



**UNIVERSITE KASDI MERBAH  
OUARGLA**

N° d'ordre  
N° de série

**FACULTE DES SCIENCES ET  
TECHNOLOGIE**

**Département de Génie des Procédés**

**Thèse**

*Présenté pour l'obtention du diplôme de*

**DOCTORAT**

**Spécialité : Génie des Procédés Option : Génie des Procédés**

**Par : SOUYEI Belgacem**

**Thème**

**Modélisation de la Relation Structure Chimique /  
Coefficient de Partage Octanol/Eau: Application  
pour la génération de nouveaux pesticides**

**Soutenu publiquement le : .. / .. /2019**

**Devant le jury composé de :**

Mr.	D. ZERROUKI	Pr	Université d'Ouargla	Président
Mr.	A.KABOUCHE	Pr	Université d'Oum El Bouaghi	Examineur
Mr.	S. LAOUNI	M.C.A	Université d'El Oued	Examineur
Mr.	A. HADJ SEYD	M.C.A	Université d'Ouargla	Rapporteur

**Année universitaire : 2019 - 2020**

## *Dédicaces*

*Je dédie ce modeste travail aux esprits de mon défunt père et mère et de ma femme, mes enfants, mes frères et sœurs chacun par son nom, à toute ma famille et à tous mes amis.*

*Belgacem Souyei*

## Remerciements

Je tiens d'abord à remercier très sincèrement Monsieur HADJ SEYD.ABD EL KADER, Maître de conférence A à l'université de Ouargla, mon encadreur et directeur de thèse. J'ai eu vraiment le plus grand plaisir de travailler avec lui, durant la préparation de ma thèse de Doctorat. Je lui serai éternellement reconnaissant de la confiance qu'il m'a témoignée, et du soutien qu'il m'a porté, pour sa plus grande disposition et ses conseils qui m'ont permis de progresser dans les meilleures conditions pendant mes travaux de recherche. Qu'il soit ici vivement remercié.

Je tiens aussi à remercier toute l'équipe du Laboratoire de recherche VTRS de l'université d'El Oued pour l'accueil qui m'a été réservé, J'ai trouvé au sein du Laboratoire une ambiance particulièrement favorable pour réaliser la partie pratique de mes travaux.

J'adresse des remerciements en particulier au professeur KORICHI .M pour ses efforts, conseils et ses précieux conseils dont j'ai bénéficié dans ce travail.

Je remercie vivement Monsieur ZERROUKI Djamal Professeur à l'université de Ouargla qui m'a fait l'honneur de présider le jury.

Je suis également très reconnaissant envers les membres de jury, le Professeur : KABOUCHE Azedine., de l'université de Oum El Bouaghi .Dr. LAOUINI Salah, de l'université d'El Oued, qui m'ont fait le plaisir d'examiner et évaluer ce travail.

Je remercie vivement Dr . ZAIZ. F. de l'université d'El Oued et Dr . LABBI Y. de l'université d'El Oued pour leur aide et encouragement à la réussite de mon travail.

Mes remerciements vont encore à ceux qui ont participé de près ou de loin à la réalisation de ce travail, je cite en particulier ma femme et mes enfants pour leur soutien qu'ils m'ont apporté durant toute la période de réalisation de cette thèse.

تُعد الشراة اتجاه الشحوم الدهنية معلمة مهمة يستخدمها الكيميائيون في العديد من المجالات العلمية ، مثل اكتشاف مركبات كيميائية جديدة بما في ذلك الأدوية ومنتجات الصحة النباتية.

قصد توليد هياكل جديدة من المبيدات الحشرية، وهذا في ضوء الاستخدام (QSAR/QSPR-Inverse) هذا العمل يتمحور حول تطبيق طريقة المهم للغاية والواسع النطاق لهذه الطريقة في العديد من المجالات، سيما في المجال الزراعي. لهذا الغرض قمنا بتطبيق منهجية تحتوي على لبناء نموذج رياضي يربط الهياكل (QSAR/QSPR) الخاصة بتسع خطوات متتالية مفصلة تتضمن دراسة العلاقة الكمية للهياكل مع النشاط ( $\log_{ow}$ ) استخدمنا البصمات (التوقيعات) الذرية بينما الخاصية (n-octanol-eau)  $\log_{ow}$ . مبيد حشري بمعاملات توزيعها 091 البنويوية ل للقدرة التنبؤية لهذا (Test) مبيد حشري كمجموعة اختبار 01 مبيد حشري كمجموعة عمل لبناء النموذج و 031 هي المتغير التابع. أخذنا بدالة الواصفات المتمثلة في التوقيعات الذرية بتطبيق الطريقة الإحصائية وهي النحدار الخطي  $\log_{ow}$  النموذج. تمت نمذجة الخاصية (QSAR/QSPR) تحصلنا على معادلة رياضية Leave-one-out (Loo) cross validation . باستخدام التحقق من الصحة (MLR) المتعدد بصمة (توقيع) ذرية حيث: 01 تحتوي على

هذا  $Q^2 = 0.65617$  ,  $R^2 = 0.659273$  و  $S = 1.37297$  ,  $RMSE_{training} = 0.930182$  وجد: (Leave-one-out (Loo)) باستخدام طريقة (test) لمجموعة البناء أما لمجموعة الاختبار على التوالي. بالإضافة إلى كل ما سبق تم إنشاء (توليد) هياكل جديدة لمجموعة من المبيدات  $RMSE_{loo} = 1.0936$  ,  $q^2 = 0.605676$  التي يمكن إدراجها كموضوعات بحثية في المستقبل.

## Résumé

La Lipophilicité est un paramètre important utilisé par les chimistes dans plusieurs domaines scientifiques, telle que, la découverte de nouveaux composés chimiques notamment les médicaments et les produits phytosanitaires.

Le présent travail concerne l'application du problème QSAR / QSPR-inverse pour générer de nouvelles structures de pesticides; ceci est dû à leur utilisation extrêmement importante et répandue dans plusieurs domaines, en particulier dans le domaine agricole. Pour cette raison, nous avons mis en œuvre une méthodologie comprenant neuf étapes successives détaillées comprenant une étude de la relation quantitative structure-activité / propriété (QSAR / QSPR) pour développer un modèle reliant les structures de 190 composés pesticides à leurs coefficients de partage n-octanol-eau ( $\log_{ow}$ ). Nous avons utilisé les signatures atomiques uniques qui représentent la structure moléculaire et agissent comme des variables indépendantes, tandis que la propriété ( $\log_{ow}$ ) est la variable dépendante. Le modèle a été construit en utilisant 130 molécules en tant qu'ensemble de criblage, et la capacité prédictive a été testée avec 60 composés. La modélisation du  $\log_{ow}$  de ces composés en fonction des descripteurs de signatures a été établie par la régression linéaire multiple (MLR) à l'aide de la validation croisée (LOO). conséquence, une équation QSAR / QSPR avec 14 signatures atomiques a été obtenue avec un  $R^2 = 0.659273$ ,  $Q^2 = 0.65617$  et  $RMSE_{training} = 0.930192$ ,  $s = 1.37297$  pour l'ensemble d'apprentissage et dans la valeur définie de l'expérience de validation croisée Leave-one-out (LOO),  $q^2 = 0.605676$ ,  $RMSE_{LOO} = 1.0936$ , respectivement. En plus de tout ce qui précède, des nouvelles structures ont été créées pour une gamme de pesticides pouvant être incluses en tant que futurs sujets de recherche.

## Abstract

Lipophilicity is an important parameter used by chemists in several scientific fields, such as the discovery of new chemical compounds including drugs and phytosanitary products. The present work has focused on the application of the inverse-QSAR/QSPR problem for generating new structures of pesticides; this is in view of its extremely important and widespread use in several areas, particularly the agricultural field. For this reason, we implemented a methodology containing nine detailed successive steps that include a quantitative structure-activity/property relationship (QSAR/QSPR) study performed to develop a model that relates the structures of 190 pesticides compounds to their n-octanol-water partition coefficients ( $\log_{ow}$ ). We used the unique atomic signatures which represent the structures and acts as independent variables while the property ( $\log_{ow}$ ) as the dependent variable. The model was constructed using 130 molecules as training set, and predictive ability tested using 60 compounds. Modeling of  $\log_{ow}$  of these compounds as a function of the signatures descriptors was established by multiple linear regression (MLR) using (LOO) cross-validation. As a result, a QSAR/QSPR equation with 14 atomic signatures was hereby obtained with cides that can be included a  $R^2 = 0.659273$ ,  $Q^2 = 0.65617$  and  $RMSE_{training} = 0.930192$ ,  $s = 1.37297$  for the training set and in leave-one-out (LOO) cross-validation experiment set value,  $q^2 = 0.605676$ ,  $RMSE_{LOO} = 1.0936$  respectively. In addition to all of the above, new structures have been generated for a range of pesti as future search topics.

## Liste des figures

<b>Figure</b>	<b>Page</b>	
I-1	Mesure du coefficient de partage n-octanol /eau <b>logK<sub>ow</sub></b>	06
I-2	Relation de <b>logK<sub>ow</sub></b> avec d'autres propriétés physico-chimiques	07
I-3	Diagramme de la relation de <b>logK<sub>ow</sub></b> avec d'autres propriétés physico-chimiques	07
I-4	Méthode expérimentale de détermination de <b>logK<sub>ow</sub></b>	08
I-5	Méthode par agitation du flacon système de deux phases	09
I-6	Le principe de la méthode HPLC	10
I-7	Le principe de la méthode par agitation douce	11
I-8	Le principe de la méthode sur colonne	12
I-9	Exemples d'application de <b>logK<sub>ow</sub></b> dans le domaine pharmaceutique	20
I-10	Classification des pesticides	24
I-11	Concept d'optimisation	31
II-1	Illustre les signatures atomiques et moléculaires d'Ethéphon avec sa signature moléculaire au niveau-1	35
II-2	Méthodologie de développement des modèles QSPR	38
II-3	Schéma de l'algorithme I-QSPR	42
III-1	le fichier de format texte de l'insecticide « Thiram »	48
III-2	La signature moléculaire de Diméthoate	49
III-3	La signature moléculaire de Formothion	50
III-4	Equations de contrainte dérivées de Signatures de niveau-1 pour les deux composés d'ensemble de criblage : Diméthoate et Formation	53
VI-1	Impact des signatures atomiques de niveau-1 de pesticides sur les statistiques QSPR, tracées en fonction du nombre de variables indépendantes	64
VI-2	Les valeurs expérimentales par rapport aux prévisions pour l'équation QSPR basée sur le <b>logk<sub>ow</sub></b> avec <b>29</b> signatures atomiques de niveau-1	66
VI-3	Illustre la traduction de la structure qui possède la formule brute <b>C<sub>7</sub>H<sub>17</sub>O<sub>2</sub>PS<sub>4</sub></b> en utilisant la méthode développée par Faulon et al	74



## Liste des tableaux

Tableau		Page
I-1	valeurs des lipophiles de différents fragments	15
I-2	Termes correctifs de la formule de REKKER	16
I-3	Valeurs des $K_{ow}$ pour certains composés organiques	18
II-1	la base de données signatures atomiques uniques de niveau-1 pour l'Ethéphon avec les numéros d'occurrence	36
III-1	Base de données des signatures atomiques pour le Dimethoate et le Formothion élaborée au niveau-1 avec les nombres d'occurrences minimum et maximum dans une molécule	50
III-1	Base de données des signatures atomiques pour le Dimethoate et le Formothion élaborée au niveau-1 avec les nombres d'occurrences minimum et maximum dans une molécule	50
III -2	Matrice de descripteurs pour l'ensemble d'apprentissage des deux insecticides : Dimethoate et Formothion	52
III-3	ordre d'exécution et nombre d'itérations pour toutes les variables suivies par l'algorithme pour l'exemple d'ensemble de criblage composé des deux composés Dimethoate et Formothion	56
IV- 1	Signatures atomiques uniques de niveau -1 utilisées dans l'analyse QSPR	63
IV- 2	Equations de contrainte	67
IV- 2	Equations de contrainte	68
IV- 2	Equations de contrainte	69
IV- 3	Exemple de signature moléculaire résolue, notez que seules les signatures atomiques dont le nombre d'occurrences est supérieur à 0 sont illustrées	71
IV- 4	Solutions (nouvelles signatures moléculaires) de la structure ayant pour formule brute : $C_7H_{17}O_2PS_4$	71
IV- 5	Types, nombre d'occurrence et forme de signatures avec l'interconnectes des atomes	72



---

---

IV- 6	Structures 3D des 20 nouveaux pesticides et valeurs des <b>log<sub>k<sub>ow</sub></sub></b> calculées par la théorie I-QSPR et prédites par le logiciel Hyperchem, avec d'autres propriétés QSAR simulées Hyperchem	75
IV- 6	Structures 3D des 20 nouveaux pesticides et valeurs des <b>log<sub>k<sub>ow</sub></sub></b> calculées par la théorie I-QSPR et prédites par le logiciel Hyperchem, avec d'autres propriétés QSAR simulées Hyperchem	76
IV- 6	Structures 3D des 20 nouveaux pesticides et valeurs des <b>log<sub>k<sub>ow</sub></sub></b> calculées par la théorie I-QSPR et prédites par le logiciel Hyperchem, avec d'autres propriétés QSAR simulées Hyperchem	77

---

---

## Liste des équations

<b>Equation</b>		<b>Page</b>
I-1	Méthode de substitution de HUNSCH Formule pour déduire le logK de la molécule RX	13
I-2	Méthode fragmentale de REKKER Formule d'expression de logK	14
II-1	Le calcul de la signature moléculaire d'un composé	37
II-2	La représentation plus formelle des relations structure-propriété	38
II-3	l'équation de la régression linéaire multiple	39
II-4	Expression de l'équation de graphicalité	44
III-1	Expression de l'équation de graphicalité	54
III-2	Formule de calcul du coefficient de détermination (le carré de coefficient de corrélation) $R^2$	59
III -3	Formule de calcul du coefficient de la validation croisée $Q^2$	60
III -4	Formule de calcul de l'erreur quadratique moyenne (RMSE)	60
III -5	Formule de calcul de l'écart-type (la déviation standard)	60

## **Introduction générale**

Le développement scientifique observé dans divers domaines a beaucoup contribué dans l'évolution et la progression de la qualité et du niveau de vie des êtres vivants et les êtres humains. Les conditions de vie se sont considérablement améliorées, notamment en matière de santé, cela est dû au développement nutritionnel en termes de quantité, de qualité et de sources de nutriments.

Au cours des dernières décennies, le développement technologique et industriel a connu une croissance rapide, dans tous les domaines, visant comme but le confort et le bien-être de l'humanité. Ce développement a inclut le domaine agricole, dans le but de mettre dans la voie de consommation des denrées alimentaires et de produits, principalement fruits et légumes, exempts de matières polluantes ou toxiques ayant des effets néfastes sur la santé humaine.

Les pesticides, sources de ces polluants, étaient les plus visés, puisqu'ils ont un impact direct sur le rendement de production alimentaire, leur utilisation doit être rationnelle et bien contrôlée.

Les pesticides sont devenus des produits précieux dans l'agriculture moderne: les fongicides sont utilisés pour prévenir les maladies des cultures, les insecticides minimisent certains dommages causés par les insectes et les herbicides sont utilisés pour lutter contre les mauvaises herbes qui seraient autrement en concurrence avec les cultures pour obtenir des ressources précieuses.

Sur la base de tout ce qui précède, les pesticides ont un rôle crucial dans notre vie et ils doivent être prêtés d'une attention spéciale en ce qui concerne le suivi de leur développement, leur production et leur utilisation. Pour suivre ce développement tangible et rapide, différentes techniques ont été utilisées ayant pour but de contrôler, étudier, estimer et prédire les structures, les effets ainsi que les propriétés physico-chimiques de ces composés.

Le rôle de ces techniques ne se limite pas à ce qui est mentionné, mais s'étend à la création et la génération de nouveaux composés de la famille des pesticides.

Une des méthodes les plus importantes et les plus avancées dans ce domaine, est la conception moléculaire assistée par ordinateur (CAMD). Cette méthode engendre une multitude de théories qui ont été mises en évidence afin de faciliter le travail des chercheurs et leur permettre d'aboutir aux nouveaux composés chimiques et de les mettre en voie de synthèse et de commercialisation.

Parmi ces méthodes, la méthode de la relation inverse- structure propriétés I-QSAR/ QSPR qui consiste à résoudre un ensemble d'équations aboutissant à trouver des valeurs de variables indépendantes en fonction de l'activité/propriété souhaitée, qui aboutissent à la génération de nouveaux composés.

L'estimation des propriétés hydrophobes / hydrophiles de composés chimiques est pertinente pour de nombreux domaines, telle que, la médecine, la pharmacologie, les aliments, les parfums, l'industrie chimique et la protection de l'environnement. Ces propriétés sont mesurées par le coefficient de partage octanol / eau ( $k_{ow}$ ) appelée aussi lipophilie qui représente la distribution d'une substance entre deux phases [1-4]. Il s'agit d'une propriété physico-chimique importante dans l'environnement pharmaceutique, biochimiques, et sciences toxicologiques [5-9].  $K_{ow}$  est également utilisé pour fournir des informations précieuses pour la compréhension globale de l'absorption, la distribution, la biotransformation et l'élimination d'une grande variété de produits chimiques [10]. En raison de son utilisation croissante dans l'estimation de nombreuses autres propriétés,  $K_{ow}$  est considéré comme une propriété requise dans les études de produits chimiques nouveaux ou problématiques. Le coefficient de partage octanol / eau ( $K_{ow}$ ), généralement exprimée en valeurs de  $\log k_{ow}$  est un paramètre clé dans l'évaluation environnementale des produits chimiques, car il est lié directement au caractère lipophile / hydrophobe [11-12]. Il est utilisé comme prédicateur de base dans de nombreux modèles d'estimation pour la solubilité de l'eau, la biodisponibilité, la bioaccumulation, la toxicité / l'écotoxicité et l'évaluation / dépistage de PBT [13-14].  $\log k_{ow}$  est important à la fois pour la prédiction pharmacocinétique et la pharmacodynamique de médicaments et de substances toxiques [15]. Dans le règlement REACH, fournissant une valeur de  $\log k_{ow}$  ou encore ( $\log P$ ) est nécessaire pour toutes les bandes de tonnage des produits chimiques [16-17].

Compte tenu de l'extrême importance et du fait que la détermination expérimentale des coefficients de partage d'un grand ensemble de composés est fastidieuse et prend beaucoup de temps et exige des échantillons de haute pureté et des opérateurs qualifiés, le développement d'une méthode alternative comme la relation quantitative structure-propriété/ activité (QSPR, QSAR) serait utile et nécessaire pour le calcul théorique des valeurs de ces entités.

Dans cette étude, un effort a été fait pour concevoir de nouvelles structures chimiques des pesticides en utilisant un algorithme de conception moléculaire assisté par ordinateur (CAMD) faisant intervenir un puissant descripteur moléculaire appelé signature. Partant d'une base de données de plus de 190 pesticides, l'application de la théorie inverse-QSAR, nous a permis de générer une vingtaine de nouveaux composés, classifiés comme pesticides. Leurs

$\log k_{ow}$  et d'autres propriétés ont été encore prédites vérifiées et comparées à celles calculées par d'autres logiciels de modélisation, fiables, tel que Hyperchem.

La thèse est divisée en quatre grands chapitres:

Le premier chapitre, a été consacré à une synthèse bibliographique qui comprend des généralités sur le phénomène de partage, coefficient partage n-octanol/eau ( $\log k_{ow}$ ), détermination expérimentale, prédiction et domaines d'application de  $\log k_{ow}$ , ainsi qu'un aperçu sur les pesticides, leurs propriétés et leur classification. Le chapitre comprend aussi un aperçu historique, bien détaillée, sur les travaux antérieurs ayant utilisé les méthodes I-QSAR pour la prédiction et la génération de nouveaux composés chimiques pour des applications environnementales.

Dans le deuxième chapitre, nous avons abordé le concept des étapes de l'algorithme de conception moléculaire assisté par ordinateur (CAMD) utilisant le descripteur signature moléculaire et la méthodologie de développement de modèles QSPR, et enfin une description des étapes de l'algorithme de conception QSPR-inverse (I-QSPR) et leurs applications.

Le troisième chapitre comprend un détail des étapes de l'algorithme de conception de la QSPR-inverse avec un exemple d'application et la validation du modèle QSPR ainsi que les équations mathématiques qui permettent de déterminer les paramètres statistiques de nos modèles.

Le quatrième chapitre est consacré à l'interprétation des résultats de nos calculs sur les signatures atomiques de niveau-1, il contient les nouvelles structures de pesticides générés, leurs coefficients de partage octanol/eau ( $\log k_{ow}$ ) prédits

Le manuscrit est terminé par une conclusion générale avec quelques recommandations.

---

---

**CAPITRE I : SYNTHÈSE BIBLIOGRAPHIQUE****I-1 Introduction**

Une relation quantitative structure - activité (QSAR), parfois désignée sous le nom de relation quantitative structure - propriété (QSPR) est le procédé par lequel une structure chimique est corrélée avec un effet bien déterminé comme l'activité biologique ou la réactivité chimique. L'activité biologique peut être exprimée de manière quantitative, tel est le cas du coefficient de partage  $\log K_{ow}$ , qui représente le partage d'une molécule entre une phase hydrophobe (octanol) et une phase hydrophile (l'eau). Cette propriété physicochimique est devenue très utile et une caractéristique très importante d'une molécule du point de vue chimique et biologique. Lorsque les propriétés physicochimiques des molécules sont exprimées par des chiffres, on peut proposer une relation mathématique, ou tout simplement une relation quantitative structure à activité (QSAR), entre les deux. L'expression mathématique obtenue peut alors être utilisée comme moyen prédictif de la réponse biologique pour des structures similaires.

Ce chapitre comprend une analyse bibliographique succincte sur l'importance du coefficient de partage, son application dans le domaine environnemental tel que les pesticides et enfin un historique sur l'utilisation des relations I-QSAR pour la génération de nouvelles structures moléculaires.

**I-2 Définition du phénomène de partage**

De nombreux contaminants organiques couramment présents dans les eaux de surface et les eaux souterraines sont des substances hydrophobes stables qui se répartissent et se concentrent facilement dans divers compartiments environnementaux (biote, sédiments, sols, etc) [18]. En ce sens, on peut dire que le partage d'une substance entre une phase aqueuse et une phase organique constitue une propriété physico-chimique très importante de cette substance, notamment en ce qui concerne sa nature lipophile ou hydrophile, la propriété de partage conditionne en partie les propriétés biologiques de la molécule considérée, telles que le transport, la diffusion à travers les membranes (une molécule qui présente un partage élevée sera très lipophile et diffusera donc beaucoup à travers la membrane et inversement), la biodisponibilité (distribution et accumulation), l'affinité pour un récepteur et la fixation par une protéine, l'activité pharmacologique ou encore la

toxicité, s'agissant de contaminants, ce même partage conditionne leur devenir dans notre environnement en particulier leur accumulation dans les organismes aquatiques [17].

### I-3 Coefficient de partage n-octanol/eau

Le coefficient de partage n-octanol/eau est le rapport de la concentration d'un produit chimique dans le n-octanol à celle dans l'eau, dans un système à deux phases à l'équilibre [19-20]. Le logarithme de ce coefficient  $\log P$ , désigné encore par  $\log k_{ow}$ , a été montré pour être un des paramètres clés [21] dans l'étude quantitative des relations structure-activité (QSAR) et structure-propriété (QSPR) et est utilisé pour fournir des informations précieuses pour la compréhension globale de l'absorption, la distribution, la biotransformation et l'élimination pour une grande variété de produits chimiques.

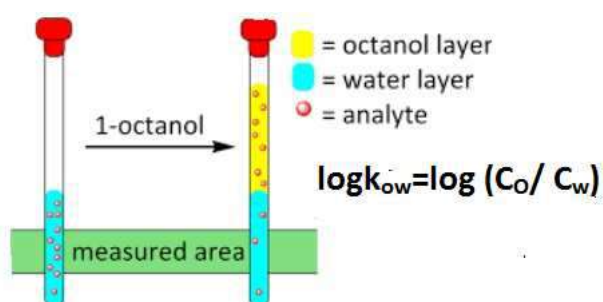
Ainsi, il existe un besoin d'avoir des programmes fiables qui peuvent être utilisés pour prédire le caractère lipophile des composés chimiques à partir de leur structure.

De nombreuses approches ont été développées pour la prédiction de  $\log P$  sur la base de paramètres structuraux non expérimentaux [22-23]. La plupart de ces méthodes utilisent des sous-structures (fragments / fragments d'atomes) [1,24] ou paramètres de chimie quantique (charges, potentiels électroniques, volumes moléculaires, forme ... etc.) [25-26] et l'analyse de régression linéaire multiple pour adapter les modèles aux données expérimentales ou encore les méthodes dites méthodes d'atomes-additifs [27-28]. Une vue d'ensemble exhaustive des différentes méthodes d'estimation du coefficient de partage octanol/eau ainsi que d'autres propriétés physiques ont été récemment publiées par Katritzky et al [29]. La principale caractéristique des méthodes par fragments est que la découverte commence par l'identification de composés ou de fragments de faible poids moléculaire (en général environ 200 Da) qui se lient à la molécule biologique d'intérêt [30]. Les méthodes à base de fragments fournissent de bons résultats pour un grand nombre de composés. Toutefois, des difficultés peuvent surgir dans la décomposition des structures en fragments appropriés dont les constantes sont disponibles. Le nombre de fragments dans de tels procédés peut varier de quelques centaines, dans les méthodes CLOGP et KOWWIN, jusqu'à plusieurs milliers, dans les méthodes ACD / LogP. Ces procédés sont considérés comme étant les plus précis sur le terrain en fonction de la comparaison des 14 méthodes différentes et réalisées par Mannhold et Dross [31].

Les paramètres de chimie quantique ont été utilisés pour prédire plusieurs centaines de composés [32], et on ne sait pas si ces approches peuvent être utilisées comme méthodes

d'estimation générales. En outre, les calculs chimiques quantiques prennent beaucoup de temps, ce qui est en fait un facteur limitant pour l'estimation d'un grand nombre de composés.

Les indices de l'état électrotopologiques (E -state) ont été récemment introduits par Hall et Kier [33-34] pour la description de molécules. Ces indices combinent les caractéristiques électroniques et topologiques des molécules analysées. Pour chaque type d'atome dans une molécule les indices E-état sont additionnés et peuvent être utilisés d'une manière de contribution de groupe. Ces indices sont connus comme indices d'atome- type état-E. De la même manière, les indices E-état peuvent également être utilisés pour décrire les liaisons entre les atomes spécifiques, à savoir, pour décrire les propriétés de deux atomes. Ces indices sont connus comme indices de liaison type E-état. Le calcul des types de deux indices E-état est très simple et directe. Le nombre d'application réussie de ces indices est de plus en plus rapidement, et plus de 100 articles avec application de ces indices pour prédire les propriétés physiques et de corrélérer différents types d'activités biologiques ont été publiés Kier et Hall [35]. La documentation systématique a été faite aux indices E-étatiques dans un nouveau book [36]. La figure I-1 explique le concept et comment calculer le coefficient de partage n-octanol /eau  $\log K_{ow}$



**Figure I-1:** Mesure du coefficient de partage n-octanol /eau  $\log K_{ow}$  [37]

#### I-4 Relation de $\log K_{ow}$ avec d'autres propriétés physico-chimiques:

Au cours des dernières années,  $K_{ow}$  est devenu un paramètre clé dans les études du devenir de produits chimiques organiques dans l'environnement. Il a été trouvé être liée à la solubilité dans l'eau coefficients d'adsorption huile / sédiments, et les facteurs de bioconcentration pour la vie aquatique. En raison de son utilisation croissante dans



l'estimation de ces autres propriétés,  $K_{ow}$  est considéré comme un bien nécessaire dans les études de problématiques des nouveaux produits chimiques [38].

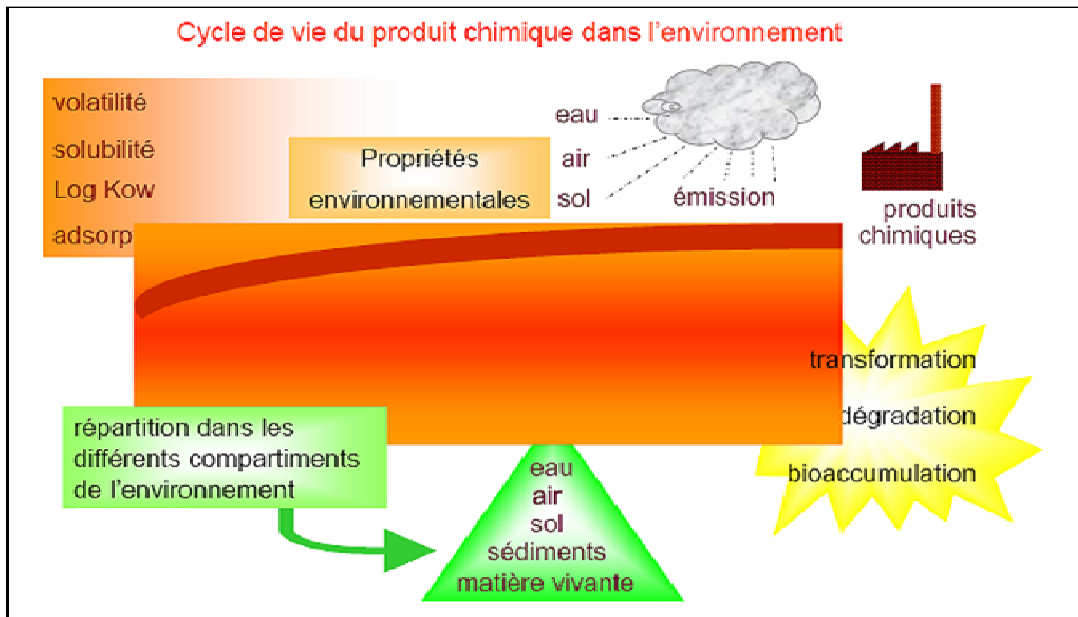


Figure I-2: Relation de  $\log K_{ow}$  avec d'autres propriétés physico-chimiques:

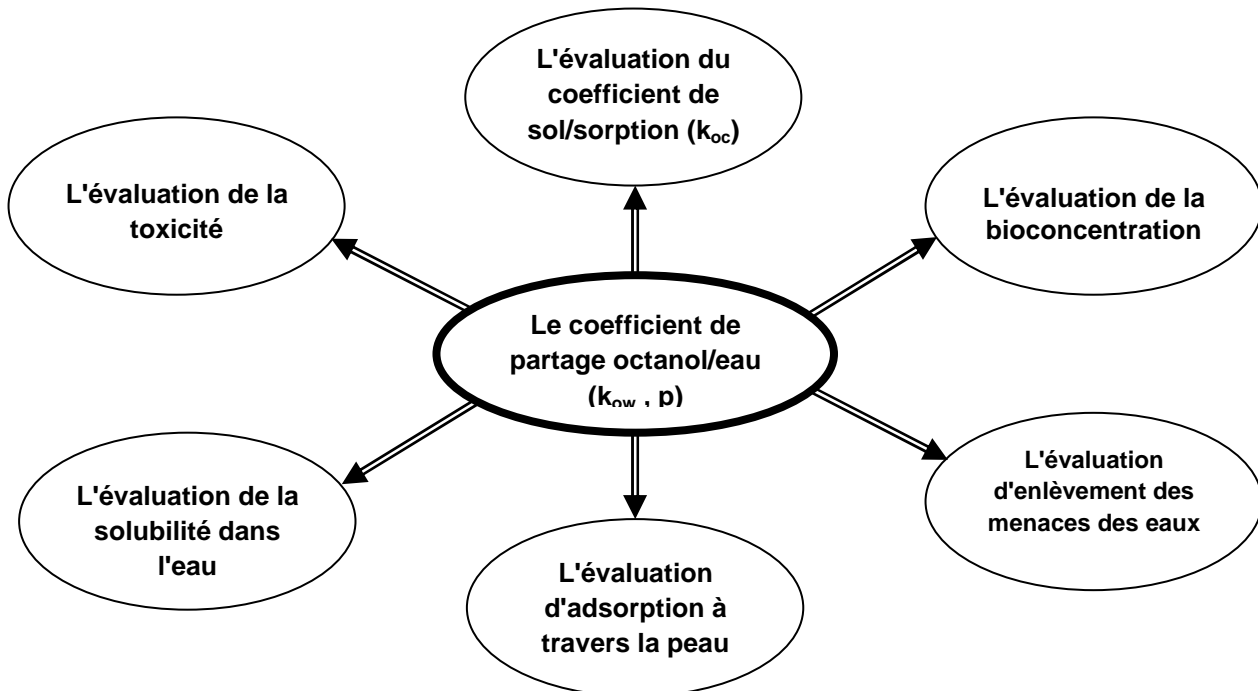


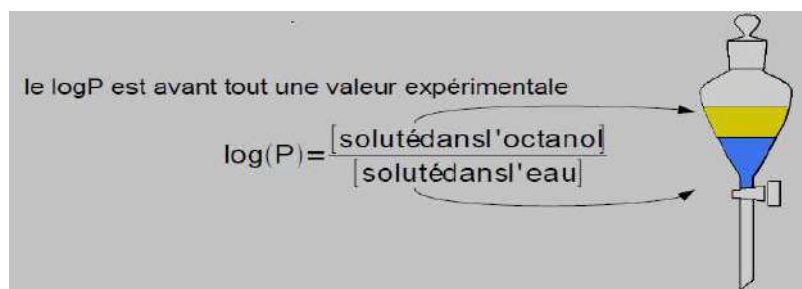
Figure I- 3: Diagramme de la relation de  $\log K_{ow}$  avec d'autres propriétés physico-chimiques:

## I-5 Détermination de $K_{ow}$

La détermination du coefficient de partage octanol-eau ( $K_{ow}$ ) est indispensable pour constituer les données de base qui doivent être soumises lors de la notification de nouvelles substances et de substances existantes prioritaires au sein de l'UE. Le coefficient de partage peut être déterminé au moyen des méthodes expérimentales, ou par prédiction à l'aide des relations empiriques, on calcule généralement le logarithme décimal du coefficient de partage  $\log K_{ow}$  ( $\log P$ ). Les deux méthodes peuvent être aussi combinées afin de prédire la valeur de cette entité. La détermination expérimentale de  $K_{ow}$  n'étant pas toujours indiquée, par exemple, pour les substances très hydrosolubles et pour les substances très lipophiles, il est possible d'utiliser une valeur de  $K_{ow}$  obtenue à partir de QSAR. Il convient cependant d'être extrêmement prudent dans l'utilisation des QSAR dans le cas de substances pour lesquelles la détermination expérimentale du coefficient de partage est impossible (tensioactifs, par exemple).

### I -5-1 Méthodes expérimentales

Pour déterminer expérimentalement des valeurs de  $K_{ow}$ , en général deux méthodes sont utilisées, la première c'est par agitation en flacon, la seconde par chromatographie en phase liquide à haute performance HPLC, ces deux méthodes sont décrites dans les lignes directrices normalisées, notamment la Ligne directrice 107 de l'OCDE (1995), la Ligne directrice 117 de l'OCDE (1983), et les documents CEE A.8. (1992), EPA-OTS (1982), EPA-FIFRA (1982) et ASTM (1993).



**Figure I- 4:** Méthode expérimentale de détermination de  $\log K_{ow}$

Les données obtenues selon la méthode par agitation du flacon ou de la méthode par HPLC, conformément aux lignes directrices normalisées, ne sont pas les seules

recommandées. Pour les substances hautement lipophiles, qui se dissolvent lentement dans l'eau, les données obtenues selon une méthode par agitation douce sont généralement plus fiables [39]. La méthode par agitation douce fait actuellement l'objet d'un essai circulaire, en vue de la mise au point définitive d'une ligne directrice de l'OCDE.

#### I-5-1-1 Méthode par agitation du flacon

Le principe de base de cette méthode consiste à mesurer la dissolution de la substance dans deux phases différentes, l'eau et l'octanol. Pour déterminer le coefficient de partage, l'équilibre entre tous les composants qui inter-réagissent dans le système doit être atteint; après quoi on détermine la concentration des substances dissoutes dans les deux phases. La méthode par agitation du flacon est applicable lorsque  $\log K_{ow}$  est compris entre -2 et 4 [40]. Elle ne s'applique qu'aux substances pratiquement pures, solubles dans l'eau et l'octanol, et doit être mise en œuvre à une température constante ( $\pm 1^\circ\text{C}$ ), dans l'intervalle 20-25°. Les avantages de cette méthode sont: sa facilité à l'utilisation, fiable pour les substances qui ont la valeur de  $K_{ow} < 4$ , et n'a pas besoin d'équipement coûteux.

Mais, ses désavantages, c'est qu'elle n'est pas utilisable pour les substances qui ont la valeur de  $K_{ow} > 4$ , l'agitation peut former des micro-émulsions qui conduit à des résultats inexacts.



Figure I -5: montre un système de deux phases [41]

### I-5-1-2 Méthode par HPLC

La plupart des méthodes de HPLC appliquées à la quantification de l'OMZ dans des échantillons pharmaceutiques et biologiques utilisent des colonnes conventionnelles C8 ou C18, un tampon phosphate (pH 7 à 7,6), l'acétonitrile en phase mobile et une détection UV [42]. La chromatographie en phase liquide à haute performance (HPLC) est devenue une approche couramment utilisée pour atteindre ces objectifs [43]. Elle se pratique sur des colonnes analytiques garnies d'une phase solide disponible dans le commerce, présentant de longues chaînes d'hydrocarbures (par exemple, C<sub>8</sub>, C<sub>18</sub>) chimiquement liées au gel de silice. Les substances injectées sur une telle colonne migrent à des vitesses différentes en raison de leur degré de partage différent entre la phase aqueuse mobile et la phase hydrocarbonée stationnaire. La méthode par HPLC n'est pas applicable aux bases et aux acides forts, aux complexes métalliques, aux matériaux tensioactifs ou aux substances qui réagissent avec l'éluant. Elle est applicable lorsque la valeur de  $\log K_{ow}$  est comprise entre 0 et 6 [44]. Cette méthode est moins sensible à la présence d'impuretés dans le composé d'essai que la méthode par agitation du flacon.

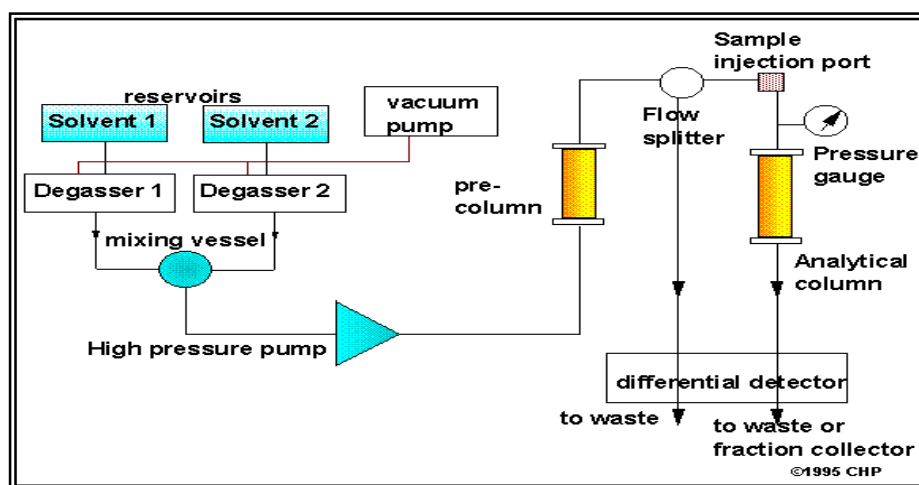


Figure I -6:Montre le principe de l' HPLC. [45].

### I-5-1-3 Méthode par agitation douce

La méthode par agitation douce permet une détermination exacte et précise de  $K_{ow}$  pour des composés dont le  $\log K_{ow}$  va jusqu'à 8.2 [39]. Dans le cas de composés fortement

lipophiles, la méthode par agitation du flacon a tendance à produire des artefacts (formation des microgouttelettes) et la méthode par HPLC nécessite une extrapolation au-delà de la gamme d'étalonnage pour obtenir une estimation de  $K_{ow}$ . Pour déterminer un coefficient de partage, on met en contact l'eau, l'octanol et le composé à tester jusqu'à ce que l'équilibre entre eux soit atteint, après quoi on détermine la concentration du composé testé dans les deux phases. Dans la méthode par agitation douce, on peut surmonter jusqu'à un certain point les difficultés expérimentales liées à la formation des microgouttelettes caractéristiques de la méthode en amenant à l'équilibre l'eau, l'octanol et le composé testé dans un réacteur soumis à une agitation douce. Cette agitation crée un écoulement plus ou moins laminaire entre l'octanol et l'eau, et améliore l'échange entre les phases, sans qu'il y ait formation de microgouttelettes.

Les avantages de cette méthode, facile à utiliser, relativement rapide, n'a pas besoin d'équipement coûteux, fiable essentiellement pour toutes les substances. Mais les désavantages sont, cette méthode a besoin d'excitation attentive et contrôle fermé de température pour éviter la formation des micro-émulsions [46].

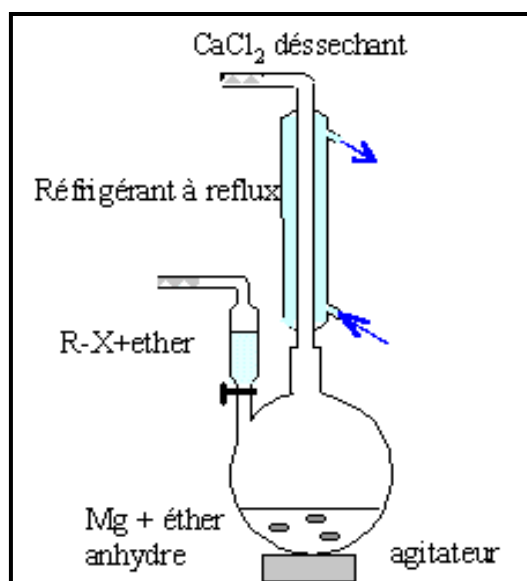


Figure I-7: Principe de la méthode par agitation douce. [46]

#### I-5-1-4 Méthode sur colonne

La méthode sur colonne constitue une autre méthode très polyvalente pour déterminer  $\log K_{ow}$ . Dans cette méthode, on utilise une colonne pour obtenir le partage de la substance d'essai entre les phases octanol et eau. La colonne est garnie avec un support solide et saturée avec une concentration donnée de la substance à tester dans de l'octanol. La substance à tester est diluée à partir de la colonne saturée en octanol par de l'eau.

La solution aqueuse sortant de la colonne représente la concentration d'équilibre de la substance d'essai partagée entre la phase octanol et la phase aqueuse. Le principal avantage de la méthode sur colonne par rapport à la méthode par agitation en flacon est qu'elle évite totalement la formation de microémulsions.

Elle est donc particulièrement utile et fiable pour mesurer le  $K_{ow}$  de substances pour lesquelles  $\log K_{ow}$  est inférieur à 4.5. L'un de ses inconvénients tient à ce qu'elle nécessite un équipement complexe et coûteux.



Figure I-8: Principe de la méthode sur colonne [45]

### I-5-2 Prédiction de $\log k_{ow}$

La plupart des méthodes expérimentales de détermination de  $\log K_{ow}$  souffrent du même inconvénient, à savoir que leur domaine d'application est relativement étroit. D'autre part, du fait de la nature intrinsèque de certaines molécules, leurs  $\log K_{ow}$  sont inaccessibles à l'expérience.

Avec le développement des moyens de calcul notamment l'informatique, la détermination de  $\log K_{ow}$  dans le domaine de la conception assistée par ordinateur ou dans le domaine de la chimie combinatoire est devenue possible [47], les chercheurs travaillent sur des modèles moléculaires avant même que les molécules aient été synthétisées. Les plus anciennes des méthodes d'estimation des  $\log K_{ow}$  sont les méthodes de substitution et les méthodes fragmentables. La valeur de  $\log K_{ow}$  est obtenue par sommation des lipophiles des substituant ou des fragments [48].

#### I-5-2-1 Méthode de substitution de HUNSCH

Hansch considère que la substitution d'un hydrogène d'un radical R pouvant être le cycle benzénique, par un substituant X ( X= O, NH<sub>2</sub>, CH<sub>2</sub>...) est équivalente à l'insertion d'un constituant de type X-H à ce radical, c'est-à-dire d'un OH pour X=O, d'un NH pour X=NH<sub>2</sub>, d'un CH<sub>3</sub> pour X=CH<sub>2</sub> ...etc. Il a attribué à chaque substituant sa propre lipophilie appelé paramètre de Hansch :  $\Pi_X$ , connaissant donc le coefficient de partage de la molécule RH, on peut facilement en déduire le logK de la molécule RX en utilisant la formule suivante :

$$\log K_{RX} = \log K_{RH} + \Pi_X + \Pi_{Corr} \quad \text{I-1}$$

$K_{RX}$  et  $K_{RH}$  sont les coefficient de partage des molécules RH et RX .

$\Pi_X$  est le paramètre de lipophilie du substituant X.

$\Pi_{Corr}$  est un terme correctif tenant compte de l'effet entraîné par les ramifications, les doubles liaisons, les fermetures de cycle, les liaisons intramoléculaires ... etc.

**I-5-2-2 Méthode fragmentale de REKKER**

La méthode consiste à utiliser la contribution atomes/fragments dans la lipophilie d'une substance, cette dernière est donnée par sommation des valeurs des lipophilies de ces atomes ou fragments dont les valeurs sont connues et déterminées expérimentalement, appelées constantes fragmentales hydrophobes tableau **I-1**, et d'introduire les facteurs de corrections nécessaires qui sont fonction de la structure chimique du composant considéré.

La valeur de logK peut être exprimée de la façon suivante [17-49].

$$\log K = \sum_n a_n f_n + k.C_M$$

**I-2**

- $a_n$  est le nombre de fragments identiques dans la molécule.
- $f_n$  est la constante fragmentale hydrophobe.
- $k$  est un entier qui dépend de la structure de la molécule.
- $C_M$  est un terme correctif appelé constante magique ( $C_M=0.219$ ).



Tableau I- 1: valeurs des lipophiles de différents fragments [49]

Fragment	f <sub>aliphatique</sub>	f <sub>aromatique</sub>
H		0,204
C		0,110
CH		0,315
CH <sub>2</sub>		0,519
CH <sub>3</sub>		0,724
C <sub>6</sub> H <sub>5</sub>		1,902
N	-2,074	-0,979
O	-1,545	-0,450
S	-0,558	0,099
F	-0,213	0,444
Cl	0,057	0,933
Br	0,258	1,134
I	0,570	1,446
NO <sub>2</sub>	-0,915	-0,039
OH	-1,448	-0,353
SH	-0,046	0,611
NH <sub>2</sub>	-1,340	-0,902
CO	-1,633	-0,976
CH=O	-0,990	-0,33
CO-O	-1,200	-0,543
O-CO		-0,981
COOH	-0,942	-0,066
CONH <sub>2</sub>	-2,011	-1,135
OCONH <sub>2</sub>	-1,405	-0,967
OCH <sub>3</sub>	-0,821	0,274

Tableau I- 2: Termes correctifs de la formule de REKKER [49]

Structure chimique	Correction : $k \cdot C_M$	Remarques
<b>Hydrocarbure aliphatiques non conjugués</b>		
Alcanes	$+2C_M$	sauf méthane et cyclopropane $:-1C_M$
Alcènes	0	pas de correction
Alcynes	$-1C_M$	par triple liaison
<b>Cycles aromatiques</b>		
Cycles aromatiques	$+1C_M$	Si l'on calcule à partir de $f(C)$ et $f(H)$
Cycles aromatiques	0	Si l'on utilise les constantes $f_{\text{aromatiques}}$
Cycles aromatiques accolés	$+1C_M$	Par paire de sommets communs, ainsi que pour chaque liaison unissant deux cycles.
Fragment basique lié à deux cycles	$+1C_M$	
<b>Hydrocarbures conjugués</b>		
Liaisons conjuguées	$+1C_M$	Par paire de liaison double ou triple
<b>Composés de type <math>X-(CH_2)_n-Y</math></b>		
Composés de type $X-(CH_2)_n-Y$	$+4C_M$ $+2C_M$ 0	Pour $n=1$ Pour $n=2$ Pour $n>2$

### I-6 Paramètres influençant sur la constante de partage

Le coefficient de partage  $K$  est une constante thermodynamique qui dépend :

- De la nature de la substance qui se partage entre les deux phases
- Du système de solvants utilisés

Partant du principe « qui se ressemble s'assemble », un solvant dissout bien un composé qui lui ressemble, les solvants polaires et dissociant tels que l'eau dissolvent les composés ioniques et / ou polaires hydrophiles. Les molécules apolaires présentent souvent un  $K > 10$  entre un solvant hydrophobe et l'eau, les solvants apolaires et peu dissociant dissolvent les molécules et les composés hydrophobes (solvants chlorés ou hydrocarbures).

Le solvant intervient aussi par son caractère protique ou aprotique, les solvants protiques tel que l'eau, ammoniac, alcools, phénols, acides et amides non substituées influent sur l'extraction à cause de l'existence d'un hydrogène labile.

D'autre part, la nature du soluté intervient aussi dans le phénomène de partage, en effet, la structure d'une molécule organique joue un rôle important sur son coefficient de partage, l'accroissement de la chaîne augmente la valeur de  $K$  d'environ 4 unités pour chaque groupement méthylénique ( $\text{CH}_2$ ) incorporé dans la molécule .

Un composé ramifié présente un  $K$  inférieur à celui de son isomère linéaire, il en va de même pour un composé non saturé par rapport au composé saturé correspondant.

La présence d'hétéro atome ( $\text{O}, \text{N}$ ) diminue parfois considérablement la valeur de  $K$  à cause des liaisons hydrogène avec l'eau [50] .

L'introduction d'un halogène au contraire favorise le passage en phase organique donc l'augmentation de  $K$  [50].

L'inclusion d'une molécule simple dans un complexe chargé diminuera fortement le coefficient de partage, en effet la forme ionique ainsi obtenue est du fait de sa charge beaucoup plus fortement retenue par l'eau, solvant polaire.

Une molécule possédant un moment dipolaire ( $\mu$ ) non nul, définissant en quelque sorte l'énergie de liaisons Van Der Waals , la constante diélectrique ( $\epsilon$ ) elle, définit la

polarisabilité moléculaire ( ou polarisation moléculaire ), elle permet de se rendre compte de l'affinité des solvants entre eux même [50] .

En résumé, pour une substance donnée, K augmente avec:

- la longueur de chaîne.
- la linéarité de la chaîne.
- la saturation de la chaîne.
- le contenu en chlorure.
- une diminution du contenu en oxygène ou azote.

Le tableau suivant illustre l'ordre de grandeur de K de différents composés organiques dans le système: octanol/eau

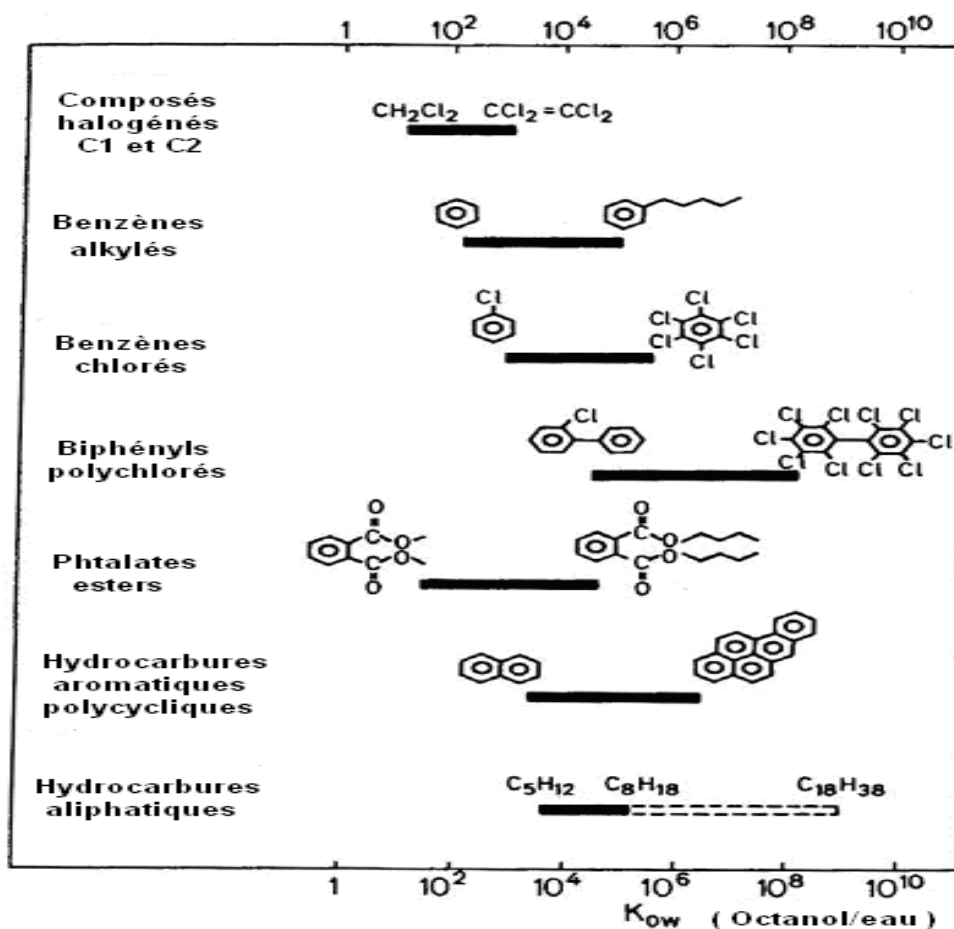
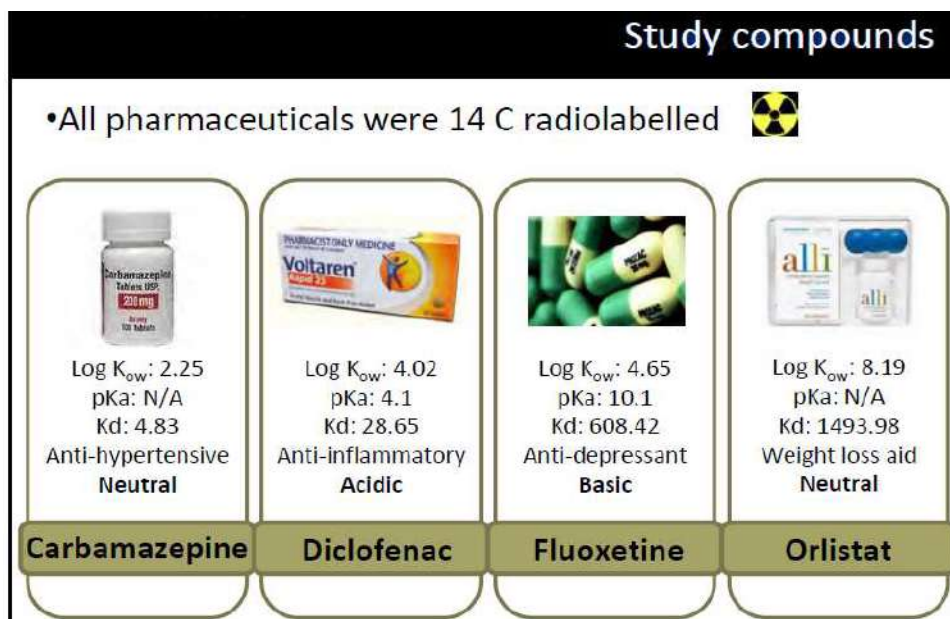


Tableau I- 3 : Valeurs des  $K_{ow}$  pour certains composés organiques [51]

### I-7 Application de $\log K_{ow}$ dans les domaines de médecine, pharmacie, environnement et autres

Le coefficient de partage octanol-eau ( $K_{ow}$ ) est une mesure de la concentration d'équilibre d'un composé entre l'octanol et l'eau qui indique le potentiel pour le partitionnement en matière organique du sol (soit un  $K_{ow}$  élevé indique un composé qui de préférence se répartir dans le sol organique question plutôt que de l'eau). Il est connu pour être l'une des propriétés physiques quantitatives les plus corrélées avec l'activité biologique, car le système octanol saturé d'eau est considéré comme un modèle raisonnable de l'environnement physico-chimique dans les organismes vivants [52-53]. Ce paramètre nous montre que les produits chimiques dont  $\log K_{ow} < 2,5$ , comme le sulfaméthoxazole et la sulfadimidine, sont considérés comme ayant un faible potentiel de sorption hydrophobe [54]. Ce coefficient est inversement proportionnel à la solubilité du composé dans l'eau. Le  $\log K_{ow}$  est modélisé pour estimer les facteurs de bioaccumulation des plantes et d'invertébrés du sol. Le  $\log K_{ow}$  est couramment utilisé dans les études QSAR / QSPR et la conception de médicaments, car cette propriété est liée à l'absorption du médicament, la biodisponibilité, le métabolisme et la toxicité. Ce paramètre est également utilisé dans de nombreuses études environnementales pour aider à déterminer le devenir dans l'environnement de produits chimiques [55-56].  $\log K_{ow}$  a fait un grand usage dans la médecine et la chimie médicinale. Biodégradation (TBD en mol / h) est un autre des facteurs utiles et importants dans les études chimiques et biochimiques [56].  $\log P$  ou  $\log K_{ow}$  est aussi un indice très important dans le domaine médical, qui indique la distribution d'un médicament dans les tissus et les lipides dans l'être humain et la capacité d'un médicament à atteindre rapidement un site actif. Ajouter à cela les modèles pharmacocinétiques à base physiologique (PBPK) utilisent des coefficients de partage et des volumes de tissu pour décrire la distribution du médicament [57].



**Figure I- 9:** exemples d'application de logKow dans le domaine pharmaceutique.

### I-8 Importance d'application du log $k_{ow}$ dans le domaine environnemental

Le coefficient de partage est le paramètre le plus important pour une grande variété de produits chimiques organiques à caractère non ionique telle que les pesticides [58]. Dans la chimie environnementale des pesticides,  $K_{ow}$  a prouvé son utilité pour prédire l'adsorption dans le sol [59], la mobilité et la persistance dans le sol [60-61], la bioconcentration par des organismes aquatiques [62, 63-64], et accumulation par le biote terrestre. L'importance de  $K_{ow}$  est mise en évidence par sa corrélation avec un certain nombre de processus impliquant le partitionnement lipide / eau, tels que le comportement du médicament, la biosynthèse des caroténoïdes [65], l'absorption des racines et la translocation de produits chimiques non ionisés et de phosphate dans des plantes [66-67].

En plus de ce qui précède le taux de pesticides transférés de la solution du sol dans la sève de xylème de la plante [68-69]. Briggs et al [67], ont mis en évidence une forte corrélation entre les facteurs de FCR ou de TSCF et les coefficients de partage octanol-eau ( $K_{ow}$ ) des pesticides. Ces corrélations ont principalement été prises en compte dans les études prédisant l'absorption dynamique par les plantes de non électrolytes et d'électrolytes faibles, d'acides et de bases [70-67].

## I-9 Les pesticides

Les pesticides sont des produits agrochimiques [71] utilisés pour tuer les insectes nuisibles, notamment les insectes, les rongeurs, les champignons et les plantes indésirables (mauvaises herbes), et pour lutter contre les organismes considérés nocifs. Ce terme englobe les substances utilisées comme régulateurs de la croissance végétale, défoliants, exciccateurs, agents d'ébourgeonnement ou inhibiteurs de germination, ainsi que les substances appliquées aux cultures avant ou après la récolte pour protéger la production contre toute détérioration pendant le stockage et le transport recommandés par la FAO et l'OMS. Ils sont devenus une composante importante des systèmes agricoles mondiaux et du secteur de la santé pour les individus et les communautés au cours du siècle dernier, permettant une augmentation notable du rendement des cultures et de la production alimentaire [72]. Ils sont considérés comme un élément essentiel de l'agriculture moderne jouant un rôle majeur dans la sécurité alimentaire et la santé publique [73-74]. En outre, ces produits chimiques améliorent directement la santé publique en atténuant certains agents pathogènes tels que le paludisme [74-75], les poux [76], les puces et les tiques [77] et le virus Zika [73]. Leurs principales propriétés incluent une faible polarité, une faible solubilité dans l'eau et une lipophilie élevée. En conséquence, ils se bioaccumulent et se bioamplifient tout au long de la chaîne alimentaire et constituent donc une menace pour l'environnement et la santé humaine.

Les pesticides ont de nombreux effets bénéfiques. Celles-ci incluent la protection des cultures, la conservation des aliments et des matériaux et la prévention des maladies à transmission vectorielle. Par exemple, les pesticides peuvent être utilisés dans la prévention du paludisme, qui tue jusqu'à 1 million d'enfants par an, et dans la prévention d'autres maladies à vecteur telles que la dengue, la leishmaniose et l'encéphalite japonaise.

Les pesticides sont souvent considérés comme une solution facile, rapide et peu coûteuse pour lutter contre les mauvaises herbes et les insectes nuisibles en agriculture et dans les paysages urbains. Les pesticides sont intrinsèquement toxiques- ce sont des BIOCIDES, conçus pour tuer, réduire ou repousser les insectes, les mauvaises herbes, les rongeurs, les champignons ou d'autres organismes qui peuvent menacer la santé publique et l'économie. La contamination par les pesticides pose des risques considérables pour l'environnement et les organismes non ciblés [78]. Leur mode d'action consiste à cibler chez

les organismes nuisibles des systèmes ou des enzymes qui peuvent être identiques ou très similaires aux systèmes ou aux enzymes chez l'homme et qui, par conséquent, présentent des risques pour la santé humaine et l'environnement [79].

L'exposition des enfants aux pesticides et leur sensibilité particulière suscitent de plus en plus d'inquiétudes. Les enfants ne sont pas de petits adultes et peuvent avoir des expositions plus élevées et une plus grande vulnérabilité à des niveaux d'exposition élevés et faibles [80].

Les pesticides sont classés sur la base de divers critères tels que toxicité (effets dangereux), organisme nuisible qu'ils tuent et la fonction de pesticide, composition chimique, mode d'entrée, mode d'action et comment ou quand ils fonctionnent [81].

### **I-9 -1 Classification des pesticides sur la base de la toxicité**

Cette Classification donne la réponse aux questions suivantes : quelle quantité de substance est impliquée (dose) et combien de fois l'exposition à la substance se produit (temps) donner lieu à deux types différents de toxicité - toxicité aiguë et toxicité chronique [82]. La toxicité aiguë fait référence à la toxicité d'un pesticide pour un être humain, un animal ou une plante après une exposition unique à court terme. Pendant que La toxicité chronique est un effet toxique retardé de l'exposition à un pesticide et concerne le grand public ainsi que ceux travaillant directement avec les pesticides. Dans sa classification des pesticides, l'organisation mondiale de la santé (OMS) a mis l'accent sur et se concentre uniquement sur la toxicité aiguë. Après révision en 2009, de ces classes ont été harmonisées avec les catégories de danger de toxicité aiguë du «Système général harmonisé (SGH) [83]».

### **I-9 -2 Classification des pesticides sur la base de leur utilisation [82-84].**

Cette classification se base sur les organismes nuisibles que les pesticides tuent ainsi que la fonction de ceux-ci. Elle consiste en 30 types qui sont : Acaricides (substances utilisées pour tuer les acariens et les tiques, ou pour perturber leur croissance ou leur développement), Algicide (substances utilisées pour tuer ou inhiber les algues), Antifeedants (Substances chimiques empêchant l'alimentation des insectes ou autres nuisibles), Avicides (nuisibles ciblés sont des produits chimiques utilisés pour tuer les oiseaux), Bactericides (tuent ou inhibent les bactéries dans les plantes ou le sol), Bird



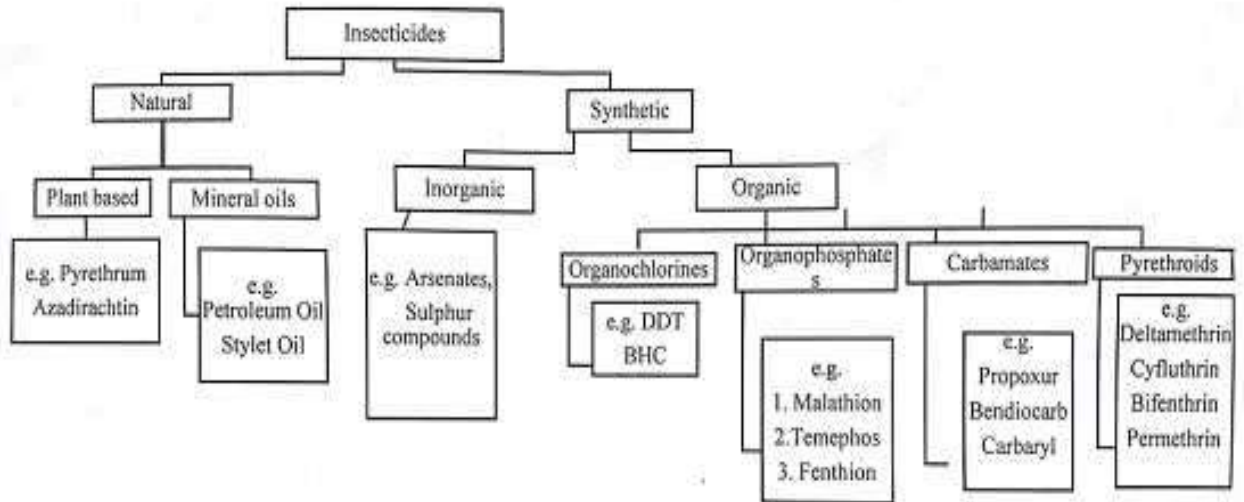
repellents (les insectes cibles sont des produits chimiques qui repoussent les oiseaux), Chemosterillant (les organismes nuisibles visés sont des produits chimiques rendant un insecte stérile et l'empêchant ainsi de se reproduire), Desiccants (agir sur les plantes en séchant leurs tissus), Fongicide (pour tuer les moisissures ou les champignons), Herbicide softener (Un produit chimique qui protège les cultures contre les dommages causés par les herbicides, mais n'empêche pas les herbicides de tuer les mauvaises herbes.), Herbicides (pour tuer les mauvaises herbes), Insect attractant (Un produit chimique qui attire les animaux nuisibles pour les piéger, les retirant ainsi des animaux cultivés et des produits stockés), ), Insect growth regulator (substance qui perturbe la croissance ou le développement d'un insecte), Insecticide (pesticide utilisé pour tuer les insectes ou perturber leur croissance ou leur développement), Larvicides (inhiber la croissance des larves), Lampricides (Larves ciblées des lamproies qui sont des poissons sans mâchoires comme les vertébrés), Mammal repellent (produit chimique qui dissuade les mammifères de s'approcher des cultures ou des produits stockés, de se nourrir d'eux); Mating disrupters (substances chimiques qui nuisent à la façon dont les insectes mâles et femelles se localisent à l'aide de produits en suspension dans l'air, les empêchant ainsi de se reproduire), Molluscicides (Substances utilisées pour tuer les limaces et les escargots.), Moth balls (Arrête tout endommagement des chiffons par les larves de papillons), Nematicides (Produits chimiques utilisés pour lutter contre les nématodes), Ovicides (Empêche la croissance des œufs d'insectes et d'acariens), Piscicides (Agit contre les poissons), Plant growth regulators (les substances modifient le taux de croissance, de floraison ou de reproduction prévu des plantes), Rodenticides (substances utilisées pour tuer les rats et les animaux apparentés), Silvicides (agit contre la végétation ligneuse), Synergists (un produit chimique augmente la toxicité d'un pesticide pour un organisme nuisible, mais n'est pas auto-toxique pour le parasite), Termiticides (élimine les termites), Virucide (agent capable de détruire des virus inactivés ).

### **I-9 -3 Classification des pesticides sur la base de la composition chimique [85]**

La méthode la plus courante et la plus utile de classification des pesticides est basée sur leur composition chimique et la nature de leurs ingrédients actifs. C'est ce type de classification qui donne des indices sur l'efficacité, les propriétés physiques et chimiques des pesticides respectifs. Cette méthode s'applique à la fois insecticides, fongicides, herbicides et rodenticides comme suit :

• **Insecticides** : sont classés sur la base de la composition chimique comme suit : Carbamates (carbaryl), Organochlorés (Endosulfan), Organophosphorés (Monocrotophos), Pyréthroïdes (perméthrine) Néonicotinoïdes (Imidaclopride), divers pesticides tels que Spinosyns (Spinosad), Benzolureas (diflubenzuron), Antibiotiques (abamectin) [81].

En raison de son extrême importance, Les insecticides qui peuvent être en outre classé en plusieurs sous-classes comme le montre la figure ci-dessous [84].



**Figure I-10:** Classification des pesticides

- **Fongicides:** les fongicides sont classés en tant que fongicides azotés aliphatiques (dodine), fongicides amides (carpropamide), fongicides aromatiques (chlorothalonil), fongicides dicarboximide (famoxadone), fongicides dinitrophénol (dinocap), etc.
- **Herbicides** : Les herbicides sont les herbicides anilides (flufenacet), phénoxyacétiques (2, 4-D), les herbicides à l'ammonium quaternaire (Paraquat), les herbicides à la chlorotriazine (atrazine), les herbicides à la sulfonilurée (chlorimuron), etc.
- **Rodenticides** : Ils sont classés en tant que rodenticides inorganiques (phosphure de zinc, phosphure d'aluminium), rodenticides coumariniques (organiques) (bromadiolone, coumatétralyle).

#### I-9 -4 Classification des pesticides sur la base du mode d'entrée [84].

Les façons dont les pesticides entrent en contact avec la cible ou s'y introduisent sont appelées modes d'entrée. L'illustration est : **type de pesticide** (description).

- **Pesticides Systémiques** (Ce sont des pesticides qui sont absorbés par les plantes ou les animaux et transférés vers des tissus non traités),
- **Pesticides de contact** (Ils agissent sur les parasites ciblés quand ils entrent en contact),
- **Poisons de l'estomac** (Ils pénètrent dans le corps du ravageur par la bouche et par le système digestif),
- **Fumigants** (Pesticides agissant ou susceptibles de tuer les organismes nuisibles visés en produisant de la vapeur et en pénétrant dans le corps de cet organisme par la trachée),
- **Répulsifs** (Les répulsifs ne tuent pas mais sont suffisamment désagréables pour éloigner les parasites de la zone traitée. Ils entravent également la capacité de l'organisme nuisible à localiser les cultures).

#### **I-9 -5 - Classification des pesticides sur la base du mode d'action [84].**

Les pesticides sont également classés en fonction de leur mode d'action :

- **Poison physique** ( Les pesticides provoquent la mort d'un insecte en exerçant un effet physique),
- **Poison protoplasmique** (Les pesticides sont responsables de la précipitation des protéines),
- **Poison respiratoire** (Produits chimiques qui inactivent les enzymes respiratoires),
- **Poison nerveux** (Les produits chimiques inhibent la conduction des impulsions),
- **Inhibition de la chitine** (Les produits chimiques inhibent la synthèse de chitine chez les animaux nuisibles).

#### **I-9-6 Les pesticides peuvent également être classés en fonction de comment ou quand ils fonctionnent**

- **Pesticides de contact** : Ces pesticides contrôlent un organisme nuisible à la suite d'un contact. Les insectes sont destructifs lorsqu'ils sont pulvérisés directement ou lorsqu'ils rampent sur des surfaces traitées avec un insecticide de contact à effet résiduel.

- **Pesticides systémiques** : Les pesticides qui sont absorbés par les plantes ou les animaux et déplacent les tissus non traités. Les insecticides ou les fongicides systémiques se déplacent dans les plantes traitées et détruisent certains insectes ou champignons.
- **Pesticides Foliaires** : Ceux-ci sont appliqués aux plantes, feuilles, tiges et branches
- **Pesticides appliqués au sol** : Ceux-ci sont appliqués au sol. Certaines sont absorbées par les racines et transplantées à l'intérieur de la plante.
- **Fumigants** : Les produits chimiques appliqués sous forme de gaz toxique ou sous forme de solide ou de liquide forment un gaz toxique et pénètrent dans les fissures et les crevasses des structures ou du sol.
- **Herbicides de pré-plantation** : Ceux-ci sont appliqués sur le sol avant l'ensemencement ou le repiquage.

**Herbicides en prélevée-** Celles-ci sont appliquées au sol après la plantation mais avant la levée de la culture ou de la mauvaise herbe.

- **Herbicides de post levée** : Ceux-ci sont appliqués après que la culture ou la mauvaise herbe est apparue.
- **Pesticides sélectifs** : Ils ne contrôleront que certains parasites.
- **Pesticides non sélectifs (ou à large spectre)** : Ils contrôleront un large éventail de parasites.
- **Pesticides suffocants** : ils obstruent le système respiratoire des insectes et peuvent affecter les œufs.
- **Pesticides résiduels** : Ils ne se décomposent pas rapidement et peuvent contrôler les ravageurs pendant une longue période.
- **Pesticides non résiduels** : Ils sont rapidement rendus inactifs après l'application et n'affectent pas les cultures ou les parasites à venir.

**I-10 Aperçu historique sur l'utilisation des méthodes I- QSAR/QSPR**

**En 1990**, V. Ekaterina et al [86], ont tenté pour la première fois de s'attaquer à un problème inverse intrigant dans QSAR / QSPR. Ils considèrent la solution de cette tâche pour les indices de connectivité moléculaire de premier ordre.

S'agissant de la méthodologie générale suivie, ils se sont appuyés sur les notions de base de l'approche: **i**) restauration des structures chimiques à partir des valeurs des indices topologiques ( ${}^1\chi$  and  ${}^1\chi'$ ), **ii**) distribution des types de bord (DET), **iii**) distribution des types de sommet (DVT). Cette méthodologie a été résumée dans un algorithme en plus des étapes de base de la reconstruction. Ces étapes comprennent: génération exhaustive de DET pour une valeur d'indice donnée, génération de VDTs pour un DET donné puis génération et sélection de structures chimiques.

**En 2002**, P. Donald et al [98], ont publié un article dans lequel ils présentent au début un nouveau travail concernant l'introduction de la signature en tant que nouveau descripteur moléculaire et divers indices topologiques utilisés dans les relations quantitatives structure-activité (QSARs) sont exprimées en tant que fonctions du nouveau descripteur. À cet égard, ils fournissent un exemple simple montrant qu'il suffit de compter l'apparition d'une signature particulière dans une molécule en tant que descripteur par rapport au calcul de certains TIs (indices topologiques) et démontrer l'équivalence des QSPRs résultants. Ces QSPRs ont été établis en corrélant l'activité de 121 inhibiteurs de la protéase du VIH-1 à partir de signatures de différentes hauteurs. Ils comparent cela à un QSAR déterminé à partir d'un emballage disponible dans le commerce par exemple, des descripteurs (Molconn-Z). Comme résultats préliminaires, tous les TIs choisis peuvent être calculés à partir de la signature-**h** du graphe moléculaire inconnu pour **h** (la hauteur des signatures atomiques) non supérieur à **n**, nombre d'atomes calculé à partir de la signature-**0**.

En ce qui concerne l'application du problème inverse (I-QSAR) pour trouver la conception de la bibliothèque ciblée (les solutions), l'équation QSAR utilisée, a été établie à l'aide de la signature pour résoudre les ensembles de signatures atomiques correspondantes à une activité souhaitée.

**En 2004**, J. Carla et al. [102] ont publié un article dans lequel ils présentent une méthodologie pour résoudre le problème inverse -QSAR à l'aide du descripteur moléculaire

appelé signature. Cette méthodologie est détaillée en quatre parties. Tout d'abord, ils ont créé une équation QSAR qui corrèle l'occurrence d'une signature avec les valeurs d'activité à l'aide d'une technique de régression linéaire multiple par étapes. Deuxièmement, ils ont construit des équations de contrainte, en particulier les équations de graphicalité et de cohérence, qui facilitent la reconstruction des composés de la solution directement à partir des signatures. Troisièmement, ils ont résolu l'ensemble des équations de contrainte à la fois linéaires et diophantiennes. Enfin, ils ont reconstruit et énuméré les molécules de la solution et calculé leurs valeurs d'activité à partir de l'équation QSAR. Ils ont appliqué cette méthode inverse -QSAR à un petit groupe d'inhibiteurs du peptide LFA-1 / ICAM-1 afin de faciliter la recherche et la conception de composés inhibiteurs plus puissants. De nombreux nouveaux inhibiteurs ont été prédits, nombre d'entre eux devraient être plus puissants que l'inhibiteur le plus puissant de l'ensemble de criblage. Deux des inhibiteurs les plus puissants ont été synthétisés et testés *in vivo*, ce qui confirme qu'ils sont les peptides inhibiteurs les plus puissants à ce jour. Certains de ces composés peuvent être recyclés pour former un nouveau QSAR et développer une bibliothèque plus ciblée de composés de plomb.

Un petit ensemble de ces peptides dérivés comprend quatorze de ces peptides ont été utilisés en tant qu'ensemble de criblage pour le processus d'inverse -QSAR; les peptides 4 et 13 ont été utilisés comme ensemble de test. Une équation QSAR linéaire pourrait être créée en fonction d'une matrice de descripteurs  $47$  (signatures atomiques uniques de hauteur-1)  $\times$   $14$  (valeurs expérimentales des peptides  $IC_{50}$ ).

**En 2006**, V. Donald et al. [87], ont appliqué la méthode Inverse-QSAR au développement d'inhibiteurs de phosphate Cdc25B. Ce travail a été sélectionné en fonction de l'importance du rôle des phosphates Cdc25 dans la croissance et le développement des cellules eucaryotes. Une base de données de 28 inhibiteurs de Cdc25B connus avec des données d' $IC_{50}$  a été obtenue pour créer un QSAR.

En raison de la difficulté de résoudre les équations de contrainte Cdc25B, ceci est dû au fait que les composés de l'ensemble de criblage étaient plus volumineux (77 à 100 atomes) avec des liaisons aromatiques et doubles. En conséquence, ils ont développé des algorithmes modifiés pour gérer la solution des équations de Diophantine générées. Des

étapes supplémentaires de l'algorithme I-QSAR ont été développées pour appliquer la règle de Fins de Lipinski.

**En 2009**, W. L. William et al. [88] ont appliqué la méthode Inverse-QSAR pour la découverte d'un nouveau médicament (drug) à l'aide d'une méthodologie par noyau. Cette approche inverse-QSAR peut être décrite en cinq étapes. Les deux premières étapes consistent en une analyse QSAR. Dans la première étape, ils génèrent un VSMMD (descripteur moléculaire de modèle d'espace vectoriel) pour chaque composé de l'ensemble d'apprentissage (criblage). Ensuite, lors de la deuxième étape, ils utilisent une fonction du noyau pour mapper chaque VSMMD vers un espace de caractéristiques généralement utilisé pour la classification. La troisième étape consiste à concevoir ou à générer un nouveau point dans l'espace des caractéristiques du noyau en utilisant un algorithme d'espace des caractéristiques du noyau (par exemple, le centre des composés très actifs). Dans la quatrième étape, ils mappent ce point à partir de l'espace de fonctionnalités retour à l'espace d'entrée en utilisant un algorithme d'approximation de pré-image. Dans la dernière étape, la matrice de structure moléculaire sera construite par leur algorithme de récupération de molécules VMD.

**En 2010**, C. Derick et al. [117] ont appliqué l'algorithme CAMD à l'aide du descripteur moléculaire Signature. Ils ont ainsi identifié de nouveaux solvants respectueux de l'environnement, en complément du guide de sélection des solvants de GlaxoSmithKline (GSK). Ils génèrent des QSPRs en utilisant la signature pour classer les solutions inverses par rapport aux composés connus pour les déchets environnementaux, l'impact environnemental, la santé, la sécurité et l'évaluation du cycle de vie (LCA). Ensuite, ils résolvent le problème de conception inverse, les marquent avec le QSPR, appliquent divers filtres et, finalement, génèrent les structures. En bref, ils appliquent l'algorithme lié à ce problème. Ils nous ont expliqué que l'application de cette approche est beaucoup plus rapide, moins coûteuse et permet une recherche plus complète pour les candidats les plus appropriés que de travailler uniquement avec des données expérimentales. Ils présentent des résultats pour des solvants aux propriétés prédites optimales qui couvrent les classes des 47 composés du guide de sélection des solvants de GSK et en inclure plusieurs qui sont des hybrides qui se recoupent parmi les classes. De plus, leur technique «redécouvre» le solvant vert connu du lactate d'éthyle par ce procédé en combinant différentes classes de solvants.

**En 2012**, B. K. Kristian [122] a mis au point une relation quantitative -inverse structure-activité à partir des tensions de surface critiques des polymères. Ceci dans le but d'une nouvelle recherche sur les matériaux membranaires. La fabrication de nouvelles membranes pour la séparation des dispersions secondaires d'eau dans le diesel. Les tensioactifs situés dans le biodiesel diminuent la distribution de la taille des gouttes et augmentent la concentration en eau totale, ce qui diminue l'efficacité de la combustion et la durée de vie des moteurs.

La séparation des dispersions secondaires d'eau dans le diesel a déjà été démontrée comme étant dépendante du caractère hydrophobe de la surface de la membrane. L'hydrophobicité d'une surface dépend à la fois de l'énergie et la morphologie de la surface.

L'équipe de travail a utilisé deux modèles i-QSAR où le premier contenait les angles de contact avec de l'eau de 44 films de polymère, le second était basé sur les tensions superficielles critiques des polymères et consistait en 39 polymères, il a été utilisé dans les deux cas des signatures atomiques de niveau 1 et de niveau 2, en tant que variables indépendantes pour les matrices de descripteurs QSAR, l'évaluation ayant abouti à l'achat de poly (styrène-co-a-méthylstyrène), PSMS, produit caractérisé comme étant hydrophobe.

**En 2014**, K. Hamed [101] a mis au point de nouveaux adjuvants chimiques pour béton, la nouvelle approche Conception moléculaire assistée par ordinateur (CAMD) spécifiquement en ce qui concerne l'application de l'I-QSPR a été présenté dans la demande de conception de deux types d'adjuvants chimiques présentant un intérêt: (1) les adjuvants réducteurs d'eau (WRA) et (2) les adjuvants réducteurs de retrait (SRA).

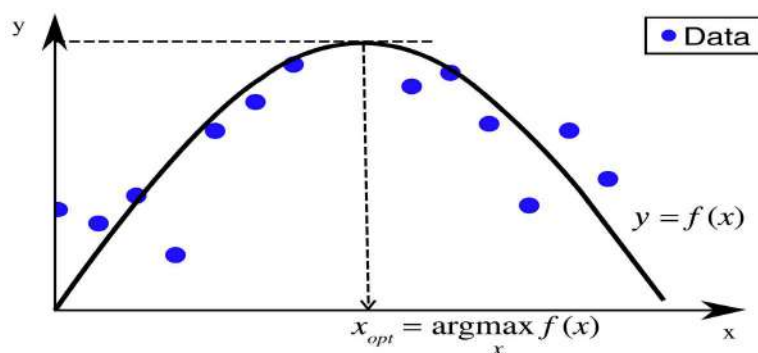
I-QSPR pour la conception de nouvelles méthodologies potentielles de WRAs / RMAs a été présenté à l'aide de dix-huit produits chimiques à bas poids moléculaires appartenant à différentes classes chimiques ont été testés à l'aide d'un rhéomètre à plaques parallèles pour leur effet sur les propriétés rhéologiques de la pâte de ciment. En particulier, l'effet des acides carboxyliques hydroxylés, des acides de sucre et des sucres a été évalué. Quatre modèles de statistiques QSPR pour les signatures de hauteur-1, la maniabilité, le temps de prise, la viscosité et la contrainte de rendement en utilisant la régression de la machine à vecteurs de support (SVR).



En tant que résultats expérimentaux, quatre conjectures structure-propriété ont déjà été développées en comparant les valeurs de pente de tension superficielle calculées à partir des données de la littérature. Pour tester ces conjectures, 14 nouveaux composés dans chacune des quatre conjectures ont été sélectionnés et testés pour leur effet sur la réduction de la tension superficielle de l'eau. Chaque composé a été testé trois fois et une valeur moyenne a été rapportée.

**En 2017**, M. Tomoyuki et al. [89] n'ont pas abordé la méthodologie de la phase de construction du graphe chimique de la inverse-QSAR. Leur objectif principal et leur plus grand défi était d'atteindre l'optimiseur et l'optimisation du descripteur de la méthode inverse-QSAR en deux étapes. Parallèlement, l'accent a été mis sur l'élaboration d'une nouvelle méthodologie permettant d'optimiser les paramètres de descripteur en ce qui concerne l'activité du composé supérieure à celle observée, en tant que condition préalable à la génération de la structure candidate. À cette fin, une approche plus avancée est introduite pour identifier les coordonnées du descripteur correspondant à l'activité prédite la plus élevée dans le domaine d'applicabilité d'un modèle QSAR donné.

Puisque l'inverse-QSAR basé sur la construction de la dérivation des coordonnées du descripteur pour un modèle et un ensemble de données. Pour cela, il est nécessaire de trouver des coordonnées souhaitables dans un espace de descripteur prédéfini (espace  $\mathbf{x}$ ) à partir d'une fonction de régression  $\mathbf{f}(\mathbf{x})$  représentant un modèle QSAR. Lorsque vous limitez la recherche dans le champ d'applicabilité d'un modèle QSAR, vous devez entrer dans un problème d'optimisation contrainte (COP). L'idée de l'optimisation a étudié est illustrée dans la figure 2 suivante.



**Figure I-11:** Concept d'optimisation.

Une fonction de régression  $f(\mathbf{x})$  ajuste les données d'apprentissage pour déterminer de nouvelles coordonnées dans l'espace des descripteurs. Les coordonnées optimisées basées sur  $f(\mathbf{x})$  tombent dans la plage de données d'apprentissage, mais produisent une valeur  $y$  supérieure à celle de tout autre point de données.

Les coordonnées nouvellement dérivées devraient être plus souhaitables en ce qui concerne les critères d'évaluation prédéfinis que tout autre point de données utilisé pour construire le modèle de régression.

Les principaux résultats obtenus par cette équipe sont les suivants: L'optimisation des coordonnées du descripteur pour les valeurs préférées d'un modèle donné est un aspect central du processus de la méthode inverse-QSAR en deux étapes, pour lequel il n'existe actuellement que des solutions approximatives. Cette approche affichait de meilleures performances que les méthodes évolutives alternatives, telles que les algorithmes génétiques (GA) ou l'optimisation d'essaims de particules (PSO). Cet algorithme simple à appliquer aux problèmes où plusieurs contraintes doivent être équilibrées, comme c'est le cas en inverse-QSAR.

## I-11 Conclusion

Dans ce chapitre nous avons mis en exergue le coefficient de partage en termes de  $\log K_{ow}$  son importance et son utilité ainsi que son application dans les domaines de la santé, l'environnement et sa relation avec d'autres propriétés physico-chimiques. Les différentes méthodes de détermination expérimentale de  $\log K_{ow}$  et par modélisation ont été aussi mis en évidence dans ce chapitre. L'application de  $\log k_{ow}$  dans le domaine environnemental surtout pour les composés chimiques appartenant à la classe des pesticides qui ont été encore étudié d'une façon particulière. Le chapitre est terminé par un aperçu historique sur l'évolution des méthodes I-QSAR dédiées à la prédiction des propriétés physicochimiques et à la génération de nouvelles structures moléculaires appliquées dans le domaine environnemental.

**CHAPITRE II : CONCEPTION MOLECULAIRE ASSISTEE PAR ORDINATEUR (CAMD) ET LEUR UTILISATION DANS LES EQUATIONS INVERSE-QSPR**

**II-1 Introduction**

Les logiciels sont des outils qui aident les chercheurs et les chimistes à la conception d'une grande variété de produits chimiques, ceux-ci doivent satisfaire aux multiples spécifications de haute performance et de compatibilité environnementale. Les logiciels préférés sont ceux qui peuvent afficher en tridimensionnelles des images informatiques vives sur des stations de travail de hautes performances. Cette classe de logiciels dite conception moléculaire assistée par ordinateur, en Anglais CAMD (Computer Aided Molecular Design) est devenue nécessaire et elle est très répandue dans la recherche dans le domaine pharmaceutique, biologique et environnemental. Elle fournit un excellent aperçu de la structure et de l'activité des médicaments et des composés étudiés. Les logiciels CAMD génèrent toutes les structures moléculaires possibles [90]. Au cours de cette étape, nous pouvons limiter les types de produits chimiques conçus. L'utilisation de groupes structuraux comme blocs de construction nous permet d'utiliser des techniques d'estimation de la contribution d'un groupe pour prédire les propriétés de toutes les structures générées. Pour l'estimation des propriétés de composés purs, les méthodes de contribution de groupe telles que celles données par Joback et Reid [91], Lydersen [92], Ambrose [93], Klinecicz et Reid [94], Lyman et al [95] et Horvath [96] ont été largement utilisés. Dans ces méthodes, la propriété d'un composé est fonction de paramètres structurellement dépendants, qui sont déterminés en faisant la somme de la fréquence de chaque groupe dans la molécule multiplié par sa contribution. Ces méthodes offrent l'avantage des estimations rapides sans nécessiter des ressources informatiques considérables. L'utilisation de techniques d'estimation de la contribution de groupe permet aux logiciels CAMD d'évaluer de nouveaux composés.

Une approche raisonnable du problème du CAMD devrait être capable de prendre en compte une grande diversité de structures sans rencontrer de difficultés de calcul importantes [97]. La méthode QSPR qui capture les aspects des méthodes basées sur GC (Group Contribution) et TI (Topological Indices) est celle des descripteurs de signature (SD). Les descripteurs de signature sont beaucoup plus jeunes que les autres méthodes discutées. Ils ont été développés par Visco et al [98] et par Faulon et al [99].

Afin de comprendre le concept, le rôle et les caractéristiques de la signature, nous avons défini dans ce chapitre la signature atomique et le nombre d'occurrences, suivi d'une définition complète des spécifications de la signature moléculaire, puis nous avons discuté les caractéristiques et la manière de construire les modèles QSPR à l'aide de la méthode statistique de régression linéaire multiple. Enfin on a fait un rappel sur les principales caractéristiques de la signature.

## II-2 La conception de signature

La signature est une ramification où les atomes d'une molécule sont liées selon un système de codification donné, tel que le voisinage est décrit par un alphabet de types d'atomes [100] et où la connectivité des atomes, dans une molécule, obéit à la théorie de graphe simple et à une méthode efficace d'encodage de topologie locale dont le degré de ramification est réglé par une distance prédéfinie appelée « niveau » de la signature.

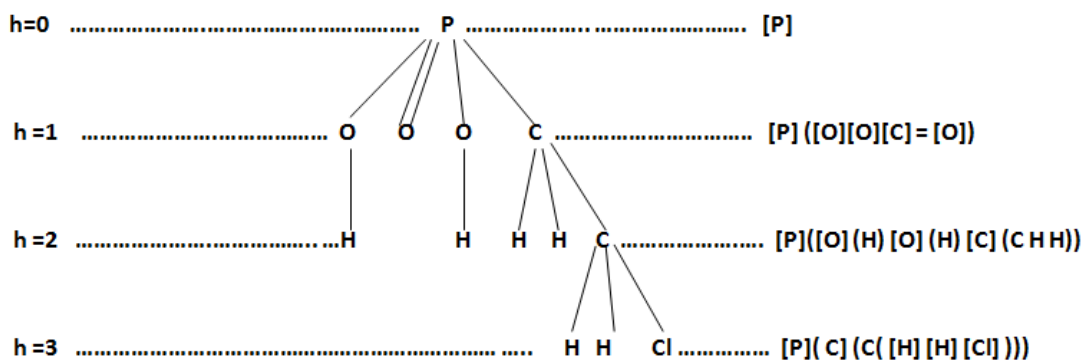
## II-3 Définition de la signature atomique

La signature, qui trouve son origine dans les études d'élucidation structurelle de Faulon, est basée sur le graphe moléculaire 2D d'une molécule,  $G = (V_G, E_G)$ , où les éléments en  $V_G$  désignent les atomes de la molécule, et les arêtes de  $E_G$  correspondent aux liens entre ces atomes. Nous définissons une signature atomique,  $^h\sigma_G(\mathbf{x})$ , comme le sous-graphe canonique de  $G$  composé de tous les atomes situés à une distance  $h$  de la racine  $\mathbf{x}$ .

Certaines signatures atomiques pouvant apparaître plusieurs fois dans le même composé, elles sont pondérées par leur occurrence avec laquelle elles apparaissent dans le graphe moléculaire [101].

Une fois le niveau de signature spécifié, la signature moléculaire de chacun des  $N$  composés identifiés à l'étape 1 est calculée à l'aide d'un programme informatique.

La méthode de construction d'une signature moléculaire à partir d'un ensemble de signatures atomiques uniques est illustrée dans l'exemple d'arbre de signature de la structure de l'Ethephon représentée dans la figure II-1, ci-dessous.



**Molecular Signatures for height-1:**

$$\sigma^1 = [P]([O][O][C]=[O]) + [O]([P]) + 2[O]([P][H]) + [C]([P][C][H][H]) + [C]([C][H][H][Cl]) + 2[H]([O]) + 4[H]([C]) + [Cl]([C])$$

**Figure II-1.** Illustre les signatures atomiques et moléculaires d'Ethéphon avec sa signature moléculaire au niveau-1.

De ce qui précède, toutes les liaisons atomiques s'étendent jusqu'à un niveau prédéterminé "h" qui s'étend du haut en bas sous forme arborescente, sans revenir en arrière. Le nombre de signatures atomiques augmente au fur et à mesure que le nombre de niveaux augmente. La figure II-1 montre qu'au niveau 0, la signature atomique est la racine elle-même, l'atome de phosphore. Pour le niveau-1, on a quatre atomes (C, O, O et =O), de ces atomes on a d'autres branchements au 2<sup>ème</sup> niveau et ainsi de suite..

Dans ce cas, la signature atomique de l'atome de phosphore terminal représentant la racine, apparaît sous la forme [P]([C][O][O]=[O]), où l'atome de phosphore a une liaison simple à un atome de carbone et à deux atomes d'oxygène, pendant qu'elle a une double liaison avec le troisième atome d'oxygène. Les nombres précédant les signatures atomiques représentent le nombre d'occurrences de chaque signature atomique unique dans une signature moléculaire. Comme exemple, [H]([C]) se trouve quatre fois dans le Ethéphon et est donc précédé de «4» dans la signature moléculaire. Les liaisons entre les atomes aliphatiques sont supposées être uniques, sauf indication contraire (i.e. "=" pour les doubles liaisons et "p[C]" pour liaison aromatique etc.). L'application de la procédure pour obtenir une base de données de toutes les signatures atomiques uniques avec des nombres d'occurrence pour tous les atomes d'une molécule constitue la base de données de la signature de la molécule, après sommation des signatures atomiques (par exemple, tous les atomes de phosphore, carbone, d'oxygène, d'hydrogène et de chlore), comme le cas de l' Ethéphon ( cf. **Tableau**

$$\text{II-1) ,on a : } \sigma^1 = [\text{P}][\text{O}][\text{O}][\text{C}]=[\text{O}] + =[\text{O}][\text{P}] + 2[\text{O}][\text{P}][\text{H}] + [\text{C}][\text{P}][\text{C}][\text{H}][\text{H}] + [\text{C}][\text{C}][\text{H}][\text{H}][\text{Cl}] + 2[\text{H}][\text{O}] + 4[\text{H}][\text{C}] + [\text{Cl}][\text{C}]$$

**Tableau II-1** : la base de données signatures atomiques uniques de hauteur-1 pour l'Éthéphon avec les numéros d'occurrence.

Niveau- 1 : Signatures Atomiques	Numéro d'occurrence
[P]([C][O][O]=[O])	1
[C]([ P][H][H][C])	1
[C]([ C][H][H][Cl])	1
[O]([P][H])	2
= [O]([P])	1
[H]([O])	2
[H]([C])	4
[Cl]([C])	1

## II – 4 Définition des descripteurs de signature moléculaire

Tout d'abord la signature moléculaire est définie comme la sommation de toutes les signatures atomiques uniques d'une molécule. Exactement, la définition la plus précise d'une signature moléculaire est la combinaison linéaire de ses signatures atomiques uniques [98]

En conséquence, des descripteurs codant des informations structurales significatives sont utilisés pour présenter les caractéristiques physicochimiques des composés afin de construire la relation entre structure et propriété dans cette étude. Le descripteur moléculaire utilisé dans ce projet était le descripteur moléculaire appelé Signature en raison de son succès à résoudre le problème I-QSAR. Le succès de Signature est triple. Tout d'abord, Signature effectue l'analyse QSAR ainsi que les descripteurs moléculaires conventionnels [102-98]. Deuxièmement, Signature a une dégénérescence plus faible que les autres descripteurs moléculaires et peut être contrôlé par l'utilisateur par une variable appelée hauteur. La signature moléculaire d'un composé est la somme de chaque signature

atomique unique multipliée par le vecteur d'apparition de cette signature atomique dans le composé donné. Elle peut être calculée à l'aide de l'équation suivante **II-1** ci-après [103].

$$h_{\sigma}(G) = \sum_{x \in V} h_{\sigma G}(x) \quad \text{II-1}$$

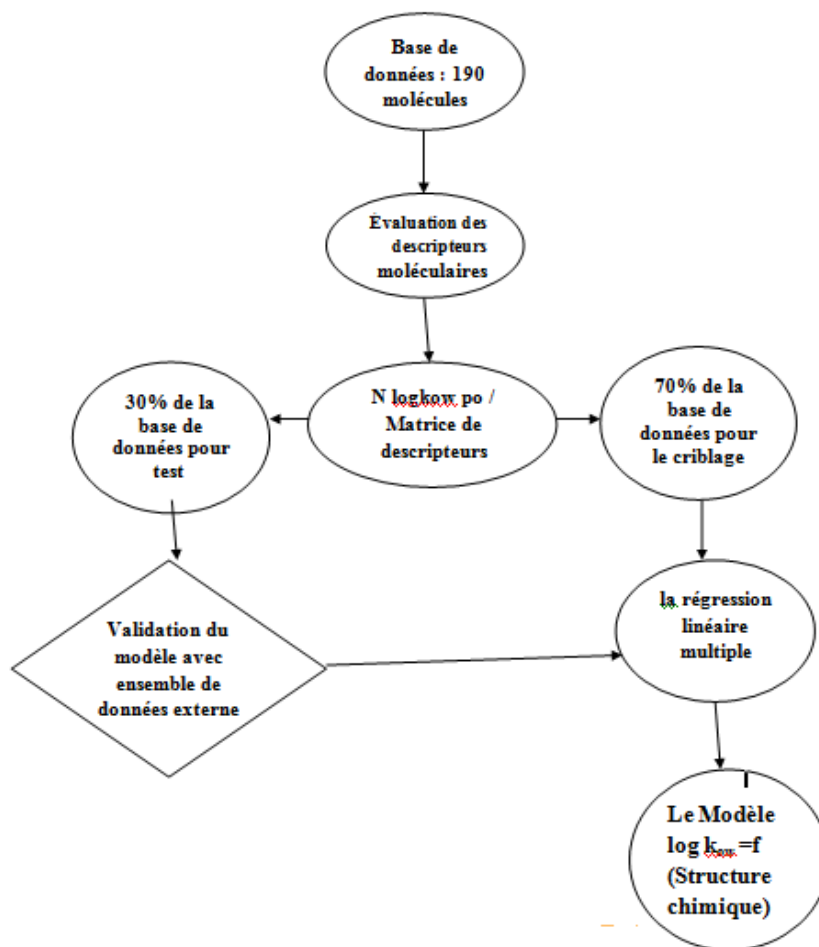
Où les éléments de  $V_G$  (matrice des sommets) sont les atomes ( $\mathbf{X}$ ), est l'ensemble de base de toutes les signatures atomiques de hauteur  $\mathbf{h}$  et  $h_{\sigma G}$  est le vecteur du nombre d'occurrences de signatures- $\mathbf{h}$  atomiques du graphe  $\mathbf{G}$ .

## II – 5 Modèles QSPR

Les modèles QSPR permettent aux utilisateurs de corrélérer les propriétés des composés pour les utiliser dans des algorithmes de conception. En se basant sur des méthodes théoriques et informatiques, les modèles QSPR complètent la recherche expérimentale [104]. En particulier, ils relient les propriétés complexes de différents matériaux ou composants moléculaires à leurs structures. La modélisation QSPR est basée sur l'hypothèse qu'il existe une relation entre la structure d'une molécule et les changements observés dans ses propriétés.

La relation structure-propriété quantitative QSPR en tant que domaine important de la chimiométrie a fait l'objet d'une série d'enquêtes [105-106]. Cette relation ou fonction est ensuite utilisée pour prédire les mêmes propriétés des composés non impliqués dans l'ensemble d'apprentissage à partir de leurs descripteurs structurels.

Les étapes d'élaboration d'un modèle QSPR est résumé à la **figure II-2**



**Figure II-2** : Méthodologie de développement des modèles QSPR.

La première étape consiste à choisir un ensemble de composés de la base de donnée (dans notre cas 190 pesticides) pour mesurer la propriété d'intérêt pertinente ( $\log k_{ow}$ ) pour une application spécifique. Les informations sur la structure des molécules de l'ensemble sont ensuite corrélées aux propriétés expérimentales obtenues. En particulier, cette étape implique la génération de relations linéaires ou non linéaires entre la propriété des molécules et les descripteurs de la structure (ici, les signatures atomiques). L'équation II-2 est une représentation plus formelle des relations structure-propriété:

$$y = f(x_1, x_2, \dots, x_N) \quad \text{II - 2}$$

Où une fonction (f) met en corrélation une propriété d'intérêt (y) avec les numéros d'occurrence de Signature atomiques uniques ( $x_i$ ). Enfin, la dernière étape consiste à valider le modèle QSPR développé (normalement via expérimentation) afin d'évaluer sa puissance prédictive, sa fiabilité et sa robustesse. Presque tous les modèles QSPR sont basés sur une



sorte de modèle de régression [107]. La régression peut être aussi simple qu'une régression linéaire multiple (MLR) ou aussi complexe que la régression vectorielle de support (SVR). Dans certains cas, toutefois, la relation existante entre la structure et la propriété d'intérêt est non linéaires. Par conséquent, une méthodologie non linéaire est parfois utilisée. Dans d'autres cas, la relation structure-propriété est basée sur une classe ou une catégorie dans laquelle des techniques d'apprentissage de machine à vecteurs de support (SVM) ou des arbres de décision sont utilisés. La section suivante présente la technique MLR.

### II-5-1 Régression linéaire multiple : MLR

RLM est une méthode la plus simple, reproductible et facilement interprétable pour développer des modèles QSPR [108]. La M RL est un outil statistique qui régresse les variables indépendantes par rapport à une variable dépendante. L'objectif de MLR est de trouver un modèle linéaire de la propriété d'intérêt [109]. De tels modèles sont basés sur l'hypothèse qu'il existe une fonction linéaire qui corrèle la propriété d'intérêt aux descripteurs. Dans le cas linéaire l'équation II-3 sera comme suite :

$$y = b + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_N x_N \quad \text{II-3}$$

Où  $y$  est la propriété d'intérêt,  $x_i$  sont les nombres d'occurrence de signatures atomiques,  $\alpha_i$  sont les coefficients de régression de l'ajustement obtenu à l'aide de la méthode des moindres carrés et  $b$  est la constante d'ajustement (y-intercept). Lorsque le modèle est non linéaire, l'ordre d'un ou de plusieurs descripteurs est supérieur à 1, ce qui étend le modèle à la régression polynomiale.

Dans certains cas, l'utilisation de la méthodologie MLR pour développer des modèles QSPR peut conduire à un problème de sur-ajustement en raison d'un plus grand nombre de variables indépendantes (signatures atomiques) par rapport au nombre de variables dépendantes (propriété d'intérêt).

Étant donné que les modèles QSPR visent à prédire la performance des structures, le sur-ajustement est souvent un problème car il réduit le pouvoir de prédiction du modèle. En particulier le sur-ajustement génère différents modèles qui utilisent un ensemble distinct de descripteurs, dont aucun ne peut être fiable pour prédire les propriétés de nouvelles structures. Diverses méthodes sont utilisées pour tenter de résoudre le problème de sur-ajustement et de réduire le nombre de variables indépendantes (descripteurs) utilisées dans

un modèle de régression. Ces techniques comprennent les moindres carrés partiels (PLS) et les composantes principales [110]. En outre, la régression linéaire par étapes à sélection directe ou inverse est souvent utilisée en tant que technique de RLM dans la mesure où il s'agit d'une méthode efficace en termes de calcul pour gérer le nombre de variables indépendantes incluses dans le QSPR. En particulier, cet algorithme ajoute les descripteurs les plus influents du modèle QSPR permettent d'abord d'observer l'effet de chaque descripteur [103]. Par exemple, il est possible d'éviter d'ajouter des descripteurs qui ont peu d'effet sur l'ajustement global ou le pouvoir prédictif du modèle afin d'éviter les sur-ajustements.

## **II- 6 Principales caractéristiques de la signature**

La principale motivation qui nous a incité à utiliser les descripteurs moléculaires signature sur d'autres méthodes disponibles utilisées dans les techniques de CAMD réside dans les principales caractéristiques bénéfiques de ce descripteur moléculaire.

Premièrement, Signature s'est révélée être le moins dégénéré par rapport aux autres descripteurs lorsqu'il a été évalué pour les alcanes, les alcools et les structures de fullerène. En particulier, Signature est capable de différencier des composés qui sont des isomères, bien qu'ils possèdent le même nombre d'atomes et la même masse moléculaire. En outre, Signature a une dégénérescence ajustable par laquelle le paramètre de niveau (h) est ajustable. En conséquence, Faulon et al. ont montré que la dégénérescence diminuait lorsque la hauteur de la signature augmentait [100]. Cependant, une dégénérescence plus faible peut produire des modèles QSPR de haute qualité qui sont la clé de l'algorithme CAMD. En particulier, un descripteur moléculaire de faible dégénérescence est essentiel et permet la différenciation entre les molécules actives et inactives.

Deuxièmement, Signature encapsule les informations d'ensemble de formation de molécules à partir desquelles de nouveaux descripteurs moléculaires peuvent être calculés.

Troisièmement, Signature permet clairement de passer des solutions numériques du problème inverse (voir les détails au chapitre III), avec de nombreuses signatures atomiques distinctes, aux structures réelles qui correspondent à des solutions.

## II- 7 Algorithme de conception inverse

Pour la conception moléculaire assistée par ordinateur, CAMD est une combinaison d'outils de chimie informatique et de technologie de l'information qui nous aident à découvrir de nouveaux composés utiles [111]. Sur cette base, l'utilisation généralisée du QSAR / QSPR-inverse est l'incarnation de ce concept.

QSAR / QSPR-inverse est connu comme les valeurs d'utilisation technique pour les variables indépendantes d'un composé particulier dans le QSAR / QSPR afin de résoudre l'activité / la propriété de ce composé (la variable dépendante). En revanche, l'objectif du problème QSAR / QSPR-inverse est de déterminer les valeurs des variables indépendantes en fonction de l'activité / propriété souhaitée [103]. Un problème de QSPR-inverse (I-QSPR) est un algorithme CAMD basé sur la signature qui identifie les composés possédant une certaine performance (ou propriété) d'intérêt prédit à l'aide d'un modèle QSPR développé [103]. L'I-QSPR est une technique interchangeable avec l'algorithme CAMD du descripteur de signature moléculaire. Un aperçu général de l'algorithme I-QSPR est fourni à la **figure II-3**, qui explique que cet algorithme est exécuté en neuf étapes fondamentales qui sont:

- (1) la sélection des composés de base de données;
- (2) génération des structures 2D;
- (3) traduction des composés de la base de données en signatures en plus de l'analyse QSPR;
- (4) génération d'équations de contrainte;
- (5) Résolution d'équations de contraintes puis en recherchant des solutions inverses;
- (6) Vérification des solutions pour la base de données;
- (7) stockage des solutions dans la plage souhaitée;
- (8) génération d'une nouvelle structure;
- (9) base de données ciblée.



### **Étape 3: traduction des composés de la base de données en signatures et analyse QSPR.**

La représentation structurelle des composés étudiés revêt une grande importance pour décrire, faire circuler et expliquer les informations structurelles significatives sur leurs caractéristiques. Sur la base de cette représentation, dans quelle mesure cette structure est liée à l'activité / propriétés physico-chimiques de la molécule étudiée. L'information structurelle d'une molécule est évaluée par des entités appelées descripteurs moléculaires.

Dans cette étape les données d'entrée de l'étape 2 sont transformés en signatures atomiques de niveau -1 pour chacun des N composés identifiés à l'aide d'un programme informatique conçu à cet effet. Ces signatures atomiques constituent le type de descripteur qui est utilisable dans le problème de conception inverse I-QSPR

Le descripteur distinctif et conforme aux conditions applicables à cette technique (I-QSPR) appelé signature [105-100]. C'est à cette étape que la base de données de signature atomique est utilisée comme variable indépendante lors de l'élaboration du QSPR ainsi que lors de la formulation du problème de conception de l'inverse-QSPR (étape 4).

#### **L'analyse QSPR.**

La QSPR prend des données expérimentales pour une propriété d'intérêt et les régresse en fonction des occurrences des signatures atomiques dans la base de données atomique Signature. En ce qui concerne cette partie, elle est détaillée dans la section modèle QSPR.

### **Étape 4: Génération des équations de contraintes**

La prochaine étape de l'analyse I-QSPR est la construction des équations de contrainte, qui servent à la construction de nouveaux composés en reconnectant des signatures atomiques en signatures moléculaires avec les propriétés souhaitées déterminées par l'équation QSPR [112]. Les équations de contrainte sont générées à partir de la base de données de signature atomique. L'équation de graphicalité est développée à partir des signatures atomiques de niveau-0 est une condition nécessaire pour un graphe connecté [98]. Le concept et la formule de L'équation II-4 de graphicalité est exprimé sous la forme suivante.

$$\text{Mod} \left( \sum_{i=2}^z (i-2)n_i - n_1 + 2, 0 \right) = 0 \quad \text{II-4}$$

Où  $z$  est le nombre maximum de sommets d'atomes dans l'ensemble de données, tandis que  $n_i$  est le degré de la racine de la signature atomique  $i$  [113].

Le second type d'équations est connu sous le nom d'équations de consistance. Ces équations sont des exigences relatives à la manipulation manuelle qui sont écrites pour chaque type de liaison dans le système. Fondamentalement, les signatures atomiques (enracinées en C) qui ont, par exemple, un C simple lié à H doit être apparié avec les signatures atomiques (enracinées en H) qui ont un H simple lié à C. L'ensemble des équations de cohérence (consistance) et graphicalité forment ensemble le système d'équations de contrainte. Ce système est sous-spécifié car il aura plus de variables (signatures atomiques) que d'équations. Étant donné que tous les coefficients dans les équations, ainsi que les solutions, sont des entiers, il s'agit d'équation diophantienne.

#### **Étape 5: Résoudre les équations de contrainte**

Puisque l'espace des solutions étant infini, nous limitons la plage que ces solutions (variables indépendantes) peuvent prendre, en fonction de leur plage dans l'ensemble de criblage d'origine. Les valeurs minimale et maximale (par signature atomique) fournissent les contraintes supplémentaires nécessaires à la résolution du système. En raison du grand nombre d'équations, nous avons utilisé les valeurs min / max dans l'algorithme appelé **PSO** (Particle Swarm Optimization) [114]. Cet algorithme semble satisfaire les équations de contrainte de manière progressive, de sorte que les itérations impliquant les variables qui apparaissent dans les équations vont du plus petit au plus grand nombre.

#### **Étape 6: Vérifier les solutions pour la base de données**

Puisque les équations de contrainte sont dérivées du nombre de composés utilisés, ce nombre devrait représenter des solutions aux équations de contrainte, elles sont évaluées en fonction de la liaison des atomes dans la molécule. C'est pour le but de vérifier la sortie de l'étape 5 et est important pour le débogage.

### **Étape 7: Conserve les solutions avec la plage souhaitée**

Dans cette étape, les solutions qui émergent de l'étape 5 doivent être notées pour l'adéquation par rapport à une valeur de propriété souhaitée (ou une plage de valeurs). Les solutions qui présentent l'aptitude souhaitée sont conservées, tandis que la signature atomique inappropriée doit être retirée de la solution. En conséquence, les solutions sont filtrées à travers la sortie QSPR de l'étape 3 et celles qui ont la forme souhaitée sont conservées. C'est à ce stade que diverses heuristiques peuvent être appliquées pour concentrer l'espace de la solution sur la base de connaissances expertes ou d'autres moyens [115].

### **Étape 8: Générer de nouvelles structures**

Les signatures moléculaires (solutions) qui émergent de l'étape 7 sont les signatures moléculaires à partir desquelles la génération de structure se produira. La génération de structure est réalisée à l'aide d'un algorithme développé par Faulon et Coworkers [116], qui repose sur un algorithme précédent de dénombrement d'isomères développé par Faulon (annexé dans des fichiers supplémentaires). L'algorithme est itératif, ce qui nécessite de commencer par une signature moléculaire de tous les atomes et de ne pas comporter de liaisons, puis d'essayer d'ajouter des liaisons de toutes les manières possibles pour correspondre à la signature moléculaire cible. Dans cette étape, nous avons sélectionné des structures appropriées après différents filtres pour éliminer ces structures candidates indésirables.

### **Étape 9: Base de données ciblée**

Les structures qui ont survécu jusqu'à ce point font partie de la base de données ciblée. Ce sont des structures de grande qualité qui méritent d'être approfondies. C'est ici que des expériences menées sur un certain nombre de composés pour vérifier les prédictions de l'algorithme seraient utilisées. Souvent, les résultats de l'expérimentation peuvent être utilisés pour affiner le QSPR et la base de données ciblée elle-même [115].

## **II- 8 Applications de la méthodologie I-QSPR**

En raison de la grande importance de la technique I-QSPR dans le domaine appliqué et qu'elle peut être utilisée pour concevoir de nouvelles structures pour n'importe quelle application souhaitée, quel que soit le domaine d'utilisation. À cet égard, Weis et Visco ont montré leur application dans la sélection des solvants et l'identification de nouveaux solvants écologiques [117]. En outre, il a été utilisé pour concevoir des agents gonflants pour les mousses écologiques [118]. Churchwell et ses collaborateurs ont utilisé les Signatures pour concevoir les inhibiteurs d'ICAM-1 les plus puissants [113]. Jackson et ses collaborateurs ont utilisé les Signatures pour concevoir les ligands des récepteurs des glucocorticoïdes [119]. Brown et ses collaborateurs ont utilisé la technique I-QSPR pour concevoir de nouveaux polymères aux propriétés ciblées [97].



---

---

**CHAPITRE III****METHODOLOGIE I-QSPR UTILISANT LA SIGNATURE****III -1 Introduction**

Le concept d'I-QSPR est un chemin, connu pour posséder une caractéristique particulière, c'est un moyen de créer un ensemble de données à partir de structures bidimensionnelles. Le résultat final de la méthodologie I-QSPR est un ensemble de composés (par exemple des structures 2D) avec des propriétés attendues améliorées en dehors de l'ensemble initial de composés spécifiques. Pour l'illustration de l'algorithme I-QSPR, un exemple est considéré, pour cela deux structures Dimethoate et la Formothion, forment l'ensemble de criblage. Cet exemple simple permettrait au lecteur de comprendre précisément la méthodologie.

**III-2 Sélection des composés de la base de données.**

Dans ce travail, seront avons sélectionné une base de données contenant 190 composés chimiques de la famille de pesticides, de classes différentes, ce choix est justifié, de point de vue que ces composées jouent un rôle important dans la vie humaine. Les valeurs des données expérimentales correspondantes à la propriété physico-chimique ( $\log K_{ow}$ ) ont été obtenues de la littérature [143] en plus de (structures chimiques [www.chemspider.com](http://www.chemspider.com) et [www.pubchem.com](http://www.pubchem.com)).

**III -3 Dessin des structures 2D.**

Après l'identification et la sélection de la base de données, nous avons enregistré les informations structurelles 2D des molécules, telles que les coordonnées atomiques et les liaisons chimiques sur des fichiers de format texte. Ces fichiers représentent les données d'entrée identifiées à l'étape 1 et qui seront calculés et transformés et sortie sous forme de signature moléculaire de niveau-1 pour chacun des N composés identifiés à l'aide d'un programme de traduction interne que nous avons développé. La figure III -1 montre une illustration du format texte de données d'entrée identifiées de la molécule Thiram.

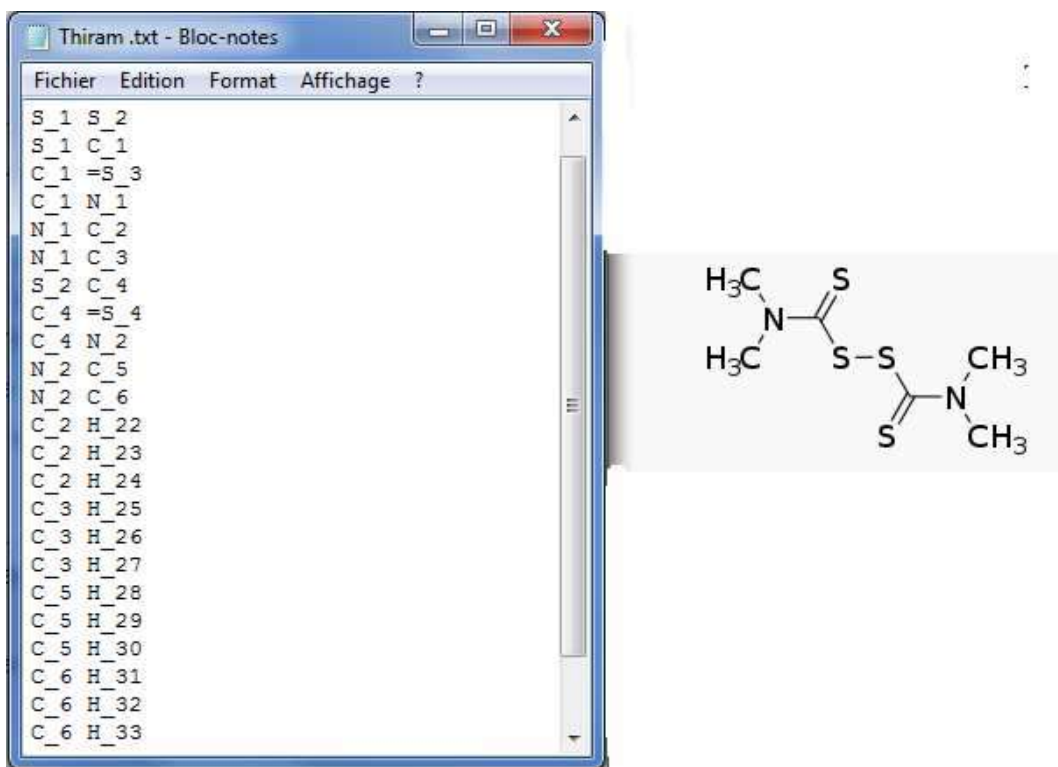
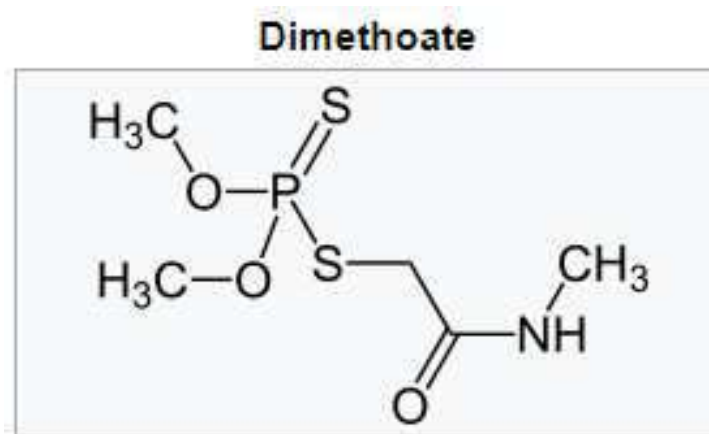


Figure III -1 Fichier en format texte de l'insecticide « Thiram ».

#### III-4 Traduction des structures 2D en signatures atomiques

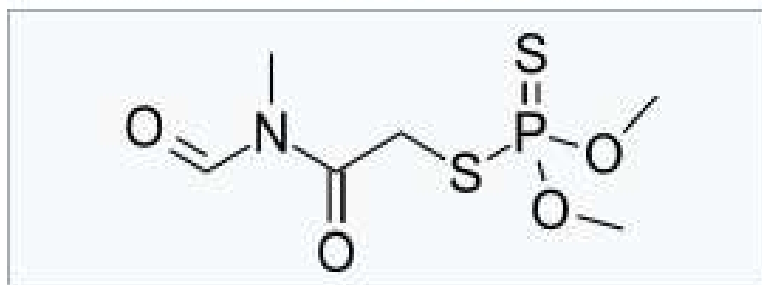
Les informations de structure 2D incorporées dans les fichiers de format texte sont utilisées pour traduire chaque structure en signatures atomiques correspondantes. Cette traduction commence par la première étape de sélectionner les atomes de racine. Les figures III -2 et III -3 illustrent différents atomes de racine à sélectionner dans les structures des insecticides « Dimethoate » et « Formothion » respectivement. Une fois qu'un atome de racine est sélectionné, l'algorithme vérifie tous les autres atomes connectés à l'atome de racine à une distance prédéfinie (c'est-à-dire au niveau "h" ). Par exemple, l'atome de carbone terminal de la Dimethoate et la Formothion illustré à les figures III -2 et III -3 apparaît trois fois dans les structures et est connecté d'abord à deux atomes d'oxygène et un autre atome d'azote d'une part et d'autre part lié à trois atomes d'hydrogène dans les trois cas. Tous les atomes de racine des carbones terminaux sont entourés de Dimethoate et de Formothion (voir Figure III-2 et Figure III-3). En conséquence, au niveau-1, les signatures atomiques uniques sont écrites en tant que [C]([O] [H] [H] [H]) et [C]([N] [H] [H] [H])). Tous les atomes connectés à un atome de racine apparaissent entre parenthèses indiquant la distance entre les atomes et la structure. Étant donné que toutes les

liaisons présentées pour ces deux exemples sont simples et doubles entre les atomes de racine et les atomes qui leur sont connectés. Une fois que tous les atomes d'une structure sont traités comme des atomes de racine, un ensemble de signatures atomiques uniques est obtenu. La signature moléculaire peut alors être écrite comme une somme de toutes les signatures atomiques uniques d'une structure. En conséquence, le nombre de fois où une signature atomique spécifique apparaît dans la structure est représentée par le nombre précédant chaque signature atomique unique. Ceci est connu comme le numéro d'occurrence. Par exemple, l'atome de carbone terminal Signature ([C] ([O] [H] [H] [H])) apparaît deux fois dans le Dimethoate et dans le Formothion (représenté par «2» sur les figures III-2 et III-3), tandis que la signature atomique unique [C] ([N] [H] [H] [H]) apparaît une fois dans les deux Dimethoate aussi bien dans Formothion (indiqué par «1» au figures III-2 et III-3). En conséquence, le même algorithme est appliqué à tous les atomes de chaque structure et finalement à toutes les structures de l'ensemble de criblage afin d'obtenir une base de données de signatures atomiques uniques au niveau spécifié.



**Figure III -2 Dimethoate = C( C S H H)+C( C =O N)+S( C P)+11 H(C)+P( S =S O O)+S(P)+2 O( P C)+=O(C)+N( C H C)+H(N)+C( N H H H)+2 C( O H H H)**

## Formothion



**Figure III -3 Formothion** = C( C S H H)+C( C =O N)+S( C P)+12 H(C)+2 =O(C)+N( C C C)+C( N H H H)+P( S =S O O)+C( N =O H)+=S(P)+2 O( P C)+2 C( O H H H)

Toutes les signatures atomiques de niveau-1 pour l'ensemble de criblage à deux composés sont présentées dans le **tableau III-1** ci-dessous, ainsi que le nombre d'occurrences minimum et maximum de chaque signature atomique dans l'ensemble d'apprentissage. Comme la base de données de signature atomique doit être utilisée plus avant dans l'algorithme, une variable est attribuée à chaque signature atomique unique pour des raisons de simplicité et de commodité.

**Tableau III-1.** Base de données des signatures atomiques pour le Dimethoate et le Formothion élaborée au niveau-1 avec les nombres d'occurrences minimum et maximum dans une molécule.

Signature atomique au niveau-1	Variable	Numéro d'occurrence [min, max]
$[C] ([C] [S] [H] [H])$	X1	[0,1]
$[C] ([C] = [O] [N])$	X2	[0,1]
$[S] ([C] [P])$	X3	[0,1]
$[H] ([C])$	X4	[11,12]
$[P] ([S] = [S] [O] [O])$	X5	[0,1]
$= [S] ([P] )$	X6	[0,1]
$[O] ([P] [C] )$	X7	[0,2]

$=[O]([C])$	X8	[1,2]
$[N]([C][C][H])$	X9	[0,1]
$[H]([N])$	X10	[0,1]
$[C]([N][H][H][H])$	X11	[0,1]
$[C]([O][H][H][H])$	X12	[0,2]
$[N]([C][C][C])$	X13	[0,1]
$[C]([N]=[O][H])$	X14	[0,1]

Pour souligner l'importance et le rôle de la base de données [120] de signatures atomiques, nous trouvons: premièrement permet de construire une matrice de descripteur utilisée dans l'élaboration des modèles QSPR et ensuite une base pour la dérivation des équations de contrainte nécessaires à l'obtention de solutions inverses.

### III -5 Elaboration de la matrice des descripteurs

La matrice de descripteurs est composée de lignes qui représentent les différents composés de l'ensemble de criblage, tandis que les colonnes représentent les signatures atomiques uniques. Les entrées de la matrice sont les numéros d'occurrence pour une signature atomique dans un composé particulier. En peut dire que la matrice de descripteurs nous donne des informations sur les signatures atomiques pour tous les composés entrant dans la construction du modèle QSPR écrit au format matriciel.

Pour illustrer le concept de la matrice de descripteur, prenons l'ensemble constitué des deux composés Dimethoate et Formothion où les composés (rangées) de l'ensemble de criblage apparaissent dans la matrice de descripteur dans l'ordre ASCII [121] illustrés dans le **tableau III -2**, ci-après.

**Tableau III -2** Matrice de descripteurs pour l'ensemble d'apprentissage des deux insecticides : Dimethoate et Formothion .

	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>	X <sub>10</sub>	X <sub>11</sub>	X <sub>12</sub>	X <sub>13</sub>	X <sub>14</sub>
<b>Dimethoate</b>	1	1	1	11	1	1	2	1	1	1	1	2	0	0
<b>Formothion</b>	1	1	1	12	1	1	2	2	0	0	1	2	1	1

Les valeurs zéro signifient qu'une signature moléculaire se produit zéro fois dans un composé particulier et cela veut dire que toutes les signatures atomiques ne sont pas présentes dans tous les composés. Par exemple, X<sub>9</sub> qui correspond à la signature atomique [N] ([C] [C] [H]) (voir tableau **III-1**) ne se produit pas dans la Formothion. Bien que nous trouvions cela se produit une fois dans la Dimethoate. En conséquence, les variables indépendantes (entrées de la matrice de descripteur) dans les modèles QSPR sont les occurrences des signatures atomiques et la variable dépendante est la propriété d'intérêt.

### III -6 Génération des équations de contraintes

Après compilation de la base de données de signatures atomiques, le processus est poursuivi par la génération des équations de contraintes associées à la base de données élaborée. Il est à noter qu'I-QSPR cherche toutes les combinaisons possibles de signatures atomiques uniques dans la base de données pour générer de nouvelles structures. Il est très peu probable qu'une combinaison aléatoire de signatures atomiques conduise à des structures réelles sans appliquer certaines contraintes pour imposer des restrictions de liaison et de valence aux atomes. Ces contraintes sont présentés dans des équations appelées équations de contraintes qui sont directement dérivés de la base de données de signatures atomiques. **La figure III-4** représente l'ensemble des équations de contrainte pour l'ensemble des deux composés comprenant la Dimethoate et la Formothion. Ces équations (l'équation de graphicalité et système d'équations de cohérence) forment un système d'équations dites diophantiennes où tous les coefficients et solutions du système sont des entiers.

Les équations de contrainte dérivées de Signatures de niveau-1 pour les deux composés Diméthoate et Formation s'expriment comme suit :

$$\text{Equation}<1> = \text{mod}(+1X_1+1X_2, 2) = 0 \quad \text{C-C}$$

$$\text{Equation}<2> = -1X_1+1X_3 = 0 \quad \text{C-S}$$

$$\text{Equation}<3> = -2X_1+1X_4-3X_{11}-3X_{12}-1X_{14} = 0 \quad \text{C-H}$$

$$\text{Equation}<4> = -1X_2+1X_8-1X_{14} = 0 \quad \text{C=O}$$

$$\text{Equation}<5> = -1X_2+2X_9-1X_{11}+3X_{13}-1X_{14} = 0 \quad \text{C-N}$$

$$\text{Equation}<6> = -1X_3+1X_5 = 0 \quad \text{S-P}$$

$$\text{Equation}<7> = -1X_5+1X_6 = 0 \quad \text{P=S}$$

$$\text{Equation}<8> = -2X_5+1X_7 = 0 \quad \text{P-O}$$

$$\text{Equation}<9> = -1X_7+1X_{12} = 0 \quad \text{O-C}$$

$$\text{Equation}<10> = -1X_9+1X_{10} = 0 \quad \text{N-H}$$

$$\text{Équation de graphicalté}<11> = \text{mod}(2X_1+X_2+X_3-1X_4+2X_5-1X_6+X_7-1X_8+X_9-1X_{10}+2X_{11}+2X_{12}+X_{13}+X_{14}, 2) = 0$$

**Figure III-4** : Equations de contrainte dérivées de Signatures de niveau-1 pour les deux composés d'ensemble de criblage Diméthoate et Formation

### III -6 -1 Équation de graphicalté

Sur la base de l'équivalence d'atomes individuels dans la base de données de signatures atomiques, l'équation de graphicalté décrit la connexion entourant l'atome en garantissant qu'un atome donné a exactement le même nombre de liaisons que l'équivalence. Cette équation est basée sur la théorie de graphes et satisfait à une condition nécessaire à la connectivité (c'est-à-dire un graphe connecté). Cependant, une solution qui satisfait l'équation de graphicalté peut entraîner la création de deux sous-graphes déconnectés, par conséquent, une telle équation n'est pas une condition suffisante. L'équation de graphicalté est un modulo, ainsi, pour qu'une signature moléculaire soit connectée, il faut que les signatures atomiques uniques soient complètes et les liaisons entre les atomes soient

réciroques [113], la formule III -1 ci-dessous montre comment établir l'équation de graphicalité :

$$\text{Mod} \left( \sum_{i=2}^z (i-2)n_i - n_1 + 2, 0 \right) = 0 \quad \text{III - 1}$$

L'équation de graphicalité est dérivée directement des signatures atomiques de niveau-0. Cela se fait essentiellement en réduisant la hauteur des signatures atomiques du tableau III-1 au niveau-0. les signatures atomiques suivantes [C] ([C] [S] [H] [H]) ; [C] ([N] [H] [H] [H]) et [C] ([O] [H] [H] [H]) ont le carbone comme signature atomique au niveau-0 lorsqu'il est réduit. Comme le carbone a une valence de 4, son sommet est 4 et donc  $i = 4$ . Le terme  $(i - 2)$  devient «2» pour les trois signatures atomiques, qui sont indiqués dans les coefficients précédents  $X_1$ ,  $X_{11}$ ,  $X_{12}$  dans la figure III-4. D'autre part, les signatures atomiques de niveau-1 [N] ([C] [C][C]) se transforment en [N] lorsqu'il est réduit au niveau-0, en prenant en considération que l'azote a une valence de 3. En conséquence, le premier terme de la graphicalité devient «1» pour cette signature atomique spécifique (c'est-à-dire  $X_{13}$  précédé de «1»). Toutes les signatures atomiques impliquant de l'hydrogène (H) en tant qu'atome racine résultent en un «-1» précédant la variable dans l'équation de graphicalité, comme indiqué par le terme «- $n_1$ » dans l'équation III -1.

À la suite de ce qui a été mentionné on peut dire que l'équation de graphicalité doit être satisfaite pour tous les composés de l'ensemble de criblage, car ils sont utilisés dans le développement de la base de données de signature atomique. En conséquence, l'équation de graphicalité donne une valeur de zéro, qui satisfait à l'exigence de module. Cependant, si un composé ne fait pas partie de l'ensemble de criblage et possède une signature atomique qui ne peut pas être attachée, l'équation de graphicalité ne peut pas être satisfaite et le résultat est un graphe déconnecté.

### III -6 -2 Équations de cohérence

Les équations de cohérence sont considérées comme le deuxième type d'équations de contrainte. Elles assurent que chaque signature atomique aura une place connectée dans la structure, ceci est basé sur l'information qui dit que les signatures moléculaires sont des collections (somme) de signatures atomiques interconnectées. Ces signatures moléculaires sont créées d'une manière qui assure qu'une liaison dans une signature



atomique aura une liaison correspondante dans une autre signature atomique, mais dans l'ordre inverse [122].

Les équations de cohérence mettent en application les exigences de valence pour tous les types de liaison qui apparaissent dans les composés de l'ensemble de criblage. Sur cette base tous les types de liaisons sont identifiés d'un atome de racine à son atome voisin dans la base de données [72]. A titre d'exemple, pour l'illustration, on considère que la troisième équation de cohérence sur la **figure III-4** écrite pour le collage **C-H**. Il convient de noter que les coefficients précédant chaque variable correspondent à une signature moléculaire qui nous indique combien de fois cette liaison (C-H) particulière se produit dans une signature moléculaire. Afin de clarifier cela, nous passons l'examen des signatures atomiques [C]([C] [S] [H] [H]), [C] ([N] [H] [H] [H] ), qui révèle que la liaison C-H apparaît deux fois et trois fois, respectivement.

Ceci est reflété dans les coefficients de l'équation de cohérence obtenue pour C-H où «-2» précède ( $x_1$ ) tandis que «-3» précède ( $x_{11}$ ). Cette liaison C-H apparaît également dans la douzième signature atomique, qui est également exprimé dans l'équation de cohérence correspondante. Le collage H-C représente l'inverse de C-H et apparaît dans la base de données des signatures sous la forme [H] ([C]) ou variable  $x_4$ . En conséquence, dans l'équation de cohérence de liaison C-H,  $x_4$  est précédé de «1». Il est à noter que lorsqu'un type spécifique de liaison ne se produit pas dans une signature atomique, la variable correspondante n'apparaît pas dans l'équation de cohérence. Par exemple, le type de liaison C-H n'est pas présent dans les signatures  $X_2, X_3, X_5, X_6, X_7, X_8, X_9, X_{10}$  et  $X_{13}$  et donc ils ne sont pas dans la troisième équation de cohérence. La même approche est utilisée pour dériver les autres équations de cohérence pour tous les autres types de liaison. Lorsqu'il s'agit de liaisons impliquant les mêmes atomes (C-C par exemple), l'équation devient une équation de module indiquant que le résultat doit être pair. Ceci est illustré par la première équation de la **figure III-4**.

### III -7 Résolution du problème inverse

Les solutions attendues au problème inverse doivent être satisfaites pour toutes les équations de cohérence ainsi que l'équation de graphicalité. Puisque le nombre de variables inconnus est beaucoup plus élevé que le nombre d'équations, ces solutions seront infinies. Par conséquent, les valeurs minimales / maximales des signatures atomiques sur

lesquelles la méthode de Weis et Visco [117] est utilisée, nous l'avons utilisé dans l'algorithme appelé **PSO** (Particle Swarm Optimization). Cet algorithme semble satisfaire les équations de contrainte de manière progressive, de sorte que les itérations impliquant les variables qui apparaissent dans les équations vont du plus petit au plus grand nombre.

Afin de résoudre ce système d'équations de contraintes pour l'exemple étudié ou l'ensemble de criblage constitué des deux composés Dimethoate et Formothion. En particulier, l'algorithme commence par le plus petit nombre d'occurrences (c'est-à-dire zéro pour cet exemple) et le plus petit nombre d'itérations possibles. Dans cet exemple, la méthodologie commence par une itération sur les variables  $X_1$  et  $X_3$ , les variables allant de 0 à 1, en fonction des nombres d'occurrences minimum et maximum. Une fois la solution satisfaite, l'algorithme passe à l'étape suivante où  $X_3$  et  $X_5$  sont itérés. Cela continue jusqu'à ce qu'une solution ne puisse pas être satisfaite et que l'algorithme recommence.

**Tableau III-3** Ordre d'exécution et nombre d'itérations pour toutes les variables suivies par l'algorithme pour l'exemple d'ensemble de criblage composé des deux composés Dimethoate et Formothion.

Étape	Variables dans l'équation	Variables à itérer à l'étape	Plage Min/Max pour les variables	Type de liaison	Nombre d'itérations
1	$X_1$ et $X_3$	$X_1$ et $X_3$	[0-1] et [0-1]	C-S	4
2	$X_3$ et $X_5$	$X_5$	[0-1]	S-P	2
3	$X_5$ et $X_6$	$X_6$	[0-1]	P=S	2
4	$X_1$ et $X_2$	$X_2$	[0-1]	C-C	2
5	$X_9$ et $X_{10}$	$X_9$ et $X_{10}$	[0-1] et [0-1]	N-H	4
6	$X_5$ et $X_7$	$X_7$	[0-2]	P-O	3
7	$X_7$ et $X_{12}$	$X_{12}$	[0-2]	O-C	2
8	$X_2$ , $X_8$ et $X_{14}$	$X_8$ et $X_{14}$	[1-2] et [0-1]	C=O	4

9	X <sub>2</sub> , X <sub>9</sub> , X <sub>14</sub> X <sub>11</sub> et X <sub>13</sub>	X <sub>11</sub> et X <sub>13</sub>	[0-1] et [0-1]	C-N	4
10	X <sub>1</sub> , X <sub>11</sub> X <sub>12</sub> , X <sub>14</sub> et X <sub>4</sub>	X <sub>4</sub>	[0,1]	C-H	2
	Une équation de graphicalité				1

Selon le tableau ci-dessus, toutes les itérations résultant des valeurs minimale / maximale de chacune des quatorze variables générées par les deux composées Dimethoate et Formothion sont égales à **24576** itérations.

### III-8 Modèles QSPR et enregistrement des solutions inverses

Dans les modèles QSPR les occurrences de signatures atomiques sont les variables indépendantes, tandis que la propriété d'intérêt est la variable dépendante. Les lignes de la matrice de descripteurs sont constituées de composés étudiés et de leurs colonnes représentées par le numéro d'occurrence de signature atomique donné dans un composé particulier. Dans les problèmes I-QSPR, tous les modèles QSPR sont créés en corrélant l'occurrence des signatures atomiques avec la propriété d'intérêt. Pour arriver à une bonne application du concept de QSPR, la méthodologie la plus simple pour créer ce modèle nécessite l'utilisation de l'algorithme MLR à pas en avant [113]. Le coefficient de corrélation  $R^2$  est une mesure de l'ajustement du modèle de régression.  $R^2$  représente la partie de la variation des données observées (expérimentales) expliquée par le modèle [123]. La valeur de  $R^2$  augmente avec l'augmentation du nombre de descripteurs ajoutés, en partie dans la mesure où elle résulte parfois de la valeur élevée de  $R^2$  un modèle suréquipé (sur-ajustement) [124]. Par conséquent, cela conduit à un modèle QSPR avec une faible puissance prédictive en dehors de l'ensemble de criblage.

L'acquisition d'un modèle QSAR de bonne qualité dépend de nombreux facteurs, tels que la qualité des données, le choix des descripteurs, sélection de variables, méthodes statistiques et validation [98-125]. Cette dernière est considérée comme une technique importante pour évaluer le modèle QSPR développé. La technique de validation croisée "leave-one-out", utilisée dans l'étude, garantit des modèles aussi stables et fiables que possible [126]. La technique de validation croisée implique la scission des données en

différents sous-ensembles, où le modèle avec la meilleure capacité prédictive calculée est sélectionné [127].

Une variante largement utilisée de la validation croisée est la validation croisée Leave-one-out (LOO-CV), où les observations sont omises une à la fois puis prédites en fonction de l'ajustement du modèle aux données restantes [128]. Cela peut être répété jusqu'à ce que tous les composés de l'ensemble de criblage soient utilisés et que la métrique de sur-ajustement soit ensuite calculée.

La qualité du modèle développé est très importante pour déterminer la capacité prédictive en dehors de l'ensemble de criblage car cela se reflète sur la nature du QSPR supporté à travers lequel les nouvelles structures générées sont évaluées, aussi bien fait l'enregistrement rapide des solutions obtenues à partir de résolution du problème inverse.

### III -9 Génération des nouvelles structures

Les solutions inverses obtenues en résolvant le système d'équations de contraintes et qui sont prédites par le modèle QSPR, représentent des structures potentielles qui doivent être construites à partir de signatures moléculaires. Il est à noter que pour une même signature moléculaire, il existe plusieurs structures, selon la dégénérescence liée à la hauteur de la signature. Dans le but de faire générer toutes les structures possibles on a utilisé l'algorithme de numérotation développé par Faulon et al [113].

Dans le but de générer une structure moléculaire correctement connectée représentant une structure correspondante à une signature moléculaire spécifique. L'algorithme de numérotation élaboré par Faulon, commence par un graphe moléculaire avec des sommets ou des atomes déconnectés et ajoute des arêtes ou des liaisons de toutes les manières possibles ou en conjuguant différentes combinaisons.

#### III -9-1. Développement de modèle QSPR requis

L'utilisation du modèle de la technique QSPR est nécessaire et essentielle à la réalisation de l'I-QSPR, car elle enregistre les solutions inverses, qui représentent les structures potentielles et elle est développée pour chaque propriété ou performance d'intérêt. Pour cette raison, le modèle QSPR a été développé à l'aide de la technique de régression linéaire multiple (RLM). La RLM est généralement utilisée comme outil statistique pour prédire l'événement de variable dépendante basée sur un ensemble de prédicateurs [129-130]. Cette technique statistique est utilisée pour étudier la relation entre une variable dépendante et plusieurs variables indépendantes. C'est une technique

mathématique qui minimise les différences entre les valeurs réelles et prédites [131]. Le modèle de régression linéaire multiple (MLR) a été généré à l'aide du logiciel Molegro, version 10, pour évaluer la capacité de prédiction de coefficient de partage octanol-eau (**logKow**), en utilisant la validation croisée Leave-One-Out (LOOCV).

### III-9-2 Validation du modèle et paramètres statistiques

L'ensemble de données doit être divisé en trois nouveaux ensembles de données, l'un pour le calibrage et la prévision (formation) et l'autre pour les ensembles de validation. L'ensemble de calibration a été utilisé pour la génération du modèle. L'ensemble de prédiction a été appliqué à la suralimentation du réseau, alors que l'ensemble de validation dont les molécules ne jouent aucun rôle dans la construction du modèle a été utilisé pour évaluer la capacité prédictive des modèles pour l'ensemble externe [132-133]. Un modèle réussi doit avoir la capacité de prédire non seulement la propriété des molécules de calibrage (validation interne) mais également des sources externes (validation externe). Parfois, un ensemble de données supplémentaires, nommé ensemble de test externe, est également utilisé. Les ensembles de criblage et de prédiction sont utilisés dans les étapes de développement du modèle et la capacité de prédiction globale du modèle est évaluée par application pour prédire la propriété de l'ensemble de test externe [134]. Par conséquent, les modèles QSPR sont dérivés de la partition initiale des composés et, par conséquent, la ventilation des données influence les performances du modèle développé [135]. Un modèle MLR est efficace lorsqu'il fonctionne bien sur un ensemble de données qui n'a pas été utilisé pour entraîner le modèle. Pour l'ajustement de l'équation du modèle en doit se baser sur des paramètres statistiques. Les paramètres statistiques mentionnés ci-dessous ont été utilisés pour évaluer la signification statistique du modèle QSPR.

L'équation III-2 exprime le coefficient de détermination (le carré de coefficient de corrélation)  $R^2$  qui est basé sur la somme d'erreur de prédiction des carrés.

$$R^2 = \frac{\sum_{i=1}^n (y_i^{\wedge} - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad \text{III-2}$$

$y_i$  est le  $\log K_{ow}$  expérimental dans l'échantillon  $i$ ,  $\hat{y}_i$  représente le  $\log K_{ow}$  prévu dans l'échantillon  $i$ ,  $\bar{y}$  est la moyenne du  $\log K_{ow}$  expérimental dans l'ensemble de prédiction et  $n$  est le nombre total d'échantillons utilisés dans l'ensemble de validation.

L'équation III-3 exprime le coefficient de la validation croisée  $Q^2$  qui est généralement plus petit que le  $R^2$  global pour une équation QSPR. (La différence entre  $R^2$  et  $Q^2$  ne devrait pas être plus de 0,3) [136]. Il est utilisé comme outil de diagnostic pour évaluer le pouvoir prédictif d'une équation.

$$Q^2 = 1 - \frac{\sum (y_i - y_{ipred})^2}{\sum (y_i - y_{mean})^2} \quad \text{III -3}$$

Où  $Y_i$  est la valeur expérimentale de  $\log K_{ow}$ ,  $Y_{ipred}$  est la valeur prédite de  $\log K_{ow}$  et  $Y_{mean}$  est la moyenne de la valeur expérimentale de  $\log K_{ow}$ .

L'équation III-4 montre l'erreur quadratique moyenne (RMSE) pour les ensembles de criblage et de test, est une mesure de la différence moyenne entre les valeurs prédites et les valeurs expérimentales, à l'étape de prédiction. La mesure de RMSE est utilisée pour évaluer la proximité de la ligne de régression de prévision avec un ensemble de valeurs réelles de la variable dépendante [137]. RMSE peut être interprété comme l'erreur de prédiction moyenne, exprimée dans les mêmes unités que les valeurs de réponse d'origine. Sa faible valeur indique que le modèle prédit mieux que le hasard et peut être considéré comme statistiquement significatif [138].

$$RMSE = \left[ \frac{1}{n} \sum_{i=1}^n (y_i^{\wedge} - \bar{y})^2 \right]^{\frac{1}{2}} \quad \text{III -4}$$

L'équation III -5 exprime l'écart-type (la déviation standard)  $S$  qui est probablement la mesure la plus communément déclarée et la plus importante de la propagation d'un ensemble de données. Il est étroitement lié à la variance car il est calculé en prenant sa racine carrée. Comme la variance est calculée différemment pour la population et pour les données de l'échantillon, il en est de même de l'écart type.

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

III -5

Où

$\sum$  somme des moyennes

$x_i$  représente chaque valeur de x dans la base de données

$\bar{x}$  la moyenne des valeurs  $x_i$

n le total des valeurs  $x_i$

## CHAPITRE IV

### RESULTATS ET DISCUSSION

Ce chapitre contient les résultats du modèle I-QSAR basé sur l'ensemble de criblage de coefficient de partage octanol-eau  $\log k_{ow}$ . Ces résultats sont basés sur les relations mentionnées aux chapitre **III** (section **III-1** jusqu'à **III-7**), auxquelles on a présenté les aspects théoriques et la méthodologie de l'élaboration des modèles I-QSAR avec un exemple détaillé pour les deux insecticides (Dimethoate et Formothion).

Le modèle QSPR a été construit en utilisant des signatures atomiques de niveau-1 pour l'ensemble de criblage. Sur la base des valeurs de  $R^2$  et de  $q^2$ , comme décrit au chapitre **III** (section **III-7-1** et **III-7-2**), ce qui nous a permis de construire le modèle I-QSAR.

#### IV- 1 Calcul des signatures atomiques

Sur la base de la méthodologie de l'élaboration du modèle I-QSPR mentionné et détaillé au chapitre **II** (section **II-5**), la traduction des structures moléculaires 2D en signatures atomiques peut être illustré dans le **tableau IV- 1** ci-dessous.

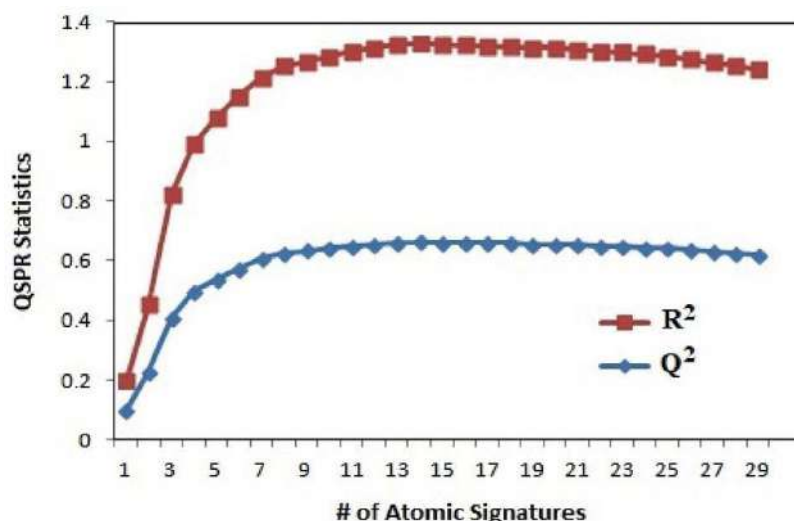
L'analyse QSPR a été calculée sur la base d'une matrice de descripteurs. La matrice de descripteurs pour la signature atomique de niveau-1 contenait **190** lignes et **253** colonnes, une colonne pour le  $\log k_{ow}$  et **252** colonnes pour les signatures atomiques uniques. Cependant, l'équation QSPR a été calculée uniquement sur la base de 29 signatures atomiques, éliminant ainsi 223 signatures atomiques, afin de réaliser la méthode de validation LOOCV [139]. Pour effectuer la régression multi linéaire pas en avant, 223 signatures atomiques uniques ont été retirées et nous en avons laissé 29 avec des nombres d'occurrences supérieurs ou égaux à 3 afin d'effectuer l'analyse LOOCV.



Tableau IV- 1. Signatures atomiques uniques de niveau -1 utilisées dans l'analyse QSPR.

Variable	Signature atomique au niveau-1	Occurrence [Min, max]
X <sub>2</sub>	[H]( [N])	[0,5]
X <sub>5</sub>	[O](=[C])	[0,3]
X <sub>6</sub>	[C]([C] [H] [H] H)	[0,5]
X <sub>8</sub>	[S]([P] [C])	[0,3]
X <sub>9</sub>	[O]([P][C])	[0,4]
X <sub>11</sub>	[C] ([O] [H] [H ] [H])	[0, 4]
X <sub>12</sub>	[H] ([C])	[0,29]
X <sub>14</sub>	p[C] ([N] p[C] p[C])	[0,4]
X <sub>18</sub>	p[C] ( p[C] p[C] [C])	[0,5]
X <sub>19</sub>	[C] ([C])	[0,4]
X <sub>20</sub>	p[C] ( p[C] p[C] [H])	[0,10]
X <sub>21</sub>	[H](p[C])	[0, 11]
X <sub>23</sub>	pC( p[N] p[N] [N])	[0,3]
X <sub>24</sub>	p[N] ( p[C] p[C])	[0,3]
X <sub>33</sub>	[O] ( p[C] [C])	[0,3]
X <sub>35</sub>	[O] (=S)	[0,4]
X <sub>37</sub>	C( [N] H] H] H])	[0,4]
X <sub>39</sub>	[N] ([C] [C] [C])	[0,3]
X <sub>43</sub>	[C](p[C][H][H][H])	[0,4]
X <sub>44</sub>	[pN] ( pC] pN])	[0,4]
X <sub>49</sub>	[C]([C] H] [H]S))	[0,3]
X <sub>52</sub>	p[C](p[C]p[C]C!)	[0,5]
X <sub>53</sub>	[C](p[C])	[0,5]
X <sub>57</sub>	p[C]([pC] p[C][O])	[0,3]
X <sub>64</sub>	[C]([O][C][H][H])	[0,4]
X <sub>74</sub>	[C]([C][C][H][H])	[0,7]
X <sub>75</sub>	[F] ([C])	[0,6]
X <sub>89</sub>	[H] ([O])	[0,3]
X <sub>104</sub>	[H](=[C])	[0,3]

Les signatures atomiques les plus significatives ont ensuite été ajoutées une à la fois, sur la base des valeurs  $R^2$  et  $Q^2$  calculées pour chaque étape, la **figure IV-1**, impact des signatures atomiques de niveau-1 de la base de données des pesticides étudiée, sur les statistiques QSPR, tracées en fonction du nombre de variables indépendantes, représente les valeurs de  $R^2$  et  $Q^2$  en fonction du nombre de variables indépendantes, c'est-à-dire les signatures atomiques au niveau-1.



**Figure IV-1.** Impact des signatures atomiques de niveau-1, des pesticides étudiés, sur les statistiques QSPR, tracées en fonction du nombre de variables indépendantes.

Les statistiques QSPR consistent à analyser l'amélioration de la corrélation avec l'augmentation du nombre de variables du modèle. La représentation des valeurs de  $R^2$  et  $Q^2$  en fonction du nombre de descripteurs (**Fig IV-1**) fait apparaître un comportement asymptotique, le modèle est considéré comme optimal lorsque l'amélioration de la corrélation devient maximale, c'est-à-dire représentant le meilleur compromis entre la corrélation et le paramétrage.

Les calculs ont été arrêtés à **14** signatures atomiques, ainsi la 15<sup>ème</sup> signature atomique était insignifiante. On a obtenu statistiquement le modèle QSPR validé en utilisant RLM pour prédire quantitativement les propriétés  $\log k_{ow}$  des structures candidates [140].

L'équation QSPR a été choisie sur la base du meilleur modèle de prévision, c'est-à-dire la valeur  $Q^2$  la plus élevée. Une équation QSPR à 14 signatures atomiques a été obtenue avec un

$R^2 = 0,659273$ ,  $Q^2 = 0,65617$  et  $RMSE_{\text{training}} = 0,930192$ ,  $s = 1,37297$  pour l'ensemble de criblage (training), et dans la valeur de l'ensemble de validation croisée Leave-one-out (LOO) avec une  $q^2 = 0,605676$  et une  $RMSE_{\text{LOO}} = 1,0936$ . Il est à noter que :  $R^2$  représente le coefficient de détermination,  $Q^2$  : le carré de coefficient de validation,  $s$ : écart type,  $RMSE$ : l'erreur quadratique moyenne.

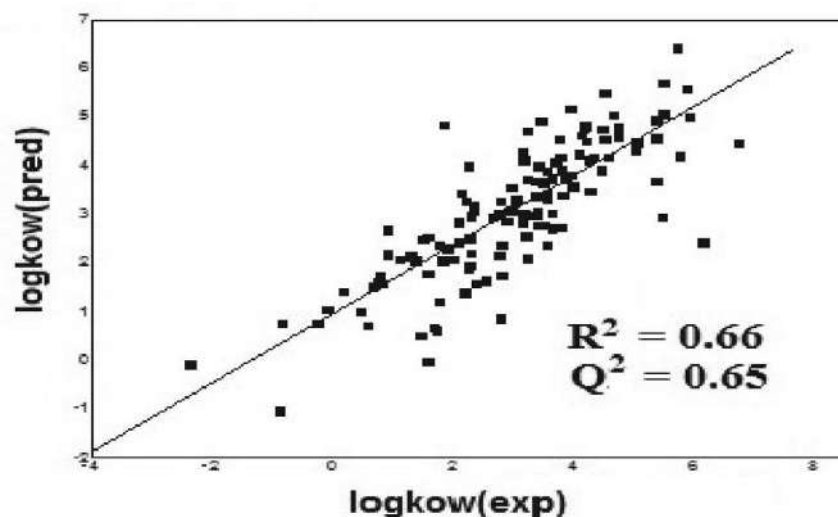
#### IV- 2 Equation modèle

L'équation modèle exprimant le coefficient de partage en termes de  $\log k_{ow}$  peut être écrite comme suit:

$$\log k_{ow} = - 0.167497 + 0.444669 * X_{52} + 0.417366 * X_6 + 0.785521 * X_{20} + 0.461849 * X_{75} + 0.0716288 * X_{12} - 0.419674 * X_5 - 0.409265 * X_{21} + 0.528737 * X_{19} - 0.220096 * X_{44} + 0.251859 * X_{74} + 0.365269 * X_9 + 0.226272 * X_{14} + 0.445708 * X_{53} + 0.0357723 * X_{24}.$$

Les 14 signatures atomiques incluses dans l'équation QSPR sont indiquées en gras dans le **tableau IV 1**. Les résultats statistiques et la signification de ce modèle final illustrent le fait que le coefficient de valeur positive élevée, correspond à la signature atomique  $X_{20}$ . Il a également été suggéré dans ce modèle que la signature atomique  $X_{24}$  était un contributeur nécessaire de  $\log k_{ow}$ , de sorte que la signature atomique  $X_5$  est affectée en tant que variable effective de  $\log k_{ow}$ , mais avec un coefficient négatif. L'utilisation de l'équation QSPR pour prédire  $\log k_{ow}$  des pesticides dans le même ensemble de criblage et la représentation graphique de ces valeurs par rapport aux données expérimentales est illustrée dans la **Figure IV- 2**. Le graphique montre les valeurs de  $\log k_{ow}$  prévues sur la base de l'équation du modèle qui est validée comme étant statistiquement significative par la validation croisée "leave-one-out cross-validation" par rapport aux valeurs expérimentales.

De toute évidence, les valeurs de  $\log k_{ow}$  prédites sont en bon accord avec les valeurs expérimentales. Les 14 paramètres du modèle fournissent une qualité statistique élevée:  $R^2 = 0,66$  et  $Q^2 = 0,65$ , et cela montre que la condition de la prévisibilité selon la considération de R. Veerasamy et al. [139], et A. Golbraykh et al. [141] sont satisfaits.



**Figure IV- 2:** Valeurs de  $\log k_{ow}$  prédites par rapport aux valeurs expérimentales pour l'équation QSPR pour 29 signatures atomiques de niveau-1.

### IV-3 Equations de contrainte

Les équations de contrainte sont élaborées dans l'ordre du plus petit nombre de paramètres au plus grand, c'est-à-dire, dans l'ordre dans lequel elles ont été résolues:

- (i) Équations de cohérence qui assurent l'alignement des atomes dans la construction des signatures moléculaires.
- (ii) Équation de graphicalité qui représente la valence de chaque atome.

Pour résoudre ces équations, une méthode développée par Weis et Visco [142] a été adoptée, et à cause de la large base de données et du grand nombre d'équations de contrainte, nous l'avons utilisée dans un programme basé sur l'algorithme PSO (Particle Swarm Optimization).

Les équations sont dressées dans le tableau **IV-2** ci-dessous.

Tableau IV-2: Equations de contrainte

N°	équations de contrainte
Eq.1	$\text{Mod}(+X_{217}, 2) = 0$
Eq.2	$\text{Mod}(+X_{242}, 2) = 0$
Eq.3	$+X_{44}+X_{169} = 2$
Eq.4	$-X_{45}+x_{48} = 0$
Eq.5	$-X_{91}+X_{92} = 0$
Eq.6	$-2X_{136}+X_{137} = 0$
Eq.7	$-X_{138}+X_{139} = 0$
Eq.8	$-X_{144}+X_{145} = 0$
Eq.9	$-X_{191}+X_{193} = 0$
Eq.10	$\text{Mod}(+X_{213}+X_{215}, 2) = 0$
Eq.11	$-X_{243}+X_{244} = 0$
Eq.12	$-X_{41}+X_{42}-X_{63} = 0$
Eq.13	$-X_{70}+X_{72}-X_{185} = 0$
Eq.14	$-X_{70}+X_{73}-X_{185} = 0$
Eq.15	$-X_{93}+X_{95}+X_{159} = 0$
Eq.16	$-X_{113}+X_{114}-X_{208} = 0$
Eq.17	$-X_{173}+2X_{174}-X_{175} = 0$
Eq.18	$-X_{206}+X_{207}+X_{209} = 0$
Eq.19	$-X_{52}+X_{53}-X_{119}-X_{187} = 0$
Eq.20	$-X_{88}+X_{89}-X_{153}-X_{195} = 0$
Eq.21	$-X_{149}+X_{150}+X_{226}+X_{250} = 0$
Eq.22	$-X_{168}+X_{169}-X_{186}-X_{246} = 0$
Eq.23	$-X_{121}+2X_{122}-X_{123}-X_{234}-X_{235} = 0$
Eq.24	$-X_{151}+X_{152}+X_{170}-X_{176}-X_{252} = 0$
Eq.25	$-X_1+X_3+X_{160}-X_{172}+X_{188}-X_{205} = 0$
Eq.26	$-X_{20}+X_{21}-X_{47}-X_{58}-X_{123}-X_{175} = 0$
Eq.27	$-X_{50}+X_{51}-X_{116}-X_{144}-X_{170}-X_{188} = 0$
Eq.28	$-X_{62}+X_{63}+X_{198}-X_{199}-X_{212}-X_{247} = 0$

Eq.29	$-X_{101}+X_{104}-X_{150}-2X_{155}-X_{165}-X_{226} = 0$
Eq.30	$-3X_{71}+X_{75}-X_{133}-3X_{167}-2X_{219}-X_{220}-2X_{239} = 0$
Eq.31	$-X_{109}+X_{112}-X_{142}-X_{207}-X_{215}-X_{250}+X_{251} = 0$
Eq.32	$-X_3+X_7-X_{59}-X_{96}-X_{134}-X_{152}-X_{160}-X_{232} = 0$
Eq.33	$-X_{17}+X_{19}-2X_{80}-3X_{103}-X_{115}-3X_{117}-2X_{133}-X_{245} = 0$
Eq.34	$-X_{32}+X_{33}-X_{57}+2X_{85}+X_{93}+X_{118}-X_{124}+X_{195} = 0$
Eq.35	$-X_3+X_8-X_{50}-X_{59}-2X_{96}+X_{146}-X_{170}-3X_{232}+X_{241} = 0$
Eq.36	$-X_{27}+X_{28}-X_{67}+X_{120}+X_{146}-X_{218}+X_{230}+2X_{236}+2X_{240} = 0$
Eq.37	$\text{Mod} (+X_{60}+X_{62}+X_{140}+X_{141}+X_{184}+X_{185}+X_{201}+X_{202}+X_{238}, 2) = 0$
Eq.38	$-2X_{34}+X_{35}-2X_{38}-2X_{95}-2X_{120}-2X_{159}-2X_{223}-X_{224}-2X_{230}-2X_{240}-2X_{241} = 0$
Eq.39	$-X_4+X_5-X_{30}-X_{56}-X_{61}-X_{65}-X_{66}-X_{82}-X_{163}-X_{171}-X_{177}-X_{222} = 0$
Eq. 40	$-X_{31}+2X_{34}-2X_{36}+X_{38}-X_{83}-X_{84}+X_{95}+X_{99}+2X_{107}+X_{120}-2X_{132}-X_{221} = 0$
Eq.41	$-X_1+X_2-X_{25}-X_{31}-2X_{46}-X_{69}-X_{79}-X_{83}-X_{162}-X_{184}-2X_{202}-2X_{205}-2X_{221}-X_{238} = 0$
Eq.42	$-X_{13}+X_{14}+X_{23}-X_{25}-X_{46}+X_{55}-X_{70}-X_{76}-X_{77}-X_{83}-X_{132}-X_{140}-2X_{162}+X_{234}-X_{238} = 0$
Eq.43	$-X_3+X_9-2X_{50}-2X_{59}-X_{96}-3X_{116}+X_{118}-3X_{134}-2X_{144}-2X_{152}+X_{153}-2X_{160}-X_{170}-2X_{188}+2X_{231}+X_{251} = 0$
Eq.44	$-2X_{23}+2X_{24}-2X_{27}-X_{32}+X_{44}+2X_{45}-2X_{47}-X_{55}-X_{58}-X_{94}-X_{119}-2X_{124}-2X_{125}-2X_{126}+X_{169}-2X_{187}-X_{218}-X_{234}-X_{235} = 0$
Eq.45	$-X_{40}+X_{41}-X_{105}+X_{149}-X_{177}+2X_{178}-X_{181}-X_{182}-X_{183}-X_{192}+X_{193}-X_{196}-X_{197}+X_{198}-X_{210}-X_{216}+X_{217}-X_{228}-2X_{248}-X_{249} = 0$
Eq.46	$-X_{76}-X_{77}+3X_{78}-X_{79}-X_{141}+X_{142}+X_{150}+X_{165}-2X_{166}-X_{180}+X_{194}-X_{199}+2X_{200}-X_{201}+X_{207}-X_{208}+X_{209}-X_{211}-X_{212}-2X_{214}+X_{215}+X_{250} = 0$
Eq.47	$+2X_{14}+2X_{18}+2X_{20}+X_{32}+2X_{42}+2X_{52}+X_{55}+2X_{57}+X_{58}+2X_{67}+3X_{86}+2X_{91}+X_{94}+X_{119}+X_{121}+X_{123}+X_{126}+2X_{138}+2X_{145}+X_{173}+X_{175}+X_{218} = 2$
Eq.48	$-X_{18}+X_{22}+X_{43}+X_{61}+X_{71}+X_{81}+X_{82}+2X_{87}-X_{94}-X_{121}-X_{125}+X_{127}+2X_{130}+X_{131}+X_{147}+X_{148}+X_{157}+3X_{161}+X_{164}-X_{173}+X_{179}+X_{197}-X_{235} = 0$
Eq.49	$-X_8+X_{10}-X_{28}-X_{38}+X_{49}+X_{54}+X_{66}+X_{81}+X_{97}-X_{99}+X_{103}+X_{108}-2X_{111}+X_{133}+2X_{156}-X_{159}+X_{179}+X_{182}+X_{196}-X_{206}+X_{210}-2X_{223}-2X_{224}-X_{230}+X_{237}-X_{241}-X_{242}+X_{243} = 0$
Eq.50	$-X_{100}+2X_{101}-X_{102}-X_{108}+2X_{109}-X_{110}-X_{135}+X_{136}+X_{142}-X_{143}-X_{154}+X_{155}-X_{164}+X_{165}-X_{181}-X_{189}+2X_{190}-X_{192}+2X_{194}+X_{200}-2X_{204}+X_{209}+X_{213}-X_{222}-X_{225}+X_{226}-X_{227}-X_{228}-2X_{229}-X_{245} = 0$

Eq. 51	$  \begin{aligned}  & -X_9+X_{11}+X_{15}-2X_{16}-X_{33}+X_{56}+X_{64}+X_{65}+2X_{68}+X_{82}+X_{87}-X_{88}+X_{90}+X_{106}-X_{112} \\  & X_{114}+X_{115}+X_{135}+X_{147}+X_{148}+2X_{158}+X_{161}+X_{167}+X_{189}- \\  & X_{191}+X_{196}+X_{219}+X_{220}+X_{228}+X_{246}+X_{252} = 0  \end{aligned}  $
Eq.52	$  \begin{aligned}  & -X_1+X_4-2X_{13}+X_{15}-X_{25}+X_{26}+X_{29}+2X_{30}-X_{31}-X_{36}+X_{37}-3X_{39}+X_{40}+X_{54}+X_{56}- \\  & 2X_{60}+X_{61}+X_{66}-2X_{69}-X_{76}-X_{79}-2X_{84}+2X_{105}-2X_{113}+X_{127}+X_{129}-X_{140}-X_{166}+X_{171}- \\  & 2X_{172}+X_{176}+X_{177}-2X_{180}-X_{184}+X_{189}-X_{201}+2X_{203}+X_{210}-X_{211}- \\  & X_{212}+X_{222}+X_{225}+2X_{227}+X_{233}+X_{243}-2X_{247}+X_{249} = 0  \end{aligned}  $
Eq.53	$  \begin{aligned}  & -3X_6-3X_{10}-3X_{11}+X_{12}-2X_{15}-2X_{17}-2X_{22}-2X_{26}-X_{29}-3X_{37}-X_{40}-3X_{43}-2X_{49}- \\  & 2X_{54}-2X_{64}-2X_{74}-X_{80}-2X_{81}-X_{90}-X_{97}-X_{98}-X_{100}-2X_{102}-3X_{110}-X_{115}-2X_{127}- \\  & X_{130}-X_{131}-X_{135}-X_{148}-2X_{151}-2X_{156}-X_{158}-2X_{168}-X_{171}-2X_{176}-X_{179}-X_{181}- \\  & 2X_{183}-X_{186}-2X_{203}-X_{204}-X_{220}-X_{225}-X_{239}-X_{246}-2X_{249}-X_{252} = 0  \end{aligned}  $
Eq.54	$  \begin{aligned}  & +X_4+X_6+X_{17}+X_{22}+X_{26}+2X_{29}+X_{49}+X_{64}+X_{65}+2X_{68}+2X_{74}+X_{80}+X_{87}+2X_{90}+2X_{97}+3X_9 \\  & 8+X_{100}+X_{102}+3X_{106}+X_{108}+X_{115}+X_{117}+4X_{128}+2X_{129}+X_{130}+2X_{131}+3X_{143}+2X_{147}+ \\  & X_{148}+X_{151}+2X_{154}+3X_{157}+X_{158}+2X_{163}+X_{164}+X_{168}+X_{179}+X_{182}+X_{183}+2X_{186}+X_{192}+ \\  & X_{197}+2X_{216}+X_{219}+X_{220}+X_{229}+3X_{233}+3X_{237}+X_{239}+X_{245}+X_{246}+X_{248}+X_{252} = 2  \end{aligned}  $
Eq.55	$  \begin{aligned}  & \text{Mod}( X_1-X_2+2X_3+X_4-X_5+2X_6-X_7+X_8+X_9+2X_{10}+2X_{11}- \\  & X_{12}+X_{13}+X_{14}+2X_{15}+X_{16}+2X_{17}+X_{18}-X_{19}+X_{20}- \\  & X_{21}+2X_{22}+X_{23}+X_{24}+X_{25}+2X_{26}+X_{27}+X_{28}+2X_{29}+X_{30}+X_{31}+X_{32}+X_{33}+2X_{34}- \\  & X_{35}+X_{36}+2X_{37}+2X_{38}+X_{39}+X_{40}+X_{41}+X_{42}+2X_{43}+X_{44}+X_{45}+X_{46}+X_{47}- \\  & X_{48}+2X_{49}+2X_{50}-X_{51}+X_{52} \\  & X_{53}+2X_{54}+X_{55}+X_{56}+X_{57}+X_{58}+2X_{59}+X_{60}+X_{61}+X_{62}+X_{63}+2X_{64}+X_{65}+X_{66}+X_{67}+2X_{68} \\  & +X_{69}+X_{70}+2X_{71}-X_{72}-X_{73}+2X_{74}- \\  & X_{75}+X_{76}+X_{77}+X_{78}+X_{79}+2X_{80}+2X_{81}+X_{82}+X_{83}+X_{84}+X_{85}+X_{86}+2X_{87}+X_{88}- \\  & 1X_{89}+2X_{90}+X_{91}- \\  & X_{92}+X_{93}+X_{94}+2X_{95}+2X_{96}+2X_{97}+2X_{98}+X_{99}+X_{100}+X_{101}+2X_{102}+2X_{103}- \\  & X_{104}+X_{105}+2X_{106}+X_{107}+X_{108}+X_{109}+2X_{110}+X_{111}+X_{112}+X_{113}+X_{114}+2X_{115}+2X_{116}+ \\  & 2X_{117}+X_{118}+X_{119}+2X_{120}+X_{121}+X_{122}+X_{123}+X_{124}+X_{125}+X_{126}+2X_{127}+2X_{128}+X_{129}+ \\  & 2X_{130}+2X_{131}+X_{132}+2X_{133}+2X_{134}+X_{135}+X_{136}-X_{137}+X_{138}- \\  & X_{139}+X_{140}+X_{141}+X_{142}+2X_{143}+2X_{144}+X_{145}+X_{146}+2X_{147}+2X_{148}+X_{149}+X_{150}+2X_{151} \\  & +2X_{152}+X_{153}+X_{154}+X_{155}+2X_{156}+2X_{157}+2X_{158}+2X_{159}+2X_{160}+2X_{161}+X_{162}+X_{163}+ \\  & X_{164}+X_{165}+X_{166}+2X_{167}+2X_{168}+X_{169}+2X_{170}+X_{171}+X_{172}+X_{173}+X_{174}+X_{175}+2X_{176}+ \\  & X_{177}+X_{178}+2X_{179}+X_{180}+X_{181}+X_{182}+2X_{183}+X_{184}+X_{185}+2X_{186}+X_{187}+2X_{188}+X_{189}+ \\  & X_{190}+X_{191}+X_{192}+X_{193}+X_{194}+X_{195}+X_{196}+X_{197}+X_{198}+X_{199}+X_{200}+X_{201}+X_{202}+2X_{20}  \end{aligned}  $

$ \begin{aligned} &3+X_{204}+X_{205}+X_{206}+X_{207}+X_{208}+X_{209}+X_{210}+X_{211}+X_{212}+X_{213}+X_{214}+X_{215}+X_{216}+X_{217}+X_{218}+2X_{219}+2X_{220}+X_{221}+X_{222}+2X_{223}+X_{224}+X_{225}+X_{226}+X_{227}+X_{228}+X_{229}+2X_{230}+X_{231}+2X_{232}+2X_{233}+X_{234}+X_{235}+X_{236}+2X_{237}+X_{238}+2X_{239}+2X_{240}+2X_{241}+X_{242} \\ &+X_{243}-X_{244}+X_{245}+2X_{246}+X_{247}+X_{248}+2X_{249}+X_{250}+X_{251}+2X_{252} \quad , 2) = 0 \end{aligned} $
---

#### IV-4 Génération de nouvelles structures de pesticides

La génération de nouvelles structure est basés sur les solutions inverses obtenues en résolvant le système d'équations de contraintes, de nouvelles structures peuvent être construites à partir de signatures moléculaires (solutions inverses). Il est à noter que pour une même signature moléculaire, il existe plusieurs structures.

En résolvant les équations de contrainte (**tableau. IV-2**), un nombre total de 5500 solutions (nouvelles signatures moléculaires) qui seront de nouvelles structures probables. Etant donné qu'il est difficile d'examiner plus de 5500 structures, les structures nouvellement créées ont été affinées en fonction de la structure chimique différente de tous les pesticides existants. Toutes les structures nouvellement générées ont été vérifiées et confirmées qu'elles sont inexistantes, via des sites de recherche de structure on line , tels que : ChemSpider, PUB Chem et LookChem, qui sont des bases de données reconnues et fiables pour identifier les composés disponibles en littérature.

Après toutes ces étapes, et la résolution des 55 équations de contraintes et l'acceptation des signatures moléculaires avec leurs valeurs de  $\log_{ow}$  correspondantes, qui ont abouti aux solutions de ces équations qui sont de nouvelles signatures moléculaires. Le **tableau IV-3** en donne un exemple.

Il est à noter que seules les signatures atomiques ayant une valeur supérieure à 0 sont décrites dans le tableau. Dans la phase finale, un ensemble de 20 échantillons a été sélectionné en fonction de leur  $\log_{ow}$ , ce qui est proche de la valeur moyenne de la base de données des 190 pesticides étudiée ( $\log_{ow} = 2,94$ ).



**Tableau IV-3.** Exemple de signature moléculaire résolue, notez que seules les signatures atomiques dont le nombre d'occurrences est supérieur à 0 sont illustrées.

Signature Moléculaire	X <sub>6</sub>	X <sub>8</sub>	X <sub>9</sub>	X <sub>10</sub>	X <sub>11</sub>	X <sub>12</sub>	X <sub>49</sub>	X <sub>50</sub>	X <sub>51</sub>	X <sub>64</sub>
une seule structure moléculaire	1	1	2	1	1	15	1	1	1	1

Les 20 composés identifiés ont été intégrés à la base de données ciblée. Afin d'évaluer le niveau de diversité parmi les structures nouvellement créées, celles-ci ont été comparées aux structures des ensembles de formation. En plus du travail effectué, nous avons comparé les valeurs de  $\log k_{ow}$  prévues par notre modèle à celles calculées par le logiciel Hyperchem (version 8). Il a été conclu que les valeurs des deux résultats (**Tableau IV-6**) sont proches dans la plupart des cas.

Pour plus d'illustration de la méthode de génération des 20 nouvelles structures, par exemple, si on prend la première qui possède la formule brute **C<sub>7</sub>H<sub>17</sub>O<sub>2</sub>PS<sub>4</sub>**, la solution (nouvelle signature moléculaire) de cette formule est déterminées en appliquant l'algorithme I-QSPR (**tableau IV-4**).

**Tableau IV-4** Solution (nouvelle signature moléculaire) de la structure ayant pour formule brute : **C<sub>7</sub>H<sub>17</sub>O<sub>2</sub>PS<sub>4</sub>**.

Signature Moléculaire	X <sub>6</sub>	X <sub>8</sub>	X <sub>9</sub>	X <sub>10</sub>	X <sub>11</sub>	X <sub>12</sub>	X <sub>49</sub>	X <sub>50</sub>	X <sub>51</sub>	X <sub>237</sub>	X <sub>242</sub>
une seule structure de formule C <sub>7</sub> H <sub>17</sub> O <sub>2</sub> PS <sub>4</sub>	2	1	2	1	2	17	1	1	1	1	2

#### IV-5 Explication de la génération de structure

La reconstruction d'une signature moléculaire commence par des signatures atomiques ne contenant que des atomes parents, c'est-à-dire sans liaisons. Des liaisons sont ensuite ajoutées par tous les moyens possibles pour saturer la valence de tous les atomes dans la signature moléculaire, cela est effectué dans l'ordre de l'atome avec la valence la plus basse à l'atome de valence la plus haute. Ceci est illustré à la **figure IV- 3**, où les atomes parents dans chaque signature atomique de **C<sub>7</sub>H<sub>17</sub>O<sub>2</sub>PS<sub>4</sub>** sont 3·H, 3·H, 3·H, 3·H, 3·H, 2·H,OO, SSSS, C,C,C,CC,CC, P.

Les formes, le nombre, les types et la manière de l'interconnexions de leur atomes, ainsi que le nombre d'occurrence des signatures atomiques permettant de générer la formule brute de la nouvelle structure, sont présentés sur le tableau IV- 5.

**Tableau IV- 5** Types, nombre d'occurrence et forme de signatures avec l'interconnexion des atomes.

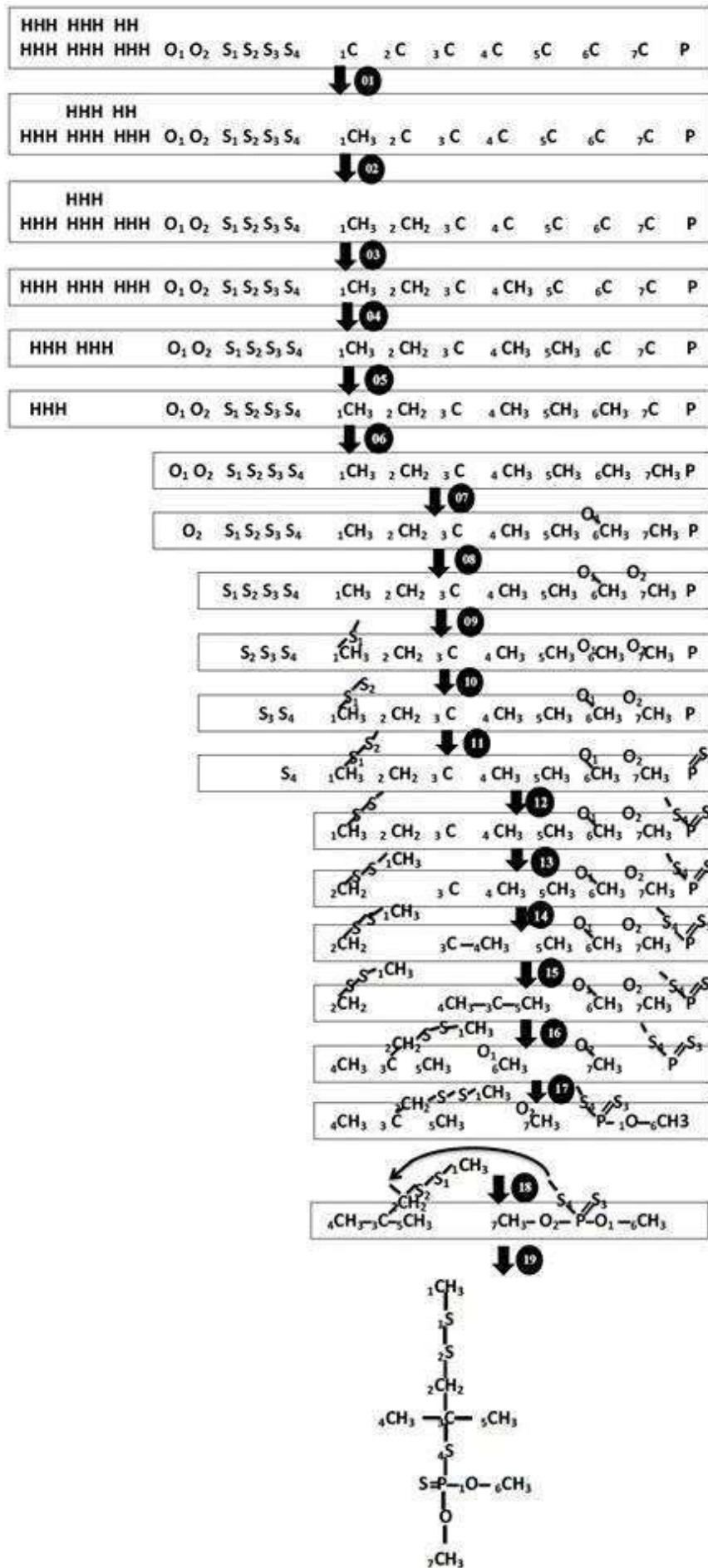
Signature	Nombre d'occurrence	Forme avant la numérotation	Forme après la numérotation
X <sub>6</sub>	2	C(CHHH)	<sub>4</sub> C ( <sub>3</sub> CHHH) et <sub>5</sub> C ( <sub>3</sub> CHHH)
X <sub>10</sub>	1	C(SHHH)	<sub>1</sub> C(S <sub>1</sub> HHH)
X <sub>11</sub>	2	C(OHHH)	<sub>6</sub> C(O <sub>1</sub> HHH) et <sub>7</sub> C(O <sub>2</sub> HHH)
X <sub>49</sub>	1	C(SCHH)	<sub>2</sub> C (S <sub>2</sub> <sub>3</sub> CHH)
X <sub>237</sub>	1	C(CCCS)	<sub>3</sub> C ( <sub>2</sub> C <sub>5</sub> C <sub>4</sub> CS <sub>4</sub> )
X <sub>8</sub>	1	S(PC)	S <sub>4</sub> (P <sub>3</sub> C)
X <sub>9</sub>	2	O(PC)	O <sub>1</sub> (P <sub>6</sub> C) et O <sub>2</sub> (P <sub>7</sub> C)
X <sub>50</sub>	1	P (=SOOS)	P (=S <sub>3</sub> O <sub>1</sub> O <sub>2</sub> S <sub>4</sub> )
X <sub>242</sub>	2	S(CS)	S <sub>1</sub> ( <sub>1</sub> C S <sub>2</sub> ) et S <sub>2</sub> (S <sub>1</sub> <sub>2</sub> C)
X <sub>51</sub>	1	=S(P)	=S <sub>3</sub> (P)

Comme indiqué dans ce tableau, l'algorithme sature en une fois tous les atomes d'une orbite choisie jusqu'à ce qu'il n'y ait plus d'orbitales à saturer. Toutes les étapes de saturation sont indiquées par une flèche avec un numéro (**Fig IV- 3**). A chaque étape de l'algorithme, l'orbite à saturer est toujours celle qui reste la plus à gauche. Le graphique initial est composé de 31 atomes isolés, 17 atomes d'hydrogène, deux atomes d'oxygènes, quatre atomes de soufres, sept atomes de carbone et un atome de phosphore. Les atomes sont partitionnés en fonction de leurs signatures cibles (tableau d'illustration IV- 5).

La première orbite à être saturée, c'est celle qui contient les atomes (H), chacun des 17 atomes de l'orbite doit être lié à un carbone lié à trois hydrogènes au total jusqu'à l'étape 6

sauf :  $_3C$  et  $_2C$  qui est saturé par 2H seulement. La deuxième orbite à être saturée, c'est celle qui contient les deux atomes d'(O) lié aux deux carbones  $_6C$  et  $_7C$  les étapes 7 et 8. La troisième orbite à être saturée qui contient les quatre atomes de soufre ( $S_1 S_2 S_3 S_4$ ), lié au carbone  $_1C$  tel que ( $S_1..._1C$ ), ensuite le soufre  $S_2$ , tel que ( $S_2...S_1$ ), le phosphore P, tel que ( $S_3==P$ ) ainsi le phosphore P, tel que ( $S_4...P$ ) de l'étape 9 jusqu'à l'étape 13.

Les phases de saturation se poursuivent en fonction du nombre de valence spécial attribué à chaque atome, jusqu'à l'étape 18, où tous les atomes sont saturés et complètent toutes les liaisons qui génèrent la nouvelle structure **Figure IV- 3**.






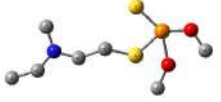


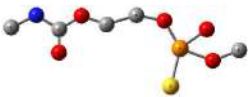

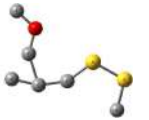

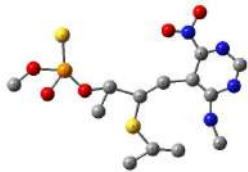

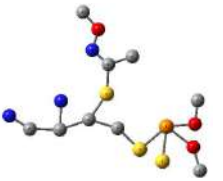


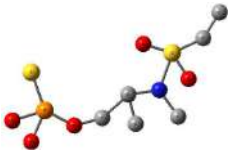
**Figure IV- 3 :** Illustre la traduction de la structure du composé ayant pour formule brute  $C_7H_{17}O_2PS_4$  en utilisant la méthode développée par Faulon et al. [100].

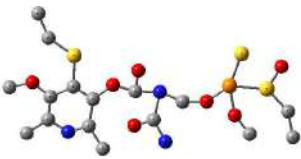
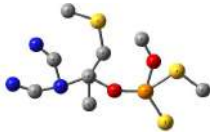


Le tableau suivant, comprend les structures 3D des 20 nouveaux pesticides générés en utilisant la théorie I-QSPR avec leurs valeurs de  $\log k_{ow}$  basées sur cette théorie et celles prédites par le logiciel Hyperchem avec d'autres propriétés QSAR simulées par des calculs par mécanique moléculaire MM<sup>+</sup> et semi-empiriques PM3.







On remarque que les valeurs des lipophiles prédits par le QSPR des nouveaux pesticides générés sont dans la plupart des cas proches de celles simulées par Hyperchem qui est un logiciel très utilisé en mécanique moléculaire pour le calculs de ces entités. Cela vérifie bien l'authenticité de nos résultats basé sur les calculs par la théorie de I-QSPR dédiée à la génération de nouvelles structures moléculaires de pesticides.

**Tableau IV- 6.** Structures 3D des 20 nouveaux pesticides et valeurs des  $\log k_{ow}$  calculées par la théorie I-QSPR et prédites par le logiciel Hyperchem, avec d'autres propriétés QSAR simulées Hyperchem

Formule brute	Structure 3D <sup>(a)</sup>	$\log k_{ow}$ <sup>(b)</sup>	$\log k_{ow}$ <sup>(c)</sup>	Surface area	Volume	Réfractivité	Masse Mol.	énergie MM+ (KJ/mol)	énergie PM3 (KJ/mol)
$C_7H_{17}O_2PS_4$		2.93	2.62	536.71	760.29	76.60	292.42	22.67	-2641.14
$C_6H_{15}O_2PS_3$		2.17	2.61	531.90	730.74	64.50	246.34	24.518	-2302.97
$C_7H_{15}N_2O_4PS_3$		6.05	2.19	580.95	882.26	77.57	318.36	146.48 8	-2905.23
$C_8H_{17}N_2O_4PS_3$		6.74	2.20	577.82	882.76	81.80	332.39	131.31	-3174.98
$C_9H_{17}Cl_2N_2O_3PS_2$		6.30	2.19	564.48	874.47	77.39	353.22	26.47	-2922.12
$C_7H_{18}NO_2PS_2$		1.87	1.36	573.67	483.36	65.30	243.32	54.01	-2703.03

$C_5H_{12}NO_3PS$		2.65	2.61	451.91	657.32	52.07	229.19	21.28	-2288.44
$C_7H_{16}NO_4PS$		4.56	2.20	510.26	729.30	59.29	241.24	19.674	-2737
$C_6H_{14}OS_2$		1.46	1.88	410.80	566.15	47.31	166.30	-0.46	-1997.52
$C_9H_{19}N_2O_4PS_4$		6.00	2.20	651.1	992.59	95.46	378.47	32.8	-3488
$C_{13}H_{23}N_4O_5PS_2$		3.34	2.61	600.25	1074.63	106.29	366.51	44.67	-4527.12
$C_6H_{16}NO_2PS_2$		1.55	2.61	606.55	978.56	81.77	347.45	31.30	3549.116
$C_9H_{22}N_3O_3PS_3$		5.02	2.62	651.7	982.59	95.42	346.48	32.8	-3458
$C_6H_{14}NO_3PS_4$		4.71	2.19	499.29	791.32	73.09	307.40	14.102	2539.87
$C_{10}H_{15}N_2O_4PS_2$		5.2	2.20	488.23	853.74	76.81	322.33	27.922	-3329.07
$C_6H_{16}NO_3PS_2$		4.27	2.61	493.77	731.	63.56	277.23	49.9	-2743.26

$C_{16}H_{26}N_3O_7PS_3$		8.66	1.47	693	1249	115.85	499.5	46.36	-5356.36
$C_8H_{18}N_3O_2PS_3$		2.59	2.61	506.16	858.53	74.30	315.40	43.37	-3075.79
$C_8H_{20}NO_3PS_2$		1.51	1.46	543.54	803.91	70.45	273.35	26.001	-3074.00
$C_5H_{15}N_2O_4PS$		5.52	2.20	492.04	72.69	57.54	242.23	40.15	-26.21

(<sup>a</sup>):  : Carbon (C);  : Chlorine (Cl);  : Nitrogen (N);  : Oxygen (O);  : Phosphorus;  : Sulfur (S); (the H atoms are hidden)  
 (b): logk<sub>ow</sub> predicted by Hyperchem software ; (<sup>c</sup>): logk<sub>ow</sub> calculated by I-QSPR theory

## Conclusion et perspectives

Le grand intérêt de l'utilisation des pesticides dans divers domaines, notamment en agriculture, nous oblige à étudier ces composés de manière largement approfondie. Cela se fait par leur identification et en se concentrant sur la mise en évidence de toutes leurs caractéristiques, y compris les propriétés physico-chimiques, puis en essayant d'établir de nouvelles structures chimiques.

L'identification de nouveaux pesticides avec des caractéristiques bien spécifiques et des propriétés souhaitées, peut être effectué en utilisant des méthodes et des modèles mathématiques basés sur les relations structures propriétés de ces composés.

Dans cette étude, nous avons développé une relation inverse-quantitative structure-propriété (I-QSPR) sur la base du coefficient de partage octanol-eau (**log<sub>k</sub><sub>ow</sub>**). Nous avons traité une base de données de 190 composés de la classe des pesticides, après avoir développé des signatures moléculaires calculées à partir de signatures atomiques de ces entités. Après réalisation de la LOOCV, seules 29 signatures atomiques sur 252 ont été utilisées en tant que variables indépendantes et les valeurs du coefficient de partage en termes de **log<sub>k</sub><sub>ow</sub>** en tant que variable dépendantes.

L'élaboration de la QSPR et la résolution des équations de contraintes, au nombre de 55, ont été réalisés par un code de calcul développé à cet effet, basé sur l'algorithme PSO, qui nous a permis, par excellence, d'aboutir à 5500 solutions, représentant éventuellement les nouvelles structures probables de pesticides non encore connus sur le marché.

La méthode inverse-QSPR a été utilisée pour prédire les nouvelles structures de composés pesticides possédant une valeur de **log<sub>k</sub><sub>ow</sub>** assez proches de celles de la base de données de départ.

On a réussi à mettre en évidence 20 nouvelles structures de composés qu'on peut classer comme nouveaux pesticides.

Ces nouvelles structures qui n'existent pas encore, en tant que pesticides ou en produits chimiques, cela a été vérifié et confirmé en consultant les bases de données de composés chimiques disponibles en littérature, d'après une recherche approfondie dans les data-bases les plus reconnues des produits chimiques.

Nous avons conclu que la méthode inverse-QSPR est très efficace dans la prédiction et la mise en évidence de nouvelles structures et peut être utilisée comme une approche fiable pour



générer de nouveaux composés de différentes classes de familles chimiques. Pour le cas de notre étude, concernant la génération de nouvelles structures de pesticides, les résultats statistiques montrent que les résultats sont satisfaisants ayant de bons coefficients de corrélation. D'une autre part, la comparaison de nos résultats prédits pour certaines propriétés prédites, tel que la lipophilie, sont très proches des valeurs calculées par d'autres logiciels puissant et bien reconnus dans le domaine de simulation en mécanique moléculaire et quantique, tel que Hyprchem.

Comme perspective, on recommande que les nouvelles structures des 20 pesticides, générés, dans cette étude, doivent servir comme base de données à d'autres études, dans le domaine des pesticides. Ces nouveaux composés doivent être pris en considération pour leur synthèse et leur mise en voie de commercialisation dans le marché des pesticides, et cela, après d'autres études complémentaires, notamment en ce qui concerne leur toxicité et leur impact sur l'environnement.

## Nomenclature

- ACD / LogP** : Advanced Chemistry Development/ LogP
- ASCII** : American Standard Code for Information Interchange
- ASTM** : American Society for Testing and Materials
- CAMD** : Computer Aided Molecular Design
- CEE** : Communauté économique européenne
- CLOGP** : Calculated log P
- KOWWIN** : Octanol-Water Partition Coefficient *Program* for Microsoft Windows[modèle]
- QSAR** : Relation quantitative structure-activité
- OCDE** : Organisation de coopération et de développement économiques
- Da** : Dalton = Un dalton est défini comme égal à **1/12** de la masse d'un atome de carbone 12
- DET** : Distribution des types de bord
- DVT** : Distribution des types de sommet
- Eq. Diophantienne** : équation polynomiale dans laquelle seules les solutions entières sont autorisées.
- EPA** : Environmental Protection Agency's
- FAO** : Organisation des Nations Unies pour l'alimentation et l'agriculture
- FIFRA** : Federal Insecticide, Fungicide, and Rodenticide
- HPLC** : La chromatographie en phase liquide à haute performance
- ICAM-1** : Molécule d'adhésion intercellulaire (InterCellular Adhesion Molecule)
- I-QSPR** : Inverse Quantitative Structure-Property Relationships
- LOO-CV** : Validation croisée Laisser un sort ( Leave-one-out cross-validation)
- MLR** : Régression linéaire multiple ou modèle gaussien

**OMS** : Organisation mondiale de la santé

**OMZ** : Omeprazole

**OTS** : Office of Toxic Substances

**PBPK** : La pharmacocinétique basée sur la physiologie

**PSO** : Optimisation des essaims de particules ( Particle Swarm Optimization)

**RMSE** : Erreur quadratique moyenne

**SD** : Descripteurs de signature

**SGH** : Système général harmonisé

**SVM** : Support Vector Machine Regression

**SVR** : Support Vector Regression

**TBD** : Une base azotée, le 1,5,7-triazabicyclo[4.4.0]déc-5-ène.

**TI** : Topological Indices

**UE** : L'Union européenne

*Références bibliographiques*

- 01 **Hansch C., Leo A.**, Substituent Constants for Correlation Analysis in Chemistry and Biology. Wiley: New York, **1979**.
- 02 **Tewari Y. B., Miller M. M., Wasik S. P., Martire, D. E.**, J. Chem. Eng. Data ,**1982**, 27:451–454.
- 03 **Papa E., Dearden J. C., Gramatica P.**, Chemosphere **2007**, 67, 351.
- 04 **Gestel V., Oterman, K., Canton J. H.**, Regulatory Toxicology and Pharmacology ,**1985**, 5,422.
- 05 **Mackay D. , Shiu W. , Ma K. ,** Illustrated Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemicals; Lewis Publishers: MI, **1993**.
- 06 **Lyman W. J., Reehl W. F. , Rosenblatt D. H.**, Handbook of Chemical Property Estimation Methods; American Chemical Society: Washington, DC, **1990**.
- 07 **Leo A., Hansch C., Elkins D.**, Partition Coefficients and their Uses. Chem. Rev. **1971**, 71: 525-616.
- 08 **Meylan W., Howard P. H.**, Estimating log P with Atom/Fragments and Water Solubility with log P. Perspectives Drug Discovery Design **2000**, 19, 67.
- 09 **Verschueren K.**, Handbook of Environmental Data on Organic Chemicals, 2nd ed.; Van Nostrand Reinhold: New York, **1983**.
- 10 **de Bruijn J., Hermens J.**, Environ. Toxicol. Chem. **1991**, 10: 791 – 804.
- 11 **Lombardo A. , Roncaglioni A., Boriani E., Milan C., Benfenati E.**, Assessment and validation of the CAESAR predictive model for bioconcentration factor (BCF) in fish. Chem. Cent. J, **2010**:4(1): S1.
- 12 **Sabljić A., Guesten H., Hermens J., Opperhuizen A.**, Modeling octanol/water partition coefficients by molecular topology: chlorinated benzenes and biphenyls. Environ. Sci. Technol, **1993**, 27, 1394–1402.
- 13 **Oszwaldowski S., Timerbaev A. R.**, Development of quantitative structure–activity

- relationships for interpretation of the migration behavior of neutral platinum(II) complexes in microemulsion electrokinetic chromatography. *J. Chromatogr, A*, **2007**, 1146: 258–263.
- 14 **Klopman G., Li J.-Y., Wang S.**, *J Chem Inf Comput Sci*, **2002**, 42: 109-142.
- 15 REACH - Environment - European Commission, accessed Apr 21, 2013.  
[http://ec.europa.eu/environment/chemicals/reach/reach\\_intro.htm](http://ec.europa.eu/environment/chemicals/reach/reach_intro.htm)
- 16 **European Chemicals Agency Guidance on information requirements and chemical safety assessment**. Chapter R.7a: Endpoint specific guidance; European Chemicals Agency: Annankatu 18, Helsinki, Finland, **2012**.
- 17 **Capry.A**, Importance de la lipophilie en modélisation moléculaire, LPTC, UFR de chimie, **1999**, Université de Bordeaux I, France.
- 18 **Yu-PING CHIN., WALTER J., Weber Jr., VOICE T . C.**, Determination of Partition Coefficients and Aqueous Solubilities By Reverse Phase Chromatography—II. *Wat. Res.* Vol. 20, No. 11, pp, **1986**, 1443-1450.
- 19 **Goudarzi N ., Goodarzi M .,** QSPR study of partition coefficient ( $K_{o/w}$ ) of some organic compounds using radial basic function-partial least square (RBF-PLS) . *J. Braz. Chem. Soc.* vol.21 no.9 São Paulo **2010**.
- 20 **Bayat Z., Emadiyan M.**, Evaluation of the 1-octanol/water partition coefficient of derivatives of the anti-HIV drug in quantum chemical calculations . *Der Chemica Sinica*, **2011**, 2(6):341-350.
- 21 **Heidari A.**, Cancer Risk Prediction and Assessment in Human Cells under Synchrotron Radiations Using Quantitative Structure Activity Relationship (QSAR) and Quantitative Structure Properties Relationship (QSPR) Studies . *International Journal of Clinical & Medical Imaging* ,**2016**.
- 22 **Milanetti E ., Raimondo D ., Tramontano A.**, Prediction of the permeability of neutral drugs inferred from their solvation properties. **2016**, 32(8): 1163–1169.
- 23 **Plante J., Werner S., JPlogP: an improved logP predictor trained using predicted data** .*Journal of Cheminformatics* December 2018, 10:61.
- 24 **Petrauskas A. A., Kolovanov E. A.**, ACD/LogP Method Description, *Persp. Drug DiscoV. Design*, **2000**, 19: 1-19.
- 25 **Klopman G., Iroff L.**, Calculation of Partition Coefficients by the Charge Density Method. *J Comput. Chem*, **1981**, 2: 157 –160.
- 26 **Buchwald P., Bodor N.**, Octanol-Water Partitioning: Searching for Predictive Models.

- Current Medicinal Chemistry, **1998**, 5: 353-380.
- 27 **Broto P., Moreau G., Vandycke C.**, Molecular Structures Perception, Autocorrelation Descriptor and SAT Studies; System of Atomic Contributions for the Calculation of Octanol-Water Partition Coefficient. Eur, Journal of Medicinal Chemistry, **1984**, 19:71-78.
- 28 **Wang R., Fu Y., Lai L.**, A New Atom-Additive Method for Calculating Partition Coefficients. Journal of Chemical Information and Computer Sciences, **1997**, 37: 615-621.
- 29 **Katritzky A. R., Maran U., Lobanov V. S., Karelson M.**, Structurally Diverse Quantitative Structure-Property Relationship Correlations of Technologically Relevant Physical Properties. Journal of Chemical Information and Computer Sciences, **2000**, 40: 1-18.
- 30 **Lamoree Bas ., Hubbard R. E.**, Current perspectives in fragment-based lead discovery (FBLD). Essays Biochem. **2017** , 61(5): 453-464.
- 31 **Mannhold R., Dross K.**, Calculation Procedures for Molecular Lipophilicity: A Comparative Study. Quantitative structure-activity relationship, **1996**, 15: 403-409.
- 32 **Erdemir A., Shenghua Li ., Jin Y. S.**, Relation of Certain Quantum Chemical Parameters to Lubrication Behavior of Solid Oxides. Int. J. Mol. Sci. **2005**, 6, 203-218
- 33 **Kier L. B., Hall, L. H.**, An Electrotopological State Index for Atoms in Molecules. Pharmaceutical Research, **1990**, 7:801-807 .
- 34 **Hall L. H., Kier L. B.**, Electrotopological State Indices for Atom Types: A Novel Combination of Electronic, Topological, and Valence State Information. Journal of chemical information and computer science, **1995**, 35: 1039-1045.
- 35 A continuously growing list of these publications is available from the home page of MolconnZsoftware at <http://www.eslc.vabiotech.com/molconn/mconpubs.html>.
- 36 **Kier L. B., Hall L. H.**, Molecular Structure Description: The Electrotopological State, Academic Press: London, **1999**.
- 37 **Hemi C., and Christoph R.**, Octanol-Water Partition Coefficient Measurement by a Simple  $^1\text{H}$  NMR Method DOI: 10.1021/acsomega.7b01102 ACS Omega **2017**, 2: 6244-6249.
- 38 **Hiroshi Y.**, Properties Estimation: logP, logKow: Octanol-water partition coefficient: Lecture note of for Yokohama National University (YNU) **2011**.
- 39 **DeBruijn et al, 1989, Tolls et Sijm, 1993.**, Determination of octanol/water partition coefficients for hydrophobic organic chemicals with the "slow-stirring" method. Environmental Toxicology and Chemistry , 8:499-512.

- 40 **Ligne directrice 107 de l'OCDE, 1995.**
- 41 Partition Coefficient From Wikipedia, the free encyclopedia [edit] Applications.
- 42 **Flor S., Tripodi V., Scioscia S., Revello L., Lucangioli S.,** Fast and Sensitive New Hplc-Uv Method For Determination Of Omeperazole and Major Related Substances in Pharmaceutical Formulation .. Journal of Liquid Chromatography Related Technologies **2010** 33:1666–1678.
- 43 **Petrova O. E ., Sauer K.,** High-Performance Liquid Chromatography (HPLC)-Based Detection and Quantitation of Cellular c-di-GMP. Methods Mol Biol. **2017**,1657: 33–43.
- 44 **Ligne directrice 117 de l'OCDE, 1989.**
- 45 **Chemicool Dectinary:** Definition of High-Performance Liquid Chromatography (HPLC), [https://www.chemicool.com/definition/high\\_performance\\_liquid\\_chromatography\\_hplc.html](https://www.chemicool.com/definition/high_performance_liquid_chromatography_hplc.html).
- 46 **Bayles J. P.;** 400 manipulations commentées de chimie des solutions, vol. 1, ed Ellipses, 2011, pp 352.
- 47 **Lanez T., Hadj Seyd A .,** Simulation du rendement dans une extraction liquide-liquide principe et application .Herstellung :Siehe letzete Seite : 978-613-1-59059-7 , **2013.**
- 48 **RAHMANI Z., SAIDI M., YOUSFI M. and DAKMOUCHE M.,** Experimental and Theoretical Study on Lipophilicity of Novel 1,2-Dithiole-3-thiones Synthetic .Asian Journal of Chemistry; Vol. 25, No. 15,**2013**
- 49 **Lanez. T,** Polycopier de cours, module : Séparation et purification en chimie organique ,PG-Ouargla **2005-2006.**
- 50 **Goulon A., Picot T., Duprat, A. et Dreyfus, G.** Predicting activities without computing descriptors: graph machines for QSAR. SAR and QSAR in Environmental Research, **2007**, 18 (1 & 2), p. 141-153.
- 51 Environmental Organic Chemistry, Schwarzenbach et al., **1999**
- 52 **Danielssso L.G., Zhang Y.H .,** Methods for determining n-octanol-water partition constants. Tr.A.C 1996 ,15: 188-196.
- 53 **Deng Y., Besse-Hoggan P., Sancelme M., Delort A. M., Husson P.et al.,** Influence of oxygen functionalities on the environmental impact of imidazolium based ionic liquids. J Hazard Mater ,**2011**, 198: 165-174.
- 54 **Wang Y., Wang Xu , Li M., Dong J., Sun C., and Chen G.,** Removal of Pharmaceutical and Personal Care Products (PPCPs) from Municipal Waste Water with Integrated Membrane Systems, MBR-RO/NF. Int J Environ Res Public Health. **2018** , 15(2): 269.

- 55 **Hansch C., Leo A. and Hoekman D.**, Exploring QSAR: Hydrophobic, Electronic, Steric Constants, ACS, Washington, DC, USA, **1995**.
- 56 **Bundy J. G., Morriss A. W. J., Durham D. G., Campbell C. D. and Paton G. I.**, (and the literature cited there in), *Chemosphere*, **2001**, 42: 885-892.
- 57 **Korzekwa K., Nagar S.**, Drug Distribution Part 2. Predicting Volume of Distribution from Plasma Protein Binding and Membrane Partitioning. *Pharm Res.* **2017**, 34(3): 544–551.
- 58 **Fernando S. and Attilio A.M. Del Re.**, " Relationships of Pesticide Octanol/Water Partition Coefficients to Their Physicochemical Properties *Reviews of Environmental Contamination and Toxicology*, **1993**, Vol. 133 Pages 59-93.
- 59 **Pussemier L., Szabo G., Bulman R. A.**, Prediction of the soil adsorption coefficient Koc for aromatic pollutants. *Chemosphere*, **1990**, 21: 1199–1212.
- 60 **Ellgehausen H., Guth J. A., Esser H. O.**, International congress on pesticide chemistry, 4th ed, v32, IUP AC, Zurich, **1978**.
- 61 **Green G., Karichoff S. W.**, Pesticide in the soil environment: Processes impacts and modelling. In: Cheng H (ed), SSSA book series, vol 2. Madison, WI, **1990**, pp 431–432.
- 62 **Kenaga E.E., Goring C.A.I.**, Relationship between water solubility, soil sorption, octanol-water partitioning, and concentration of chemicals in Biota. In: Eaton JG, Parrish PR, Hendricks AC (eds) *Aquatic toxicology*. ASTM, STP 707, Philadelphia, PA, **1980**, pp 78–115.
- 63 **Isnard P., Lambert S.**, Estimating bioconcentration factors from octanol-water partition coefficient and aqueous solubility. *Chemosphere*, **1988**, 17: 21–34.
- 64 **de Wolf W., de Bruijn J.H.M., Seinen W., Hermens J.L.M.**, Influence of biotransformation on the relationship between bioconcentration factors, and octanol-water partition coefficients. *Environ Sei Technol*, **1992**, 26: 1197–1201.
- 65 **Poling S.M., Hsu W.J., Yohoyama H.**, Structure-activity relationships of chemical inducers of carotenoid biosynthesis. *Phytochemistry*, **1975**, 14: 1933–1938
- 66 **Glass A.D.M.**, Inhibition of phosphate uptake in barley roots by hydroxy-benzoic acids. *Phytochemistry*, **1975**, 14: 2127–2130.
- 67 **Briggs G.G., Bromilow R.H., Evans A.A.**, Relationship between lipophilicity and root uptake and translocation of non-ionized chemical by barley. *Pestic Sci*, **1982**, 13: 495–504. and 1-Octanol. *J. Chem. Eng. Data*, 2002, 47: 941-943.
- 68 **Shone M.G., Wood A.V.**, A comparison of the uptake and translocation of some organic



- herbicides and a systemic fungicide by barley I. Absorption in relation to physico-chemical properties. *J Exp Bot*, **1974**, 25: 390–400.
- 69 **Shone M.G., Bartlett B.O., Wood A.V.**, A comparison of the uptake and translocation of some organic herbicides and a systemic fungicide by barley II. Relationship between uptake by roots and translocation to shoots. *J Exp Bot*. **1974**, 25: 401–409.
- 70 **Trapp S.**, Modelling uptake into roots and subsequent translocation of neutral and ionisable organic compounds. *Pest Manag Sci*, **2000**, 56: 767–778.
- 71 **Nicolopoulou-Stamati P., Maipas S., Kotampasi C., Stamatis P., Hens L.**, Chemical Pesticides and Human Health: The Urgent Need for a New Concept in Agriculture *Front Public Health*. **2016**, 4: 148.
- 72 **Alexandratos N., Bruinsma J.**, Global Perspective Studies Team FAO Agricultural Development Economics Division, World agriculture towards 2030/2050: the **2012** revision.
- 73 **Zhang M., Zeiss M.R., Geng S.**, Agricultural pesticide use and food safety: California's model. *Journal of Integrative Agriculture*, **2015**, 14(11):2340-57.
- 74 **Bonner M.R., Alavanja M.C.R.**, Pesticides, human health, and food security. *Food and Energy Security*, **2017**, 6(3):89-93.
- 75 **Rehwagen C.**, WHO recommends DDT to control malaria. *BMJ*, **2006**, 333(7569):622.3.
- 76 **Franklin L.U., Cunnington G.D., Young D.E.**, Terpene based pesticide treatments for killing terrestrial arthropods including, amongst others, lice, lice eggs, mites and ants. *Google Patents*, **2000**.
- 77 **Dashtbozorgi Z., Golmohammadi H., Konoz E.**, Support vector regression based QSPR for the prediction of retention time of pesticide residues in gas chromatography–mass spectroscopy. *Microchemical Journal*, **2013**, 106:51-60.
- 78 **Aktar M. W., Sengupta D., Chowdhury A.**, Impact of pesticides use in agriculture: their benefits and hazards. *Interdisc Toxicol*. **2009**, 2(1):1-12
- 79 **Proshad R., Kormoker T., Islam S . Md., Haque A. Md., Rahman M. Md., Mithu R . M. Md.**, Toxic effects of plastic on human health and environment: A consequences of health risk assessment in Bangladesh. *International. Journal of Health* , **2018**, 6 (1) :1-5
- 80 **Afful S., Anim A. K. and Serfor-Armah Y.**, Spectrum of organochlorine pesticide residues in fish samples from Densu Basin. *Res. J. Environ. Earth Sci*, **2010**, 2 (3): 133–138.
- 81 **ZHAO Z., ZHANG L., WU J. and FAN C.**, Residual levels, tissue distribution and risk assessment of organochlorine pesticides (OCPs) in edible fishes from Taihu Lake, China. *Environ. Monit. Assess*, **2013**, 185 :9265–9277.

- 82 **ZHAO Z., ZHANG L., WU J. and FAN C.,** Distribution and bioaccumulation of organochlorine pesticides in surface sediments and benthic organisms from Taihu Lake, China. *Chemosphere* ,**2009**, 77: 1191–1198.
- 83 **Algimantas M., Ramunas M., Gedimanas V. A.,** Agriculture use of pesticides/ plant protection products, Forestry operations and OSH issues OSHWiki. [Cited on **2018** April 24]. Available from <https://oshwiki.eu/wiki/Agriculture>.
- 84 **Yadav I.C., Devi N.L.,** pesticides classification and its impact on Human and environment environ.sci &engg, *Toxicology* , **2017** ,6:p.140-158.
- Pesticide Management division, NIPHM, syllabus,** Pesticide classification on use, chemical nature, formulation toxicity and action etc, Hyderabad p.1-17, [Cited on **2018** May 5] available from <http://www.niphm.gov.in./ASO.PMD.pdf>.
- 85 **Gordeeva E. V., Molchanova M.S., Zefirov N.S.,** General Methodology and Computer Program for the Exhaustive Restoring of Chemical Structures by Molecular Connectivity Indexes. Solution of the Inverse Problem in QSAR/QSPR. *Tetrahedron Computer Methodology*, **1990**, 3 (6): 389-415.
- 86 **Weis D.C., Visco D.P.,** Brown M., Martin S., Inverse-QSAR for Inhibitors of Phosphate Cdc25B .Published 2006.
- 87 **Wong W. W., Burkowski F.J.,** A constructive approach for discovering new drug leads: Using a kernel methodology for the inverse-QSAR problem. *Journal of Cheminformatics* ,**2009**, 1:4 .
- 88 **Miyao T ., Funatsu K., Bajorath J.** Exploring differential evolution for inverse QSAR analysis **2017**; 6: Chem Inf Sci-1285 .
- 89 **Kevin G. J.,** Computer Aided Molecular Design (CAMD): Designing Better Chemical Products. Molecular Knowledge Systems, Inc. Computerizing Group Contribution Techniques: Issues and Opportunities Copyright © 1998-2019 Molecular Knowledge Systems, Inc
- 90 **Joback K.G., Reid R.C.,** *Chem. Eng. Comm.* **1987**,57 :233–243.
- 91 **Lydersen A.L.,** Estimation of critical properties of organic compounds, College Engineering University Wisconsin, Engineering Experimental Station Report 3, Madison, WI, April, **1955**.
- 92 **Ambrose D.,** Correlation and estimation of vapor–liquid critical properties. I. Critical temperatures of organic compounds, National Physical Laboratory, Teddington, UK, NPL Report Chem., 92, September **1978**

- 94 **Klincewicz K.M., Reid R.C.**, *AIChE J.* **1984**, 30:137–142.
- 95 **Lyman W.J., Reehl W.F., Rosenblatt D.H.**, *Handbook of Chemical Property Estimation Methods*, American Chemical Society, Washington, DC, **1990**.
- 96 **Horvath A.L.**, *Molecular Design*, Elsevier, Amsterdam, **1992**.
- 97 **Brown W.M., Martin S., Rintoul M.D., Faulon J.-L.**, Designing novel polymers with targeted properties using the signature molecular descriptor. *J Chem Inf Model*, **2006**, 46(2):826-35.
- 98 **Visco D. P., Pophale R. S., Rintoul M. D., and Faulon J.-L.**, Developing a methodology for an inverse quantitative structure-activity relationship using the signature molecular descriptor. *Journal of Molecular Graphics and Modelling*, **2002**,20:429–438.
- 99 **Faulon J.-L., Visco D. P. and Pophale R. S.**,The signature molecular descriptor. 1. Using extended valence sequences in QSAR and QSPR studies. *Journal of Chemical Information and Computer Sciences*, **2003**,43:707– 720.
- 100 **Faulon J.-L., Churchwell C.J., Visco D. P.**, The signature molecular descriptor. 2. Enumerating molecules from their extended valence sequences. *J Chem Inf Comput Sci*, **2003**,43(3):721-34.
- 101 **Hamed M. K.**, *Computer-Aided Molecular Design Using The Signature Molecular Descriptors : An Application to Design Novel Chemical Admixtures for Concrete*. Thesis August, **2014**.
- 102 **Churchwell C., Rintoul M.D., Martin S., Visco D.P., Kotu A., Larson R.S., Sillerud L.O., Brown D.C., and Faulon J.L.**, "The signature molecular descriptor 3. Inverse-quantitative structure-activity relationship of ICAM-1 inhibitory peptides", *Journal of Molecular Graphics and Modelling*, **2004**: 22, 263-273.
- 103 **Faulon J. L., Visco D. P. Jr., et Pophale R. S.**, The Signature Molecular Descriptor. 1. Extended Valence Sequences vs. Topological Indices in QSAR and QSPR studies, *J. Chem. Inf. Comput. Sci*, **2003**,43: 707 – 720.
- 104 **Fayet G., Rotureau P.**, QSPR Model for Regulatory Purpose: from Development to Integration into the QSAR Toolbox. *CHEMICAL ENGINEERING TRANSACTIONS VOL. 48*, **2016**.
- 105 **Hansch C., Muir R .M., Fujita T., Maloney P. P., Geiger F., Streich M. J.**, *Am. Chem. Soc*, **1963**, 85:2817-2825.

- 106 **Bodor N.**, Current Medicinal Chemistry, 1988, 5, 353-380. From book: Biochemistry of Redox Reactions, by Bernard Testa, editor: London [u, a], Acad. Press, **1995**.
- 107 **Dudek A.Z., Arodz T., Galvez J.**, Computational Methods in Developing Quantitative Structure-Activity Relationships (QSAR): A Review. Comb Chem High Throughput Screen, **2006**,9(3):213-228.
- 108 **Shamsara J.**, Ezqsar: An R Package for Developing QSAR Models Directly From Structures . Published online ,**2017** .
- 109 **Darnag R., Minaoui B., Fakir Md.**, QSAR models for prediction study of HIV protease inhibitors using support vector machines, neural networks and multiple linear regression. Arabian Journal of Chemistry . **2017**,10 ( 1 ) : 600-608.
- 110 **Le T., Epa V.C., Burden F.R., Winkler D. A.**, Quantitative structure-property relationship modeling of diverse materials properties. Chem Rev, **2012**;,112(5):2889-919.
- 111 **Mishra R., Mishra B.**, Drug Discovery Tools : Computer Aided Molecular Design and Chemo-Informatics. Chemoinformatics and drug delivery. Journal of Pharmacology and Biomedicine. **2017**, 1(4): 103-114.
- 112 **Brown ,W.M., Martin, S., Rintoul ,M.D., Faulon J.L.**, "Designing Novel Polymers with Targeted Properties Using the Signature Molecular Descriptor". J. Chem. Inf. Model, **2006**,46:826-835
- 113 **Churchwell C. , Rintoul M. D., Martin S. , Visco J. D. P. , Kotu, A., Larson R. S., Sillerud L. O., Brown D. C., Faulon, J.-L.**, The signature molecular descriptor. 3. Inverse quantitative structure relationship of ICAM-1 inhibitory peptides. J. Molecular Graphics and Modeling, **2003**, 22, 263 – 273
- 114 **Kennedy J., Eberhart R.**, Particle Swarm Optimization . modeled by. 0-7803-2768-3/**95**/\$4.00 0 **1995** IEEE .
- 115 **Miyao T., Kaneko H., Funatsu K.**, J Chem Inf Model, **2016**, 56, 286–299.
- 116 **Faulon J. L., Churchwell C. J., Visco D. P. Jr**, J. Chem. Inf. Sci,**2003**, 43, 721–734.
- 117 **Weis D.C., Visco D.P.**, Computer-aided molecular design using the Signature molecular descriptor: Application to solvent selection. Comput Chem Eng,**2010**,34(7):1018-1029.
- 118 **Weis D.C., Faulon J. L., LeBorne R.C., Visco D. P.**, The Signature Molecular

- Descriptor. 5. The Design of Hydrofluoroether Foam Blowing Agents Using Inverse-QSAR. *Ind Eng Chem Res*, **2005**, 44 (23):8883-8891.
- 119 **Jackson J.D., Weis D.C., Visco D.P.**, Potential glucocorticoid receptor ligands with pulmonary selectivity using I-QSAR with the signature molecular descriptor. *Chem Biol Drug Des*, **2008**, 72(6):540-50
- 120 **Saeed A. M.**, Role of Database Management Systems (DBMS) in Supporting Information Technology in Sector of Education. *International Journal of Science and Research (IJSR)* Volume 6 Issue 5, May **2017**.
- 121 **Suseela V.J.**, Unicode Applications in the Digital Libraries of India. January **2013**  
DOI: 10.5958/j.0973-7081.7.1.006
- 122 **Kristian B. K.**, Master Thesis: Separation of secondary dispersed diesel-water emulsions by electrospun membranes identified by inverse-quantitative structure-activity relationship. Issues: 5 Departments of Biotechnology, Chemistry and Environmental Engineering, Aalborg University, 9000 Aalborg, Denmark, **May 2012**
- 123 **Vahdani S. and Bayat Z.**, Computational approaches to the predication of the octanol-water partition coefficient (LogPo/w). *J. Chem. Pharm. Res*, **2011**, 3(4): 565-575.
- 124 **Draper N .R., Smith H.**, *Applied Regression Analysis*. 3rd ed. New York: Wiley, **1998**.
- 125 **Achenie L.E.K., Gani R., Venkatasubramanian V.**, Challenges and opportunities for CAMD. In: *Computer Aided Molecular Design: Theory and Practice*. Elsevier Ltd, **2002**:357-377.
- 126 **Hawkins D.M.**, The problem of overfitting. *J Chem Inf Comput Sci*, **2004**, 44(1):1-12.
- 127 **Fengping L., Chenzhong C., Bin C.**, A Quantitative Structure-Property Relationship (QSPR) Study of Aliphatic Alcohols by the Method of Dividing the Molecular Structure into Substructure *Int. J. Mol. Sci*, **2011**, 12, 2448-2462; doi: 10.3390/ijms12042448.
- 128 **Bürkner P.C ., Gabry J. , Vehtari A.**, Bayesian leave-one-out cross-validation for non-factorizable normal models\*(Submitted on 24 Oct 2018 (v1), last revised 14 Mar **2019** (this version, v2).
- 129 **Sales J.**, The use of linear regression to predict digestible protein and available amino acid contents of feed ingredients and diets for fish. **2008**, 278(1–

- 4444):128–142
- 130 **Yamanaka N., Okamoto E., Kuwata K., Tanaka N.,** A multiple regression equation for prediction of posthepatectomy liver failure. *Ann Surg.* **1984**, 200(5):658–63.
- 131 **Mbarki S., Dguigui K., El Hallaoui M.,** Construction of 3D-QSAR models to predict antiamebic activities of pyrazoline and dioxazoles derivatives. *J. Mater. Environ. Sci.* **2011**, 2 (1)61-70.
- 132 **Deeb O.,** *Chemom. Intell. Lab. Syst.*, **2010**, 104:181.
- 133 **Pran Kishore D., Balakumar C., Raghuram Rao A., Partha Pratim R., Roy K.** *Bioorg. Med. Chem. Lett.*, **2011**, 21:818 -823.
- 134 **Ghasemi J., Piravi-vanak Z., Khavarian F. A.D.,** Quantitative Structure Property Relationship Modeling for Prediction of Retention index for a set of Essential Oils and Organic Substances of Plant and Animal Origin. *J. Sci. I. A. U (JSIAU)*, Vol 18, No. 68, summer **2008**.
- 135 **Hemmateenejad, B., Javadnia K. and Elyasi M.,** *Anal. Chem. Acta.*, **592**, 72, 2007.
- 136 **Leach A.R.,** *Molecular modeling: Principles and applications.* Pearson Education Ltd. Harlow, England, **2001**.
- 137 **YANG S. J.H ., LU O. H.T., HUANG A. Y.Q., HUANG J. C.H., OGATA H., LIN A. J.Q.,** Predicting Students Academic Performance Using Multiple Linear Regression and Principal Component Analysis . *Journal of information processing* **2018**, 25, 170- 176.
- 138 **Mehrdad S., Sharmin E.,** Quantitative structure activity relationship (QSAR) study of octanol-water partition coefficients of some of environmental toxic of petroleum substances. *Asian Journal of Green Chemistry* 1, **2017**, 116-129
- 139 **Veerasamy R., Rajak H., Jain A., Sivadasan S., Varghese C. P., Agrawal R.K.,** *International Journal of Drug Design and Discovery* ,**2011**, 2, 511–519.
- 140 **Akbaria S., Zebardasta T., Zarghib A. and Hajimahdi Z.,** QSAR Modeling of COX -2 Inhibitory Activity of Some Dihydropyridine and Hydroquinoline Derivatives Using Multiple Linear Regression (MLR) Method. *Iranian Journal of Pharmaceutical Research* .**2017**, 16 (2): 525-532
- 141 **Golbraikh A., Tropsha A.,** *J. Mol. Graphics Mod*, 2002, 20,269–276.
- 142 **Faulon J. L.,** *J. Chem. Info. Comput. Sci.* **1994**, 34, 1204–1218.

- 143 **Benfenati E. , Gini G., Piclin N., Roncaglioni A., Var M.R.,** Predicting logP of pesticides using different softwareChemosphere 53 (2003) 1155–1164





# SOMMAIRE

Liste des figures	I
Liste des tableaux	II
Liste des équations	IV
Introduction générale	1

## Chapitre I: Coefficient de partage (importance et applications)

I-1 Introduction	4
I-2 Définition du phénomène de partage	4
I-3 Coefficient de partage n-octanol/eau	5
I-4 Relation de $\log K_{ow}$ avec d'autres propriétés physico-chimiques	6
I-5 Détermination de $K_{ow}$	8
I-5-1 Méthodes expérimentales	8
I-5-1-1 Méthode par agitation du flacon	9
I-5-1-2 Méthode par HPLC	10
I-5-1-3 Méthode par agitation douce	10
I-5-1-4 Méthode sur colonne	12
I-5-2 Prédiction de $\log k_{ow}$	13
I-5-2-1 Méthode de substitution de HUNSCH	13
I-4-2-2 Méthode fragmentale de REKKER	14
I-6 Paramètres influençant sur le coefficient de partage	17
I-7 Application de $\log K_{ow}$ dans les domaines Médecine, Pharmacie, environnement et Autres	19
I-8 Importance d'application du $\log k_{ow}$ dans le domaine environnemental	20
I-9 Les pesticides	21
I-9 -1 Classification des pesticides sur la base de la toxicité	22
I-9 -2 Classification des pesticides sur la base de leur utilisation	22
I-9 -3 Classification des pesticides sur la base de la composition chimique	23
I-9 -4 Classification des pesticides sur la base du mode d'entrée	24
I-9 -5 - Classification des pesticides sur la base du mode d'action	25
I-9-6 Les pesticides peuvent également être classés en fonction de comment ou quand ils fonctionnent	25
I-10 Aperçu historique sur l'utilisation des méthodes I- QSAR/QSPR	27
I-11 Conclusion	32

## Chapitre II : Conception moléculaire assistée par ordinateur (CAMD) et leur utilisation dans les équations inverse-QSPR

II-1 Introduction	33
II-2 La conception de signature	34
II-3 Définition de la signature atomique	34

II – 4 Définition des descripteurs de signature moléculaire	36
II – 5 Modèles QSPR	37
II-5-1 Régression linéaire multiple : RLM	39
II- 6 Principales caractéristiques de la signature	40
II- 7 Algorithme de conception inverse	41
II- 8 Applications de la méthodologie I-QSPR	46

### **Chapitre III : Méthodologie I-QSPR utilisant la signature**

III.1 Introduction	47
III-2 Sélection des composés de la base de données.	47
III -3 Dessin des structures 2D.	47
III-4 Traduction des structures 2D en signatures atomiques	48
III -5 Elaboration de la matrice des descripteurs	51
III -6 Génération des équations de contraintes	52
III -6 -1 Équation de graphicalité	53
III -6 -2 Équations de cohérence	54
III -7 Résolution du problème inverse	55
III-8 Modèles QSPR et enregistrement des solutions inverses	57
III -9 Génération des nouvelles structures	58
III -9-1. Développement de modèle QSPR requis	58
III-9-2 Validation du modèle et paramètres statistiques	59

### **Chapitre IV : Résultats et discussion**

IV- 1 Calcul des signatures atomiques	62
IV- 2 Equation modèle	65
IV-3 Equations de contrainte	66
IV-4 Génération de nouvelles structures de pesticides	70
IV-5 Explication de la génération de structure	71
Conclusion et perspectives	78
Bibliographie	80
Annexe 1	92
Annexe 2	108
Annexe 3	115

## Annexe 1: Résultats de calcul des Signatures moléculaires relatives aux pesticides de la base de données

**Acephate** = N( H P C)+H(N)+P( N =O S O)+C( N =O C)+=O(C)+C( C H H H)+=O(P)+S( P C)+O( P C)+C( S H H H)+C( O H H H)+9 H(C)

**Alachlor** = N( pC C C)+pC( N pC pC)+C( N H H O)+C( N =O C)+17 H(C)+O( C C C)+=O(C)+C( C H H Cl)+2 pC( pC pC C)+C( O H H H)+Cl(C)+3 pC( pC pC H)+3 H(pC)+2 C( pC H H C)+2 C( C H H H)

**Ametryn** = 2 pC( pN pN N)+3 pN( pC pC)+2 N( pC H C)+2 H(N)+C( N C H H)+3 C( C H H H)+15 H(C)+pC( pN pN S)+S( pC C)+C( S H H H)+C( N H C C)

**Amidosulfuron**= C( N N =O)+N( C S H)+N( C H pC)+=O(C)+2 H(N)+pC( N pN pN)+2 pN( pC pC)+2 pC( pN pC O)+pC( pC pC H)+2 O( pC C)+H(pC)+S( N =O =O N)+4 =O(S)+N( S C S)+C( N H H H)+S( N =O =O C)+C( S H H H)+2 C( O H H H)+12 H(C)

**Amitraz.** = N( C C C)+2 C( N =N H)+C( N H H H)+17 H(C)+2 =N( C pC)+2 pC( =N pC pC)+4 pC( pC pC C)+6 pC( pC pC H)+6 H(pC)+4 C( pC H H H)

**Amitrole**= pC( pN pN N)+2 pN( pC pN)+pN( pC pC)+N( pC H H)+pC( pN pN H)+ +H(pC)+2 H(N)

**Anilofos.** = C( C =O N)+C( C H H S)+=O(C)+N( C C pC)+C( N C C H)+pC( N pC pC)+2 C( C H H H)+15 H(C)+S( C P)+P( S =S O O)+=S(P)+2 O( P C)+2 C( O H H H)+4 pC( pC pC H)+pC( pC pC Cl)+4 H(pC)+Cl(pC)

**Azamethiphos** = N( C pC C)+C( N H H S)+pC( N pC pN)+C( N =O O)+8 H(C)+pC( pC pC O)+pN( pC pC)+=O(C)+O( C pC)+pC( pC pC H)+pC( pC pC Cl)+pC( pC pN H)+2 H(pC)+Cl(pC)+S( C P)+P( S O O =O)+2 O( P C)+=O(P)+2 C( O H H H)

**Azinphos-ethyl**= C( N S H H)+N( C C N)+C( N =O pC)+S( C P)+N( N =N)+=N( N pC)+=O(C)+pC( C pC pC)+4 pC( pC pC H)+pC( pC pC =N)+4 H(pC)+P( S =S O O)+=S(P)+2 O( P C)+2 C( O C H H)+12 H(C)+2 C( C H H H)

**Azinphos-methyl**= C( S N H H)+S( C P)+N( C N C)+8 H(C)+N( N =N)+C( N =O pC)+=N( N pC)+=O(C)+pC( C pC pC)+P( S =S O O)+=S(P)+2 O( P C)+4 pC( pC pC H)+pC( pC pC =N)+4 H(pC)+2 C( O H H H)

**Benalaxyl**= N( C C pC)+C( N H C C)+C( N =O C)+pC( N pC pC)+15 H(C)+C( C H H H)+C( C =O O)+2 =O(C)+O( C C)+C( O H H H)+C( C pC H H)+3 pC( C pC pC)+8 pC( pC pC H)+8 H(pC)+2 C( pC H H H)

**Benazolin-ethyl**= N( C pC C)+C( N H H C)+pC( N pC pC)+C( N =O S)+7  
H(C)+pC( pC pC S)+pC( pC pC Cl)+2 =O(C)+S( C pC)+3 pC( pC pC H)+3  
H(pC)+Cl(pC)+C( C =O O)+O( C C)+C( O H H C)+C( C H H H)

**Bendiocarb**= 3 pC( pC pC O)+3 pC( pC pC H)+3 O( pC C)+3 H(pC)+C( O O C  
C)+2 C( C H H H)+C( O =O N)+=O(C)+N( C C H)+C( N H H H)+H(N)+9 H(C)

**Benfluralin**= 3 pC( N pC pC)+N( pC C C)+2 pC( pC pC H)+pC( pC pC C)+2 N(  
pC =O eO)+2 H(pC)+C( pC F F F)+2 =O(N)+2 eO(N)+2 C( N C H H)+2 C( C C H  
H)+2 C( C H H H)+14 H(C)+3 F(C)

**Benomyl**= 2 pC( N pC pC)+N( pC =C C)+4 pC( pC pC H)+N( pC =C)+=C( N N  
N)+C( N =O N)+4 H(pC)+N( =C H C)+2 H(N)+C( N =O O)+2 =O(C)+O( C C)+C(  
O H H H)+N( C C H)+C( N C H H)+2 C( C C H H)+C( C H H H)+12 H(C)

**Benoxacor**= pC( pC pC N)+4 pC( pC pC H)+pC( pC pC O)+N( pC C C)+C( N C H  
C)+C( C H H H)+7 H(C)+4 H(pC)+O( pC C)+C( O C H H)+C( N =O  
C)+=O(C)+C( C H Cl Cl)+2 Cl(C)

**Bensulferon-methyl**= C( N N =O)+N( C H S)+N( C H pC)+2 =O(C)+2 H(N)+S( N  
=O =O C)+2 =O(S)+C( S H H pC)+11 H(C)+2 pC( C pC pC)+5 pC( pC pC H)+C(  
pC =O O)+O( C C)+3 C( O H H H)+5 H(pC)+pC( N pN pN)+2 pN( pC pC)+2 pC(  
pN pC O)+2 O( pC C)

**Bentazone**= pC( C pC pC)+C( pC =O N)+pC( pC pC N)+4 pC( pC pC H)+N( pC S  
H)+4 H(pC)+=O(C)+N( C S C)+S( N =O =O N)+H(N)+2 =O(S)+C( N H C C)+7  
H(C)+2 C( C H H H)

**Bifenox**= 2 pC( pC pC O)+6 pC( pC pC H)+O( pC pC)+N( =O eO  
pC)+=O(N)+eO(N)+pC( pC pC N)+pC( pC pC C)+6 H(pC)+C( pC =O  
O)+=O(C)+O( C C)+C( O H H H)+2 pC( pC pC Cl)+2 Cl(pC)+3 H(C)

**Biphenyl** = 2 pC( pC pC pC)+10 pC( pC pC H)+10 H(pC)

**Bromopropylate**= C( C O pC pC)+C( C =O O)+O( C H)+2 pC( C pC  
pC)+H(O)+=O(C)+O( C C)+C( O C C H)+2 C( C H H H)+7 H(C)+8 pC( pC pC  
H)+2 pC( pC pC Br)+8 H(pC)+2 Br(pC)

**Bupirimate** = pC( pC pN O)+pC( pC pC C)+2 pN( pC pC)+O( pC S)+pC( pN pN  
N)+pC( pN pC C)+S( O =O =O N)+2 =O(S)+N( S C C)+2 C( N H H H)+N( pC H  
C)+H(N)+C( N C H H)+C( pC H H H)+C( pC C H H)+23 H(C)+2 C( C H H H)+2  
C( C C H H)

**Cadusafos**= P( S =O S O)+2 S( P C)+=O(P)+O( P C)+C( O C H H)+5 C( C H H  
H)+23 H(C)+2 C( S C C H)+2 C( C C H H)

**Captan**= 2 C( C C H C)+2 C( C =O N)+C( C C H H)+7 H(C)+2 =O(C)+N( C S C)+S( N C)+C( C =C H)+=C( C C H)+C( =C C H H)+C( S Cl Cl Cl)+3 Cl(C)+H(=C)

**Carbendazim**= pC( pC pC N)+4 pC( pC pC H)+pC( pC pC =N)+N( pC H C)+2 H(N)+C( N =N N)+4 H(pC)+=N( pC C)+N( C H C)+C( N =O O)+=O(C)+O( C C)+C( O H H H)+3 H(C)

**Carbofuran**= 2 pC( pC pC O)+pC( pC pC C)+2 O( pC C)+3 pC( pC pC H)+C( pC C H H)+3 H(pC)+C( O =O N)+=O(C)+N( C H C)+H(N)+C( N H H H)+C( C C C O)+2 C( C H H H)+11 H(C)

**Carbosulfan**= C( =O O N)+=O(C)+2 O( C pC)+2 N( C C S)+2 pC( O pC pC)+3 pC( pC pC H)+pC( pC pC C)+C( pC C H H)+C( C O C C)+4 C( C H H H)+29 H(C)+3 H(pC)+C( N H H H)+S( N N)+2 C( N C H H)+4 C( C C H H)

**Carboxin**= C( N =O C)+N( C H pC)+=O(C)+C( C =C S)+H(N)+pC( N pC pC)+=C( C C O)+C( =C H H H)+S( C C)+O( =C C)+C( O C H H)+C( C S H H)+5 pC( pC pC H)+5 H(pC)+7 H(C)

**Chlorbromuron**= N( C H pC)+C( N N =O)+H(N)+pC( N pC pC)+3 pC( pC pC H)+pC( pC pC Cl)+pC( pC pC Br)+3 H(pC)+Cl(pC)+Br(pC)+N( C C O)+=O(C)+C( N H H H)+O( N C)+C( O H H H)+6 H(C)

**Chlorethoxyfos**= 3 O( C P)+C( O C Cl H)+P( O O O =S)+=S(P)+2 C( O C H H)+2 C( C H H H)+C( C Cl Cl Cl)+4 Cl(C)+11 H(C)

**Chlorotoluron**= pC( N pC pC)+N( pC C H)+C( N =O N)+H(N)+ =O(C)+N( C C C)+2 C( N H H H)+3 pC( pC pC H)+pC( pC pC Cl)+pC( pC pC C)+3 H(pC)+Cl(pC)+C( pC H H H)+9 H(C)

**Chlorpyrifos** = P( =S O O O)+=S(P)+2 O( P C)+O( P pC)+pC( O pC pN)+2 pC( pC pC Cl)+pN( pC pC)+pC( pC pC H)+pC( pC pN Cl)+3 Cl(pC)+H(pC)+2 C( O C H H)+2 C( C H H H)+10 H(C)

**Chlorpyrifos-methyl**= pC( O pC pN)+O( pC P)+2 pC( pC pC Cl)+pN( pC pC)+P( O =S O O)+=S(P)+2 O( P C)+pC( pC pC H)+pC( pC pN Cl)+3 Cl(pC)+H(pC)+2 C( O H H H)+6 H(C)

**Chlorthalidimethyl**= 2 pC( pC pC C)+4 pC( pC pC Cl)+2 C( pC =O O)+2 =O(C)+2 O( C C)+2 C( O H H H)+4 Cl(pC)+6 H(C)

**Chlozolinate**= N( C C pC)+C( N O =O)+C( N C =O)+pC( N pC pC)+2 O( C C)+C( O C C C)+2 C( C H H H)+C( C =O O)+3 =O(C)+C( O C H H)+3 pC( pC pC H)+2 pC( pC pC Cl)+3 H(pC)+2 Cl(pC)+8 H(C)

**Chorsulfuron**= C( N N =O)+N( C H S)+N( C H pC)+2 =O(C)+2 H(N)+S( N =O =O pC)+2 =O(S)+pC( S pC pC)+pC( pC pS C)+pC( pC pC H)+pS( pC pC)+pC( pS

pC H)+C( pC =O O)+2 H(pC)+O( C C)+2 C( O H H H)+pC( N pN pN)+3 pN( pC pC)+pC( pN pN O)+pC( pN pN C)+C( pC H H H)+O( pC C)+9 H(C)

**Cinosulfuron**= C( N N =O)+N( C H S)+N( C H pC)+=O(C)+2 H(N)+S( N =O =O pC)+2 =O(S)+pC( N pN pN)+3 pN( pC pC)+2 pC( pN pN O)+3 O( pC C)+pC( S pC pC)+4 pC( pC pC H)+pC( pC pC O)+4 H(pC)+2 C( O C H H)+13 H(C)+3 C( O H H H)+O( C C)

**Clofentezine**= 2 pC( pN pC pN)+4 pN( pC pN)+2 pC( pC pC pC)+8 pC( pC pC H)+2 pC( pC pC Cl)+8 H(pC)+2 Cl(pC)

**Clomazone**= C( N pC H H)+N( C C O)+pC( C pC pC)+10 H(C)+C( N =O C)+O( N C)+=O(C)+C( C C C C)+C( C O H H)+2 C( C H H H)+pC( pC pC Cl)+4 pC( pC pC H)+Cl(pC)+4 H(pC)

**Clomeprop**= N( C H pC)+C( N =O C)+H(N)+=O(C)+C( C C O H)+C( C H H H)+O( C pC)+7 H(C)+pC( O pC pC)+7 pC( pC pC H)+2 pC( pC pC Cl)+pC( pC pC C)+7 H(pC)+2 Cl(pC)+C( pC H H H)+pC( N pC pC)

**Cycloate**= N( C C C)+C( N =O S)+C( N C H H)+C( N C CH)+2 C( C H H H)+=O(C)+S( C C)+C( S C H H)+5 C( C C H H)+20 H(C)

**Cyromazine**= 3 pC( pN pN N)+3 pN( pC pC)+N( pC H C)+2 N( pC H H)+5 H(N)+C( N C C H)+2 C( C C H H)+5 H(C)

**DDT 399** = C( pC pC C H)+2 pC( C pC pC)+C( C Cl Cl Cl)+3 Cl(C)+H(C)+8 pC( pC pC H)+2 pC( pC pC Cl)+8 H(pC)+2 Cl(pC)

**Demeton-S-methyl**= 3 C( H H C S)+15 H(C)+S( C P)+S( C C)+C( C H H H)+P( S =O O O)+=O(P)+2 O( P C)+2 C( O H H H)

**Desmedipham**= pC( O pC pC)+O( pC C)+9 pC( pC pC H)+2 C( O =O N)+2 =O(C)+2 N( C pC H)+2 pC( N pC pC)+2 H(N)+9 H(pC)+O( C C)+C( O H H C)+5 H(C)+C( C H H H)

**Desmetryn**= 2 pC( pN pN N)+3 pN( pC pC)+2 N( pC H C)+pC( pN pN S)+S( pC C)+2 H(N)+C( N H H H)+C( S H H H)+C( N C C H)+2 C( C H H H)+13 H(C)

**Diazinon** = O( pC P)+pC( O pC pN)+P( O O O =S)+2 O( P C)+=S(P)+pC( pC pC H)+2 pN( pC pC)+pC( pC pN C)+pC( pN pN C)+H(pC)+C( pC H H H)+C( pC H C C)+20 H(C)+4 C( C H H H)+2 C( O C H H)

**Dicamba** = pC( C pC pC)+C( pC =O O)+pC( pC pC O)+2 pC( pC pC Cl)+2 pC( pC pC H)+O( pC C)+2 Cl(pC)+2 H(pC)+C( H H H O)+3 H(C)+=O(C)+O( C H)+H(O)

**Dichlofluamid**= N( pC S S)+pC( N pC pC)+S( N =O =O N)+S( N C)+5 pC( pC pC H)+5 H(pC)+2 =O(S)+N( S C C)+2 C( N H H H)+C( S Cl Cl F)+2 Cl(C)+F(C)+6 H(C)

**Dichlorprop** = O( C pC)+C( O C H C)+pC( O pC pC)+3 pC( pC pC H)+2 pC( pC pC Cl)+3 H(pC)+2 Cl(pC)+C( C H H H)+4 H(C)+C( C =O O)+=O(C)+O( C H)+H(O)

**Dichlorvos**= P( =O O O O)+=O(P)+3 O( P C)+2 C( O H H H)+C( O =C H)+=C( C Cl Cl)+2 Cl(=C)+7 H(C)

**Diclofop-methyl**= 3 pC( O pC pC)+O( pC pC)+7 pC( pC pC H)+2 pC( pC pC Cl)+7 H(pC)+O( pC C)+2 Cl(pC)+C( O C C H)+C( C H H H)+C( C =O O)+7 H(C)+=O(C)+O( C C)+C( O H H H)

**Dicloran**= 2 pC( N pC pC)+N( pC H H)+2 H(N)+2 pC( pC pC Cl)+2 pC( pC pC H)+2 Cl(pC)+2 H(pC)+N( pC =O eO)+=O(N)+eO(N)

**Dicofol**= C( pC pC C O)+2 pC( C pC pC)+C( C Cl Cl Cl)+3 Cl(C)+O( C H)+H(O)+8 pC( pC pC H)+2 pC( pC pC Cl)+8 H(pC)+2 Cl(pC)

**Diflubenzuron 243**= C( =O N N)+2 =O(C)+N( C H pC)+N( C C H)+2 H(N)+pC( N pC pC)+7 pC( pC pC H)+pC( pC pC Cl)+7 H(pC)+Cl(pC)+C( N =O pC)+pC( C pC pC)+2 pC( pC pC F)+2 F(pC)

**Diflufenican**= C( N =O pC)+N( C H pC)+=O(C)+2 pC( C pC pC)+H(N)+pC( N pC pC)+9 pC( pC pC H)+pC( pC pN O)+pC( pC pN H)+pN( pC pC)+10 H(pC)+O( pC pC)+pC( O pC pC)+C( pC F F F)+2 pC( pC pC F)+2 F(pC)+3 F(C)

**Dimefuron**= 2 pC( N pC pC)+N( pC N C)+N( N =C)+C( N =O O)+2 =O(C)+O( C =C)+=C( N C O)+C( =C C C C)+3 pC( pC pC H)+pC( pC pC Cl)+3 H(pC)+N( pC H C)+Cl(pC)+H(N)+C( N =O N)+N( C C C)+2 C( N H H H)+3 C( C H H H)+15 H(C)

**Dimethachlor**= N( pC C C)+pC( N pC pC)+C( N H H C)+C( N =O C)+15 H(C)+C( C O H H)+O( C C)+=O(C)+C( C H H Cl)+2 pC( pC pC C)+C( O H H H)+Cl(C)+3 pC( pC pC H)+3 H(pC)+2 C( pC H H H)

**Dimethoate** = C( C S H H)+C( C =O N)+S( C P)+11 H(C)+P( S =S O O)+=S(P)+2 O( P C)+=O(C)+N( C H C)+H(N)+C( N H H H)+2 C( O H H H)

**Dinitramine** = 4 pC( N pC pC)+N( pC H H)+pC( pC pC C)+pC( pC pC H)+2 N( pC =O eO)+N( pC C C)+H(pC)+C( pC F F F)+2 =O(N)+2 eO(N)+2 H(N)+3 F(C)+2 C( N C H H)+2 C( C H H H)+10 H(C)

**Disulfoton** = S( P C)+P( S =S O O)+3 C( S C H H)+19 H(C)+S( C C)+=S(P)+2 O( P C)+3 C( C H H H)+2 C( O C H H)

**Diuron** = pC( N pC pC)+N( pC C H)+C( N =O N)+H(N)+=O(C)+N( C C C)+2 C( N H H H)+3 pC( pC pC H)+2 pC( pC pC Cl)+3 H(pC)+2 Cl(pC)+6 H(C)

**EPN**= P( =S pC O O)+=S(P)+pC( P pC pC)+O( P C)+O( P pC)+C( O C H H)+C( C H H H)+pC( O pC pC)+9 pC( pC pC H)+9 H(pC)+pC( pC pC N)+N( pC eO =O)+5 H(C)+eO(N)+=O(N)

**EPTC** = C( N =O S)+N( C C C)+=O(C)+S( C C)+2 C( N C H H)+C( S C H H)+3 C( C H H H)+2 C( C C H H)+19 H(C)

**Edifenphos** = P( S S =O O)+2 S( P pC)+=O(P)+O( P C)+2 pC( S pC pC)+C( O C H H)+C( C H H H)+10 pC( pC pC H)+10 H(pC)+5 H(C)

**Epoxiconazole** = C( C O pC C)+C( C O H pC)+O( C C)+2 pC( C pC pC)+C( C H H N)+4 H(C)+N( C N C)+N( N =C)+C( N =N H)+=N( C =C)+=C( N =N H)+H(=C)+8 pC( pC pC H)+pC( pC pC F)+8 H(pC)+F(pC)+pC( pC pC Cl)+Cl(pC)

**Etephon** = C( H H C P)+4 H(C)+C( C H H Cl)+P( C =O O O)+=O(P)+2 O( P H)+2 H(O)+Cl(C)

**Ethalfuralin** = 3 pC( N pC pC)+N( pC C C)+2 C( N C H H)+2 C( C H H H)+C( C C =C)+=C( C H H)+2 pC( pC pC H)+pC( pC pC C)+2 N( pC =O eO)+2 H(pC)+C( pC F F F)+2 =O(N)+2 eO(N)+3 F(C)+10 H(C)+2 H(=C)

**Ethiofencarb** = pC( O pC pC)+O( pC C)+C( O =O N)+=O(C)+N( C H C)+4 pC( pC pC H)+pC( pC pC C)+4 H(pC)+C( pC S H H)+S( C C)+C( S C H H)+C( C H H H)+H(N)+C( N H H H)+10 H(C)

**Ethion** = C( H H S S)+22 H(C)+2 S( C P)+2 P( S O O =S)+4 O( P C)+2 =S(P)+4 C( O C H H)+4 C( C H H H)

**Ethofumesate** = pC( pC pC C)+3 pC( pC pC H)+2 pC( pC pC O)+C( pC C C C)+3 H(pC)+O( pC S)+O( pC C)+3 C( C H H H)+C( C O O H)+O( C C)+15 H(C)+C( O C H H)+S( O =O =O C)+2 =O(S)+C( S H H H)

**Ethoprophos** = P( O =O S S)+O( P C)+=O(P)+2 S( P C)+2 C( S C H H)+2 C( C C H H)+3 C( C H H H)+C( O C H H)+19 H(C)

**Fenamiphos** = pC( pC pC O)+3 pC( pC pC H)+O( pC P)+P( O =O N O)+=O(P)+N( P C H)+O( P C)+pC( pC pC S)+pC( pC pC C)+3 H(pC)+S( pC C)+C( pC H H H)+C( S H H H)+C( O H H C)+18 H(C)+3 C( C H H H)+C( N H C C)+H(N)

**Fenarimol 365** = C( pC pC pC O)+3 pC( C pC pC)+O( C H)+H(O)+2 pC( pC pN H)+2 pN( pC pC)+pC( pN pN H)+11 H(pC)+8 pC( pC pC H)+2 pC( pC pC Cl)+2 Cl(pC)



**Fencloirim** = pC( pC pN pN)+pC( pC pC pC)+2 pN( pC pC)+2 pC( pN pC Cl)+6 pC( pC pC H)+2 Cl(pC)+6 H(pC)

**Fenitrothion 359** = O( pC P)+pC( O pC pC)+P( O O O =S)+2 O( P C)+2 C( O H H H)+=S(P)+3 pC( pC pC H)+pC( pC pC C)+pC( pC pC N)+3 H(pC)+C( pC H H H)+N( pC =O eO)+=O(N)+eO(N)+9 H(C)

**Fenoxycarb** = 3 pC( O pC pC)+O( pC C)+2 C( O C H H)+C( C N H H)+N( C C H)+C( N =O O)+H(N)+=O(C)+O( C C)+C( C H H H)+9 pC( pC pC H)+9 H(pC)+O( pC pC)+9 H(C)

**Fenthion** = O( pC P)+pC( O pC pC)+P( O O O =S)+2 O( P C)+2 C( O H H H)+=S(P)+3 pC( pC pC H)+pC( pC pC S)+pC( pC pC C)+3 H(pC)+S( pC C)+C( S H H H)+C( pC H H H)+12 H(C)

**Flamprop-Mmethyl** = N( C C pC)+C( N H C C)+C( N =O pC)+pC( N pC pC)+7 H(C)+C( C H H H)+C( C =O O)+2 =O(C)+O( C C)+C( O H H H)+pC( C pC pC)+8 pC( pC pC H)+8 H(pC)+pC( pC pC Cl)+pC( pC pC F)+Cl(pC)+F(pC)

**Fluazinam** = N( pC H pC)+pC( N pC pN)+H(N)+3 pC( N pC pC)+2 pC( pC pC H)+2 pC( pC pC C)+2 pC( pC pC Cl)+2 N( pC =O eO)+3 H(pC)+2 C( pC F F F)+2 Cl(pC)+2 =O(N)+2 eO(N)+6 F(C)+pN( pC pC)+pC( pN pC H)

**Fluazuron** = pC( N pC pC)+N( pC C H)+C( N =O N)+2 =O(C)+N( C H C)+2 H(N)+C( N =O pC)+2 pC( C pC pC)+7 pC( pC pC H)+pC( pC pC O)+2 pC( pC pC Cl)+8 H(pC)+O( pC pC)+2 Cl(pC)+pC( O pC pN)+pN( pC pC)+pC( pN pC H)+C( pC F F F)+3 F(C)+2 pC( pC pC F)+2 F(pC)

**Flumetralin** = N( C C pC)+C( N H H pC)+C( N H H C)+7 H(C)+3 pC( N pC pC)+5 pC( pC pC H)+2 pC( pC pC C)+2 N( pC =O eO)+5 H(pC)+C( pC F F F)+2 =O(N)+2 eO(N)+pC( pC pC F)+pC( pC pC Cl)+F(pC)+Cl(pC)+C( C H H H)+3 F(C)

**Fluometuron 191** = N( C pC H)+C( N =O N)+pC( N pC pC)+H(N)+4 pC( pC pC H)+=O(C)+N( C C C)+pC( pC pC C)+2 C( N H H H)+4 H(pC)+C( pC F F F)+6 H(C)+3 F(C)

**Flurenol** = C( C pC pC O)+C( C =O O)+2 pC( C pC pC)+2 O( C H)+2 H(O)+=O(C)+8 pC( pC pC H)+2 pC( pC pC pC)+8 H(pC)

**Fluridone** = C( C =O C)+2 C( C =C pC)+=O(C)+2 =C( C H N)+2 H(=C)+N( =C C =C)+C( N H H H)+3 pC( C pC pC)+9 pC( pC pC H)+9 H(pC)+C( pC F F F)+3 F(C)+3 H(C)

**Flurprimidol** = C( C O pC pC)+C( C C H C)+O( C H)+H(O)+2 pC( C pC pC)+2 pC( pC pN H)+2 pN( pC pC)+pC( pN pN H)+7 H(pC)+4 pC( pC pC H)+pC( pC pC O)+O( pC C)+C( O F F F)+3 F(C)+2 C( C H H H)+7 H(C)

**Flutriafol 188** = C( O pC pC C)+O( C H)+2 pC( C pC pC)+C( C pN H H)+ N( C pC pN)+2 H(C)+2 pC( pC pC F)+8 pC( pC pC H)+H(O)+2 pC( pN pN H)+pN( pN pC)+pN( pC pC)+2 F(pC)+10 H(pC)

**Fonofos** = pC( S pC pC)+S( pC P)+5 pC( pC pC H)+5 H(pC)+P( S O =S C)+O( P C)+=S(P)+C( O C H H)+2 C( C H H H)+C( P C H H)+10 H(C)

**Formothion** = C( C S H H)+C( C =O N)+S( C P)+12 H(C)+2 =O(C)+N( C C C)+C( N H H H)+P( S =S O O)+C( N =O H)+=S(P)+2 O( P C)+2 C( O H H H)

**Fosthiazate** = P( O =O S N)+O( P C)+=O(P)+S( P C)+N( P C C)+C( O C H H)+3 C( C H H H)+C( S C H C)+18 H(C)+C( C C H H)+C( N C H H)+C( N =O S)+C( C S H H)+S( C C)+=O(C)

**Furalaxyl** = N( C C pC)+C( N =O pC)+C( N H C C)+2 =O(C)+pC( C pC pO)+5 pC( pC pC H)+pO( pC pC)+pC( pC pO H)+6 H(pC)+pC( N pC pC)+2 pC( pC pC C)+13 H(C)+C( C =O O)+C( C H H H)+O( C C)+C( O H H H)+2 C( pC H H H)

**Furathiocarb** = 2 C( O =O N)+2 O( C pC)+2 =O(C)+2 N( C C S)+2 C( N H H H)+S( N N)+2 pC( O pC pC)+3 pC( pC pC H)+pC( pC pC C)+3 H(pC)+C( pC C H H)+C( C C C O)+3 C( C H H H)+23 H(C)+O( C C)+C( O C H H)+2 C( C C H H)

**Glyphosate** = N( H C C)+H(N)+C( N C H H)+C( N P H H)+P( C =O O O)+C( C =O O)+=O(C)+O( C H)+3 H(O)+=O(P)+2 O( P H)+4 H(C)

**Haloxyfop etotyl** = 2 pC( O pC pC)+O( pC C)+C( O C H C)+2 C( C H H H)+13 H(C)+C( C =O O)+=O(C)+2 O( C C)+3 C( O C H H)+5 pC( pC pC H)+6 H(pC)+O( pC pC)+pC( O pC pN)+pC( pC pC Cl)+pN( pC pC)+pC( pC pC C)+pC( pC pN H)+Cl(pC)+C( pC F F F)+3 F(C)

**Hexaconazole** = C( C O pC C)+C( C pN H H)+O( C H)+pC( C pC pC)+3 C( C C H H)+C( C H H H)+ N( C pC pN)+11 H(C)+H(O)+3 pC( pC pC H)+2 pC( pC pC Cl)+5 H(pC)+2 Cl(pC)+2 pC( pN pN H)+pN( pN pC)+pN( pC pC)

**Hexazinone**= 3 N( C C C)+C( N H C C)+C( N =O N)+C( N =O =N)+2 =O(C)+20 H(C)+C( N =N N)+=N( C C)+3 C( N H H H)+5 C( C C H H)

**Hexythiazox** = N( C C C)+C( N =O S)+C( N C H H)+C( N =O N)+2 =O(C)+N( C H C)+C( C S H pC)+S( C C)+13 H(C)+pC( C pC pC)+H(N)+C( N C C)+5 C( C C H H)+5 pC( pC pC H)+5 H(pC)

**Imazalil** = C( H O pC C)+9 H(C)+O( C C)+pC( C pC pC)+C( C N H H)+N( C C =C)+C( N =N H)+=C( N C H)+=N( C C)+C( =N =C H)+3 H(=C)+2 pC( pC pC Cl)+3 pC( pC pC H)+3 H(pC)+2 Cl(pC)+C( O H H C)+C( C H H =C)+=C( C H H)

**Imibenconazole** = C( C =N S)+C( C pN H H)+=N( C pC)+S( C C)+C( S H H pC)+pC( =N pC pC)+4 H(C)+pC( C pC pC)+ N( C pC pN)+pN( pN pC)+2 pC( pN pN H)+pN( pC pC)+3 pC( pC pC Cl)+7 pC( pC pC H)+3 Cl(pC)+9 H(pC)

**Imidacloprid** = C( pC N H H)+pC( C pC pC)+N( C C C)+C( N =N N)+C( N C H H)+C( C =N H H)+=N( C C)+N( C H N)+H(N)+N( N =O eO)+6 H(C)+pC( pC pN H)+2 pC( pC pC H)+pN( pC pC)+pC( pN pC Cl)+3 H(pC)+Cl(pC)+=O(N)+eO(N)

**Iprodione** = N( C C pC)+2 C( N N =O)+C( N C =O)+pC( N pC pC)+N( C C C)+C( C N H H)+3 =O(C)+9 H(C)+3 pC( pC pC H)+2 pC( pC pC Cl)+3 H(pC)+2 Cl(pC)+N( C H C)+H(N)+C( N C C H)+2 C( C H H H)

**Isazofos** = pC( pN O pN)+pN( pC pC)+O( pC P)+pN( pC pN)+ N( C pC pN)+C( pN C C H)+4 C( C H H H)+pC( pN pN Cl)+Cl(pC)+P( O =S O O)+=S(P)+2 O( P C)+2 C( O C H H)+17 H(C)

**Isofenphos** = pC( O pC pC)+O( pC P)+pC( pC pC C)+4 pC( pC pC H)+P( O =S N O)+=S(P)+N( P C H)+O( P C)+C( O C H H)+C( pC =O O)+4 H(pC)+5 C( C H H H)+C( N C C H)+H(N)+=O(C)+O( C C)+C( O H C C)+19 H(C)

**Isoproturon** = pC( N pC pC)+N( pC H C)+H(N)+C( N =O N)+=O(C)+4 pC( pC pC H)+pC( pC pC C)+4 H(pC)+C( pC C C H)+2 C( C H H H)+13 H(C)+N( C C C)+2 C( N H H H)

**Isoxaben** = N( C H C)+C( N =C O)+H(N)+C( N =O pC)+=O(C)+pC( C pC pC)+2 pC( pC pC O)+3 pC( pC pC H)+2 O( pC C)+3 H(pC)+2 C( O H H H)+=C( C C)+O( C =N)+C( =C =N C)+=N( C O)+C( C C C C)+3 C( C H H H)+2 C( C C H H)+19 H(C)

**Lenacil** = N( C C C)+C( N =O C)+C( N =O N)+C( N H C C)+17 H(C)+2 =O(C)+C( C C =C)+7 C( C C H H)+=C( C N C)+N( C H =C)+H(N)+C( =C C H H)

**Linuron** = N( pC C H)+pC( N pC pC)+C( N =O N)+H(N)+=O(C)+N( C C O)+3 pC( pC pC H)+C( N H H H)+O( N C)+2 pC( pC pC Cl)+C( O H H H)+6 H(C)+3 H(pC)+2 Cl(pC)

**Maleic hydrazide 604** = 2 pC( O pN pC)+2 O( pC H)+2 H(O)+2 pN( pC pN)+2 pC( pC pC H)+2 H(pC)

**Mecoprop 142** = pC( pC pC O)+3 pC( pC pC H)+pC( pC pC C)+O( pC C)+pC( pC pC Cl)+3 H(pC)+Cl(pC)+C( pC H H H)+7 H(C)+C( O C C H)+C( C H H H)+C( C =O O)+=O(C)+O( C H)+H(O)

**Mefenacet** = C( S =N O)+S( C pC)+=N( C pC)+O( C C)+C( O C H H)+C( C =O N)+=O(C)+N( C C pC)+C( N H H H)+pC( N pC pC)+9 pC( pC pC H)+9 H(pC)+pC( S pC pC)+pC( pC pC =N)+5 H(C)

**Mepronil** = C( N =O pC)+N( C H pC)+=O(C)+2 pC( C pC pC)+H(N)+pC( N pC pC)+8 pC( pC pC H)+C( pC H H H)+8 H(pC)+pC( pC pC O)+O( pC C)+C( O C C H)+2 C( C H H H)+10 H(C)

**Metalaxyl 67** = N( C pC C)+C( N =O C)+pC( N pC pC)+C( N C C H)+2 =O(C)+C( C O H H)+C( C =O O)+C( C H H H)+18 H(C)+2 pC( pC pC C)+2 O( C C)+3 pC( pC pC H)+2 C( O H H H)+2 C( pC H H H)+3 H(pC)

**Metamitron** = C( =N pC C)+=N( C N)+pC( C pC pC)+C( C =O N)+5 pC( pC pC H)+N( =N =C)+=C( N C N)+=O(C)+C( =C H H H)+N( =C C N)+N( N H H)+2 H(N)+3 H(C)+5 H(pC)

**Metazachlor** = C( N H H N)+N( C C pC)+11 H(C)+N( C =C N)+C( N =O C)+pC( N pC pC)+2 =C( N C H)+N( N =C)+C( =C =C H)+2 H(=C)+=O(C)+C( C H H Cl)+2 pC( pC pC C)+3 pC( pC pC H)+2 C( pC H H H)+3 H(pC)+Cl(C)

**Methamidophos** = P( =O S O N)+=O(P)+S( P C)+O( P C)+N( P H H)+C( S H H H)+2 H(N)+C( O H H H)+6 H(C)

**Methidathion** = C( N S H H)+N( C C N)+S( C P)+11 H(C)+C( N =O S)+N( N =C)+=O(C)+S( C =C)+=C( N S O)+O( =C C)+3 C( O H H H)+P( S =S O O)+=S(P)+2 O( P C)

**Methiocarb** = pC( pC pC O)+2 pC( pC pC H)+O( pC C)+2 pC( pC pC C)+pC( pC pC S)+2 H(pC)+2 C( pC H H H)+S( pC C)+C( S H H H)+C( O =O N)+=O(C)+N( C H C)+H(N)+C( N H H H)+12 H(C)

**Methomyl** = N( =C O)+=C( N C S)+O( N C)+C( =C H H H)+S( =C C)+C( O =O N)+=O(C)+N( C H C)+H(N)+C( N H H H)+C( S H H H)+9 H(C)

**Metobromuron** = N( C H pC)+C( N N =O)+H(N)+pC( N pC pC)+4 pC( pC pC H)+pC( pC pC Br)+4 H(pC)+Br(pC)+N( C C O)+=O(C)+C( N H H H)+O( N C)+C( O H H H)+6 H(C)

**Metolachlor (ANSIWSSA)** = N( C C pC)+C( N H C C)+C( N =O C)+pC( N pC pC)+19 H(C)+C( C H H O)+2 C( C H H H)+O( C C)+C( O H H H)+=O(C)+C( C H H Cl)+Cl(C)+2 pC( pC pC C)+3 pC( pC pC H)+C( pC H H H)+3 H(pC)+C( pC H H C)

**Metosulam** = C( N =N S)+N( C =C)+=N( C N)+S( C =O =O N)+=C( N N N)+N( =N =C C)+C( N =C O)+=C( C =C)+N( =C =C)+=C( N =C O)+O( C C)+O( =C C)+2 =O(S)+N( S H pC)+H(N)+pC( N pC pC)+2 pC( pC pC Cl)+pC( pC pC C)+2 pC( pC pC H)+2 H(pC)+2 Cl(pC)+C( pC H H H)+2 C( O H H H)+9 H(C)

**Metoxuron** = N( C H pC)+C( N N =O)+H(N)+pC( N pC pC)+3 pC( pC pC H)+pC( pC pC O)+pC( pC pC Cl)+3 H(pC)+O( pC C)+Cl(pC)+C( O H H H)+N( C C C)+=O(C)+2 C( N H H H)+9 H(C)

**Metribuzin** = C( C =N C)+C( C C C C)+2 =N( C =N)+C( C =O N)+=O(C)+N( C N C)+N( N H H)+C( N =N S)+S( C C)+C( S H H H)+3 C( C H H H)+2 H(N)+12 H(C)

**Metsulfuronmethyl** = C( N N =O)+N( C H S)+N( C H pC)+2 =O(C)+2 H(N)+S( N =O =O pC)+2 =O(S)+pC( S pC pC)+4 pC( pC pC H)+pC( pC pC C)+4 H(pC)+C( pC =O O)+O( C C)+2 C( O H H H)+pC( N pN pN)+3 pN( pC pC)+pC( pN pN O)+pC( pN pN C)+C( pC H H H)+O( pC C)+9 H(C)

**Monolinuron** = C( =O N N)+=O(C)+N( C pC H)+N( C C O)+pC( N pC pC)+H(N)+4 pC( pC pC H)+pC( pC pC Cl)+4 H(pC)+Cl(pC)+C( N H H H)+O( N C)+C( O H H H)+6 H(C)

**Nicosulfuron** = C( N N =O)+N( C H S)+N( C H pC)+2 =O(C)+2 H(N)+S( N =O =O pC)+2 =O(S)+pC( N pN pN)+3 pN( pC pC)+2 pC( pN pC O)+3 pC( pC pC H)+2 O( pC C)+4 H(pC)+pC( S pN pC)+pC( pC pC C)+pC( pN pC H)+C( pC N =O)+N( C C C)+2 C( N H H H)+12 H(C)+2 C( O H H H)

**Novaluron** = pC( N pC pC)+N( pC C H)+C( N =O N)+2 =O(C)+N( C H C)+2 H(N)+C( N =O pC)+pC( C pC pC)+6 pC( pC pC H)+pC( pC pC O)+pC( pC pC Cl)+6 H(pC)+O( pC C)+Cl(pC)+C( O C F F)+C( C O F H)+O( C C)+C( O F F F)+6 F(C)+H(C)+2 pC( pC pC F)+2 F(pC)

**Ofurace** = N( C C pC)+C( N C H C)+C( N =O C)+pC( N pC pC)+2 =O(C)+C( C H H Cl)+C( C =O O)+13 H(C)+C( C C H H)+O( C C)+C( O C H H)+2 pC( pC pC C)+3 pC( pC pC H)+2 C( pC H H H)+3 H(pC)+Cl(C)

**Omethoate** = C( C S H H)+C( C =O N)+S( C P)+=O(C)+N( C C H)+P( S O O =O)+2 O( P C)+=O(P)+2 C( O H H H)+11 H(C)+C( N H H H)+H(N)

**Oryzalin** = 3 pC( N pC pC)+N( pC C C)+2 C( N C H H)+2 C( C C H H)+2 C( C H H H)+14 H(C)+2 pC( pC pC H)+pC( pC pC S)+2 N( pC =O eO)+2 H(pC)+S( pC =O =O N)+2 =O(S)+N( S H H)+2 H(N)+2 =O(N)+2 eO(N)

**Oxadiazon** = N( N pC C)+N( N =C)+pC( N pC pC)+C( N O =O)+2 pC( pC pC Cl)+2 pC( pC pC H)+=C( N O C)+O( =C C)+C( =C C C C)+5 C( C H H H)+=O(C)+pC( pC pC O)+O( pC C)+2 Cl(pC)+2 H(pC)+C( O C C H)+15 H(C)

**Oxadixyl** = N( C N pC)+C( N =O C)+N( N C C)+pC( N pC pC)+2 =O(C)+2 C( C O H H)+2 O( C C)+C( O H H H)+2 pC( pC pC C)+3 pC( pC pC H)+2 C( pC H H H)+3 H(pC)+C( N C H H)+C( N O =O)+15 H(C)

**Oxamyl** = N( O =C)+O( N C)+=C( N S C)+C( O =O N)+2 =O(C)+N( C H C)+H(N)+3 C( N H H H)+S( =C C)+C( S H H H)+C( =C =O N)+N( C C C)+12 H(C)

**Oxycarboxin** = C(N=O C)+N(C pC H)+=O(C)+pC(N pC pC)+H(N)+C(C=C S)+=C(C O C)+S(C C=O=O)+O(=C C)+C(=C H H H)+C(O C H H)+C(C S H H)+2=O(S)+5 pC(pC pC H)+5 H(pC)+7 H(C)

**Oxydemetonmethyl** = 3 C(C S H H)+S(C P)+15 H(C)+S(C=O C)+=O(S)+C(C H H H)+P(S O O=O)+2 O(P C)+=O(P)+2 C(O H H H)

**Parathion 396** = O(pC P)+pC(O pC pC)+P(O O O=S)+2 O(P C)+2 C(O C H H)+2 C(C H H H)+=S(P)+4 pC(pC pC H)+pC(pC pC N)+4 H(pC)+N(pC=O eO)+=O(N)+eO(N)+10 H(C)

**Parathion methyl** = O(pC P)+pC(O pC pC)+P(O O O=S)+2 O(P C)+2 C(O H H H)+=S(P)+4 pC(pC pC H)+pC(pC pC N)+4 H(pC)+N(pC=O eO)+=O(N)+eO(N)+6 H(C)

**Pentachloronitrobenzene 446** = pC(N pC pC)+N(pC=O eO)+=O(N)+eO(N)+5 pC(pC pC Cl)+5 Cl(pC)

**Pentachlorophenol 363** = pC(O pC pC)+O(pC H)+H(O)+5 pC(pC pC Cl)+5 Cl(pC)

**Phorate** = P(=S S O O)+=S(P)+S(P C)+2 O(P C)+2 C(O C H H)+3 C(C H H H)+C(S S H H)+S(C C)+C(S C H H)+17 H(C)

**Phosalone** = N(C pC C)+C(N=O O)+pC(N pC pC)+C(N H H S)+=O(C)+O(C pC)+3 pC(pC pC H)+pC(pC pC O)+pC(pC pC Cl)+3 H(pC)+Cl(pC)+12 H(C)+S(C P)+P(S=S O O)+=S(P)+2 O(P C)+2 C(O C H H)+2 C(C H H H)

**Phosmet** = 2 C(N=O pC)+N(C C C)+2 =O(C)+2 pC(C pC pC)+C(N S H H)+S(C P)+P(S=S O O)+=S(P)+2 O(P C)+2 C(O H H H)+4 pC(pC pC H)+4 H(pC)+8 H(C)

**Piperophos** = C(N=O C)+N(C C C)+=O(C)+C(C S H H)+C(N C H H)+C(N C H C)+5 C(C C H H)+S(C P)+P(S=S O O)+=S(P)+2 O(P C)+2 C(O C H H)+3 C(C H H H)+28 H(C)

**Pirimicarb** = pC(pC pN O)+pC(pC pC C)+2 pN(pC pC)+O(pC C)+C(O=O N)+=O(C)+N(C C C)+4 C(N H H H)+18 H(C)+pC(pN pN N)+pC(pN pC C)+N(pC C C)+2 C(pC H H H)

**Pirimiphos-ethyl** = O(pC P)+pC(O pC pN)+P(O O O=S)+2 O(P C)+=S(P)+pC(pC pC H)+2 pN(pC pC)+pC(pC pN C)+pC(pN pN N)+H(pC)+C(pC H H H)+N(pC C C)+2 C(N C H H)+4 C(C H H H)+23 H(C)+2 C(O C H H)

**Pirimiphos-methyl** = pC(N pN pN)+N(pC C C)+2 pN(pC pC)+pC(pN pC O)+pC(pC pC H)+pC(pC pN C)+O(pC P)+C(pC H H H)+H(pC)+P(O=S O)+2 C(N C H H)+2 C(C H H H)+=S(P)+2 O(P C)+2 C(O H H H)+19 H(C)

**Pretilachlor** = N( C pC C)+C( N =O C)+pC( N pC pC)+C( N C H H)+=O(C)+C( C H Cl H)+2 C( C O H H)+23 H(C)+2 pC( pC pC C)+Cl(C)+O( C C)+C( C C H H)+3 C( C H H H)+3 pC( pC pC H)+2 C( pC C H H)+3 H(pC)

**Prochloraz** = 2 C( N C H H)+2 N( C C C)+C( C O H H)+O( C pC)+C( N =O N)+=O(C)+pC( O pC pC)+3 pC( pC pC Cl)+2 pC( pC pC H)+3 Cl(pC)+2 H(pC)+C( N =N H)+C( N =C H)+=C( C =N H)+=N( C =C)+13 H(C)+H(=C)+C( C C H H)+C( C H H H)

**Profenofos** = O( pC P)+pC( O pC pC)+P( O S O =O)+S( P C)+C( S C H H)+O( P C)+C( O C H H)+=O(P)+3 pC( pC pC H)+pC( pC pC Cl)+pC( pC pC Br)+3 H(pC)+Br(pC)+Cl(pC)+C( C C H H)+2 C( C H H H)+12 H(C)

**Prometon** = pC( O pN pN)+O( pC C)+3 pN( pC pC)+2 pC( pN pN N)+2 N( pC C H)+2 C( N C C H)+2 H(N)+4 C( C H H H)+17 H(C)+C( O H H H)

**Prometryn** = 2 pC( N pN pN)+2 N( pC H C)+3 pN( pC pC)+pC( pN pN S)+S( pC C)+C( S H H H)+2 H(N)+2 C( N C C H)+4 C( C H H H)+17 H(C)

**Propachlor 822** = pC( pC pC N)+5 pC( pC pC H)+N( pC C C)+C( N C C H)+2 C( C H H H)+9 H(C)+5 H(pC)+C( N =O C)+=O(C)+C( C H H Cl)+Cl(C)

**Propazine 222** = 2 pC( pN pN N)+3 pN( pC pC)+2 N( pC H C)+pC( pN pN Cl)+Cl(pC)+2 C( N C C H)+2 H(N)+4 C( C H H H)+14 H(C)

**Propisochlor** = N( C pC C)+C( N =O C)+pC( N pC pC)+C( N H H O)+=O(C)+C( C H Cl H)+19 H(C)+2 pC( pC pC C)+Cl(C)+O( C C)+C( O C C H)+3 C( C H H H)+3 pC( pC pC H)+C( pC H H H)+3 H(pC)+C( pC C H H)

**Propoxur 96** = 2 pC( pC pC O)+4 pC( pC pC H)+2 O( pC C)+C( O =O N)+=O(C)+N( C H C)+H(N)+C( N H H H)+10 H(C)+C( O C H C)+2 C( C H H H)+4 H(pC)

**Prosulfocarb** = C( pC S H H)+pC( C pC pC)+S( C C)+16 H(C)+C( S =O N)+=O(C)+N( C C C)+5 pC( pC pC H)+5 H(pC)+2 C( N C H H)+2 C( C C H H)+2 C( C H H H)

**Prothiofos** = P( S O O =S)+S( P C)+O( P pC)+O( P C)+=S(P)+C( O C H H)+2 C( C H H H)+C( S C H H)+C( C C H H)+pC( O pC pC)+3 pC( pC pC H)+2 pC( pC pC Cl)+3 H(pC)+2 Cl(pC)+12 H(C)

**Pyrazophos** = C( N N =C)+N( C =C)+N( C =C =N)+=C( C C H)+C( =C =N O)+2 H(=C)+=C( N C C)+C( =C =C C)+=C( C N H)+=N( C N)+3 O( C P)+P( O O =S O)+C( =C H H H)+C( C =O O)+=O(C)+O( C C)+3 C( O C H H)+3 C( C H H H)+=S(P)+18 H(C)

**Pyrifenoxy** = C(=N pC C)+=N(C O)+2 pC(C pC pC)+C(C pC H H)+O(=N C)+C(O H H H)+2 pC(pC pC Cl)+5 pC(pC pC H)+2 Cl(pC)+7 H(pC)+2 pC(pC pN H)+pN(pC pC)+5 H(C)

**Rimsulfuron** = C(N N =O)+N(C H S)+N(C H pC)+=O(C)+2 H(N)+S(N =O =O pC)+4 =O(S)+pC(S pC pN)+pC(pC pC S)+3 pN(pC pC)+pC(pN pC H)+3 pC(pC pC H)+4 H(pC)+S(pC =O =O C)+C(S H H C)+11 H(C)+C(C H H H)+pC(N pN pN)+2 pC(pN pC O)+2 O(pC C)+2 C(O H H H)

**Simazine** = 2 pC(pN pN N)+3 pN(pC pC)+2 N(pC H C)+pC(pN pN Cl)+Cl(pC)+2 C(N C H H)+2 H(N)+2 C(C H H H)+10 H(C)

**Sulfotep** = O(P P)+2 P(O =S O O)+2 =S(P)+4 O(P C)+4 C(O C H H)+4 C(C H H H)+20 H(C)

**Sulprofos** = pC(O pC pC)+O(pC P)+P(O =S S O)+=S(P)+4 pC(pC pC H)+pC(pC pC S)+4 H(pC)+S(pC C)+C(S H H H)+S(P C)+O(P C)+C(S C H H)+C(C C H H)+2 C(C H H H)+C(O C H H)+15 H(C)

**TRIBUFOS** = P(S S S =O)+3 S(P C)+=O(P)+3 C(S C H H)+6 C(C C H H)+3 C(C H H H)+27 H(C)

**Tebuconazole** = C(C C H H)+C(C H H pC)+C(C O C C)+O(C H)+C(C pN H H)+ N(C pC pN)+pN(pN pC)+2 pC(pN pN H)+pN(pC pC)+6 H(pC)+15 H(C)+pC(C pC pC)+4 pC(pC pC H)+pC(pC pC Cl)+Cl(pC)+H(O)+C(C C C C)+3 C(C H H H)

**Tebufenozide** = N(C N C)+2 C(N =O pC)+N(N C H)+C(N C C C)+2 =O(C)+H(N)+5 pC(C pC pC)+7 pC(pC pC H)+7 H(pC)+2 C(pC H H H)+C(pC C H H)+4 C(C H H H)+20 H(C)

**Tebuthiuron** = pC(pS pN N)+pS(pC pC)+2 pN(pC pN)+N(pC C C)+pC(pS pN C)+C(pC C C C)+3 C(C H H H)+C(N =O N)+2 C(N H H H)+=O(C)+N(C H C)+H(N)+15 H(C)

**Tecnazene** = pC(N pC pC)+N(pC =O eO)+=O(N)+eO(N)+4 pC(pC pC Cl)+pC(pC pC H)+4 Cl(pC)+H(pC)

**Temephos** = S(pC pC)+2 pC(S pC pC)+8 pC(pC pC H)+2 pC(pC pC O)+8 H(pC)+2 O(pC P)+2 P(O O O =S)+4 O(P C)+2 =S(P)+4 C(O H H H)+12 H(C)

**Terbufos** = P(=S O O S)+=S(P)+2 O(P C)+S(P C)+2 C(O H H C)+21 H(C)+5 C(C H H H)+C(S S H H)+S(C C)+C(S C C C)

**Terbumeton** = 2 pC(pN pN N)+3 pN(pC pC)+N(pC C H)+pC(pN pN O)+N(pC N H)+N(N C H)+O(pC C)+C(O H H H)+C(N C H H)+3 H(N)+4 C(C H H H)+17 H(C)+C(N C C C)



**Terbutylazine** = pC( pN pN Cl)+3 pN( pC pC)+Cl(pC)+2 pC( pN pN N)+2 N( pC H C)+2 H(N)+C( N H H C)+14 H(C)+4 C( C H H H)+C( N C C C)

**Terbutryn 210** = 2 pC( N pN pN)+2 N( pC H C)+3 pN( pC pC)+2 H(N)+C( N C C C)+pC( pN pN S)+S( pC C)+4 C( C H H H)+C( S H H H)+C( N H H C)+17 H(C)

**Tetraconazole** = C( C H pC C)+C( C pN H H)+6 H(C)+pC( C pC pC)+C( C O H H)+O( C C)+ N( C pC pN)+C( O C F F)+C( C F F H)+3 pC( pC pC H)+2 pC( pC pC Cl)+5 H(pC)+2 Cl(pC)+2 pC( pN pN H)+pN( pN pC)+pN( pC pC)+4 F(C)

**Tetradifon** = S( =O =O pC pC)+2 =O(S)+2 pC( S pC pC)+6 pC( pC pC H)+4 pC( pC pC Cl)+6 H(pC)+4 Cl(pC)

**Thifensulfuronmethy** = C( N N =O)+N( C H S)+N( C H pC)+=O(C)+2 H(N)+S( N =O =O pC)+2 =O(S)+pC( S pC pC)+pC( pC pC Cl)+4 pC( pC pC H)+Cl(pC)+4 H(pC)+pC( N pN pN)+3 pN( pC pC)+pC( pN pN O)+pC( pN pN C)+C( pC H H H)+O( pC C)+C( O H H H)+6 H(C)

**Thiodicarb** = 2 N( C S C)+2 C( N H H H)+S( N N)+2 C( N =O O)+2 =O(C)+2 O( C N)+2 N( O =C)+2 =C( N S C)+2 S( =C C)+2 C( =C H H H)+2 C( S H H H)+18 H(C)

**Thiometon** = S( P C =O =O)+P( S =S O O)+3 C( S H H C)+2 =O(S)+19 H(C)+S( C C)+3 C( C H H H)+=S(P)+2 O( P C)+2 C( O C H H)

**Thiram** = 2 S( S C)+2 C( S =S N)+2 =S(C)+2 N( C C C)+4 C( N H H H)+12 H(C)

**Tolyfluanid** = N( S S pC)+S( N =O =O N)+S( N C)+pC( N pC pC)+4 pC( pC pC H)+pC( pC pC C)+4 H(pC)+C( pC H H H)+2 =O(S)+N( S C C)+2 C( N H H H)+C( S Cl Cl F)+9 H(C)+2 Cl(C)+F(C)

**Tri-allate** = C( C S H H)+C( C Cl =C)+S( C C)+C( S =O N)+=O(C)+N( C C C)+2 C( N H C C)+16 H(C)+4 C( C H H H)+Cl(C)+=C( C Cl Cl)+2 Cl(=C)

**Triadimefon** = C( N O C H)+N( C pC pN)+O( C pC)+C( C =O C)+10 H(C)+=O(C)+pC( O pC pC)+4 pC( pC pC H)+pC( N pN H)+pN( N pC)+C( C C C)+3 C( C H H H)+pN( pC pC)+pC( pN pN H)+6 H(pC)+pC( pC pC Cl)+Cl(pC)

**Triazamate** = C( S =N N)+S( C C)+=N( C C)+N( C =N C)+=N( N C)+C( =N C =N)+C( C C C C)+C( S C H H)+C( C =O O)+2 =O(C)+O( C C)+C( O C H H)+4 C( C H H H)+C( N =O N)+N( C C C)+2 C( N H H H)+22 H(C)

**Triazophos** = N( C N pC)+C( N =N H H)+N( N =C)+pC( N pC pC)+=N( C =C)+=C( =N N O)+O( =C P)+5 pC( pC pC H)+5 H(pC)+P( O O O =S)+2 O( P C)+=S(P)+2 C( O C H H)+2 C( C H H H)+12 H(C)

**Tribenuron methyl** = C(N N=O)+N(C C pC)+N(C S H)+2=O(C)+C(N H H H)+S(N=O=O pC)+H(N)+2=O(S)+pC(S pC pC)+pC(N pN pN)+3 pN(pC pC)+pC(pN pN C)+pC(pN pN O)+C(pC H H H)+O(pC C)+pC(pC pC C)+4 pC(pC pC H)+C(pC=O O)+4 H(pC)+O(C C)+2 C(O H H H)+12 H(C)

**Trichlorfon** = C(H P O C)+7 H(C)+P(C O=O O)+O(C H)+C(C Cl Cl Cl)+H(O)+2 O(P C)+=O(P)+3 Cl(C)+2 C(O H H H)

**Triflumuron** = C(N N=O)+N(C H pC)+N(C C H)+2=O(C)+2 H(N)+C(N=O pC)+pC(C pC pC)+pC(N pC pC)+8 pC(pC pC H)+pC(pC pC O)+8 H(pC)+O(pC C)+C(O F F F)+3 F(C)+pC(pC pC Cl)+Cl(pC)

**Vinclozolin** = N(pC C C)+pC(N pC pC)+C(N=O C)+C(N=O O)+3 pC(pC pC H)+2 pC(pC pC Cl)+3 H(pC)+2 Cl(pC)+2=O(C)+C(C O C C)+O(C C)+C(C H=C)+C(C H H H)+4 H(C)+=C(C H H)+2 H(=C)

**butachlor [ANSI, JMAF, WSSA]** = N(C pC C)+C(N O H H)+pC(N pC pC)+C(N=O C)+O(C C)+23 H(C)+2 pC(pC pC C)+=O(C)+C(C H H Cl)+C(O C H H)+2 C(C C H H)+3 C(C H H H)+3 pC(pC pC H)+Cl(C)+2 C(pC C H H)+3 H(pC)

**trifluralin** = N(C C pC)+2 C(N C H H)+3 pC(N pC pC)+2 pC(pC pC H)+pC(pC pC C)+2 N(pC=O eO)+2=O(N)+2 eO(N)+2 H(pC)+C(pC F F F)+2 C(C C H H)+2 C(C H H H)+3 F(C)+14 H(C)

## Annexe 2 : Valeurs expérimentales de $\log K_{ow}$ des pesticides de la base de données étudiée

N°	Pesticides	Formule	$\log k_{ow}$
1	Acephate	$C_4H_{10}NO_3PS$	-0,85
2	Alachlor	$C_{14}H_{20}ClNO_2$	3,52
3	Ametryn	$C_9H_{17}N_5S$	2,98
4	Amidosulfuron	$C_9H_{15}N_5O_7S_2$	1,64
5	Amitraz	$C_{19}H_{23}N_3$	5,5
6	Amitrole	$C_2H_4N_4$	-0,87
7	Anilofos	$C_{13}H_{19}ClNO_3PS_2$	4,2
8	Azamethiphos	$C_9H_{10}ClN_2O_5PS$	0,82
9	Azinphos-ethyl	$C_{12}H_{16}N_3O_3PS_2$	3,4
10	Azinphos-methyl	$C_{10}PN_3H_{12}S_2O_3$	2,75
11	Benalaxyl	$C_{20}H_{23}NO_3$	3,88
12	Benazolin-ethyl	$C_{11}H_{10}ClNO_3S$	2,8
13	Bendiocarb	$C_{11}H_{13}NO_4$	1,86
14	Benfluralin	$C_{13}H_{16}F_3N_3O_4$	5,41
15	Benomyl	$C_{14}H_{18}N_4O_3$	2,12
16	Benoxacor	$C_{11}H_{11}Cl_2NO_2$	3,19
17	Bensulfuron-methyl	$C_{16}H_{18}N_4O_7S$	1,6
18	Bentazone	$C_{10}H_{12}N_2O_3S$	2,80
19	Bifenox	$C_{14}H_9Cl_2NO_5$	4,48
20	Biphenyl	$C_{12}H_{10}$	4,01
21	Bromopropylate	$C_{17}H_{16}Br_2O_3$	5,4
22	Bupirimate	$C_{13}H_{24}N_4O_3S$	2,7
23	Cadusafos	$C_{10}H_{23}O_2PS_2$	3,9
24	Captan	$C_9H_8Cl_3NO_2S$	2,80
25	Carbendazim	$C_9H_9N_3O_2$	1,52
26	Carbofuran	$C_{12}H_{15}NO_3$	2,32
27	Carbosulfan	$C_{20}H_{32}N_2O_3S$	6,05
28	Carboxin	$C_{12}H_{13}NO_2S$	3,003

29	Chlorbromuron	C <sub>9</sub> H <sub>10</sub> BrClN <sub>2</sub> O <sub>2</sub>	3,26
30	Chlorethoxyfos	C <sub>6</sub> H <sub>11</sub> Cl <sub>4</sub> O <sub>3</sub> PS	3,97
31	Chlorotoluron	C <sub>10</sub> H <sub>13</sub> ClN <sub>2</sub> O	2,46
32	Chlorpyrifos 398	C <sub>9</sub> H <sub>11</sub> Cl <sub>3</sub> NO <sub>3</sub> PS	4,77
33	Chlorpyrifos-methyl	C <sub>7</sub> H <sub>7</sub> Cl <sub>3</sub> NO <sub>3</sub> PS	4,31
34	Chlorthalidimethyl	C <sub>10</sub> H <sub>6</sub> Cl <sub>4</sub> O <sub>4</sub>	4,4
35	Chlozolate	C <sub>13</sub> H <sub>11</sub> Cl <sub>2</sub> NO <sub>5</sub>	3,11
36	Chorsulfuron	C <sub>12</sub> H <sub>12</sub> ClN <sub>5</sub> O <sub>4</sub> S	2,3
37	Cinosulfuron	C <sub>15</sub> H <sub>19</sub> N <sub>5</sub> O <sub>7</sub> S	1,4
38	Clofentezine	C <sub>14</sub> H <sub>8</sub> Cl <sub>2</sub> N <sub>4</sub>	3,27
39	Clomazone	C <sub>12</sub> H <sub>14</sub> ClNO <sub>2</sub>	2,17
40	Clomeprop	C <sub>16</sub> H <sub>15</sub> Cl <sub>2</sub> NO <sub>2</sub>	4,78
41	Cycloate	C <sub>11</sub> H <sub>21</sub> NOS	3,83
42	Cyromazine	C <sub>6</sub> H <sub>10</sub> N <sub>6</sub>	-0,06
43	DDT 399	C <sub>14</sub> H <sub>9</sub> Cl <sub>5</sub>	6,91
44	Demeton-S-methyl	C <sub>6</sub> H <sub>15</sub> O <sub>3</sub> PS <sub>2</sub>	1,32
45	Desmedipham	C <sub>16</sub> H <sub>16</sub> N <sub>2</sub> O <sub>4</sub>	3,53
46	Desmetryn	C <sub>8</sub> H <sub>15</sub> N <sub>5</sub> S	2,56
47	Diazinon	C <sub>12</sub> H <sub>21</sub> N <sub>2</sub> O <sub>3</sub> PS	3,81
48	Dicamba	C <sub>8</sub> H <sub>6</sub> Cl <sub>2</sub> O <sub>3</sub>	2,76
49	Dichlofluanid	C <sub>9</sub> H <sub>11</sub> Cl <sub>2</sub> FN <sub>2</sub> O <sub>2</sub> S <sub>2</sub>	3,45
50	Dichlorprop	C <sub>9</sub> H <sub>8</sub> Cl <sub>2</sub> O <sub>3</sub>	2,932
51	Dichlorvos	C <sub>4</sub> H <sub>7</sub> Cl <sub>2</sub> O <sub>4</sub> P	0,71
52	Diclofop-methyl	C <sub>16</sub> H <sub>14</sub> Cl <sub>2</sub> O <sub>4</sub>	4,24
53	Dicloran	C <sub>6</sub> H <sub>4</sub> Cl <sub>2</sub> N <sub>2</sub> O <sub>2</sub>	3,538
54	Dicofol	C <sub>14</sub> H <sub>9</sub> Cl <sub>5</sub> O	5,74
55	Diflubenzuron 243	C <sub>14</sub> H <sub>9</sub> ClF <sub>2</sub> N <sub>2</sub> O <sub>2</sub>	3,68
56	Diflufenican	C <sub>19</sub> H <sub>11</sub> F <sub>5</sub> N <sub>2</sub> O <sub>2</sub>	3,65
57	Dimefuron	C <sub>15</sub> H <sub>19</sub> ClN <sub>4</sub> O <sub>3</sub>	2,29
58	Dimethachlor	C <sub>13</sub> H <sub>18</sub> ClNO <sub>2</sub>	2,13
59	Dimethoate	C <sub>5</sub> H <sub>12</sub> NO <sub>3</sub> PS <sub>2</sub>	0,48
60	Dinitramine	C <sub>11</sub> H <sub>13</sub> F <sub>3</sub> N <sub>4</sub> O <sub>4</sub>	5,86

61	Disulfoton	$C_8H_{19}O_2PS_3$	4,02
62	Diuron	$C_9H_{10}Cl_2N_2O$	2,784
63	EPN	$C_{14}H_{14}NO_4PS$	4,78
64	EPTC	$C_9H_{19}NOS$	3,37
65	Edifenphos	$C_{14}H_{15}O_2PS_2$	3,48
66	Epoxiconazole	$C_{17}H_{13}ClFN_3O$	3,2
67	Etephon	$C_2H_6ClO_3P$	-0,22
68	Ethalfuralin	$C_{13}H_{14}F_3N_3O_4$	5,1
69	Ethiofencarb	$C_{11}H_{15}NO_2S$	2,04
70	Ethion	$C_9H_{22}O_4P_2S_4$	5,07
71	Ethofumesate	$C_{13}H_{18}O_5S$	2
72	Ethoprophos	$C_8H_{19}O_2PS_2$	3,59
73	Fenamiphos	$C_{13}H_{22}NO_3PS$	3,68
74	Fenarimol 365	$C_{17}H_{12}Cl_2N_2O$	3,22
75	Fenclorim	$C_{10}H_6Cl_2N_2$	4,02
76	Fenitrothion 359	$C_9H_{12}NO_5PS$	3,24
77	Fenoxycarb	$C_{17}H_{19}NO_4$	3,83
78	Fenthion	$C_{10}H_{15}O_3PS_2$	3,21
79	Flamprop-Mmethyl	$C_{17}H_{15}ClFNO_3$	3,2
80	Fluazinam	$C_{13}H_4Cl_2F_6N_4O_4$	8,19
81	Fluazuron	$C_{20}H_{10}Cl_2F_5N_3O_3$	5,41
82	Flumetralin	$C_{16}H_{12}ClF_4N_3O_4$	5,91
83	Fluometuron 191	$C_{10}H_{11}F_3N_2O$	2,36
84	Flurecol	$C_{14}H_{10}O_3$	3,581
85	Fluridone	$C_{19}H_{14}F_3NO$	1,87
86	Flurprimidol	$C_{15}H_{15}F_3N_2O_2$	2,37
87	Flutriafol 188	$C_{16}H_{13}F_2N_3O$	2,31
88	Fonofos	$C_{10}H_{15}OPS_2$	3,94
89	Formothion	$C_6H_{12}NO_4PS_2$	1,48
90	Fosthiazate	$C_9H_{18}NO_3PS_2$	0,94
91	Furalaxyl	$C_{17}H_{19}NO_4$	2,52
92	Furathiocarb	$C_{18}H_{26}N_2O_5S$	4,52

93	Glyphosate	$C_3H_8NO_5P$	-2,36
94	Haloxyfop etotyl	$C_{19}H_{19}ClF_3NO_5$	4,7
95	Hexaconazole	$C_{14}H_{17}Cl_2N_3O$	3,66
96	Hexazinone	$C_{12}H_{20}N_4O_2$	1,85
97	Hexythiazox	$C_{17}H_{21}ClN_2O_2S$	3,41
98	Imazail	$C_{14}H_{14}Cl_2N_2O$	3,58
99	Imibenconazole	$C_{17}H_{13}Cl_3N_4S$	5,8
100	Imidacloprid	$C_9H_{10}ClN_5O_2$	0,199
101	Iprodione	$C_{13}H_{13}Cl_2N_3O_3$	2,8
102	Isazofos	$C_9H_{17}ClN_3O_3PS$	2,99
103	Isofenphos	$C_{15}H_{24}NO_4PS$	4,12
104	Isoproturon	$C_{12}H_{18}N_2O$	2,32
105	Isoxaben	$C_{18}H_{24}N_2O_4$	3,25
106	Lenacil	$C_{13}H_{18}N_2O_2$	2,862
107	Linuron	$C_9H_{10}Cl_2N_2O_2$	3,2
108	Maleic hydrazide 604	$C_4H_4N_2O_2$	-0,84
109	Mecoprop 142	$C_{10}H_{11}ClO_3$	2,835
110	Mefenacet	$C_{16}H_{14}N_2O_2S$	3,85
111	Mepronil	$C_{17}H_{19}NO_2$	4,12
112	Metalaxyl 67	$C_{15}H_{21}NO_4$	2,15
113	Metamitron	$C_{10}H_{10}N_4O$	0,83
114	Metazachlor	$C_{14}H_{16}ClN_3O$	2,11
115	Methamidophos	$C_2H_8NO_2PS$	-0,82
116	Methidathion	$C_6H_{11}N_2O_4PS_3$	2,03
117	Methiocarb	$C_{11}H_{15}NO_2S$	2,88
118	Methomyl	$C_5H_{10}N_2O_2S$	1,6
119	Metobromuron	$C_9H_{11}O_2N_2Br$	2,4
120	Metolachlor (ANSIWSSA)	$C_{15}H_{22}ClNO_2$	3
121	Metosulam	$C_{14}H_{13}Cl_2N_5O_4S$	3,08
122	Metoxuron	$C_{10}H_{13}ClN_2O_2$	1,8
123	Metribuzin	$C_8H_{14}N_4OS$	1,7
124	Metsulfuronmethyl	$C_{14}H_{15}N_5O_6S$	2.20

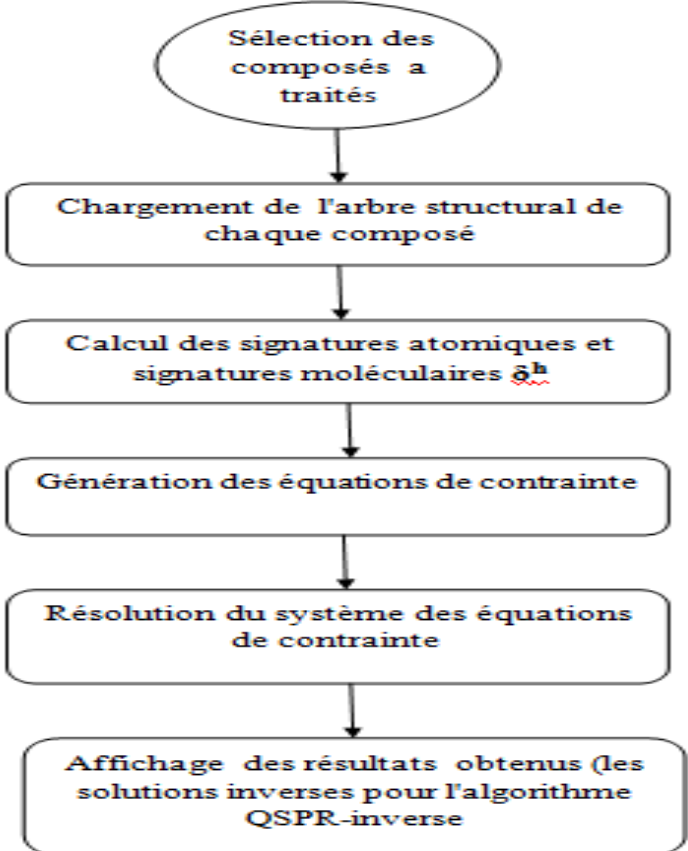
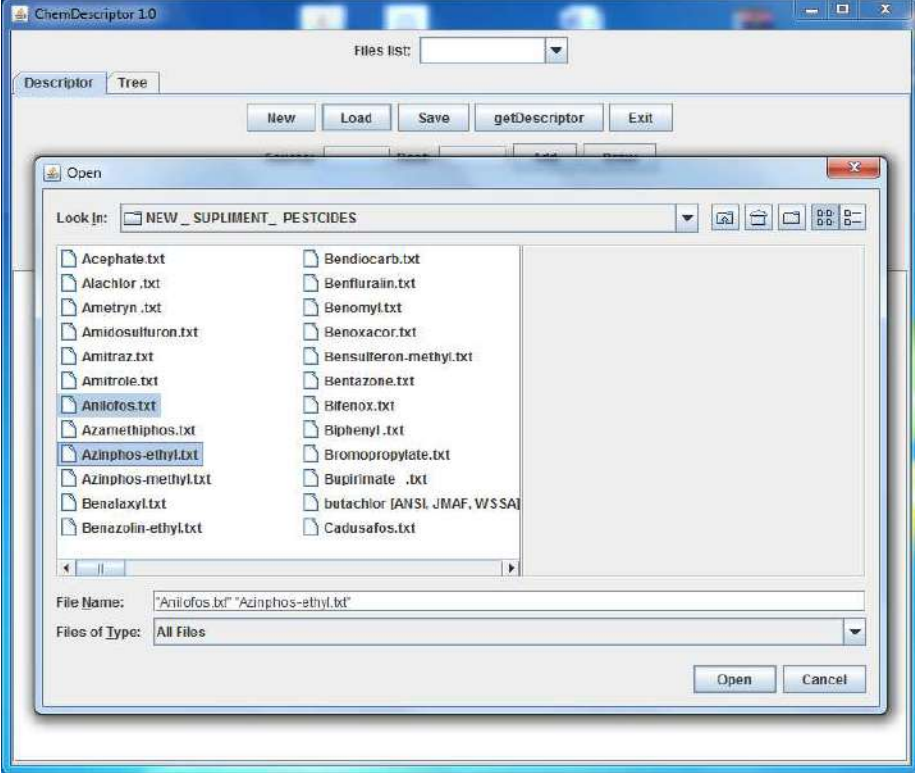
125	Monolinuron	$C_9H_{11}ClN_2O_2$	2,3
126	Nicosulfuron	$C_{15}H_{18}N_6O_6S$	0,6
127	Novaluron	$C_{17}H_9ClF_8N_2O_4$	6,78
128	Ofurace	$C_{14}H_{16}ClNO_3$	0,93
129	Omethoate	$C_5H_{12}NO_4PS$	-0,74
130	Oryzalin	$C_{12}H_{18}N_4O_6S$	3,38
131	Oxadiazon	$C_{15}H_{18}Cl_2N_2O_3$	4,55
132	Oxadixyl	$C_{14}H_{18}N_2O_4$	0,68
133	Oxamyl	$C_7H_{13}N_3O_3S$	-0,47
134	Oxycarboxin	$C_{12}H_{13}NO_4S$	1,14
135	Oxydemetonmethyl	$C_6H_{15}O_4PS_2$	-0,74
136	Parathion 396	$C_{10}H_{14}NO_5PS$	3,84
137	Parathion methyl	$C_8H_{10}NO_5PS$	2,78
138	Pentachloronitrobenzene 446	$C_6H_5NO_2$	4,16
139	Pentachlorophenol 363	$C_6Cl_5OH$	4,78
140	Phorate	$C_7H_{17}O_2PS_3$	3,67
141	Phosalone	$C_{12}H_{15}ClNO_4PS_2$	4,28
142	Phosmet	$C_{11}H_{12}NO_4PS_2$	2,84
143	Piperophos	$C_{14}H_{28}NO_3PS_2$	4,2
144	Pirimicarb	$C_{11}H_{18}N_4O_2$	1,7
145	Pirimiphos-ethyl	$C_{13}H_{24}N_3O_3PS$	5,06
146	Pirimiphos-methyl	$C_{11}H_{20}N_3O_3PS$	4
147	Pretilachlor	$C_{17}H_{26}ClNO_2$	4,25
148	Prochloraz	$C_{15}H_{16}Cl_3N_3O_2$	3,8
149	Profenofos	$C_{11}H_{15}BrClO_3PS$	4,6
150	Prometon	$C_{10}H_{19}N_5O$	2,91
151	Prometryn	$C_{10}H_{19}N_5S$	3.436
152	Propachlor 822	$C_{11}H_{14}ClNO$	2,28
153	Propazine 222	$C_9H_{16}ClN_5$	2,98
154	Propisochlor	$C_{15}H_{22}ClNO_2$	3,27
155	Propoxur 96	$C_{11}H_{15}NO_3$	1,6

156	Prosulfocarb	C <sub>14</sub> H <sub>21</sub> NOS	3,99
157	Prothiofos	C <sub>14</sub> H <sub>21</sub> NOS	5,53
158	Pyrazophos	C <sub>14</sub> H <sub>20</sub> N <sub>3</sub> O <sub>5</sub> PS	3,01
159	Pyrifenox	C <sub>14</sub> H <sub>12</sub> Cl <sub>2</sub> N <sub>2</sub> O	3,42
160	Rimsulfuron	C <sub>14</sub> H <sub>17</sub> N <sub>5</sub> O <sub>7</sub> S <sub>2</sub>	2,22
161	Simazine	C <sub>7</sub> H <sub>12</sub> ClN <sub>5</sub>	2.284
162	Sulfotep	C <sub>8</sub> H <sub>20</sub> O <sub>5</sub> P <sub>2</sub> S <sub>2</sub>	3,99
163	Sulprofos	C <sub>12</sub> H <sub>19</sub> O <sub>2</sub> PS <sub>3</sub>	4,64
164	TRIBUFOS	C <sub>12</sub> H <sub>27</sub> OPS <sub>3</sub>	3,23
165	Tebuconazole	C <sub>16</sub> H <sub>22</sub> ClN <sub>3</sub> O	3,58
166	Tebufenozide	C <sub>22</sub> H <sub>28</sub> N <sub>2</sub> O <sub>2</sub>	4,24
167	Tebuthiuron	C <sub>9</sub> H <sub>16</sub> N <sub>4</sub> OS	1,79
168	Tecnazene	C <sub>6</sub> HCl <sub>4</sub> NO <sub>2</sub>	3,73
169	Temephos	C <sub>16</sub> H <sub>20</sub> O <sub>6</sub> P <sub>2</sub> S <sub>3</sub>	5,96
170	Terbufos	C <sub>9</sub> H <sub>21</sub> O <sub>2</sub> PS <sub>3</sub>	4,37
171	Terbumeton	C <sub>10</sub> H <sub>19</sub> N <sub>5</sub> O	2,91
172	Terbuthylazine	C <sub>9</sub> H <sub>16</sub> ClN <sub>5</sub>	2,98
173	Terbutryn 210	C <sub>10</sub> H <sub>19</sub> N <sub>5</sub> S	3,44
174	Tetraconazole	C <sub>13</sub> H <sub>11</sub> Cl <sub>2</sub> F <sub>4</sub> N <sub>3</sub> O	3,19
175	Tetradifon	C <sub>12</sub> H <sub>6</sub> Cl <sub>4</sub> O <sub>2</sub> S	5,52
176	Thifensulfuronmethyl	C <sub>12</sub> H <sub>13</sub> N <sub>5</sub> O <sub>6</sub> S <sub>2</sub>	1,52
177	Thiodicarb	C <sub>10</sub> H <sub>18</sub> N <sub>4</sub> O <sub>4</sub> S <sub>3</sub>	1,52
178	Thiometon	C <sub>6</sub> H <sub>15</sub> O <sub>2</sub> PS <sub>3</sub>	2,25
179	Thiram	C <sub>6</sub> H <sub>12</sub> N <sub>2</sub> S <sub>4</sub>	1,75
180	Tolyfluanid	C <sub>10</sub> H <sub>13</sub> Cl <sub>2</sub> FN <sub>2</sub> O <sub>2</sub> S <sub>2</sub>	3,91
181	Tri-allate	C <sub>10</sub> H <sub>16</sub> Cl <sub>3</sub> NOS	6,18
182	Triadimefon	C <sub>14</sub> H <sub>16</sub> ClN <sub>3</sub> O <sub>2</sub>	2,77
183	Triazamate	C <sub>13</sub> H <sub>22</sub> N <sub>4</sub> O <sub>3</sub> S	1,93
184	Triazophos	C <sub>12</sub> H <sub>16</sub> N <sub>3</sub> O <sub>3</sub> PS	3,34
185	Tribenuron methyl	C <sub>15</sub> H <sub>17</sub> N <sub>5</sub> O <sub>6</sub> S	0,78
186	Trichlorfon	C <sub>4</sub> H <sub>8</sub> Cl <sub>3</sub> O <sub>4</sub> P	0,48
187	Triflumuron	C <sub>15</sub> H <sub>10</sub> ClF <sub>3</sub> N <sub>2</sub> O <sub>3</sub>	4,55



188	Vinclozolin	$C_{12}H_9Cl_2NO_3$	3,19
189	butachlor [ANSI, JMAF, WSSA]	$C_{17}H_{26}ClNO_2$	4,51
190	trifluralin	$C_{13}H_{16}F_3N_3O_4$	5,412

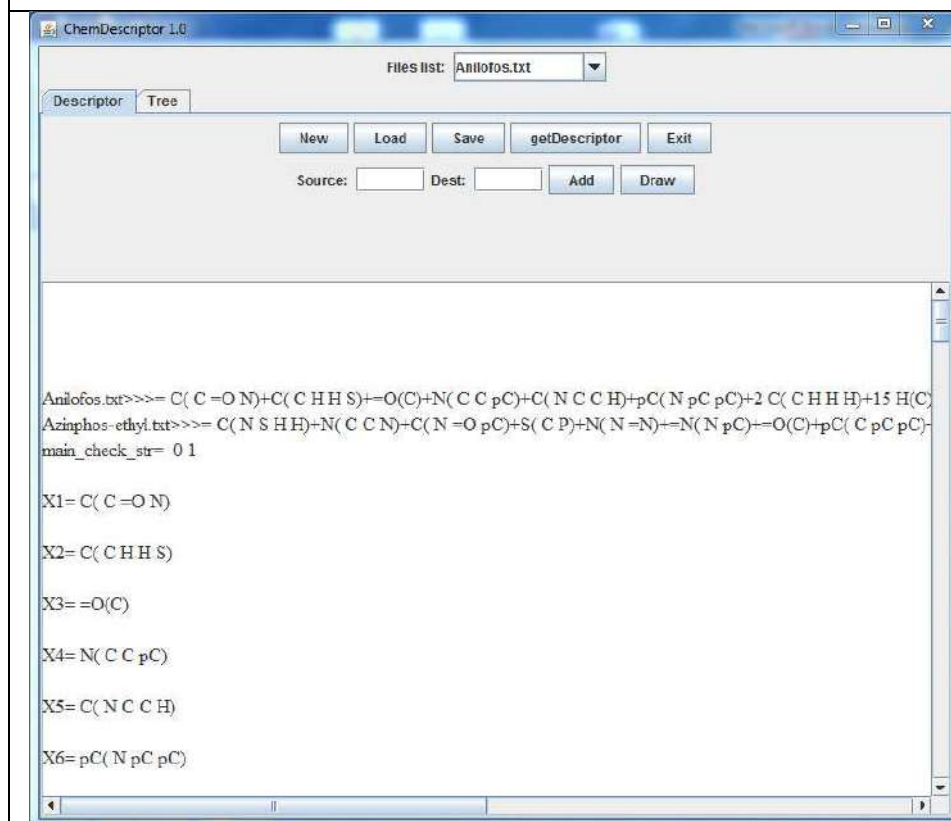
## Annexes 3

Figure	Explication
 <pre> graph TD     A([Sélection des composés à traités]) --&gt; B[Chargement de l'arbre structural de chaque composé]     B --&gt; C[Calcul des signatures atomiques et signatures moléculaires <math>\alpha^h</math>]     C --&gt; D[Génération des équations de contrainte]     D --&gt; E[Résolution du système des équations de contrainte]     E --&gt; F[Affichage des résultats obtenus (les solutions inverses pour l'algorithme QSPR-inverse)]           </pre> <p>The flowchart illustrates the QSPR-inverse algorithm process. It begins with the selection of compounds to be treated, followed by loading the structural tree for each compound. The next steps are calculating atomic and molecular signatures (<math>\alpha^h</math>), generating constraint equations, resolving the system of constraint equations, and finally displaying the results, which are the inverse solutions for the QSPR-inverse algorithm.</p>	<p>Algorithme des étapes de QSPR – inverse pour la génération des nouvelles structures chimiques.</p>
 <p>The screenshot shows the ChemDescriptor 1.0 interface. An 'Open' dialog box is displayed, showing a list of files in the directory 'NEW_SUPLIMENT_PESTICIDES'. The files listed include various pesticide-related text files such as Acephate.txt, Alachlor.txt, Ametryn.txt, Amidosulfuron.txt, Amitraz.txt, Amitrole.txt, Anilofos.txt, Azamethiphos.txt, Azinphos-ethyl.txt, Azinphos-methyl.txt, Benalaxyl.txt, Benazofin-ethyl.txt, Bendiocarb.txt, Benflurain.txt, Benomyl.txt, Benoxacor.txt, Bensulfuron-methyl.txt, Bentazone.txt, Bifenox.txt, Biphenyl.txt, Bromopropylate.txt, Bupirimate.txt, butachlor [ANSI, JMAF, WSSAI], and Cadusafos.txt. The 'File Name' field contains 'Anilofos.txt *Azinphos-ethyl.txt' and the 'Files of Type' is set to 'All Files'.</p>	<p>Cette figure représente la sélection des composés à traités</p>



## Figure

## Explication



Cette figure représente les signatures atomiques et signatures moléculaires  $\delta^h$  calculés des composés a traités

C:\PROGRA~1\XINOXS~1\CREAT~1\GE2001.exe

```
Equation<11>= -1X2+1X9-1X18 check_vector[i]= 0
Equation<12>= +2X6+2X14+2X15+2X23+2X24 check_vector[i]= 2
Equation<13>= -1X1+1X3-1X20 check_vector[i]= 0
Equation<14>= +1X1+1X2+2X5+1X7+1X25 check_vector[i]= 2
Equation<15>= -1X1+2X4-1X5-1X18+2X19-1X20 check_vector[i]= 0
Equation<16>= -2X2-1X5-3X7+1X8-3X13-2X18-2X25 check_vector[i]= 0
Equation<17>= X1+X2+X7+X14+X15+X21+X24 check_vector[i]= 2
Equation<18>= X1+2X2-1X3+X4+2X5+X6+2X7-1X8+X9+2X10-1X11+X12+2X13+X14+X15-1X16-1X17+2X18+X19+X20+X21+X22+X23+X24+2X25 check_vector[i]= 2
```

Cette figure représente les équations decontrainte générées des composés a traités