Cyclooxygénase-2, QSAR/QSPR, MobyDigs, Descriptors.



Sommaire



Sommaire

PARTIE I : Partie théorique

1	Introduction générale	1
2	Méthodologie QSAR	2
	Historique	2
	Principe	3
	La stratégie globale	3
	Models QSAR/QSPR	4
	Les descripteurs	5
	Descripteurs constitutionnels.....	5
	Descripteurs topologiques.....	6
	Descripteurs géométriques	8
	Descripteurs quantiques.....	8
	Descripteurs de réactivité issus de la DFT conceptuelle	9
3	Méthodes utilisées pour le développement de modèles QSAR/QSPR	9
	Introduction	9
	Méthodes de régressions linéaires et multilinéaire	10
	Evaluation préliminaire des données	11
	Régression linéaire multiple.....	11
4	Les Principes de l'OCDE	15
5	Choix de l'ensemble de calibrage et de validation	15
6	Test de randomisation	16
7	Les activités biologiques.....	17
	ACHE	17
	Introduction	17
	Acétylcholinestérase	17
	Role.....	17
	Dosage.....	17
	Intérêt de dosage.....	17
	Butyrylcholinestérase	18

	Les inhibiteurs de AchE	18
7.2	COX 2	18
7.2.1	La cyclooxygénase (COX)	18
7.2.2	COX 2	19
	Similitude et différence	19
	Fonctions physiologiques et pathologiques de COX 1 et COX 2	19

PARTIE II : Partie Pratique

1	Introduction	22
2	ChemDraw	24
	Interface utilisateur ChemDraw.....	24
	Barres d'outils.....	25
	La barre d'outils principale	25
	Enregistrement.....	25
3	HyperChem.....	26
	Optimisation géométrique	26
	Mécanique moléculaire MM+.....	26
	Semi Empirique PM3.....	28
4	DRAGON	29
	Calcul de descripteurs.....	29
5	MobyDigs	32
	Définition	32
	Méthode utilisée dans MobyDigs	32
6	Application de l'approche QSAR dans la modélisation de l'AchE	37
	Origine des données	38
7	Résultats et discussion	38
	Matrice de corrélation	38
	Equation et analyse de régression	38
	Analyse des résidus et diagnostics d'influence	39
	Qualité de l'ajustement	41
	Le domaine d'application	41
	Test de randomisation	42

Conclusion	43
8 Application de l'approche QSAR dans la modélisation de la Cox-2	45
Origine des données	45
9 Résultats et discussion	45
Matrice de corrélation	45
Equation et analyse de régression	45
Analyse des résidus et diagnostics d'influence	46
Qualité de l'ajustement	48
Le domaine d'application	49
Test de randomisation	49
Conclusion	50
Conclusion générale	51
Références bibliographiques	
ANNEXE	
Références de l'annexe	

A teal horizontal bar spans the top of the page, with a thin teal vertical line extending downwards from its left edge. The text is centered below these elements.

Liste des tableaux

Liste des tableaux

Tableau 1 matrice de corrélation pour le modèle 01	38
Tableau 2 valeurs essentielles pour le modèle 01	38
Tableau 3 définitions des descripteurs	38
Tableau 4 Valeurs des pCIC50, observées (piC50) et calculées, prédites (pred y), des résidus standardisés (R1. R2. R3) ainsi que des leviers (hat) (std.Err.Pred).	39
Tableau 5 matrice de corrélation pour le modèle 02	45
Tableau 6 les valeurs essentielles du modèle 02	45
Tableau 7 définition des descripteurs moléculaires	45
Tableau 8 Valeurs des pCIC50, observés (piC50) et calculés, prédites (pred y), des résidus standardisés (R1. R2. R3) ainsi que des leviers (hat) (std.Err.Pred).	46



Liste des figures



Listes des figures

Figure 1 Schémagénéral QSAR.....	4
Figure 2 Distances topologiques au sein de la molécule de nitrobenzène.....	7
Figure 3 Angle dièdre dirigeant l'activité catalytique des complexes bis (arylimino) pyridine du fer.....	7
Figure 4 Illustration de la méthode « Y-scrambling » (randomisation de Y).....	16
Figure 5 ChemDraw interface.....	24
Figure 6 barre d'outils principale.....	25
Figure 7 ajout d'hydrogènes.....	26
Figure 8 choix de mode d'optimisation.....	27
Figure 9 choix du mode mm+.....	27
Figure 10 lancement de l'optimisation.....	28
Figure 11 choix de méthode semi empirique.....	28
Figure 12 sélection du mode PM3.....	29
Figure 13 import du fichier et sélection des molécules.....	30
Figure 14 choix des descripteurs et lancement du calcul.....	31
Figure 15 fin de sauvegarde.....	31
Figure 16 importer le fichier.....	32
Figure 17 sélection de Y = pIC50.....	33
Figure 18 sélection de variable X = descripteurs.....	33
Figure 19 division des groupes.....	34
Figure 20 choix de méthode Q2 leave one out.....	34
Figure 21 lancement du calcul.....	35
Figure 22 template des molécules AchE.....	39
Figure 23 Droites d'ajustement pour le modèle 01.....	41
Figure 24: le domaine d'applications chimiques des composés.....	42
Figure 25 Test de randomisation du modèle AchE.....	43
Figure 26 template des molécules.....	46
Figure 27 Droites d'ajustement pour le modèle 02.....	48
Figure 28: le domaine d'applications chimique des composés.....	49
Figure 29 test de randomisation du modèle COX2.....	50

Symboles et abréviations

AG: Algorithme génétique

CAS : Chemical Abstracts Service

EQM: Ecart quadratique moyen.

EQMC: Ecart quadratique moyen calculé sur l'ensemble de calibrage.

EQMP Ecart quadratique moyen de prédiction.

EQMPext:: Ecart quadratique moyen calculé sur l'ensemble de validation externe.

e_i : Résidu : différence entre les valeurs observée (y_i) et estimée (\hat{y}_i).

e_{i_std} : Résidu standardisé.

F : Statistique de Fisher.

FIT: Fonction de KUBINYI.

FIV: Facteur d'inflation de la variance.

GA: Algorithme génétique (Genetic Algorithm).

HF : Hartree -Fock

h_{ii} : Eléments diagonaux de la matrice chapeau.

LMO: Validation croisée par omission d'un ensemble d'observations: Cross-validation by leave-many-out.

LOO: Validation croisée par omission d'une observation: Cross-validation by leaveone-out.

MM+: Mécanique Moléculaire (+).

MM2 Mécanique moléculaire 2

n: Dimension de la population (échantillon).

n-p : Nombre de degrés de liberté.

OCDE : Organisation de Coopération et de Développement Economiques.

PM3 : ParametrizationMethod 3.

PRESS : Somme des carrés des erreurs de prédiction.

p : Nombre de descripteurs en comptant la constante (Nombre de paramètres

Q2LOO: Coefficient de prédiction.

QSAR: Relations Quantitatives Structure/ Activité.

QSPR : Relations Quantitatives Structure/ Propriété.

Q²*Yscr* Coefficient de prédiction des modèles où les Y sont randomisés.

R2 : Coefficient de détermination.

REACH : en **R**egistrement, **E**valuation et **A**utorisation des produits **C**himiques.

RLM (MLR): Régression linéaire multiple.

RMSE: Racine de l'écart quadratique moyen (RootMeanSquaredError).

RNA: Réseaux de neurones artificiels.

R^2_{Yscr} Coefficient de détermination des modèles où les Y sont randomisés.

S : Erreur standard.

SCE : Somme des carrés des écarts.

SCT : Somme des carrés totale.

t : t de Student.

t_i : Résidu studentisé externe.

y_i : Valeur observée.

y_i : Valeur estimée.

$y(i)$: Valeur prédite.

tr : trining

ts : test



Introduction générale



1 Introduction générale

Le développement d'un nouveau médicament pour des maladies telles qu'Alzheimer nécessite le plus souvent des investissements considérables en plus leurs processus est long. Même si les dépenses consacrées à la découverte de médicaments ont été multipliées par plusieurs fois, le nombre de nouveaux médicaments approuvés reste considérablement réduit.

La conception de médicaments *in silico* est un nouveau paradigme qui a un impact significatif sur le processus global de découverte de médicaments. L'intégration de diverses approches de découverte de médicaments *in silico*, telles que les approches basées sur les réseaux, les ligands et les relations structure-activité.

Suite au développement de l'informatique et de l'existence sur le marché de logiciels professionnels adaptés peu coûteux et rapide, il y a une croissance exceptionnelle de l'intérêt pour les relations quantitatives structure-Activité (QSAR), qui utilisent des méthodes d'analyses multidimensionnelles afin de modéliser des Activité en fonction des paramètres structuraux moléculaires (appelés descripteurs).[1]

L'objectif de ce travail est donc de développer des voies d'accès à des analogues des inhibiteurs de cholinestérase (AChE) et COX-2 et les études de relations structure activité (QSAR) seront menées et leur résultat nous aidera à nous orienter vers les structures les plus prometteuses.

Le manuscrit est divisé en deux parties : une partie théorique et une partie expérimentale.

Dans la partie théorique, les chapitres présentent l'histoire de QSAR, son principe et sa stratégie globale pour développer un model QSAR, son application et les différentes étapes à respecter pour le réaliser, ainsi que la détermination des activités biologiques que nous avons choisies pour appliquer cette approche.

Les deux chapitres de la deuxième partie sont consacrés à la méthodologie de travail réalisé dans ce mémoire dans laquelle nous avons travaillé sur deux series d'inhibiteurs dont 41 molécules d'AChE et 64 molécules de COX 2, aux résultats obtenus et à leurs commentaires, et enfin l'établissement d'une conclusion.



Partie I

Partie théorique



2 Méthodologie QSAR

Historique

Il y a plus d'un siècle et demi, en 1863, des scientifiques ont observé que des propriétés physico-chimiques des alcanes augmentent quand le nombre d'atomes de carbone augmentent. Ils ont observé également une diminution de la solubilité dans l'eau des alcools avec l'augmentation du nombre d'atomes de carbone, cela est considéré depuis, comme la première formulation générale en QSAR.

Cinq ans après, en 1868, Crum-Brown et Fraser [2-3] postulèrent que « l'activité biologique d'une molécule est une fonction de sa constitution chimique » selon la forme suivante :

$$\emptyset = f(C) \quad (\text{eq.01})$$

Quelques décennies plus tard, en 1893, Richet [4] a découvert que la toxicité des composés organiques a un rapport avec leur solubilité dans l'eau est définie par :

$$\chi\emptyset = f(\chi C) \quad (\text{eq.02})$$

A la fin du 19ème siècle, Meyer en 1899 et Overton en 1901 [5-6], ont observé « une relation linéaire entre l'activité des narcotiques et leur coefficient de partage huile-eau ».

Six ans après, en 1907, Fühner et Neubauer [7] ont montré pour une série de narcotiques homologues, que l'activité augmente en fonction de la progression géométrique de la série de composés, cela montrait l'importance de la contribution d'additivité de groupements fonctionnels pour l'activité biologique.

En 1962, Hansen [8] a montré l'existence d'une corrélation entre la toxicité des acides benzoïques substitués et les constantes électroniques « σ » des substituants.

L'année 1964 est considérée comme le début des méthodes QSAR modernes. Hansch et Fujita [8] ont établi les premières corrélations entre les propriétés physico-chimiques (log P, pKa, Approches de base, développement, validation et application des méthodes QSAR/QSPR 21 paramètres stériques et électroniques) et l'activité biologique (activité enzymatique, pharmacologique).

L'approche QSAR résulte de la combinaison de différents paramètres physico-chimiques de façon linéaire additive (eq. 03) ; $\log (1/C)$ est le logarithme de l'inverse de la dose molaire qui produit ou empêche une certaine réponse biologique, $\log P$ est le logarithme du coefficient de partage de n-octanol/ eau). D'autres méthodes utilisent un paramètre connu sous le nom de paramètre de lipophilie calculée π (eq. 4), il est employé au lieu des valeurs

mesurées de $\log P$, et la formulation d'une équation parabolique pour la description quantitative non-linéaire des rapports lipophilie-activité (eq. 5)

$$\text{Log}1/C = a \log P + b\sigma + \dots + \text{const} \quad (\text{eq 03})$$

$$\pi^x = \log Pr_X - \log Pr_h \quad (\text{eq 04})$$

$$\text{Log } 1/C = a (\log P)^2 + b \log P + c \sigma + \dots + \text{const} \quad (\text{eq 05})$$

Sept ans plus tard, Hansch et Lien ont réalisé une étude QSAR sur différentes familles biologiques. Ils ont observé que quelque soit la famille utilisée, l'activité antifongique dépend du coefficient de partage Eau-Octanol, expérimental ou calculé [9].

Ces études ont été extrapolées aux techniques séparatives en corrélant les propriétés physico-chimiques des analyses avec les temps de rétention obtenus expérimentalement : c'est l'étude quantitative des relations structure temps de rétention noté QSRR [10].

Récemment, des méthodes basées sur des structures 3D comme l'étude CoMFA (Comparative Molecular Field Analysis) et CoMSIA (Comparative Molecular Similarity Indices Analysis) [11-12] permettent de traiter les relations structure-activité en trois dimensions, 3D-QSAR/QSPR.

Principe

L'étude QSAR/QSPR consiste à définir une relation mathématique entre une activité biologique, ou une propriété mesurée pour une série de composés similaires dans les mêmes conditions expérimentales, avec des descripteurs moléculaires à l'aide des méthodes statistiques. L'objectif de ces études est d'analyser les données structurales afin de détecter les facteurs déterminants pour l'activité ou la propriété étudiée. Pour ce faire, différents types de méthodes statistiques peuvent être employées.

L'expression mathématique obtenue est utilisée comme un moyen prédictif de l'activité/propriété étudiée pour de nouvelles molécules ou bien des molécules pour lesquelles les données expérimentales ne sont pas disponibles.[13]

La stratégie globale

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

relier à l'activité/propriété expérimentale étudiée. Une fois développé, le modèle doit être validé en termes de corrélation.

L'influence des composés du jeu d'entraînement sur le modèle est estimée par des méthodes de validation interne. Pour estimer le pouvoir prédictif du modèle, il est nécessaire de disposer de données expérimentales supplémentaires, c'est la validation externe.[13]

Models QSAR/QSPR

Au cours des années passées, les Relations Quantitatives Structure Activité, Quantitatives Structure Propriété (QSAR/QSPR) sont devenues un puissant outil théorique, à la place de la mécanique quantique, pour la description et la prédiction des propriétés des systèmes moléculaires complexes dans différentes circonstances.

L'approche QSAR procède à l'hypothèse d'une relation dépendante entre n'importe quelle propriété physique, affinité chimique, ou activité biologique d'un composé chimique et sa structure moléculaire. Cette correspondance univoque peut être représentée par la composition chimique, la connectivité des atomes, la surface d'énergie potentielle, et la fonction d'onde électronique d'un composé. (Figure 1)

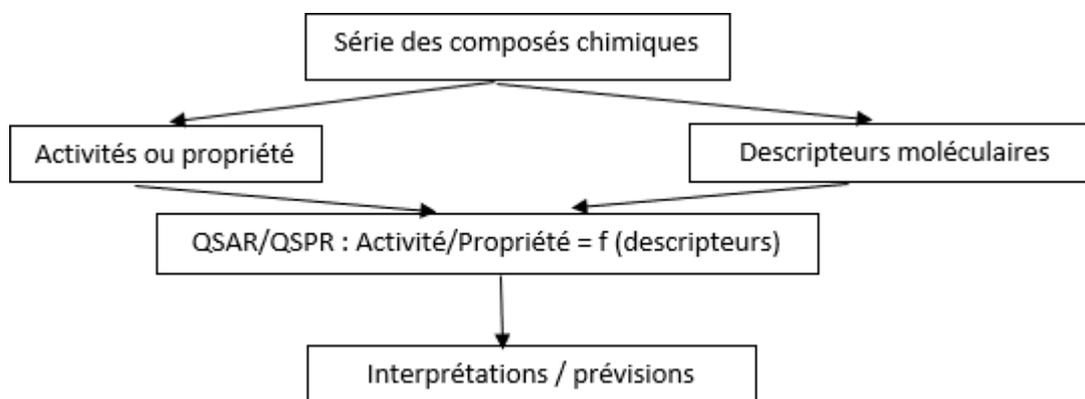


Figure 1 Schéma général QSAR

Différents descripteurs moléculaires physico- chimiques représentant la structure peuvent être déterminés empiriquement ou en utilisant des méthodes théoriques et computationnelles de différents niveaux de complexité. Il a fallu mettre en évidence que la connaissance de la constitution chimique exacte et/ou de la structure moléculaire tridimensionnelle des composés chimiques étudiés est une condition essentielle à l'application de l'approche QSAR.

La réalisation de l'approche QSAR, dépend de de l'utilisation appropriée des descripteurs moléculaires. [1]

Les descripteurs

Un descripteur moléculaire est un paramètre (une valeur numérique) propre à une structure chimique donnée.

Ces valeurs peuvent être obtenues expérimentalement ou calculées à partir de la structure de la molécule. Les descripteurs calculés, permettent d'effectuer des prédictions sans avoir à synthétiser les molécules, ce qui est l'un des objectifs de la modélisation moléculaire. Les descripteurs moléculaires jouent un rôle fondamental dans les études de la relation quantitative structure activité/propriété.

Ils sont utilisés en tant que variables indépendantes pour prédire une variable dépendante (activité ou propriété).

L'utilisation des descripteurs moléculaires dans le développement de modèles QSAR/QSPR n'est pas une tâche aisée. Tout d'abord, un très grand nombre de descripteurs moléculaires, de différentes complexités et de conceptions diverses a été introduit au cours des dernières années. Ensuite, pendant ce temps, aucune règle stricte n'a été établie pour la sélection de descripteurs adaptés parmi le grand nombre de descripteurs disponibles. Ce choix a souvent été basé sur l'intuition chimique des chercheurs, ou en se pliant à la tradition.

Descripteurs constitutionnels

Dans le but de représenter un système moléculaire d'une façon simple, on considère sa composition chimique seulement, sans considérer son côté géométrique ou électronique.

On prend comme exemples les caractéristiques de descripteurs constitutionnels : 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, 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/QSPR les plus simples, pour une bonne interprétation des mécanismes d'interaction mis en jeu pour la

propriété étudiée, ces derniers peuvent poser problème. Par exemple, on ne peut pas distinguer les isomères (sans géométrie). 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.

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.

Ces descripteurs, ou indices topologiques, viennent 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.

Le système moléculaire est considéré comme un graphe $G [V, R]$ dans lequel les atomes sont des sommets V et les liaisons chimiques des arrêtes R . La distance topologique d entre deux atomes est alors définie comme le nombre de liaisons minimum qui relie ces atomes.

L'indice de Wiener, par exemple, 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 et donc définie par cette équation :

$$w = \sum_{i=j}^n d_{ij} \quad (\text{eq 6})$$

Parmi les autres indices topologiques courants, on retrouve l'indice de Randic, qui peut être interprété comme une mesure de l'aire de la molécule accessible au solvant, les indices de Kier-Hall ou encore l'indice de Balaban. Enfin, des indices de contenu d'information appliquent les fondements de la théorie de l'information de Shannon aux distances topologiques. (Figure 2)

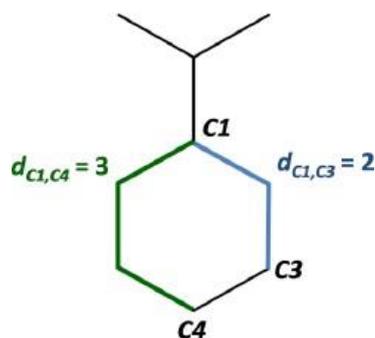


Figure 2 Distances topologiques au sein de la molécule de nitrobenzène

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 numériques. Cela dit, dans la plupart des cas, l'interprétation des équations QSAR/QSPR qui en résulte n'est pas aisée, puisqu'il est difficile de les relier aux mécanismes sous-jacents. (Figure 3)

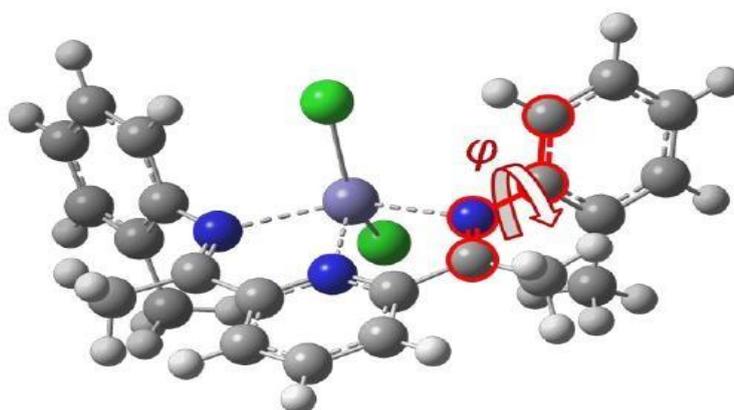


Figure 3 Angle dièdre dirigeant l'activité catalytique des complexes bis (arylimino) pyridine du fer

Descripteurs 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.

Un tel paramètre géométrique a pu notamment montrer son rôle central pour caractériser l'activité des complexes bis (arylimino) pyridine du fer (II) pour l'oligomérisation de l'éthylène. Il a pu être montré que l'activité de ces complexes est liée à un angle dièdre particulier φ qui régit l'accessibilité au cœur réactif métallique du précurseur du complexe catalytique par son activateur.

Descripteurs quantiques

Afin d'aller plus loin dans la description, des structures moléculaires des caractéristiques supplémentaires de la structure moléculaire peuvent encore être quantifiées. Ces descripteurs qui, pour être calculés de manière fiable, nécessitent des calculs plus sophistiqués sont regroupés ici sous l'étiquette de « descripteurs quantiques ».

Les approches de chimie quantique nous donnent accès à des informations supplémentaires telles que des données énergétiques, vibrationnelles et orbitales du système.

Il est alors possible, par exemple, de calculer des énergies de dissociation. Pour ce faire, considérons une dissociation homolytique sachant que A et B sont 2 composés d'une même molécule :



L'énergie de dissociation de la liaison A-B est alors la différence d'énergie entre la molécule AB et la somme des radicaux A et B issus de la dissociation de la liaison considérée.

$$E_{\text{diss}}(A-B) = E(A) + E(B) - E(AB) \quad \text{(eq 8)}$$

A partir de tels calculs, sont également évaluées des propriétés électroniques telles que le moment dipolaire ou la polarisabilité. Des descripteurs permettent également de quantifier les différentes interactions inter- et intramoléculaires, ces dernières ayant une grande influence sur la plupart des propriétés des systèmes moléculaires.

Ce sont notamment des descripteurs de réactivité pouvant être dérivés de la DFT conceptuelle ou encore des charges partielles pouvant être calculées selon différentes approches.

Descripteurs de réactivité issus de la DFT conceptuelle

La DFT conceptuelle permet de caractériser les propriétés de réactivité des composés chimiques. Les descripteurs qui en sont issus représentent un moyen simple de rationaliser le comportement chimique des molécules, sur la base de la théorie HSAB (Hard and Soft Acides and Bases) de Pearson. Leur fiabilité a d'ailleurs été démontrée via différentes analyses théoriques, dédiées principalement à la réactivité chimique. [14]

3 Méthodes utilisées pour le développement de modèles QSAR/QSPR

Introduction

La réalisation d'un modèle QSAR n'est pas une tâche aisée, on rencontre plusieurs difficultés liées à plusieurs domaines.

La première difficulté rencontrée et due à la différence d'échelle existant entre les données à corréliser car la structure est à une échelle moléculaire alors que les propriétés à prédire sont à une échelle macroscopique.

De plus il tient compte des problèmes d'incertitude à la fois au niveau des structures moléculaires liées niveau du calcul et de données expérimentales protocole de mesure

Parmi les problèmes importants on cite le traitement de grande quantité de données, on analyse un grand nombre de descripteurs et de molécules. Or il n'existe aucune règle stricte concernant le choix des paramètres structuraux les plus importants parmi la totalité de ceux disponibles.

Afin de mettre en place des modèles QSAR fiables à partir de données disponibles il faut trouver le moyen le plus adapté parmi plusieurs outils existants.

Plusieurs approches sont envisageables, selon chaque cas il faut alors choisir celles permettant de mieux caractériser le système. [14]

En dernier ressort, les modèles QSAR peuvent être développés selon des modèles mathématiques différents, généralement en relation avec l'analyse statistique multivariée.

Le premier modèle, et le plus largement utilisé, consiste en une équation (multi) linéaire obtenue par régression des données expérimentales en fonction d'un ensemble de descripteurs pré-sélectionnés (ou d'un seul), en utilisant la méthode des moindres carrés ordinaires (MCO).

Dans quelques cas, les modèles physiques ou chimiques connus du phénomène étudié laissent prévoir certaines formes mathématiques non linéaires (exponentielles ou logarithmiques) de la dépendance entre les données expérimentales et les descripteurs moléculaires. Les modèles QSAR peuvent alors être établis à l'aide de la technique de régression par les moindres carrés non linéaires. [1]

Méthodes de régressions linéaires et multilinéaire

Comme cité plus haut, le chercheur choisit dans chaque cas un ou plusieurs descripteurs supposé(s) refléter les interactions physiques ou chimiques à la base de la propriété moléculaire ou de la caractéristique du phénomène étudié.

Ce choix est généralement fondé sur l'intuition chimique, la tradition ou simplement la disponibilité du descripteur.

Néanmoins pour établir un modèle QSAR, 5 principes peuvent aider à la sélection des descripteurs convenables pour l'établissement de modèles QSAR. Ce sont :

- a) Un nombre maximal de données expérimentales (de préférence toutes) doivent être caractérisées par des valeurs de descripteurs originaux complémentaires.
- b) Les valeurs des descripteurs doivent être obtenues de la même source et, de préférence, mesurées selon le même protocole expérimental ou calculées en utilisant le même logiciel.
- c) Le nombre de descripteurs dans les modèles de régression multiples doit être minimisé, sans perte d'information, ce qui met en évidence les critères statistiques (valeurs des tests t et F...).
- d) Dans les modèles RLM, les descripteurs utilisés doivent être statistiquement orthogonaux.
- e) Pourvu que les autres critères soient similaires, la nature physique ou chimique du descripteur sélectionné doit être la plus proche de la propriété ou du phénomène étudié.

En réalité, il est difficile de se conformer pratiquement aux 5 principes énoncés. Cependant, la négligence de plusieurs d'entre eux peut conduire à des équations inutiles sans aucun pouvoir prédictif sinon très limité. [1]

Evaluation préliminaire des données

Il est hautement recommandé d'examiner la qualité statistique des données de départ. A la fois les données à corrélérer et les descripteurs utilisés dans la corrélation avant d'entamer le développement effectif des équations de régression QSAR.

Dans ce genre de pré-traitement des données on distingue généralement les analyses univariées et les analyses bivariées.

Dans l'analyse univariée, il est recommandé de vérifier la conformité des données à la distribution normale. Une précaution particulière doit être prise lors de la procédure de régression subséquente si les valeurs de la propriété étudiée, ou d'un descripteur, ne suivent pas la loi de Laplace- Gauss.

Pour un ensemble de descripteurs différents, il est nécessaire d'effectuer une analyse des données bivariées, c'est-à-dire de calculer le coefficient de corrélation linéaire R entre chacune des paires de l'ensemble des descripteurs. Si R est statistiquement significatif ($R > 0,95$), ces deux descripteurs ne peuvent être utilisés simultanément lors de l'analyse par RLM. [1]

Régression linéaire multiple

Un modèle de régression linéaire multiple entre une variable expliquée Y et p variables explicatives X_1, \dots, X_p , s'écrit pour tout $i=1, \dots, n$:

$$y_i = \beta_0 + \beta_j x_{ij} + \varepsilon \quad (\text{eq.9})$$

ou les $y_i, x_{i1}, x_{i2}, \dots, x_{ip}$ sont des données respectivement relatives aux variables Y, X_1, \dots, X_p .

Les estimateurs β_j sont calculés en utilisant la méthode des moindres carrés ordinaires. Les variables aléatoires ε_i représentent les termes d'erreur non observables du modèle. On peut estimer ces erreurs par les résidus ordinaires e_i , différence entre les valeurs observées y_i et les valeurs estimées \hat{y}_i .

Pour construire le modèle et admettre que les coefficients de la régression sont sans biais et convergents, on montre qu'il faut poser comme hypothèses :

- a) Les résidus (\mathbf{e}) ont une espérance mathématique nulle :

$$E(\mathbf{e}) = \mathbf{0} \quad (\text{eq.10})$$

- b) Le modèle choisi est correct (aucune variable explicative n'a été omise).
- c) Les résidus sont indépendants entre eux :

$$E(e_i, e_j) = 0 \quad \text{si } i \neq j \quad (\text{eq. 11})$$

Leurs covariances sont nulles.

- d) Les résidus ont tous même variance σ^2 (propriété d'homoscédasticité).

Par ailleurs, l'emploi de tests statistiques pour analyser la variation expliquée par la régression conduit à admettre que :

- e) Les résidus suivent une distribution normale (de Laplace- Gauss).

L'analyse des résidus présente un intérêt à plusieurs égards. Elle permet en effet de vérifier, a posteriori, la validité du modèle utilisé, en ce qui concerne, d'une part la forme de celui-ci (linéarité ou non linéarité de la relation, par exemple) et d'autre part, certaines hypothèses plus spécifiques, telles que l'égalité des variances résiduelles, la normalité des résidus ou l'absence d'auto- corrélation.

Pour minimiser l'influence des erreurs de détermination des valeurs explicatives (ou régresseurs) sur la précision des résultats de la régression 4 à 5 données (variables dépendantes, ou encore observations) doivent, à la limite, être associées à chaque variable explicative. Le nombre de degré de liberté final $(n-p-1)$ doit être [15] tel que :

$$n- p- 1 \geq 10 \quad (\text{eq. 12})$$

n étant la dimension de l'échantillon, et p le nombre de variables explicatives entrant dans la construction du modèle.

Pour les modèles à plus de deux descripteurs, de faibles coefficients de corrélation croisés n'assurent pas forcément l'orthogonalité des descripteurs. Une indépendance globale acceptable des descripteurs sera vérifiée lorsque les facteurs d'inflation de la variance (FIV) calculés pour chacun d'eux obéissent [15] à la condition $FIV < 5$.

Deux paramètres statistiques sont couramment utilisés pour l'évaluation de la qualité du modèle :

- Le coefficient de détermination multiple :

$$R^2 = 1 - \frac{\text{SCE}}{\text{SCT}} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (\text{eq. 13})$$

Où \bar{y} est la valeur moyenne des valeurs observées pour l'ensemble de calibrage.

- La racine de l'écart quadratique moyen de calcul :

$$\sigma_N = \text{EQMC} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (\text{eq. 14})$$

Il est intéressant de considérer, également, la racine de l'écart quadratique moyen de prédiction (EQMP), et celle calculée sur l'ensemble de validation externe (EQMP_{ext}) :

$$\text{EQMP} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_{(i)})^2} = \sqrt{\frac{\text{PRESS}}{n}} \quad (\text{eq. 15})$$

$$\text{EQMP}_{ext} = \sqrt{\frac{\sum_{i=1}^{n_{ext}} (y_i - \hat{y}_i)^2}{n_{ext}}} \quad (\text{eq. 16})$$

La validation croisée par « leave – one - out » (LOO) [16] consiste à recalculer le modèle sur (n-1) observations, et à utiliser le modèle ainsi obtenu pour calculer la grandeur d'intérêt du composé écarté, notée $\hat{y}_{(i)}$. On répète le procédé pour chacune des grandeurs d'intérêt. La somme des carrés des erreurs de prédiction, désignée par le symbole PRESS (eq.15), est une mesure de la dispersion des estimations. On l'utilise pour définir le coefficient de prédiction [16] :

$$Q_{\text{LOO}}^2 = \frac{\text{SCT} - \text{PRESS}}{\text{SCT}} \quad (\text{eq. 17})$$

Contrairement à R^2 qui augmente avec le nombre de paramètres du modèle, le facteur Q_{LOO}^2 affiche une courbe avec maximum (ou avec palier) obtenue pour un certain nombre de descripteurs, puis décroît de façon monotone. Ce qui confère une grande importance au coefficient Q_{LOO}^2 . Une valeur de $Q_{\text{LOO}}^2 > 0,5$ est considérée comme satisfaisante, et une valeur supérieure à 0,9 est excellente. [17]

Si de petites valeurs de Q_{LOO}^2 indiquent des modèles peu robustes, caractérisés par de faibles capacités prédictives internes, le contraire n'est pas nécessairement vrai. En fait, si une forte valeur de Q_{LOO}^2 est une condition nécessaire de robustesse et d'une possible capacité prédictive élevée d'un modèle, cette condition seule n'est pas suffisante, et peut conduire à une surestimation de la capacité prédictive du modèle lorsqu'il est appliqué à des composés réellement externes.

Evidemment, on peut être amené à écarter 2, 3 ou un plus grand nombre d'éléments à la fois, ce qui conduit aux procédures LMO (leave – many- out).

Dans le cas où on a suffisamment de données qui n'ont pas servi dans la création du modèle ou après collecte de nouvelles données, on peut ou on doit procéder à la validation de ce dernier, c'est la validation externe. La statistique se rapportant à ce procédé, notée Q_{ext}^2 , est calculée comme suit :

$$Q_{\text{ext}}^2 = 1 - \frac{\sum_{i=1}^{n_{\text{ext}}} (y_i - \hat{y}_i)^2 / n_{\text{ext}}}{\sum_{i=1}^n (y_i - \bar{y})^2 / n} \quad (\text{eq. 18})$$

Pour une grande valeur de Q_{LOO}^2 , une valeur élevée de Q_{ext}^2 permet de présager d'une bonne capacité prédictive du modèle.

La validation interne peut être également réalisée en utilisant la technique du bootstrap : Q_{boot}^2 (bootstrapping). Elle consiste à simuler m échantillons de même taille n que l'échantillon initial. Ils sont obtenus par tirage au hasard avec remise parmi les n individus observés au départ, ceux-ci ayant tous la même probabilité $1/n$ d'être choisis [16,18]. Contrairement aux validations croisées par LOO et LMO, les méthodes de bootstrap sont plus efficaces et plus stables.

4 Les Principes de l'OCDE

Les principes de l'OCDE sont les meilleurs éléments possibles des points essentiels à résoudre tout en développant des modèles QSAR fiables et reproductibles [19]. Les principes ont été formulés par des experts du QSAR lors d'une réunion tenue à Setúbal (Portugal) en mars 2002 comme lignes directrices pour la validation des modèles QSAR. Les 5 principes sont :

- _ Principe 1— Définition précise de la propriété prédite,
- _ Principe 2— Equation mathématique (ou algorithme) sans équivoque (reproductible),
- _ Principe 3— Domaine d'applicabilité défini,
- _ Principe 4— Mesures appropriées des performances,
- _ Principe 5— Interprétation des mécanismes sous-jacents (si possible).

Le défi actuel de QSAR/QSPR n'est plus au développement d'un modèle statistiquement solide, c'est l'élaboration d'un modèle capable de prédire avec précision l'activité ou la propriété de nouveaux produits chimiques [19].

5 Choix de l'ensemble de calibrage et de validation

Avant que nous débutions le développement des modèles, nous cherchons d'abord à diviser les données en deux sous-ensembles :

Un pour le calibrage et un pour la validation externe du modèle.

Le sous-ensemble de calibrage représente des données initiales.

Le sous ensemble de la validation doit être choisis pour évaluer la qualité du modèle.

Pour sélectionner ces deux sous-ensembles, il existe de nombreux algorithmes de sélection, qui se différencient par leurs techniques de base.

Sélection aléatoire : cette sélection est parmi les méthodes les plus utilisés [19], on divise l'ensemble de données par un simple processus de sélection aléatoire de l'ensemble de calibrage et de test (pour la validation externe).

Les critères de validation d'un modèle

Golbraikh et Tropsha [20]. Ont proposé un ensemble de paramètres pour déterminer la validation et la prédiction externe du modèle QSAR. Selon ces auteurs, les modèles sont considérés comme satisfaisants, si l'ensemble des conditions suivantes sont simultanément réalisées :

$$Q_{2training} > 0.5 \quad (\text{eq. 19})$$

$$R_{ztest} > 0.6 \quad (\text{eq. 20})$$

$$(R^2 - R_0^2) / R^2 < 0.1 \quad (\text{eq. 21})$$

$$K_{XY} > K_X \text{ (QUIK rule)} \quad (\text{eq. 22})$$

6 Test de randomisation

Afin de s'assurer qu'un modèle QSPR est fiable, les tests de Y-randomisation (Tropsha *et al.*)[21] sont une des techniques les plus employées. En effet, il n'est pas rare d'obtenir des corrélations fortuites (ou « chance correlation »), c'est-à-dire un modèle affichant de bons résultats statistiques (R^2 , Q^2) pour l'apprentissage, mais impliquant des descripteurs qui dans la réalité ne sont pas reliés à la propriété modélisée. Ces modèles aléatoires peuvent être détectés par la procédure Y-randomisation. Elle consiste à mélanger aléatoirement les propriétés expérimentales pour le jeu d'apprentissage et, en utilisant les mêmes descripteurs (figure 4), à entraîner à nouveau l'algorithme d'apprentissage pour tenter d'obtenir un modèle. Normalement, les modèles obtenus doivent avoir des performances très faibles. La distribution des modèles obtenus permet de fixer un seuil heuristique de signification des modèles.

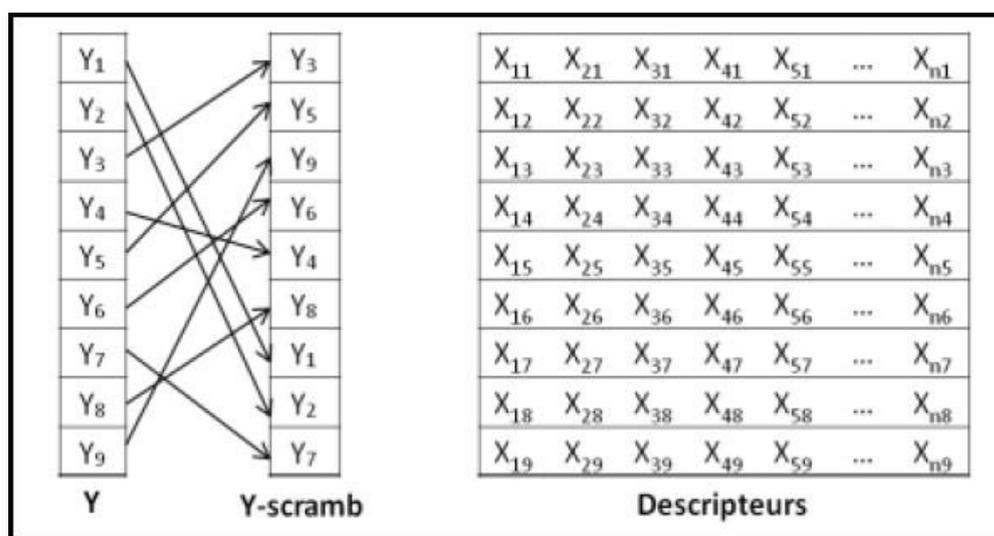


Figure 4 Illustration de la méthode « Y-scrambling » (randomisation de Y)

7 Les activités biologiques

ACHE

Introduction

On peut distinguer chez l'homme 2 types de cholinestérases différents l'acétylcholinestérase (appelé aussi cholinestérase globulaire ou cholinestérase vraie) et le butyrylcholinestérase (appelé aussi cholinestérase sérique ou pseudocholinestérase) Ces deux types sont différents par leur Origine, lieu de synthèse, structure, spécificité d'action, rôle physiologique et le taux de concentration dans l'organisme.

Acétylcholinestérase

C'est une enzyme de la classe des hydrolases synthétisée dans le globule rouge et le tissu nerveux, Elle est présente dans les synapses essentiellement, a la jonction neuromusculaire, dans la substance grise, les poumons et la rate. Elle a une affinité presque exclusivement spécifique pour son substrat naturel, l'acétylcholine.[22]

Role

Pour éviter l'accumulation de l'acétylcholine dans les synapses acétylcholinergiques et assurer le bon fonctionnement de ces derniers Acétylcholinestérase clive l'acétylcholine, un neurotransmetteur, en ion acétate et choline dans les synapses selon la réaction suivante :

$$\text{CH}_3\text{COOCH}_2\text{CH}_2\text{N} + (\text{CH}_3)_3 + \text{H}_2\text{O} \rightarrow \text{CH}_3\text{COO}^- + \text{HOCH}_2\text{CH}_2\text{N} + (\text{CH}_3)_3 \text{ (eq 23)}$$

Dosage

On utilise une méthode colorimétrique, par mesure de la cinétique d'hydrolyse à l'aide de son substrat (l'acétylthiocholine)

Les valeurs usuelles sont :

Acétylcholinestérase : de 5,14 à 8,94 U/ml de globules rouges.

Intérêt de dosage

Il permet de confirmer une intoxication aiguë ou chronique par un pesticide organophosphoré. L'acétylcholinestérase est aussi présente dans le liquide amniotique dans le cas de défaut de fermeture du tube neural (pathologie grave de l'embryogénèse)

Butyrylcholinestérase

Synthétisée par le foie, elle se trouve dans le plasma ou le sérum, le pancréas, l'intestin et d'autres tissus ..., elle présente une affinité beaucoup plus large pouvant hydrolyser un grand nombre d'esters synthétiques et naturels dont l'acétylcholine et la succinylcholine. Mais son rôle physiologique n'est actuellement pas connu.[22]

Les inhibiteurs de AChE

L'AChE est la cible d'un éventail de toxine d'origine animale, végétale et fongique, et de poisons synthétisés par l'homme. C'est une cible thérapeutique qui peut combattre des maladies neuromusculaires comme la myasthénie et la maladie d'Alzheimer. [23]

Parmi les inhibiteurs de l'AChE, les molécules de la famille des organophosphorés (OP) et des carbamates, ce sont les plus répandues. Elles agissent au niveau de la sérine du site actif.

Les carbamates sont des dérivés de l'acide carbamique. Il existe un carbamate Naturel, l'ésérine (ou physostigmine) extraite d'une fève et autrefois utilisée à des fins judiciaires dans certaines provinces d'Afrique de l'ouest [40]. Lorsque le soufre substitue un ou deux atomes d'oxygène de la structure de l'acide carbamique, on parle de thio- ou dithiocarbamates respectivement. [24]

L'inhibition de l'AChE par un carbamate est un mécanisme identique à celui d'organophosphorés. Comme les OP, les carbamates ont un groupe partant qui est hydrolysé lors de la formation du complexe ligand-AChE. Il s'ensuit une carbamylation du groupement hydroxyle de la sérine.

La différence majeure dans le processus est que l'AChE est régénérée directement à partir du complexe carbamate/AChE, et qu'il n'existe pas de processus de vieillissement. Il en résulte une régénération plus rapide de l'AChE lors de l'inhibition par un carbamate que par un organophosphoré. Une autre différence notable de leur mode d'action est que les OP, contrairement aux carbamates, passent par une étape d'activation métabolique avant de mener l'inhibition de l'AChE. [25]

COX 2

La cyclooxygénase (COX)

Appelée aussi la prostaglandine H2 synthase (PGHS) est une enzyme de la classe des oxydoréductases qui catalyse les deux premières étapes de la biosynthèse des prostaglandines (PG) à partir du substrat l'acide arachidonique (AA) ces 2 réactions sont :

L'oxydation de AA à la hydroperoxyendoperoxyde PGG₂ et sa réduction subséquente en hydroxy endoperoxyde PGH₂.

Cette enzyme a été purifiée la 1^{ère} fois en 1976 et clonée en 1988.

La découverte de l'isoforme COX 2 en début des années 90 était une grande découverte dans la biologie des prostanoïdes dans près de 20 ans. [26]

COX 2

Cette enzyme présente une grande similarité structurale et catalytique avec COX 1 avec un poids moléculaire de 71 KDA et plus de 600 acides aminés dont 63% sont dans une séquence identique.

L'aspirine et d'autres anti-inflammatoires non stéroïdiens peuvent inhiber l'activité biosynthétique des deux isoformes COX 1 et COX 2. [26]

Similitude et différence

Malgré cette similarité structurale on remarque une nette différence biochimique dans la sélectivité de substrat et de l'inhibiteur, on ne peut citer que :

COX-2 oxygène d'autres substrats d'acides gras comme l'acide eicosapentaénoïque, acide linoléique, acide β -linoléique et acide linoléique de manière plus efficace que la COX 1.

De plus COX 2 acétylé sur le Ser 530 peut toujours oxyder les AA contrairement à COX 1 acétylé de la même façon.

Fonctions physiologiques et pathologiques de COX 1 et COX 2

Il est préférable d'utiliser la classification appliquée aux isoformes COX 1 et COX 2 :

Constitutive ou induite que de classer la biosynthèse PG en physiologique et pathologique.

L'activité de la COX-1 est constitutive, présente dans presque tous les types de cellules à un niveau constant, alors que l'activité COX 2 est dans le cas normal absente dans les cellules et les niveaux augmentent et diminuent en quelques heures après un seul stimulus.

La majorité des stimuli connus pour induire la COX-2 sont ceux associés à une inflammation à titre d'exemple : les lipopolysaccharide bactérien (LPS), les cytokines (il-1 et

il-2), facteur de nécrose tumorale (TNF). Et c'est la raison principale pour étiqueter COX 1 et COX 2 comme physiologique et pathologique respectivement. [26]



Partie II

Partie expérimentale





Matériels et Méthodes



1 Introduction

L'élaboration de ce mémoire a été réalisée au Centre de Recherche en Biotechnologies Constantine (C.R.Bt), au niveau du laboratoire Bio-informatique et Bio-statistique.

Le laboratoire Bio-informatique et Bio-statistique équipé par deux stations de calculs type HP Z620 dotée d'un processeur XEON R E-2620 et 16 go destinées au calcul et traitement des données, permettant ainsi la réalisation des calculs rapidement et avec précision.

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

1. **Dessin des molécules** : Pour le dessin des structures des molécules on a utilisé le logiciel « ChemDraw » version 7.
2. **Optimisation** : tous les calculs d'optimisation des structures des molécules ont été effectués en utilisant le logiciel HyperChem 08.
3. **Modélisation** : la modélisation passe par deux étapes :
 - ◆ Calcul de descripteurs : Les géométries ainsi optimisées ont été transférées dans le logiciel Dragon version 5.5 pour le calcul de plus de 1500 descripteurs.
 - ◆ Conception du modèle : Dans le présent travail, la sélection du modèle a été réalisée par algorithme génétique, en utilisant le logiciel MOBY DIGS.

2 ChemDraw

Est un outil complet destiné aux chimistes et biologistes, intégrant toute une gamme d'outils intelligents permettant de faciliter les travaux des chercheurs, il est développé en 1985 par David A. Evans et Stewart Rubenstein. C'est un outil essentiel et préféré pour illustrer les concepts chimiques et biologiques et Il est imposé comme une référence pour le dessin de structures moléculaires. Puisqu' il est simple à utiliser, puissant et permet de dessiner de manière intuitive et efficace en deux et trois dimensions.[27]

Interface utilisateur ChemDraw

Par défaut, l'interface utilisateur affiche les barres d'outils, le menu principal et la barre d'état des documents couramment utilisés, qui nous facilitent le dessin de nos molécules. La figure 5 montre l'interface utilisateur ChemDraw.

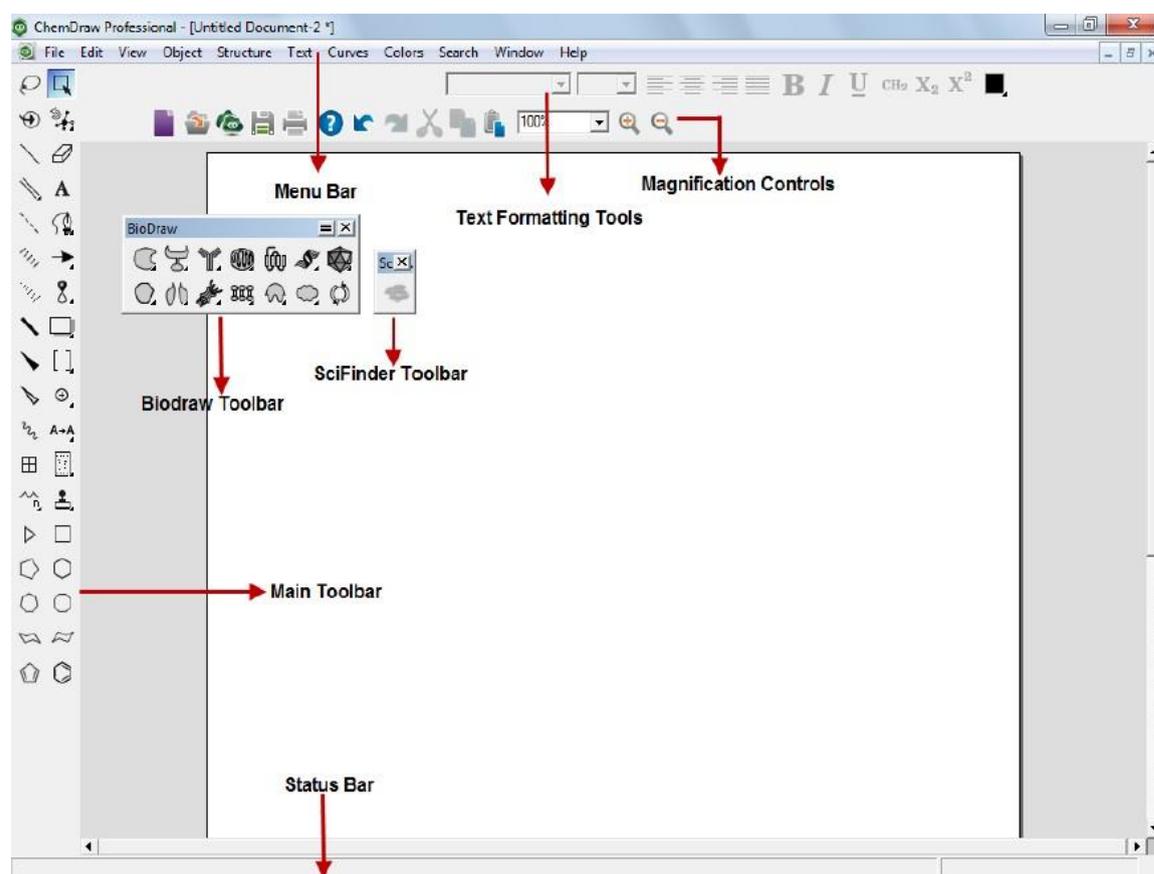


Figure 5 ChemDraw interface

Barres d'outils

Dans ChemDraw, plusieurs barres d'outils sont introduites, telles que Recherche, Structure, Courbes, etc. Nous utilisons chacune dans une tâche spécifique toujours dans le but de dessiner nos molécules de manière plus pratiques.

ChemDraw, nous permet aussi d'afficher ou masquer une barre d'outils, selon nos besoins en la sélectionnant dans le menu Affichage.

La barre d'outils principale

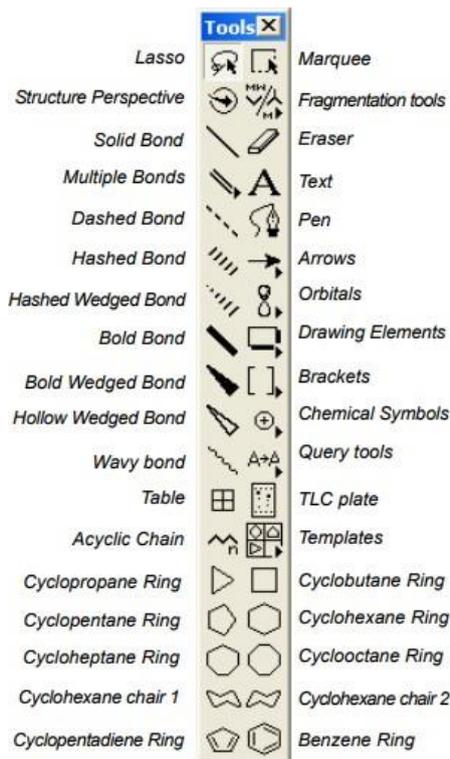


Figure 6 barre d'outils principale

La barre d'outils principale montrée dans la figure 6 inclut les outils les plus couramment utilisés pour dessiner des structures. On prend comme exemple :

Le cycle benzène, au lieu de faire 6 liaisons successives, parmi eux 3 doubles liaisons, il suffit juste de cliquer sur « benzène ring », la même chose pour le cyclohexane et cyclo octane...etc.

2.3 Enregistrement

Une fois le dessin achevé, il sera enregistré sous format MOL, pour ensuite modéliser la structure 3D de la molécule en utilisant le logiciel HyperChem.

3 HyperChem

HyperChem est un environnement de modélisation moléculaire sophistiqué, reconnu pour sa qualité, flexibilité et facilité d'utilisation. Unification de la visualisation 3D et de l'animation avec la chimie quantique, calcul la mécanique moléculaire et la dynamique, HyperChem met plus de modélisation moléculaire à votre portée que tout autre programme Windows. Il comprend tous les composants de structure, thermodynamique, spectres et cinétiques.[28]

Optimisation géométrique

Mécanique moléculaire MM+

Les molécules déjà dessinées à l'aide du logiciel ChemDraw ont été importées à HyperChem.

Tout d'abord, il faut préparer la molécule en ajoutant les atomes d'hydrogène. (Voir figure 7)

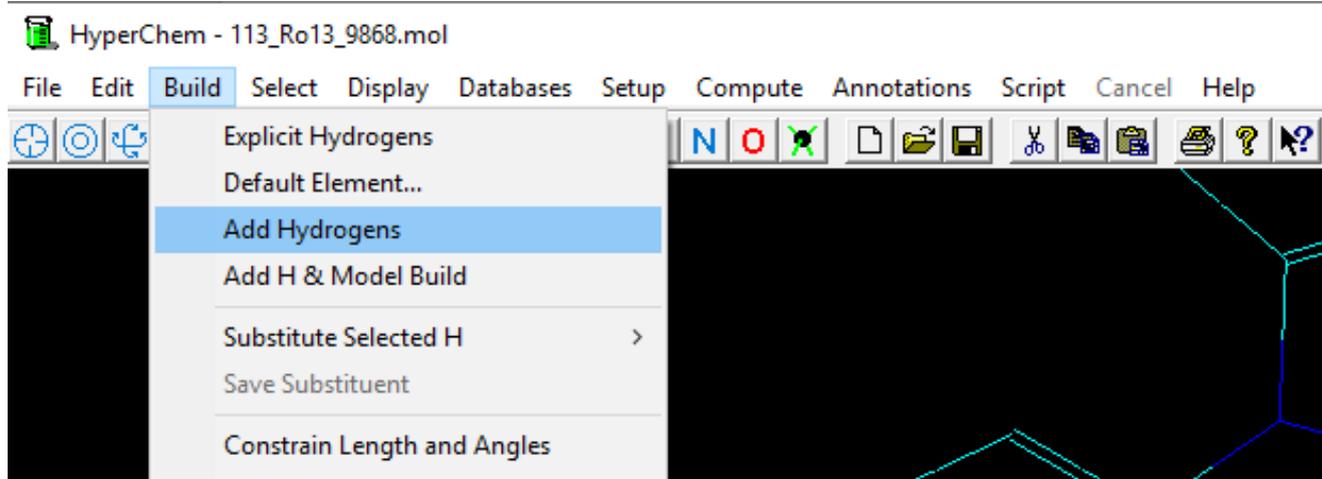


Figure 7 ajout d'hydrogènes

Maintenant la molécule est prête pour être optimisée comme suit :

Sélection du menu "Setup" dans la ligne supérieure de la fenêtre HyperChem puis dans la liste, « Mécanique moléculaire" est sélectionné comme le montre la figure 8.

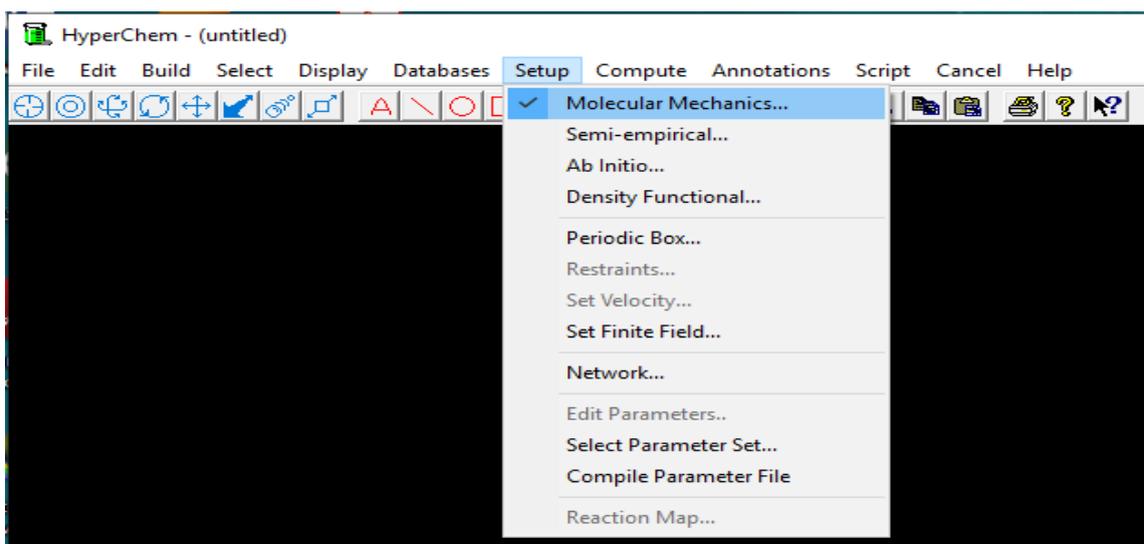


Figure 8 choix de mode d'optimisation

Dans la fenêtre "MolecularMechanics Force Field" qui apparaît "MM +" est sélectionné. (Voir figure 9)

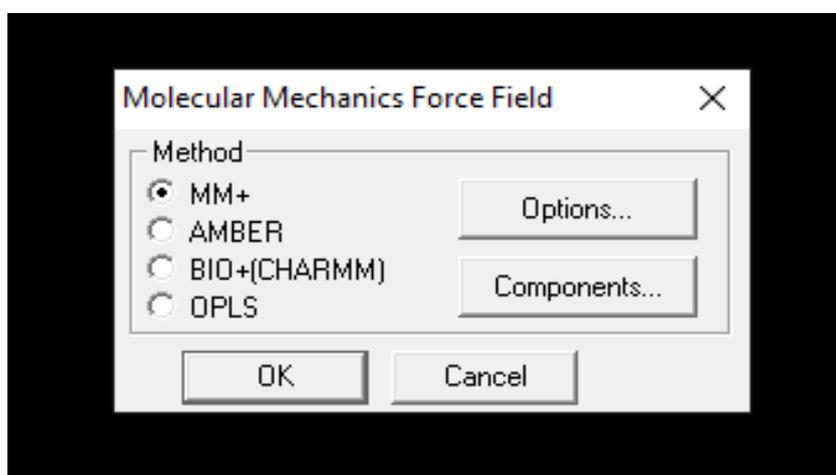


Figure 9 choix du mode mm+

Le logiciel est maintenant prêt pour lancer l'optimisation de la géométrie. (Voir figure 10)

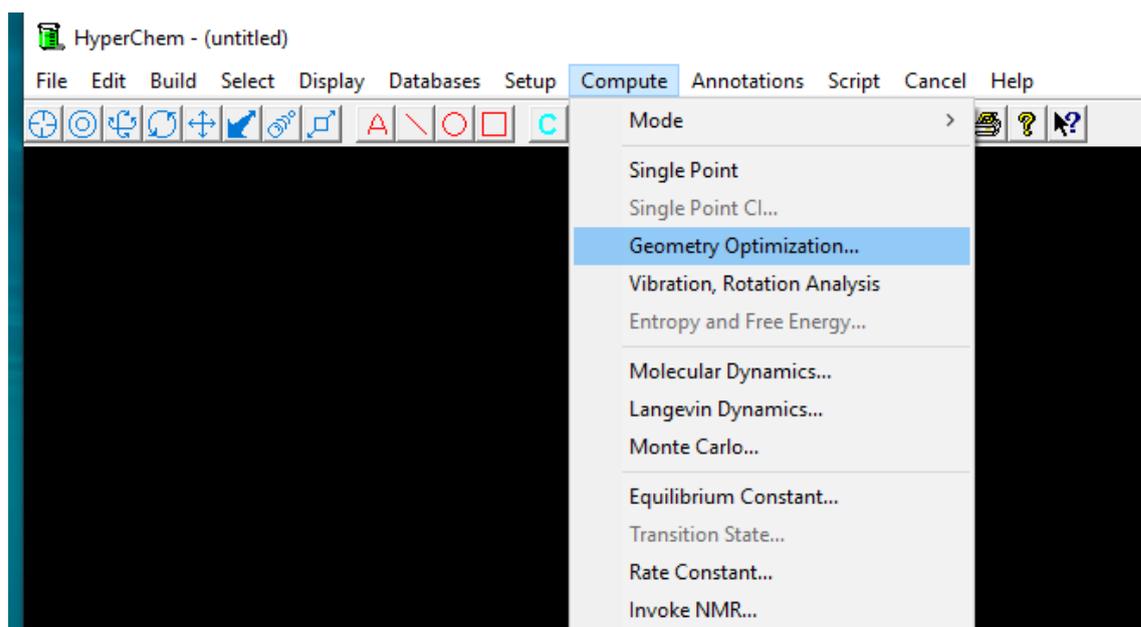


Figure 10 lancement de l'optimisation

Semi Empirique PM3

Nous pouvons maintenant passer à la deuxième étape d'optimisation dite « semi empirique ». (Voir figure 11)

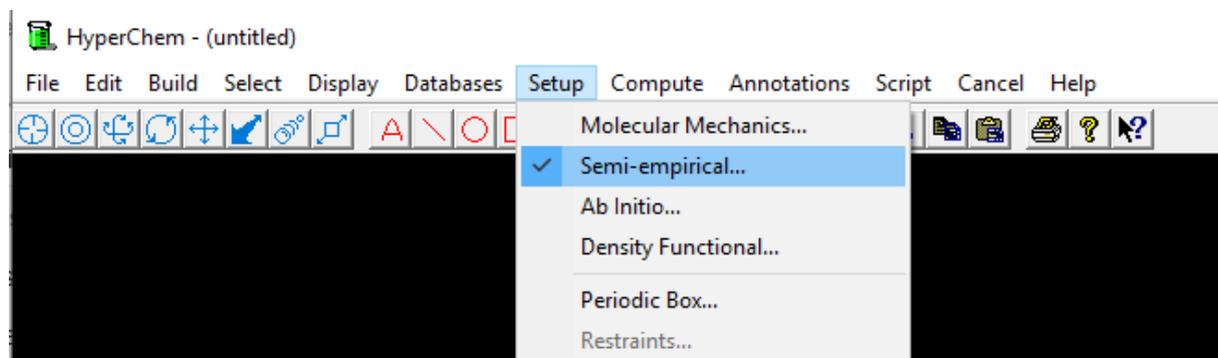


Figure 11 choix de méthode semi empirique

Une fenêtre de méthodes semi-empirique apparait et il faut choisir le paramètre PM3 comme nous avons démontré dans la figure 12.

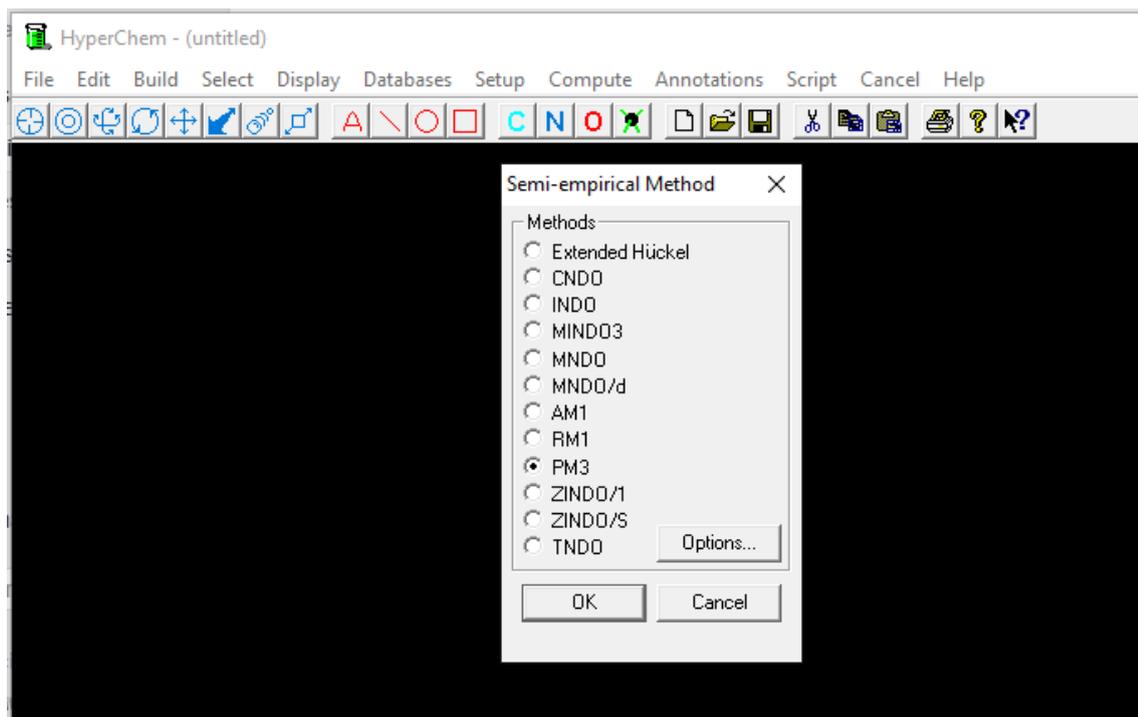


Figure 12 sélection du mode PM3

4 DRAGON

DRAGON est une application de calcul de descripteurs moléculaires, ces descripteurs peuvent être utilisés pour évaluer les molécules et la relations structure-activité ou structure-propriété, ainsi que pour l'analyse de similarité et criblage à haut débit des bases de données moléculaires.

Les versions de DRAGON ont évolué avec le temps afin de faire avancer la recherche dans le QSAR.

Sachant que le DRAGON n'a pas été conçu comme un logiciel QSAR, il ne fournit que des descripteurs moléculaires et ne peut pas effectuer une analyse QSAR (au contraire le MobyDigs).

Cependant, DRAGON permet la fusion des descripteurs moléculaires calculés et des propriétés définies par l'utilisateur pour un ensemble de molécules, fournissant un fichier de sortie complet qui est facilement chargé.[29]

Calcul de descripteurs

Après avoir dessiné les molécules données par le ChemDraw et modéliser leurs structures 3D avec le logiciel HyperChem nous les avons enregistrées sous format HIN. (Car c'est le seul format que Dragon peut traiter en bloque).

En ce qui suit les étapes que nous avons suivies pour le calcul des descripteurs de notre ensemble de molécules :

Le fichier contenant les molécules en question est importé au logiciel en le localisant dans la section « look in » de l'interface du logiciel, puis les molécules choisies sont sélectionnées dans le volet « select molécule files » à gauche dans la même fenêtre. (Voir figure 13)

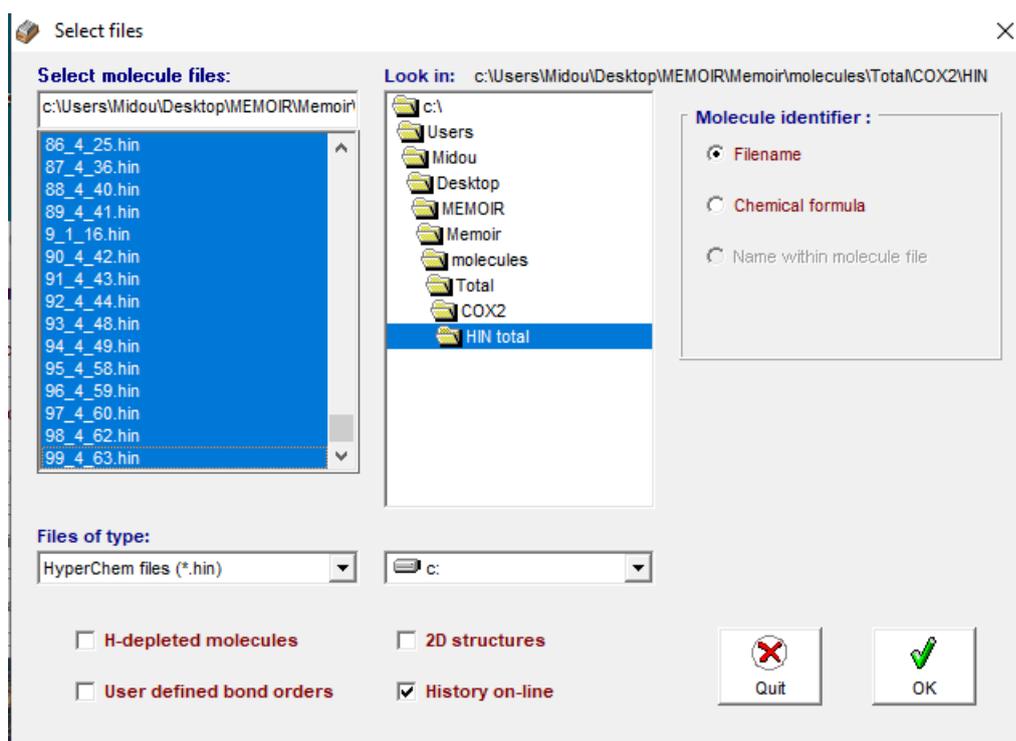


Figure 13 import du fichier et sélection des molécules

Une fois les molécules sont sélectionnées, nous passons alors au choix des descripteurs calculés, on coche tous les descripteurs disponibles pour le dragon 5 (1664 descripteurs), vu que chaque information donnée par les descripteurs est importante dans l'analyse QSAR/QSPR. Et lancer le calcul par le bouton « Run ». (Voir figure 14)

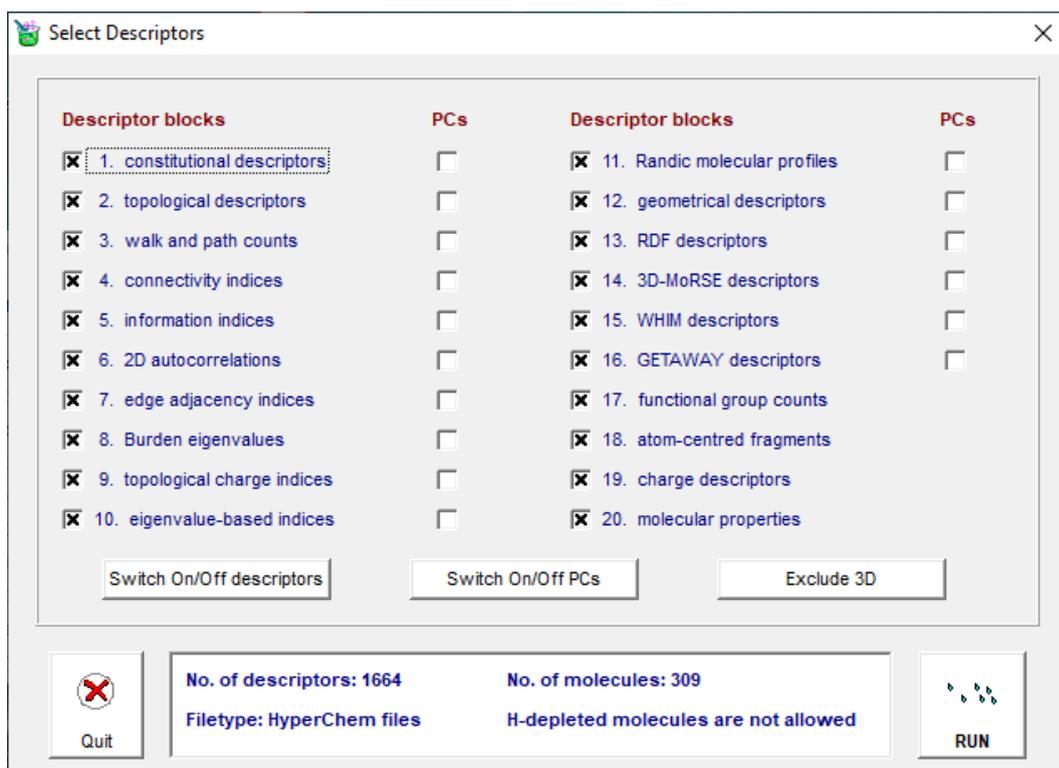


Figure 14 choix des descripteurs et lancement du calcul

Après avoir terminé toutes ces étapes, Dragon affiche un dernier message indiquant la fin de la procédure de sauvegarde qui contient le nom du fichier, la racine de la sauvegarde... (Voir figure 15).

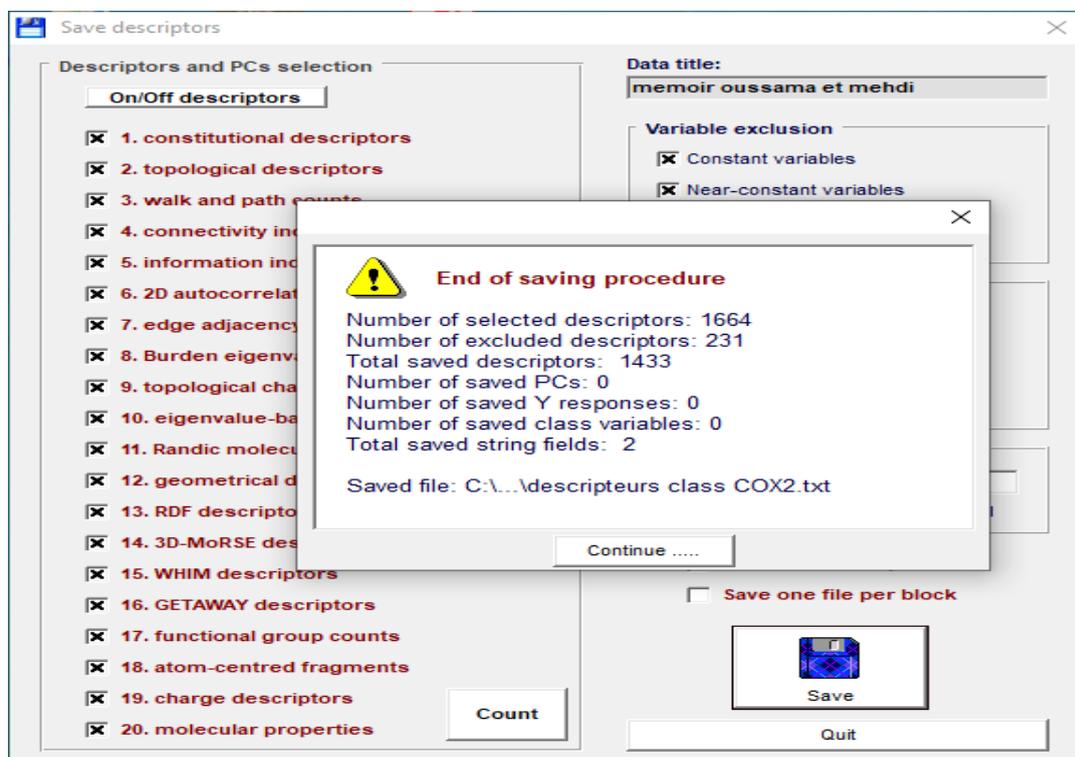


Figure 15 fin de sauvegarde

5 MobyDigs

Après avoir dessiné les molécules avec « ChemDraw » et optimiser leurs structures via le logiciel « HyperChem », ces derniers sont passés par un troisième logiciel Dragon pour calculer les descripteurs.

Les résultats ont été traités et ajoutés aux valeurs iC_{50} expérimentaux des molécules en question avec Excel, puis enregistrés sous forme TXT supporté par le logiciel de calcul « MobyDigs ».

Définition

C'est un logiciel qui permet de calculer un model QSAR avec des régressions linéaires multiples via des équations citées en partie théorique.

Méthode utilisée dans MobyDigs

Le fichier contenant les descripteurs et les activités des molécules est importé dans le menu « wizard » (Voir figure 16)

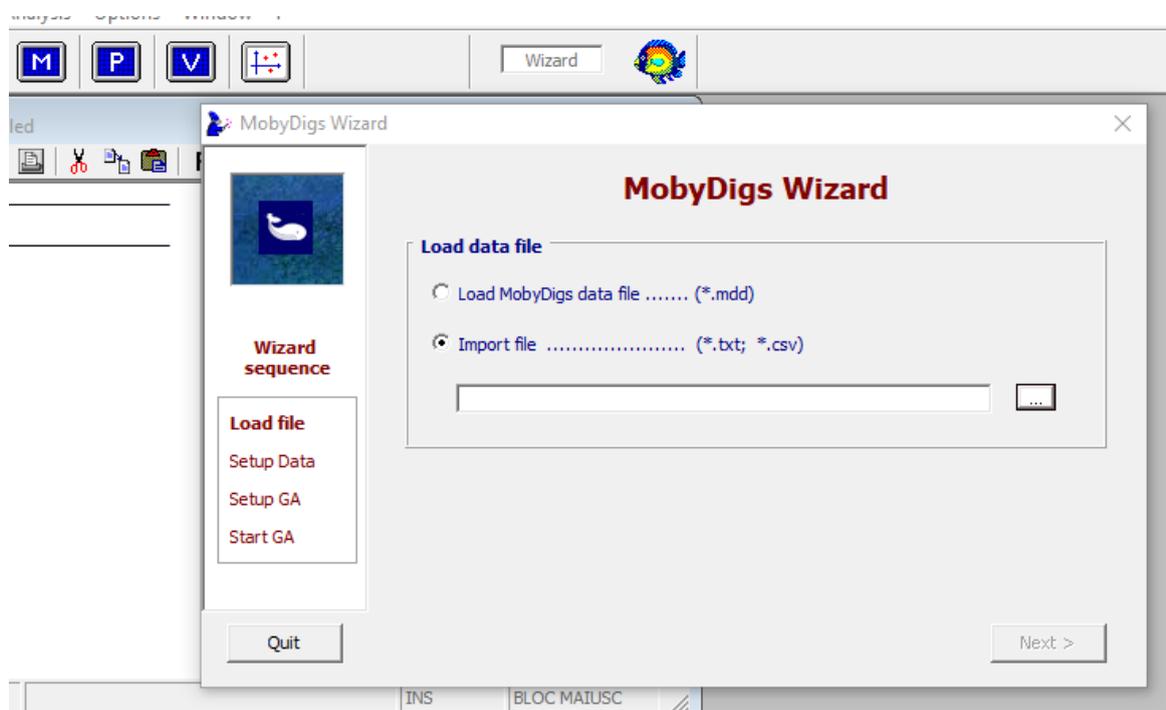


Figure 16 importer le fichier

La variable Y qui représente le « pIC50 » est sélectionnée. (Figure 17)

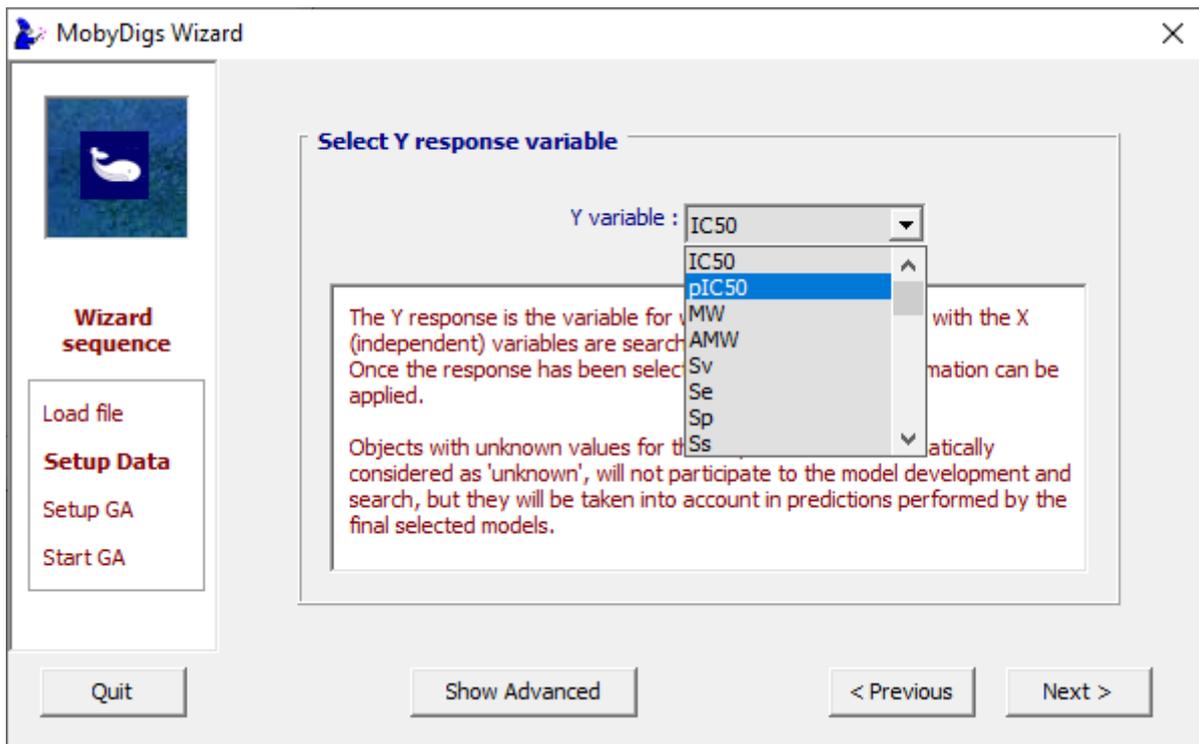


Figure 17 sélection de Y = pIC50

Les descripteurs sont sélectionnés comme variable X. (Figure 18)

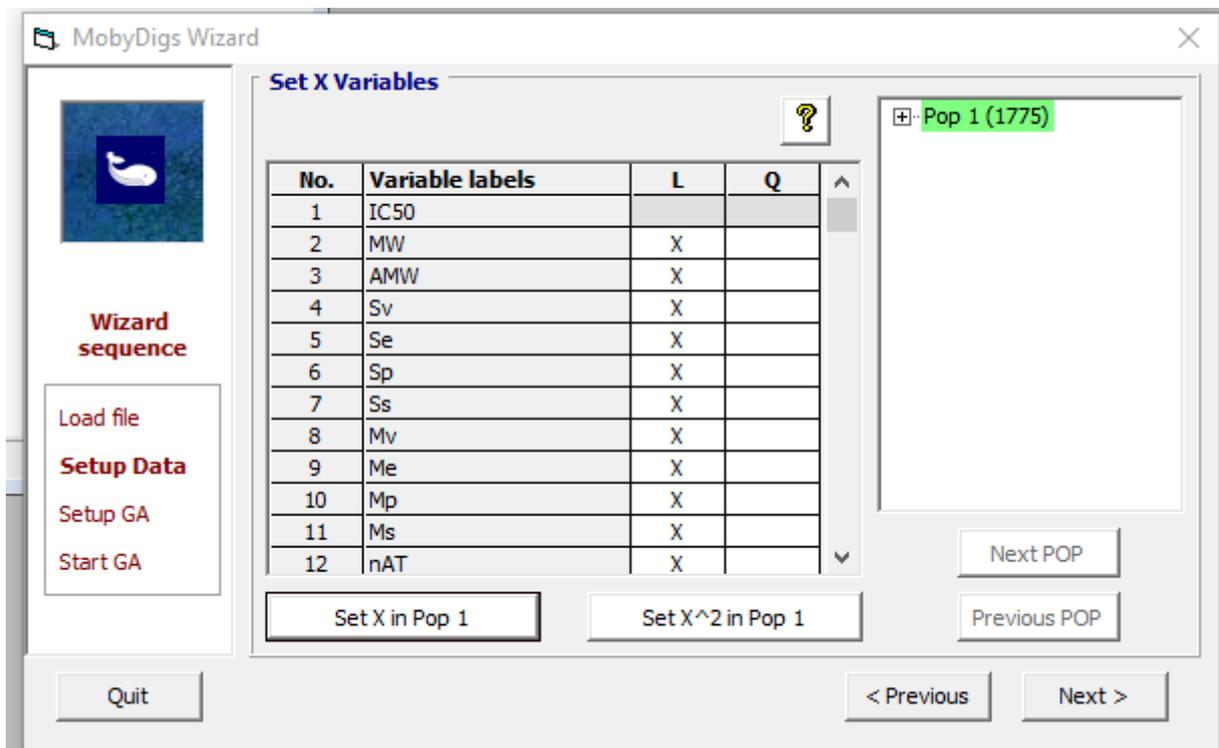


Figure 18 sélection de variable X = descripteurs

Nous avons divisé l'ensemble des molécules en deux groupes (training 80%, test 20%) via un choix aléatoire afin de réaliser une validation interne et externe. (Figure 19)

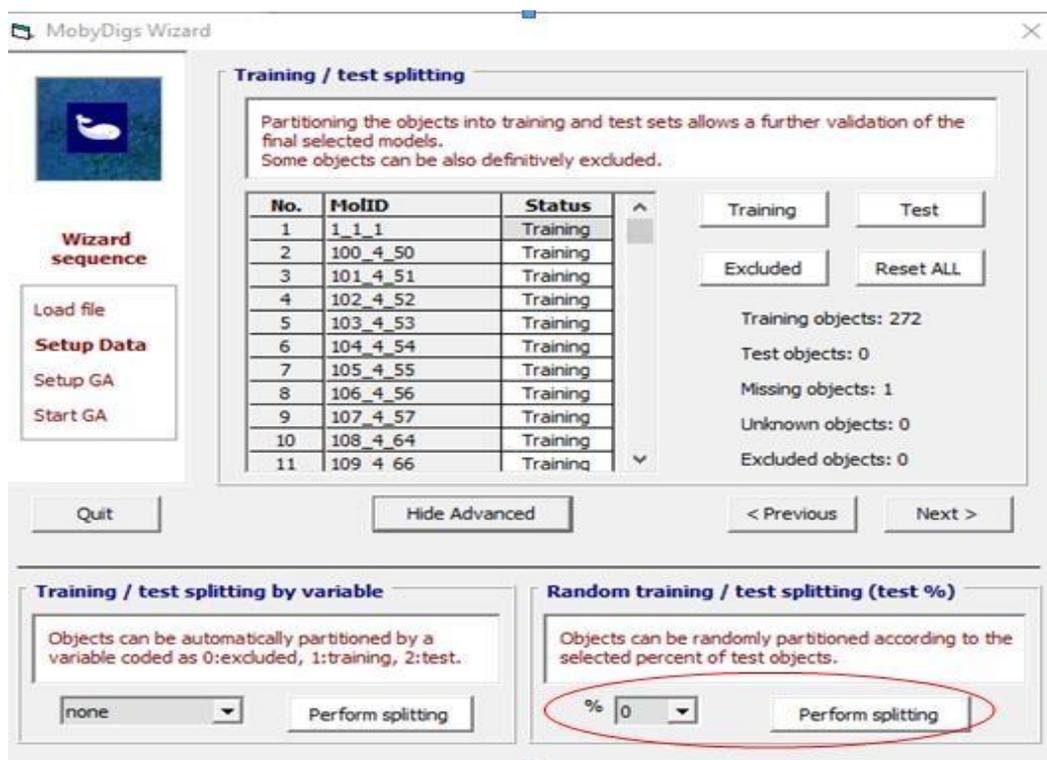


Figure 19 division des groupes

Nous avons choisi la méthode « Q2 leave on out » précédemment décrit (voir partie théorique) et 3 descripteurs. (Figure 20)

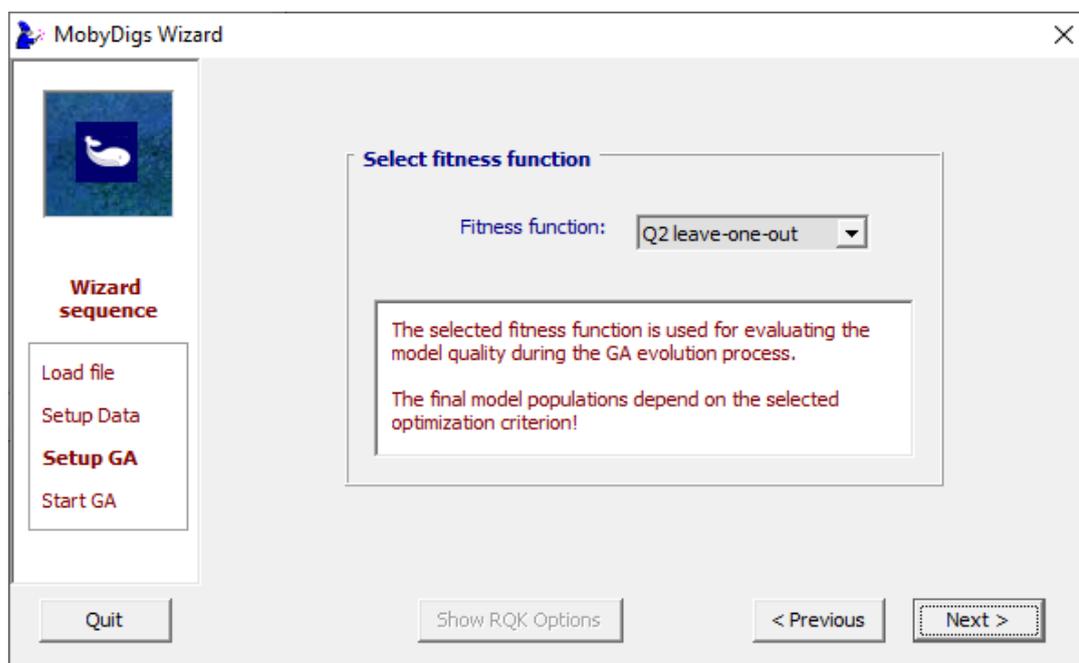


Figure 20 choix de methode Q2 leave one out

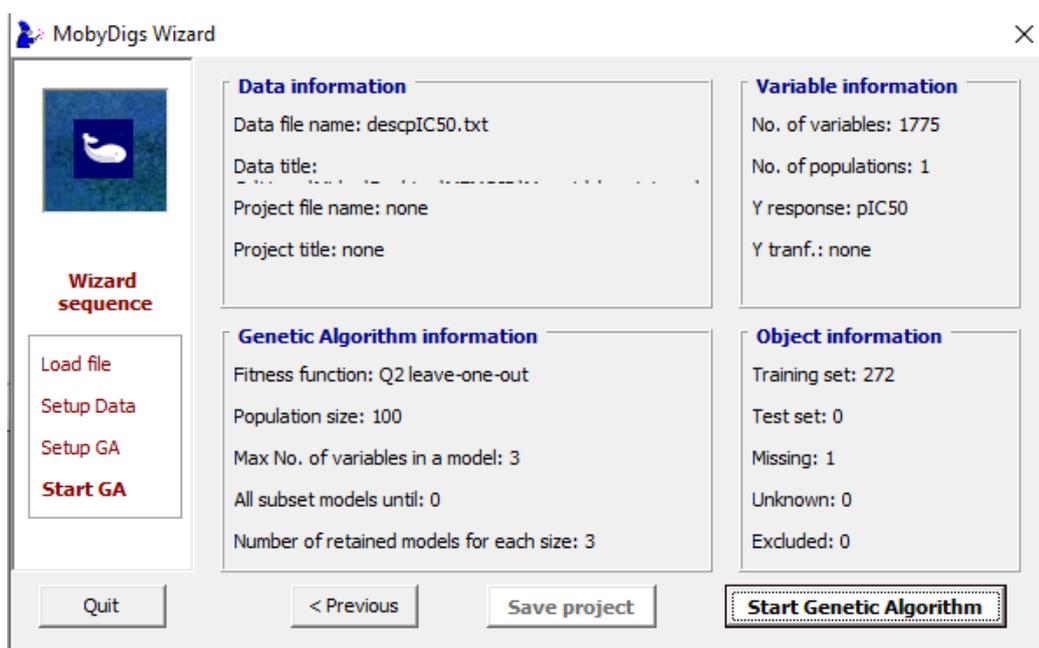


Figure 21 lancement du calcul

MobyDigs commence à calculer (figure 21) et les résultats finals pour les deux familles de molécules sur lesquelles nous avons travaillé (ACHE, COX2) sont démontrés et interprétés dans la partie « résultat et discussions ».

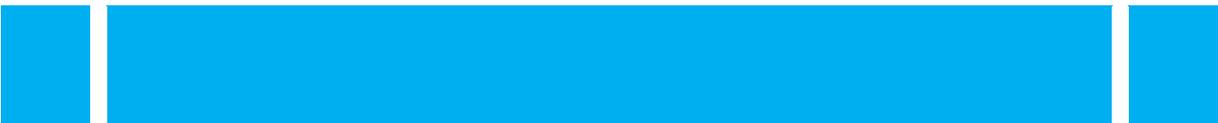


Résultats et discussion





***RELATIONS
STRUCTURE ACTIVITE
DES AchE***



6 Application de l'approche QSAR dans la modélisation de l'AchE

La concentration d'inhibition (IC₅₀) d'inhibiteur acétylcholinestérase des dérivés de : [2-(1-Benzyl-piperidin-4-yl)-ethyl]-diméthyl-amine a été modélisée par régression multilinéaire (MLR). Le modèle a été validé par validations interne et externe.

Origine des données

Les valeurs de l'activité AchE (IC₅₀) rapportées dans [30-32], ont été utilisées pour l'analyse des relations structure / propriétés.

7 Résultats et discussion

Matrice de corrélation

Tableau 1 matrice de corrélation pour le modèle 01

	MATS5v	nH	RDF150m
nH	0,232		
RDF150m	0,452	0,561	
F06[C-O]	0,478	0,651	0,7

Equation et analyse de régression :

Le modèle QSAR sélectionné est représenté par l'équation :

$$pIC_{50} = (\pm 0.91)0.83 - nH(\pm 0.03)0.19 - MATS5v(\pm 0.88)2.84 + RDF150m(\pm 0.21)0.83 + F06[C-O](\pm 0.06)0.54 \text{ (eq 24)}$$

Avec les valeurs des paramètres statistiques ci-après :

Tableau 2 valeurs essentielles pour le modèle 01

R ²	Q ₂	Q _{2ext}	R _{2adj}	SDEP	SDEP _{ext}	K _x	K _y	F	S
79.18	75.07	75.12	76.41	0.552	0.551	40.92	41.44	28.5275	0.5449

Tableau 3 définitions des descripteurs

Descripteur	Classe	Définition
nH	Constitutional indices	Nombre d'atomes d'hydrogènes
MATS5v	2D autocorrelations	Autocorrélation de Moran du lag 5 pondérée par le volume de van der Waals
RDF150m	RDF descriptors	Fonction de distribution radiale - 150 / pondérée en masse
F06[C-O]	2D Atom Pairs	Fréquence de C - O à la distance topologique 6

Analyse des résidus et diagnostiques d'influence

Les valeurs des résidus de prédiction standardisés, et les valeurs de Hat sont présentées dans le tableau 4 dont les colonnes 7 et 8 reproduisent les valeurs expérimentales et calculées, prédites de pCIC₅₀ des composés considérés.

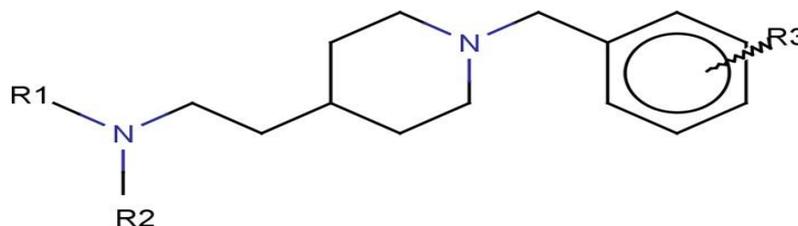


Figure 22 template des molécules AchE

Tableau 4 Valeurs des pCIC₅₀, observées (piC₅₀) et calculées, prédites (pred y), des résidus standardisés (R1. R2. R3) ainsi que des leviers (hat) (std.Err.Pred).

Name	R1	R2	R3	Famille	Ic50	piC50	Pred y	Hat	Std.Err.Pred
1-1	PhCO-	H-	H	A.1	560	-2.748	-2.41	0.065	0.64
1-2	PhCO-	Me-	H	A.1	170	-2.23	-2.91	0.041	-1.28
1-3a	PhCO-	Ph-	H	A.1	35	-1.544	-2.81	0.049	-2.38
1-3b	p-OCH ₃ PhCO-	Ph-	H	A.1	590	-2.771	-2.94	0.06	-0.32
1-3c	p-FPhCO-	Ph-	H	A.1	18	-1.255	-1.86	0.111	-1.18
1-3d	p-NO ₂ PhCO-	Ph-	H	A.1	5.4	-0.732	-0.39	0.231	0.72
1-3e	p-pyridineCO-	Ph-	H	A.1	64	-1.806	-1.64	0.815	0.69
1-3f	cyclohexylCO-	Ph-	H	A.1	9400	-3.973	-3.65	0.227	0.68
1-3g	CH ₃ CO-	Ph-	H	A.1	52	-1.716	-2.4	0.046	-1.29
1-3h	CH ₃ CH ₂ CO-	Ph-	H	A.1	830	-2.919	-2.79	0.048	0.24
1-3i	CH ₃ CO-	m-OCH ₃ Ph-	H	A.1	46	-1.663	-1.57	0.126	0.19
1-3j	CH ₃ CO-	p-OCH ₃ Ph-	H	A.1	700	-2.845	-2.01	0.101	1.61
1-3k	CH ₃ CO-	m-FPh-	H	A.1	65	-1.813	-2.18	0.059	-0.7
1-3l	CH ₃ CO-	p-FPh-	H	A.1	205	-2.312	-2.93	0.132	-1.21
1-3m	CH ₃ CH ₂ -	Ph-	H	A.1	12000	-4.079	-4.66	0.191	-1.18
1-3n	CH ₃ CO-	p-pyridine-	H	A.1	108	-2.033	-2.65	0.052	-1.16
1-3o	CH ₃ CO-	Me-	H	A.1	660	-2.82	-2.65	0.053	0.32
2-3b	o-MePhCO-	H-	H	A.1	1000	-3	-2.82	0.045	0.34
2-3c	m-MePhCO-	H-	H	A.1	470	-2.672	-2.61	0.077	0.13
2-3d	p-MePhCO-	H-	H	A.1	180	-2.255	-2.28	0.055	-0.05
2-3e	o-NO ₂ PhCO-	H-	H	A.1	880	-2.944	-2.6	0.634	1.05
2-3f	m-NO ₂ PhCO-	H-	H	A.1	230	-2.362	-2.62	0.077	-0.5
2-3g	p-NO ₂ PhCO-	H-	H	A.1	55	-1.74	-1.2	0.165	1.08
2-3h	p-OMePhCO-	H-	H	A.1	88	-1.944	-1.83	0.206	0.24
2-3i	p-CHOPhCO-	H-	H	A.1	120	-2.079	-1.31	0.13	1.52
2-3j	p-ClPhCO-	H-	H	A.1	180	-2.255	-2.27	0.082	-0.03
2-3k	p-FPhCO-	H-	H	A.1	85	-1.929	-2.36	0.077	-0.82

2-3l	p-MeCOPhCO-	H-	H	A.1	51	-1.708	-1.8	0.083	-0.18
2-3m	p-(PhCH ₂ SO ₂)PhCO-	H-	H	A.1	29	-1.462	-1.34	0.202	0.24
2-3p	p-pyridineCO-	H-	H	A.1	39	-1.591	-1.28	0.249	0.66
2-3r	cyclohexylCO-	H-	H	A.1	1600	-3.204	-3.19	0.091	0.03
2-12	PhCH ₂ -	H-	H	A.1	46000	-4.663	-4.29	0.239	0.78
2-14	PhCO-	Et-	H	A.1	130	-2.114	-2	0.300	0.26
2-15	PhCO-	PhCH ₂ -	H	A.1	940	-2.973	-3.16	0.183	-0.38
2-21	p-(PhCH ₂ SO ₂)PhCO-	Me-	H	A.1	0.6	0.222	-0.26	0.348	-1.09
2-22	p-(PhCH ₂ SO ₂)PhCO-	Et-	H	A.1	0.3	0.523	0.72	0.473	0.49
2-25	PhCO-	Me-	2-Me	A.1	770	-2.886	-3.11	0.052	-0.42
2-26	PhCO-	Me-	3-Me	A.1	145	-2.161	-3.1	0.052	-1.76
2-27	PhCO-	Me-	4-Me	A.1	41000	-4.613	-3.01	0.052	3.02
2-28	PhCO-	Me-	2-NO ₂	A.1	14000	-4.146	-2.87	0.05	2.41
2-29	PhCO-	Me	3-NO ₂	A.1	370	-2.568	-2.95	0.05	-0.72
2-30	PhCO-	Me-	4-NO ₂	A.1	3300	-3.519	-2.93	0.05	1.11

Qualité de l'ajustement

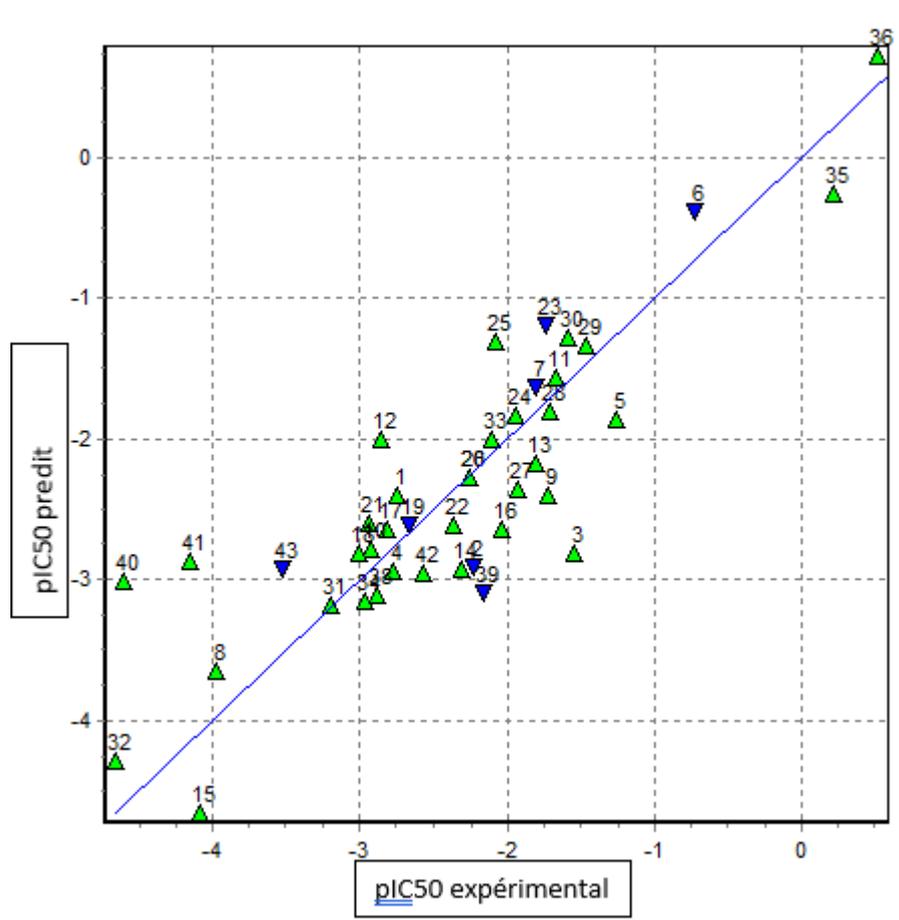


Figure 23 Droites d'ajustement pour le modèle 01

Les valeurs prédites ACHE pour les composés dans les ensembles de calibrage et de validation externe en utilisant l'équation (24) ont été présentées en fonction des valeurs expérimentales dans la (figure 23). D'après le tableau 4 et la (figure 23), les valeurs prédites pour les pIC50 sont en bon accord avec celles des valeurs expérimentales.

Le domaine d'application

Comme le montre le diagramme de Williams de la (figure 24), la majorité des composés de l'ensemble de données se situe dans le domaine d'application, à l'exception des composés 21, 36 qui dépassent le seuil $h^* = 0,43$.

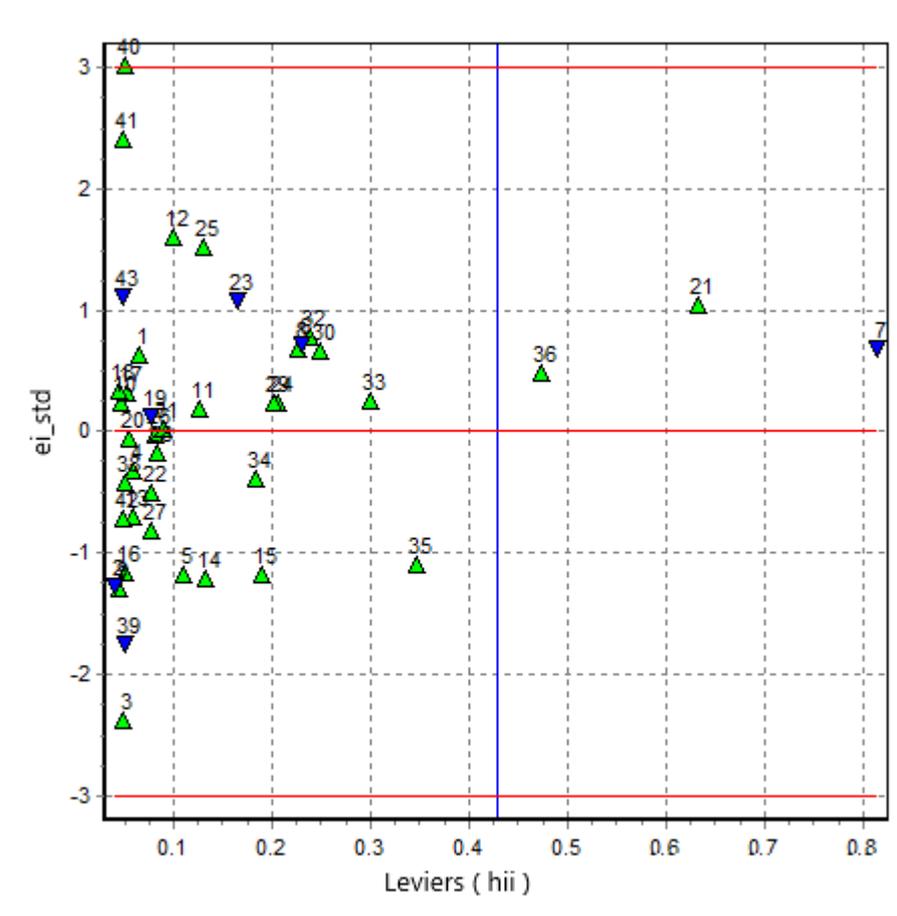


Figure 24: le domaine d'applications chimiques des composés

Test de randomisation

Le modèle a été validé en utilisant un test de randomisation (figure 25). Il est clair que les statistiques obtenues pour les vecteurs modifiés de (pIC50) sont inférieures à celles du modèle réel (cercle bleu et vert). Ceci permet d'assurer que le modèle établi n'est pas dû au hasard.

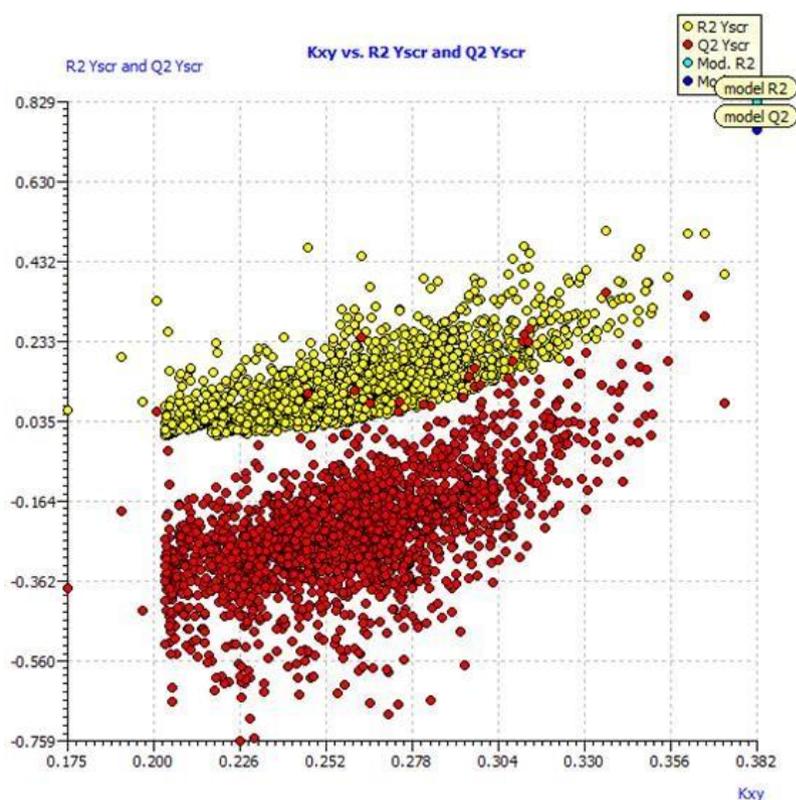


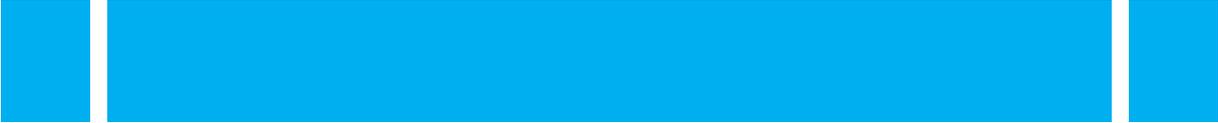
Figure 25 Test de randomisation du modèle AchE

Conclusion

Nous avons développé une équation utile de QSAR qui relie des descripteurs théoriques au pIC50 expérimental de 42 molécules.

Pour chaque composé plus de 1500 descripteurs sont calculés par le logiciel dragon, et l'ensemble des molécules a été divisé en 2 parties « training » et « test » avec un choix aléatoire, puis les meilleurs descripteurs ont été sélectionnés par « algorithme génétique » de Mobydigs.

Le model obtenu à une qualité statistique élevée et fiable et de faible erreur de prédiction, on peut alors conclure que pour cet ensemble de données, les combinaisons des techniques de modélisation ont comme conséquence une amélioration des modèles linéaires.



***RELATIONS
STRUCTURE ACTIVITE
DES COX2***



8 Application de l'approche QSAR dans la modélisation de la Cox-2

La concentration d'inhibition (IC_{50}) d'inhibiteur cyclooxygénase-2 des dérivés 1-Phenyl-1H-imidazole a été modélisée par régression multilinéaire (MLR). Le modèle a été validé par validations interne et externe.

Origine des données

Les valeurs de l'activité Cox-2 (IC_{50}) rapportées dans [33-34], ont été utilisées pour l'analyse des relations structure / propriétés.

9 Résultats et discussion

Matrice de corrélation

Tableau 5 matrice de corrélation pour le modèle 02

	ATS6v	nSO2N
nSO2N	0.076	
ALOGP	0.761	-0.184

Equation et analyse de régression :

Le modèle QSAR sélectionné est représenté par l'équation :

$$pIC_{50} = (\pm 3.25)20.85 + AlogP (\pm 0.12)1.43 - (\pm 1.08)7.95 ATS6v + (\pm 0.18)1.63 nSO2N \quad (\text{eq 25})$$

Avec les valeurs des paramètres statistiques ci-après :

Tableau 6 les valeurs essentielles du modèle 02

R2	Q2	Q2ext	R2adj	SDEP	SDEPext	Kx	Ky	F	S
80.05	75.46	74.93	78.77	0.499	0.504	40.02	45.72	62.844	0.4687

Tableau 7 définition des descripteurs moléculaires

Descripteur	Classe	Définition
ATS6v	Volume 2D autocorrelations	Autocorrélation de Broto-Moreau du décalage 6 (fonction logarithmique) pondérée par van der Waals
nSO2N	Functional group counts	nombre de sulfonamides (thio / dithio)
ALOGP	Molecularproperties	Partition Ghose-Crippen octanol-eau coeff. (logP)

Analyse des résidus et diagnostics d'influence

Les valeurs des résidus de prédiction standardisés, et les valeurs de hii sont présentées dans le tableau 8 dont les colonnes 7 et 8 reproduisent les valeurs expérimentales et calculées, prédites de pCIC50 des composés considérés.

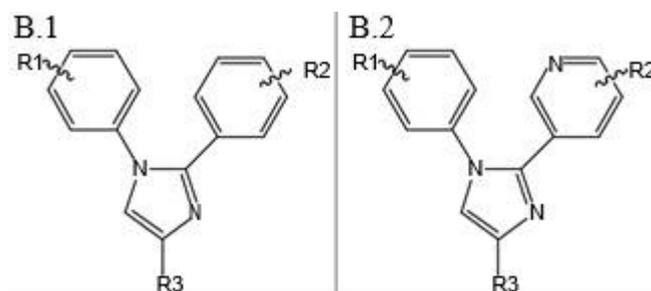


Figure 26 template des molécules

Tableau 8 Valeurs des pCIC50, observés (pic50) et calculés, prédites (pred y), des résidus standardisés (R1. R2. R3) ainsi que des leviers (hat) (std.Err.Pred).

Name	R1	R2	R3	Famille	Ic50	Pic50	Pred y	Hat	Std.Err.Pred.
3-5	-4-SO ₂ Me	-4-Cl	-Me	B.1	0.24	0.620	1.43	0.214	1.94
3-6	-4-SO ₂ Me	-4-Cl	-CF ₃	B.1	0.11	0.959	1.34	0.085	0.85
3-10	-4-SO ₂ Me	-H	-CF ₃	B.1	0.12	0.921	0.75	0.064	-0.39
3-12	-4-SO ₂ Me	-4-OMe	-CF ₃	B.1	0.57	0.244	0.25	0.027	0.01
3-13	-4-SO ₂ Me	-4-NHMe	-CF ₃	B.1	1.47	-0.167	-	0.025	0.2
3-14	-4-SO ₂ Me	-4-NMe ₂	-CF ₃	B.1	0.7	0.155	0.16	0.025	0.01
3-15	-4-SO ₂ Me	-4-SMe	-CF ₃	B.1	0.16	0.796	0.83	0.044	0.08
3-17	-4-SO ₂ Me	-4-SO ₂ Me	-CF ₃	B.1	5.7	-0.756	2.01	0.062	-0.32
3-18	-4-SO ₂ NH ₂	-4-Cl	-CF ₃	B.1	0.01	2.000	1.47	0.145	0.03
3-20	-4-SO ₂ NH ₂	-H	-CF ₃	B.1	0.04	1.398	0.86	0.139	0.16
3-22	-4-SO ₂ Me	-3-Cl	-CF ₃	B.1	0.06	1.222	0.08	0.045	-0.8
3-27	-4-SO ₂ Me	-3-OMe	-CF ₃	B.1	0.35	0.456	0.27	0.025	-1.16
3-28	-4-SO ₂ Me	-3-SMe	-CF ₃	B.1	0.35	0.456	0.95	0.034	-0.41
3-29	-4-SO ₂ Me	-3-CH ₂ OMe	-CF ₃	B.1	68.1	-1.833	0.32	0.104	1.99
3-30	-4-SO ₂ Me	-3-NMe ₂	-CF ₃	B.1	3.2	-0.505	0.49	0.046	0.41
3-31	-4-SO ₂ Me	-3-NHMe	-CF ₃	B.1	0.92	0.036	0.91	0.038	-1.15
3-32	-4-SO ₂ Me	-3-NH ₂	-CF ₃	B.1	5.89	-0.770	0.36	0.053	-0.3
3-33	-4-SO ₂ Me	-3-NO ₂	-CF ₃	B.1	0.58	0.237	0.91	0.034	-1.29

3-38	-4-SO ₂ Me	-2-Cl	-CF ₃	B.1	0.9	0.046	0.64	0.045	1.89
3-40	-4-SO ₂ Me	-2-Me	-CF ₃	B.1	0.8	0.097	0.69	0.034	1.17
3-42	-4-SO ₂ NH ₂	-2-F	-CF ₃	B.1	0.1	1.000	0.21	0.045	-0.69
3-43	-4-SO ₂ NH ₂	-2-Me	-CF ₃	B.1	0.2	0.699	0.99	0.114	0.52
3-44	-4-SO ₂ Me	-4-OMe-3-F	-CF ₃	B.1	0.15	0.824	0.36	0.025	-1.33
3-46	-4-SO ₂ Me	-4-SMe-3-Cl	-CF ₃	B.1	0.04	1.398	0.1	0.062	-0.91
3-47	-4-SO ₂ Me	-4-NMe ₂ -3-Cl	-CF ₃	B.1	0.32	0.495	1.19	0.049	-0.3
3-49	-4-SO ₂ Me	-4-NHMe-3-Cl	-CF ₃	B.1	0.66	0.180	0.77	0.037	-0.18
3-50	-4-SO ₂ Me	-4-Me-3-Cl	-CF ₃	B.1	0.03	1.523	0.51	0.067	-0.74
3-52	-4-SO ₂ Me	-3-Me-4-F	-CF ₃	B.1	0.17	0.770	0.31	0.04	-0.01
3-54	-4-SO ₂ Me	-3-OMe-4-Cl	-CF ₃	B.1	0.25	0.602	0.83	0.035	-0.2
3-55	-4-SO ₂ Me	-3-NMe ₂ -4-Cl	-CF ₃	B.1	1.04	-0.017	0.98	0.053	0.72
3-57	-4-SO ₂ Me	-3,4-F	-CF ₃	B.1	0.12	0.921	0.78	0.047	-0.2
3-58	-4-SO ₂ Me	-3,4-Me	-CF ₃	B.1	0.33	0.481	1.66	0.052	1.09
3-61	-4-SO ₂ Me	-3-OMe-5-F	-CF ₃	B.1	0.96	0.018	1.81	0.032	-0.26
3-66	-4-SO ₂ NH ₂	-4-OMe-3-F	-CF ₃	B.1	0.03	1.523	1.52	0.059	-1.64
3-69	-4-SO ₂ NH ₂	-4-SMe-3-Cl	-CF ₃	B.1	0.01	2.000	0.8	0.129	-0.77
3-70	-4-SO ₂ NH ₂	-4-Me-3-Cl	-CF ₃	B.1	0.003	2.523	0.14	0.128	-1.62
3-71	-4-SO ₂ NH ₂	-3-OMe-4-Cl	-CF ₃	B.1	0.02	1.699	0.24	0.116	-1.28
3-72	-4-SO ₂ NH ₂	-3,4-F	-CF ₃	B.1	0.03	1.523	0.23	0.118	0.02
3-73	-4-SO ₂ NH ₂	-3-Me-5-Cl	-CF ₃	B.1	0.04	1.398	1.41	0.131	0.28
3-75	-4-SO ₂ NH ₂	-3-OMe-5-Cl	-CF ₃	B.1	0.46	0.337	0.52	0.139	1.07
3-79	-4-SO ₂ Me	-4-OMe-3,5-Me	-CF ₃	B.1	0.72	0.143	0.39	0.079	-0.02
3-80		-4-Ome-2.5-Me	-CF ₃	B1	12.2	-1.086	-0.5	0.079	2.95
3-87	-4-SO ₂ Me	4-Cl	-CHF ₂	B.1	0.61	0.215	0.06	0.041	-0.97
3-88	-4-SO ₂ Me	-4-Cl	-CH ₂ F	B.1	0.41	0.387	0.06	0.163	2.39
3-89	-4-SO ₂ Me	-4-Cl	-CHO	B.1	1.6	-0.204	0.36	0.079	1.6
3-90	-4-SO ₂ Me	-4-Cl	-CN	B.1	0.23	0.638	0.29	0.06	-0.54
3-91	-4-SO ₂ Me	-4-Cl	-CO ₂ Et	B.1	5.7	-0.756	0.64	0.047	0.56
3-96	-4-SO ₂ Me	-4-Cl	-Ph	B.1	0.24	0.620	0.38	0.139	-1.65
3-99	-4-SO ₂ Me	-4-Cl	-CH ₂ OMe	B.1	3.72	-0.571	1.47	0.032	1.12
3-100	-4-SO ₂ Me	-4-Cl	-CH ₂ OH	B.1	8.35	-0.922	0.75	0.063	1.9
3-101	-4-SO ₂ Me	-4-Cl	-CH ₂ SMe	B.1	0.32	0.495	0.03	0.026	-0.3
3-103	-4-SO ₂ Me	-4-Cl	-CH ₂ CN	B.1	1.54	-0.188	0.38	0.028	-0.23
4-25	-4-SO ₂ Me	-H	-CF ₃	B.2	1.69	-0.228	0.01	0.058	-0.91

4-36	-4-SO ₂ Me	-2-Me	-CF ₃	B.2	9.6	-0.982	0.24	0.065	-0.22
4-40	-4-SO ₂ Me	-6-OMe	-CF ₃	B.2	1.2	-0.079	-	0.033	-0.66
4-41	-4-SO ₂ Me	-5-OMe	-CF ₃	B.2	37.6	-1.575	-1.9	0.101	0.24
4-42	-4-SO ₂ Me	-5-Br	-CF ₃	B.2	0.95	0.022	-	0.048	-1.68
4-43	-4-SO ₂ NH ₂	-H	-CF ₃	B.2	0.44	0.357	1.43	0.171	-0.77
4-44	-4-SO ₂ NH ₂	-2-Me	-CF ₃	B.2	2.8	-0.447	1.34	0.187	0.16
4-49	-4-SO ₂ NH ₂	-5-Br	-CF ₃	B.2	0.34	0.469	0.25	0.16	-1.13
4-58	-4-SO ₂ Me	-H	-CHF ₂	B.2	20.7	-1.316	0.07	0.162	3.62
4-59	-4-SO ₂ Me	-H	-CN	B.2	24.4	-1.387	0.16	0.146	-0.21
4-62	-4-SO ₂ Me	-H	-CH ₂ OH	B.2	93.4	-1.970	-0.9	0.207	0.18
4-63	-4-SO ₂ NH ₂	-H	-CHF ₂	B.2	1.83	-0.262	2.01	0.098	-1.36

Qualité de l'ajustement

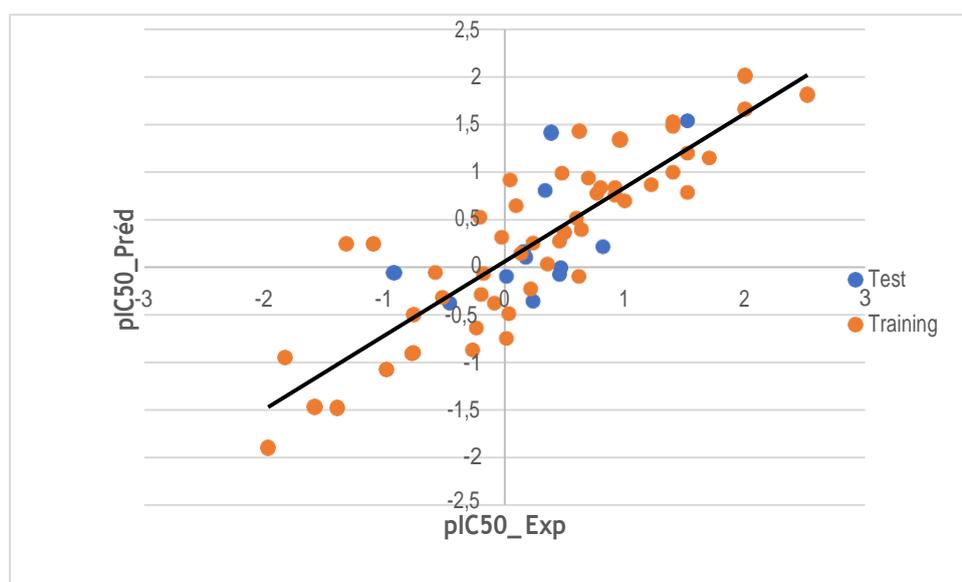


Figure 27 Droites d'ajustement pour le modèle 02

Les valeurs prédites de pIC50 pour les composés dans les ensembles de calibrage et de validation externe en utilisant l'équation (25) ont été présentées en fonction des valeurs expérimentales dans la figure 2. D'après le tableau 8 et la figure 27, les valeurs prédites pour la solubilité sont en bon accord avec celles des valeurs expérimentales.

Le domaine d'application

Le Domaine d'application du modèle a été analysé à l'aide du diagramme de Williams (figure 28). La Figure montre clairement la bonne prédictivité pour les ensembles d'apprentissage et de validation. Tous les composés de validation sont dans le domaine chimique, ce qui suggère qu'il n'y a pas de valeurs aberrantes (en X et/ou en Y) et que la prédiction du modèle est fiable.

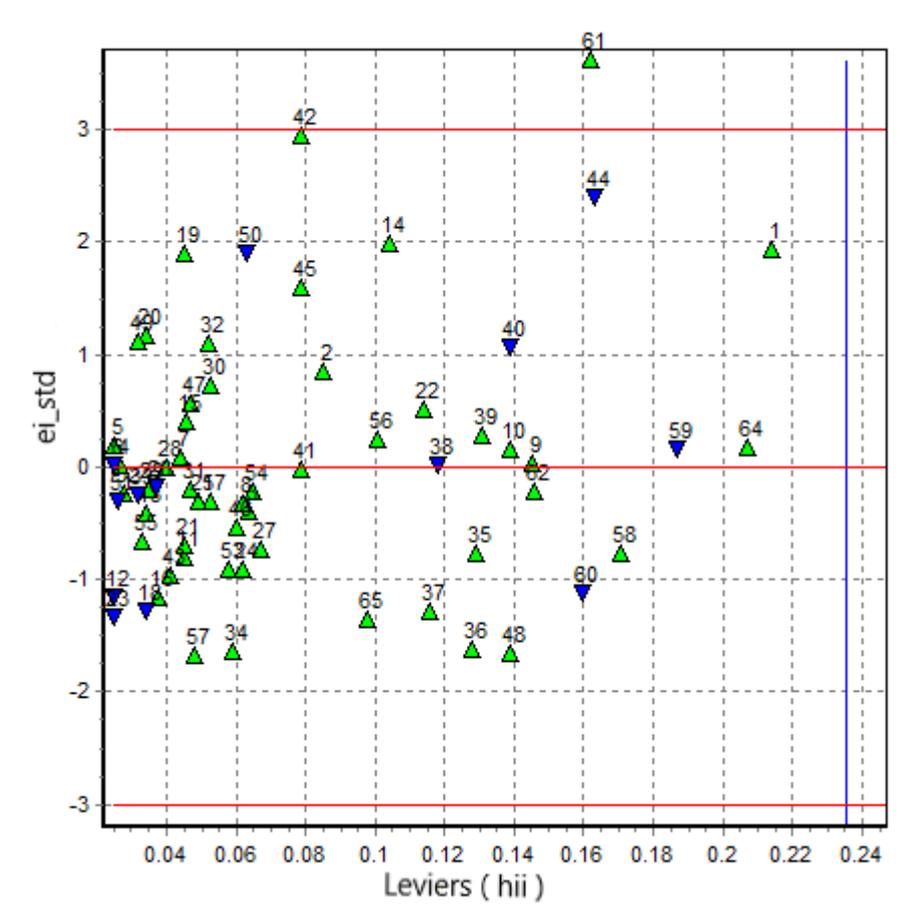


Figure 28: le domaine d'applications chimique des composés

Test de randomisation

Le modèle a été validé en utilisant un test de randomisation (figure 29). Il est clair que les statistiques obtenues pour les vecteurs modifiés de (pIC50) sont inférieures à celles du modèle réel (cercle bleu et vert). Ceci permet d'assurer que le modèle établi n'est pas dû au hasard.

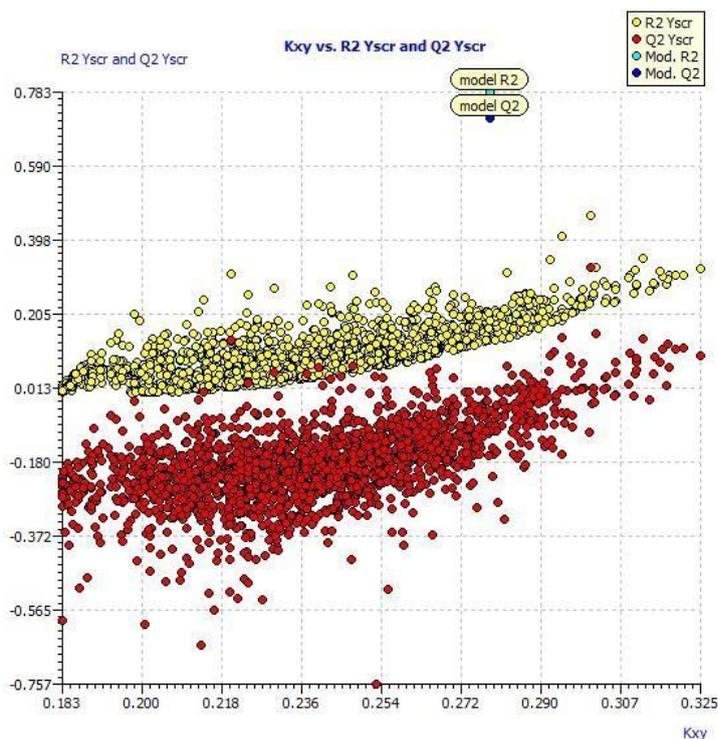


Figure 29 test de randomisation du modèle COX2

Conclusion

Les valeurs élevées des paramètres statistiques (Q2 ...) montrent la fiabilité du modèle proposé et sa possession d'une capacité prédictive acceptable, pouvant ainsi être utilisé pour estimer les pIC50 sans données expérimentales connues.

Comparativement aux résultats obtenus au paravant par d'autres auteurs, le model QSAR présenté dans notre travail a montré de meilleurs résultats de prédictions.

Les valeurs de pIC50 expérimentales ont été corrélées avec les descripteurs de structures moléculaires par l'approche QSAR en exploitant le logiciel MobyDigs et un model bidimensionnelle a été calculé.

Conclusion générale

Parmi les grandes voies d'étude et de conception de médicaments par modélisation moléculaire est celle qui se fonde sur la relation entre l'activité et la structure moléculaire.

Dans ce contexte, l'objectif de cette thèse était de développer des modèles QSAR Fiabiles pour la prédiction des pIC50 de deux familles d'inhibiteurs (AchE et COX2).

Un grand nombre de descripteurs moléculaires a été calculé (Descripteurs constitutionnels, électroniques, topologiques, géométriques, physicochimiques,...).

La méthodologie basée sur la MLR, a été utilisée principalement dans la prédiction.

Les principales techniques de validation ont été utilisées (les tests statistiques standards, la validation interne, la validation externe, la randomisation des Y, les domaines d'applications...).

La concentration d'inhibition de la croissance 50 % (IC50) des deux familles d'inhibiteurs (AchE et COX2) un modèle a été développé par la méthode MLR, pour chaque famille.

Des validations rigoureuses internes et externes ont été utilisées pour juger la stabilité, la justesse et la capacité prédictive des modèles obtenus pour les deux familles.

La qualité de l'ajustement des modèles développés a été vérifiée en procédant à la représentation des valeurs calculées en fonction du celles observées. Le domaine d'application des modèles a été étudié à l'aide du diagramme de Williams, ce dernier fait ressortir parmi les composés de l'ensemble de calibrage et de validation les composés influents et aberrants.

Finalement, les objectifs principaux de notre travail ont été remplis, nous prévoyons : reprendre les mêmes bases de données et élaborer des modèles en utilisant d'autres méthodes telles que : le 3D QSAR et d'autres logiciels de calculs comme Qsarins ... afin de poursuivre notre chemin de recherche dans cette discipline.

Références bibliographiques

- [1] K BOUHEDJAR, Toxicité des phénols : Une étude QSAR basée sur la méthode préconisée par l'OCDE, université Annaba décembre 2010
- [2] A.C. Crum-Brown and T.R. Fraser, —On the Connection Between Chemical Constitution and Physiological Action, Part I: On the Physiological Action of the Salts of the Ammonium Bases, Derived from Strychnia, Brucia, Thebia, Codeia, Morphia, Nicotia, *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, 25, **1868**, 151–203;
- [3] A.C. Crum-Brown and T.R. Fraser, —On the Connection between Chemical Constitution and Physiological Action. Part II: On the Physiological Action of the Ammonium Bases derived from Atropia and Conia, *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, 25, **1869**, 693–739.
- [4] J.A. Pople, —Nobel Lecture: Quantum chemical models, *Review of Modern Physics*, 71, **1999**, 1267–1274.
- [5] H. Meyer, La théorie de la narcose alcoolique. Premier message Quelle propriété des anesthésiques est la cause de leur effet narcotique, *Archive pour Pathologie experimental et pharmacologie*, **1899**, 109–118.
- [6] E. Overton et G. Fischer, Études sur l'anesthésie en même temps, une contribution à la pharmacologie générale, **1901**.
- [7] H. Fühner et E. Neubauer, "Immolation par des séries homologues de substances", *Archive pour Pathologie experimental et pharmacologie*, 333-345, 1907.
- [8] C. Hansch and T. Fujita, — ρ - σ - π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure, *Journal of the American Chemical Society*, 86(8), **1964**, 1616–1626.
- [9] C. Hansch and E.J. Lien, —Structure-activity relationships in antifungal agents. A survey, *Journal of Medicinal Chemistry*, 14(8), **1971**, 653–670.

- [10] S.Y. Tham and S. Agatonovic-Kustrin, —Application of the artificial neural network in quantitative structure-gradient elution retention relationship of phenylthiocarbamyl amino acids derivatives, *Journal of Pharmaceutical and Biomedical Analysis*, 28(3), **2002**, 581-590
- [11] R.D. Cramer, D.E. Patterson, and J.D. Bunce, —Comparative molecular field analysis (CoMFA): Effect of shape on binding of steroids to carrier proteins, *Journal of the American Chemical Society*, 110(18), **1988**, 5959-5967.
- [12] G. Klebe, U. Abraham, and T. Mietzner, —Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity, *Journal of Medicinal Chemistry*, 37(24), **1994**, 4130-4146.
- [13] S Chtita, Modélisation de molécules organiques hétérocycliques biologiquement actives par des méthodes QSAR/QSPR. Recherche de nouveaux médicaments, **2017**
- [14] Guillaume F, Développement de modèles QSPR pour la prédiction des propriétés d'explosibilité des composés nitroaromatiques , **2010**
- [15] R. Tomassone, E. Lesquoy, C. Miller. La régression : nouveaux regards sur une ancienne méthode statistique. Masson, INRA (1983).
- [16] R. Wehrens, H. Putter, L. M. C. Buydens. Chemom. Int. Lab. Syst., 54, **2000**, 35- 52
- [17] L. Eriksson, J. Jaworska, A. P. Worth, M. T. D. Cronin, R. M. McDowell, P. Gramatica. Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification and Regression Based QSARs. Environmental Health Perspectives III, **2003**, 1361-1375 ().
- [18] N. R. Draper, H. Smith. Applied Regression Analysis, 3rd ed; Wiley Series in Probability and Statistics (**1998**).

- [19] N BOUARRA, Etudes QSPR des propriétés contrôlant l'évolution de quelques HAP dans l'environnement **2018**
- [20] Golbraikh, A., M. Shen, Z. Xiao, Y.-D. Xiao, K.-H. Lee, and A. Tropsha, Rational selection of training and test sets for the development of validated QSAR models. *Journal of Computer-Aided Molecular Design*, **2003**. 17(2-4): p. 241-253.
- [21] Tropsha A, Gramatica P, Gombar V K, The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSAR Models. *QSAR. Comb. Sci.* **2003**. 22 69.
- [22] Lauwerys RR. Toxicologie industrielle et intoxications professionnelles : Insecticides organophosphorés. —4e édition. Paris:Masson, **1999** ;pp. 774-782.
- [23] Birks, J. S. Cholinesterase inhibitors for Alzheimer's disease. *Cochrane Libr.* (2006).
- [24] Jackson, C. J., Oakeshott, J. G., Sanchez-Hernandez, J. C. & Wheelock, G. E. in *Anticholinesterase Pestic. Metab. Neurotox. Epidemiol.* (John Wiley & Sons, **2011**).
- [25] Clemence R, Application de la démarche de drug-design pour la conception de nouveaux médicaments vétérinaires contre le parasite *Varroa destructor* (AcariVarroidae), **2018**
- [26] J. R. Vane, Y. S. Bakhle¹, and R. M. CYCLOOXYGENASES 1 AND 2 *Annu. Rev. Pharmacol. Toxicol.* **1998**. 38:97–120
- [27] Kadrilege et Datacross. CHEMDRAW 18.0, LE DESSIN MOLÉCULAIRE PAR EXCELLENCE [enligne] <https://ritme.com/fr/logiciels/chemdraw-chemoffice/> (consulté le 03/07/2019)
- [28] MakoLab, Hypercube HyperChem Professional 8.0 [enligne] <http://www.hyper.com/?tabid=360> (consulté le 03/07/2019)

- [29] Talete s.r.l. Dragon 7 available now [enligne]
http://www.talete.mi.it/products/dragon_description.htm (consulté le 03/07/2019)
- [30] Sugimoto, H.; Tsuchiya, Y.; Sugumi, H.; Higurashi, K.; Karibe, N. et al. Synthesis and structure-activity relationships of acetylcholinesterase inhibitors: 1-benzyl-4-(2-phthalimidoethyl) piperidine and related derivatives. *J. Med. Chem.* **1992**, 35, 4542-4548.
- [31] Sugimoto, H.; Tsuchiya, Y.; Sugumi, H.; Higurashi, K.; Karibe, N. et al. Novel piperidine derivatives. Synthesis and anti-acetylcholinesterase activity of 1-benzyl-4-[2-(N-benzoylamino) ethyl] piperidine derivatives. *J. Med. Chem.* **1990**, 33, 1880-1887.
- [32] Sugimoto, H.; Iimura, Y.; Yamanishi, Y.; Yamatsu, K. Synthesis and structure-activity relationships of acetylcholinesterase inhibitors: 1-benzyl-4- [(5,6-dimethoxy-1-oxoindan-2- Yl) methyl] piperidine hydrochloride and related compounds. *J. Med. Chem.* **1995**, 38, 4821-4829.
- [33] Khanna, I. K.; Weier, R. M.; Yu, Y.; Xu, X. D.; Koszyk, F. J. et al. 1,2-diarylimidazoles as potent, cyclooxygenase-2 selective, and orally active anti-inflammatory agents. *J. Med. Chem.* **1997**, 40, 1634-1647.
- [34] Khanna, I. K.; Yu, Y.; Huff, R. M.; Weier, R. M.; Xu, X. D. et al. Selective cyclooxygenase-2 inhibitors: Heteroaryl modified 1,2-diarylimidazoles are potent, orally active anti-inflammatory agents. *J. Med. Chem.* **2000**, 43, 3168-3185.



Annexe



ANNEXE

Coefficient de partage Octanol/eau (logP):**I- Définition :**

Le transport, le passage à travers les membranes, la bioaccumulation ou encore l'activité pharmacologique d'une molécule peuvent être conditionnés par son partage entre une phase lipidique et une phase aqueuse, c'est-à-dire son caractère hydrophile. Celui-ci peut être quantifié par le coefficient de partage octanol-eau, noté logP, qui mesure la solubilité différentielle d'un soluté dans ces deux solvants non miscibles [35].

$$\log P = \log \left(C_{\text{octanol}} / C_{\text{H}_2\text{O}} \right) \quad (\text{eq. 26})$$

C_{octanol} et $C_{\text{H}_2\text{O}}$ sont les concentrations du soluté dans l'octanol et l'eau.

La lipophilie, échelle continue, peut avoir aussi bien une valeur positive qu'une valeur négative. Donc, les composés qui ont les valeurs de $P > 1$ ou $\log P > 0$ sont lipophiles, et les composés qui ont les valeurs de $P < 1$ ou $\log P < 0$ sont hydrophiles.

La lipophilie reflète la balance entre deux classes d'interactions intermoléculaires :

La première est proportionnelle à la taille du soluté, et la seconde est associée aux groupes fonctionnels. Ainsi, la lipophilie peut être décomposée en deux contributions principales (Carrupt *et al*, 1997) [36] :

$$\text{Lipophilie} = \text{Hydrophobie} - \text{Polarité}. \quad (\text{eq. 27})$$

Comme la détermination expérimentale de $\log P$ est souvent délicate, de nombreuses méthodes ont été développées pour prédire ce paramètre. Parmi celles-ci, les plus couramment utilisées sont celles qui sont basées sur l'additivité des contributions lipophiles de fragments moléculaires ou atomiques. Elles ne tiennent compte ni de l'aspect tridimensionnel des molécules ni de leur stéréochimie. Pour pallier ce défaut, d'autres techniques dites « Tridimensionnelles » ont été proposées.

Prédiction du coefficient de partage octanol/eau :

Cette propriété physico-chimique peut être mesurée, mais ces mesures sont généralement longues et coûteuses. Par conséquent, différentes méthodes de prédiction du logP ont été mises au point, et il existe un nombre important de logiciels de prédiction de cette propriété. Ceux-ci s'appuient aussi bien sur des méthodes de contribution de groupes

(ACD/LogP, KOWWIN, CLOGP ...) que sur des régressions multilinéaires à partir de descripteurs (VLogP) ou sur des réseaux de neurones (AUTOLogP).

Dans les bases de données, on peut trouver près de 30 000 structures organiques dont les valeurs de logP expérimental sont disponibles, ce nombre est cependant négligeable comparé au nombre des structures pour lesquelles les données de log P sont fortement demandées. En effet, il y a un intérêt continu pour le développement de méthodes de calculs de log P à partir de la structure moléculaire.

La première méthode de calcul de logP a été développée par Hansch et Fujita (système π) [37]. Les imperfections relevées ont conduit Rekker à développer la première approche de contribution fragmentaire [38-39]. Puisque la définition de fragment peut être ambiguë, (Broto et al) [40], suivis par d'autres, ont développé des systèmes de calcul basés sur des contributions atomiques. Toutes les méthodes qui divisent les molécules en sous-structures, sont appelées approches de sous-structure. L'utilisation exclusive des sous-structures atomiques caractérise des méthodes de contribution d'atomes, tandis que l'utilisation additionnelle de plus grands groupes est typique pour des méthodes fragmentaires. L'addition des contributions des sous-structures conduit finalement à la valeur de logP.

En revanche, les approches basées sur la molécule entière emploient, les potentiels moléculaires de lipophilie (PML), les indices topologiques, ou les propriétés moléculaires pour quantifier log P.

Tableau : Classification des logiciels de calcul de logP.

Approches par Sous-Structure	Approches par Molécule-Entière
<i>Méthodes Fragmentales</i>	<i>Potentiels Moléculaires de Lipophilie (PML)</i>
CLOGP Cf-SYBYL AB/LogP	CLIP HINT
ACD/LogP KOWWIN	<i>Indices Topologiques</i>
<i>Méthode des Contributions d'Atomes</i>	MLOGP AUTOLOGP
TSAR, MOLCAD	SciLogP
ALOGP XLOGP	<i>Propriétés Moléculaires</i>
	BLOGP QLOGP

Approche CLOGP :

Hansch et Leo [41-42] ont développé un système de fragmentation connu sous le nom de CLOGP, basé sur les principes du "constructionisme". Les valeurs des fragments de base ont été déduites à partir de données de log P mesurées, d'un petit ensemble de molécules simples telles que l'hydrogène et le méthane. Cette méthode a été adaptée la première fois pour l'usage informatique par Chou et Jurs [43] ; elle a été lancée sur le marché sous l'appellation CLOGP.

Approche Ghose/Crippen :

Le groupe de Crippen [44,45] a développé une " *procédure basée sur les contributions atomiques pure* " qui applique exclusivement les contributions d'atomes et évite les facteurs de correction (Eq. 24), où N_k est l'occurrence du k ème type d'atome).

$$\log P = \sum a_k \cdot N_k \quad (\text{eq. 24})$$

Le carbone, l'hydrogène, l'oxygène, l'azote, le soufre, et les halogènes sont classés dans 110 types d'atome ; après plusieurs révisions, le nombre de classifications d'atome qui arrive jusqu'à 120 [44] est obtenu à partir d'un ensemble de calibration de 893 structures. L'hydrogène et les halogènes sont classés par l'état d'hybridation et d'oxydation du carbone auquel ils sont liés ; les atomes de carbone sont classés selon leur état d'hybridation et la nature chimique des atomes voisins. La complexité de cette procédure de classification est certifiée par un total de 44 types de carbone. L'approche de Ghose/Crippen qui utilise largement la contribution d'atomes, a été utilisée par le logiciel qu'on a employé pour le calcul de l'hydrophilie notée ALOGP par les développeurs de ce logiciel.

Références de l'annexe

- [35] P.J. Taylor, Hydrophobic Properties of Drugs. In *Quantitative Drug Design. Vol. 4* (Ramsden, C.A., ed.), Pergamon Press, Oxford (UK), 241-294 (1990).
- [36] F. A. Carrupt, B. Testa, P. Gaillard, Computational Approaches to Lipophilicity: Methods and Applications. In *Reviews in Computational Chemistry, Vol. 11* Wiley-VCH, New York (NY), 241-315 (1997).
- [37] ACD/LogP, v. 10, Advanced Chemistry Development, Inc., Toronto, ON, Canada, www.acdlabs.com, (2007).

- [38] C. G. Nys, R. F. Rekker, *Chim. Ther.*, 8, 521-535 (1973).
- [39] R. F. Rekker, The hydrophobic fragmental constant. Its derivation and application. A means of characterizing membrane systems. *Pharmacochem. Library Vol.1*, Elsevier, Amsterdam, (1977).
- [40] P. Broto.G, Moreau, C. vandycke, *Eur. J. Med. Chem.*, 19, 71-78 (1984).
- [41] A. J. Leo, P. Y. C. Jow, C. Silipo, C.Hansch, *J. Med. Chem.*, 18, 865-868 (1975).
- [42] A. J. Leo, *Chem. Rev.*, 93, 1281- 1306 (1993).
- [43] J. T. Chou, P. C. Jurs, *J. Chem, Inf Comput. Sci.*, 19, 172-178 (1979).
- [44] A. K. Ghose, M. CRIPPEN, *J. Comp. Chem.*, 7, 565-577 (1986).
- [45] V. N. Viswanadhaan, A.K. Ghose, *Inf. Comput. Sci.*, 29, 163-172 (1989).

Année universitaire : 2018/2019

Présenté par : KHAINNAR Mehdi et

AIMEUR Oussama

Etude des relations Quantitatives Structure-Activité (QSAR) de plusieurs familles de composés sur des données biologiques

Mémoire de fin de cycle présenté en vue de l'obtention du diplôme de Master en Biochimie appliquée.

Résumé

Le travail présenté dans ce mémoire a pour objectif d'élaborer des modèles QSAR fiables, stables et prédictifs pour la prédiction des propriétés pIC_{50} (concentration inhibitrice médiane) d'inhibiteurs de deux familles de molécules (AchE et COX2)

Le logiciel de modélisation moléculaire ChemDraw a été utilisé pour représenter les molécules, puis à l'aide de la méthode semi-empirique PM3 dans le logiciel HyperChem les géométries finales ont été obtenues. Différents descripteurs moléculaires sont calculés à l'aide du logiciel Dragon (V.5.5).

Des modèles QSAR ont été développés pour la prédiction des pIC_{50} de divers ensembles de d'inhibiteurs. Des approches basées sur la régression linéaire multiple (RLM), conduisent à des modèles de qualités différentes. Les algorithmes génétiques (GA), ont été associés pour sélectionner les descripteurs les plus importants,

L'approche hybride algorithme génétique/ régression multilinéaire a été utilisée pour modéliser les pIC_{50} .

Les modèles établis, ont été validés selon les cinq principes avancés par l'Organisation de Coopération et de Développement Économiques (OCDE). Le domaine d'application des modèles est étudié à l'aide du diagramme de Williams pour détecter les composés aberrants en X et/ou en Y.

Des validations rigoureuses internes et externes ont été considérées pour juger la stabilité et la capacité prédictive de ces modèles afin de combler les lacunes dans les données des pIC_{50} des molécules.

Mots clés : Acétylcholinestérase, Cyclooxygénase-2, QSAR/QSPR, MobyDigs, Descripteurs.

Laboratoire de recherche : laboratoire 18 de bio-informatique et biostatistiques au CRBt en collaboration avec le laboratoire de Biochimie à l'UFM CONSTANTINE 1

Jury d'évaluation :

Président du jury : BENSEGUENI A. (Professeur à UFM Constantine 1),

Examineur : MOKRANI E.H. (Maitre-assistant A à UFM Constantine 1),

Rapporteur : DEMS M.A.E. (Maître de recherche classe A au CRBt).

CO-ENCADREUR : BOUHEDJAR K (Chercheur Permanent au CRBt).

Date de soutenance : 09/07/2019