

République Algérienne Démocratique et Populaire

**Ministère de l'Enseignement Supérieur et de la
Recherche Scientifique**

CENTRE UNIVERSITAIRE D'EL-OUED

INSTITUT DES SCIENCES ET TECHNOLOGIE

Mémoire de fin d'étude

Présenté pour l'obtention du diplôme de

LICENCE ACADEMIQUE

Domaine : Sciences et technologie

Filière : hydrocarbure et chimie

Spécialité : raffinage

Thème

**Contribution a la prédiction de coefficient
de partage octanol/eau par la technique
QSPR en utilisant les descripteurs
moléculaires**

Présenté par :

GHOUAR Ziad.

GHOUAR Larbi.

Encadrer par :

Mr. SOUYEI Belgacem

2010-2011



Remerciements

En premier lieu, nous tenons à remercier notre DIEU, notre Créateur pour m'avoir donné la force pour accomplir ce travail.

Nous tenons à notifier un remerciement spécial à tous nos professeurs .

Nous tenons à remercier Mr. SOUYEI BELGACEM notre promoteur pour son grand soutien et ses conseils considérables.

Nous remercions également tous les professeurs de CENTRE UNIVERSITAIRE D'EL-OUED et surtout les professeurs de département de génie des procédés.

Qu'ils trouvent ici toute notre gratitude et toute notre reconnaissance pour leur aide, leur encadrement dont j'ai pu bénéficier durant toute notre scolarité.

Nous désirons aussi, exprimer nos essentiels remerciements :

Mr. AHMADI RIDHA , Mr. LAOUINI Salaheddine , tous les professeurs de raffinage .

Enfin nos remerciements vont à tous ceux qui ont contribué de près ou de loin pour l'aboutissement de ce travail.



DEDICACE

Je dédie ce modeste travail à :

- En premier lieu à ma mère et mon père qui ont consenti beaucoup de sacrifices pour me permettre de réaliser mes objectifs Qu'ils trouvent ici toute ma reconnaissance et ma gratitude.

- A mes grandes mères.

- A mon frère et mes sœurs.

- A tous mes amis, particulièrement : larbi, khaled, Brahim

- A tous mes camarades des groupes RAFFINAGE.

- A tous mes enseignants depuis le primaire jusqu'à maintenant.

Et à tous ceux qui m'ont aidé de près ou de loin à accomplir ce travail.

ZIAD.





DEDICACE

*Je dédie ce modeste travail aux deux êtres qui me sont
très chers dans cette vie, à savoir mon père et ma mère. Je leur
dis merci papa du fond de mon cœur pour ton éducation, ton
sacrifice, ton assistance et pour ce tu m'as fait et qui m'a
permis d'avoir cette réussite et ce bonheur.*

*Je te dis merci et mille fois merci maman pour ta patience, ton
courage et ton sacrifice pour moi A tous mes proches Surtout
ma famille, mes amis d'enfance, sans oublier toutes mes
connaissances durant ma vie*

Larbi.

SOMMAIRE

Introduction générale	1
------------------------------	----------

Chapitre I : Définition et détermination expérimentale de Kow

I-1-Définition Coefficient de partage octanol-water log Kow	3
I-1-1- Définition générale	3
I-1-2- Coefficient de partage log P	3
I-1-3- Définition Coefficient de partage octanol-water log Kow	4
I-1-4- La relation de calcul	5
I-2-Quelques exemples de log Kow à 25°C	7
I-3-Relations LOG KOW avec d'autres propriétés physico-chimiques	7
I-4- Détermination expérimentale de Kow	8
I-4-1-méthode du brassage lent	8
I-4-2-Méthode par agitation du flacon	8
I-4-3-Méthode par HPLC	8
I-4-4-Méthode sur colonne	9
I-5-Les méthodes (QSAR/QSPR)	10
I-5-1- Définition de QSAR / QSPR	10
I-5-2- Méthode quantitative structure-propriété relation (QSPR)	10
I-6-Les descripteurs moléculaires	12
I-6-1-La sélection des descripteurs moléculaire	13
I-6-2- Mapping (plan)	13

Chapitre II : La Régression Simple et Multiple

II-1-La régression Linéaire Simple (RLS)	15
II-2 Régression Linéaire Multiple (RLM)	17
II-2-1 Principes de la Régression Linéaire Multiple	17
II-3 Valider un modèle	18

Chapitre III : Diagramme de travail et la base des données

III-1- Les étapes de travail	21
III-2- base de données	22
III-3- Génération des descripteurs moléculaires	22
III-3-1- Descripteurs constitutionnels	23
III-3-2- Descripteurs topologiques	24
III-3-3- Propriétés moléculaires	24
III-3-4- fragments de l'atome centrée	24
III-3-5- comptes de groupe fonctionnels	25
III-4- Organigramme	25
III-4-1- La préparation la base de données	26

III-4-2- Logiciel DRAGON	27
III-4-2-1 – définition Logiciel DRAGON	27
III-4-2-2- Travail par Logiciel DRAGON	28
III-4-3- Logiciel XLSTAT	29
III-4-3-1- définition Logiciel XLSTAT	29
III-4-3-2- Travail par Logiciel XLSTAT	29
III-5-Déterminer les éléments au-delà du domaine de confiance	30

Chapitre IV : Résultats et discussion

IV-1-le résultat de la bloc comptes de groupe fonctionnels	31
IV-1-1-Equation du modèle	31
IV-1-2-Coefficients d'ajustement	31
IV-1-3-Les paramètres du modèle	32
IV-1-4-Analyse des résultats	33
IV-1-5- discussion	34
IV-2-le résultat de la bloc Descripteurs topologiques	34
IV-2-1-Equation du modèle	34
IV-2-2- Les paramètres du modèle	34
IV-2-3- Coefficients d'ajustement	35
IV-2-4-Analyse des résultats	36
IV-2-5- discussion	36
IV-3-le résultat de la bloc fragments de l' atome centrale	36
IV-3-1-Equation du modèle	36
IV-3-2-Les paramètres du modèle	37
IV-3-3-Coefficients d'ajustement	37
IV-3-4-Analyse des résultats	39
IV-3-5- discussion	39
IV-4-Propriétés moléculaires	39
IV-4-1-Equation du modèle	39
IV-4-2-Les paramètres du modèle	39
IV-4-3- Coefficients d'ajustement	40
IV-4-4-Analyse des résultats	41
IV-4-5- discussion	41
IV-5-Descripteurs constitutionnels	41
IV-5-1- Equation du modèle	41
IV-5-2- Les paramètres du modèle	42
IV-5-3-Coefficients d'ajustement	42
IV-5-4-Analyse des résultats	43
IV-5-5- discussion	43
IV-6-Analyse globale	44
IV-7-Discussion générale	44
Conclusion générale	46
Références bibliographique	47

Liste des Figures

Figure	Titre	Page
Figure I-1	l'équilibre de coefficient de partage Kow	5
Figure I-2	Relations LOG K _{OW} avec d'autres propriétés physico-chimiques	7
Figure I-3	principe du méthode par agitation du flacon	8
Figure I-4	principe de fonctionnement de HPLC	9
Figure I-5	Montre le principe du méthode sur colonne	9
Figure I-6	Outil de prédiction des propriétés LOG K _{OW}	12
Figure II-1	Montre la meilleure droite de régression linéaire Simple	15
Figure II-2	Montre la réalisation de valeur ϵ_i et la droite de régression	16
Figure III-1	Les étapes de travail (Diagramme de travail)	21
Figure III-2	relation descripteurs moléculaires Enter QSPR et log kow	23
Figure III-3	décrit les étapes de la préparation de la base de données	26
Figure III-4	résulta graphique de XLSTAT (regression)	30

Liste des Tableaux

Tableaux	Titre	Page
Tableau I-1	Valeurs des Coefficients de partage octanol/eau et solubilités pour quelques produits chimiques organiques synthétiques et des produits non-organiques.	7
Tableau III-1	Montre les résultats de calcul des descripteurs	25

Nomenclature

Symboles

Définitions

log P	coefficient de partage
QSPR	méthodes quantitatives relation structure-activité
QSPR	méthode quantitative relation structure-propriété
log K_{ow}	coefficient de partage octanol-water
OR	octanol-riches
WR	water –riches
PEC	Concentration Environnementale Prévisible.
PNEC	Concentration Prévisible sans Effet
GC	contribution du groupe
C_i^o	concentrations de soluté <i>i</i> dans la phase octanol
C_i^w	concentrations de soluté <i>i</i> dans la phase aqueuse
$\gamma_i^{w,\infty}$	l'activité infinie de dilution coefficients dans les phases de l'eau
$\gamma_i^{o,\infty}$	l'activité infinie de dilution coefficients dans les phases de l'octanol
K_{oc}	le coefficient de partage le sédiment/eau carbone-normalisés organiques
γ_p	la propriété cible
GA	l'algorithme génétique
BMLR	best molecular liner régresse
P.C.A	principal component
P.L.S	partical leost squar
NNS	neural networks
RLS	régression linéaire simple
MLS	modèle linéaire simple
TPSAL	topologique surface polaires

MLogP	partition moriguchi octanol-eau
ALogP	partition Ghose-Crippen-Viswanadhan octanol-eau
AMR	réfractivité molaire
LAI	l'indice d'alerte Lipinski
CAS	chemical abstract services
SMILLE	simplified molecular input line entry specification

Résumé :

L'étude de la relation quantitative structure-propriété (QSPR) est réalisée pour établir des corrélations qui relient les structures moléculaires des composés organiques à leur coefficients de partage Octanol-eau, **K_{ow}** , en utilisant les descripteurs moléculaires .

Les corrélations sont simples en application avec une précision ,qui fournissent une manière simple, directe et relativement précise pour calculer **K_{ow}** .Ces calcul nous donnent un modèle qui nous amène aux résultats remarquables en corrélation avec les descripteurs des blocs fragments de l'atome centré et groupes fonctionnels respectivement. ($R^2 = 0.941$) ($R^2 = 0.905$).

Introduction générale :

Aujourd'hui, l'industrie chimique joue un rôle majeur et divers dans notre vie, ce rôle se diffère suivant la nature des composés chimiques utilisés. Il existe plusieurs variétés de composés chimiques classifiés comme produits dangereux (Toxiques), qui ont un impact sérieux sur la santé de l'homme et de l'environnement. Parmi ces composés, on cite: les organochlorés comme les PCB (polychlorobiphenyls) sont des composés aromatiques possédant une ossature biphenyle, substituée par des atomes de chlore en position et nombre variables [1], le polychlorure de vinyle (PVC), les hydrocarbures aromatiques polycycliques (PAHs) sont souvent résistants à la dégradation biologique et ne sont pas efficacement éliminés par des techniques classiques de séparation [2]. Les pesticides comme le DDT (dichlorodiphényltrichloroéthane), les éthers de glycol et les organophosphorés [3].

Pour l'évaluation du risque environnemental, la législation européenne sur les produits chimiques (substances existantes, substances nouvelles, biocides) impose une évaluation du risque chimique afin d'assurer la protection de la santé humaine et celle de l'environnement.

La caractérisation du risque est basée sur la valeur du rapport PEC/PNEC, Tel que:

PEC : Concentration Environnementale Prévisible.

PNEC : Concentration Prévisible sans Effet [4].

D'après tout cela, on comprend qu'il y a une nécessité urgente et sérieuse qui demande et exige le traitement avec une grande prudence et exactitude de ces composés Chimiques. Cela ne peut se faire qu'après avoir élaboré une manière pour l'étude et le calcul avec précision des propriétés physico-chimiques qui ont une influence sur L'environnement.

Le but de ce travail, est centré sur le développement d'un modèle mathématique basé sur les données expérimentales disponibles et en fonction de la structure moléculaire, permettant ainsi d'estimer, d'étudier et d'identifier la propriété physico-chimiques notamment le coefficient de partage octanol/eau (log_{K_{ow}}), qui est un paramètre clé pour l'évaluation de l'impact des composés chimiques sur l'environnement, sans recours aux activités expérimentales.

Pour la réalisation de notre travail, nous avons construits une base de données qui a été arrange a partir des variétés de familles chimiques et biochimiques, suivi du choix des blocs des descripteurs moléculaires qui sont (Descripteurs constitutionnels, Descripteurs topologiques, Indices de connectivité, Indices de charge topologiques, Descripteurs

Propriétés moléculaires , Descripteurs du groupe fonctionnel , Descripteurs des fragments d'atome centré). Ces blocs et cette base de donnée subits a un calcul statistique de la régression linéaire multiple par la méthode « stepwise » a l' aide du logiciel XLSTAT. Et finalement les résultats du travail

I- Définition et détermination expérimentale de K_{ow}

Introduction:

L'importance de la bioconcentration dans l'évaluation de danger pour l'environnement [5] et l'utilité du paramètre hydrophobe dans sa prédiction [6], mène à un intérêt immense pour la mesure et la prévision de $\log K_{ow}$ à partir de la structure moléculaire [7]. La persistance est également une préoccupation profonde dans l'évaluation du danger pour l'environnement des produits chimiques, est une voie importante par laquelle on les élimine à travers la biodégradation [8].

I-1-Définition de Coefficient de partage octanol-water ($\log K_{ow}$):

I-1-1- Définition générale :

Pendant de nombreuses années, la partition solvant-eau ($\log P$) a été utilisée comme une mesure de la lipophilie. Un travail de pionnier par Hansch et Leo a conduit à l'utilisation de $\log P$ en ce qui concerne les QSAR (méthodes quantitatives relation structure-activité), QSPR (Méthode quantitative relation structure-propriété) [9] comme une description générale de la perméabilité cellulaire.

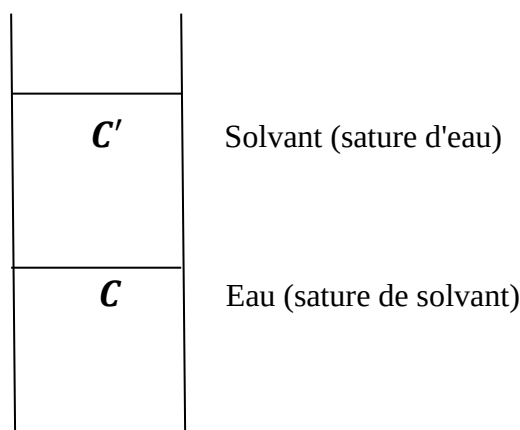
Plusieurs références ont été faites pour ambiguïté (nébuliseur) dans le système de classification le grand nombre de types d'atomes (et donc des paramètres réglables), irréaliste les valeurs de certaines contributions atome, un constat d'échec au prévision et les préjugés envers la sous-estimation de $\log P$. Nous présente une nouvelle définition des types de $\log P$ et contributions conçu en partie pour répondre à ces plaintes. Un Descripteur comptables communes pour la taille moléculaire et polarisabilité, [10] est réalisé de la même manière.

I-1-2- Coefficient de partage $\log P$:

On considère le partage d'un soluté entre l'eau et un solvant non miscible (hexane, éther, chloroforme, octanol...)

Le coefficient de partage P est égal au rapport des concentrations du soluté dans les deux phases :

$$P = \frac{C'}{C} \quad (I - 1)$$



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

$C' < C \Rightarrow P < 1 \Rightarrow \log P < 0$: le soluté est dit hydrophile [11]

I-1-3- Définition de Coefficient de partage octanol-water log Kow:

Le coefficient de partage octanol/eau (**Kow**) est le paramètre physico-chimique le plus facile à mesurer au laboratoire et utilisé pour caractériser quantitativement la nature hydrophobe des composées organiques. Il est l'indicateur majeur concernant le partage entre l'eau et la matière organique.

A température et pression ambiantes, l'eau et 1-octanol sont partiellement miscibles, formant un mélange octanol-riches phase (72,5% en mole 1-octanol), désigné comme ici est **OR**, et une phase riche en eau (99,99% d'eau mol), indiqué par le **WR** exposant. Le coefficient partage octanol-eau d'un soluté **i**, **Kow**, est défini comme l'équilibre rapport des concentrations de ce soluté entre les deux phases liquides quand une très petite quantité est ajoutée à une eau équilibrée et le mélange de 1-octanol, qui est

$$K_{ow,i} = \frac{C_i^{OR}}{C_i^{WR}} \quad (I - 2)$$

OR : octanol-riches

WR : water -riches

Ce coefficient est une mesure de l'hydrophobie de soluté. Un produit chimique hydrophobe préférentiellement partitions dans la phase octanol-riche, résultant dans une grande valeur de **K_{ow}**, tout un produit chimique hydrophile aura une faible valeur **K_{ow}**. Par conséquent, ce paramètre est utilisé pour décrire l'interaction hydrophobe dans la bien connue relations quantitatives structure-activité (**QSAR**), [12,13] qui a été largement utilisé pour prédire les activités biologique des composés bioactifs et la toxicité des produits chimiques[14, 15] et pour concevoir de nouveaux médicaments.[16, 17] En outre, 1-octanol est un bon substitut pour les lipides dans les milieux aquatiques biote des animaux, et la matière organique dans les sols et sédiments en raison de la similitude chimique entre 1-octanol et la plupart des lipides; tell que les deux ont une longue chaîne hydrocarbonée hydrophobe et une extrémité groupe hydrophile. Sur la base de cela, Mackay [18] et co-workers [19] mis au point un modèle de fugacité de prédire la distribution de longue durée de vie des produits chimiques organiques dans l'environnement à l'aide des informations sur **K_{ow}**. En outre, les valeurs de **K_{ow}** ont été trouves de corrélér la phénomène de bioaccumulation dans la chaîne alimentaire. Par exemple, Thomann [20] utilisé un simple modèle de l'état d'équilibre et a constaté que la bioamplification (ou accumulation dans la chaîne alimentaire) se produit lorsque **K_{ow}** est supérieure à environ 10 [21].

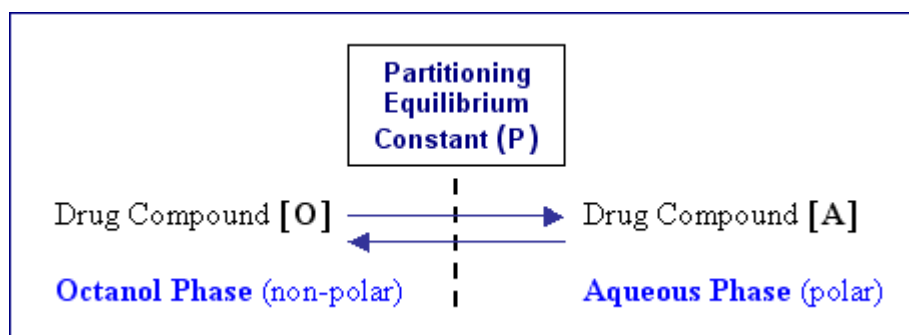


Figure I-1: l'équilibre de coefficient de partage **K_{ow}**

I-1-4- La relation de calcule :

Notre objectif dans ce travail a été d'appliquer et d'enquêter sur l'applicabilité des modèles existants basés sur les quasis chimiques la théorie et la contribution du groupe (**GC**) concept pour les solutions aqueuses à dilution infinie. En particulier, nous sommes intéressés dans le calcul de coefficient de partage octanol- eau, a partir des valeurs expérimentales disponibles dans la littérature. **K_{ow}** est défini comme :

$$k_{ow} = \frac{C_i^o}{C_i^w} \quad (I-3)$$

où C_i^o et C_i^w sont les concentrations de soluté i dans la phase octanol (o) et la phase aqueuse (w), respectivement. L'unité de la concentration est mol / L ou mol/cm³. Les variables de concentrations peut être aisément transformées à cette coefficient d'activité de sorte que l'application d'un modèle thermodynamique est possible pour l'estimation de **Kow**. L'équation I-3 peut donc être réécrite comme [22]

$$k_{ow} = 0.151 \left(\frac{\gamma_i^{w,\infty}}{\gamma_i^{o,\infty}} \right) \quad (I-4)$$

où $\gamma_i^{w,\infty}$ et $\gamma_i^{o,\infty}$ sont les coefficients d'activité de dilution infinie dans les phases de l'eau et l'octanol, respectivement. Le coefficient de l'équation I-4 (0,151) et le calcul des coefficients d'activité dans la phase octanol ont été menées dans ce travail à 25 ° C. Nous avons supposé que la solubilité de l'eau dans l'octanol est de 27,5 mol %, comme déclarés par Apelblat. La phase aqueuse est supposée être de l'eau pure raison de la solubilité négligeable d'octanol dans l'eau. Plusieurs auteurs ont étudié les solubilités de l'eau dans le 1-octanol, et les valeurs ne sont en accord les uns avec les autres modérés, allant de 20,7 à 27,5 mol %. La majorité des valeurs sont proches à 27,5 mol %. La pureté de 1-octanol sur lequel Apelblat [23] base de sa mesure est supérieure à 99%.

Le logarithme du coefficient de partage octanol-eau **log Kow** est la valeur le plus souvent rapportée dans la littérature. **log Kow** est utilisé dans le calcul des facteurs de bioaccumulation, pour estimer le coefficient de partage le sédiment/eau carbone-normalisés organiques (**Kco**) et pour calculer la solubilité dans l'eau et d'autres propriétés en utilisant des corrélations de réactivité de propriété. [24]

I-2-Quelques exemples de log K_{ow} à 25°C :

Tableau I-1: Valeurs des Coefficients de partage octanol/eau et solubilités pour quelques produits chimiques organiques synthétiques et des produits non-organiques. [25]

Composé (mg/L)	Solubilité à 25°C, dans l'eau	Coefficient de partition K_{ow} log octanol/ /Eau
Benzidine	400	1.81
Toluene	535	2.69
Cblorobenzene	472	2.84
N-nitrosodimethylamine	Miscible	0.06
Acrylonitrile	73500	-0.14

I-3-Relation de K_{ow} avec d'autres propriétés physico-chimiques :

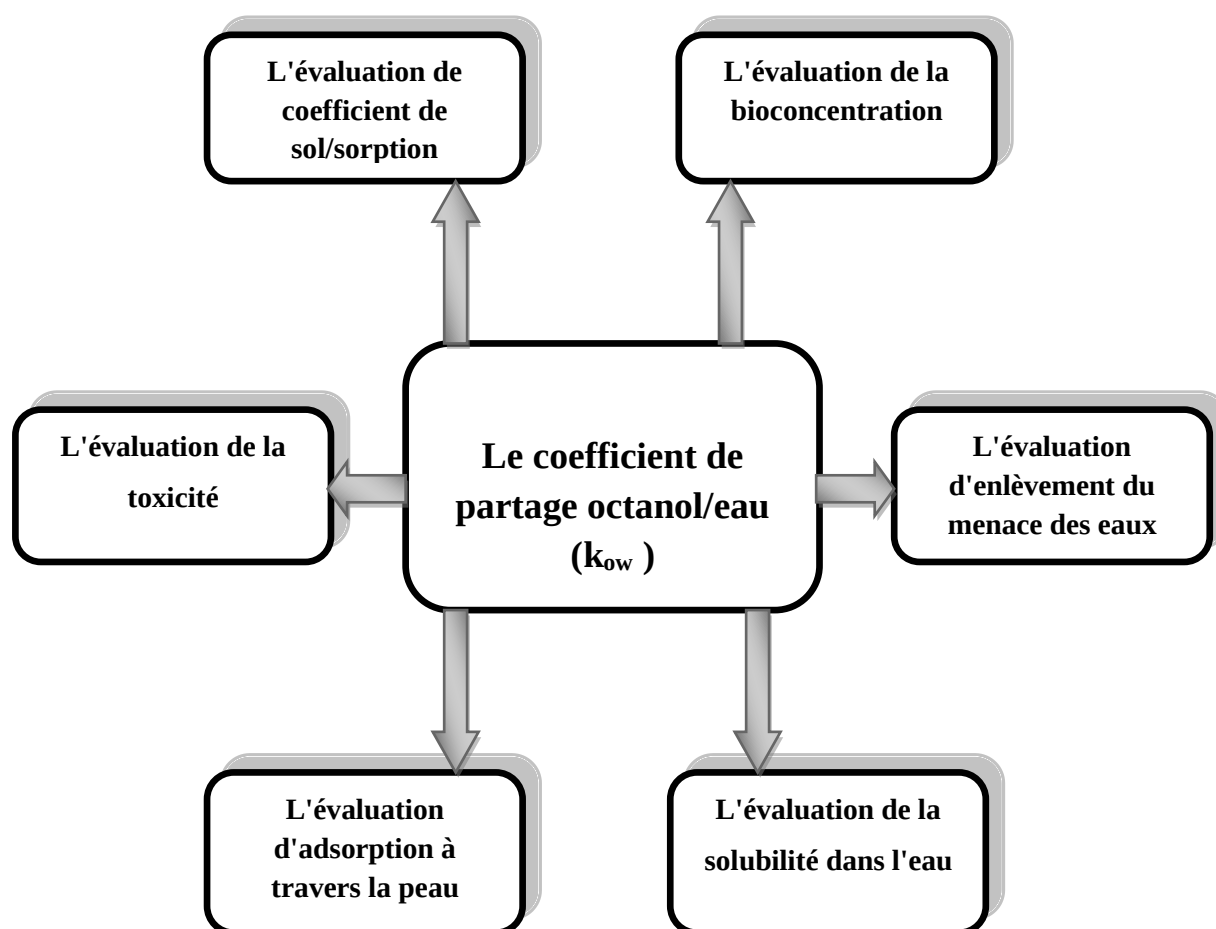


Figure I-2: Relation de K_{ow} avec d'autres propriétés physico-chimiques

I-4- Détermination expérimentale de K_{ow} :

I- 4 -1-méthode du brassage lent

La méthode du brassage lent, proposée ici, permet de réduire les difficultés expérimentales associées à la formation de micro-gouttelettes dans la méthode par agitation en flacon. Avec le brassage lent, l'eau, le 1-octanol et la substance d'essai s'équilibrent dans un réacteur thermostaté soumis à une agitation douce. Les échanges entre les phases sont accélérés par le brassage. Ce dernier produit une légère turbulence qui favorise les échanges entre le 1-octanol et l'eau sans provoquer la formation de gouttelettes [26].

I-4-2-Méthode par agitation du flacon

La loi de partage de Nernst s'applique aux solutions diluées et à température, pression et PH constants. Elle n'est valable que pour une substance pure, dispersée entre deux solvants purs, et lorsque la concentration du soluté ne dépasse pas 0,01 mole par litre dans chacune des phases. Les résultats peuvent être affectés par la présence concomitante de plusieurs solutés différents dans l'une des deux ou dans les deux phases.

La dissociation ou l'association des molécules dissoutes entraîne des déviations par rapport à la loi de partage. Ces déviations se manifestent par le fait que le coefficient de partage devient dépendant de la concentration. Les substances ionisables ne doivent être mesurées que sous leur forme non ionique (acide libre ou base libre), obtenue grâce à un tampon approprié dont le pH a été ajusté à au moins une unité en dessous (acide libre) ou au-dessus (base libre) du pK [27].

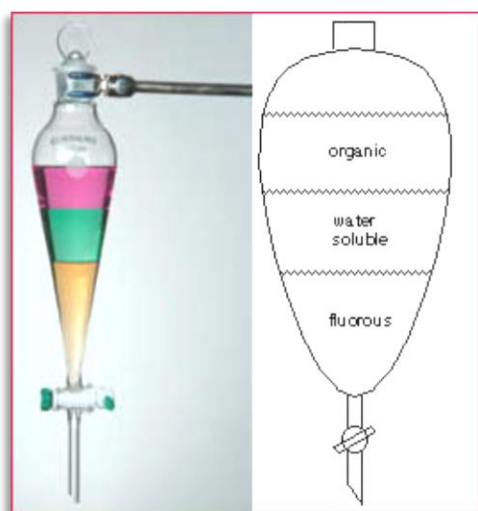


Figure I-3: principe du méthode par agitation du flacon[28].

I-4-3-Méthode par HPLC :

Il détermine en une analyse trois propriétés majeures des phases stationnaires:

- l'hydrophobicité, capacité générale de rétention de la phase qui dépend du taux de greffage et de la surface spécifique
- la sélectivité stérique liée à son aptitude à séparer des molécules dont la structure

tridimensionnelle est différente (isomères)

- l'accessibilité aux silanols résiduels, notamment pour les produits basiques

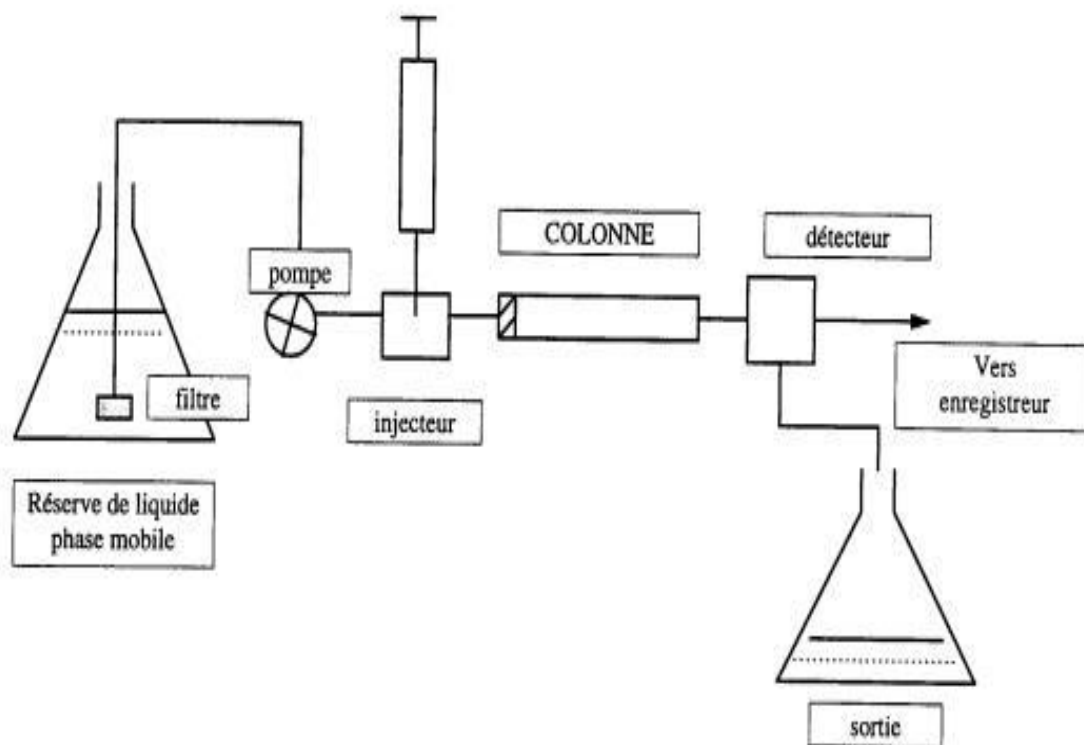


Figure I-4: principe de fonctionnement de HPLC [29].

I-4-4-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 saturé avec une concentration donnée de la substance à tester dans de l'octanol.



Figure I-5: Montre le principe du méthode sur colonne.[30]

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 micro-émulsions.

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

I-5-Les méthodes (QSAR/QSPR) :

I-5-1- Définition de QSAR / QSPR :

Quantitatives structure-activité et propriété de la structure les études de relations (**QSAR / QSPR**) sont sans aucun doute d'une grande importance dans la chimie moderne et de la biochimie. Le concept de **QSAR / QSPR** est de transformer les recherches pour les composés avec souhaitée en utilisant les propriétés chimiques intuition et l'expérience en mathématique et informatique quantifiés formulaire. Une fois une corrélation entre la structure et activité / établissement se trouve, un certain nombre de composés, y compris ceux non encore synthétisé, peut être facilement projeté sur l'ordinateur afin de sélectionner les structures avec les propriétés souhaitées. Il est alors possible de sélectionner les composés les plus prometteurs pour synthétiser et de tester en laboratoire. Ainsi, les **QSAR / QSPR** approche qui préserve les ressources et accélère la processus de développement de nouvelles molécules pour les utiliser comme médicaments, les matériaux, les additifs, ou pour toute autre fin.

I-5-2- Méthode quantitative structure-propriété relation (QSPR):

Les données expérimentales de la propriété ne sont disponibles que pour une petite fraction des composés purs ayant trait à la chimie, génie chimique, génie de l'environnement et l'évaluation de l'impact de l'environnement ,analyse de risque opérationnelle, etc. Par conséquent, des méthodes de prédiction fiable de données de propriété sont nécessaires.

L'introduction des descripteurs topologique, courbe théorique, topochimique, chimie quantique, structure, physico-chimiques, electrotopologique , et d'autres descripteurs moléculaires ont permis considérables amélioration de la présentation des structures chimiques et par conséquent, de l'exactitude des prédictions de propriétés.

QSPRs ont été développés afin de profiter de nouveaux descripteurs moléculaires sophistiqués pour la prédiction des propriétés physico-chimiques. Une étude récente a été publiée dans ce domaine par Dearden.[31]

Dans cette étude, nous allons concentrés sur la plupart des caractéristiques importantes communes des méthodes **QSPR**, telles que définies par Wakeham et d'autres [32].

La méthode **QSPR** peut être schématisé par l'équation suivante:

$$y_p = f(x_{s1}, x_{s2}, \dots, x_{sk}, x_{p1}, x_{p2}, \dots, x_{pm}, \beta_0, \beta_1, \dots, \beta_n) \quad (I - 5)$$

où $x_{s1}, x_{s2}, \dots, x_{sk}$ sont les descripteurs de la structure moléculaire des composés particuliers pures, $x_{p1}, x_{p2}, \dots, x_{pm}$ sont mesurables propriétés d'un même composé (tels que la température d'ébullition, la température de fusion, la toxicité, etc.), $\beta_0, \beta_1, \dots, \beta_n$ sont les **QSPR** paramètres, et y_p est la propriété cible (à prévoir) de le même composé.

Pour calculer la **QSPR**, les données disponibles sont divisées en un "ensemble d'apprentissage" et un "ensemble de validation". En utilisant l'ensemble d'apprentissage pour la régression linéaire multiple ou non linéaire et des techniques de moindres carrés partiels sont utilisées pour sélectionner les descripteurs moléculaires et / ou les propriétés à être inclure dans la partie droite de l'équation (I – 5) et de calculer les valeurs des paramètres du modèle. La validation du modèle est réalisée en utilisant l'ensemble de validation. Applications avec **QSPRs** utilisés pour la prédiction des propriétés physiques et thermodynamiques sont rapportés par de nombreux auteurs.[32.33]

Ne faut pas s'attendre prédire de façon fiable la propriété modélisée pour l'univers entier de produits chimiques. Sur la base de cette observation, ils ont recommandé la définition d'un "domaine d'applicabilité du modèle" afin de obtenir une prévision fiable de propriétés le domaine d'applicabilité du modèle est déterminé par la variation des structures chimiques dans la base de données et testé par l'ensemble d'apprentissage. Par conséquent, la sélection des composés dans le jeu de la formation ne doit pas être faite au hasard, mais il devrait contenir des composés structurellement liée à le composés ciblé, dont les propriétés sont à prévoir. En outre, de nouvelles méthodes pour l'identification de structure apparentée composés sont nécessaires non seulement pour les ensembles d'apprentissage, mais aussi pour la conception assistée par ordinateur moléculaire des composés structurellement semblable à un composé dont on connaît des propriétés intéressantes. Ces problèmes et les approches pour traiter avec eux ont été récemment passés en revue par Basak et d'autres.[34] Tropsha et d'autres.[35]

Une nouvelle **QSPRs**, est basé sur l'hypothèse que la structure (le vecteur de descripteurs moléculaires) d'un composé cible peut être représenté par une combinaison linéaire des vecteurs de descripteurs moléculaires des composés structurellement liés prédictive (à structure-structure " corrélation ")Shacham et d'autres.[36] et Brauner et d'autres.[37] .

Une régression linéaire pas à pas algorithm a été utilisée pour identifier les paramètres de la structure-structure corrélation. Les paramètres de cette corrélation sont ensuite utilisés dans les corrélations propriété des biens, capable de prédire toutes les propriétés nécessaires d'un composé cible et de fournir une évaluation réaliste estimation de la prédiction accuracy.

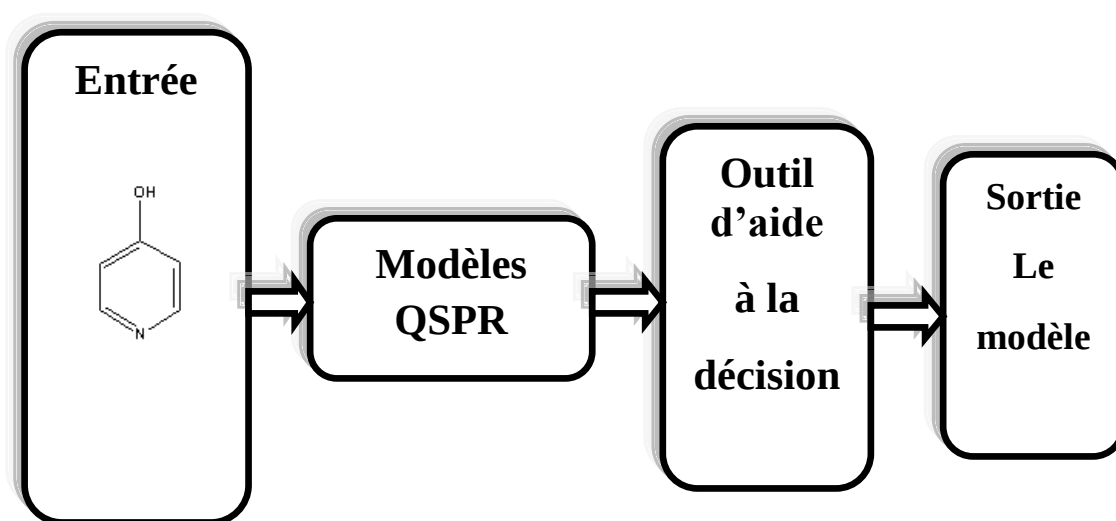


Figure I-6: *Outil de prédiction des propriétés physico-chimique*

I-6-Les descripteurs moléculaires :

Un descripteur moléculaire est le résultat final d'une logique et d'un procédé mathématique qui transforme l'information chimique codée dans une représentation symbolique d'une molécule en nombre utile ou résultat d'une certaine expérience normalisée [38]. Les descripteurs moléculaires jouent un rôle fondamental dans la chimie, les sciences pharmaceutiques, la politique de protection de l'environnement, la recherche de santé et le contrôle de qualité, étant obtenu quand des molécules sont transformées en représentation moléculaire permettant un certain traitement mathématique. On a proposé beaucoup de descripteurs moléculaires qui dérivent de différentes théories et approches dans le but de prévision des propriétés biologiques et physico-chimiques des molécules. [39] Les descripteurs moléculaires sont calculés par des logiciels conçus à cet effet, telle que le logiciel **DRAGON** qui possède 20 blocs de Descripteurs qui sont [40].

- ✓ Descripteurs constitutionnels (Constitutional descriptors (block 1))
- ✓ Descripteurs topologiques (Topological descriptors (block 2))
- ✓ Comptes de promenade et de chemin (Walk and path counts (block 3))
- ✓ Indices de connectivité (Connectivity indices (block 4))
- ✓ Indices de l'information (Information indices (block 5))
- ✓ autocorrélations 2D (List of 2D autocorrelation indices (block 6)-
- ✓ Indices de contiguïté de bord (Edge adjacency indices (block 7))
- ✓ Descripteurs de valeur propre de fardeau (Burden eigenvalue descriptors (block8))
- ✓ Indices de charge topologiques (Topological charge indices (block 9))
- ✓ Indices de Valeur-basés (Eigenvalue-based indices (block 10))
- ✓ Profils moléculaires de Randic (Randic molecular profiles (block 11))
- ✓ Descripteurs géométriques (Geometrical descriptors (block 12))
- ✓ Descripteurs de RDF (RDF descriptors (block 13))
- ✓ Descripteurs 3D-MoRSE (3D-MoRSE descriptors (block 14)-
- ✓ Descripteurs Etranges (Bisard)(WHIM descriptors (block 15))
- ✓ Descripteurs de manière d'obtention (GETAWAY descriptors (block 16))
- ✓ comptes de groupe fonctionnels (Functional group counts)
- ✓ fragments de l' atome centrale (Atom-centred fragments (block 18))
- ✓ Descripteurs du charge (Charge descriptors (block 19))
- ✓ Propriétés moléculaires (Molecular properties (block 20))

I-6- 1-La sélection des descripteurs moléculaire :

La sélection des descripteurs: implique de trouver le sous ensemble le plus informatif de descripteurs des modèles . L'algorithme génétique (GA) étés présenté pour être plus effectif dans la sélection des descripteurs performants dans le cas d'une large gamme de descripteurs .

I-6-2- Mapping (plan):

implique l'application des descripteurs utilisant les analyses de régression multi linéaire tel que heuristique et meilleur régression multilinéaire par méthodes (**BMLR**) .

méthodes multi variables statistique tel que les analyses de composant principal (**P.C.A**) et la moindre carré partiel (**P.L.S**) .

les méthodes non linéaires tel que réseaux de computational neural (nerveux) (**NNS**) pour construire les méthodes mathématiques et d'enchaîner directement les descripteurs au propriété a investiguer. [41]

Conclusion:

Dans ce chapitre nous avons donné une définition générale et spéciale de coefficient de partage. Et nous avons vu les méthodes les plus importantes qui déterminent $\log K_{OW}$ expérimentale et la méthode QSPR qui relie entre $\log K_{OW}$ et les descripteurs moléculaires.

II- La Régression Simple et Multiple

Introduction:

Dans ce chapitre nous avons étudiés l'analyse statistique qui est basé sur la régression simple et multiple pour avoir obtenir un modèle mathématique suivit d'une validation de cet modèle.

II-1-La régression Linéaire Simple (RLS) :

La Régression Linéaire Simple (**RLS**) est un modèle particulier de régression dans lequel :

- ✓ Il n'y a qu'une seule variable explicative (numérique),
- ✓ Le modèle est linéaire dans la variable dans les paramètres.

Comme toutes les techniques prédictives, elle a deux objectifs :

- ✓ Construire un modèle dont les paramètres soient interprétables par le praticien en termes de propriétés de la population dont est extrait l'échantillon. On espère bien entendu que les paramètres du modèle seront de bons estimateurs des paramètres de la population sous-jacente.
- ✓ Utiliser le modèle pour faire des prédictions.

La Régression Linéaire Simple traite la question suivante :

- ✓ Une grandeur y est mesurée
- ✓ Pour un certain nombre de valeurs d'une autre grandeur x .

Ces mesures se traduisent par le diagramme de dispersion suivant :

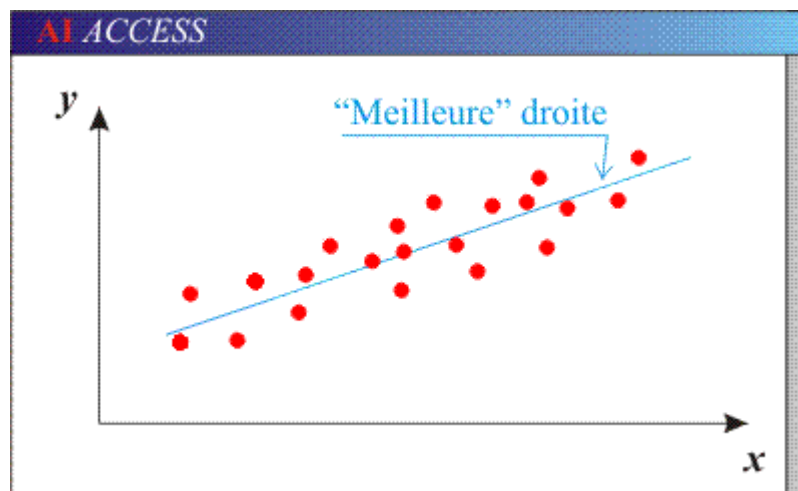


Figure II-1: Montre la meilleure droite de régression linéaire Simple.

Le Modèle Linéaire Simple (MLS) précise les conditions suivantes :

1) Il existe une droite d'équation :

$$y = \alpha + \beta.x \quad (\text{II-1})$$

Telle que, pour chaque valeur x_i de x , la valeur y_i de Y_i est obtenue de la façon suivante :

- ✓ La hauteur de cette droite (donc y) est la valeur de la mesure si celle-ci n'était pas entachée d'erreur.
- ✓ Mais à cette valeur "parfaite" vient s'ajouter une **erreur de mesure**, réalisation d'une valeur. ε_i de Moyenne nulle et de variance σ^2 .

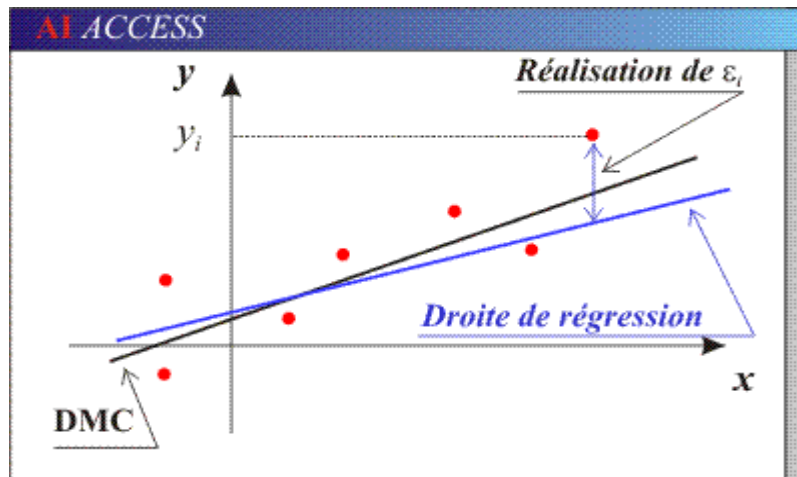


Figure II-2 :Montre la réalisation de valeur ε_i et la droite de régression

Cette droite (inconnue) s'appelle la **droite de régression** du mécanisme qui a engendré les données. Dans la littérature, la Droite des Moindres Carrés est souvent, appelée Droite de Régression. Nous réservons cette appellation à l'authentique droite de régression. Il est par contre justifié de baptiser la DMC "droite de régression empirique".

Les valeurs. Y_i sont donc définies par :

$$Y_i = \alpha + \beta.x + \varepsilon_i \quad (\text{II-2})$$

Les ε_i étant de moyenne nulle, la moyenne de Y_i (pour un i donné) est égale à $\alpha + \beta.x_i$. Les valeurs, Y_i ont toutes leur moyenne sur la droite de régression (ce qui est d'ailleurs la définition de la régression).

Le **MLS** est donc défini par trois paramètres :

- β , la pente de la droite de régression,
- α , l'ordonnée à l'origine de la droite de régression,
- σ^2 , la variance des erreurs.

L'objectif central de la RLS est de trouver de bons estimateurs de ces quantités ainsi que d'identifier les liens de dépendance (covariance) entre ces estimateurs.

II-2 Régression Linéaire Multiple (RLM) :

La **RLM** est une généralisation de la Régression Linéaire Simple (**RLS**) au cas où il y a plusieurs variables explicatives : on cherche alors dans un espace à plus de deux dimensions la "meilleure droite" passant dans un nuage de points de forme si possible effilée et rectiligne.

Dans la régression Linéaire Multiple. Le caractère linéaire dans les variables et les paramètres du modèle peut être maintenu lorsque y dépend non pas d'une seule variable x , mais de plusieurs variables $\{x_1, x_2, \dots, x_n\}$ [42].

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_px_p . \quad (\text{II-3})$$

II-2-1 Principes de la Régression Linéaire Multiple :

Le principe des moindres carrés appliqué à l'hyperplan formé par l'ensemble des variables explicatives (chaque variable constitue un axe de l'hyperplan) est le même que pour le test de régression linéaire simple. Seules les méthodes de calcul changent car on fait ici appel aux calculs matriciels pour lesquels nous ne rentrerons pas dans les détails.

Toutefois, le calcul du vecteur colonne des paramètres du modèle se calcule de la façon suivante :

$$\beta = (X'X)^{-1} X'y \quad (\text{II-4})$$

où " X " est la matrice ($n \times p$) des données relatives aux variables explicatives avec une colonne de 1 pour le calcul de la constante, " y " est le vecteur colonne ($p \times 1$) des données

relatives à la variable expliquée, n est le nombre d'observations et p le nombre de paramètres du modèle.

La conditions d'utilisation du test est :

La distribution normale des mesures sur chaque variable étudiée [43]

II-3 Valider un modèle :

Rappelons que la qualité réelle d'un modèle ne peut pas être estimée au vu de ses résultats sur les données qui ont servi à le construire, sauf cas exceptionnel (p. ex. Régression Linéaire dans les conditions idéales). En fait, beaucoup de types de modèles peuvent, lorsqu'ils sont mal utilisés, conduire à des résultats excellents (et totalement illusoires) sur les données "de construction", et des résultats Catastrophiques sur les Donnée "d'utilisation".

Valider un modèle est difficile, et la qualité de la validation dépend grandement du temps que l'on est prêt à y consacrer. Les approches classiques sont :

- ✓ La théorie, qui, dans les cas les plus simples, permet d'estimer les performances en généralisation.
- ✓ Diverses techniques de simulation des conditions réelles d'utilisation à partir des données disponibles. Citons En particulier .
 - L'utilisation d'un ou de plusieurs ensemble de validation aléatoirement choisis.
 - L'utilisation de plusieurs ensembles de validation choisis de façon particulière par les techniques dites de "_Validation_Croisée", de "Leave_One Out" et plus récemment, de "Bootstrap". Ces techniques (dites de "ré-échantillonnage") permettent en fait d'estimer le pouvoir de généralisation d'un certain type d'architecture de modèle, plutôt que d'un modèle particulier.
- ✓ Des approximations mathématiques de "linéarisation de modèle dans l'espace des paramètres", que nous ne pouvons malheureusement pas aborder dans ce glossaire, et qui donnent toute leur puissance dans le contexte des Réseaux de neurones supervisés.

Note : La phase de validation est parfois appelée phase de "test".[43]

Conclusion:

Après les activités de la préparation de la base des données, nous avons arrivés premièrement a un ensemble de valeurs expérimentales de **log K_{ow}** de 600 composé chimique et biochimique, et deuxièmement a un ensemble de 235 descripteurs moléculaires a l'aide du programme **DRAGON**.

III - Diagramme de travail et la base des données

Introduction:

Pour la réalisation de cette étude il faut arriver premièrement la préparation de la base des données, qui est représentée par les valeurs expérimentales de **log k_{ow}** et les descripteurs moléculaires qui sont sélectionnés par le logiciel **DRAGON**.

Construction d'un plan de travail illustrer les étapes nécessaires pour atteindre la base des données présente.

III-1- Les étapes de travail :

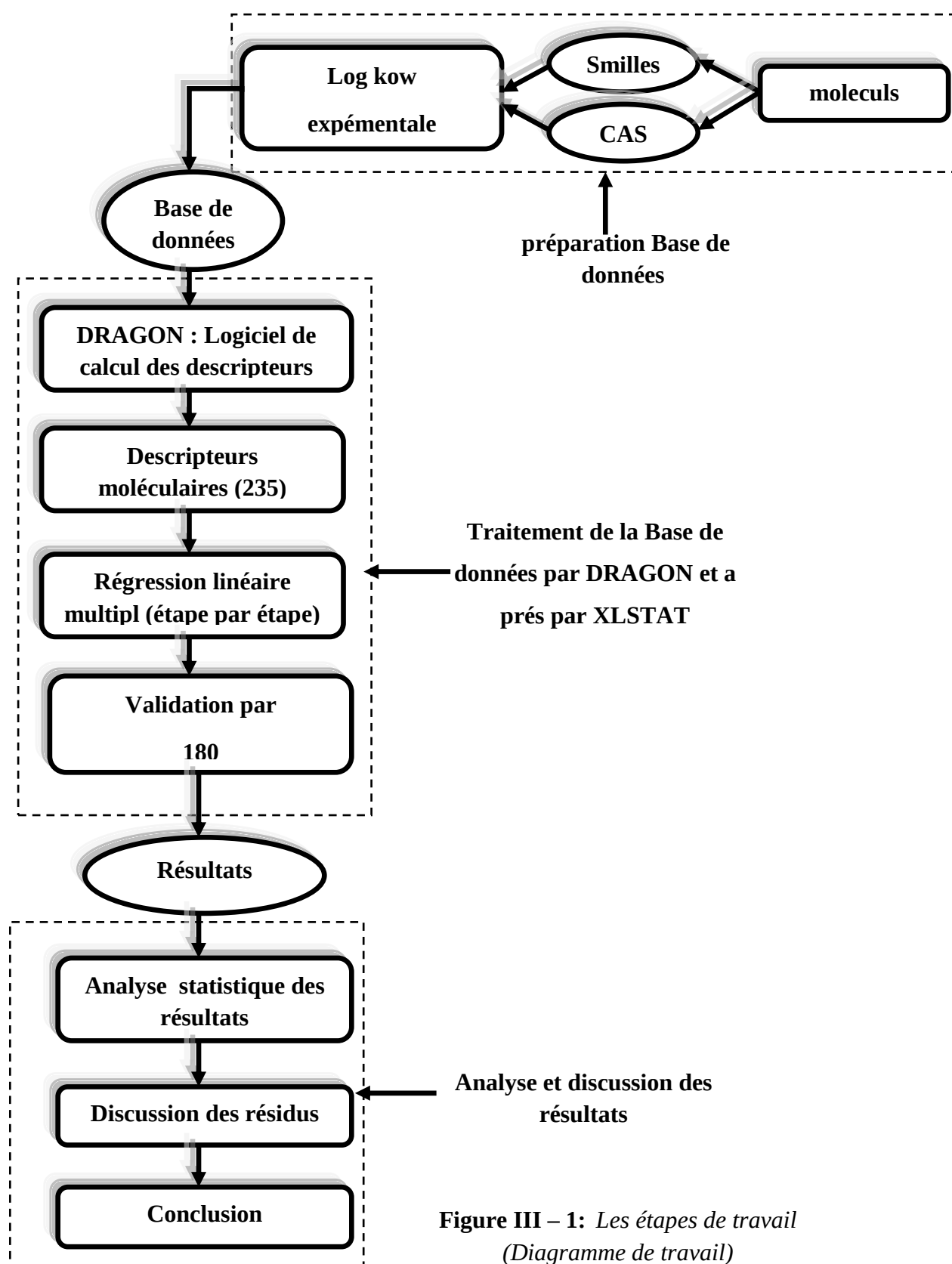


Figure III – 1: Les étapes de travail
(Diagramme de travail)

III-2- base de données :

la base des données utilisées contient 600 composés de diverses natures chimiques, biochimiques et toxiques. Dans cette base de données on a trouvé plusieurs familles chimiques.

cette base de données est présentée par les deux codifications **CAS**(Chemical Abstract Services) et **smiles** (Simplified molecular input line entry specification). Nous avons cherché les valeurs expérimentales de **log K_{ow}** à l'aide de **CAS** et **smiles**. Ces valeurs expérimentales forment la base de données de notre travail

III-3- Génération des descripteurs moléculaires

Afin d'exploiter au maximum les informations contenues dans les structures moléculaires, celles-ci sont traduites en une série de grandeurs (en général scalaires) qui quantifient leurs caractéristiques physico-chimiques et structurales. Ces grandeurs sont appelées descripteurs. Depuis des décennies, de nombreux travaux ont été menés afin de développer des descripteurs capables de décrire de la manière la plus exhaustive possible les structures moléculaires. On en dénombre aujourd'hui des milliers [44,45]. Ils peuvent être calculés ou empiriques. Ces derniers nécessitant d'être mesurés, les descripteurs calculés seront privilégiés. Les détails les concernant sont disponibles dans l'ouvrage de Karelson [46].

Dans ce travail nous avons utilisé 5 blocs de descripteurs moléculaires qui sont :

- ✓ **Descripteurs constitutionnels (Constitutional descriptors (block 1))**
- ✓ **Descripteurs topologiques (Topological descriptors (block 2))**
- ✓ **Propriétés moléculaires (Molecular properties (block 20))**
- ✓ **fragments de l'atome central (Atom-centred fragments (block 18))**
- ✓ **comptes de groupe fonctionnels (Functional group counts)**

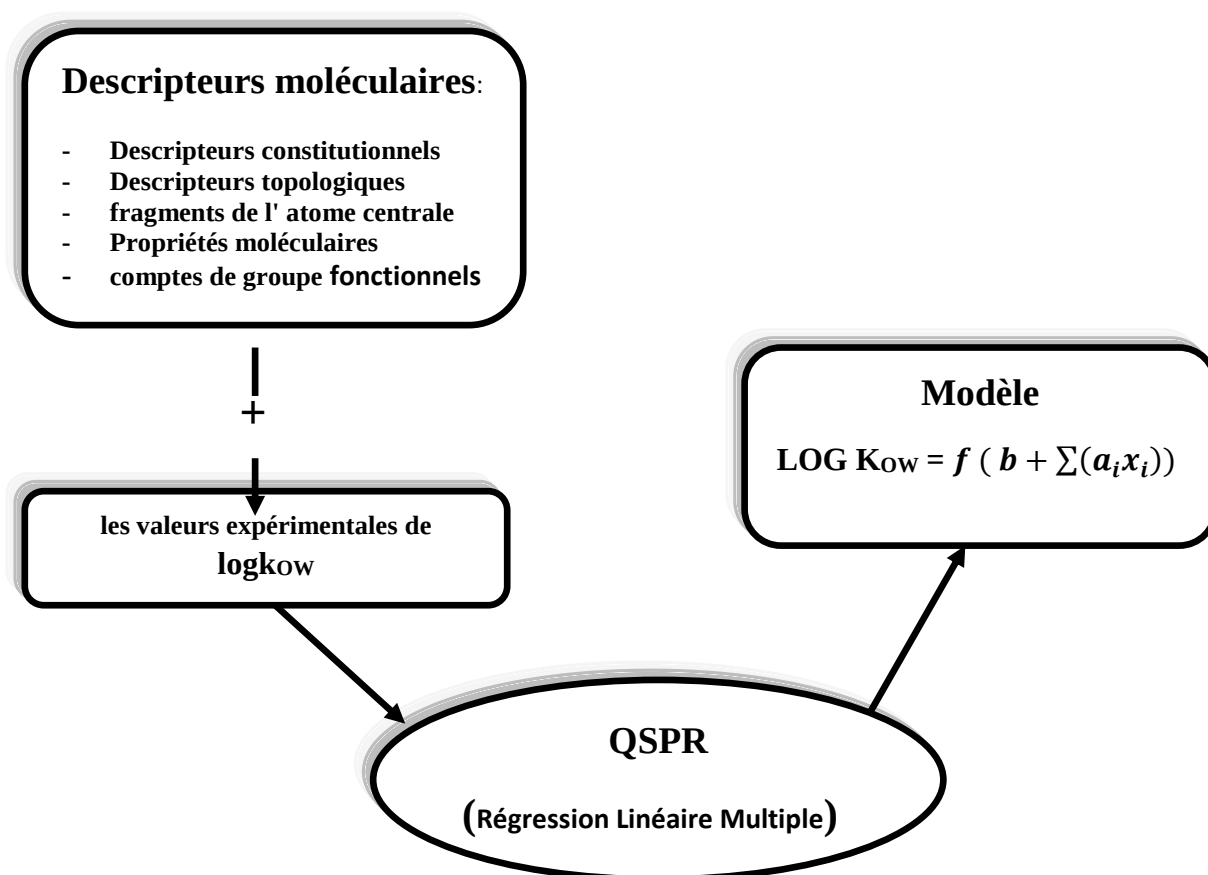


Figure III-2: relation descripteurs moléculaires Enter QSPR et log K_{ow}

III-3-1- Descripteurs constitutionnels

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

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

Ces descripteurs sont très couramment utilisés du fait de leur extrême simplicité non seulement d'un point de vue conceptuel mais surtout calculatoire. Ils sont notamment à la base des analyses par groupes de contribution, comme la méthode de Benson [47], dans la

quelle l'effet des substituants est introduit sous la forme d'une valeur tabulée (sous forme de tableaux).

III-3-2- Descripteurs topologiques

Les descripteurs topologiques sont des indices obtenus à partir d'une structure **2D** de la molécule, à savoir une simple table de connectivité des atomes dans la molécule. Ils contiennent en leur sein des informations sur la taille globale du système, sa forme globale et ses ramifications [44].

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

Parmi les autres indices topologiques courants, on retrouve l'indice de Randic [48], qui peut être interprété comme une mesure de l'aire de la molécule accessible au solvant, les indices de Kier-Hall [49] ou encore l'indice de Balaban [50]. Enfin, des indices de contenu d'information appliquent les fondements de la théorie de l'information de Shannon aux distances topologiques [51].

III-3-3- Propriétés moléculaires

Ces descripteurs comprennent un ensemble de descripteurs moléculaires hétérogènes décrivant physico-chimiques et biologiques ainsi que certaines caractéristiques moléculaires obtenues par les modèles de la littérature.

- ✓ La surface topologique polaires (**TPSA**)
- ✓ La partition Moriguchi octanol-eau (**MLogP**)
- ✓ La partition Ghose-Crippen-Viswanadhan octanol-eau (**ALogP**)
- ✓ La réfractivité molaire (**AMR**)
- ✓ L'indice d'alerte Lipinski (**LAI**)
- ✓ Huit descripteurs drug-like

III-3-4- fragments de l' atome centrée

Ce sont de simples descripteurs moléculaires défini comme le nombre de types d'atomes dans une molécule spécifique. Ils sont calculés en connaissant la composition moléculaire et connectivités atome. **DRAGON** calcule 120 fragments atome-centrée qui sont ceux définis par Ghose et Crippen. Chaque type d'atome est un atome dans la molécule décrite par ses atomes voisins. L'hydrogène et les atomes d'halogène sont classés par l'état d'hybridation et

de l'oxydation de l'atome de carbone auquel ils sont liés, car hydrogènes hétéroatomes liés à un atome de carbone dans une position sont en outre estimé. Les atomes de carbone sont classés par leur état d'hybridation et selon que leurs voisins sont de carbone ou des hétéroatomes.

III-3-5- comptes de groupe fonctionnels

Ce sont de simples descripteurs moléculaires défini comme le nombre de groupes fonctionnels spécifiques dans une molécule. Ils sont calculés en connaissant la composition moléculaire et connectivités atome. [40]

Les blocs des descripteurs moléculaires choisis et les résultats de calcul des descripteurs (paramètres) par le Logiciel DRAGON sont résumés dans le tableau suivant III-1

Les blocs des descripteurs	Le nombre de descripteurs dans le bloc .	Le nombre de descripteurs après la réduction.
comptes de groupe fonctionnels	154	90
Descripteurs topologiques	119	13
fragments de l' atome centrale	120	83
Propriétés moléculaires	29	21
Descripteurs constitutionnels	48	28

Tableau III-1: Montre les résultats de calcul des descripteurs[52]

III-4- Organigramme:

Pour la contribution à la prédiction de **log K_{ow}** des composés chimiques et d'identifier les propriétés physiques et chimiques de ces composés, nous avons divisé cette étude en deux parties expérimental et par l'ordinateur.

Le travail expérimental est-ce la préparation de la base de données à un groupe spécifique de composés issus différentes familles chimiques, où chaque composé de la base de données sera défini par sa valeur expérimental de **log K_{ow}**. Le travail que nous faisons par l'ordinateur est-ce les deux programmes **DRAGON** et **XLSTAT** . Lorsque le programme **DRAGON** calcule les descripteurs moléculaires pour les composé inclus dans la base de données. Et le programme **XLSTAT** doit tenir compte statistiquement les résultats de **DRAGON** .

III-4-1- La préparation la base des données :

Nous apportons 600 composés chimiques différents (Composés de différentes familles chimiques) , Où chaque compose est présenté par **CAS** et **SMILLE**

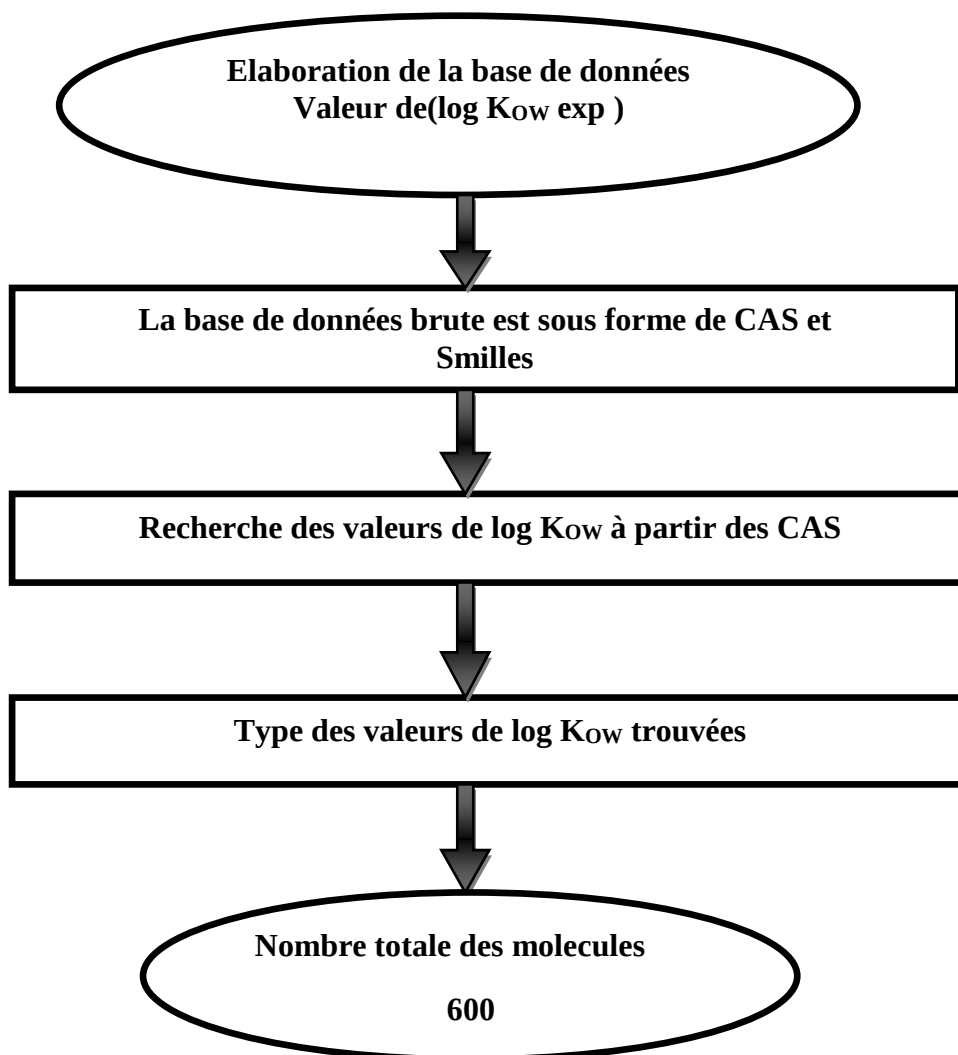
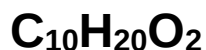


Figure III-3: décrit les étapes de la préparation de la base de données

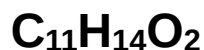
- Exemple de la base de données :



Name: Decanoic acid

CAS #: 334-48-5

SMILES: CCCCCCCCCC(=O)O



Name: Benzoic acid, butyl ester

CAS #: 136-60-7

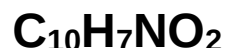
SMILES: c1ccccc1C(=O)OCCCC



Name: Benzenamine, N-phenyl-

CAS #: 122-39-4

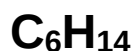
SMILES: c1ccccc1Nc2ccccc2



Name: Naphthalene, 1-nitro-

CAS #: 86-57-7

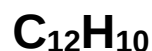
SMILES: c1cc2cccc(N(=O)=O)c2cc1



Name: Butane, 2,2-dimethyl-

CAS #: 75-83-2

SMILES: CCC(C)(C)C



Name: 1,1'-Biphenyl

CAS #: 92-52-4

SMILES: c1ccccc1c2ccccc2

III-4-2- Logiciel DRAGON :

III-4-2-1 – définition Logiciel DRAGON :

DRAGON est une application pour le calcul de descripteurs moléculaires. Ces descripteurs peuvent être utilisés pour évaluer la structure moléculaire d'activité ou des relations structure-propriété, ainsi que pour l'analyse de similitude et de dépistage des bases de données des molécules.

DRAGON fournit plus de 1600 descripteurs moléculaires qui sont divisés en 20 blocs logiques. L'utilisateur peut calculer non seulement le type d'atome le plus simple, groupe fonctionnel et le nombre de fragments, mais aussi plusieurs descripteurs topologiques et géométriques. Certaines propriétés moléculaires telles que **logP**, réfractivité molaire, nombre

de titres en rotation , H-bailleurs de fonds, H-accepteurs et de la surface topologique (**TPSA**) sont également calculés en utilisant des modèles communs tirés de la littérature .

III-4-2-2- Travail par Logiciel DRAGON :

Le cadre L'exécution du programme dans la fenêtre principale de **DRAGON** contient le menu de base du programme. Lorsque le programme démarre les éléments de menu est activé sont comme suit:

1. "descripteurs Calculer " pour obtenir de nouveaux descripteurs.
2. "descripteurs Load" pour afficher et modifier les descripteurs calculés précédemment.

Une fois les descripteurs moléculaires sont calculées ou chargées, trois autres éléments de menu sont activés:

1. Les réponses Load pour charger des variables définies par l'utilisateur .
2. descripteurs View pour afficher les valeurs des descripteurs, des statistiques et des graphiques.
3. Enregistrer les descripteurs pour enregistrer les descripteurs calculés.

Les descripteurs moléculaires **DRAGON** ont été divisés en 20 blocs logiques.

Symboles de courtes définitions et des références bibliographiques de descripteurs de chaque bloc peuvent être obtenus en cliquant sur le bouton correspondant dans "blocs descripteur«le cadre dans la fenêtre principale de l'application.

Ces blocs descripteurs ont été regroupées en fonction de la dimension descripteur. proposé par **DRAGON** sont **0D**, **1D**, **2D** et **3D**, où les chiffres se réfère à **0D** type d'atome et d'obligations, **1D** chiffres fragment, la **2D** à la descripteurs topologiques et connexes, et **3D** pour tous les descripteurs qui dépendent de la géométrie des coordonnées les atomes molécule.

Pour mettre en évidence les blocs contenant des descripteurs **0D**, **1D**, **2D** ou **3D** moléculaire cliquez sur les symboles correspondants sur le haut de "blocs Descriptor» le cadre. Pour désélectionner cliquez partout dans le châssis. Le bouton «Autres» permet à l'utilisateur de mettre en lumière mixte descripteurs moléculaires comme les descripteurs de charge et les propriétés moléculaires. [53]

III-4-3- Logiciel XLSTAT :

III-4-3-1- définition Logiciel XLSTAT :

XLSTAT est un Logiciel propose un nombre important de statistiques descriptives et de graphiques qui vous permettront d'avoir un premier aperçu pertinent de vos données.

Bien que vous puissiez sélectionner plusieurs variables (ou échantillons) à la fois, **XLSTAT** calcul l'ensemble des statistiques descriptives pour chacun des échantillons indépendamment[54].

III-4-3-2- Travail par Logiciel XLSTAT :

Nous a permis de faire le calcul et l'analyse statistique par la technique **QSPR** qui est couramment utilisé dans les études linéaire et aussi comme un excellent outil pour la sélection des descripteur qui précède la construction d'un modèle linéaire ou non linéaire [55]. et la method de régression multiple « stepwise » avec les Cinq blocs des descripteurs moléculaires. Cette analyse se devise en analyse statistique simple et analyse des résidus et validation du modèle par 180 molécules qui présentent 30% du nombre globale. Pour l'évaluation de la corrélation entre les descripteurs **2D** du bloc et les valeurs expérimentales de logk_{ow} qui ont été enregistrés durant la préparation de la base de données. Nous avons réalisé les calculs à l'aide du logiciel de modélisation moléculaires **XLSTAT** et rapporté les résultats obtenus pour ces descripteurs en nous référant à la prévision de la propriété physico-chimique logk_{ow}. dont les valeurs expérimentales ont été prélevées dans la littérature.

Les modèles ont été calculés à l'aide de la méthode de sélection pas à pas des variables de la régression linéaire multiple «**STEPWISE**». Avec la probabilité pour l'entrée est $\alpha = 0.05$

III-5-Déterminer les éléments au-delà du domaine de confiance

Après avoir travaillé Programme devient notre courbe (la courbe de régression) et cette courbe est une droite de régression et le domaine de la confiance . contient cette courbe tous les composés de la base de données . nous sélectionnons les points au-delà du domaine de la confiance et ces points nous allons discuter (la discussion afin de savoir pourquoi leur présence dans le domaine de la confiance).

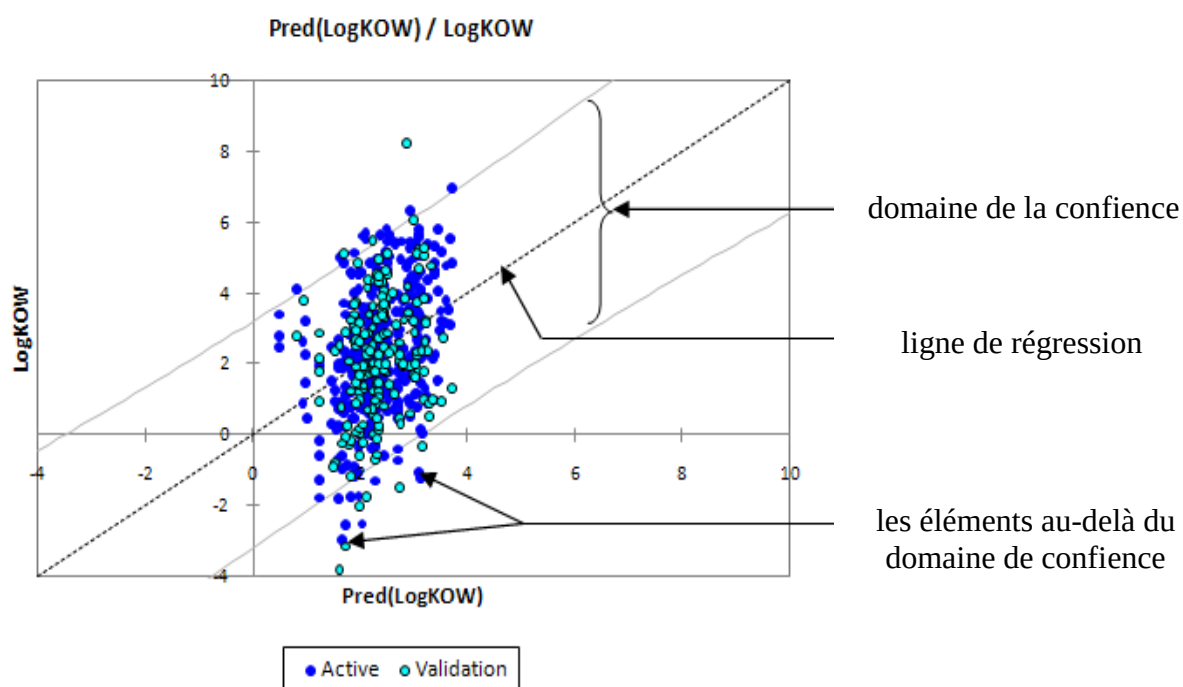


Figure III-4: résultat graphique de XLSTAT (régression)

Conclusion

Après les activités de la préparation de la base des données, nous avons arrivés premièrement à un ensemble de valeurs expérimentales de **log Kow** de 600 composé chimique et biochimique, et deuxièmement à un ensemble de 235 descripteurs moléculaires à l'aide du programme **DRAGON**.

IV- Résultats et discussion:

Introduction :

Dans ce travail nous étudions 5 blocs de descripteurs Chaque blocs séparément où étude chaque blocs les résultats obtenus par le programme XLSTAT (tableaux et graphique) puis analyser les résultats , la méthode utilisée dans cette étude sont la régression linéaire multiple pas à pas «**STEPWISE**»

Cette analyse est une analyse statistique, Pour ca nous faisons un ensemble de validation est un 180 composés de la base de données Il s'agit de déterminer et d'évaluer la relation entre les descripteurs pour chaque blocs et les valeurs de **log_{KOW}** expérimentales.

Après avoir obtenu les résultats statistiques et l'analyse des résultats on a fait une discussion pour chaque bloc, nous avons mis la discussion générale.

IV-1-le résultat de la bloc comptes de groupe fonctionnels :

IV-1-1-Equation du modèle :

$$\text{Log KOW} = b + a_1 \times nCp + a_2 \times nCs + a_3 \times nCq - a_4 \times nCrs + a_5 \times nCbH + a_6 \times nCb- + a_7 \times nR=Cp + a_8 \times nR=Cs - a_9 \times nRcooh + a_{10} \times nArCOOH - a_{11} \times nRCOOR + a_{12} \times nArCOOR - a_{13} \times nRCONH2 - a_{14} \times nArCONH2 - a_{15} \times nRCONHR + a_{16} \times nArCONHR - a_{17} \times nRCONR2 - a_{18} \times nRCHO - a_{19} \times nRCO+ a_{20} \times nC(=N)N2 - a_{21} \times nRNH2+ a_{22} \times nArNHR- a_{23} \times nRNR2+ a_{24} \times nArNR2- a_{25} \times nN-N+ a_{26} \times nN+ - a_{27} \times nOHp - a_{28} \times nOHs + a_{29} \times nRSR + a_{30} \times nRSSR - a_{31} \times nS(=O)2 + a_{32} \times nPO4 + a_{33} \times nCRX3 + a_{34} \times nArX + a_{35} \times nCXr - a_{36} \times nOxiranes - a_{37} \times nPyrrolidines + a_{38} \times nThiophenes - a_{39} \times nTriazoles + a_{40} \times nPyridines - a_{41} \times nHDon - a_{42} \times nHAcc$$

IV-1-2-Les paramètres du modèle:

Variables	Valeurs	Coef	Ecart-type
Constante	0.799	b	0.080
nCp	0.434	a ₁	0.036
nCs	0.451	a ₂	0.019
nCq	0.240	a ₃	0.120
nCrs	-0.095	a ₄	0.031
nCbH	0.226	a ₅	0.018
nCb-	0.319	a ₆	0.035
nR=Cp	0.477	a ₇	0.114
nR=Cs	0.282	a ₈	0.051
nRCOOH	-0.247	a ₉	0.102
nArCOOH	0.353	a ₁₀	0.158
nRCOOR	-0.224	a ₁₁	0.099
nArCOOR	0.397	a ₁₂	0.123

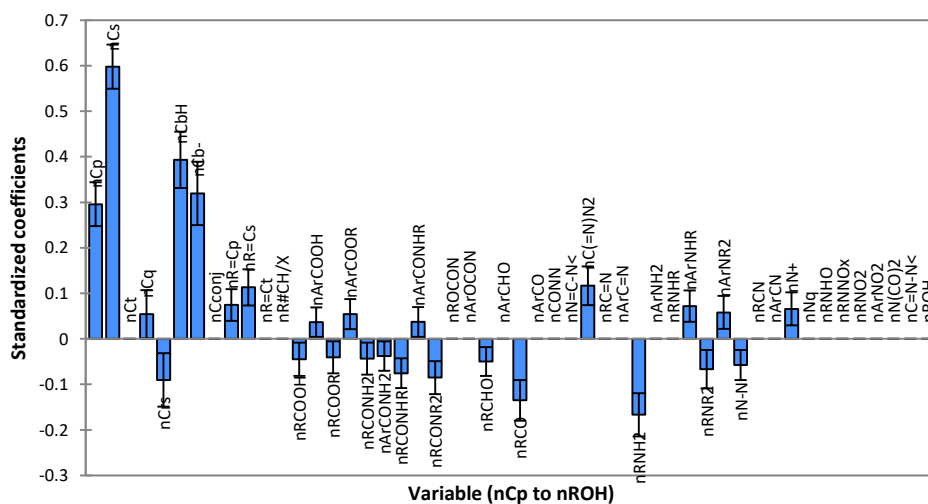
Variables	Valeurs	Coef	Ecart-type
nRCONH2	-0.871	a ₁₃	0.356
nArCONH2	-0.584	a ₁₄	0.252
nRCONHR	-0.664	a ₁₅	0.146
nArCONHR	1.281	a ₁₆	0.577
nRCONR2	-1.049	a ₁₇	0.225
nRCHO	-0.999	a ₁₈	0.322
nRCO	-0.665	a ₁₉	0.112
nC(=N)N2	2.852	a ₂₀	0.530
nRNH2	-1.676	a ₂₁	0.244
nArNHR	0.622	a ₂₂	0.152
nRNR2	-0.462	a ₂₃	0.148
nArNR2	0.674	a ₂₄	0.214
nN-N	-0.817	a ₂₅	0.238

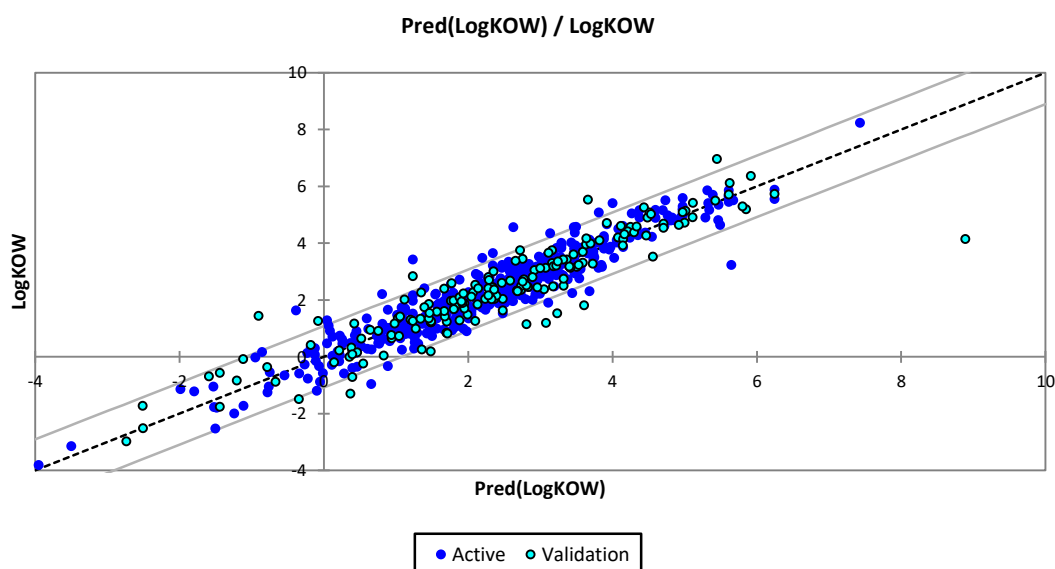
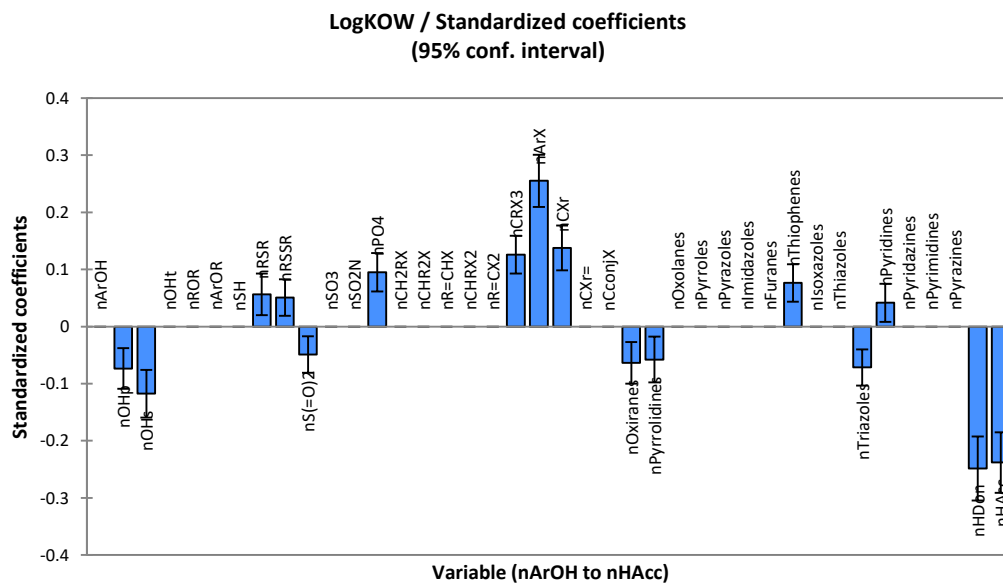
IV-1-3-Coefficients d'ajustement :

Variables	Valeurs	Coef	Ecart-type
nN+	0.460	a ₂₆	0.128
nOHp	-0.592	a ₂₇	0.146
nOHs	-0.661	a ₂₈	0.119
nRSR	0.693	a ₂₉	0.226
nRSSR	1.751	a ₃₀	0.559
nS(=O)2	-0.980	a ₃₁	0.324
nPO4	1.043	a ₃₂	0.188
nCRX3	1.096	a ₃₃	0.146
nArX	0.544	a ₃₄	0.050
nCXr	0.993	a ₃₅	0.143
nOxiranes	-1.268	a ₃₆	0.371
nPyrrolidines	-0.894	a ₃₇	0.317
nThiophenes	1.324	a ₃₈	0.289
nTriazoles	-2.472	a ₃₉	0.556
nPyridines	0.343	a ₄₀	0.140
nHDon	-0.369	a ₄₁	0.042
nHAcc	-0.203	a ₄₂	0.023

Observations	420.000
Somme des poids	420.000
DF	377.000
R ²	0.905
Ajusté R ²	0.894
MSE	0.300
RMSE	0.547
MAPE	39.595
DW	1.817
Cp	50.772
AIC	-465.474
SBC	-291.743
PC	0.117

**LogKOW / Standardized coefficients
(95% conf. interval)**





IV-1-4-Analyse des résultats:

R^2 = le coefficient de détermination donne une idée du pourcentage de la variation de la variable à modéliser, expliqué par les variables explicatives. Plus ce coefficient est proche de 1, plus le modèle est meilleur.

Dans ce cas $R^2 = 0.905$ et $R^2 \text{ ajusté} = 0.894$

Donc le pourcentage de la variation de la variable à modéliser = 90,5%.

Ce la nous indique que 90,5%. de la variabilité du **log Kow** est expliquée par les descripteurs. Le reste de la variabilité est dû à des effets (autres variables explicatives) qui ne sont pas pris en compte ou bien les descripteurs sélectionnés ne sont pas totalement descriptifs

IV-1-5- discussion :

L'étude de corrélation entre les descripteurs deux à deux ont conduit à classer les descripteurs en deux types de descripteurs moins corrélés et descripteurs de grande corrélation . Cela signifie qu'ils vont nous donner des informations différentes sur la propriété à estimer **log K_{ow}** . Le calcul de la matrice de corrélation dans ce bloc nous a permis d'éliminer **42** autres descripteurs puisqu'ils nous ramènent à la même information que les **48** restants.

L'interprétation du graphe (**périd logK_{ow} / logK_{ow}**), nous montre qu'il y a **569** observations qui présentent 95% de molécules de la base de données, et sont à l'intérieur de l'intervalle de confiance , ce qui nous permet de dire que les structures bidimensionnelles de ces molécules sont régulières. Par contre les **31** observations qui ne sont pas corrélées présentent les 5% des observations (molécules) restants de la base de données, ces molécules possèdent des structures chimiques bidimensionnelles irrégulières

IV-2-le résultat de la bloc Descripteurs topologiques :

IV-2-1-Equation du modèle:

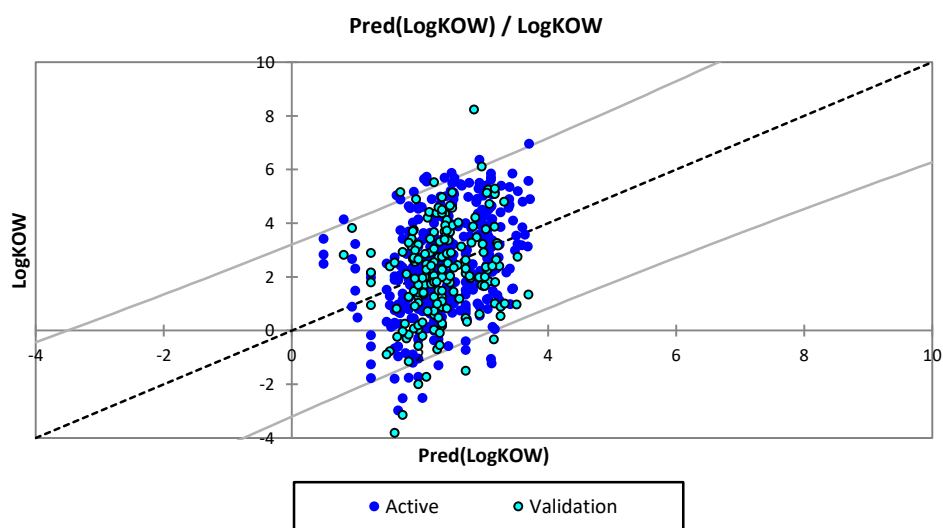
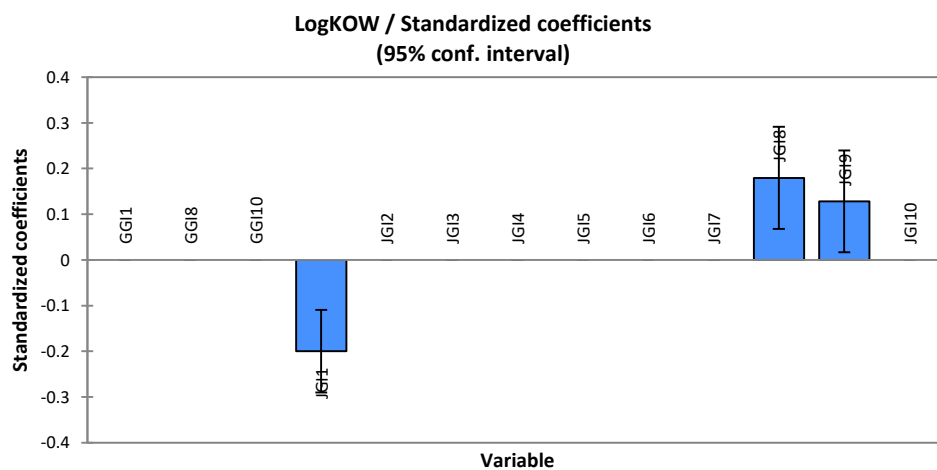
$$\text{LogKOW} = b - a_1 \times \text{JGI1} + a_2 \times \text{JGI8} + a_3 \times \text{JGI9}$$

IV-2-2- Les paramètres du modèle:

Variables	Value	Ecart-type	Coef
Constante	2.716	0.176	b
JGI1	-2.958	0.682	a ₁
JGI8	47.608	15.041	a ₂
JGI9	43.907	19.454	a ₃

IV-2-3- Coefficients d'ajustement :

Observations	420.000
Somme des poids	420.000
DF	416.000
R ²	0.119
Ajusté R ²	0.113
MSE	2.552
RMSE	1.598
MAPE	191.410
DW	0.831
Cp	3.754
AIC	397.512
SBC	413.673
PC	0.898



IV-2-4-Analyse des résultats:

R^2 = le coefficient de détermination donne une idée du pourcentage de la variation de la variable à modéliser, expliqué par les variables explicatives. Plus ce coefficient est proche de 1, plus le modèle est meilleur.

Dans ce cas $R^2 = 0.905$ et R^2 ajusté= 0.894

Donc le pourcentage de la variation de la variable à modéliser = 90,5%.

Ce la nous indique que 90,5%. de la variabilité du **log Kow** est expliquée par les descripteurs. Le reste de la variabilité est dû à des effets (autres variables explicatives) qui ne sont pas pris en compte ou bien les descripteurs sélectionnés ne sont pas totalement descriptifs.

IV-2-5- discussion:

L'étude de corrélation entre les descripteurs deux à deux ont conduit à classifier les descripteurs en deux types de descripteurs moins corrélés et descripteurs de grande corrélation. Cela signifie qu'ils vont nous donner des informations différentes sur la propriété à estimer **log Kow**. Le calcul de la matrice de corrélation dans ce bloc nous a permis d'éliminer 4 autres descripteurs puisqu'ils nous ramènent à la même information que les 9 restants.

L'interprétation du graphe (**périd log Kow / log Kow**), nous montre qu'il y a 569 observations qui pressentent 95% de molécules de la base de données, et sont à l'intérieur de l'intervalle de confiance, ce qui nous permet de dire que les structures bidimensionnelles de ces molécules sont régulières. Par contre les 31 observations qui ne sont pas corrélées présentent les 5% des observations (molécules) restants de la base de données, ces molécules possèdent des structures chimiques bidimensionnelles irrégulières.

IV-3-le résultat de la bloc fragments de l' atome centrale :

IV-3-1-Equation du modèle :

$$\text{LogKOW} = b + a_1 \times \text{C-001} + a_2 \times \text{C-002} - a_3 \times \text{C-004} + a_4 \times \text{C-013} + a_5 \times \text{C-014} + a_6 \times \text{C-015} + a_7 \times \text{C-016} - a_8 \times \text{C-018} + a_9 \times \text{C-020} + a_{10} \times \text{C-024} + a_{11} \times \text{C-025} + a_{12} \times \text{C-026} - a_{13} \times \text{C-027} + a_{14} \times \text{C-032} - a_{15} \times \text{C-036} + a_{16} \times \text{C-040} + a_{17} \times \text{C-041} - a_{18} \times \text{C-043} + a_{19} \times \text{C-044} - a_{20} \times \text{H-050} - a_{21} \times \text{H-051} - a_{22} \times \text{H-052} - a_{23} \times \text{H-053} - a_{24} \times \text{O-056} - a_{25} \times \text{O-058} - a_{26} \times \text{N-066} - a_{27} \times \text{N-072} + a_{28} \times \text{N-073} - a_{29} \times \text{N-074} + a_{30} \times \text{N-077} + a_{31} \times \text{F-081} + a_{32} \times \text{F-084} + a_{33} \times \text{Cl-086} + a_{34} \times \text{Cl-087} + a_{35} \times \text{Cl-088} + a_{36} \times \text{Cl-089} + a_{37} \times \text{I-096} + a_{38} \times \text{I-099} + a_{39} \times \text{S-107} + a_{40} \times \text{S-108} + a_{41} \times \text{B-112}$$

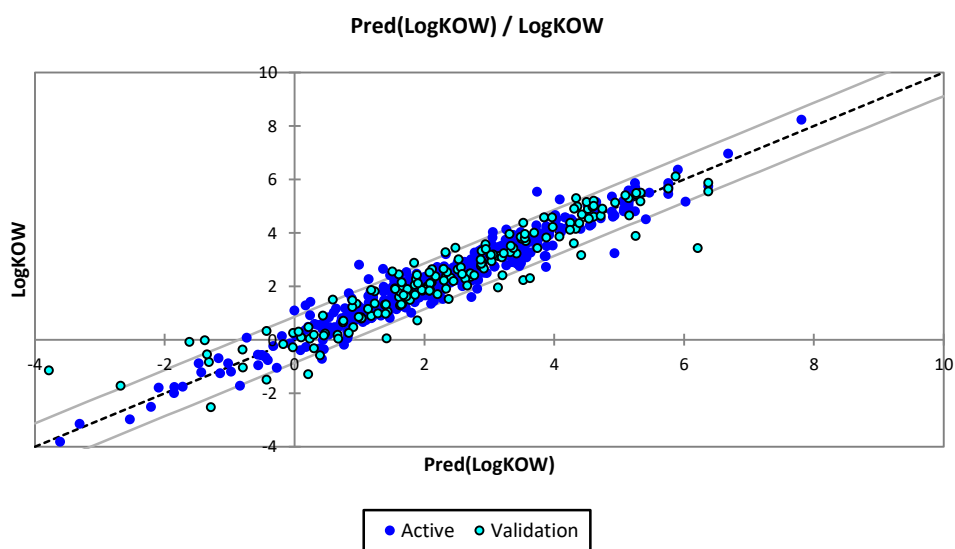
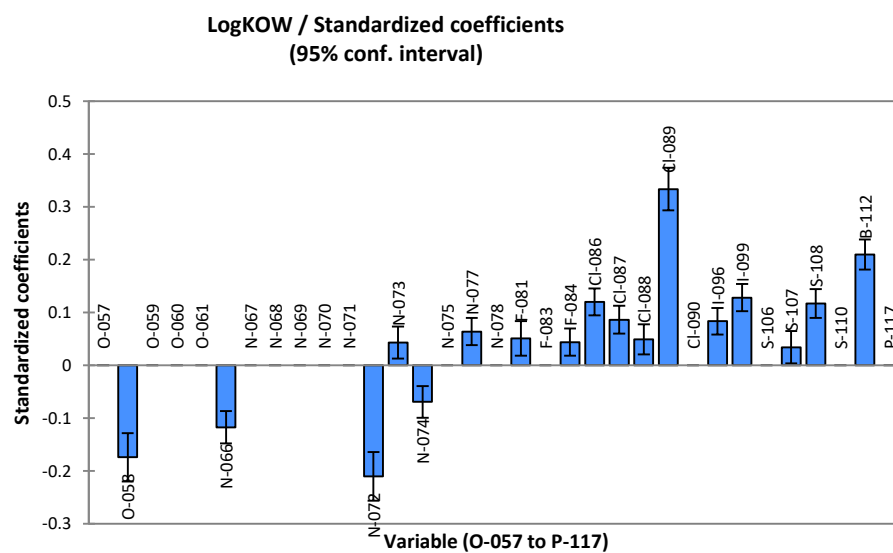
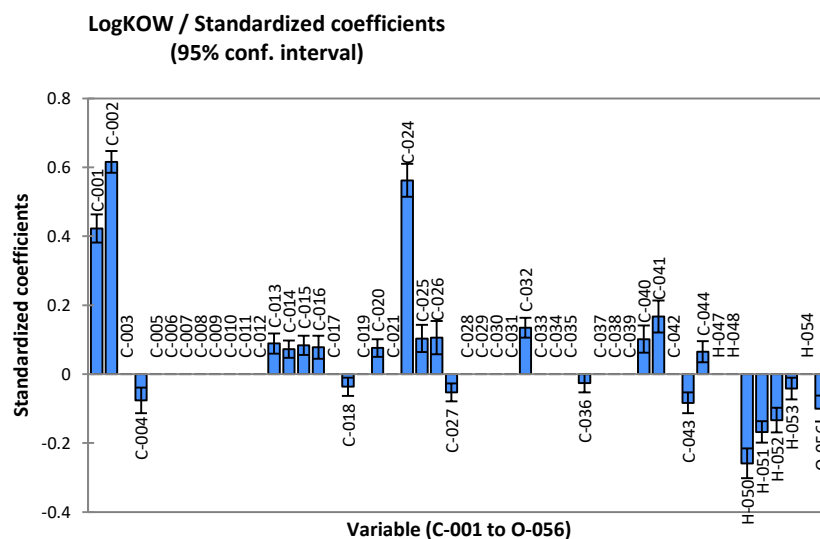
IV-3-2-Les paramètres du modèle:

Variables	Value	Coef	Ecart-type
Constante	0.301	b	0.071
C-001	0.648	a ₁	0.032
C-002	0.485	a ₂	0.013
C-004	-0.322	a ₃	0.08
C-013	0.941	a ₄	0.158
C-014	2.529	a ₅	0.442
C-015	0.572	a ₆	0.098
C-016	0.223	a ₇	0.048
C-018	-0.454	a ₈	0.166
C-020	1.532	a ₉	0.262
C-024	0.328	a ₁₀	0.014
C-025	0.173	a ₁₁	0.034
C-026	0.151	a ₁₂	0.035
C-027	-0.266	a ₁₃	0.066
C-032	0.653	a ₁₄	0.07
C-036	-0.328	a ₁₅	0.166
C-040	0.269	a ₁₆	0.053
C-041	0.92	a ₁₇	0.13
C-043	-1.191	a ₁₈	0.219
C-044	1.311	a ₁₉	0.317
H-050	-0.365	a ₂₀	0.031

Variables	Value	Coef	Ecart-type
H-051	-0.231	a ₂₁	0.022
H-052	-0.09	a ₂₂	0.012
H-053	-0.161	a ₂₃	0.061
O-056	-0.309	a ₂₄	0.059
O-058	-0.33	a ₂₅	0.044
N-066	-1.379	a ₂₆	0.183
N-072	-0.605	a ₂₇	0.067
N-073	0.371	a ₂₈	0.133
N-074	-0.526	a ₂₉	0.116
N-077	0.674	a ₃₀	0.138
F-081	0.889	a ₃₁	0.29
F-084	0.543	a ₃₂	0.162
Cl-086	0.602	a ₃₃	0.065
Cl-087	0.551	a ₃₄	0.086
Cl-088	0.359	a ₃₅	0.106
Cl-089	0.789	a ₃₆	0.048
I-096	1.46	a ₃₇	0.225
I-099	1.695	a ₃₈	0.173
S-107	0.218	a ₃₉	0.099
S-108	1.19	a ₄₀	0.142
B-112	1.576	a ₄₁	0.108

IV-3-3-Coefficients d'ajustement:

Observations	420.000
Somme des poids	420.000
DF	378.000
R ²	0.941
Ajusté R ²	0.935
MSE	0.190
RMSE	0.436
MAPE	40.148
DW	2.015
Cp	44.728
AIC	-657.422
SBC	-487.731
PC	0.072



IV-3-4-Analyse des résultats:

Dans ce cas $R^2 = 0.941$ et $R^2 \text{ ajusté} = 0.935$

Donc le pourcentage de la variation de la variable à modéliser = 94.1 %.

Cela nous indique que 94.1 % de la variabilité du **log Kow** est expliquée par les descripteurs. Le reste de la variabilité est dû à des effets (autres variables explicatives) qui ne sont pas pris en compte ou bien les descripteurs sélectionnés ne sont pas totalement descriptifs

IV-3-5- discussion:

L'étude de corrélation entre les descripteurs deux à deux ont conduit à classer les descripteurs en deux types de descripteurs moins corrélés et descripteurs de grande corrélation. Cela signifie qu'ils vont nous donner des informations différentes sur la propriété à estimer **log Kow**. Le calcul de la matrice de corrélation dans ce bloc nous a permis d'éliminer 42 autres descripteurs puisqu'ils nous ramènent à la même information que les 41 restants.

L'interprétation du graphe (**périd log Kow / log Kow**), nous montre qu'il y a 559 observations qui présentent 95% de molécules de la base de données, et sont à l'intérieur de l'intervalle de confiance, ce qui nous permet de dire que les structures bidimensionnelles de ces molécules sont régulières. Par contre les 41 observations qui ne sont pas corrélées présentent les 5% des observations (molécules) restants de la base de données, ces molécules possèdent des structures chimiques bidimensionnelles irrégulières

IV-4- le résultat de la bloc Propriétés moléculaires :

IV-4-1-Equation du modèle :

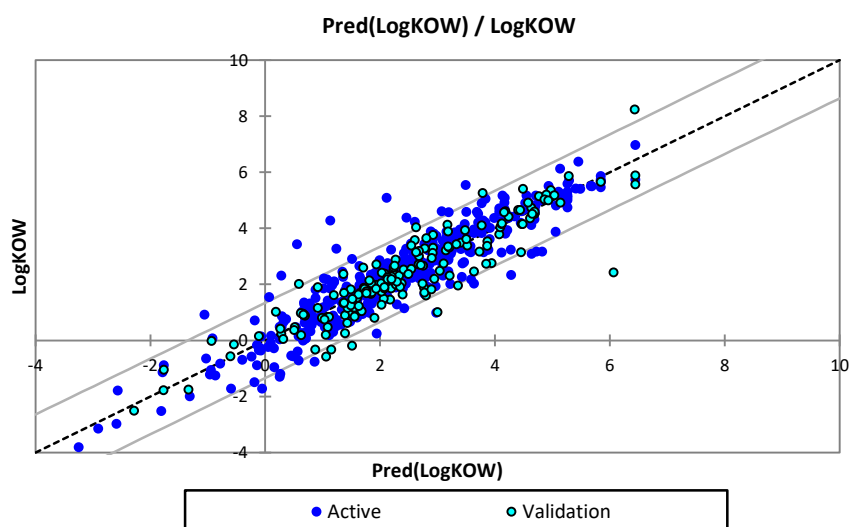
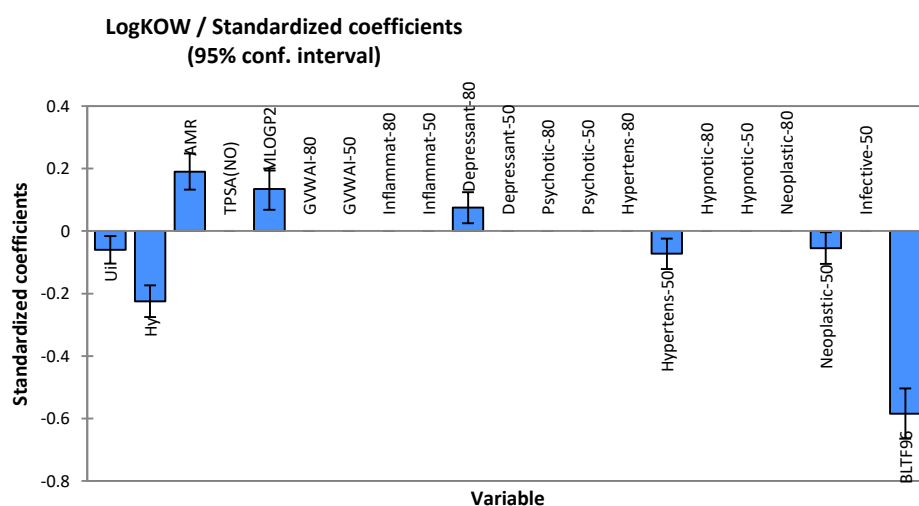
$\text{LogKOW} = b - a_1 \times U_i - a_2 \times H_y + a_3 \times \text{AMR} + a_4 \times \text{MLOGP2} + a_5 \times \text{Depressant-80} - a_6 \times \text{Hypertens-50} - a_7 \times \text{Neoplastic-50} - a_8 \times \text{BLTF96}$

IV-4-2-Les paramètres du modèle:

Variables	Value	Coef	Ecart-type
Constante	-0.854	b	0.13
U _i	-0.08	a ₁	0.03
H _y	-0.407	a ₂	0.047
AMR	0.014	a ₃	0.002
MLOGP2	0.035	a ₄	0.009
Depressant80	0.396	a ₅	0.134
Hypertens-50	-0.568	a ₆	0.193
Neoplastic-50	-0.341	a ₇	0.161
BLTF96	-0.776	a ₈	0.054

IV-4-3- Coefficients d'ajustement:

Observations	420.000
Somme des poids	420.000
DF	411.000
R ²	0.851
Ajusté R ²	0.848
MSE	0.461
RMSE	0.679
MAPE	64.930
DW	1.831
Cp	26.904
AIC	-
	315.992
SBC	-
	279.630
PC	0.156



IV-4-4-Analyse des résultats:

Dans ce cas $R^2 = 0.851$ et $R^2 \text{ ajusté} = 0.848$

Donc le pourcentage de la variation de la variable à modéliser = 85.1%.

Cela nous indique que 85.1% de la variabilité du **log K_{ow}** est expliquée par les descripteurs. Le reste de la variabilité est dû à des effets (autres variables explicatives) qui ne sont pas pris en compte ou bien les descripteurs sélectionnés ne sont pas totalement descriptifs

IV-4-5- discussion:

L'étude de corrélation entre les descripteurs deux à deux ont conduit à classer les descripteurs en deux types de descripteurs moins corrélés et descripteurs de grande corrélation. Cela signifie qu'ils vont nous donner des informations différentes sur la propriété à estimer **log K_{ow}**. Le calcul de la matrice de corrélation dans ce bloc nous a permis d'éliminer 9 autres descripteurs puisqu'ils nous ramènent à la même information que les 12 restants.

L'interprétation du graphe (**périd log K_{ow} / log K_{ow}**), nous montre qu'il y a 565 observations qui présentent 95% de molécules de la base de données, et sont à l'intérieur de l'intervalle de confiance, ce qui nous permet de dire que les structures bidimensionnelles de ces molécules sont régulières. Par contre les 35 observations qui ne sont pas corrélées présentent les 5% des observations (molécules) restants de la base de données, ces molécules possèdent des structures chimiques bidimensionnelles irrégulières

IV-5- le résultat de la bloc Descripteurs constitutionnels:

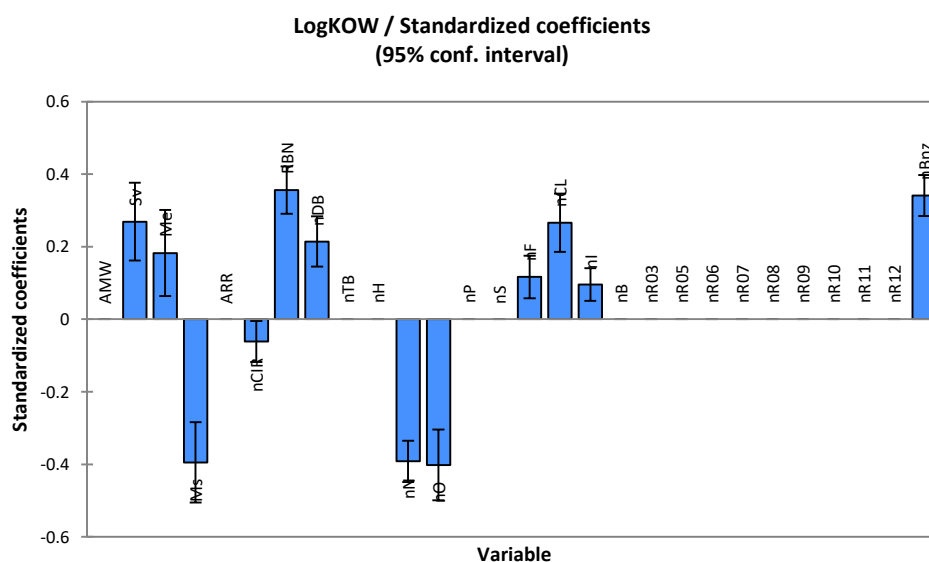
IV-5-1- Equation du modèle :

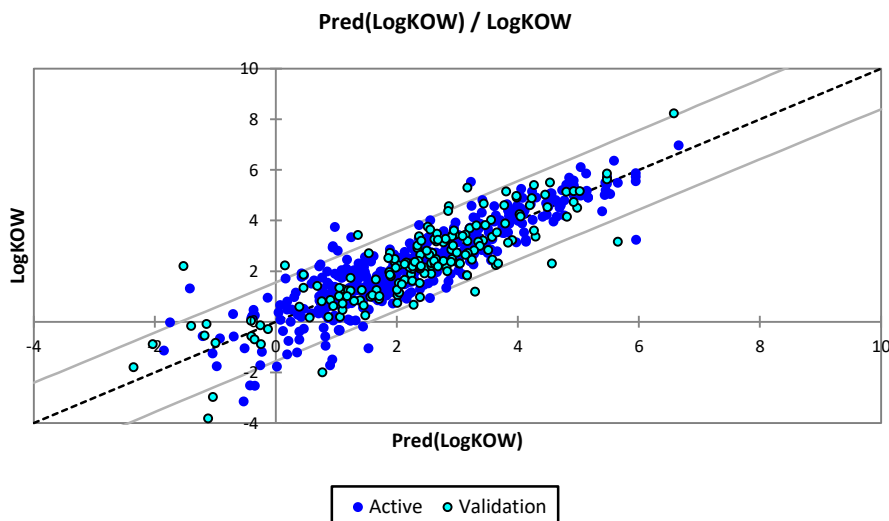
$$\text{LogKOW} = b + a_1 \times \text{Sv} + a_2 \times \text{Me} - a_3 \times \text{Ms} - a_4 \times \text{nCIR} + a_5 \times \text{RBN} + a_6 \times \text{nDB} - a_7 \times \text{nN} - a_8 \times \text{nO} + a_9 \times \text{nF} + a_{10} \times \text{nCL} + a_{11} \times \text{nI} + a_{12} \times \text{nBnz}$$

IV-5-2- Les paramètres du modèle:

IV-5-3-Coefficients d'ajustement:

Variables	Value	Observations	Coef	Ecart- Somme des poids	
Constante	-4.432	DF	b	2.598	420.000
Sv	0.067	R ²	a ₁	0.014	420.000
Me	8.641	Ajusté R ²	a ₂	2.862	407.000
Ms	-1.36	MSE	a ₃	0.194	0.796
nCIR	-0.02	RMSE	a ₄	0.009	0.790
RBN	0.252	MAPE	a ₅	0.023	0.625
nDB	0.311	DW	a ₆	0.051	0.790
nN	-0.579	Cp	a ₇	0.043	89.578
nO	-0.48	AIC	a ₈	0.059	1.620
nF	0.509	SBC	a ₉	0.13	43.716
nCL	0.465	PC	a ₁₀		-184.811
nI	1.132		a ₁₁		-132.287
nBnz	0.738		a ₁₂		0.217





IV-5-4-Analyse des résultats:

Dans ce cas $R^2 = 0.796$ et $R^2 \text{ ajusté} = 0.790$

Donc le pourcentage de la variation de la variable à modéliser = 79.6%.

Cela nous indique que 79.6% de la variabilité du **log Kow** est expliquée par les descripteurs. Le reste de la variabilité est dû à des effets (autres variables explicatives) qui ne sont pas pris en compte ou bien les descripteurs sélectionnés ne sont pas totalement descriptifs

IV-5-5- discussion :

L'étude de corrélation entre les descripteurs deux à deux ont conduit à classer les descripteurs en deux types de descripteurs moins corrélés et descripteurs de grande corrélation. Cela signifie qu'ils vont nous donner des informations différentes sur la propriété à estimer **log Kow**. Le calcul de la matrice de corrélation dans ce bloc nous a permis d'éliminer 13 autres descripteurs puisqu'ils nous ramènent à la même information que les 15 restants.

L'interprétation du graphe (**pred log Kow / log Kow**), nous montre qu'il y a 568 observations qui présentent 95% de molécules de la base de données, et sont à l'intérieur de l'intervalle de confiance, ce qui nous permet de dire que les structures bidimensionnelles de ces molécules sont régulières. Par contre les 32 observations qui ne sont pas corrélées présentent les 5% des observations (molécules) restants de la base de données, ces molécules possèdent des structures chimiques bidimensionnelles irrégulières

IV-6-Analyse globale:

Blocs des descripteurs	comptes de groupe fonctionnels	Descripteurs topologiques	fragments de l'atome centrale	Propriétés moléculaires	Descripteurs constitutionnels
R2	0.905	0.119	0.941	0.851	0.796
Nombre de variables	42	3	41	8	12
Nombre de molécules non corrélées (hors 95%)	31	31	41	35	32

IV-7-Discussion générale :

D'après les analyses statistiques des valeurs de ($\log k_{ow_{pred}}$ $\log k_{ow_{exp}}$, résidus normalisés) et l'interprétation des graphes (**périd log Kow / log Kow**), on voit que les valeurs de résidus varient de (-2+2) et les valeurs qui sont hors de cet intervalle présentent les points aberrants qui indiquent **les molécules non corrélées**. Ces molécules possèdent des structures chimiques bidimensionnelles irrégulières. Cette irrégularité est due à l'existence des halogènes comme (**Br, Cl, N, S, F**), des liaisons d'hydrogène, van der Waals, et l'aromaticité, ou bien des groupements comme (**NH₂**) qui sont capables de donner un doublet électronique. La qualité du modèle dépend du nombre de points aberrants, et du nombre des variables (paramètres).

La valeur de **R²** nous indique le degré de corrélation de la propriété (**log Kow**) avec les paramètres structuraux des différents blocs.

Si on compare les résultats des blocs entre-eux on peut dire que :

*Les paramètres **2D** des blocs (fragments de l'atome centré et les groupes fonctionnels) donnent une représentation suffisante pour la structure moléculaire des composées de la base des données utilisée.

*Les blocs (fragments de l'atome centré et les groupes fonctionnels) donnent une représentation suffisante pour la structure moléculaire des composés de la base des données utilisée.

D'après ces résultats, on observe que le degré d'estimation de la propriété (**log K_{ow}**) change suivant la qualité de descripteurs de chaque bloc.

D'après cette étude et ces résultats on peut identifier et classifier les composées chimiques ou biochimiques suivants leur toxicité (bioaccumulation) qui dépend de la valeur des résidus (**log K_{ow}_{exp} -- log K_{ow}_{pred}**) ou bien la valeur de (**log K_{ow}**)

Conclusion:

Dans ce chapitre nous avons étudié la corrélation des descripteurs moléculaires qui sont calculés par le logiciel **Dragon** (Descripteurs constitutionnels, Descripteurs topologiques, Propriétés moléculaires, fragments de l'atome centrale et comptes de groupe fonctionnels) avec la propriété physico-chimiques qui est le coefficient de partage octanol/eau (**log K_{ow}**) par la technique **QSPR**.

Conclusion générale :

Dans ce travail nous avons abordé le problème de prédiction du coefficient de partage octanol/eau (**log_{kw}**), pour cela nous avons fait la corrélation de différents types de paramètres (descripteurs) et la contribution pour le calcul de (**log_{kw}**) par la méthode **QSPR**. Après le calcul des descripteurs 2D (constitutionnel, indices de connectivité, propriétés moléculaires, groupes fonctionnelles, fragments de l'atome centré, topologique, indices de charge topologique) par le logiciel **Dragon**. Ce travail a été effectué sur **235** paramètres de **5** différents blocs et une base de données contient **600** composés

de différentes familles chimiques et biochimiques. Pendant la prédiction nous avons trouvés une relation mathématique

$$\mathbf{logk_{ow}} = f(\mathbf{paramètres\ structuraux}), \mathbf{logk_{ow}} = f(b + \sum (a_i \times x_i))$$

Ce modèle donne des résultats de corrélation remarquables avec les descripteurs des blocs **fragments de l'atome centré** et **groupes fonctionnels**, ($R^2 = 0.941$), ($R^2 = 0.905$) respectivement.

REFERENCE

- [1] Barret M. (2009). Devenir des perturbateurs endocriniens HAPs/NP/PCBs au cours de la digestion anaérobie de boues contaminées: rôle de la biodisponibilité et du cométabolisme, Thèse Doctorat, Université Montpellier II.
- [2] Crisafullyetal. 2008. page 821 Prediction of environmental parameters of polycyclic aromatic hydro carbons withCOSMO-RS BerndSchroder , Luis M.N.B.F. Santos , Marisa A.A. Rocha , Mariana B. Oliveira , Isabel M. Marrucho , JoaoA.P.Coutinho. journalhomepage:www.elsevier.com/locate/chemosphere
- [3] Dr : Khalil Hanna. Etude de la Faisabilité de Lutilisation de Molécule "cage"dans la Dépollution des Sols : Solubilisation et Extraction de Polluants Organiques par les Cyclodextrines . Thèse page 51
- [4] Marchand M. and Tissier C. Cellule Ifremer/Ineris d'Analyses des Risques Chimiques en Milieu Marin (ARC), Pages 7.52.79. Centre Ifremer de Nantes, BP 21105, 44311 Nantes Cedex 3, France.
- [5] E. E. Kenaga, 1972 Dow Chemical USA, P0 Box 1706, Midland, Michigan 48640, USA
- [6] Neely, 1974 Partition Coefficients to Measure Bioconcentration Potential of Organic Chemicals in Fish. Environ. Sci. Technol. 8, 1113-1115.
- [7] Leo, 1993; Hansch, 1995 Hansch, C. et A. Leo (1995), *Exploring QSAR* Hydrophobic, American Chemical Society .
- [8] Hansch, Hoekman, et Gao, 199 *Chem. Rev.*, 1996, 96, 1045.
- [9] Leo, A.; Hansch, C.; Elkins, D. Partition Coefficients and Their Uses. *Chem. ReV.* 1971, 71, 525.
- [10] Ghose, A. K.; Crippen, G. M. Atomic Physicochemical Parameters for Three Dimensional Structure-Directed Quantitative Structure- Activity Relationships. I. Partition Coefficients as a Measure of Hydrophobicity. *J. Comput. Chem.* 1986, 4, 565. Ghose, A. K.; Crippen, G. M. Atomic Physicochemical Parameters for Three- Dimensional Structure-Directed Quantitative Structure-Activity Relationships. II. Modelling Dispersive and Hydrophobic Interactions. *J. Chem. Inf. Comput. Sci.* 1987, 27, 21. Ghose, A.; Pritchett, A.; Crippen, G. M. Atomic Physicochemical Parameters for Three-Dimensional Structure-Directed Quantitative Structure-Activity Relationships. III. Modeling Hydrophobic Interactions *J. Comput. Chem.* 1988, 9, 80.
- [11] Jean debord , lipophilie coefficient de partage , Page 5 , (14 février 2004)

- [12] Hansch, C.; Fujita, T. A Method for the Correlation of Biological Activity and Chemical Structure. *J. Am. Chem. Soc.* 1984, 86, 1616.
- [13] Leo, A.; Hansch, C.; Elkins, D. Partition Coefficients and Their Uses. *Chem. Rev.* 1971, 71, 525.
- [14] Hansch, C.; Kim, D.; Leo, A.; Novellino, E.; Silipo, C.; Vittoria, A. Toward a Quantitative Comparative Toxicology of Organic Compounds. *Crit. Rev. Toxicol.* 1989, 19, 185.
- [15] Hansch, C.; Leo, A. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*; American Chemical Society: Washington, DC, 1995.
- [16] Martin, Y. C. *Quantitative Drug Design*; Marcel Dekker Inc.: New York, 1978.
- [17] Franke, R. *Theoretical Drug Design Methods*; Akademie Verlag: Berlin, 1984.
- [18] Mackay, D. *Multimedia Environmental Models; the fugacity approach*; Lewis: Chelsea, MA, 1991.
- [19] Mackay, D.; Shiu, W. Y.; Ma, K. C. *Illustrated Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemicals*; Lewis: Chelsea, MA, 1992-1995.
- [20] Thomann, R. V. Boiaccumulation Model of Organic Chemical Distribution in Aquatic Food Chains. *Environ. Sci. Technol.* 1989, 23, 699.
- [21] Martin, Y. C. *Quantitative Drug Design*; Marcel Dekker Inc.: New York, 1978.
- [22] Chen, F.; Holten-Andersen, J.; Tyle, H. New developments of the UNIFAC model for environmental application. *Chemosphere* 1993, 26 (7), 1325-1354.
- [23] Apelblat, A. Correlation between Activity and solubility of water in some aliphatic alcohols. *Ber. Bunsen-Ges. Phys. Chem.* 1983, 87, 2-5.
- [24] J. Jackson Ellington and Terry L. Floyd¹ Measuring Octanol/Water Partition Coefficients by the "Slow-Stirring" Method
- [25] d'après Office of Research and Development, U. S. Environmental Protection Agency. 1981. *Treatability Manual - Vol. I Treatability Data*. EPA-600/2-82-001a.
- [26] De Bruijn JHM, Busser F, Seinen W, Hermens J. (1989). Determination of octanol/water partition coefficients with the 'slow-stirring' method. *Environ. Toxicol. Chem.* 8: 499-512.
- [27] NF T 20-043 AFNOR (1985), Produits chimiques à usage industriel - Détermination du coefficient de partage - Méthode par agitation en flacon.
- [28]. "New fluorous/organic biphasic systems achieved by solvent tuning" Qianli Chu, Marvin S. Yu, Dennis P. Curran. *Tetrahedron*. Vol 63, 39, 2007, 9890-9895 DOI:10.1016/j.tet.2007.06.092 www.fluorous.com
- [29]. Cours de chromatographie de www.chimie-sup.fr

- [30] Niemelä1993 QTOXIN-program (ver 2.0). Danish Environmental Protection Agency.
- [31] Dearden, J. C. Quantitative Structure-Property Relationships for Prediction of Boiling Point, Vapor Pressure, and Melting Point. *Environ. Toxicol. Chem.* 2003, 22, 1696.
- [32] Wakeham, W. A.; Cholakov, G. St.; Stateva, R. P. Liquid Density and Critical Properties of Hydrocarbons Estimated from Molecular Structure. *J. Chem. Eng. Data.* 2002, 47, 559.
- [33] Li, Q.; Chen, X.; Hu, Z. Quantitative structure-property relationship studies for estimating boiling points of alcohols using calculated molecular descriptors with radial basis-function neural networks. *Chemom. Intell. Lab. Syst.* 2004, 72, 93.
- [34] Basak, S. C.; Gute, B. D.; Mills, D.; Hawkins, D. M. Quantitative Molecular Similarity Methods in the Property/Toxicity Estimation of Chemicals: A Comparison of Arbitrary versus Tailored Similarity Spaces. *J. Mol. Struct.* 2003, 622 (1-2), 127.
- [35] Tropsha, A.; Gramatica, P.; Gombar, V. K. The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Combin. Sci.* 2003, 22, 69.
- [36] Shacham, M.; Brauner, N.; Cholakov, G. St.; Stateva R. P. Property Prediction by Correlations Based on Similarity of Molecular Structures. *AIChE J.* 2004, 50, 2481.
- [37] Brauner, N.; Shacham, M.; Cholakov, G. St.; Stateva, R. P. Property Prediction by Similarity of Molecular Structures - Practical Application and Consistency Analysis. *Chem. Eng. Sci.* 2005, 60, 5458
- [38] Manuel de descripteurs moléculaires, R.Todeschini et V.Consonni, Wiley-VCH, 2000)
- [39] Manuel des descripteurs moléculaires R.Todeschini, V.Consonni, , Wiley-VCH, Weinheim (GER), 2000)
- [40] Dragon_ Aide blocs des descripteurs
- [41] ELULOOKIRJELDUS Kaido Tamm QSPR Modeling Of Some Properties Of Organic Compounds
- [44] M. Karelson, *Molecular Descriptors in QSAR/QSPR*, Wiley, New York, **2000**.
- [45] R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley, Weinheim, **2000**.
- [46] M. Karelson, V.S. Lobanov, A.R. Katritzky, *Chem. Rev.*, **1996**, 96, 1027-1044.
- [47] S.W. Benson, J.H. Buss, *J. Chem. Phys.*, **1958**, 29, 546-572.
- [48] M. Randic, *J. Am. Chem. Soc.*, **2002**, 97, 6609-6615.
- [49] L.B. Kier, L.H. Hall, *J. Pharm. Sci.*, **1981**, 70, 583-589.
- [50] A.T. Balaban, *Chem. Phys. Lett.*, **1982**, 89, 399-404.
- [51] S.C. Basak, D.K. Harriss, V.R. Magnuson, *J. Pharm. Sci.*, **1984**, 73, 429-437.

[52] Rev. sci. fond. app., vol. 2 N°. 2 (2010), 23

[53]DRAGON help

[54]XLSTAT help

[55] Xia B., Liu K., Gong Z., Zheng B., Zhang X. et Fan B. Ecotoxicol. Environ. Saf.
2009, 72, 787-794