

# وزارة التعليم العالي والبحث العلمي

Université BADJI Mokhtar – Annaba

BADJI Mokhtar – Annaba University



Faculté des Sciences  
Département de Chimie

## MEMOIRE

Présenté pour l'obtention du diplôme MAGISTER

*Par M<sup>r</sup>. BOUHEDJAR KHALID*

*DES en Chimie*

**Option : Chimie et environnement**

## THEME

**Toxicité des phénols: Une étude QSAR basée sur la méthode préconisée par l'OCDE.**

*Devant le jury :*

<b>PRESIDENTE :</b>	<b>M<sup>me</sup>. ALI-MOKHNACHE Salima</b>	<b>Pr</b>	<b>UBMA</b>
<b>EXAMINATEURS :</b>	<b>M<sup>r</sup>. MESSADI Djelloul</b>	<b>Pr</b>	<b>UBMA</b>
	<b>M<sup>r</sup>. BOUKHARI Abbas</b>	<b>MC</b>	<b>UBMA</b>
<b>RAPPORTEUR :</b>	<b>M<sup>r</sup>. KHATMI Djameleddine</b>	<b>Pr</b>	<b>U 8 Mai 1945 Guelma</b>

**Année 2010/2011**

## **Dédicace**

Je dédie ce travail à :

Mes parents

Ma grand-mère

Toute ma famille

Mes amis

Et toute l'équipe du 34.

## Remerciements

Je voudrais tout d'abord exprimer ma profonde reconnaissance à Monsieur **D. Messadi** le directeur de laboratoire de sécurité environnementale et alimentaire de l'université d'Annaba qui a dirigé mon travail; Ses conseils et ses commentaires précieux m'ont permis de surmonter mes difficultés et de progresser dans mes études.

Je voudrais également exprimer mes remerciements sincères à Monsieur **D. KHATMI** pour son aide, et les membres du jury pour avoir accepté de juger ce modeste travail.

Un grand merci également à toute l'équipe du laboratoire 34 et en particulier M<sup>r</sup> N. KERTIOU et M<sup>r</sup> H. HADDAG.

## SOMMAIRE

Résumé	
Liste des tableaux	IV
Liste des figures	V
Symboles et abréviations	VI
Introduction générale	02
<b>PARTIE THEORIQUE</b>	
<i>Généralité sur les phénols</i>	05
<i>Toxicologie des phénols</i>	07
<i>Méthodologie QSAR</i>	
I- Introduction	09
II- Histoire des QSAR	10
III- Les modèles QSAR	11
IV- Méthodes utilisées pour le développement de modèles QSAR	15
V- Calcul des descripteurs moléculaires	23
VI- Collecte des données	24
<i>Coefficient de partage</i>	
I- Définition:	26
II- Mesure de coefficient de partage (logP)	27
II-1- Méthode du flacon agité	27
II-2- Détermination par HPLC	27
III- Prédiction du coefficient de partage :	28
III-1- Méthodes <u>fragmentales</u>	30
III-2- Méthodes Des Contributions d'Atomes	31
III-3- Approches par Molécule -Entière	32
<i>Interprétation du modèle</i>	
Action- Toxicité des phénols	35
1- Introduction	35
2- Mécanisme d'action	36
<b>Partie expérimentale</b>	
I- Modèle numéro 01: MLOGP	39
I-1- Equation et analyse de régression	39
I-2- Analyse des résidus et diagnostics d'influence	39
Diagramme de Williams	42
I-3- Qualité de l'ajustement	43
II- Modèle numéro 02: ALOGP	45
II-1- Equation et analyse de régression	45

II-2- Analyse des résidus et diagnostics d'influence	45
Diagramme de Williams	48
II-3- Qualité de l'ajustement	59
III- Modèle numéro 03: CLOGP	51
III-1-Equation et analyse de régression	51
III-2- Analyse des résidus et diagnostics d'influence	51
Diagramme de Williams	54
III-3- Qualité de l'ajustement	55
VI- Modèle numéro 04: LOGK <sub>OWWIN</sub>	57
VI-1-Equation et analyse de régression	57
VI-2- Analyse des résidus et diagnostics d'influence	57
Diagramme de Williams	60
VI-3- Qualité de l'ajustement	61
V- Modèle numéro 05: LOGK <sub>ow</sub>	63
1- Equation et analyse de régression	63
2- Analyse des résidus et diagnostics d'influence	63
Diagramme de Williams	63
V-3- Qualité de l'ajustement	67
<b>Chapitre 1</b>	
<b>Comparaison entre les modèles obtenus par les différentes répartitions</b>	70
Discussion	80
Conclusion	81
<b>Chapitre 2</b>	
<b>Comparaison entre les modèles obtenus par les différents logP</b>	83
Discussion	83
Conclusion	86
<b>Conclusion générale</b>	88
Annexe 01	91
Annexe 02	100
Annexe 03	103
<b>References bibliographiques</b>	112

:

العديد من نماذج QSAR ، نشرها و التي تتعلق بالتنبؤ بسمية المواد في البيئة  
تعتمد عموما على لوغار يتم معامل التجزئة - خاصية كراهية  
تعبّر عن مدى تغلغل المادة في الغشاء الليبيدي.

بما أن عملية قياس logP تجريبيا صعبة للغاية م تطوير طرق نظريه حسابية  
و لمعرفة أية طريقة تعطي أحسن نتيجة فقد تم مقارنة نتائج هذه الطرق.

من أجل اختيار مجموعتي المعايرة و التصديق قمنا بـ  
النتائج المحصل عليها تم المقارنة بينها و بين الاختيار .  
DUPLEX

#### الكلمات الجوهرية:

DUPLEX ، خاصية كراهية الذوبان في الماء،  
Tetrahymina pyriformis التراجع الخطية .

## ABSTRACT

Several QSAR models predicting acute chemical toxicity in environment have been published. They are based mainly on the logarithm of the octanol-water partition coefficient as this hydrophobicity term reproduces the ability of a substance to enter cells through the lipid membrane of living organisms.

Nevertheless the experimental determination of logP can be a complex matter, and experimental values can differ greatly even when referred to the same compound. Thus, several approaches have been developed for the theoretical calculation of logP, but also in these calculations it is not uncommon to have differences of several orders of magnitude. A comparison between these methods was made in order to select one that which can give good results.

For the selection of subsets (training and validation), we used the DUPLEX algorithm and compared there results with the several ones obtained by various randomly splitting data.

### **Key-words:**

QSAR, Hydrophobicity, DUPLEX Algorithm, Toxicity, *Tetrahymena Pyriformis*, Multilinear Regression.

## RESUME

La majorité des modèles QSAR qui sont publiées traitant de la prédiction de la toxicité dans l'environnement, sont généralement basés sur le logarithme du coefficient de partage octanol-eau comme paramètre d'hydrophobicité exprimant la possibilité qu'a une substance à pénétrer les membranes lipidiques d'organismes vivants.

Comme la mesure expérimentale du logP est une tâche très compliquée, plusieurs approches furent développées pour calculer les valeurs de logP théoriquement.

Une comparaison entre ces méthodes a été faite afin de sélectionner celle qui peut donner de bons résultats.

Pour la sélection de sous ensembles (Calibration et Validation), nous avons eu recours à l'algorithme DUPLEX en comparant ses résultats à ceux de plusieurs choix aléatoires.

### **Mots-clés:**

QSAR, hydrophobie, Toxicité, Algorithme DUPLEX, *Tetrahymena pyriformis*, Régression multilinéaire.

	<b>Titre</b>	<b>Page(s)</b>
<b>Tableau-1</b>	Propriété des phénols les plus importants	06
<b>Tableau-2</b>	Classification d'ensemble des descripteurs moléculaires empiriques	12
<b>Tableau-3</b>	Classification générale des descripteurs moléculaires théoriques	14
<b>Tableau 4</b>	Classification des logiciels de calcul de logP	29
<b>Tableau-5</b>	Adresses des différents logiciels de calcul de logP	29
<b>Tableau-6</b>	Signification des paramètres statistiques utilisés dans l' eq. (25)	33
<b>Tableau-7</b>	Valeurs des pCIC <sub>50</sub> observés (pCIC <sub>50</sub> obs) et calculés, prédites (pCIC <sub>50</sub> cal-pred), des résidus standardisés (e <sub>i std</sub> ) ainsi que des leviers (h <sub>ij</sub> ) (Cas de Mlogp).	40
<b>Tableau-8</b>	Valeurs des pCIC <sub>50</sub> observés (pCIC <sub>50</sub> obs) et calculés, prédites (pCIC <sub>50</sub> cal-pred), des résidus standardisés (e <sub>i std</sub> ) ainsi que des leviers (h <sub>ij</sub> ) (Cas de ALOGP).	46
<b>Tableau-9</b>	Valeurs des pCIC <sub>50</sub> observés (pCIC <sub>50</sub> obs) et calculés, prédites (pCIC <sub>50</sub> cal-pred), des résidus standardisés (e <sub>i std</sub> ) ainsi que des leviers (h <sub>ij</sub> ) (Cas de CLOGP).	52
<b>Tableau10</b>	Valeurs des pCIC <sub>50</sub> observés (pCIC <sub>50</sub> obs) et calculés, prédites (pCIC <sub>50</sub> cal-pred), des résidus standardisés (e <sub>i std</sub> ) ainsi que des leviers (h <sub>ij</sub> ) (Cas de LOGK <sub>owwin</sub> ).	58
<b>Tableau11</b>	Valeurs des pCIC <sub>50</sub> observés (pCIC <sub>50</sub> obs) et calculés, prédites (pCIC <sub>50</sub> cal-pred), des résidus standardisés (e <sub>i std</sub> ) ainsi que des leviers (h <sub>ij</sub> ) ( Cas de LOGK <sub>ow</sub> ).	64
<b>Tableau12</b>	Paramètres statistiques de différents choix pou le modèle MLOGP	70
<b>Tableau13</b>	Paramètres statistiques de différents choix pou le modèle ALOGP	72
<b>Tableau14</b>	Paramètres statistiques de différents choix pou le modèle CLOGP	74
<b>Tableau15</b>	Paramètres statistiques de différents choix pou le modèle LOGK <sub>owwin</sub>	76
<b>Tableau16</b>	Paramètres statistiques de différents choix pou le modèle LOGK <sub>ow</sub>	78
<b>Tableau17</b>	Variation des paramètres statistiques selon le choix	80
<b>Tableau18</b>	Valeurs de Q <sup>2</sup> et Q <sup>2</sup> <sub>ext</sub> de chaque modèle	83
<b>Tableau19</b>	Points aberrants et influents pour les modèles MLOGP, CLOGP, ALOGP et LOGK <sub>owwin</sub>	84

# LISTE DES TABLEAUX

	<b>Titre</b>	<b>Page</b>
<b>Figure-1</b>	Les structures de quelques composés phénoliques	05
<b>Figure-2</b>	Diagramme de Williams pour le modèle 01	42
<b>Figure-3</b>	Droites d'ajustement pour le modèle 01	43
<b>Figure-4</b>	Test de randomisation pour le modèle 01	44
<b>Figure-5</b>	Diagramme de Williams pour le modèle 02	48
<b>Figure-6</b>	Droites d'ajustement pour le modèle 02	49
<b>Figure-7</b>	Test de randomisation pour le modèle 02	50
<b>Figure-8</b>	Diagramme de Williams pour le modèle 03	54
<b>Figure-9</b>	Droites d'ajustement pour le modèle 03	55
<b>Figure-10</b>	Test de randomisation pour le modèle 03	56
<b>Figure-11</b>	Diagramme de Williams pour le modèle 04	60
<b>Figure-12</b>	Droites d'ajustement pour le modèle 04	61
<b>Figure-13</b>	Test de randomisation pour le modèle 04	62
<b>Figure-14</b>	Diagramme de Williams pour le modèle 05	66
<b>Figure-15</b>	Droites d'ajustement pour le modèle 05	67
<b>Figure-16</b>	Test de randomisation pour le modèle 05	68
<b>Figure-17</b>	Variation des paramètres statistiques selon les choix pour le modèle MLOGP	70
<b>Figure-18</b>	Répartition des ponts (w1, w2) pour les cinq choix aléatoires et le choix par algorithme DUPLEX du modèle MLOGP	71
<b>Figure-19</b>	Variation des paramètres statistiques selon les choix pour le modèle ALOGP	72
<b>Figure-20</b>	Répartition des ponts (w1, w2) pour les cinq choix aléatoires et le choix par algorithme DUPLEX du modèle ALOGP	73
<b>Figure-21</b>	Variation des paramètres statistiques selon les choix pour le modèle CLOGP	74
<b>Figure-22</b>	Répartition des ponts (w1, w2) pour les cinq choix aléatoires et le choix par algorithme DUPLEX du modèle CLOGP	75
<b>Figure-23</b>	Variation des paramètres statistiques selon les choix pour le modèle LOGK <sub>owwin</sub>	76
<b>Figure-24</b>	Répartition des ponts (w1, w2) pour les cinq choix aléatoires et le choix par algorithme DUPLEX du modèle LOGK <sub>owwin</sub>	77
<b>Figure-25</b>	Variation des paramètres statistiques selon les choix pour le modèle LOGK <sub>ow</sub>	78
<b>Figure-26</b>	Répartition des ponts (w1, w2) pour les cinq choix aléatoires et le choix par algorithme DUPLEX du modèle LOGK <sub>ow</sub>	79
<b>Figure-27</b>	Diagrammes de Williams pour les différents modèles	85
<b>Figure-28</b>	Modèle moléculaire de la phosphatidylcholine	105
<b>Figure-29</b>	Structure de la membrane (Têtes hydrophiles, Queues hydrophobes).	106
<b>Figure-30</b>	Succession des oxydo-réductions le log de la chaîne respiratoire	108

# LISTE DES FIGURES

<b>ACP</b>	Analyse en composantes principales
<b>ADN:</b>	Acide désoxyribonucléique.
<b>ADP:</b>	Adénosine diphosphate.
<b>AlogP:</b>	Coefficient de partage de Ghose/Crippen calculé par le logiciel DRAGON.
<b>AM1 :</b>	Austin Model 1.
<b>ATP:</b>	Adénosine triphosphate.
<b>ATPase :</b>	Adénosine triphosphatase
<b>CIC<sub>50</sub>:</b>	Concentration d'inhibition de la croissance à 50 %.
<b>CLHP:</b>	Chromatographie en Phase Liquide à Haute Performance
<b>ClogP:</b>	Coefficient de partage calculé par le logiciel CLOGP
<b>EPA:</b>	Agence pour la protection de l'environnement des États-Unis
<b>EQM:</b>	Erreur quadratique moyenne.
<b>EQMC:</b>	Ecart quadratique moyen calculé sur l'ensemble de calibration.
<b>EQMP:</b>	Ecart quadratique moyen de prédiction.
<b>EQMP<sub>ext</sub>:</b>	Ecart quadratique moyen calculé sur l'ensemble de validation externe
<b>F :</b>	Statistique de Fisher.
<b>FADH<sub>2</sub>:</b>	Flavine-adénine-dinucléotide réduite
<b>LMO:</b>	Cross-validation by leave-many-out: Validation croisée par omission d'un ensemble d'observations.
<b>LOO:</b>	Cross-validation by leave-one-out: Validation croisée par omission d'une observation.
<b>MCP</b>	Moindres carrés cartiels.
<b>MlogP:</b>	Coefficient de partage de Moriguchi calculé par le logiciel DRAGON.
<b>NAD :</b>	Nicotinamide-adénine-dinucléotide.
<b>NAD:</b>	Nicotinamide-adénine-dinucléotide réduite.
<b>NADH<sub>2</sub> :</b>	Nicotinamide-adénine-dinucléotide réduite.

<b>NADP:</b>	Nicotinamide-adénine-dinucléotide-phosphate réduite.
<b>NADPH<sub>2</sub> :</b>	Nicotinamide-adénine-dinucléotide-phosphate réduite.
<b>OCDE:</b>	Organisation de coopération et de développement économiques.
<b>PCA</b>	Analyse en composantes principales.
<b>PLS</b>	Moindres carrés partiels (Partial Least Squares) ou Projection on Latent Structures).
<b>PML:</b>	Potentiels moléculaires de lipophilie.
<b>PRESS :</b>	Somme des carrés des erreurs de prédiction.
<b>Q<sup>2</sup> LMO:</b>	Cross-validation by leave-many-out: Validation croisée par omission d'un ensemble d'observations.
<b>Q<sup>2</sup> LOO:</b>	Cross-validation by leave-one-out: Validation croisée par omission d'une observation d'une observation.
<b>Q<sup>2</sup><sub>Boot</sub>:</b>	Coefficient de prédiction par la technique du bootstrap.
<b>Q<sup>2</sup><sub>LOO</sub>:</b>	Coefficient de prédiction.
<b>QSAR :</b>	Quantitative Structure Activity Relationships (Relations Quantitatives Structure Activité).
<b>QSPR :</b>	Quantitative Structure Property Relationships (Relations Quantitatives Structure Propriété).
<b>R<sup>2</sup> :</b>	Coefficient de détermination.
<b>R<sup>2</sup><sub>adj</sub> :</b>	Coefficient de détermination ajusté.
<b>RLM:</b>	Régression linéaire multiple.
<b>RLM:</b>	Régression linéaire multiple
<b>RNA:</b>	Acide ribonucléique.
<b>SCE :</b>	Somme des carrés des écarts.
<b>SCT :</b>	Somme des carrés totale.
<b>e<sub>i std</sub> :</b>	Résidu de prédiction standardisé.
<b>h<sub>i</sub> :</b>	Éléments diagonaux de la matrice chapeau.
<b>logK<sub>owwin</sub>:</b>	Coefficient de partage de Meylan et Howard calculé par le logiciel

KOWWIN.

- logP:** Coefficient de partage octanol /eau.
- n:** Dimension de la population (échantillon).
- n<sub>ext</sub>:** Dimension de l'ensemble de validation.
- p :** Nombre de descripteurs dans le modèle.
- pCIC50 :**  $\log 1/ (CIC50)$ .
- s :** Erreur standard.
- t :** t de Student.
- y:** Valeur observée.
- $\hat{y}_{(i)}$  : Valeur prédite.
- $\hat{y}_i$  : Valeur estimée.

# SYMBOLES ET ABBREVIATIONS

# PARTIE THEORIQUE

# INTRODUCTION GENERALE

## I- INTRODUCTION GENERALE:

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

Pour cette raison, l'industrie pharmaceutique s'oriente vers de nouvelles méthodes de recherche, qui consistent à prédire les propriétés et activités de molécules avant même que celles-ci ne soient synthétisées. Deux disciplines de la « chimie computationnelle » se sont développées en réponse à ce besoin : les relations structure-activité ou QSAR (Quantitative Structure-Activity Relationships) et les relations structurepropriété ou QSPR (Quantitative Structure-Property Relationships).

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

Elle permet également de guider les synthèses de nouvelles molécules, sans avoir à les réaliser, ou à analyser des familles entières de composés [1].

Les relations entre les structures des molécules et leurs activités sont généralement établies à l'aide de méthodes de modélisation par apprentissage statistique. Les techniques usuelles reposent sur la caractérisation des molécules par un ensemble de descripteurs, nombres réels mesurés ou calculés à partir des structures moléculaires. Il est alors possible d'établir une relation entre ces descripteurs et la grandeur modélisée. Ces méthodes présentent cependant plusieurs inconvénients. Elles nécessitent en effet la sélection des descripteurs pertinents ainsi que leur calcul. De plus, les molécules sont des structures, qui peuvent être représentées par des graphes ; leur représentation par un vecteur de données induit donc une perte d'information, qui peut avoir une influence sur la qualité des modèles.

La première partie de ce mémoire présente une généralité sur les phénols ainsi que leurs toxicités, on présentant la méthodologie QSAR, nous rappelons les principaux types de descripteurs, ainsi que les méthodes expérimentales de mesure et les approches théoriques de calculs logP.

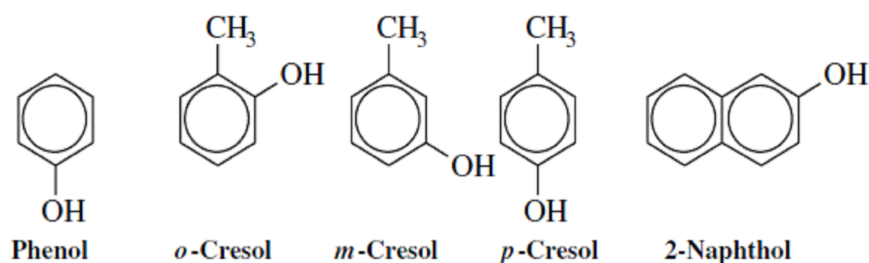
Une détermination expérimentale de  $\log P$  n'étant pas toujours possible, par exemple pour les substances très hydrosolubles, et les substances très lipophiles, il est alors possible d'utiliser une valeur de  $\log P$  déterminée par des méthodes théoriques. De nombreuses approches ont été et continuent d'être élaborées pour estimer  $\log P$ . Trois programmes sur PC, disponibles dans le commerce (CLOGP, KOWWIN, DRAGON), sont fréquemment utilisés pour évaluer les risques en l'absence de données expérimentales.

Dans le premier chapitre nous allons utiliser deux méthodes différentes pour sélectionner les échantillons de calibration et de validation, la première est la division aléatoire, l'autre méthode se fait à l'aide d'algorithmes DULEX

Dans le chapitre deux de la deuxième partie après la construction des modèles par les deux descripteurs  $\log P$  et  $pK_a$ , sachons que les valeurs de  $\log P$  sont calculées à l'aide des logiciels notés précédemment, on compare les paramètres statistiques obtenus par ces modèles.

**I- Généralités sur les phénols:**

Les phénols sont des alcools aryliques sur lesquels le groupe -OH est collé sur la partie hydrocarbure aromatique, le phénol est le plus simple de ces composés. La figure 1 montre quelques composés phénoliques importants. Les phénols ont des propriétés tout à fait différentes de celles des alcools aliphatiques et oléfiniques. Les composés phénoliques les plus importants ont les groupes nitro (-NO<sub>2</sub>) et des halogènes (en particulier Cl) collés sur les anneaux aromatiques. Ces substituants peuvent affecter le comportement chimique et toxicologique [2].



**Figure -1:** Les structures de quelques composés phénoliques

Les propriétés physiques des phénols schématisés dans la figure -1 sont résumées dans le tableau 1. Ces composés phénoliques sont des acides faibles qui s'ionisent sous forme d'ions phénolates en présence de base (eq 01) :



**Tableau-1** : Propriétés des phénols les plus importants :

Composé	Propriété	Température	
		de fusion (°C)	d'ébullition(°C)
<i>Phénol</i>	Acide carboxilique; solide blanc; odeur caractéristique.	41	102
<i>m-Cresol</i>	Se trouve souvent en mélange avec l'ortho- et le para- ; liquide jaune clair.	11	203
<i>o-Cresol</i>	Solide.	31	191
<i>p-Cresol</i>	Cristal solide avec une odeur phénolique.	36	202
<i>1-Naphtol</i>	Alpha-naphtol; solide incolore.	96	282
<i>2-Naphtol</i>	Bêta-naphtol.	122	288

Les phénols sont extraits commercialement à partir du goudron sous forme de base aqueuse, comme les ions phénolates. L'utilisation commerciale principale du phénol réside dans la fabrication des résines de polymères phénoliques, généralement avec du formaldéhyde. Les phénols et les crésols sont employés comme des produits antiseptiques et désinfectants. Le phénol a été utilisé pour la première fois pour stériliser des blessures en chirurgie, (d'après le travail de Lord Lister en 1885). Le mélange utilisé était classiquement constitué de 3 mL de phénol à 50 %, de 2 mL d'eau et de 8 gouttes de savon et 8 gouttes d'huile de croton. En application cutanée, ce mélange permettait une dépigmentation. Dans ce type d'utilisation, il a été rapporté que plus de 30 % des adultes présentaient des *dysrythmies* (Morrison *et al.* 1991) [3]. Le seul cas publié dans la littérature correspond à l'utilisation d'un mélange de composition voisine (40 % de phénol, 0,8 % d'huile de croton dans du savon à base d'hexachlorophène et d'eau) chez un enfant âgé de 10 ans. Ce mélange a été appliqué sous anesthésie sur 1,9 % de la surface corporelle. Cinquante cinq minutes après le traitement, des extrasystoles ventriculaires polymorphes sont observées en l'absence de modification de la pression artérielle et des concentrations en sodium et potassium plasmatiques (Warner et Harper, 1985) [4].

**Toxicologie des phénols**

Généralement, les phénols possèdent les mêmes effets toxicologiques. Le phénol est un poison protoplasmique, il peut endommager toutes sortes de cellules. La dernière étude médicale a démontré que la désinfection avec le phénol cause un nombre étonnant d'intoxications [5].

Des doses mortelles de phénol peuvent être absorbées par la peau. Ses effets toxicologiques sont aigus principalement pour le système nerveux central. La mort peut se produire dans la demi-heure qui suit après l'exposition; les principaux organes endommagés par l'exposition périodique au phénol sont la rate, le pancréas, et les reins et peut être les poumons [6].

# Méthodologie QSAR

## *Méthodologie QSAR :*

- I- Introduction**
- II- Histoire des QSAR**
- III- Les modèles QSAR**
- IV- Méthodes utilisées pour le développement de modèles QSAR**
- V- Calcul des descripteurs moléculaires**
- VI- Collecte des données**

## I- Introduction:

Le besoin de mesurer l'impact des polluants sur l'environnement est en constante augmentation. La mesure de cet impact nécessite de connaître non seulement la toxicité des produits chimiques rejetés, mais aussi celle des molécules issues de leur dégradation. La quantité de mesures à effectuer pour mesurer ces toxicités est donc importante, ce qui augmente les coûts et les délais de développement de nouveaux produits chimiques. Une alternative à la mesure systématique de la toxicité de tels composés sur des animaux est le recours à un modèle, pour prédire l'activité de molécules appartenant à une famille donnée.

Une base de données relative à la toxicité a été compilée par l'Agence pour la Protection de l'Environnement Américaine (EPA) [07]. Cette base, nommée ECOTOX, recense les toxicités connues de molécules diverses sur la vie aquatique, les animaux terrestres et les plantes. L'OCDE a en particulier établi une base dans le but de développer un modèle capable de prédire l'activité toxique de molécules.

Deux méthodes de modélisation peuvent dès lors être mises en place. La plus directe consiste à établir un modèle global, valable pour toutes les classes de molécules. Il semble cependant plus approprié d'utiliser des modèles distincts pour modéliser des phénomènes ou des mécanismes différents.

L'approche qui consiste à établir un modèle pour chaque mode d'action conduit ainsi à des résultats plus précis. Elle nécessite cependant deux étapes : dans un premier temps, chaque molécule est affectée à une classe donnée ; sa toxicité est ensuite prédite grâce au modèle développé pour cette classe.

Ces données ont fait l'objet de prédictions à l'aide d'autres méthodes, telles que la régression par les moindres carrés partiels [08], le logiciel ECOSAR [09], une méthode de partition floue adaptative (AFP) [10], ou les réseaux de neurones probabilistes [11]. Les descripteurs le plus souvent retenus sont le coefficient de partage octanol-eau, le pKa,...

Il est généralement admis que la toxicité de nombreuses substances, particulièrement les produits chimiques organiques industriels, est la conséquence de leur solubilité dans les lipides, alors que leurs caractéristiques moléculaires spécifiques ont peu ou pas d'influence. Leur mode d'action consisterait en la destruction des processus physiologiques associés aux membranes cellulaires.

## II- Histoire Des QSAR :

Les premiers essais de modélisation d'activités de molécules datent de la fin du 19<sup>ème</sup> siècle, lorsque Crum-Brown et Frazer [12] postulèrent que l'activité biologique d'une molécule est une fonction de sa constitution chimique C (eq. 2).

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

Richet [13] a découvert que la toxicité des composés organiques suit inversement leur solubilité dans l'eau. Un tel rapport correspond à l'eq.2, où les  $\Delta\phi$  représentent des différences entre les valeurs des activités biologiques, qui sont causées par les changements de ces propriétés chimiques et particulièrement les propriétés physico-chimiques,  $\Delta C$ .

$$\Delta\phi = f(\Delta C) \quad (\text{eq. 03})$$

Aujourd'hui il n'y a aucune méthode qui applique l'eq. 02 pour traiter les données biologiques. Toutes les équations QSAR correspondent à l'eq. 03, parce que les différences dans l'activité biologique sont seulement quantitativement corrélées avec des changements de lipophilie et/ou d'autres propriétés physico-chimiques des composés.

On peut considérer l'année 1964 comme l'année de naissance de la méthodologie QSAR moderne. Deux articles ont été publiés, un par Hansch et Fujita intitulé " *method for the correlation of biological activity and chemical structure*" [14], l'autre par Free et Wilson portant pour titre " *A mathematical contribution to structure activity studies*" [15].

Les deux contributions ont commencé par élaborer deux nouvelles méthodes pour quantifier la relation Activité Biologique / Structure (QSAR) appelée "Hansch analysis" (linear free energy-related approach, extrathermodynamic approach) et Free "Wilson analysis", respectivement.

L'approche QSAR résulte de la combinaison de différents paramètres physico-chimiques de façon linéaire additive (eq. 04 ;  $\log(1/C)$  est le logarithme de l'inverse de la dose molaire qui produit ou empêche une certaine réponse biologique,  $\log P$  est le logarithme du coefficient de partage de n-octanol/ eau). D'autres méthodes utilisent un paramètre connu sous le nom de paramètre de lipophilie calculée  $\pi$  (eq. 5), il est employé au lieu des valeurs mesurées de  $\log P$  (comme les valeurs  $\sigma$  de Hammett sont employées au lieu des constantes d'équilibre des réactions organiques), et la formulation d'une équation parabolique pour la description quantitative non-linéaire des rapports lipophilie-activité (eq. 6) [16,17].

$$\log 1/C = a \log P + b + \dots + \text{const.} \quad (\text{eq 04})$$

$$x = \log P_{R-X} - \log P_{R-H} \quad (\text{eq 05})$$

$$\log 1/C = a(\log P)^2 + b \log P + c + \dots + \text{const} \quad (\text{eq 06})$$

D'après la contribution signifiée par Fujita et Ban [18], le modèle libre de Wilson est défini par l'eq. 07, où  $a_{ij}$  est la contribution du groupe substituant  $X_I$  en position  $j$ ,  $\mu$  est la valeur de l'activité biologique (théorique) d'un composé référence dans la série ; toutes les contributions de groupe  $a_{ij}$  des différents substituants  $X_I$  se rapportent aux substituants correspondants (le plus souvent l'hydrogène) pour ce composé de référence.

$$\text{Log } 1/c = \sum a_{ij} + \mu \quad (\text{eq.07})$$

### III- Les Modèles QSAR/QSPR

Au cours des décennies passées, les Relations Quantitatives Structure- Activité/ Propriété (QSAR) sont devenues un puissant outil théorique, alternatif à la mécanique quantique, pour la description et la prédiction des propriétés des systèmes moléculaires complexes dans différents environnements. L'approche QSAR procède de l'hypothèse d'une correspondance univoque entre n'importe quelle propriété physique, affinité chimique, ou activité biologique d'un composé chimique et sa structure moléculaire [19]. Cette dernière peut être représentée par la composition chimique, la connectivité des atomes, la surface d'énergie potentielle, et la fonction d'onde électronique d'un composé. Différents descripteurs moléculaires physico- chimiques reflétant la structure peuvent être déterminés empiriquement ou en utilisant des méthodes théoriques et computationnelles de différentes complexités. Il est à souligner que la connaissance de la constitution chimique exacte et/ou de la structure moléculaire tridimensionnelle des composés chimiques étudiés est un pré-requis à l'application de l'approche QSAR.

Le succès de l'approche QSAR dépend de façon critique de la définition précise et de l'utilisation appropriée des descripteurs moléculaires. On distingue, arbitrairement, **les descripteurs moléculaires empiriques** des **descripteurs moléculaires théoriques**.

Les descripteurs empiriques peuvent être divisés en deux classes générales (tableau 1), la première reflète les interactions électroniques intramoléculaires (**descripteurs structurels**) alors que la seconde tient compte des interactions intermoléculaires dans les milieux condensés tels que les liquides et les solutions (**descripteurs de solvation**).

**Tableau -2** : Classification d'ensemble des descripteurs moléculaires empiriques

classe	Sous- classe
Descripteurs structurels	<ul style="list-style-type: none"> <li>- Constantes d'induction</li> <li>- Constantes de résonance</li> <li>- Constantes stériques</li> </ul>
Descripteurs de solvation	<ul style="list-style-type: none"> <li>- Echelles de polarité</li> <li>- Echelles de polarisabilité</li> <li>- Echelles d'acidité</li> <li>- Echelles de basicité</li> <li>- Echelles mixtes</li> </ul>

Les descripteurs structurels les plus répandus ont été définis pour quantifier les propriétés d'induction, l'effet mésomère ou de résonance, et les effets stériques des composés chimiques. Les descripteurs de solvation reflètent les interactions du soluté avec la masse du solvant environnant (**effets de solvant macroscopiques** ou **non spécifiques**), et les liaisons spécifiques, souvent des liaisons hydrogène entre le soluté et les molécules individuelles de solvant (**effets de solvant spécifiques** ou **microscopiques**). Les effets de solvant macroscopiques sont quantifiés en utilisant diverses échelles de polarité et de polarisabilité. Les descripteurs des effets de solvant microscopiques impliquent les échelles générales d'acidité et de basicité. Certaines échelles empiriques d'effets de solvant (échelles mixtes) peuvent impliquer en même temps ces deux effets macroscopique et microscopique. Le coefficient de partage octanol/ eau, log P, est le représentant typique de tels descripteurs.

Les descripteurs moléculaires théoriques peuvent, conventionnellement, être répartis en un certain nombre de classes, selon leur complexité ou leur méthode de calcul. Les descripteurs théoriques les plus simples sont des **descripteurs constitutionnels** qui peuvent être construits à partir de l'information sur la composition chimique du composé considéré. Les nombres, absolus et relatifs, des différents types d'atomes et de liaisons chimiques, la masse molaire, et le nombre de différents cycles dans le composé représentent quelques descripteurs constitutionnels typiques. **Les descripteurs, ou indices, topologiques** décrivent la connectivité des atomes dans la molécule. On a avancé [19] que les indices topologiques pouvaient encoder des interactions moléculaires subtiles et non pas seulement renseigner sur le degré de ramification des liaisons chimiques ou la distribution de la masse spécifique dans la molécule. **Les descripteurs géométriques** sont obtenus à partir de la structure tri-

dimensionnelle des molécules définie par les coordonnées des noyaux atomiques et la grosseur de la molécule représentée, par exemple, par le rayon atomique de Van der Waals. Les molécules de la plupart des composés chimiques possèdent une certaine flexibilité conformationnelle et les surfaces de potentiels moléculaires respectives possèdent de multiples minima locaux. Selon la structure de la molécule, le nombre de ces minima peut être très grand et, par conséquent, il est plutôt difficile de trouver le minimum d'énergie global pour des conditions expérimentales établies.

Evidemment, les descripteurs géométriques peuvent varier de façon significative selon les conformations utilisées dans le calcul de ces descripteurs. Dans une certaine mesure, **les descripteurs théoriques liés à la distribution de charge** peuvent également dépendre de la conformation. Ces descripteurs sont basés sur la structure tri- dimensionnelle et la distribution des charges dans la molécule. Ces dernières peuvent se présenter comme charges atomiques partielles obtenues à partir d'un schéma empirique ou en utilisant des fonctions plus sophistiquées basées sur la fonction d'onde de la molécule calculée par la chimie quantique.

Un certain nombre de **descripteurs quanto- chimiques basés sur les OM** ont été employés dans le développement d'équations QSAR. Les plus utilisés sont les énergies des OM frontières, c'est-à-dire, l'énergie calculée de la plus basse orbitale moléculaire inoccupée ( $\epsilon_{LUMO}$ ), et l'énergie de la plus haute orbitale moléculaire occupée ( $\epsilon_{HOMO}$ ), et la différence entre ces énergies. De même, différents indices de réactivité déduits de la théorie de la superdélocalisabilité de Fukui ou d'autres constructions théoriques ont gagné en popularité parmi les chercheurs.

Tous les descripteurs théoriques ne peuvent être strictement classés selon le schéma présenté dans le tableau 2. Par exemple, les indices topographiques sont déduits de l'information contenant à la fois la topologie et la géométrie des molécules. **Les indices électrotopologiques** sont fondés sur la topologie et la distribution de charge alors que les aires de surfaces partielles chargées sont des descripteurs qui encodent à la fois la distribution de charge et la géométrie des molécules. De tels descripteurs peuvent être classés comme **descripteurs moléculaires mixtes ou combinés**.

Les descripteurs moléculaires peuvent être définis pour tout le système moléculaire étudié ou pour n'importe laquelle de ses parties (fragments). Par exemple, la majorité des descripteurs empiriques structurels sont reliés à des fragments moléculaires appelés substituants. En conséquence, les molécules d'une série congénère de composés chimiques sont divisées formellement en deux ou plusieurs fragments qui correspondent à une unité

structurale constante Y (c'est-à-dire le centre de réaction) et à des unités structurales variables Xi (les substituants). Les relations QSAR/QSPR sont ainsi présentées comme suit :

$$P = P_0^{(Y)} + \sum_i \sum_k a_{ik}^{(Y)} D_{ik}^{(X)} \quad (\text{eq. 08})$$

Où  $P_0^{(Y)}$  est l'ordonnée à l'origine correspondant au fragment moléculaire constant Y, les  $D_{ik}^{(X)}$  sont les descripteurs moléculaires de type k pour les fragments variables  $X_i$ , et les  $a_{ik}^{(Y)}$  sont les coefficients de développement caractéristiques d'une série donnée de composés  $X_iY$ .

**Tableau -3** : Classification générale des descripteurs moléculaires théoriques

classe	Sous- classe
Descripteurs constitutionnels	<ul style="list-style-type: none"> <li>- Dénombrement des atomes ou des liaisons.</li> <li>- Descripteurs basés sur les masses atomiques.</li> </ul>
Descripteurs topologiques	<ul style="list-style-type: none"> <li>- Indices topologiques (connectivité).</li> <li>- Descripteurs théoriques d'information.</li> <li>- Descripteurs topochimiques.</li> </ul>
Descripteurs géométriques	<ul style="list-style-type: none"> <li>- Descripteurs liés à la distance.</li> <li>- Descripteurs liés à l'aire de la surface.</li> <li>- Descripteurs liés au volume.</li> <li>- Descripteurs du champ stérique moléculaire.</li> </ul>
Descripteurs liés à la distribution de charge	<ul style="list-style-type: none"> <li>- Charges atomiques partielles.</li> <li>- Moments électriques moléculaires</li> <li>- Polarisabilités moléculaires.</li> <li>- Descripteurs du champ électrique moléculaire.</li> </ul>
Descripteurs liés aux orbitales moléculaires	<ul style="list-style-type: none"> <li>- Energie des OM frontières</li> <li>- Ordres de liaison</li> <li>- Indices de réactivité de Fukui.</li> </ul>
Descripteurs température dépendants	<ul style="list-style-type: none"> <li>- Fonctions thermodynamiques.</li> <li>- Descripteurs facteurs de Boltzmann pondérés.</li> </ul>
Descripteurs de solvation	<ul style="list-style-type: none"> <li>- Energie électrostatique de solvation.</li> <li>- Energie de dispersion de solvation.</li> <li>- Enthalpie libre de formation de cavité.</li> <li>- Descripteurs de liaison hydrogène.</li> <li>- Entropie de solvation.</li> <li>- Descripteurs d'énergie de solvation linéaire théorique.</li> </ul>
Descripteurs mixtes	<ul style="list-style-type: none"> <li>- Descripteurs topographiques.</li> <li>- Descripteurs électrotopologiques.</li> <li>- Descripteurs de la charge partielle de l'aire de la surface.</li> </ul>

La plupart des descripteurs théoriques qui apparaissent dans le tableau 2 peuvent être calculés soit pour la molécule entière soit pour un fragment moléculaire pré- défini.

#### **IV- Méthodes utilisées pour le développement de modèles QSAR/QSPR**

##### **IV-1- Introduction**

L'application pratique des gammes des descripteurs moléculaires dans le développement de modèles QSAR/QSPR n'est pas une tâche aisée [19]. Tout d'abord, un très grand nombre (>3000) de descripteurs moléculaires, de différentes complexités et de conceptions diverses ont été imaginés et proposés au cours des (60 dernières) années. Ensuite, pendant ce temps, aucune règle stricte n'a été établie, ni même proposée, pour la sélection de descripteurs adaptés parmi la myriade de descripteurs disponibles. Ce choix a souvent été basé sur l'intuition chimique des chercheurs, ou en se pliant à la tradition.

Une autre difficulté dans la sélection des descripteurs QSAR découle de la non standardisation des gammes de descripteurs. Les gammes empiriques des constantes d'induction, de résonance et d'effet stérique des constituants, ou les échelles empiriques d'effets de solvant comportent des erreurs intrinsèques liées aux erreurs respectives des mesures expérimentales. Par ailleurs, les méthodes quanto- mécaniques appliquées aux calculs des descripteurs moléculaires et aux distributions de charges liés aux OM sont souvent basées sur différents paramètres semi- empiriques, ou l'utilisation de différents ensembles de base dans les calculs ab- initio. Naturellement, un descripteur construit à l'aide de différentes méthodes expérimentales ou théoriques, pour divers composés, ne peut être utilisé pour le calcul d'un modèle QSAR unique. Une approche systématique pour la sélection de gammes de descripteurs pour le calcul de modèles QSAR est basée sur la discrimination statistique entre de larges ensembles de descripteurs.

Dans ce qui suit nous passerons en revue diverses approches utilisées pour le développement des " meilleures" équations QSPR dans de grands espaces de descripteurs.

En dernier ressort, les modèles QSAR peuvent être développés selon des modèles mathématiques différents, généralement en relation avec l'analyse statistique multivariée. Le premier modèle, et le plus largement utilisé, consiste en une équation (multi) linéaire obtenue par régression des données expérimentales en fonction d'un ensemble de descripteurs pré-sélectionnés (ou d'un seul), en utilisant la méthode des moindres carrés ordinaires (MCO). Dans quelques cas, les modèles physiques ou chimiques connus du phénomène étudié laissent prévoir certaines formes mathématiques non linéaires (exponentielles ou logarithmiques) de la dépendance entre les données expérimentales et les descripteurs moléculaires. Les modèles

QSAR peuvent alors être établis à l'aide de la technique de régression par les moindres carrés non linéaires. D'autres modèles ont été développés en utilisant l'analyse factorielle ou l'analyse en composantes principales. L'intérêt de ces méthodes est qu'elles évacuent le problème de multicollinéarité inhérent aux méthodes de régression linéaires. Cependant, l'interprétation des équations QSAR est alors entravée par la nature formelle des facteurs ou des composantes principales. Une alternative aux méthodes très classiques de régression linéaire multiple (RLM) et d'analyse en composantes principales (ACP) est la technique de régression par les moindres carrés partiels (MCP ou PLS) [20-36].

## IV-2- Méthodes de régressions linéaire et multilinéaire

### IV-2-1 Aperçu général

Comme signalé auparavant, l'investigateur choisit dans chaque cas un ou plusieurs descripteurs supposé(s) refléter les interactions physiques ou chimiques à la base de la propriété moléculaire ou de la caractéristique du phénomène étudié. Ce choix, encore une fois, est habituellement fondé sur l'intuition chimique, la tradition, ou simplement la disponibilité du descripteur. Néanmoins, cinq principes peuvent aider à la sélection de descripteurs moléculaires convenables pour l'établissement de modèles QSAR. Ce sont:

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

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

### II-2-2- Evaluation préliminaire des données

Avant d'entamer le développement effectif des équations de régression QSAR et QSPR, il est hautement recommandé d'examiner la qualité statistique des données de départ, à la fois les données à corrélérer (variable dépendante) et les descripteurs utilisés dans la corrélation (variables indépendantes).

On distingue habituellement dans un tel pré- traitement des données les analyses univariées des analyses bivariées [31-36].

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

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

### II-2-3- Régression linéaire multiple

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

$$y_i = a_0 + \sum_{j=1}^p a_j X_{ij} + \epsilon_i \quad (\text{eq.09})$$

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

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

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

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

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

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

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

leurs covariances sont nulles.

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

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

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

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

Pour minimiser l'influence des erreurs de détermination des valeurs explicatives (ou régresseurs) sur la précision des résultats de la régression 4 à 5 données (variables dépendantes, ou encore observations) doivent, à la limite, être associées à chaque variable explicative. Le nombre de degré de liberté final ( $n-p-1$ ) doit être [37] tel que :

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

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

Pour les modèles à plus de deux descripteurs, de faibles coefficients de corrélation croisés n'assurent pas forcément l'orthogonalité des descripteurs. Une indépendance globale acceptable des descripteurs sera vérifiée lorsque les facteurs d'inflation de la variance (FIV) calculés pour chacun d'eux obéissent [37] à la condition  $FIV < 5$ .

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

- Le coefficient de détermination multiple :

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

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

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

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

Il est intéressant de considérer, également, la racine de l'écart quadratique moyen de prédiction (EQMP), et celle calculée sur l'ensemble de validation externe (EQMP<sub>ext</sub>) :

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

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

La validation croisée par « leave – one - out » (LOO) [38] consiste à recalculer le modèle sur (n-1) observations, et à utiliser le modèle ainsi obtenu pour calculer la grandeur d'intérêt du composé écarté, notée  $y_{(i)}$ . On répète le procédé pour chacune des grandeurs

d'intérêt. La somme des carrés des erreurs de prédiction, désignée par le symbole PRESS (eq.15), est une mesure de la dispersion des estimations. On l'utilise pour définir le coefficient de prédiction [38] :

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

Contrairement à  $R^2$  qui augmente avec le nombre de paramètres du modèle, le facteur  $Q_{\text{LOO}}^2$  affiche une courbe avec maximum (ou avec palier) obtenu pour un certain nombre de descripteurs, puis décroît de façon monotone. Ce fait confère une grande importance au coefficient  $Q_{\text{LOO}}^2$ . Une valeur de  $Q_{\text{LOO}}^2 > 0,5$  est considérée comme satisfaisante, et une valeur supérieure à 0,9 est excellente [39].

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

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

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

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

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

La validation interne peut être également réalisée en utilisant la technique du bootstrap :  $Q_{\text{boot}}^2$  (bootstrapping). Elle consiste à simuler  $m$  échantillons de même taille  $n$  que l'échantillon initial. Ils sont obtenus par tirage au hasard avec remise parmi les  $n$  individus

observés au départ, ceux-ci ayant tous la même probabilité  $1/n$  d'être choisis [38,40]. Contrairement aux validations croisées par LOO et LMO, les méthodes de bootstrap sont plus efficaces et plus stables.

### VI- 3- Sélection des deux ensembles (validation et calibration)

Il existe différentes méthodes de sélection des échantillons de calibration et de validation, la plus simple est la division aléatoire en deux ensemble disjoints. Une autre méthode de partage des données se fait à l'aide d'algorithmes (DULEX, CADEX, OPIST).

Dans ce travail la méthode aléatoire a été confrontée à l'éclatement par l'algorithme DUPLEX, que nous avons écrit.

Notre écriture du programme pour l'algorithme DUPLEX s'appuie sur sa description faite par R.D. Snee [41] qui mit au crédit de R.W. Kennard son développement. L'édition et l'exécution de ce programme utilisent l'environnement du logiciel MATLAB [42].

L'algorithme DUPLEX commence par une transformation des données (variables explicatives seulement) qui est faite comme suit :

1. Standardisation des données par la formule : 
$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j \sqrt{(n-1)}}$$

$s_j$  : Ecart-type du  $j$  ème régresseur.

$\bar{x}_j$  : Moyenne du  $j$  ème régresseur

$x_{ij}$  : Valeur du régresseur  $j$  pour la  $i$  ème observation.

$n$  : Nombre d'observations.

Pour les  $i = 1, 2, 3 \dots n$  et  $j=1, 2, 3 \dots k$  ; les éléments de la matrice  $\mathbf{Z}$  sont ainsi calculés.

2. On calcul la matrice symétrique définie positive  $\mathbf{A} = \mathbf{Z}'\mathbf{Z}$ .

3. La factorisation de Cholesky peut être maintenant faite pour le calcul de la matrice  $\mathbf{B}$  telle que  $\mathbf{A} = \mathbf{B}'\mathbf{B}$ .

4. Les nouvelles coordonnées des  $n$  observations c'est-à-dire les  $\mathbf{w}$  sont obtenues par le calcul de la matrice  $\mathbf{W} = \mathbf{Z}\mathbf{B}^{-1}$

Une fois la transformation faite on utilise les points orthonormalisés pour calculer les distances euclidiennes entre les paires possibles de points. La distribution des points ou leur

éclatement en un ensemble de prédiction et un second d'estimation se fait en utilisant ces distances de la manière suivante :

- i. La distance la plus grande correspond à la paire de points que l'algorithme classe comme points d'estimations, ces points (E1 et E2) sont ensuite éliminés.
- ii. Les deux points (P1 et P2) les plus éloignés, dans les  $(n - 2)$  restants, sont placés dans l'ensemble de prédiction puis automatiquement éliminés des  $(n - 2)$  points.
- iii. L'algorithme tient compte seulement des distances des  $(n - 4)$  points par rapport aux points d'estimation préalablement choisis (E1 et E2). Chaque point M à deux distances une par rapport à E1, ME1, l'autre, ME2, est celle à E2 ; mais M sera caractérisé par la plus petite distance entre elles. La plus grande distance entre ces  $(n - 4)$  petites distances, qui caractérisent ces  $(n - 4)$  points, identifie le point E3 qui est par conséquent le plus éloigné de la paire (E1, E2) et est, subséquemment à cela, classé avec eux dans l'ensemble d'estimation et éliminé des  $(n - 4)$  point.
- iv. Pour le choix du point P3 de l'ensemble de prédiction, qui sera éliminé des  $(n - 5)$  observations avant de procéder à l'étape (v); l'algorithme continue et refait l'étape (iii) mais en utilisant cette fois la paire (P1 et P2) comme référence pour les  $(n - 5)$  points restants.
- v. Le processus computationnel se poursuit en plaçant les points alternativement dans un ensemble ou dans l'autre. Chaque point (Ei ou Pi) assigné à un ensemble sera éliminé et contribuera au choix du point suivant, car la distance de ce point (Ei ou Pi) au  $((n - (i + 3))$  ou  $(n - (i + 4))$  respectivement) points restants sera prise en compte pour caractériser ces derniers. Par exemple ; le point P3 jouera un rôle, avec P1 et P2 bien sure, pour trouver le point P4 parmi les  $(n - 7)$  points restants.

Les données en main peuvent être ainsi fractionnées, et dans n'importe quel ratio, en un ensemble de prédiction et un autre d'estimation en spécifiant le nombre de point requis que nous jugerons convenable à notre étude.

Pour Snee des précautions particulières doivent être prises avant de scinder les données par cette procédure :

- Le nombre d'observations  $n$  doit être supérieur ou égale à  $(2k + 26)$  si l'ont veut des ensembles d'égales dimensions.

- L'ensemble de prédiction doit contenir au moins 15 observations pour un contrôle rigoureux du pouvoir prédictif du modèle par les statistiques usuelles

$$(Q_{ext}^2, SDEP_{ext}).$$

- Des points répliques d'autres points, ou des points étant des proches voisins, doivent en être purgées nos données d'origines avant tout traitement par DUPLEX.

## V- Calcul des descripteurs moléculaires

La représentation numérique de la structure chimique (descripteurs moléculaires) est une étape importante de l'investigation QSAR/QSPR. Les performances du modèle élaboré et la précision des résultats sont étroitement liées au mode de détermination de ces descripteurs.

Nous avons utilisé le logiciel de modélisation moléculaire Hyperchem 6.03 [43] pour représenter les molécules puis, à l'aide de la méthode semi-empirique AM1, obtenir les géométries finales. Tous les calculs ont été menés dans le cadre du formalisme RHF (pour restricted Hartree-Fock ou formalisme de Hartree-Fock avec contrainte de spin) sans interaction de configuration. Les structures moléculaires ont été optimisées à l'aide de l'algorithme Polak- Ribiere avec pour critère une racine du carré moyen du gradient égale à 0,1 kcal/mol. Les géométries ainsi optimisées ont été transférées dans le logiciel informatique DRAGON [44] pour le calcul de AlogP, MlogP.

Nous avons aussi utilisé les deux logiciels Ultra CLOGP CambridgeSoft [45], et KOWWIN [46], pour les calculs des descripteurs ClogP, et  $\log K_{owwin}$  successivement.

## VI- Collecte des données expérimentales :

Les protozoaires sont souvent utilisés pour l'évaluation de la toxicité. Les méthodes mises en œuvre sont basées sur des critères morphologiques, ultra-structuraux, éthologiques et métaboliques [47].

L'inhibition de la croissance d'une population est un indicateur très en vogue, parce qu'il peut être déterminé directement ou indirectement à l'aide d'un équipement électronique. Ce qui permet l'acquisition rapide des observations nécessaires pour les analyses de régression. Nous considérerons la concentration d'inhibition 50% de la croissance ( $CIC_{50}$ ), dont le logarithme de l'inverse, soit  $pCIC_{50} = \log(CIC_{50})^{-1}$ , servira d'indicateur de toxicité.

Les tests de toxicité ont été réalisés [48,49] en examinant la croissance d'une population de *Tetrahymena pyriformis*. Les essais ont été menés dans des erlenmeyers de 250 ml, contenant 50 ml d'un milieu dont la composition est précisée ci après :

Eau distillée	1000 ml
Proteose peptone	20 g
D-glucose	5 g
extrait de levure	1 g
FeEDTA	1 mL d'une solution à 3 % (masse/v)
pH	7,35

La température a été fixée à  $27 \pm 1^\circ\text{C}$ .

Ce milieu est inoculé avec 0,25 ml d'une culture contenant approximativement 36 000 cellules par ml. La croissance des ciliés est suivie par spectrophotométrie, en mesurant la densité optique (absorbance) à 540 nm après 48 heures d'incubation.

Plusieurs critères ont guidé au choix des composés toxiques examinés. Tous sont disponibles dans le commerce avec une pureté suffisante (95 % et plus), ce qui ne nécessite pas une re-purification préalablement au test. Des précautions ont été observées afin d'assurer une diversité concernant, à la fois, les propriétés physico-chimiques et la position des substituants.

Les solutions stocks des divers composés toxiques, ont été préparées dans le diméthylsulfoxyde (DMSO) à des concentrations de 5, 10, 25 et 50 grammes par litre. Dans chaque cas, le volume de solution stock ajouté à chaque fiole est limité par la concentration finale de DMSO qui ne doit pas excéder 0,75 % (350 µl par fiole), quantité qui n'altère pas la reproduction de *Tétrahymena* [48,49].

# Coefficient de partage

## *Coefficient de partage*

I- Définition:

II- Mesure de coefficient de partage (logP)

II-1- Méthode du flacon agité

II- 2- Détermination par HPLC

III- Prédiction du coefficient de partage octanol /eau

III- 1- Méthodes fragmentales

III-2- Méthodes Des Contributions d'Atomes

III-3- Approches par Molécule -Entière

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

Le transport, le passage à travers les membranes, la bioaccumulation ou encore l'activité pharmacologique d'une molécule peuvent être conditionnés par son partage entre une phase lipidique et une phase aqueuse, c'est-à-dire son caractère hydrophile. Celui-ci peut être quantifié par le coefficient de partage octanol-eau, noté logP, qui mesure la solubilité différentielle d'un soluté dans ces deux solvants non miscibles [50].

$$\log P = \log \left( C_{\text{octanol}} / C_{\text{H}_2\text{O}} \right) \quad (\text{eq. 19})$$

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

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

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

La première est proportionnelle à la taille du soluté, et la seconde est associée aux groupes fonctionnels. Ainsi, la lipophilie peut être décomposée en deux contributions principales (Carrupt *et al*, 1997) [51]:

$$\text{Lipophilie} = \text{Hydrophobie} - \text{Polarité}. \quad (\text{eq. 20})$$

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

## **II- Mesure du coefficient de partage (logP):**

L'octanol est un solvant qui possède une polarité très similaire à celle des membranes lipidiques, c'est pourquoi il est utilisé pour évaluer le caractère lipophile d'une molécule, donc sa capacité à se bioaccumuler dans les organismes vivants.

### **II-1- Méthode du flacon agité :**

La plus classique et la plus fiable des méthodes de détermination du logP est la méthode dite du "flacon agité" [52].

Cette méthode consiste à mélanger une quantité connue de soluté dans un volume connu d'octanol et d'eau, puis de mesurer la distribution du soluté dans chaque solvant. La méthode la plus courante pour mesurer cette distribution est la spectroscopie UV/Visible.

#### **Cette méthode comporte des avantages :**

- Méthode la plus précise, pour la gamme la plus large de solutés (neutres comme chargés)
- La structure chimique n'a pas besoin d'être connue avant de commencer la procédure.

#### **Et des inconvénients :**

- Le temps nécessaire (30 minutes par échantillon)
- L'octanol et l'eau doivent être pré-mélangés et équilibrés (cela prend au moins 24 heures)
- Il faut attendre la solubilisation complète
- Le rapport concentration/réponse UV/Visible doit être linéaire dans la gamme de concentrations du soluté (Loi de Beer-Lambert)

### **II- 2- Détermination par HPLC :**

Une méthode plus rapide de détermination du logP utilise la chromatographie liquide haute performance (HPLC) [52]. Le logP d'un soluté peut être déterminé par corrélation entre son temps de rétention et celui de composés similaires avec un logP connu.

Cette méthode a l'avantage d'être rapide (5-20 minutes par échantillon), mais présente certains inconvénients :

- La structure chimique du soluté doit être connue avant de procéder à la détermination.
- Puisque le logP est habituellement déterminé par régression linéaire, on doit connaître le logP de plusieurs composés avec une structure similaire.
- Des classes de composés chimiques différentes ont des coefficients de corrélation différents, la comparaison entre les classes n'est pas valable.

### III- Prédiction du coefficient de partage octanol/eau:

Cette propriété physico-chimique peut être mesurée, mais ces mesures sont généralement longues et coûteuses. Par conséquent, différentes méthodes de prédiction du logP ont été mises au point, et il existe un nombre important de logiciels de prédiction de cette propriété. Ceux-ci s'appuient aussi bien sur des méthodes de contribution de groupes (ACD/LogP[53], KOWWIN [46], CLOGP [45],...) que sur des régressions multilinéaires à partir de descripteurs (VLogP [54]) ou sur des réseaux de neurones (AUTOLogP[55]).

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

La première méthode de calcul de logP a été développée par Hansch et Fujita (système  $\pi$ ) [56]. Les imperfections relevées ont conduit Rekker à développer la première approche de contribution fragmentaire [57-60]. Puisque la définition de fragment peut être ambiguë, (Broto et al) [61], suivis par d'autres, ont développé des systèmes de calcul basés sur des contributions atomiques. Toutes les méthodes qui divisent les molécules en sous-structures, sont appelées approches de sous-structure. L'utilisation exclusive des sous-structures atomiques caractérise des méthodes de contribution d'atomes, tandis que l'utilisation additionnelle de plus grands groupes est typique pour des méthodes fragmentaires. L'addition des contributions des sous-structures conduit finalement à la valeur de logP.

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

Une classification des logiciels de calcul de logP les plus importants est rapportée dans le tableau 4. Les sources des programmes respectifs et leur accès à l'Internet sont énumérés dans le tableau 5.

**Tableau 4:** Classification des logiciels de calcul de logP.

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

**Tableau 5 :** Adresses des différents logiciels de calcul de logP.

<b>Logiciel</b>	<b>Fournisseur</b>	<b>Internet</b>
AB/LogP	Advanced Pharma Algorithms	www.ap-algorithms.com
ABSOLV		www.sirius-analytical.com
ACD/LogP	Advanced Chemistry Development	www.acdkzbs.com
ALOGP98	Cerius-Z/MSI	www.msi.com
AUTOLOGP	J. Devillers	jde-ctis@imuginet.j
CLIP	B. Testa, P.-A. Carrupt	bernard.testa@ict.unil.ch
CLOGP	Daylight	www.daylight.com
	Biobyte	www.biobyte.com
HINT	Edusoft	eslc@uabiotech.com
Interactive Analysis (IA)		www.logp.com
KlogP	G. Klopman	www.multicase.com
KOWWIN	Syracuse Research Corporation	meylan@syrres.com
MOLCAD	Tripes	www.tripos.com
ProLogP	F. Darvas	www.cornpudnrg.com
ProPred		www.capec.kt.dtu.dk
QikProp		www.schrodinger.com
QLOGP	P. Buchwald, N. Bodor	buchwald@ufl.edu
QMPRPlus		www.simulations-plus.com
$\sum f$ FSYBYL	Tripes	www.tripos.com
SciLogP	SciVision	www.sciuision.com
SLIPPER		www.ipac.ac.m/qsar
SPARC		ibmlcLchem.uga.edu/sparc
TSAR	Oxford Molecular	www.oxmol.co.uk
XLOGP	L. Lai	lai@ipc.pku.edu.cn

### III- 1- Méthodes fragmentales :

La lipophilie d'un composé peut être estimée par la somme de la lipophilie de ses fragments et de termes de corrections.

$$\log P = \sum_{i=1}^n a_i * f_i + \sum_{j=1}^m b_j * F_j \quad (\text{eq. 21})$$

où:  $f_i$  = constante de lipophilie du fragment  $f_i$

$a_i$  = nombre de fragments  $f_i$

$F_j$  = facteur de correction  $F_j$

$b_j$  = nombre de facteurs de correction  $F_j$

Ces approches séparent les molécules en deux parties, l'une constituée par les atomes et l'autre par des groupes plus grands appelés fragments. Pourtant les molécules ne sont pas que des atomes et des fragments. Ainsi, les Méthodes fragmentales appliquent des règles de correction couplées par la connectivité moléculaire.

Puisque n'importe quelle méthode fragmentale a des avantages et des inconvénients, la fragmentation d'une molécule est peu arbitraire, ces méthodes peuvent définir les interactions électroniques des fragments géants. Ceci représente l'avantage principal de l'emploi des fragments. D'autre part, les méthodes fragmentales ne calculent pas les structures isomères et ne considèrent pas la flexibilité conformationnelle.

**Approche CLOGP:** Hansch et Leo [62-66] ont développé un système de fragmentation connu sous le nom de CLOGP [45], basé sur les principes du "constructionisme". Les valeurs des fragments de base ont été déduites à partir de données de  $\log P$  mesurées, d'un petit ensemble de molécules simples telles que l'hydrogène et le méthane. Cette méthode a été adaptée la première fois pour l'usage informatique par Chou et Jurs [67] ; elle a été lancée sur le marché sous l'appellation CLOGP.

**Approche LOGKowwin:** La méthode de contribution d'atome/fragment', est présentée par Meylan et Howard [68]; elle est disponible dans logiciel appelé KOWWIN [46]. C'est une approche réduite déduite de l'analyse de régression. Le modèle est défini par l'eq (22).

$$\log P = 0.229 + \sum f_k \cdot N_k + \sum F_j \cdot N_j \quad (\text{eq 22})$$

$N_k$  : est l'occurrence du fragment d'ordre  $i$  ou le type d'atome.

$N_j$  : est l'occurrence du j ème facteur de correction.

$\sum f$  : est un Système qui a été développé la première fois par le groupe de Rekker. Les valeurs logP expérimentales de composés organiques simples ont été employées pour déduire les valeurs fragmentaires pour analyse de régression. Le développement du système  $\sum f$  a été développé en trois phases. La première période a conduit au calcul de 126 valeurs de fragments, la deuxième est une révision du système original, la troisième phase a été consacrée aux modèles plus complexes comportant des hydrocarbures aliphatiques.

Les constantes d'hydrophobie ont été évaluées avec une première analyse de régression sur 1120 composés, sans utilisation de facteurs de correction. La dernière analyse est déduite d'une régression linéaire sur d'autres composés (1231) en corrélant les valeurs de logP expérimentales et celles estimées par le premier modèle de régression. Le logiciel est amélioré par 144 valeurs d'atome/fragment et 235 facteurs de correction qui sont énumérés dans la version 1.54.

### III-2- Méthodes De Contribution d'Atome

La plupart des méthodes de contribution d'atome fonctionnent sans facteurs de correction (Eq. 23). Où  $n_i$ : nombre d'atomes de type  $i$  ;  $a_i$  : contribution d'un atome de type  $i$ .

$$\log P = \sum n_i \cdot a_i \quad (\text{eq. 23})$$

Dans ces méthodes plusieurs améliorations ont été apportées, mais 2 problèmes restent posés: le premier concerne le grand nombre d'atomes nécessaires pour décrire un ensemble de molécules, le second a trait aux structures isomères ces méthodes ne prennent pas en charge la flexibilité de conformation.

**Approche Ghose/Crippen:** Le groupe de Crippen [69,72] a développé une " *procédure basée sur les contributions atomiques pure* " qui applique exclusivement les contributions d'atomes et évite les facteurs de correction (Eq. 24), où  $N_k$  est l'occurrence du k ème type d'atome).

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

Le carbone, l'hydrogène, l'oxygène, l'azote, le soufre, et les halogènes sont classés dans 110 types d'atome ; après plusieurs révisions, le nombre de classifications d'atome qui arrive jusqu'à 120 [69] est obtenu à partir d'un ensemble de calibration de 893 structures. L'hydrogène et les halogènes sont classés par l'état d'hybridation et d'oxydation du carbone

auquel ils sont liés ; les atomes de carbone sont classés selon leur état d'hybridation et la nature chimique des atomes voisins. La complexité de cette procédure de classification est certifiée par un total de 44 types de carbone. L'approche de Ghose/Crippen qui utilise largement la contribution d'atomes, a été utilisée par le logiciel qu'on a employé pour le calcul de l'hydrophilie notée ALOGP par les développeurs de ce logiciel.

### III-3- Approches Molécule- Entière:

#### III-3-1 Méthodes MLP :

Les approches de Molécule- Entière utilisent les descriptions de la molécule entière, pour calculer log P. Ces modèles essaient d'éviter les imperfections des approches fragmentaires, telles que la simplification des effets stériques, et l'impossibilité de calculer logP des structures pour lesquelles les valeurs des fragments n'existent pas, ou le problème soulevé par le cas des isomères.

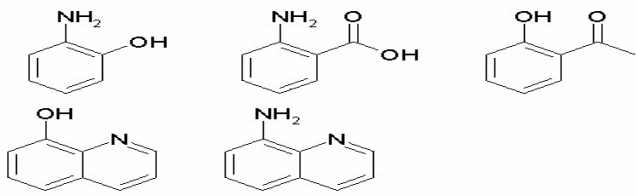
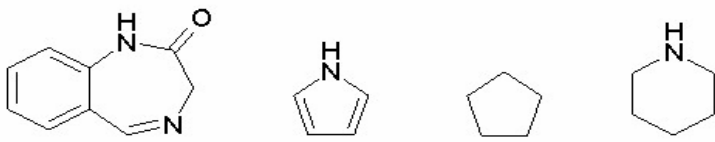
#### III-3-2 Base Topologique :

La topologie qui concerne les propriétés et les relations spatiales, est constante dans un changement de forme ou de taille. De plus la connectivité traite des liaisons entre les atomes. Si les molécules sont bien définies, les indices de connectivité moléculaires peuvent être déduits directement à partir de la structure moléculaire employée pour prédire logP.

**MLOGP:** Pour calculer logP par une "méthode simple" Moriguchi a employé une analyse de régression multiple d'un ensemble de 1230 molécules organiques comprenant, les composés aliphatiques, aromatiques, et hétérocycliques [73]. L'équation finale de régression, eq. (25) implique 13 paramètres y compris la somme des hydrophobies atomiques, la somme des hydrophilies atomiques, les effets de proximité, les liaisons insaturées, les propriétés amphotères, et les fonctions spéciales telles que la présence d'un azote quaternaire, le nombre de groupes nitro, ou un simulacre pour la présence de la p-lactame. Un coefficient de corrélation acceptable de 0.952 a été obtenu ; le logP des 1230 composés a été calculé avec un écart type de 0,411,  $s = 0.422$  [44].

$$\begin{aligned} M\log P = & -1.041 + 1.244(\mathbf{CX})^{0.6} - 1.017(\mathbf{NO})^{0.9} + 0.406(\mathbf{PRX}) - 0.145(\mathbf{UB})^{0.8} + 0.511(\mathbf{HB}) \\ & + 0.268(\mathbf{POL}) - 2.215(\mathbf{AMP}) + 0.912(\mathbf{ALK}) - 0.392(\mathbf{RNG}) - 3.684(\mathbf{QN}) + 0.474(\mathbf{NO2}) + \\ & 1.582(\mathbf{NCS}) + 0.733(\mathbf{BLM}) \end{aligned} \quad (\text{eq. 25})$$

**Tableau 6** : Signification des paramètres statistiques utilisés dans l'eq. (25).

Paramétrr	Type	Description
<b>CX</b>	N	La somme des poids des atomes de carbone et d'halogènes; les poids sont: 0.5 pour F, 1.0 pour C et Cl, 1.5 pour le Br, et 2.0 pour I.
<b>NO</b>	N	Nombre total de N et de O.
<b>PRX</b>	N	Effet de proximité de N/O: 2 pour X-Y et 1 pour X-a-y (X, Y :N et/ou O ;A :C, S, ou P ;- : Liaison saturée ou insaturée) avec une correction (-1) pour -CON< et -SO <sub>2</sub> N<
<b>UB</b>	N	Nombre de liaisons insaturées comprenant les liaisons semi-polaires tels que des N-oxydes et des sulfoxydes, sauf dans NO <sub>2</sub> .
<b>HB</b>	D	Variable factice pour la présence de liaisons hydrogène intramoléculaire comme l'ortho-OH et - CO-R, -OH et - NH <sub>2</sub> , - NH <sub>2</sub> et - COOH, ou 8-OH/NH <sub>2</sub> pour les quinolines, 5 ou le 8-OH/NH <sub>2</sub> pour les quinoxalines, etc...
		
<b>POL</b>	N	Nombre de substituants polaires aromatiques (substituants aromatiques à l'exclusion d'Ar-C(X)(Y) - et Ar-C(X)=C ;X, Y :C et/ou H).Limite supérieure = 4.
<b>AMP</b>	N	Propriété amphotère ; a-aminoacid = 1, acide aminobenzoïque = 0.5, acide pyridinecarboxylique = 0.5.
<b>ALK</b>	D	Variable factice pour l'alcane, l'alcène, le cycloalcane, ou le cycloalcene (hydrocarbures avec 0 ou 1 lien doubles).
<b>RNG</b>	D	Variable factice pour la présence de structures cycliques, sauf benzène et les structures heteroaromatique.
		
<b>QN</b>	N	Azote quaternaire > N< : 1 ;N-oxyde :0.5.
<b>NO2</b>	N	Nombre de groupes nitro.
<b>NCS</b>	N	Isothiocyanate (- N=C=S) :1.0 ;sulfocyanate (- S-C#N) :0.5.
<b>BLM</b>	D	Variable factice pour la présence de la . -lactame.

# Interprétation du modèle

## *Action- Toxicité des phénols*

1 Introduction

2 Mécanisme d'action

## Action- Toxicité des phénols:

### 1 Introduction:

Les phénols représentent une classe importante parmi les produits chimiques anthropogènes. Ils sont capables de s'ioniser sous les conditions physiologiques. La relation entre la forme ionisée et la forme neutre est déterminée par ces valeurs de  $pK_a$  à pH connu. Le groupe hydroxyle phénolique peut prendre une valeur du  $pK_a$  dans un intervalle large, selon le nombre, la position, et le caractère des substituants du cycle benzylique. Lorsque le test de toxicité de divers phénols est effectué dans un milieu à pH fixe, les résultats obtenus ne peuvent pas être directement comparés, puisque le rapport des espèces ionisées et non-ionisées change.

Pendant la dernière décennie, plusieurs articles indiquent que les composés polaires, contenant les phénols, sont sensiblement plus toxiques par rapport à une prédiction qui a été faite à partir d'un modèle basé sur des composés narcotiques apolaires [74-77]. D'après le degré d'électronégativité, ce groupe de composés sont considérés, soit comme narcotiques polaires soit comme découplants dans la chaîne respiratoire [74]. Les précédents travaux [74,77], [78-81], montrent l'importance d'utiliser un deuxième paramètre électronique/dissociation avec l'hydrophobie, pour prédire l'activité biologique des phénols.

Dans la présente étude, 93 phénols ont été choisis pour examiner l'influence de l'ionisation du groupe phénolique hydroxyle et de l'hydrophobie sur la toxicité des ces composés vis-à-vis de cellules *Tetrahymena pyriformis* par l'approche du QSAR. Tous les dérivés phénoliques ayant le groupe hydroxyle comme seule partie ionisable, sont inclus dans l'analyse. L'approche fournit également une indication sur les deux mécanismes d'action des composés narcotiques polaires et les découplants respiratoires, contribuent à la toxicité de différents dérivés du phénol dans le cas de *Tetrahymena pyriformis*.

Dans cette partie nous allons montrer l'importance d'assurer que toutes les variables dépendantes du modèle QSAR sont associées par un seul mécanisme d'action. Si on ne peut pas vérifier cette hypothèse, il ne sera pas possible de comparer le modèle QSAR pour les autres actions sur les différents systèmes biologiques, et pour cela une telle information ne peut pas être employée pour

développer le domaine des interactions biochimiques. Pour illustrer le problème une étude a été faite sur l'action d'un grand ensemble de phénols agissant sur *Tetrahymena pyriformis* par Cronin et Schultz [82].

## 2 Mécanisme d'action:

Généralement les composés qui sont à la fois non- ioniques et non- réactifs sont considérés comme des composés moins toxiques [79]), reconnue par une perturbation au niveau de la membrane ou bien par le mode d'action narcotinique. La narcose, comme "état inverse a l'arrestation de l'activité cytoplasmique [83], est le résultat de la distribution d'un composé toxique dans l'organisme. Les premiers modèles du QSAR sur les poissons concédèrent les phénols qui sont plus solubles dans l'eau, de faibles acidités ou basicités comme des produits narcotiques polaires produisant le syndrome narcotique de type I ou II [76]. La toxicité des produits narcotiques polaires est plus grande de celle des produits narcotiques apolaires parce que la valeur de dipolarité et l'acidité de la liaison hydrogène donneur des produits narcotiques polaires est plus élevée [84].

La toxicité des phénols autre que la narcose, est causée par le couplage entre la réaction de phosphorylation de l'ATP et les oxydoréductions de la chaîne respiratoire mitochondriale [85,81,86,87]. L'acide faible découplé par la phosphorylation oxydative est la conséquence de l'inhibition du couplage entre le porteur d'électron et les réactions de phosphorylation [88], conduisant à l'inhibition de la synthèse de l'ATP, sans faire un effet sur la chaîne respiratoire et le synthase de triphosphate adénosine ATP. Les acides faibles sont considérés comme des déccouplants par leur action protonophorique sur l'hydrogène ion-imperméable de membrane de la mitochondrie [89].

Certain QSAR indique que l'activité causé par le découplage est typiquement décrit comme une fonction linéaire entre l'hydrophobie mesurée par le coefficient de partage 1-octanol/eau et le pouvoir de retirer l'électron représenté par la constante de dissociation  $pK_a$  [85,81,86]. Cependant, l'utilisation d'un ensemble d'acides faibles avec une gamme d'hydrophobie et d'acidité large a montrée l'dépendance de l'activité de découplage dans les mitochondries du foie de rat [90] ou bien dans les bactéries *Paracoccus denitrificans* [91].

# PARTIE EXPERIMENTALE

Les valeurs des coefficients de partage (AlogP, ClogP, MlogP, logK<sub>owwin</sub>) qu'on a utilisées sont calculées de façons différentes avec les méthodes présentées dans la partie théorique. Les valeurs AlogP et MlogP sont calculées à l'aide du logiciel DRAGON [44], par contre les valeurs ClogP sont calculées à l'aide du logiciel ChemDraw (CambridgeSoft), celles de logK<sub>owwin</sub> avec le logiciel KOWWIN, et les valeurs logK<sub>ow</sub> sont tirées du livre "Practical Application of QSAR in Environmental Chemistry and Toxicology" [92] ; on prend le modèle logK<sub>ow</sub> comme modèle de référence.

Le modèle général qui a été avancé la première fois par Hansch dans les années 60, propose d'expliquer l'activité biologique par l'hydrophobie et un descripteur électronique. L'un des descripteurs électroniques les plus utilisés est la constante d'ionisation pKa. Shultz propose un modèle de l'activité biologique construit sur deux descripteurs, le premier est le paramètre d'hydrophobie, le deuxième est le pKa ; l'activité est représentée par la toxicité caractérisée par la concentration d'inhibition 50 % de la croissance (CIC50) d'une population de *Tetrahymena pyriformis*.

$$((pCIC_{50}) = f(\log P, pKa) \quad (\text{eq. 26})$$

D'après cette équation, on peut avoir 04 modèles avec différents logP, les valeurs de pKa exploitées dans le présent travail sont tirées de [93].

### I- Modèle numéro 01: MLOGP

Dans ce chapitre les deux ensembles de validation (19 composés) et de calibration (74 composés) sont obtenus par algorithme DUPLEX.

Matrice de corrélation:

	pCIC <sub>50</sub>	MlogP
MlogP	0,857	
	0,000	
pKa	-0,070	0,169
	0,551	0,150

#### I-1-Equation et analyse de régression :

L'équation du modèle calculé est la suivante :

$$\text{pCIC}_{50} = 0,063 (\pm 0,3413) + 0,824 (\pm 0,05182) \text{MlogP} - 0,141 (\pm 0,03582) \text{pKa} \quad (\text{eq. 27})$$

$$n = 74 \quad R^2 = 78,19 \% \quad Q^2 = 76,34 \% \quad F = 127,25 \quad s = 0,3309$$

$$R^2_{\text{adj}} = 77,57 \% \quad \text{EQMC} = 0,324 \quad \text{EQMP} = 0,338$$

#### I-2- Analyse des résidus et diagnostics d'influence :

Le calcul effectué par le logiciel MobyDigs [94] entre autres des résidus de prédiction standardisés, que nous noterons  $e_{i \text{ std}}$ , donne les résultats qui apparaissent respectivement dans la colonne 5 du tableau 7.

## Partie expérimentale

**Tableau 7- :** Valeurs des  $pCIC_{50}$  observés ( $pCIC_{50\text{ obs}}$ ) et calculés, prédites ( $pCIC_{50\text{ cal-pred}}$ ), des résidus standardisés ( $e_{i\text{ std}}$ ) ainsi que des leviers ( $h_{ij}$ ) (Cas de MLOGP).

N°	Nom du Composé	$pCIC_{50\text{ obs}}$	$pCIC_{50\text{ cal-pred}}$	$e_{i\text{ std}}$	$h_{ij}$
01*	2,4,6-tribromophénol	2,050	2,273	0,758	0,211
02*	4-acetoamidophénol	-0,820	-0,490	1,026	0,057
03*	2,6-diphénylphénol	2,110	2,301	0,616	0,124
04*	2-nitrophénol	0,670	0,155	-1,623	0,080
05*	2-tertbutyl-4-méthylphénol	1,300	1,014	-0,894	0,065
06*	3-chloro-4-fluorophénol	0,840	0,910	0,214	0,020
07*	2,3-diméthylphénol	0,120	0,410	0,885	0,023
08*	2-hydroxybenzamide	-0,240	-0,519	-0,875	0,074
09*	4-chloro-2-isopropyl-5-méthylphénol	1,860	1,427	-1,339	0,044
10*	2,3-dichlorophénol	1,270	1,242	-0,088	0,065
11*	2,5-dichlorophénol	1,130	1,242	0,350	0,065
12*	méthyl-4-hydroxybenzoate	0,080	-0,006	-0,265	0,028
13*	4-cyclopentylphénol	1,290	0,899	-1,195	0,020
14*	4-bromo-6-chloro-2-méthylphénol	1,280	1,542	0,816	0,061
15*	4-acétylphénol	-0,090	0,155	0,756	0,044
16*	4-chlorophénol	0,550	0,485	-0,199	0,014
17*	4-hexyloxyphénol	1,650	0,838	-2,497	0,034
18*	2-fluorophénol	0,280	0,440	0,489	0,021
19*	phénol	-0,430	-0,106	0,995	0,032
20	2,4,6-trichlorophénol	1,700	1,898	0,795	0,174
21	2-méthylphénol	-0,270	0,147	1,311	0,026
22	2-éthylphénol	0,180	0,431	0,780	0,020
23	2-allylphénol	0,350	0,657	0,951	0,016
24	2-isopropylphénol	0,800	0,664	-0,428	0,024
25	2-(tert)butylphénol	1,240	0,882	-1,138	0,033
26	2-phénylphénol	1,090	1,292	0,641	0,032
27	2-chlorophénol	0,280	0,609	1,029	0,023
28	2-bromophénol	0,500	0,762	0,824	0,026
29	2-cyanophénol	0,030	0,030	0,001	0,099
30	2-hydroxybenzaldehyde	0,480	-0,162	-2,095	0,050
31	2-acétylphénol	0,080	0,415	1,036	0,015
32	2-hydroxybenzyl alcohol	-0,950	-0,531	1,385	0,059
33	2-hydroxybenzaldoxime	-0,250	-0,404	-0,503	0,049
34	3-méthylphénol	-0,060	0,184	0,760	0,022
35	3-éthylphénol	0,230	0,449	0,680	0,018
36	3-isopropylphénol	0,610	0,706	0,298	0,019
37	3-tertbutylphénol	0,730	0,953	0,698	0,024
38	3-phénylphénol	1,350	1,281	-0,220	0,032
39	3-fluorophénol	0,470	0,361	-0,337	0,016

## Partie expérimentale

Tableau-7: Suite.

N°	Nom du Composé	pCIC <sub>50</sub> obs	pCIC <sub>50</sub> calc-pred	e <sub>i</sub> std	h <sub>ii</sub>
40	3-chlorophénol	0,960	0,531	-1,325	0,015
41	3-iodophénol	1,120	0,838	-0,879	0,019
42	3-nitrophénol	0,510	-0,006	-1,663	0,042
43	3-cyanophénol	-0,070	-0,200	-0,420	0,046
44	3-hydroxybenzaldahyde	0,080	-0,255	-1,079	0,042
45	3-acetylphénol	-0,380	-0,006	1,177	0,027
46	3-methoxyphénol	-0,140	-0,272	-0,425	0,039
47	3-hydroxybenzyl alcohol	-1,040	-0,519	1,721	0,057
48	methyl-3-hydroxybenzoate	-0,050	-0,029	0,067	0,028
49	ethyl-3-hydroxybenzoate	0,480	0,249	-0,718	0,019
50	4-ethylphénol	0,210	0,459	0,771	0,017
51	4-propylphénol	0,640	0,678	0,117	0,022
52	4-isopropylphénol	0,470	0,678	0,649	0,022
53	4-secbutylphénol	0,980	0,927	-0,166	0,027
54	4-tertbutylphénol	0,910	0,937	0,085	0,026
55	4-tertpentylphénol	1,230	1,168	-0,198	0,036
56	4-phenylphénol	1,380	1,292	-0,280	0,032
57	4-hydroxyphenylmethane	1,200	0,169	-3,227	0,023
58	4-tert octylphénol	2,090	1,894	-0,672	0,081
59	a,a,a-trifluoro-4-cresol	0,620	0,989	1,158	0,025
60	4-fluorophénol	0,020	0,276	0,797	0,018
61	4-bromophénol	0,680	0,637	-0,134	0,014
62	4-iodophénol	0,850	0,792	-0,178	0,016
63	4-cyanophénol	0,520	-0,108	-2,081	0,060
64	4-hydroxybenzaldehyde	0,270	-0,060	-1,113	0,071
65	4'-hydroxypropiophenone	0,060	0,303	0,758	0,021
66	4-hydroxybenzophenone	1,020	1,259	0,759	0,033
67	4-hydroxybenzylcyanide	-0,380	-0,274	0,341	0,039
68	4-methoxyphénol	-0,140	-0,350	-0,683	0,049
69	4-ethoxyphénol	0,010	-0,117	-0,409	0,041
70	4-butoxyphénol	0,700	0,380	-1,011	0,029
71	4-heptyloxyphénol	2,030	1,062	-3,117	0,042
72	4-benzyloxyphénol	1,040	1,069	0,095	0,042
73	4-hydroxybenzamide	-0,780	-0,642	0,466	0,069
74	4-hydroxyphenethyl alcohol	-0,830	-0,256	1,843	0,040
75	ethyl-4-hydroxybenzoate	0,570	0,273	-0,925	0,021
76	2,6-difluorophénol	0,400	0,975	1,908	0,060
77	2,5-dimethylphénol	0,010	0,411	1,254	0,023
78	3,4-dimethylphénol	0,120	0,414	0,918	0,022
79	3,5-dimethylphénol	0,110	0,438	1,019	0,019

## Partie expérimentale

Tableau-7: Suite et fin.

N°	Nom du Composé	pCIC <sub>50</sub> obs	pCIC <sub>50</sub> calc-pred	e <sub>i std</sub>	h <sub>ii</sub>
80	2-chloro-5-methylphénol	0,640	0,884	0,768	0,026
81	2-bromo-4-methylphénol	0,790	0,999	0,657	0,026
81	4-chloro-2-methylphénol	0,790	0,999	0,657	0,026
82	4-chloro-3-methylphénol	0,700	0,726	0,081	0,015
83	3,5-dichlorophénol	0,800	0,743	-0,176	0,015
84	2,4-dibromophénol	1,560	1,157	-1,298	0,042
85	3,4,5-trimethylphénol	1,400	1,461	0,204	0,066
86	2,3,5-trimethylphénol	0,930	0,648	-0,886	0,026
87	4-chloro-3,5-dimethylphénol	0,360	0,651	0,915	0,026
88	4-bromo-2,6-dichlorophénol	1,200	0,982	-0,680	0,020
89	2,4,5-trichlorophénol	1,780	2,017	0,965	0,181
90	4-bromo-2,6-dimethylphénol	2,100	1,754	-1,238	0,106
91	6-tertbutyl-2-4-dimethylphénol	1,280	1,065	-0,675	0,026
92	2,6ditertbutyl-4-methylphénol	1,250	1,205	-0,157	0,089
93	ethyl-4-hydroxybenzoate	1,790	1,733	-0,235	0,183

\* ensemble de validation

### Diagramme de Williams :

La figure -2 est une représentation des valeurs des résidus de prédiction standardisés  $e_{i \text{ std}}$  des composés pour les deux sous-groupes (calibration et validation) en fonction de leurs leviers respectifs,

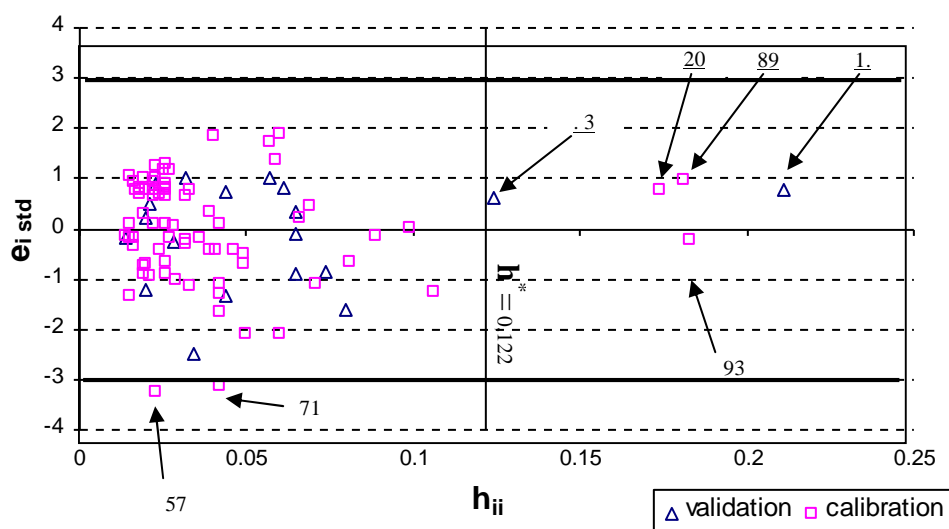


Figure -02: Diagramme de Williams pour le modèle 01.

## Partie expérimentale

Le diagramme montre qu' on a deux points aberrants [57(4-hydroxyphenylmethane), 71(4-heptyloxyphénol)] caractérisés par les valeurs des résidus standardisés, supérieures à 3 unités d'écart type ( $\pm 3$ ), et cinq points influents [3 (2,6-diphénylphénol), 20 (2,4,6-trichlorophénol), 89 (2,4,5-trichlorophénol), 1(2,4,6-tribromophénol), 93 (ethyl-4-hydroxybenzoate)] dont les  $h_i$  sont supérieurs à la valeur

$$\text{critique } h^* = \frac{3*(p+1)}{n} = \frac{3*(2+1)}{74} = 0,122.$$

### I-3- Qualité de l'ajustement :

La qualité de l'ajustement peut être appréciée dans la figure -3 par les deux droites d'ajustement dont les équations sont :

- Pour l'ensemble de calibration

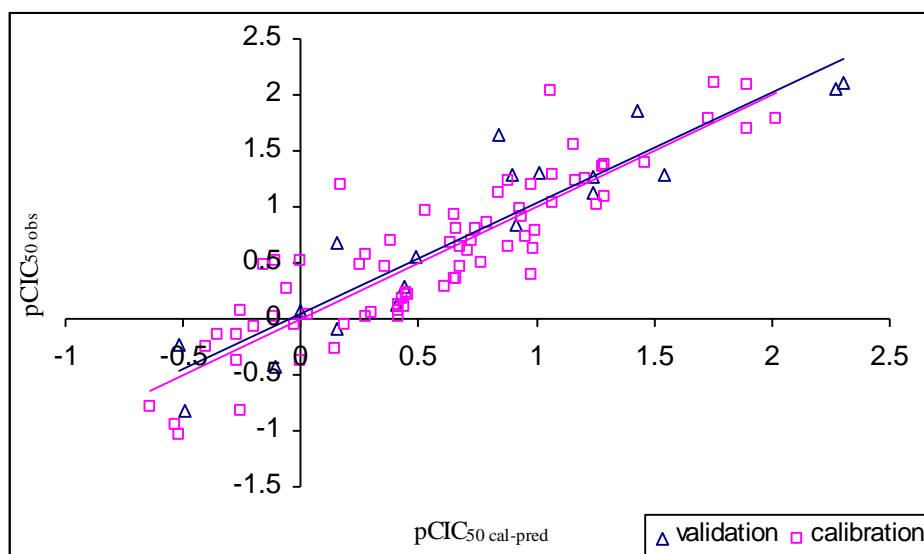
$$pCIC_{50 \text{ obs}} = -0,0000028 + 1,00000 pCIC_{50 \text{ calc}} \quad (\text{eq, 28})$$

$$s = 0,328640 \quad R^2 = 78,2 \% \quad R^2_{\text{ajust}} = 77,9 \%$$

- Pour l'ensemble de validation

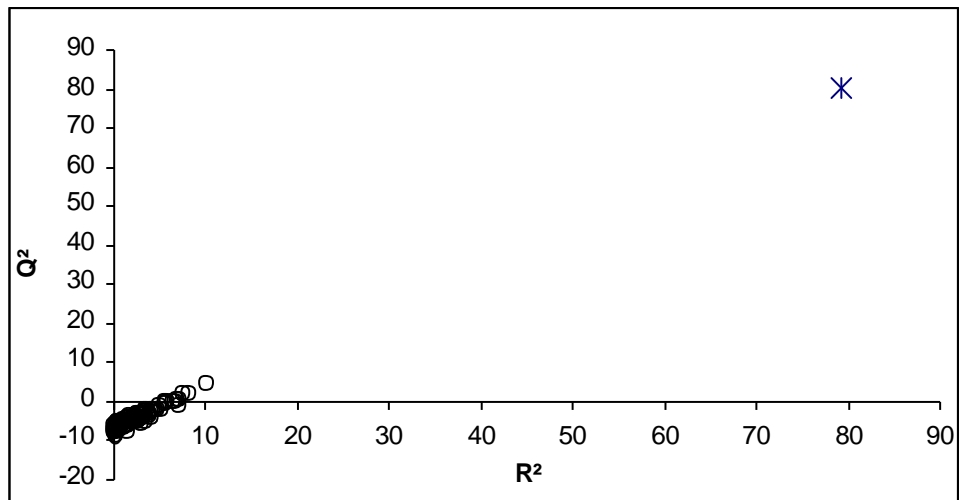
$$pCIC_{50 \text{ obs}} = 0,0463419 + 0,986691 pCIC_{50 \text{ pred}} \quad (\text{eq, 29})$$

$$s = 0,340055 \quad R^2 = 85,5 \% \quad R^2_{\text{ajust}} = 84,6 \%$$



**Figure -03 :** Droites d'ajustement pour le modèle 01,

Par le test de randomisation (figure 4), nous nous sommes assuré qu'une relation structure-activité réelle a été établie par notre modèle.



**Figure -04:** *Test de randomisation pour le modèle 01.*

Les 100 modèles pour lesquels nous avons randomisé les valeurs du  $pCIC_{50}$  ont des valeurs de  $Q^2$  faibles et des valeurs du coefficient de corrélation multiple ( $R^2$ ) petites. Seul le modèle avec le vecteur réel (étoile) offre des valeurs élevées pour les deux statistiques représentées dans la figure 7.

### II- Modèle numéro 02: ALOGP

Matrice de corrélation:

	pCIC <sub>50</sub>	AlogP
AlogP	0,844	
pKa	0,000	0,129
	-0,207	0,272
	0,076	

#### II-1- Equation et analyse de régression :

L'équation de régression est

$$\text{pCIC}_{50} = (\pm 0,3322)0,992 + (\pm 0,04040)0,694 \text{ AlogP} - (\pm 0,03501)0,219 \text{ pKa} \quad (\text{eq. 30})$$

$$n = 74 \quad R^2 = 81,42 \% \quad Q^2 = 79,86 \% \quad F = 155,59 \quad s = 0,3021$$

$$R^2_{\text{adj}} = 80,9\% \quad \text{EQMC} = 0,296 \quad \text{EQMP} = 0,308$$

#### II-2- Analyse des résidus et diagnostiques d'influence :

Les valeurs des résidus de prédiction standardisés, et les valeurs de  $h_{ii}$  sont présentées dans le tableau 8 dont les colonnes 3 et 4 reproduisent les valeurs expérimentales et calculées, prédites de pCIC<sub>50</sub> des composés considérés.

## Partie expérimentale

**Tableau - 8 :** Valeurs des  $pCIC_{50}$ , observés ( $pCIC_{50\text{ obs}}$ ) et calculés, prédites ( $pCIC_{50\text{ cal-pred}}$ ), des résidus standardisés ( $e_{i\text{ std}}$ ) ainsi que des leviers ( $h_{ii}$ ) (Cas de ALOGP).

N°	Nom du Composé	$pCIC_{50\text{ obs}}$	$pCIC_{50\text{ cal-pred}}$	$e_{i\text{ std}}$	$h_{ii}$
01*	2,6ditertbutyl-4-methylphénol	1,790	1,610	-0,676	0,223
02*	2-cyanophénol	0,030	0,466	1,520	0,097
03*	4-acetoamidophénol	-0,820	-0,744	0,261	0,076
04*	4-chloro-3,5-dimethylphénol	1,200	1,088	-0,375	0,026
05*	4-bromo-2,6-dichlorophénol	1,780	2,117	1,229	0,175
06*	2,4-dibromophénol	1,400	1,391	-0,030	0,056
07*	2-fluorophénol	0,280	0,310	0,099	0,024
08*	2-tertbutyl-4-methylphénol	1,300	0,894	-1,401	0,078
09*	2,6-diphenylphénol	2,110	2,011	-0,345	0,099
10*	2-methylphénol	-0,270	0,170	1,476	0,027
11*	4-heptyloxyphénol	2,030	1,596	-1,504	0,088
12*	4-butoxyphénol	0,700	0,668	-0,106	0,035
13*	2-bromo-4-methylphénol	0,790	1,036	0,826	0,025
14*	4-hydroxybenzylcyanide	-0,380	-0,066	1,056	0,029
15*	4-hydroxybenzamide	-0,780	-0,634	0,502	0,071
16*	4-chlorophénol	0,550	0,475	-0,251	0,014
17*	4-cyanophénol	0,520	0,252	-0,911	0,051
18*	4-benzyloxyphénol	1,040	1,156	0,394	0,056
19*	2,6-difluorophénol	0,400	0,718	1,084	0,060
20	phénol	-0,430	-0,108	1,120	0,032
21	2-ethylphénol	0,180	0,499	1,093	0,022
22	2-allylphénol	0,350	0,606	0,871	0,018
23	2-isopropylphénol	0,800	0,630	-0,586	0,028
24	2-(tert)butylphénol	1,240	0,725	-1,804	0,037
25	2-phenylphénol	1,090	1,041	-0,168	0,022
26	2-chlorophénol	0,280	0,667	1,327	0,023
27	2-bromophénol	0,500	0,747	0,851	0,025
28	2-nitrophénol	0,670	0,524	-0,549	0,081
29	2-hydroxybenzaldehyde	0,480	0,086	-1,397	0,044
30	2-acetylphénol	0,080	-0,114	-0,675	0,034
31	2-hydroxybenzamide	-0,240	-0,444	-0,764	0,079
32	2-hydroxybenzyl alcohol	-0,950	-0,512	1,582	0,057
33	2-hydroxybenzaldoxime	-0,250	-0,200	0,177	0,036
34	3-methylphénol	-0,060	0,227	0,981	0,021
35	3-ethylphénol	0,230	0,528	1,016	0,020
36	3-isopropylphénol	0,610	0,696	0,294	0,022
37	3-tertbutylphénol	0,730	0,834	0,359	0,025
38	3-phenylphénol	1,350	1,023	-1,119	0,022

## Partie expérimentale

**Tableau - 8 : Suite.**

39	3-fluorophénol	0,470	0,187	-0,966	0,020
40	3-chlorophénol	0,960	0,547	-1,398	0,015
41	3-iodophénol	1,120	0,536	-1,987	0,018
42	3-nitrophénol	0,510	0,275	-0,822	0,035
43	3-cyanophénol	-0,070	0,110	0,627	0,035
44	3-hydroxybenzaldéhyde	0,080	-0,059	-0,484	0,034
45	3-acétylphénol	-0,380	-0,114	0,930	0,034
46	3-méthoxyphénol	-0,140	-0,046	0,326	0,028
47	3-hydroxybenzyl alcool	-1,040	-0,493	1,973	0,055
48	méthyl-3-hydroxybenzoate	-0,050	-0,038	0,041	0,030
49	éthyl-3-hydroxybenzoate	0,480	0,230	-0,854	0,021
50	4-éthylphénol	0,210	0,543	1,134	0,019
51	4-propylphénol	0,640	0,794	0,531	0,029
52	4-isopropylphénol	0,470	0,652	0,627	0,026
53	4-secbutylphénol	0,980	0,968	-0,040	0,034
54	4-tertbutylphénol	0,910	0,810	-0,344	0,027
55	4-tertpentylphénol	1,230	1,111	-0,418	0,040
56	4-cyclopentylphénol	1,290	1,111	-0,622	0,030
57	4-phenylphénol	1,380	1,041	-1,161	0,022
58	4-hydroxyphenylméthane	1,200	0,205	-3,413	0,023
59	4-tertoctylphénol	2,090	2,144	0,214	0,116
60	a,a,a-trifluoro-4-cresol	0,620	0,819	0,677	0,020
61	4-fluorophénol	0,020	0,056	0,124	0,025
62	4-bromophénol	0,680	0,553	-0,430	0,014
63	4-iodophénol	0,850	0,466	-1,302	0,015
64	4-hydroxybenzaldéhyde	0,270	0,243	-0,099	0,068
65	4-acétylphénol	-0,090	0,136	0,811	0,053
66	4'-hydroxypropiophénone	0,060	0,423	1,237	0,020
67	4-hydroxybenzophénone	1,020	1,106	0,296	0,024
68	4-méthoxyphénol	-0,140	-0,166	-0,090	0,038
69	4-éthoxyphénol	0,010	0,011	0,002	0,037
70	4-hexyloxyphénol	1,650	1,279	-1,357	0,064
71	4-hydroxyphenéthyl alcool	-0,830	-1,214	-1,564	0,129
72	méthyl-4-hydroxybenzoate	0,080	-0,003	-0,288	0,030
73	éthyl-4-hydroxybenzoate	0,570	0,267	-1,036	0,022
74	2,3-diméthylphénol	0,120	0,488	1,266	0,026
75	2,5-diméthylphénol	0,010	0,490	1,650	0,026
76	3,4-diméthylphénol	0,120	0,494	1,286	0,025
77	3,5-diméthylphénol	0,110	0,531	1,440	0,021

## Partie expérimentale

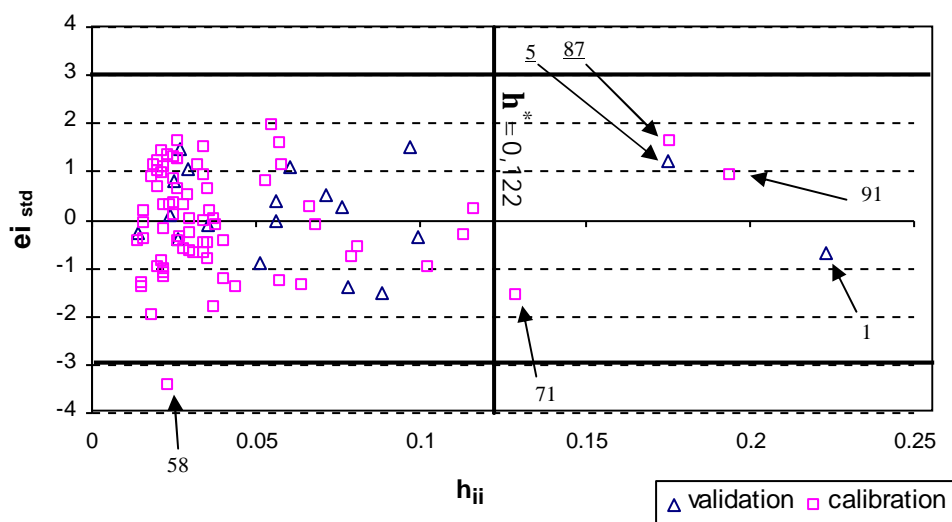
**Tableau - 8 :** Suite et fin.

N°	Nom du Composé	pCIC <sub>50</sub> obs	pCIC <sub>50</sub> calc-pred	e <sub>i std</sub>	h <sub>ii</sub>
78	3-chloro-4-fluorophénol	0,840	0,719	-0,411	0,016
79	2-chloro-5-méthylphénol	0,640	1,005	1,256	0,026
80	2,5-dichlorophénol	1,130	1,005	-0,431	0,026
81	2,3-dichlorophénol	1,270	1,340	0,256	0,066
82	4-chloro-2-méthylphénol	0,700	0,759	0,202	0,016
83	4-chloro-3-méthylphénol	0,800	0,786	-0,049	0,016
84	3,5-dichlorophénol	1,560	1,209	-1,237	0,040
85	3,4,5-triméthylphénol	0,930	0,790	-0,489	0,035
86	2,3,5-triméthylphénol	0,360	0,794	1,515	0,034
87	2,4,6-trichlorophénol	1,700	2,070	1,639	0,176
88	2,4,5-trichlorophénol	2,100	1,847	-0,985	0,102
89	4-bromo-6-chloro-2-méthylphénol	1,280	1,600	1,161	0,058
90	4-bromo-2,6-diméthylphénol	1,280	1,080	-0,693	0,031
91	2,4,6-tribromophénol	2,050	2,252	0,923	0,194
92	4-chloro-2-isopropyl-5-méthylphénol	1,860	1,509	-1,269	0,057
93	6-tertbutyl-2-4-diméthylphénol	1,250	1,167	-0,327	0,113

\* ensemble de validation

### Diagramme de Williams :

La figure-5 est une représentation des valeurs des résidus de prédiction standardisés  $e_{i \text{ std}}$  des composés pour les deux sous-groupes (calibration et validation) en fonction de leurs leviers respectifs.



**Figure -05:** Diagramme de Williams pour le modèle 02.

Le diagramme montre qu'on a un point aberrant [58 (4-hydroxyphenylmethane)] caractérisé par la valeur de son résidu de prédiction standardisé, supérieure à 3 ; et cinq points influents [71 (4-hydroxyphenethyl alcohol), 5(4-bromo-2,6-dichlorophénol), 87 (2,4,6-trichlorophénol), 91 (2,4,6-tribromophénol), 1 (2,6ditertbutyl-4-methylphénol) ]qui sont supérieurs à la valeur critique  $h^* = 0,122$ .

### II-3- Qualité de l'ajustement:

La qualité de l'ajustement peut être appréciée dans la figure 9 par les deux droites d'ajustement dont les équations sont :

- Pour l'ensemble de calibration

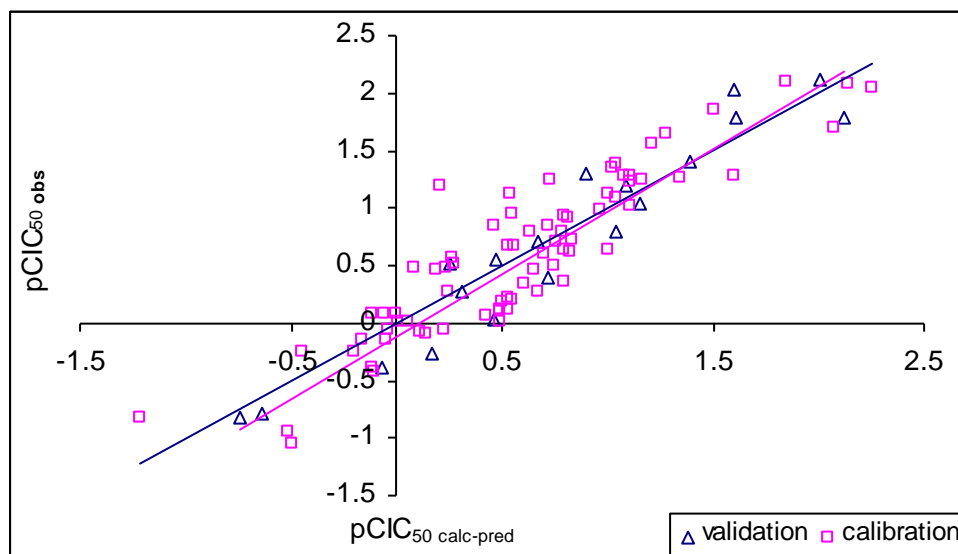
$$pCIC_{50\text{ obs}} = 0,0000032 + 0,99999 pCIC_{50\text{ Calc}} \quad (\text{eq, 31})$$

$$s = 0,299949 \quad R^2 = 81,4 \% \quad R^2_{\text{ajust}} = 81,2 \%$$

- Pour l'ensemble de validation

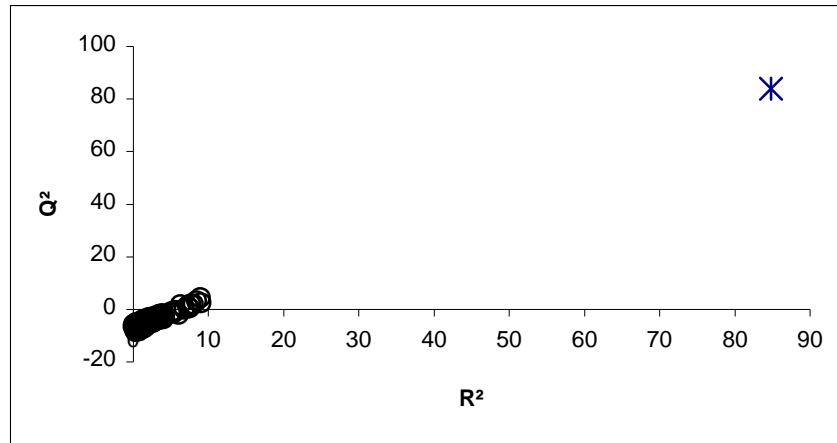
$$pCIC_{50\text{ obs}} = 0,0000032 + 0,99999 pCIC_{50\text{ cal}} \quad (\text{eq, 32})$$

$$s = 0,299949 \quad R^2 = 81,4 \% \quad R^2_{\text{ajust}} = 81,2 \%$$



**Figure -06:** Droites d'ajustement pour le modèle 02.

Par le test de randomisation (figure 7), nous nous sommes assuré qu'une relation structure-activité réelle a été établie par notre modèle.



**Figure -07:** *Test de randomisation pour le modèle 02,*

Les 100 modèles pour lesquels nous avons randomisé les valeurs du  $pCIC_{50}$  ont des valeurs de  $Q^2$  faibles et des valeurs du coefficient de corrélation multiple ( $R^2$ ) petites, Seul le modèle avec le vecteur réel (étoile) offre des valeurs élevées pour les deux statistiques représentées dans la figure 7.

### III- Modèle numéro 03: CLOGP

Matrice de corrélation:

	pCIC <sub>50</sub>	ClogP
ClogP	0,893 0,000	
pKa	-0,035 0,768	0,275 0,018

#### III-1-Equation et analyse de régression :

L'équation de régression est

$$pCIC_{50} = (\pm 0,2414)0,606 + (\pm 0,02964)0,682 \text{ ClogP} - (\pm 0,02645)0,189 \text{ pKa} \quad (\text{eq, 33})$$

$$n = 74 \quad R^2 = 88,20 \% \quad Q^2 = 87,30 \% \quad F = 265,30 \quad s = 0,2383$$

$$R^2_{\text{adj}} = 87,87\% \quad EQMC = 0,233 \quad EQMP = 0,242$$

#### III-2- Analyse des résidus et diagnostics d'influence :

Les valeurs des résidus de prédiction, standardisés de prédiction, et les valeurs de  $h_{ii}$  sont présentées dans le tableau -9 dont les colonnes 3 et 4 reproduisent les valeurs expérimentales et calculées, prédites de pCIC<sub>50</sub> des composés considérés.

## Partie expérimentale

**Tableau - 9** : Valeurs des  $pCIC_{50}$ , observés ( $pCIC_{50\text{ obs}}$ ) et calculés, prédites ( $pCIC_{50\text{ cal-pred}}$ ), des résidus, standardisés ( $e_{i\text{ std}}$ ) ainsi que des leviers ( $h_{ij}$ ) (Cas de CLOGP).

N°	Nom du Composé	$pCIC_{50\text{ obs}}$	$pCIC_{50\text{ cal-pred}}$	$h_{ij}$	$e_{i\text{ std}}$
01*	2,4,6-tribromophénol	2,050	2,100	0,194	0,232
02*	4-acetoamidophénol	-0,820	-0,970	0,093	-0,659
03*	4-heptyloxyphénol	2,030	1,821	0,087	-0,917
04*	2,6-difluorophénol	0,400	0,386	0,056	-0,062
05*	2-allylphénol	0,350	0,433	0,017	0,353
06*	4-bromo-6-chloro-2-methylphénol	1,280	1,491	0,058	0,911
07*	2-tertbutyl-4-methylphénol	1,300	0,975	0,064	-1,408
08*	3-hydroxybenzaldéhyde	0,080	-0,113	0,030	-0,820
09*	4-phenylphénol	1,380	1,212	0,028	-0,717
10*	2-chloro-5-methylphénol	0,640	0,799	0,024	0,677
11*	4-methoxyphénol	-0,140	-0,250	0,040	-0,471
12*	2,3-dichlorophénol	1,270	1,118	0,061	-0,660
13*	4-hydroxybenzamide	-0,780	-0,917	0,085	-0,602
14*	2-(tert)butylphénol	1,240	0,781	0,032	-1,959
15*	2,4,5-trichlorophénol	2,100	1,669	0,099	-1,907
16*	3-nitrophénol	0,510	0,289	0,029	-0,940
17*	4-chloro-2-isopropyl-5-methylphénol	1,860	1,581	0,053	-1,202
18*	4-benzyloxyphénol	1,040	1,189	0,046	0,641
19*	4-iodophénol	0,850	0,840	0,017	-0,041
20	phénol	-0,430	-0,278	0,039	0,678
21	2-methylphénol	-0,270	-0,023	0,032	1,090
22	2-ethylphénol	0,180	0,350	0,022	0,736
23	2-isopropylphénol	0,800	0,482	0,024	-1,386
24	2-phenylphénol	1,090	0,752	0,015	-1,449
25	2-fluorophénol	0,280	0,124	0,026	-0,680
26	2-chlorophénol	0,280	0,459	0,023	0,775
27	2-bromophénol	0,500	0,614	0,024	0,496
28	2-nitrophénol	0,670	0,505	0,069	-0,771
29	2-cyanophénol	0,030	0,375	0,083	1,650
30	2-hydroxybenzaldehyde	0,480	0,265	0,030	-0,945
31	2-acetylphénol	0,080	0,174	0,019	0,404
32	2-hydroxybenzamide	-0,240	-0,104	0,042	0,607
33	2-hydroxybenzyl alcohol	-0,950	-1,007	0,096	-0,278
34	2-hydroxybenzaldoxime	-0,250	-0,522	0,054	-1,239
35	3-methylphénol	-0,060	0,061	0,025	0,526
36	3-ethylphénol	0,230	0,408	0,019	0,770
37	3-isopropylphénol	0,610	0,675	0,019	0,281
38	3-tertbutylphénol	0,730	0,943	0,024	0,929

## Partie expérimentale

Tableau - 9: suite.

N°	Nom du Composé	pCIC <sub>50</sub> obs	pCIC <sub>50</sub> calc-pred	h <sub>ii</sub>	e <sub>i</sub> std
39	3-phenylphénol	1,350	1,078	0,023	-1,181
40	3-fluorophénol	0,470	0,155	0,019	-1,362
41	3-chlorophénol	0,960	0,580	0,015	-1,632
42	3-iodophénol	1,120	0,901	0,021	-0,948
43	3-cyanophénol	-0,070	0,067	0,030	0,599
44	3-acetylphénol	-0,380	-0,140	0,030	1,053
45	3-methoxyphénol	-0,140	-0,146	0,030	-0,026
46	3-hydroxybenzyl alcohol	-1,040	-0,956	0,090	0,407
47	methyl-3-hydroxybenzoate	-0,050	0,218	0,018	1,153
48	ethyl-3-hydroxybenzoate	0,480	0,601	0,015	0,520
49	4-ethylphénol	0,210	0,422	0,018	0,912
50	4-propylphénol	0,640	0,726	0,023	0,373
51	4-isopropylphénol	0,470	0,637	0,023	0,726
52	4-secbutylphénol	0,980	0,998	0,029	0,079
53	4-tertbutylphénol	0,910	0,923	0,026	0,055
54	4-tertpentylphénol	1,230	1,270	0,040	0,180
55	4-cyclopentylphénol	1,290	1,142	0,028	-0,650
56	4-hydroxyphenylmethane	1,200	0,042	0,027	-5,062
57	4-tertoctylphénol	2,090	2,425	0,134	1,744
58	a,a,a-trifluoro-4-cresol	0,620	0,916	0,023	1,284
59	4-fluorophénol	0,020	0,041	0,024	0,092
60	4-chlorophénol	0,550	0,517	0,014	-0,141
61	4-bromophénol	0,680	0,637	0,014	-0,187
62	4-cyanophénol	0,520	0,190	0,043	-1,481
63	4-hydroxybenzaldehyde	0,270	0,149	0,057	-0,556
64	4-acetylphénol	-0,090	0,075	0,044	0,742
65	4'-hydroxypropiophenone	0,060	0,285	0,020	0,973
66	4-hydroxybenzophenone	1,020	1,007	0,023	-0,055
67	4-hydroxybenzylcyanide	-0,380	-0,583	0,056	-0,930
68	4-ethoxyphénol	0,010	0,054	0,034	0,196
69	4-butoxyphénol	0,700	0,757	0,031	0,252
70	4-hexyloxyphénol	1,650	1,460	0,061	-0,876
71	4-hydroxyphenethyl alcohol	-0,830	-0,817	0,078	0,064
72	methyl-4-hydroxybenzoate	0,080	0,248	0,018	0,724
73	ethyl-4-hydroxybenzoate	0,570	0,633	0,017	0,272
74	2,3-dimethylphénol	0,120	0,267	0,026	0,642
75	2,5-dimethylphénol	0,010	0,269	0,025	1,128
76	3,4-dimethylphénol	0,120	0,307	0,024	0,812
77	3,5-dimethylphénol	0,110	0,373	0,020	1,137

## Partie expérimentale

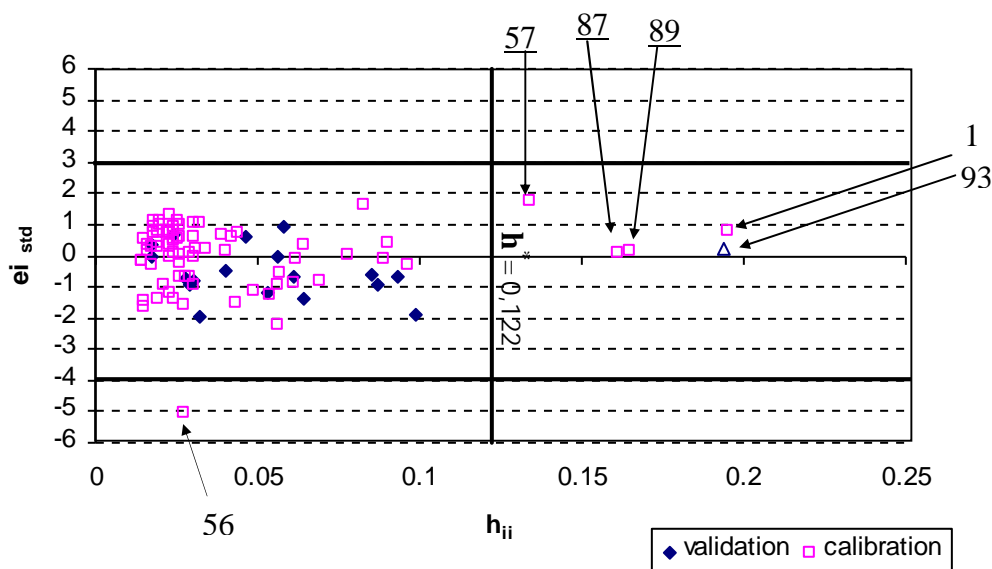
**Tableau - 9:** suite et fin.

N°	Nom du Composé	pCIC <sub>50</sub> obs	pCIC <sub>50</sub> calc-pred	h <sub>ii</sub>	e <sub>i std</sub>
78	3-chloro-4-fluorophénol	0,840	0,773	0,017	-0,289
79	2-bromo-4-méthylphénol	0,790	0,913	0,024	0,534
80	2,5-dichlorophénol	1,130	1,199	0,064	0,322
81	4-chloro-2-méthylphénol	0,700	0,778	0,016	0,335
82	4-chloro-3-méthylphénol	0,800	0,835	0,016	0,150
83	3,5-dichlorophénol	1,560	1,311	0,049	-1,126
84	2,4-dibromophénol	1,400	1,382	0,062	-0,084
85	3,4,5-triméthylphénol	0,930	0,577	0,027	-1,542
86	2,3,5-triméthylphénol	0,360	0,581	0,026	0,966
87	2,4,6-trichlorophénol	1,700	1,718	0,161	0,100
88	4-chloro-3,5-diméthylphénol	1,200	1,145	0,026	-0,238
89	4-bromo-2,6-dichlorophénol	1,780	1,811	0,165	0,172
90	4-bromo-2,6-diméthylphénol	1,280	1,122	0,029	-0,691
91	6-tertbutyl-2-4-diméthylphénol	1,250	1,227	0,089	-0,112
92	2,6-diphénylphénol	2,110	1,628	0,056	-2,204
93	2,6ditertbutyl-4-méthylphénol	1,790	1,932	0,195	0,824

\* ensemble de validation,

### Diagramme de Williams:

La figure 8 est une représentation des valeurs des résidus de prédiction standardisés  $e_{i \text{ std}}$  des composés pour les deux sous-groupes (calibration et validation) en fonction de leurs leviers respectifs.



**Figure -08 :** Diagramme de williams pour le modèle 03.

Comme le montre la figure 11, les valeurs de  $e_{i\text{ std}}$  sont toutes comprises entre les bornes  $\pm 3$  sauf le point [56(4-hydroxyphenylmethane)], on trouve aussi cinq points influents [93(2,6ditertbutyl-4-methylphénol), 57(4-tertoctylphénol), 13(4-hydroxybenzamide), 3 (4-heptyloxyphénol)] qui apparaissent en dehors de la limite fixée par la valeur critique  $h^* = 0,121$  symbolisée par la droite parallèle à l'axe des  $e_{i\text{ std}}$ .

### III-3- Qualité de l'ajustement :

La qualité de l'ajustement peut être appréciée dans la figure 09 par les deux droites d'ajustement dont les équations sont :

- Pour l'ensemble de calibration

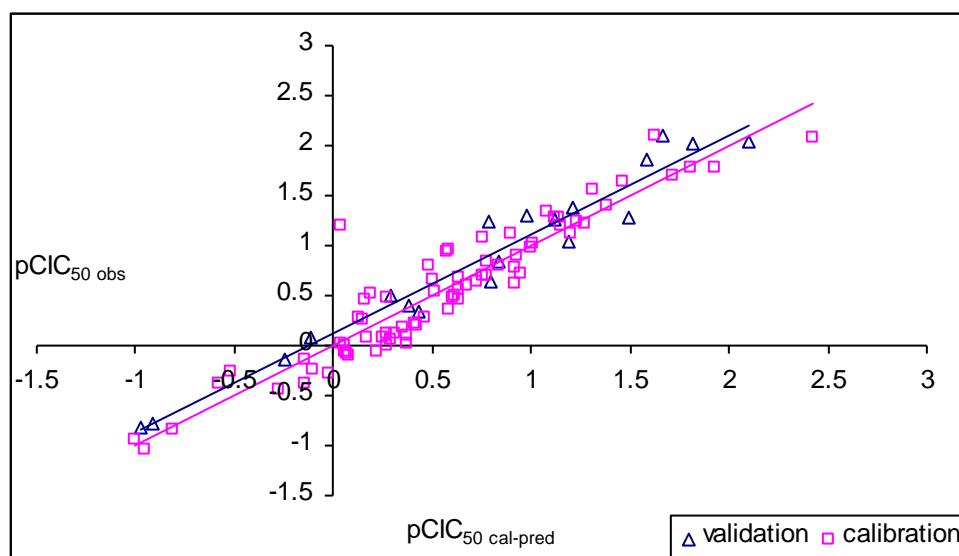
$$pCIC50_{obs} = 0,120940 + 0,993635 pCIC50_{pred} \quad (\text{eq, 34})$$

$$s = 0,197014 \quad R^2 = 95,3 \% \quad R^2_{ajust} = 95,0\%$$

- Pour l'ensemble de validation

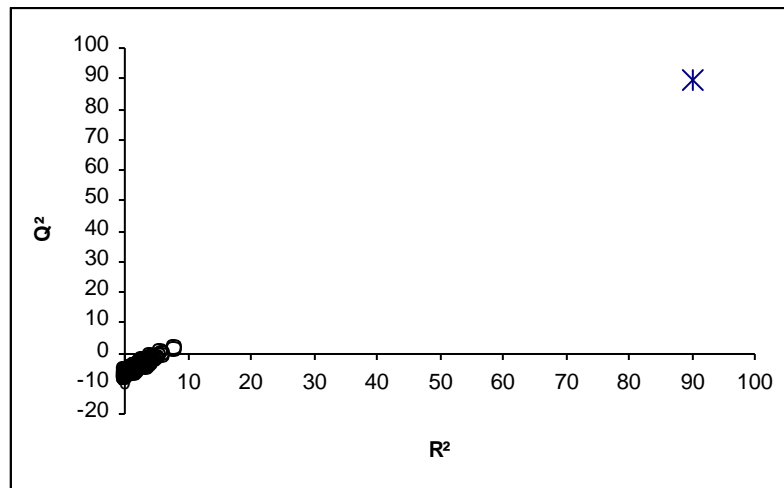
$$pCIC50_{obs} = -0,0000048 + 1,00000 pCIC50_{Calc} \quad (\text{eq, 35})$$

$$s = 0,236678 \quad R^2 = 88,2 \% \quad R^2_{ajust} = 88,0 \%$$



**Figure -09 :** Droites d'ajustement pour le modèle 03,

Par le test de randomisation (10), nous nous sommes assuré qu'une relation structure- activité réelle a été établie par notre modèle.



**Figure -10:** *Test de randomisation pour le modèle 03,*

Les 100 modèles pour lesquels nous avons randomisé les valeurs du  $pCIC_{50}$  ont des valeurs de  $Q^2$  faibles et des valeurs du coefficient de corrélation multiple ( $R^2$ ) petites. Seul le modèle avec le vecteur réel (étoile) offre des valeurs élevées pour les deux statistiques représentées dans la figure 10.

### VI- Modèle numéro 04: LOGK<sub>owwin</sub>

Matrice de corrélation:

	pCIC50	logK <sub>owwin</sub>
logK <sub>owwin</sub>	0,888	0,000
pKa	-0,088	0,261
	0,400	0,012

#### VI-1-Equation et analyse de régression :

L'équation de régression est

$$pCIC_{50} = (\pm 0,2031)0,880 + (\pm 0,02466) 0,691 \log K_{owwin} - (\pm 0,02225)0,219 pKa \quad (\text{eq. 36})$$

$$n = 74 \quad R^2 = 87,9 \% \quad Q^2 = 86,66 \% \quad F = 396,13 \quad s = 0,2376$$
$$R^2_{adj} = 87,56\% \quad EQMC = 0,236 \quad EQMP = 0,248$$

#### VI-2- Analyse des résidus et diagnostiques d'influence :

Les valeurs des résidus de prédiction standardisés, et les valeurs de  $h_{ii}$  sont présentées dans le tableau -10 dont les colonnes 3 et 4 reproduisent les valeurs expérimentales et calculées- prédites de pCIC<sub>50</sub> des composés considérés.

## Partie expérimentale

**Tableau -10** : Valeurs des  $pCIC_{50}$ , observés ( $pCIC_{50\text{ obs}}$ ) et calculés prédites ( $pCIC_{50\text{ cal-pred}}$ ), des résidus standardisés ( $e_{i\text{ std}}$ ) ainsi que des leviers ( $h_{ii}$ ) (Cas de  $LOGK_{\text{owwin}}$ )

N°	Nom du Composé	$pCIC_{50\text{ obs}}$	$pCIC_{50\text{ cal-pred}}$	$h_{ii}$	$e_{i\text{ std}}$
01*	2,6ditertbutyl-4-methylphénol	1,79	1,4623	0,19	-1,5127
02*	2,4,6-trichlorophénol	1,7	1,9602	0,153	1,1742
03*	4-hydroxybenzamide	-0,78	-0,934	0,097	-0,6731
04*	4-phenylphénol	1,38	1,0352	0,021	-1,4476
05*	2-cyanophénol	0,03	0,5517	0,085	2,266
06*	2,6-diphenylphénol	2,11	2,1349	0,107	0,1096
07*	2-fluorophénol	0,28	0,1769	0,027	-0,4344
08*	2,4-dibromophénol	1,4	1,465	0,055	0,2778
09*	4-methoxyphénol	-0,14	-0,2765	0,042	-0,5791
10*	2-tertbutyl-4-methylphénol	1,3	1,0382	0,079	-1,1331
11*	3,4-dimethylphénol	0,12	0,3856	0,024	1,1169
12*	2-chloro-5-methylphénol	0,64	0,895	0,023	1,0714
13*	4-chloro-2-isopropyl-5-methylphénol	1,86	1,5112	0,053	-1,4891
14*	2-(tert)butylphénol	1,24	0,8596	0,038	-1,6112
15*	4-bromophénol	0,68	0,491	0,014	-0,7908
16*	3-acetylphénol	-0,38	-0,2925	0,043	0,3715
17*	4-acetylphénol	-0,09	-0,0041	0,055	0,367
18*	2-hydroxybenzyl alcohol	-0,95	-0,8777	0,085	0,314
19*	2-phenylphénol	1,09	1,0352	0,021	-0,23
20	phénol	-0,43	-0,2777	0,04	0,673
21	2-methylphénol	-0,27	0,0274	0,03	1,2933
22	2-ethylphénol	0,18	0,3752	0,022	0,8386
23	2-allylphénol	0,35	0,6905	0,018	1,4533
24	2-isopropylphénol	0,8	0,6098	0,027	-0,8232
25	2-chlorophénol	0,28	0,5279	0,023	1,0658
26	2-bromophénol	0,5	0,7161	0,024	0,9305
27	2-nitrophénol	0,67	0,6946	0,07	0,1141
28	2-hydroxybenzaldehyde	0,48	0,4792	0,029	-0,0035
29	2-acetylphénol	0,08	0,237	0,019	0,6713
30	2-hydroxybenzamide	-0,24	-0,1912	0,056	0,2211
31	2-hydroxybenzaldoxime	-0,25	-0,3414	0,044	-0,4059
32	3-methylphénol	-0,06	0,0932	0,024	0,6604
33	3-ethylphénol	0,23	0,4081	0,02	0,7622
34	3-isopropylphénol	0,61	0,6856	0,021	0,3244
35	3-tertbutylphénol	0,73	0,9861	0,028	1,1108
36	3-phenylphénol	1,35	1,015	0,021	-1,4376
37	3-fluorophénol	0,47	0,0352	0,025	-1,8766
38	3-chlorophénol	0,96	0,3888	0,017	-2,433

## Partie expérimentale

**Tableau -10 : Suite.**

N°	Nom du Composé	pCIC <sub>50</sub> obs	pCIC <sub>50</sub> calc-pred	h <sub>ii</sub>	e <sub>i std</sub>
39	3-iodophénol	1,12	0,7974	0,017	-1,375
40	3-nitrophénol	0,51	0,4062	0,03	-0,4508
41	3-cyanophénol	-0,07	0,1393	0,032	0,9123
42	3-hydroxybenzaldahyde	0,08	-0,2173	0,042	-1,316
43	3-methoxyphénol	-0,14	-0,1373	0,032	0,0116
44	3-hydroxybenzyl alcohol	-1,04	-0,8549	0,084	0,8772
45	methyl-3-hydroxybenzoate	-0,05	0,2523	0,019	1,2916
46	ethyl-3-hydroxybenzoate	0,48	0,6153	0,014	0,5745
47	4-ethylphénol	0,21	0,4258	0,018	0,9219
48	4-propylphénol	0,64	0,6826	0,025	0,1838
49	4-isopropylphénol	0,47	0,6351	0,025	0,7117
50	4-secbutylphénol	0,98	0,9677	0,032	-0,0537
51	4-tertbutylphénol	0,91	0,9582	0,03	0,2098
52	4-tertpentylphénol	1,23	1,2732	0,046	0,1925
53	4-cyclopentylphénol	1,29	1,3218	0,039	0,1403
54	4-hydroxyphenylmethane	1,2	1,0726	0,031	-0,555
55	4-tertoctylphénol	2,09	2,2993	0,128	1,0676
56	a,a,a-trifluoro-4-cresol	0,62	0,6971	0,018	0,3291
57	4-fluorophénol	0,02	-0,1166	0,031	-0,5952
58	4-chlorophénol	0,55	0,3053	0,016	-1,0421
59	4-iodophénol	0,85	0,7165	0,014	-0,5667
60	4-cyanophénol	0,52	0,3038	0,045	-0,9631
61	4-hydroxybenzaldehyde	0,27	0,1318	0,067	-0,6373
62	4'-hydroxypropiophenone	0,06	0,1261	0,027	0,2863
63	4-hydroxybenzophenone	1,02	0,7881	0,017	-0,9882
64	4-hydroxybenzylcyanide	-0,38	-0,4507	0,051	-0,3176
65	4-ethoxyphénol	0,01	-0,0197	0,036	-0,1305
66	4-butoxyphénol	0,7	0,627	0,033	-0,3186
67	4-hexyloxyphénol	1,65	1,267	0,06	-1,7444
68	4-heptyloxyphénol	2,03	1,5997	0,083	-2,0366
69	4-benzyloxyphénol	1,04	0,7579	0,038	-1,2428
70	4-acetoamidophénol	-0,82	-1,1493	0,116	-1,645
71	4-hydroxyphenethyl alcohol	-0,83	-0,5451	0,057	1,2933
72	methyl-4-hydroxybenzoate	0,08	0,2928	0,019	0,9097
73	ethyl-4-hydroxybenzoate	0,57	0,6583	0,016	0,3758
74	2,6-difluorophénol	0,4	0,6192	0,056	0,9934
75	2,3-dimethylphénol	0,12	0,3785	0,025	1,1156
76	2,5-dimethylphénol	0,01	0,3805	0,025	1,5985
77	3,5-dimethylphénol	0,11	0,4286	0,021	1,3659

## Partie expérimentale

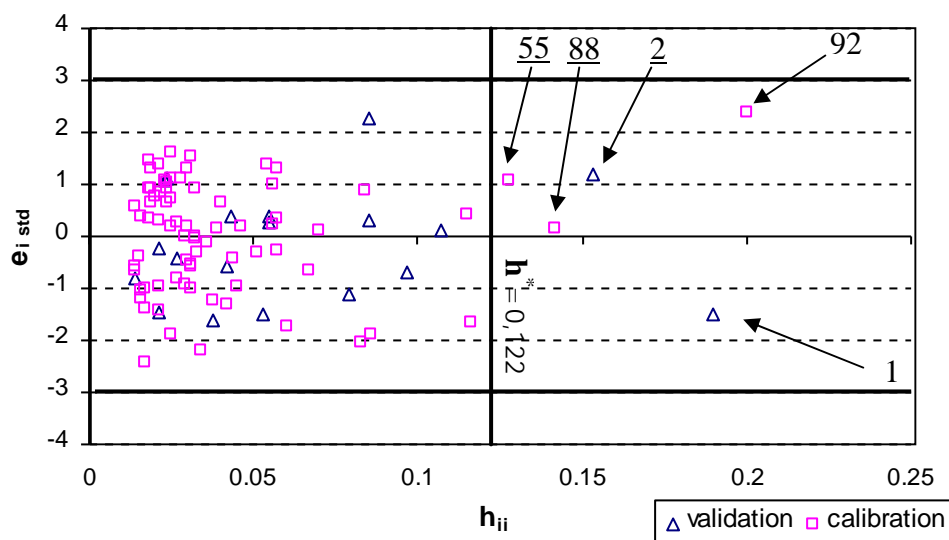
**Tableau -10 :** Suite et fin.

N°	Nom du Composé	pCIC <sub>50</sub> obs	pCIC <sub>50</sub> calc-pred	h <sub>ii</sub>	e <sub>i std</sub>
78	3-chloro-4-fluorophénol	0,84	0,5579	0,016	-1,2
79	2-bromo-4-méthylphénol	0,79	1,0339	0,024	1,0499
80	2,5-dichlorophénol	1,13	1,2077	0,057	0,3525
81	2,3-dichlorophénol	1,27	1,2077	0,057	-0,2822
82	4-chloro-2-méthylphénol	0,7	0,6111	0,015	-0,3774
83	4-chloro-3-méthylphénol	0,8	0,6415	0,014	-0,6725
84	3,5-dichlorophénol	1,56	1,056	0,034	-2,2039
85	3,4,5-triméthylphénol	0,93	0,7046	0,031	-0,9817
86	2,3,5-triméthylphénol	0,36	0,7097	0,031	1,522
87	4-chloro-3,5-diméthylphénol	1,2	0,9749	0,021	-0,9654
88	4-bromo-2,6-dichlorophénol	1,78	1,8117	0,142	0,166
89	2,4,5-trichlorophénol	2,1	1,7021	0,086	-1,8914
90	4-bromo-6-chloro-2-méthylphénol	1,28	1,5872	0,054	1,3876
91	4-bromo-2,6-diméthylphénol	1,28	1,0682	0,029	-0,9197
92	2,4,6-tribromophénol	2,05	2,4638	0,2	2,4008
93	6-tertbutyl-2-4-diméthylphénol	1,25	1,3382	0,115	0,4402

\* ensemble de validation.

### Diagramme de Williams :

La figure 11 est une représentation des valeurs des résidus de prédiction standardisés e<sub>i std</sub> des composés pour les deux sous-groupes (calibration et validation) en fonction de leurs leviers respectifs.



**Figure-11:** Diagramme de Williams pour le modèle 04.

La figure 11, montre les valeurs de  $e_{i \text{ std}}$  qui sont toutes comprises entre les bornes  $\pm 3$ . On retrouve cinq composés influents [55 (4-tertoctylphénol), 88 (4-bromo-2,6-dichlorophénol), 02 (2,4,6-trichlorophénol), 92 (2,4,6-tribromophénol) et 1 (2,6ditertbutyl-4-methylphénol)] qui apparaissent en dehors de la limite fixée par la valeur critique  $h^* = 0,122$  symbolisée par la droite parallèle à l'axe des  $e_{i \text{ std}}$ .

### VI-3- Qualité de l'ajustement :

La qualité de l'ajustement peut être appréciée dans la figure 12 par les deux droites d'ajustements dont les équations sont :

- Pour l'ensemble de calibration

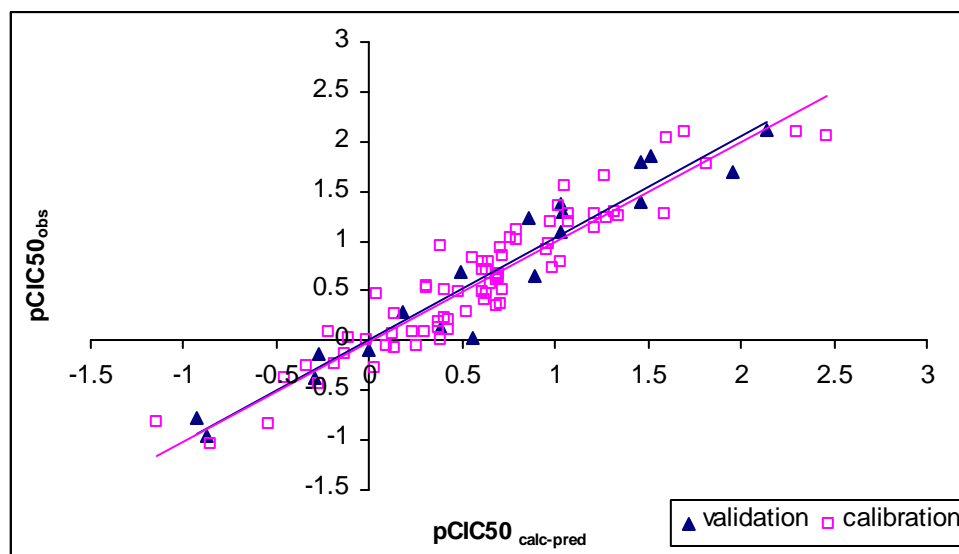
$$pCIC_{50 \text{ obs}} = 0.0000002 + 1.00001 pCIC_{50 \text{ calc}} \quad (\text{eq. 40})$$

$$s = 0.239058 \quad R^2 = 87,9 \% \quad R^2_{\text{ajust}} = 87,7 \%$$

- Pour l'ensemble de validation

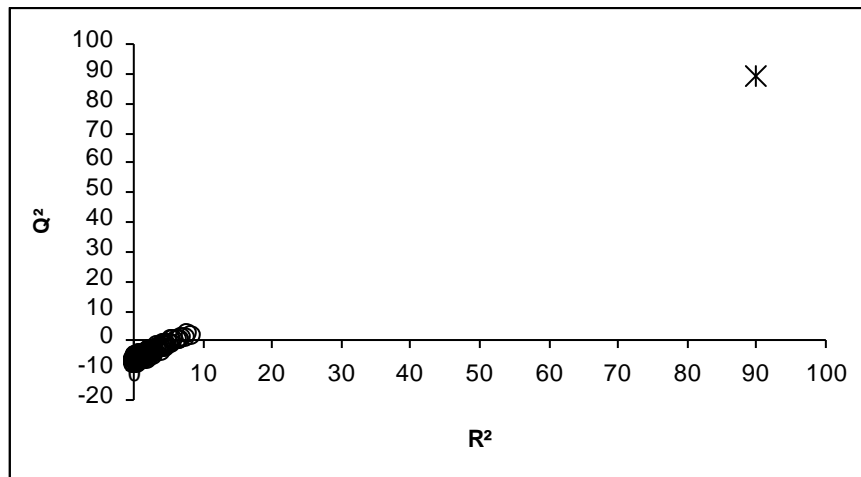
$$pCIC_{50 \text{ obs}} = 0,0195473 + 1.02310 pCIC_{50 \text{ pred}} \quad (\text{eq. 41})$$

$$s = 0.257008 \quad R^2 = 92,8 \% \quad R^2_{\text{ajust}} = 92,4 \%$$



**Figure-12:** Droites d'ajustement pour le modèle 04.

Pour le test de randomisation (figure 13), nous nous sommes assuré qu'une relation structure-activité réelle a été établie par notre modèle.



**Figure -13:** *Test de randomisation pour le modèle 04.*

Les 100 modèles pour lesquels nous avons randomisé les valeurs du  $pCIC_{50}$  ont des valeurs de  $Q^2$  faibles et des valeurs du coefficient de corrélation multiple ( $R^2$ ) petites. Seul le modèle avec le vecteur réel (étoile) offre des valeurs élevées pour les deux statistiques représentées dans la figure 13.

### V- Modèle numéro 05: LOGK<sub>ow</sub>

Matrice de corrélation:

	pCIC <sub>50</sub>	logK <sub>ow</sub>
logK <sub>ow</sub>	0,887	
	0,000	
pKa	-0,053	0,298
	0,653	0,010

#### 1- Equation et analyse de régression :

L'équation de régression est

$$pCIC50 = (\pm 0.2169)0.804 + (\pm 0.02586)0.641 \log K_{ow} - (\pm 0.02397)0.209 pKa \quad (\text{eq. 39})$$

$$n = 74 \quad R^2 = 89,67 \% \quad Q^2 = 88,73 \% \quad F = 308,12 \quad s = 0,2154$$

$$R^2_{adj} = 89,38 \% \quad EQMC = 0,211 \quad EQMP = 0,22$$

#### 2- Analyse des résidus et diagnostics d'influence :

Les valeurs des résidus de prédiction standardisés, et les valeurs de  $h_{ii}$  sont présentées dans le tableau 11 dont les colonnes 3 et 4 reproduisent les valeurs expérimentales et calculées, prédites de pCIC<sub>50</sub> des composés considérés.

## Partie expérimentale

**Tableau-11:** Valeurs des  $pCIC_{50}$ , observés ( $pCIC_{50\text{ obs}}$ ) et calculés prédites ( $pCIC_{50\text{ cal-pred}}$ ), des résidus standardisés ( $e_{i\text{ std}}$ ) ainsi que des leviers ( $h_{ii}$ ) (Cas de  $LOGK_{ow}$ ).

N°	Nom du Composé	$pCIC_{50\text{ obs}}$	$pCIC_{50\text{ calc-pred}}$	$h_{ii}$	$e_{i\text{ std}}$
01*	2,4,6-tribromophénol	2,050	2,062	0,187	0,062
02*	4-acetoamidophénol	-0,820	-0,991	0,101	-0,838
03*	6-tertbutyl-2-4-diméthylphénol	1,250	1,411	0,099	0,787
04*	3-iodophénol	1,120	0,803	0,018	-1,485
05*	4-tertoctylphénol	2,090	2,137	0,103	0,231
06*	2-hydroxybenzamide	-0,240	-0,325	0,056	-0,407
07*	4-benzyloxyphénol	1,040	0,584	0,032	-2,154
08*	2-nitrophénol	0,670	0,483	0,068	-0,899
09*	4-fluorophénol	0,020	-0,036	0,028	-0,262
10*	2,4-dibromophénol	1,400	1,320	0,057	-0,383
11*	4-bromo-2,6-diméthylphénol	1,280	1,234	0,033	-0,218
12*	3-nitrophénol	0,510	0,245	0,030	-1,249
13*	4-hydroxybenzamide	-0,780	-0,911	0,091	-0,635
14*	2,4,5-trichlorophénol	2,100	1,734	0,102	-1,795
15*	4-chlorophénol	0,550	0,426	0,014	-0,582
16*	3-acetylphénol	-0,380	-0,184	0,034	0,924
17*	2-hydroxybenzaldoxime	-0,250	-0,561	0,060	-1,489
18*	4-hexyloxyphénol	1,650	1,276	0,051	-1,784
19*	4-chloro-2-isopropyl-5-méthylphénol	1,860	1,730	0,066	-0,627
20	phénol	-0,430	-0,332	0,045	0,487
21	2-méthylphénol	-0,270	0,022	0,031	1,420
22	2-éthylphénol	0,180	0,374	0,022	0,930
23	2-allylphénol	0,350	0,426	0,017	0,362
24	2-isopropylphénol	0,800	0,589	0,025	-1,019
25	2-(tert)butylphénol	1,240	0,799	0,032	-2,151
26	2-phenylphénol	1,090	0,965	0,019	-0,600
27	2-fluorophénol	0,280	0,027	0,030	-1,230
28	2-chlorophénol	0,280	0,430	0,023	0,721
29	2-bromophénol	0,500	0,547	0,024	0,226
30	2-cyanophénol	0,030	0,373	0,083	1,814
31	2-hydroxybenzaldehyde	0,480	0,390	0,028	-0,434
32	2-acetylphénol	0,080	0,219	0,019	0,666
33	2-hydroxybenzyl alcohol	-0,950	-0,984	0,100	-0,185
34	3-méthylphénol	-0,060	0,076	0,025	0,656
35	3-éthylphénol	0,230	0,401	0,019	0,817
36	3-isopropylphénol	0,610	0,651	0,019	0,197
37	3-tertbutylphénol	0,730	0,903	0,023	0,834
38	3-phenylphénol	1,350	0,948	0,019	-1,922

## Partie expérimentale

**Tableau-11: Suite.**

N°	Nom du Composé	pCIC <sub>50</sub> obs	pCIC <sub>50</sub> calc-pred	h <sub>ii</sub>	e <sub>i</sub> std
39	3-fluorophénol	0,470	0,090	0,022	-1,826
40	3-chlorophénol	0,960	0,495	0,015	-2,210
41	3-cyanophénol	-0,070	0,033	0,032	0,501
42	3-hydroxybenzaldéhyde	0,080	-0,151	0,035	-1,131
43	3-méthoxyphénol	-0,140	-0,204	0,035	-0,311
44	3-hydroxybenzyl alcool	-1,040	-0,965	0,097	0,405
45	méthyl-3-hydroxybenzoate	-0,050	0,132	0,021	0,872
46	éthyl-3-hydroxybenzoate	0,480	0,516	0,015	0,170
47	4-éthylphénol	0,210	0,416	0,018	0,981
48	4-propylphénol	0,640	0,693	0,023	0,253
49	4-isopropylphénol	0,470	0,609	0,022	0,670
50	4-secbutylphénol	0,980	0,949	0,028	-0,150
51	4-tertbutylphénol	0,910	0,880	0,025	-0,143
52	4-tertpentylphénol	1,230	1,205	0,038	-0,121
53	4-cyclopentylphénol	1,290	1,099	0,027	-0,924
54	4-phénylphénol	1,380	0,965	0,019	-1,986
55	4-hydroxyphénylméthane	1,200	1,061	0,028	-0,672
56	a,a,a-trifluoro-4-crésol	0,620	0,826	0,021	0,987
57	4-bromophénol	0,680	0,541	0,014	-0,661
58	4-iodophénol	0,850	0,736	0,015	-0,539
59	4-cyanophénol	0,520	0,169	0,045	-1,748
60	4-hydroxybenzaldéhyde	0,270	0,137	0,058	-0,676
61	4-acétylphénol	-0,090	0,054	0,046	0,715
62	4'-hydroxypropiophénone	0,060	0,226	0,022	0,798
63	4-hydroxybenzophénone	1,020	0,923	0,021	-0,465
64	4-hydroxybenzylcyanide	-0,380	-0,606	0,062	-1,154
65	4-méthoxyphénol	-0,140	-0,318	0,047	-0,889
66	4-éthoxyphénol	0,010	-0,041	0,039	-0,252
67	4-butoxyphénol	0,700	0,617	0,030	-0,402
68	4-heptyloxyphénol	2,030	1,615	0,073	-2,158
69	4-hydroxyphényl alcool	-0,830	-0,837	0,085	-0,035
70	méthyl-4-hydroxybenzoate	0,080	0,185	0,021	0,501
71	éthyl-4-hydroxybenzoate	0,570	0,551	0,016	-0,089
72	2,6-difluorophénol	0,400	0,293	0,058	-0,545
73	2,3-diméthylphénol	0,120	0,420	0,024	1,444
74	2,5-diméthylphénol	0,010	0,422	0,024	1,981
75	3,4-diméthylphénol	0,120	0,426	0,023	1,470
76	3,5-diméthylphénol	0,110	0,461	0,020	1,681

## Partie expérimentale

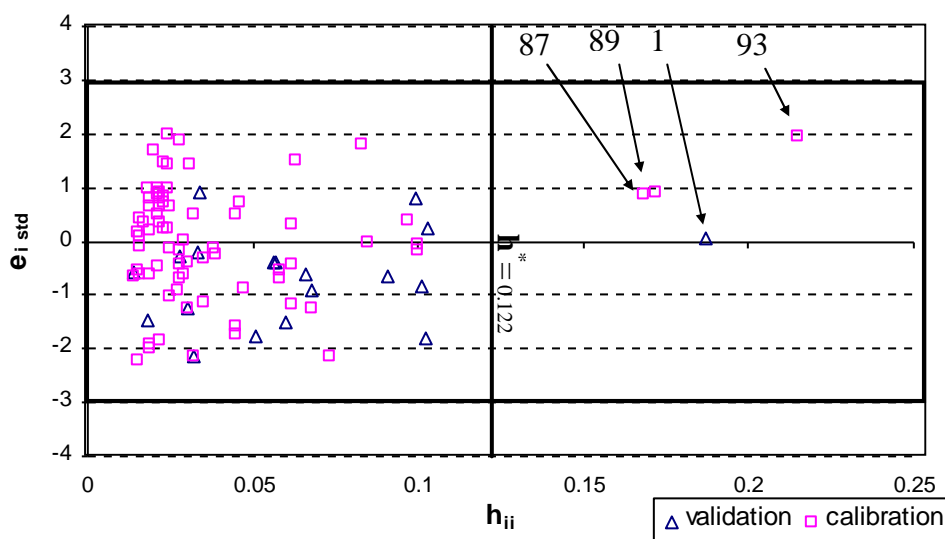
**Tableau-11:** Suite et fin.

N°	Nom du Composé	pCIC <sub>50 obs</sub>	pCIC <sub>50 calc-pred</sub>	h <sub>ii</sub>	e <sub>i std</sub>
77	3-chloro-4-fluorophénol	0,840	0,714	0,016	-0,598
78	2-chloro-5-methylphénol	0,640	0,847	0,024	0,996
79	2-bromo-4-methylphénol	0,790	0,860	0,022	0,335
80	2,5-dichlorophénol	1,130	1,190	0,062	0,306
81	2,3-dichlorophénol	1,270	1,190	0,062	-0,410
82	4-chloro-2-methylphénol	0,700	0,792	0,016	0,438
83	4-chloro-3-methylphénol	0,800	0,817	0,016	0,082
84	3,5-dichlorophénol	1,560	1,244	0,045	-1,571
85	3,4,5-trimethylphénol	0,930	0,803	0,029	-0,616
86	2,3,5-trimethylphénol	0,360	0,750	0,028	1,887
87	2,4,6-trichlorophénol	1,700	1,844	0,168	0,880
88	4-chloro-3,5-dimethylphénol	1,200	1,201	0,029	0,004
89	4-bromo-2,6-dichlorophénol	1,780	1,930	0,172	0,922
90	4-bromo-6-chloro-2-methylphénol	1,280	1,573	0,063	1,500
91	2-tertbutyl-4-methylphénol	1,300	1,055	0,068	-1,265
92	2,6-diphenylphénol	2,110	2,097	0,100	-0,071
93	2,6ditertbutyl-4-methylphénol	1,790	2,080	0,215	1,937

\* ensemble de validation,

### Diagramme de Williams :

On a représenté, sur la même figure 14; pour les deux ensembles (calibration et validation); les valeurs de e<sub>i std</sub> (tableau V-2 colonne) et celle de h<sub>ii</sub> (tableau V-2 colonne5).



**Figure -14:** Diagramme de Williams pour le modèle 05

La figure 14, montre les valeurs de  $e_{i \text{ std}}$  qui sont toutes comprises entre les bornes  $\pm 3$  , On retrouve quatre composés influents [87 (2, 4,6-trichlorophénol), 89 (4-bromo-2,6-dichlorophénol), 01(2, 4,6-tribromophénol) 93(2,6ditertbutyl-4-methylphénol)] qui apparaissent en dehors de la limite fixée par la valeur critique  $h^* = 0,122$  symbolisée par la droite parallèle à l'axe des  $e_{i \text{ std}}$ .

### V-3- Qualité de l'ajustement :

La qualité de l'ajustement peut être appréciée dans la figure 15 par les deux droites d'ajustements dont les équations sont :

- Pour l'ensemble de calibration

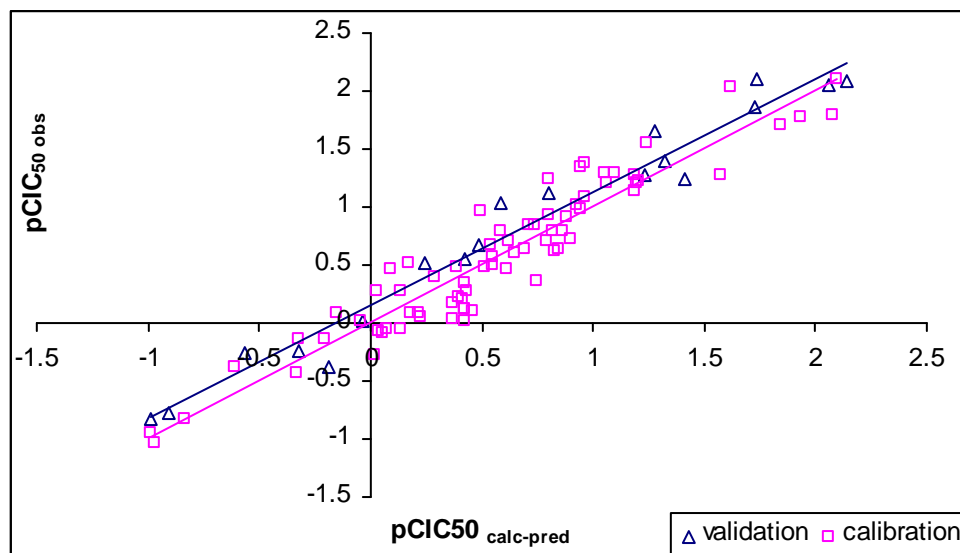
$$pCIC_{50 \text{ obs}} = -0,0000012 + 0,99999 pCIC_{50 \text{ calc}} \quad (\text{eq, 40})$$

$$s = 0,213889 \quad R^2 = 89,7 \% \quad R^2_{\text{ajust}} = 89,5 \%$$

- Pour l'ensemble de validation

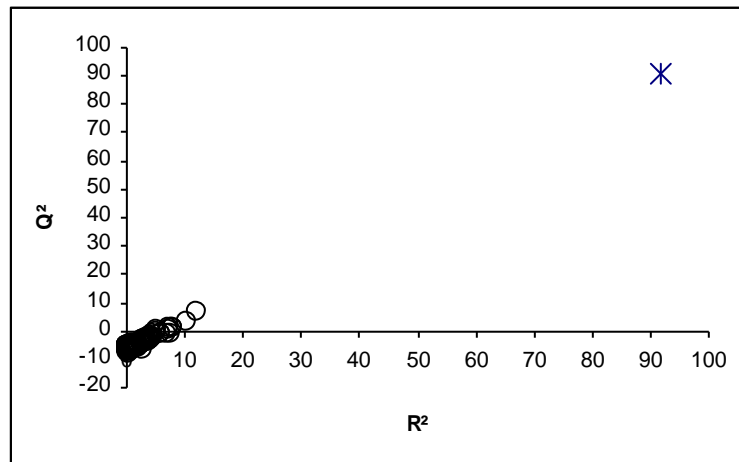
$$pCIC_{50 \text{ obs}} = 0,154020 + 0,980558 pCIC_{50 \text{ pre}} \quad (\text{eq, 41})$$

$$s = 0,182001 \quad R^2 = 96,7 \% \quad R^2_{\text{ajust}} = 96,5 \%$$



**Figure -15:** Droites d'ajustement pour le modèle 05.

Par le test de randomisation (figure 16), nous nous sommes assuré qu'une relation structure- activité réelle a été établie par notre modèle.



**Figure -16:** *Test de randomisation pour le modèle 05.*

Les 100 modèles pour lesquels nous avons randomisé les valeurs du  $pCIC_{50}$  ont des valeurs de  $Q^2$  faibles et des valeurs du coefficient de corrélation multiple ( $R^2$ ) petites. Seul le modèle avec le vecteur réel (étoile) offre des valeurs élevées pour les deux statistiques représentées dans la figure 16.

# CHAPITRE 1

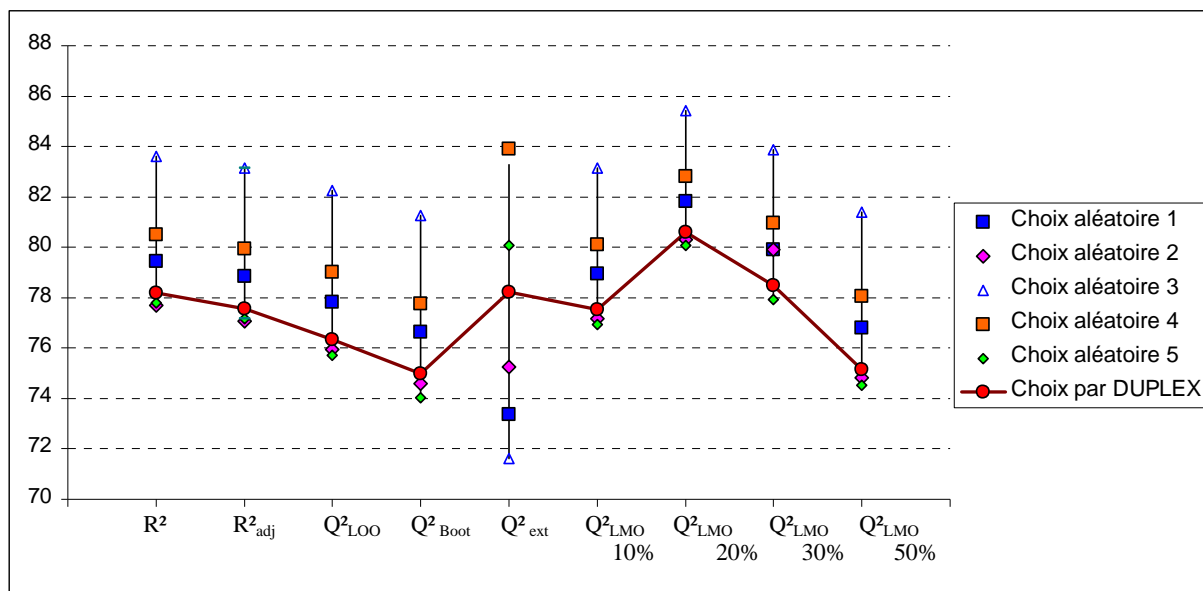
### Comparaison entre les modèles obtenus par les différentes répartitions:

Dans ce chapitre nous allons comparer les paramètres statistiques obtenus par les différentes répartitions c'est-à-dire les cinq choix aléatoires et le choix effectué à l'aide de l'algorithme DUPLEX.

#### 1- Modèle MLOGP

**Tableau-12 :** Paramètres statistiques de différents choix pour le modèle MLOGP

	Choix aléatoire 1	Choix aléatoire 2	Choix aléatoire 3	Choix aléatoire 4	Choix aléatoire 5	Choix par DUPLEX	
$n_{tr}$	19.00	19.00	19.00	19.00	19.00	19.00	
$n_{ext}$	74.00	74.00	74.00	74.00	74.00	74.00	
$R^2$	79.44	77.70	83.62	80.49	77.78	78.19	
$Q^2_{LOO}$	77.83	75.95	82.26	79.02	75.71	76.34	
$Q^2_{Boot}$	76.65	74.59	81.27	77.76	74.04	74.98	
$R^2_{adj}$	78.86	77.07	83.15	79.94	77.16	77.57	
$Q^2_{ext}$	73.37	75.26	71.62	83.89	80.07	78.23	
$F$	123.67	123.67	181.17	146.49	124.27	127.25	
$s$	0.33	0.33	0.31	0.34	0.34	0.33	
$EQMC$	0.32	0.32	0.30	0.33	0.33	0.32	
$EQMP$	0.33	0.33	0.32	0.34	0.35	0.34	
$EQMP_{ext}$	0.36	0.34	0.40	0.30	0.31	0.32	
$Q^2_{LMO}$	10%	78.95	77.15	83.16	80.10	76.92	77.52
	20%	81.83	80.32	85.43	82.82	80.08	80.59
	30%	79.91	79.90	83.88	80.96	77.91	78.49
	50%	76.81	74.83	81.41	78.05	74.53	75.15



**Figure -17 :** Variation des paramètres statistiques selon les choix pour le modèle MLOGP



## 2- Modèle ALOGP :

Tableau-13 : Paramètres statistiques de différents choix pour le modèle ALOGP

		Choix aléatoire 1	Choix aléatoire 2	Choix aléatoire 3	Choix aléatoire 4	Choix aléatoire 5	Choix par DUPLEX
$n_{tr}$		19.00	19.00	19.00	19.00	19.00	19.00
$n_{ext}$		74.00	74.00	74.00	74.00	74.00	74.00
$R^2$		84.64	82.17	87.47	84.92	82.42	81.42
$Q^2_{LOO}$		83.47	80.89	86.50	83.86	80.91	79.86
$Q^2_{Boot}$		82.55	79.76	85.72	83.09	80.01	78.69
$Q^2_{ext}$		79.15	83.15	77.31	86.72	86.85	85.64
$R^2_{adj}$		84.21	81.67	87.11	84.49	81.92	80.90
$F$		195.66	163.61	247.73	199.87	166.43	155.59
$s$		0.28	0.29	0.27	0.30	0.30	0.30
$EQMC$		0.28	0.29	0.27	0.29	0.29	0.30
$EQMP$		0.29	0.30	0.28	0.30	0.31	0.31
$EQMP_{ext}$		0.32	0.28	0.36	0.27	0.25	0.26
$Q^2_{LMO}$	10%	84.31	81.87	87.19	84.69	81.87	80.86
	20%	86.46	84.36	88.94	86.78	84.34	83.50
	30%	85.01	82.66	87.76	85.38	82.66	81.68
	50%	82.76	80.04	85.88	83.19	80.09	78.86

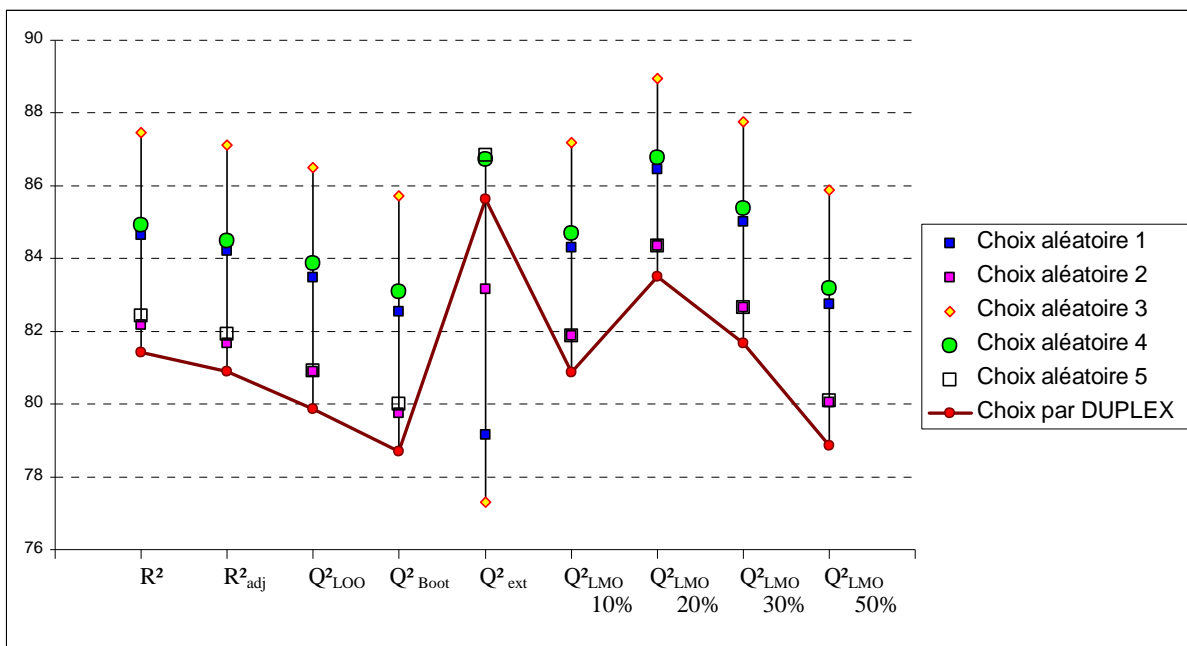


Figure-19 : Variation des paramètres statistiques selon les choix pour le modèle ALOGP.



## 3- Modèles CLOGP :

Tableau -14 : Paramètres statistiques de différents choix pour le modèle CLOGP.

	Choix aléatoire 1	Choix aléatoire 2	Choix aléatoire 3	Choix aléatoire 4	Choix aléatoire 5	Choix par DUPLEX	
$n_{tr}$	19.00	19.00	19.00	19.00	19.00	19.00	
$n_{ext}$	74.00	74.00	74.00	74.00	74.00	74.00	
$R^2$	88.69	87.98	91.64	90.04	90.04	88.20	
$Q^2_{LOO}$	87.77	87.02	90.94	89.29	89.06	87.30	
$Q^2_{Boot}$	87.01	86.36	90.51	88.68	88.47	86.61	
$Q^2_{ext}$	92.22	90.83	93.40	92.71	83.63	89.56	
$R^2_{adj}$	88.37	87.64	91.40	89.76	89.76	87.87	
$F$	278.33	259.85	389.03	321.09	320.92	265.30	
$s$	0.24	0.24	0.22	0.24	0.23	0.24	
EQMC	0.24	0.24	0.22	0.24	0.22	0.23	
EQMP	0.25	0.25	0.23	0.25	0.23	0.24	
EQMP <sub>ext</sub>	0.20	0.21	0.19	0.20	0.28	0.22	
$Q^2_{LMO}$	10%	88.38	87.66	93.49	89.83	88.59	87.93
	20%	89.99	89.36	94.39	91.23	90.15	89.59
	30%	88.90	88.20	93.77	90.31	89.14	88.50
	50%	87.22	86.46	92.82	88.85	87.49	86.74

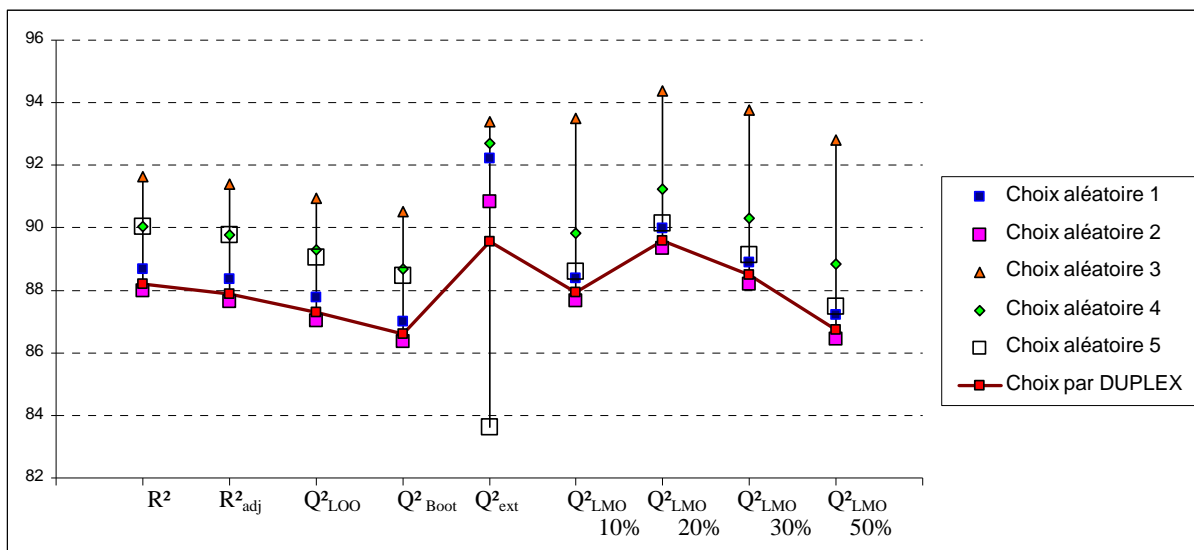
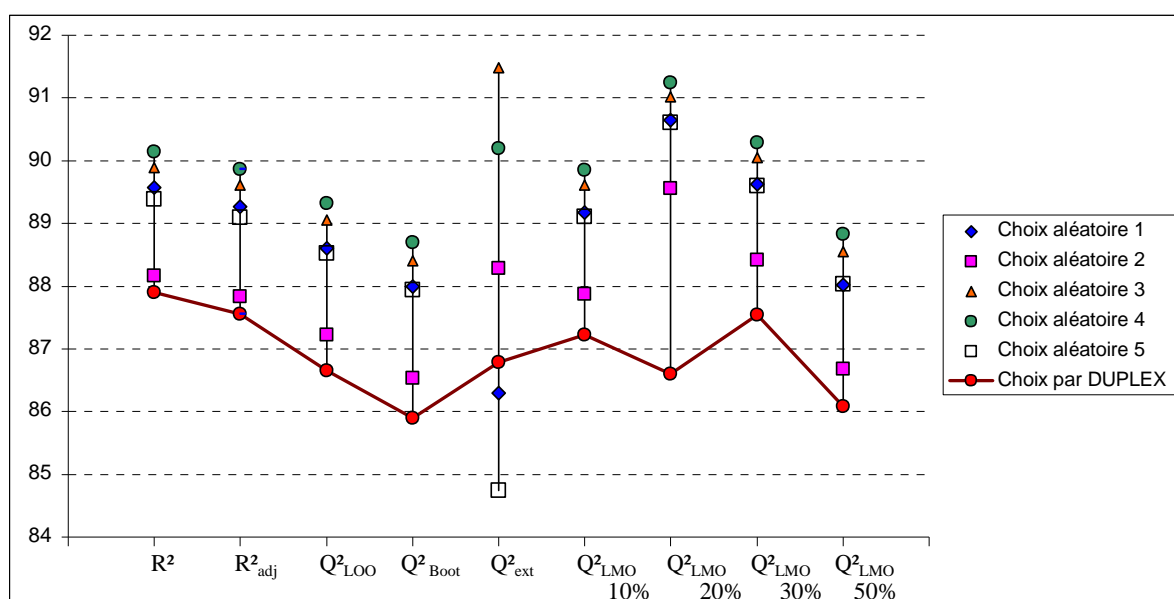


Figure-21: Variation des paramètres statistiques selon les choix pour le modèle CLOGP.



4- Modèle LOGK<sub>owwin</sub>:Tableau-15 : Paramètres statistiques de différents choix pour le modèle LOGK<sub>owwin</sub>

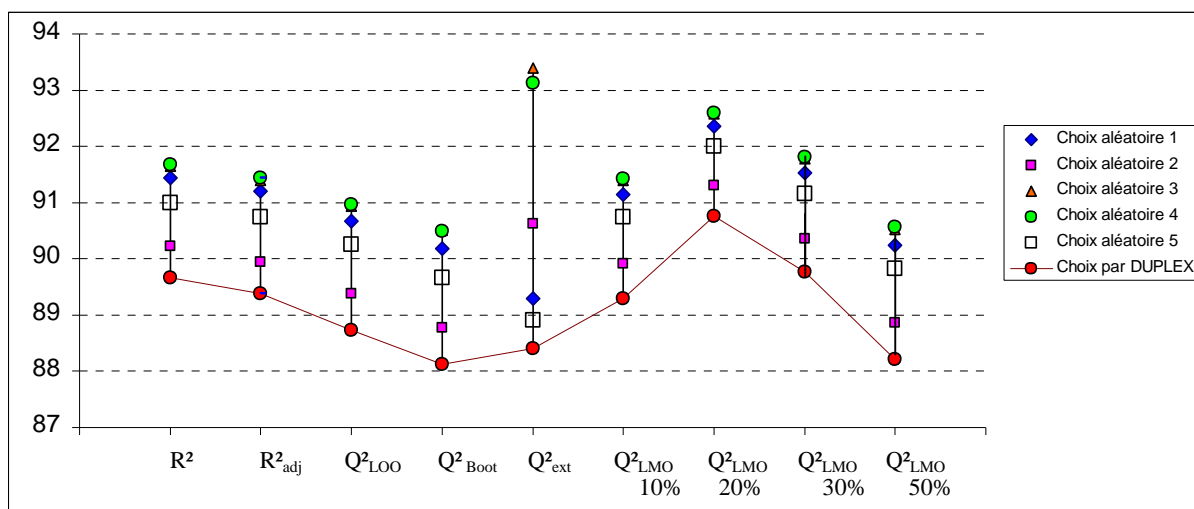
		Choix aléatoire 1	Choix aléatoire 2	Choix aléatoire 3	Choix aléatoire 4	Choix aléatoire 5	Choix par DUPLEX
$n_{tr}$		19.00	19.00	19.00	19.00	19.00	19.00
$n_{ext}$		74.00	74.00	74.00	74.00	74.00	74.00
$R^2$		89.57	88.16	89.89	90.14	89.39	87.90
$Q^2_{LOO}$		88.60	87.23	89.06	89.32	88.53	86.66
$Q^2_{Boot}$		87.99	86.53	88.40	88.70	87.94	85.90
$Q^2_{ext}$		86.30	88.29	91.48	90.20	84.74	86.79
$R^2_{adj}$		89.27	87.83	89.61	89.86	89.09	87.56
F		304.78	264.30	315.72	324.45	299.15	257.92
s		0.23	0.24	0.24	0.24	0.23	0.24
EQMC		0.23	0.24	0.24	0.24	0.23	0.24
EQMP		0.24	0.24	0.25	0.25	0.24	0.25
EQMP <sub>ext</sub>		0.26	0.23	0.22	0.24	0.27	0.25
$Q^2_{LMO}$	10%	89.17	87.87	89.61	89.85	89.11	87.22
	20%	90.65	89.56	91.02	91.25	90.60	86.61
	30%	89.63	88.42	90.06	90.29	89.59	87.55
	50%	88.02	86.68	88.56	88.84	88.03	86.08

Figure-23: Variation des paramètres statistiques selon les choix pour le modèle LOGK<sub>owwin</sub>



5- Modèle  $\text{LOGK}_{\text{ow}}$ :Tableau-16: Paramètres statistiques de différents choix pour le modèle  $\text{LOGK}_{\text{ow}}$ 

	Choix aléatoire 1	Choix aléatoire 2	Choix aléatoire 3	Choix aléatoire 4	Choix aléatoire 5	Choix par DUPLEX	
$n_{\text{tr}}$	19.00	19.00	19.00	19.00	19.00	19.00	
$n_{\text{ext}}$	74.00	74.00	74.00	74.00	74.00	74.00	
$R^2$	91.44	90.22	91.64	91.67	90.99	89.67	
$Q^2_{\text{LOO}}$	90.67	89.38	90.94	90.97	90.25	88.73	
$Q^2_{\text{Boot}}$	90.18	88.78	90.51	90.49	89.67	88.13	
$Q^2_{\text{ext}}$	89.29	90.62	93.40	93.12	88.91	88.41	
$R^2_{\text{adj}}$	91.20	89.94	91.40	91.44	90.74	89.38	
$F$	379.25	327.34	389.03	390.77	358.52	308.12	
$s$	0.21	0.22	0.22	0.22	0.21	0.22	
EQMC	0.21	0.21	0.22	0.22	0.21	0.21	
EQMP	0.22	0.22	0.23	0.23	0.22	0.22	
EQMP <sub>ext</sub>	0.23	0.21	0.19	0.20	0.23	0.22	
$Q^2_{\text{LMO}}$	10%	91.14	89.91	91.39	91.43	90.75	89.29
	20%	92.35	91.31	92.57	92.60	92.01	90.75
	30%	91.52	90.35	91.78	91.81	91.15	89.77
	50%	90.24	88.86	90.52	90.57	89.82	88.22

Figure-25: Variation des paramètres statistiques selon les choix pour le modèle  $\text{LOGK}_{\text{ow}}$ .



**Discussion :**

D'après les résultats obtenus pour les différents choix (résumés dans les tableaux 12, 13, 14, 15 et 16) on trouve qu'il y a une différence notable, visible dans les figures 17, 19, 21, 23, et 25. On remarque aussi que l'intervalle de variation pour les paramètres statistiques  $R^2$ ,  $R^2_{adj}$ ,  $Q^2_{LOO}$ ,  $Q^2_{Boot}$ , et  $Q^2_{LMO}$  (10%,20%,30%et 50%) varie d'une manière constante selon le choix. Par contre, l'intervalle de variation pour les différentes valeurs de  $Q^2_{ext}$  est plus grand. Il est presque le double des premiers ( $R^2$ ,  $R^2_{adj}$ ,  $Q^2_{LOO}$ ,  $Q^2_{Boot}$  et  $Q^2_{LMO}$ ). Le tableau-19 montre les variations des paramètres selon les choix

**Tableau-17** : Variation des paramètres statistiques selon le choix.

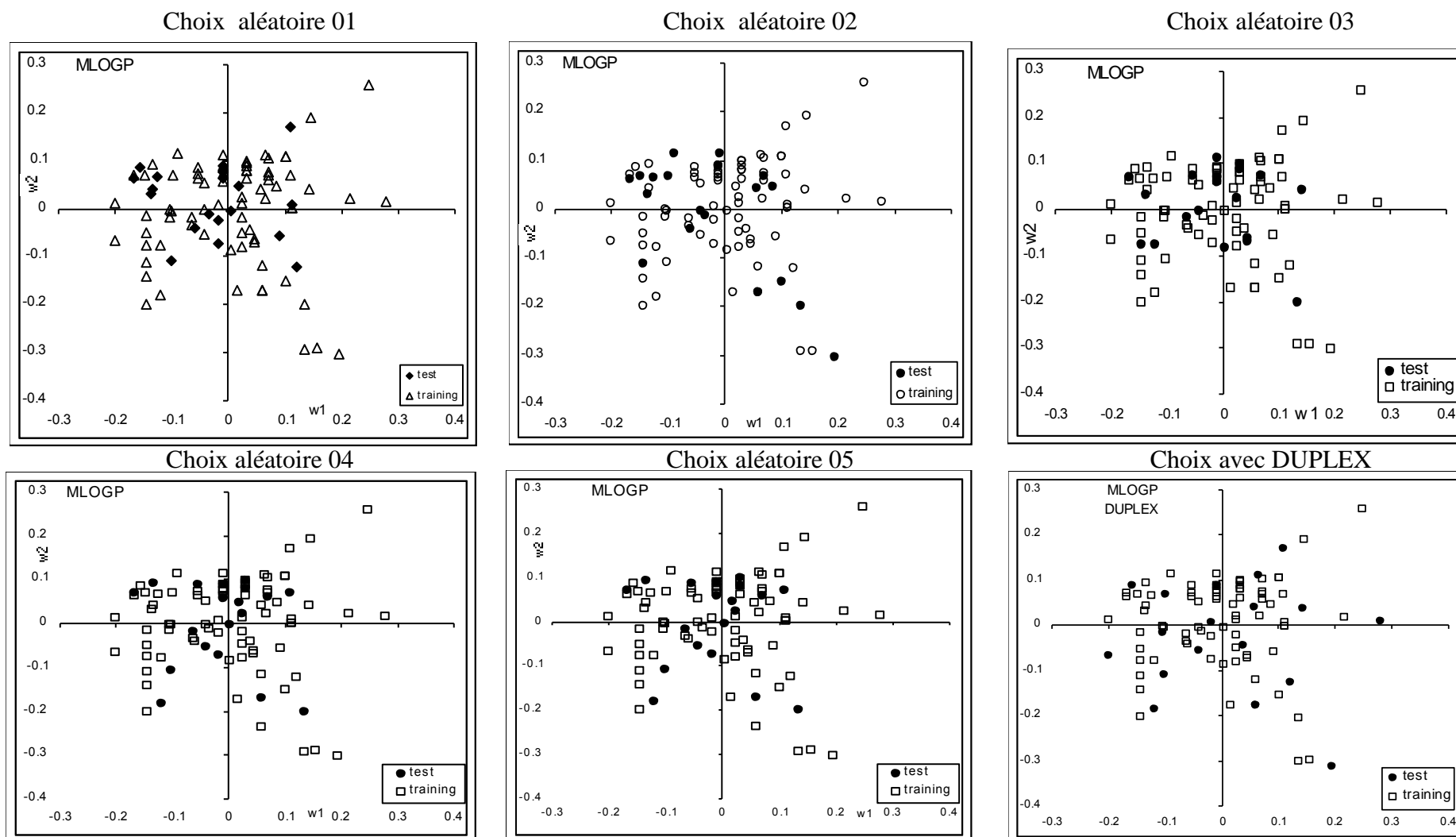
		MLOGP	ALOGP	CLOGP	LOGK <sub>owwin</sub>	LOGK <sub>ow</sub>
$R^2$		5,84	6,05	3,66	2,24	2
$R^2_{adj}$		5,99	6,21	3,76	2,3	2,06
$Q^2_{LOO}$		6,55	6,64	3,92	2,4	2,24
$Q^2_{Boot}$		7,23	7,03	4,15	2,8	2,36
$Q^2_{ext}$		11,67	9,54	9,77	6,74	4,99
$Q^2_{LMO}$	10 %	6,23	6,33	5,83	2,63	2,14
	20 %	4,63	5,44	4,8	4,64	1,85
	30 %	5,97	6,08	5,57	2,74	1,04
	50%	6,88	7,02	6,36	2,76	2,35

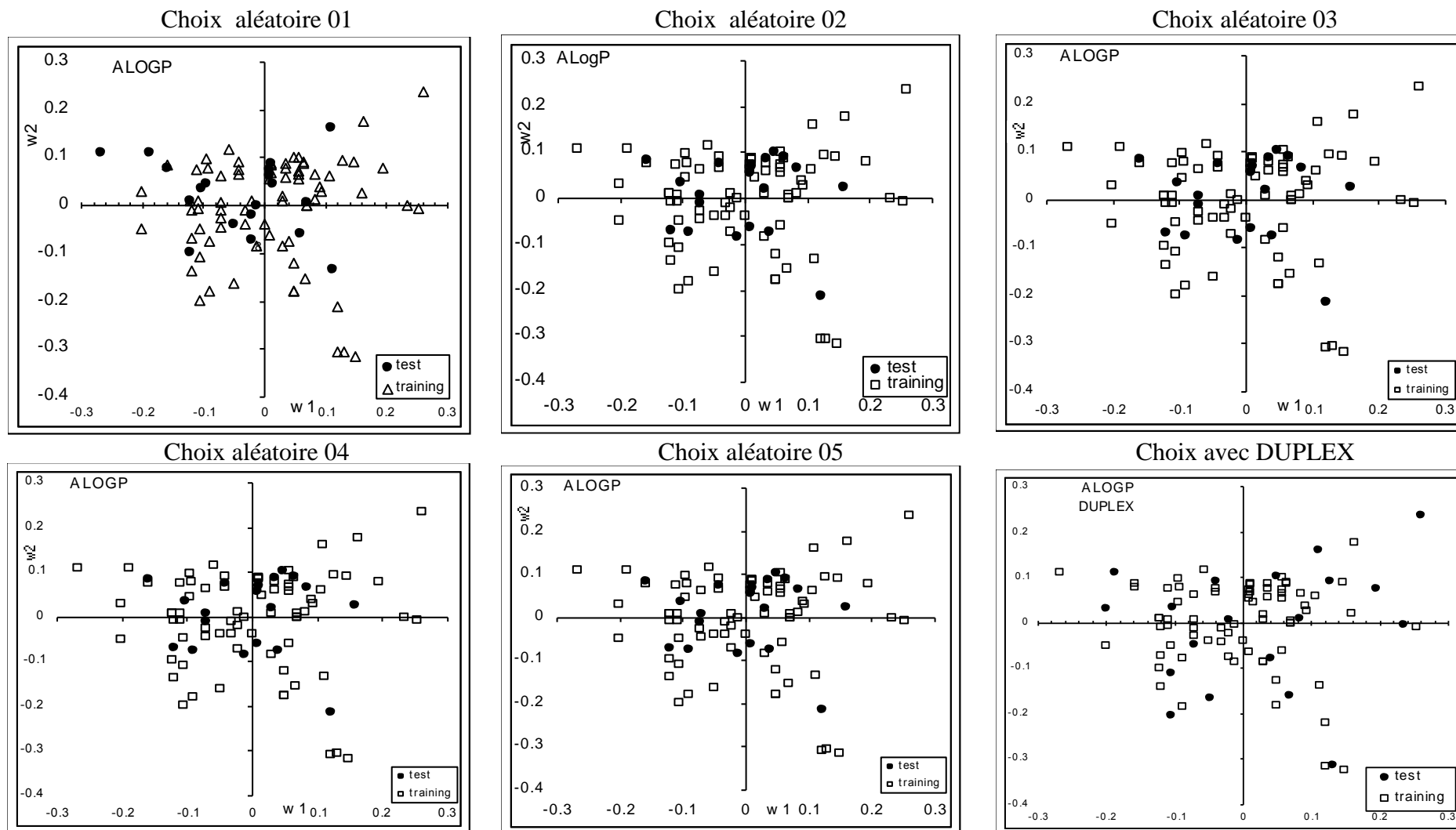
$\Delta$  : est la différence entre la meilleure et la mauvaise valeur de  $R^2$ ,  $R^2_{adj}$ ,  $Q^2_{LOO}$ ,  $Q^2_{Boot}$ ,  $Q^2_{ext}$  et  $Q^2_{LMO}$

**Conclusion :**

Les choix aléatoires peuvent donner de bons ou de mauvais résultats, mais l'algorithme DULEX donne un unique résultat. La grande différence entre les valeurs de  $Q^2_{\text{ext}}$  ( $Q^2_{\text{ext}} = 10\%$ ), par rapport aux autres paramètres est expliquée par la manière de sélectionner des deux ensembles (validation, calibration), les figures 18, 20, 22, 24, et 26 montrent la distribution des points de validation et calibration. Dans le cas du DUPLEX, les points sont répartis de façon homogène il sont distribués et représentés presque dans tout le domaine de définition des variables explicatives.

Mais pour les choix aléatoires les points sont regroupés soit au centre ou bien dans une seule région. On conclut que le modèle par DUPLEX est validé sur un ensemble de composés homogène, qui explique l'utilisation de l'algorithme DUPLEX pour la sélection de l'ensemble de validation. L'ensemble de calibration lui aussi est plus homogène, pour cette raison on a des paramètres statistiques de quelques choix aléatoires meilleurs que celui de DUPLEX, ce sont des choix heureux, on dit que ces modèles (choix aléatoire) sont construits et validés sur des ensembles serrés.

Figure-17 Répartition des ponts ( $w_1, w_2$ ) pour les cinq choix aléatoires et le choix par algorithme DUPLEX du modèle MLOGP.

Figure-20: Répartition des ponts  $(w_1, w_2)$  pour les cinq choix aléatoires et le choix par algorithme DUPLEX du modèle ALOGP.

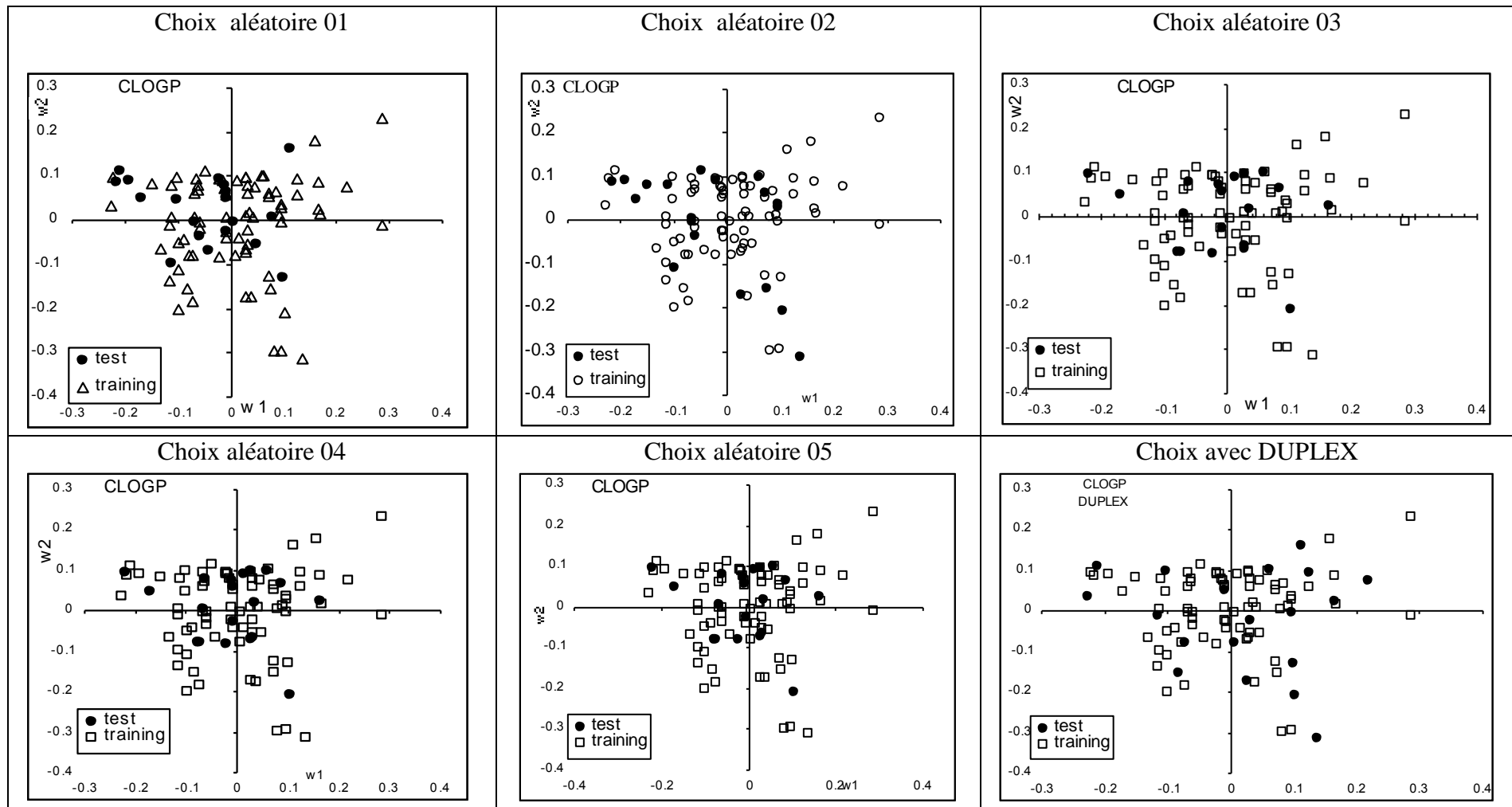
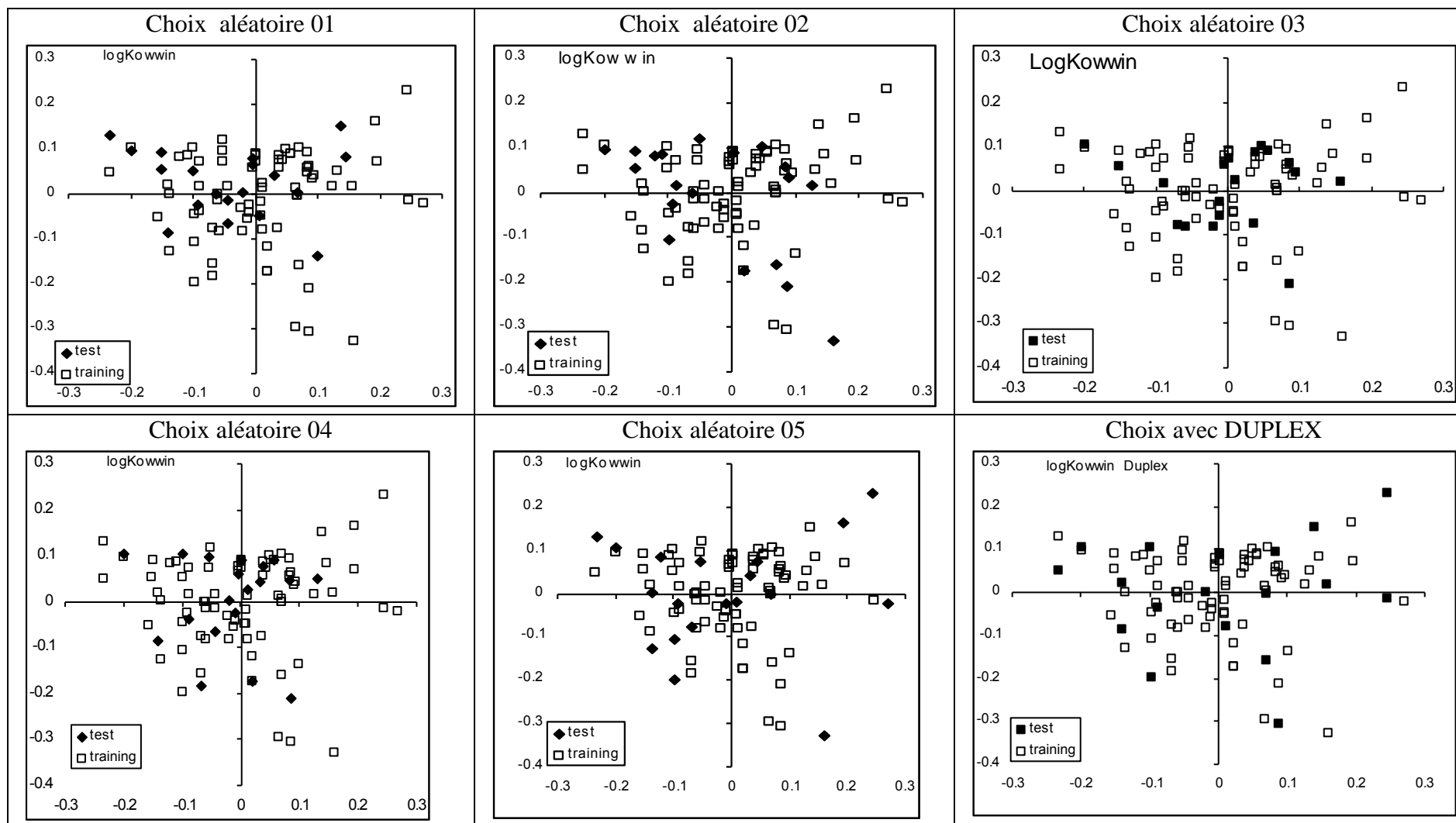
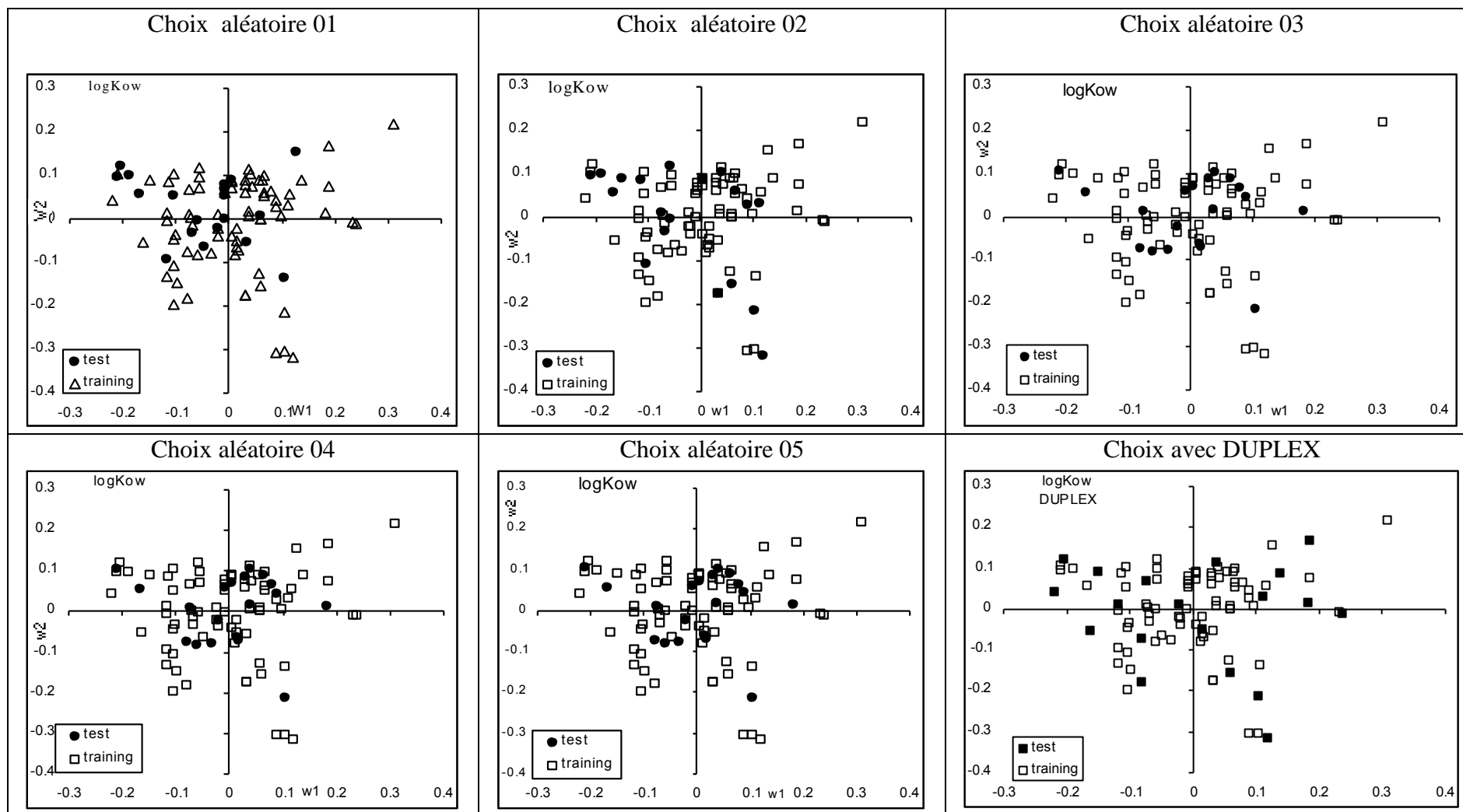


Figure-22: Répartition des ponts ( $w_1$ ,  $w_2$ ) pour les cinq choix aléatoires et le choix par algorithme DUPLEX du modèle CLOGP.

Figure-24: Répartition des ponts ( $w_1, w_2$ ) pour les cinq choix aléatoires et le choix par algorithme DUPLEX du modèle LOGKowwin.

Figure-26: Répartition des ponts ( $w_1, w_2$ ) pour les cinq choix aléatoires et le choix par algorithme DUPLEX du modèle LOGKow



# CHAPITRE 2

### Comparaison entre les modèles obtenus par les différents logP:

D'après les principes de base pour l'établissement des modèles QSAR, les valeurs des descripteurs doivent être obtenues de la même source ou calculées en utilisant le même logiciel, donc il est impératif de sélectionner, dès le départ une seule méthode théorique sur laquelle on calcule un descripteur qui peut donner les meilleurs résultats. Pour cela, et comme on a utilisé quatre logiciels différents qui sont basés sur différentes approches théoriques pour calculer les valeurs de logP, on fait une comparaison entre les différents modèles (MLOGP, ALOGP, CLOGP et LOGK<sub>owwin</sub>) afin d'obtenir le meilleur modèle. Le modèle LOGK<sub>ow</sub>, puisé dans la littérature, est employé comme référence.

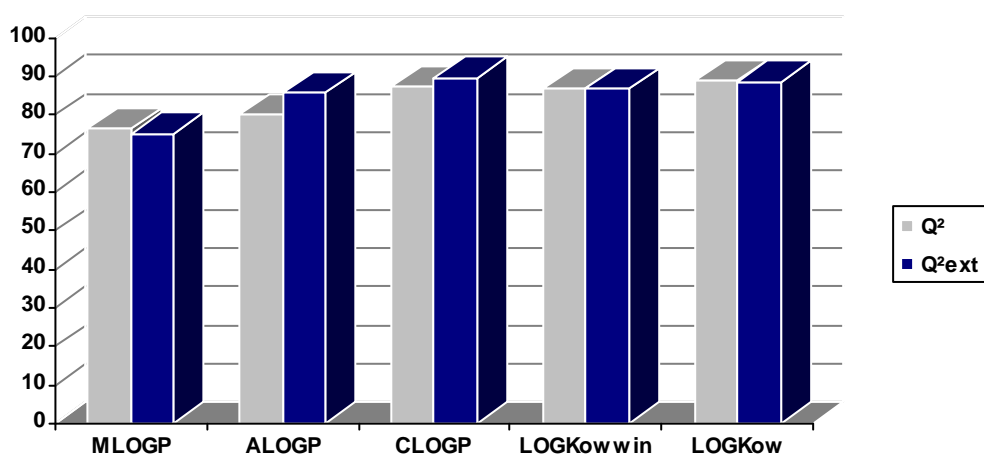
Puisque la sélection des deux ensembles (19 pour calibration et 74 pour validation), avec l'algorithme DUPLEX donne des résultats stables et uniques, nous allons utiliser les modèles précédemment développés.

#### 1- Discussion:

**Tableau-18:** Valeurs de  $Q^2$  et  $Q^2_{ext}$  de chaque modèle.

	MLOGP	ALOGP	CLOGP	LOGK <sub>owwin</sub>	LOGK <sub>ow</sub>
$Q^2$	76,34	79,86	87,3	86,66	88,73
$Q^2_{ext}$	74,98	85,64	89,56	86,79	88,41

**Histogramme:** Représentation des Valeurs  $Q^2$  et  $Q^2_{ext}$ .



D'après l'histogramme on remarque que les valeurs  $Q^2$  et  $Q^2_{ext}$  pour les deux modèles  $LOGK_{ow}$  et  $LOGK_{owwin}$  sont plus grandes que celles des autres modèles.

On utilise les diagrammes de Williams des modèles MLOGP, CLOGP, ALOGP et  $LOGK_{owwin}$  (figure-27), pour détecter les points aberrants et les points influents. Les résultats obtenus sont regroupés dans le tableau 13.

**Tableau-19** : Points aberrants et influents pour les modèles MLOGP, CLOGP, ALOGP et  $LOGK_{owwin}$ .

	Points aberrants	Points influents
MLOGP	4-hydroxyphenylmethane 4-heptyloxyphenol	2,4,6-tribromophenol 2,4,5-trichlorophenol 2,6-diphenylphenol ethyl-4-hydroxybenzoate
ALOGP	4-hydroxyphenylmethane	2,4,6-tribromophenol 2,4,6-trichlorophenol 4-bromo-2,6-dichlorophenol 2,6-ditertbutyl-4-methylphenol 4-hydroxyphenethyl alcohol
CLOGP	4-hydroxyphenylmethane	2,4,6-tribromophenol 2,4,6-trichlorophenol 4-bromo-2,6-dichlorophenol 2,6ditertbutyl-4-methylphenol 4-tertoctylphenol
$LOGK_{owwin}$	aucun	2,4,6-tribromophenol 2,4,6-trichlorophenol 4-bromo-2,6-dichlorophenol 2,6ditertbutyl-4-methylphenol 4-tertoctylphenol
$LOGK_{ow}$	aucun	2, 4,6-tribromophenol 2, 4,6-trichlorophenol 4-bromo-2,6-dichlorophenol 2,6ditertbutyl-4-methylphenol

On trouve qu'il n'y a aucun point aberrant dans le modèle  $LOGK_{owwin}$ , par contre il y en a un dans le modèle CLOGP, un dans le modèle ALOGP et deux dans le modèle MLOGP, le point représentant le composé (4-hydroxyphenylmethane), est aberrant dans les trois premier modèle mais pas pour le modèle  $LOGK_{owwin}$ ,



## 2- Conclusion:

Parmi les quatre modèles développés, le meilleur est  $\text{LOGK}_{\text{owwin}}$ , car ne présente aucun point aberrant, et ces paramètres statistiques sont bons par rapport aux autres (MLOGP, ALOGP, et CLOGP). On trouve aussi que les résultats obtenus par le modèle  $\text{LOGK}_{\text{owwin}}$  sont plus proches des résultats du modèle de référence ( $\text{LOGK}_{\text{ow}}$ ).

Il y a d'autres travaux qui font la comparaison entre les différents logP, basée sur le mode d'action de la toxicité (MDA), dans ce cas le choix n'est pas facile, parce qu'il faut classer les composés selon leur mode d'action. Le problème est que ces modes ne sont pas clairement définis, ce sont, généralement, des hypothèses [95]. Dans ces travaux les composés sont d'abord classés selon leurs modes d'action, puis le meilleur logP est sélectionné en complétant avec d'autres descripteurs.

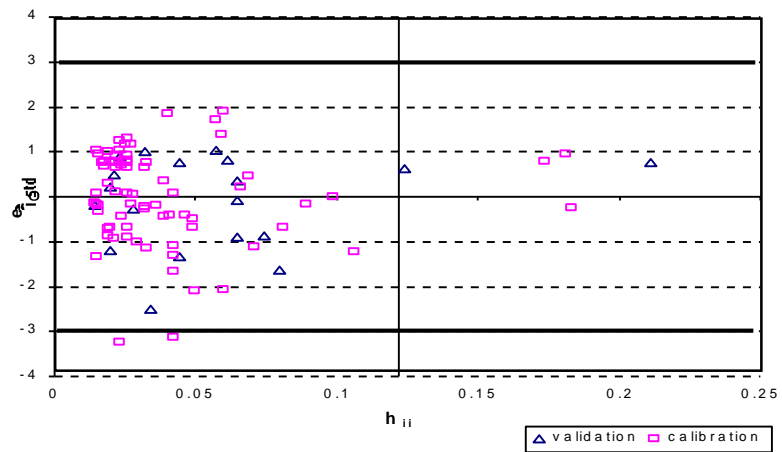


Diagramme de Williams pour le modèle MLOGP.

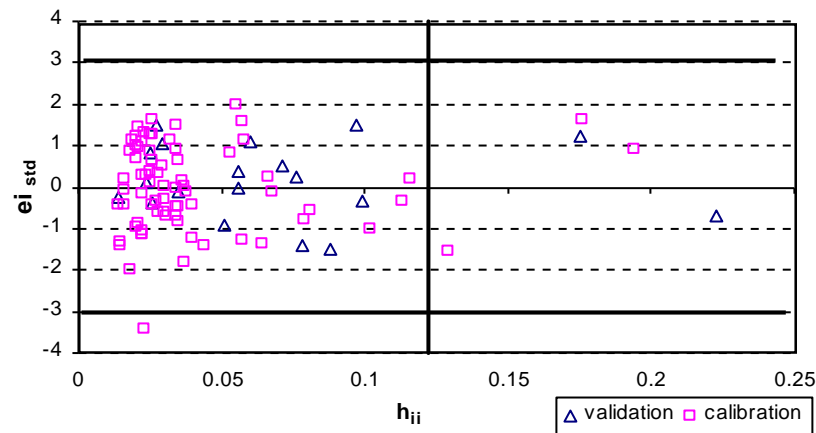


Diagramme de Williams pour le modèle ALOGP.

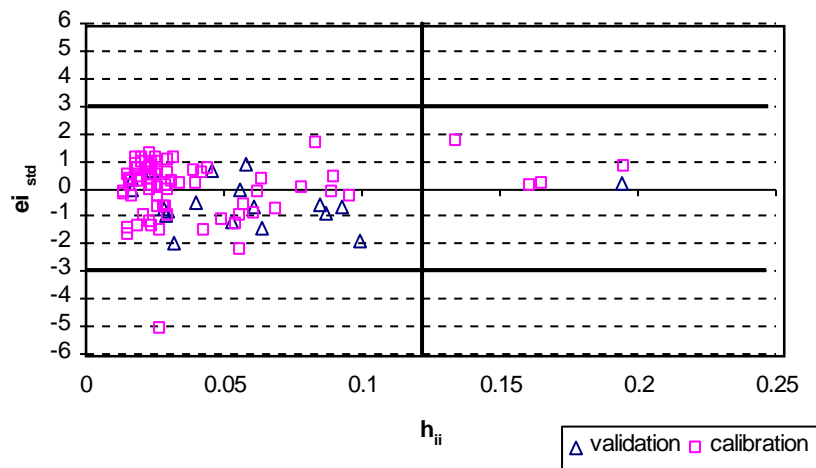


Diagramme de williams pour le modèle CLOGP

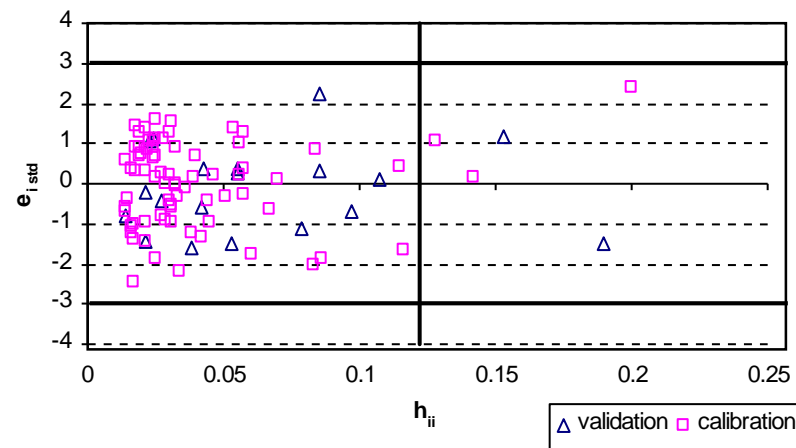


Diagramme de Williams pour le modèle LOGK<sub>owwin</sub>.

**Figure 27:** Diagrammes de Williams pour les différents modèles.



# CONCLUSION GENERALE

## Conclusion:

En conclusion, les QSAR constituent un outil puissant d'information sur les caractéristiques probables d'une substance à condition d'être maniés et interprétés par un utilisateur averti. Ils peuvent permettre d'éviter des tests expérimentaux ou de ne les effectuer que lorsque l'interprétation des résultats laisse planer une suspicion. Ils peuvent permettre aussi de compléter l'information expérimentale. Les QSAR sont faciles à mettre en oeuvre, ils fournissent un résultat immédiat et leurs coûts sont sans commune mesure avec ceux engendrés par la réalisation de tests expérimentaux.

Ils peuvent être réalisés par des laboratoires équipés des logiciels adaptés ou par le demandeur lui-même s'il dispose des outils informatiques et des personnes compétentes pour valider les résultats car on ne saurait trop insister sur les risques de conclusions hâtives.

Dans ce travail, la concentration d'inhibition de la croissance CIC50 d'une série de phénol a été reliée à deux paramètres physico-chimiques; le pKa reflétant l'acidité de ces phénols et le logarithme du coefficient de partage octanol-eau qui simule l'affinité des composés à la phase liquide ou organique.

La validation des QSAR passant d'abord (dans notre cas) par le choix des ensembles de training et test, vu que ce choix peut influencer les résultats. On utilise la méthode décrite par R. D. Snee pour cet éclatement. La méthode DUPLEX éprouve autant que possible (les données disponibles) les modèles construits. Il reflète avec plus de justesse, que le choix aléatoire, la qualité des modèles discutés.

Parmi les différentes mesures de l'affinité (logP), le meilleur modèle est celui avec LOGKowwin dont la disponibilité gratuite du logiciel, le calculant est un sérieux avantage. Le même résultat lorsque on a sélectionné les deux ensembles par l'algorithme DUPLEX.

Finalement, les objectifs principaux de cette thèse ont donc été remplis dans le sens où ils ont permis de mettre en évidence des méthodologies efficaces, en accord avec les règles mises en place par l'OCDE quant à la validation des modèles QSAR :

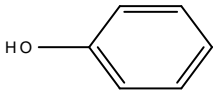
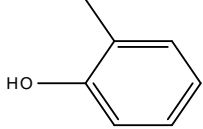
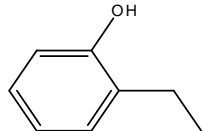
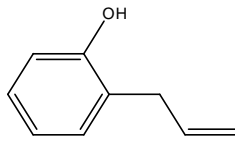
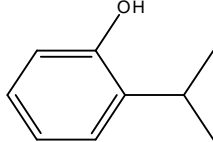
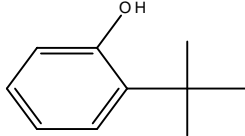
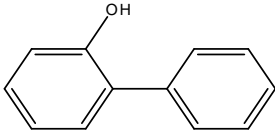
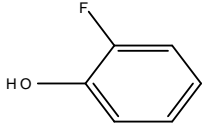
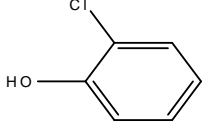
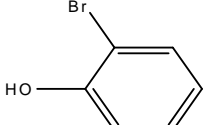
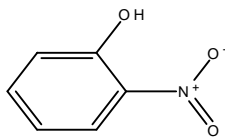
- importance de la définition et de la fiabilité de la base de données ;
- développement de modèles non ambigus à partir d'une équation et de descripteurs obtenus à un niveau de calcul suffisant et défini ;
- définition des domaines d'applicabilité des modèles ;

- mesures de la corrélation et de la robustesse des modèles et de leur pouvoir prédictif ;

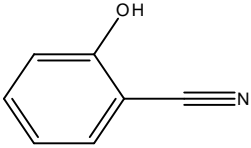
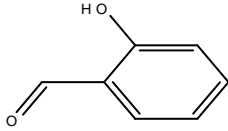
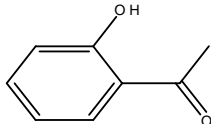
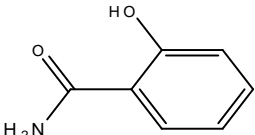
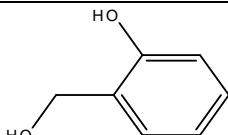
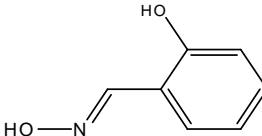
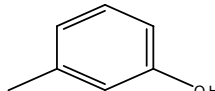
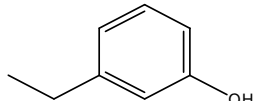
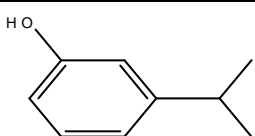
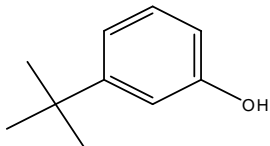
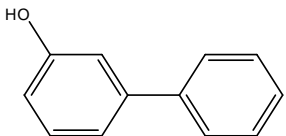
- mise en évidence des mécanismes sous-jacents par l'intégration de descripteurs quantiques d'une part et à l'aide de l'analyse des mécanismes réactionnels d'autre part.

# ANNEXES

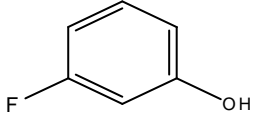
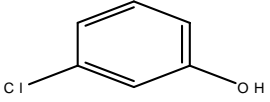
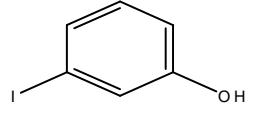
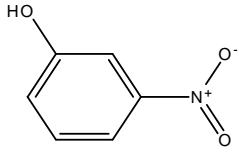
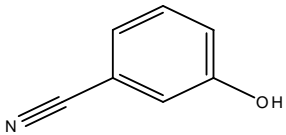
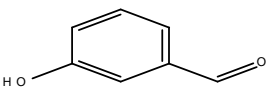
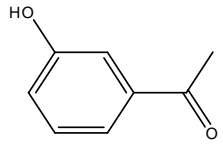
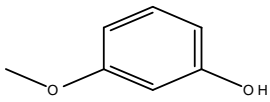
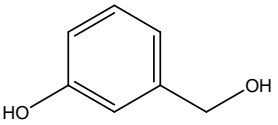
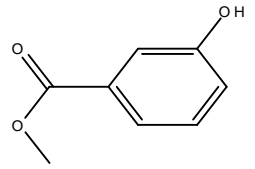
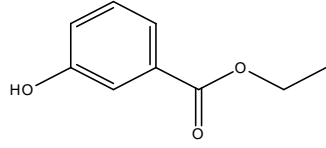
**Annexe-1** ; Toxicité relative, hydrophobie et valeur de pKa pour les phénols choisis.

N:	Numéro de CAS	Nom du composé	pCIC50	logK <sub>ow</sub>	pKa	Structure
01	108-95-2	Phénol	-0,43	1,48	9,99	
02	95-48-7	2-méthylphénol	-0,27	2,12	10,26	
03	90-00-6	2-éthylphénol	0,18	2,65	10,2	
04	1745-81-9	2-allylphénol	0,35	2,64	9,92	
05	88-69-7	2-isopropylphénol	0,8	3,05	10,4	
06	88-18-6	2-(tert)butylphénol	1,24	3,45	10,62	
07	90-43-7	2-phénylphénol	1,09	3,36	9,55	
08	367-12-4	2-fluorophénol	0,28	1,63	8,73	
09	95-57-8	2-chlorophénol	0,28	2,2	8,55	
10	95-56-7	2-bromophénol	0,5	2,35	8,45	
11	88-75-5	2-nitrophénol	0,67	1,85	7,22	

**Annexe-1 Suite**

N:	Numéro de CAS	Nom du composé	pCIC50	logK <sub>ow</sub>	pKa	Structure
12	611-20-1	2-cyanophénol	0,03	1,6	6,98	
13	90-02-8	2-hydroxybenzaldehyde	0,48	2,07	8,34	
14	118-93-4	2-acetylphénol	0,08	2,08	9,19	
15	65-45-2	2-hydroxybenzamide	-0,24	0,96	8,36	
16	90-01-7	2-hydroxybenzyl alcohol	-0,95	0,44	9,92	
17	94-67-7	2-hydroxybenzaldoxime	-0,25	1,1	9,92	
18	108-39-4	3-méthylphénol	-0,06	2,12	10	
19	620-17-7	3-éthylphénol	0,23	2,65	10,07	
20	618-45-1	3-isopropylphénol	0,61	3,05	10,1	
21	585-34-2	3-tertbutylphénol	0,73	3,45	10,12	
22	580-51-8	3-phénylphénol	1,35	3,36	9,63	

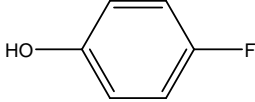
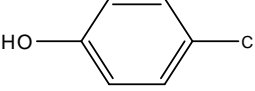
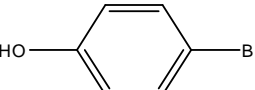
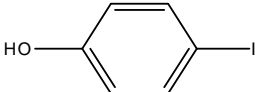
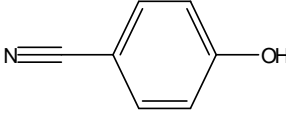
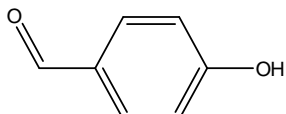
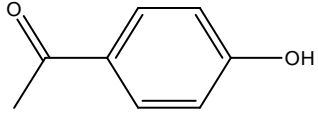
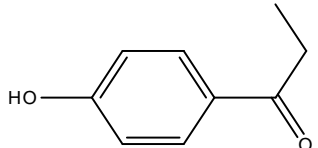
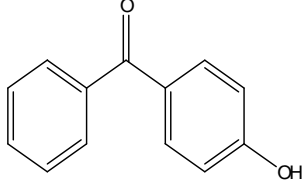
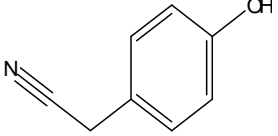
**Annexe-1 Suite**

N:	Numéro de CAS	Nom du composé	pCIC50	logK <sub>ow</sub>	pKa	Structure
23	372-20-3	3-fluorophénol	0,47	1,91	9,29	
24	108-43-0	3-chlorophénol	0,96	2,48	9,1	
25	626-02-8	3-iodophénol	1,12	2,89	8,88	
26	554-84-7	3-nitrophénol	0,51	1,85	8,36	
27	873-62-1	3-cyanophénol	-0,07	1,6	8,61	
28	100-83-4	3-hydroxybenzaldéhyde	0,08	1,44	9	
29	121-71-1	3-acétylphénol	-0,38	1,45	9,19	
30	150-19-6	3-méthoxyphénol	-0,14	1,57	9,65	
31	620-24-6	3-hydroxybenzyl alcool	-1,04	0,44	9,83	
32	19438-10-9	méthyl-3-hydroxybenzoate	-0,05	1,95	9,21	
33	7781-98-8	éthyl-3-hydroxybenzoate	0,48	2,51	9,09	

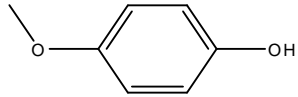
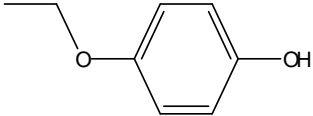
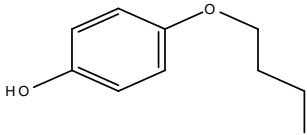
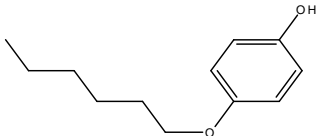
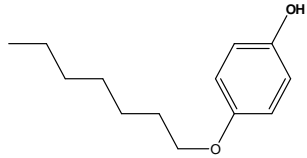
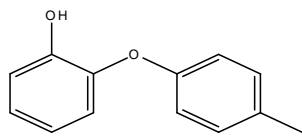
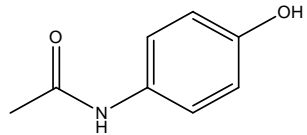
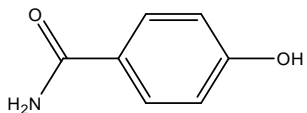
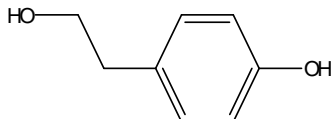
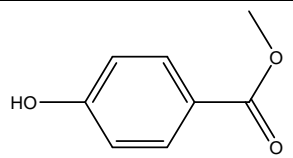
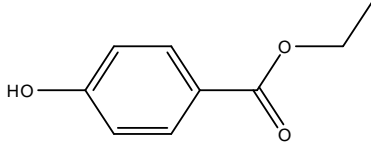
**Annexe-1 Suite**

N:	Numéro de CAS	Nom du composé	pCIC50	logK <sub>ow</sub>	pKa	Structure
34	123-07-9	4-ethylphénol	0,21	2,65	10	
35	645-56-7	4-propylphénol	0,64	3,18	10,3	
36	99-89-8	4-isopropylphénol	0,47	3,05	10,3	
37	99-71-8	4-secbutylphénol	0,98	3,58	10,3	
38	98-54-4	4-tertbutylphénol	0,91	3,45	10,23	
39	80-46-6	4-tertptylphénol	1,23	3,98	10,3	
40	1518-83-8	4-cyclopentylphénol	1,29	3,69	9,92	
41	92-69-3	4-phenylphénol	1,38	3,36	9,55	
42	101-53-1	4-hydroxyphenylmethane	1,2	3,69	10,1	
43	140-66-9	4-tertoctylphénol	2,09	5,31	9,92	
44	402-45-9	a,a,a-trifluoro-4-cresol	0,62	2,88	8,74	

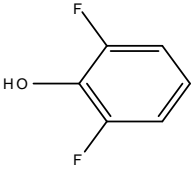
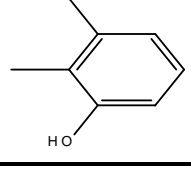
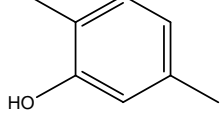
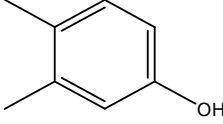
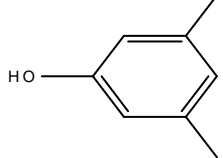
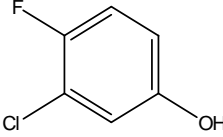
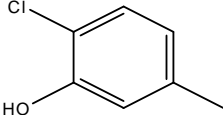
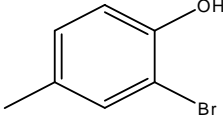
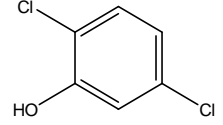
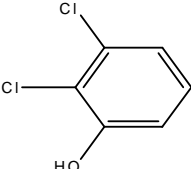
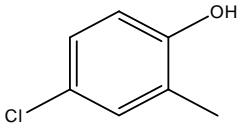
**Annexe-1 Suite**

N:	Numéro de CAS	Nom du composé	pCIC50	logK <sub>ow</sub>	pKa	Structure
45	371-41-5	4-fluorophénol	0,02	1,91	9,89	
46	106-48-9	4-chlorophénol	0,55	2,48	9,43	
47	106-41-2	4-bromophénol	0,68	2,63	9,34	
48	540-38-5	4-iodophénol	0,85	2,89	9,2	
49	767-00-0	4-cyanophénol	0,52	1,6	7,96	
50	123-08-8	4-hydroxybenzaldehyde	0,27	1,44	7,62	
51	99-93-4	4-acetylphénol	-0,09	1,45	8,05	
52	70-70-2	4'-hydroxypropiophenone	0,06	1,98	8,85	
53	1137-42-4	4-hydroxybenzophenone	1,02	3,08	8,89	
54	14191-95-8	4-hydroxybenzylcyanide	-0,38	0,9	9,52	

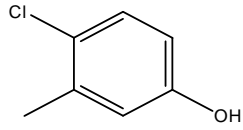
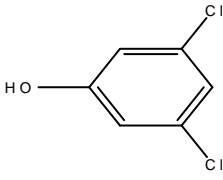
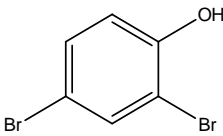
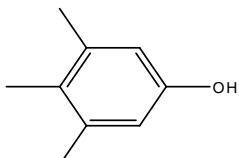
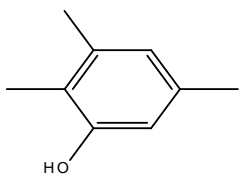
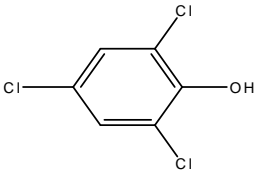
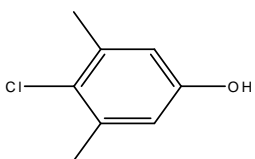
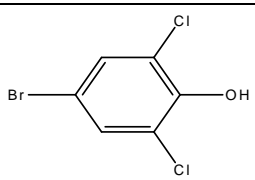
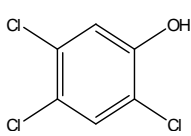
**Annexe-1 Suite**

N:	Numéro de CAS	Nom du composé	pCIC50	logK <sub>ow</sub>	pKa	Structure
55	150-76-5	4-methoxyphénol	-0,14	1,57	10,2	
56	622-62-8	4-ethoxyphénol	0,01	2,1	10,5	
57	122-94-1	4-butoxyphénol	0,7	3,16	10,6	
58	18979-55-0	4-hexyloxyphénol	1,65	4,22	10,7	
59	13037-86-0	4-heptyloxyphénol	2,03	4,75	10,7	
60	103-16-2	4-benzyloxyphénol	1,04	3,14	10,7	
61	103-90-2	4-acetoamidophénol	-0,82	0,49	10,108	
62	619-57-8	4-hydroxybenzamide	-0,78	0,33	9,23	
63	501-94-0	4-hydroxyphenethyl alcohol	-0,83	0,67	9,92	
64	99-76-3	methyl-4-hydroxybenzoate	0,08	1,98	9,05	
65	120-47-8	ethyl-4-hydroxybenzoate	0,57	2,51	8,92	

**Annexe-1 Suite**

N:	Numéro de CAS	Nom du composé	pCIC50	logK <sub>ow</sub>	pKa	Structure
66	28177-48-2	2,6-difluorophénol	0,4	1,65	7,518	
67	526-75-0	2,3-diméthylphénol	0,12	2,77	10,348	
68	95-87-4	2,5-diméthylphénol	0,01	2,77	10,34	
69	95-65-8	3,4-diméthylphénol	0,12	2,77	10,32	
70	108-68-9	3,5-diméthylphénol	0,11	2,77	10,15	
71	2613-23-2	3-chloro-4-fluorophénol	0,84	2,78	8,968	
72	615-74-7	2-chloro-5-méthylphénol	0,64	2,85	8,548	
73	6627-55-0	2-bromo-4-méthylphénol	0,79	2,91	8,67	
74	583-78-8	2,5-dichlorophénol	1,13	3,07	7,58	
75	576-24-9	2,3-dichlorophénol	1,27	3,07	7,58	
76	1570-64-5	4-chloro-2-méthylphénol	0,7	3,13	9,67	

**Annexe-1 Suite**

N:	Numéro de CAS	Nom du composé	pCIC50	logK <sub>ow</sub>	pKa	Structure
77	59-50-7	4-chloro-3-methylphénol	0,8	3,13	9,55	
78	591-35-5	3,5-dichlorophénol	1,56	3,35	8,18	
79	615-58-7	2,4-bromophénol	1,4	3,37	7,878	
80	527-54-8	3,4,5-trimethylphénol	0,93	3,42	10,508	
81	697-82-5	2,3,5-trimethylphénol	0,36	3,33	10,488	
82	88-06-2	2,4,6-trichlorophénol	1,7	3,69	6,35	
83	88-04-0	4-chloro-3,5-dimethylphénol	1,2	3,78	9,708	
84	697-86-9	4-bromo-2,6-dichlorophénol	1,78	3,84	6,4	
85	95-95-4	2,4,5-trichlorophénol	2,1	3,85	7,37	

**Annexe-1 Suite**

N:	Numéro de CAS	Nom du composé	pCIC50	logK <sub>ow</sub>	pKa	Structure
86	7530-27-0	4-bromo-6-chloro-2-méthylphénol	1,28	3,87	8,2	
87	2374-05-2	4-bromo-2,6-diméthylphénol	1,28	3,93	10,01	
88	118-79-6	2,4,6-tribromophénol	2,05	4,02	6,318	
89	2409-55-4	2-tertbutyl-4-méthylphénol	1,3	4,1	11,39	
90	89-68-9	4-chloro-2-isopropyl-5-méthylphénol	1,86	4,71	10,03	
91	1879-09-0	6-tertbutyl-2,4-diméthylphénol	1,25	4,75	11,68	
92	3-11-2423	2,6-diphénylphénol	2,11	5,25	9,928	
93	128-37-0	2,6ditertbutyl-4-méthylphénol	1,79	6,08	12,558	

**Annexe-2 : Valeurs de différents logP**

<b>N</b>	<b>Nom du composé</b>	<b>AlogP</b>	<b>MogP</b>	<b>ClogP</b>	<b>logK<sub>ow</sub>win</b>
01	Phénol	1,56	1,51	1,48	1,51
02	2-méthylphénol	2,05	1,86	1,92	2,06
03	2-éthylphénol	2,51	2,19	2,45	2,55
04	2-allylphénol	2,57	2,42	2,50	2,91
05	2-isopropylphénol	2,76	2,51	2,70	2,97
06	2-(tert)butylphénol	2,96	2,81	3,20	3,42
07	2-phenylphénol	3,08	3,13	2,86	3,28
08	2-fluorophénol	1,77	1,95	1,71	1,71
09	2-chlorophénol	2,23	2,13	2,15	2,16
10	2-bromophénol	2,31	2,30	2,35	2,40
11	2-nitrophénol	1,60	1,35	1,85	1,91
12	2-cyanophénol	1,44	1,16	1,60	1,61
13	2-hydroxybenzaldehyde	1,32	1,16	1,81	2,01
14	2-acetylphénol	1,30	2,00	1,91	1,97
15	2-hydroxybenzamide	0,57	0,73	1,28	1,03
16	2-hydroxybenzyl alcohol	0,96	0,98	0,39	0,60
17	2-hydroxybenzaldoxime	1,41	1,13	1,10	1,39
18	3-méthylphénol	2,05	1,86	1,97	2,06
19	3-éthylphénol	2,51	2,19	2,50	2,55
20	3-isopropylphénol	2,76	2,51	2,90	2,97
21	3-tertbutylphénol	2,96	2,81	3,30	3,42
22	3-phenylphénol	3,08	3,13	3,36	3,28
23	3-fluorophénol	1,77	1,95	1,91	1,71
24	3-chlorophénol	2,23	2,13	2,48	2,16
25	3-iodophénol	2,14	2,46	2,89	2,68
26	3-nitrophénol	1,60	1,35	1,85	1,91
27	3-cyanophénol	1,44	1,16	1,60	1,61
28	3-hydroxybenzaldahyde	1,32	1,16	1,44	1,23
29	3-acetylphénol	1,30	1,49	1,45	1,19
30	3-méthoxyphénol	1,55	1,25	1,57	1,59
31	3-hydroxybenzyl alcohol	0,96	0,98	0,44	0,60
32	méthyl-3-hydroxybenzoate	1,42	1,47	1,98	2,00
33	ethyl-3-hydroxybenzoate	1,77	1,78	2,51	2,49
34	4-éthylphénol	2,51	2,19	2,50	2,55
35	4-propylphénol	2,96	2,51	3,03	3,04
36	4-isopropylphénol	2,76	2,51	2,90	2,97
37	4-secbutylphénol	3,21	2,81	3,43	3,46

**Annexe-2: Suite,**

N°:	Nom du composé	AlogP	MogP	ClogP	logKowwin
38	4-terbutylphénol	2,96	2,81	3,30	3,42
39	4-terpentyphénol	3,42	3,11	3,83	3,91
40	4-cyclopentylphénol	3,30	2,71	3,54	3,84
41	4-phenylphénol	3,08	3,13	3,54	3,28
42	4-hydroxyphenylmethane	2,05	1,86	1,97	3,54
43	4-tertoctylphénol	4,79	3,92	5,42	5,28
44	a,a,a-trifluoro-4-cresol	2,51	2,62	2,88	2,48
45	4-fluorophénol	1,77	1,95	1,91	1,71
46	4-chlorophénol	2,23	2,13	2,48	2,16
47	4-bromophénol	2,31	2,30	2,63	2,40
48	4-iodophénol	2,14	2,46	2,89	2,68
49	4-cyanophénol	1,44	1,16	1,60	1,61
50	4-hydroxybenzaldehyde	1,32	1,16	1,44	1,23
51	4-acetylphénol	1,30	1,49	1,45	1,19
52	4'-hydroxypropiophenone	1,97	1,81	1,98	1,68
53	4-hydroxybenzophenone	2,97	2,97	3,05	2,67
54	4-hydroxybenzylcyanide	1,48	1,22	0,90	1,08
55	4-methoxyphénol	1,55	1,25	1,57	1,59
56	4-ethoxyphénol	1,90	1,58	2,10	2,08
57	4-butoxyphénol	2,88	2,20	3,16	3,07
58	4-hexyloxyphénol	3,79	2,77	4,22	4,05
59	4-heptyloxyphénol	4,24	3,05	4,75	4,54
60	4-benzyloxyphénol	3,61	3,05	3,82	3,30
61	4-acetoamidophénol	0,68	1,06	0,49	0,27
62	4-hydroxybenzamide	0,57	0,73	0,33	0,26
63	4-hydroxyphenethyl alcohol	0,05-	1,31	0,67	1,09
64	methyl-4-hydroxybenzoate	1,42	1,47	1,98	2,00
65	ethyl-4-hydroxybenzoate	1,77	1,78	2,51	2,49
66	2,6-difluorophénol	1,97	2,40	1,76	1,91
67	2,3-dimethylphénol	2,54	2,19	2,37	2,61
68	2,5-dimethylphénol	2,54	2,19	2,37	2,61
69	3,4-dimethylphénol	2,54	2,19	2,42	2,61
70	3,5-dimethylphénol	2,54	2,19	2,47	2,61
71	3-chloro-4-fluorophénol	2,43	2,56	2,73	2,36
72	2-chloro-5-methylphénol	2,71	2,46	2,65	2,70
73	2-bromo-4-methylphénol	2,80	2,62	2,85	2,95

**Annexe-2: Suite et fin**

N°:	Nom du composé	AlogP	MogP	ClogP	logKowwin
74	2,5-dichlorophénol	2,89	2,73	2,97	2,80
75	2,3-dichlorophénol	2,89	2,73	2,85	2,80
76	4-chloro-2-méthylphénol	2,71	2,46	2,93	2,70
77	4-chloro-3-méthylphénol	2,71	2,46	2,98	2,70
78	3,5-dichlorophénol	2,89	2,73	3,30	2,80
79	2,4-bromophénol	3,06	3,05	3,32	3,29
80	3,4,5-triméthylphénol	3,02	2,51	2,87	3,15
81	2,3,5-triméthylphénol	3,02	2,51	2,87	3,15
82	2,4,6-trichlorophénol	3,56	3,31	3,39	3,45
83	4-chloro-3,5-diméthylphénol	3,20	2,78	3,48	3,25
84	4-bromo-2,6-dichlorophénol	3,64	3,47	3,54	3,25
85	2,4,5-trichlorophénol	3,56	3,31	3,60	3,45
86	4-bromo-6-chloro-2-méthylphénol	3,46	3,20	3,57	3,59
87	4-bromo-2,6-diméthylphénol	3,28	2,93	3,53	3,50
88	2,4,6-tribromophénol	3,81	3,76	3,94	4,18
89	2-tertbutyl-4-méthylphénol	3,45	3,11	3,70	3,97
90	4-chloro-2-isopropyl-5-méthylphénol	3,91	3,37	4,21	4,16
91	6-tertbutyl-2-4-diméthylphénol	3,94	3,39	4,15	4,52
92	2,6-diphénylphénol	4,60	4,42	4,25	5,04
93	2,6ditertbutyl-4-méthylphénol	4,85	4,18	5,43	5,03

# Annexe (III)

## *Activité Biologique*

**I- INTRODUCTION**

**II- LA MITOCHONDRIE**

**III- LA MEMBRANE**

**IV- LA CHAÎNE RESPIRATOIRE**

## I- Introduction

Le mécanisme d'action des molécules organiques (dans notre cas les phénols) n'est pas connu de façon précise mais les hypothèses les plus sûres suggèrent que ces molécules perturbent le fonctionnement des membranes lipidiques, ou agissent en se liant à des protéines lipidiques.

Pour mener à bien cette étude, on doit connaître quelques principes de base de la biologie cellulaire, comme la membrane, la mitochondrie, afin de passer par un phénomène important pour expliquer l'action de toxicité.

## II- Les mitochondries :

La mitochondrie est souvent décrite comme «la centrale énergétique de la cellule ». Comme cette analogie est bonne, mais il est aussi bon d'insister sur le fait que, comme toutes les centrales énergétiques, la mitochondrie ne *produit* pas d'énergie : elle *transforme* une certaine forme d'énergie en une autre forme d'énergie. Les dimensions des mitochondries sont de l'ordre de 1 à 2  $\mu\text{m}$  de longueur pour 0.5 à 1  $\mu\text{m}$  de diamètres, limitée par deux membranes séparées par un espace inter-membranaire, la membrane interne formant des replis, ou *crêtes mitochondriales*, vers l'intérieur de l'organite (matrice). La microcinématographie de cellules en culture *in vivo* a prouvé que le mitochondriome (ensemble des mitochondries) est un *ensemble dynamique* pouvant être déformées, être déplacées, se diviser ou fusionner.

En moyenne, le volume du chondriome représente 15 à 20 % du volume cellulaire, mais en fait une cellule peut réguler le nombre de ces mitochondries en fonction de son activité métabolique [96].

### Une membrane externe très perméable :

La membrane externe de la mitochondrie est proche par sa composition chimique de la membrane plasmique. Sa principale caractéristique est la présence abondante d'une protéine transmembranaire.

**La membrane interne imperméable et riche en protéines :** La membrane interne de la mitochondrie forme des crêtes, qui sont souvent en forme de lamelles. Avant tout caractérisée par sa richesse en protéines (80 % en poids). On y trouve :

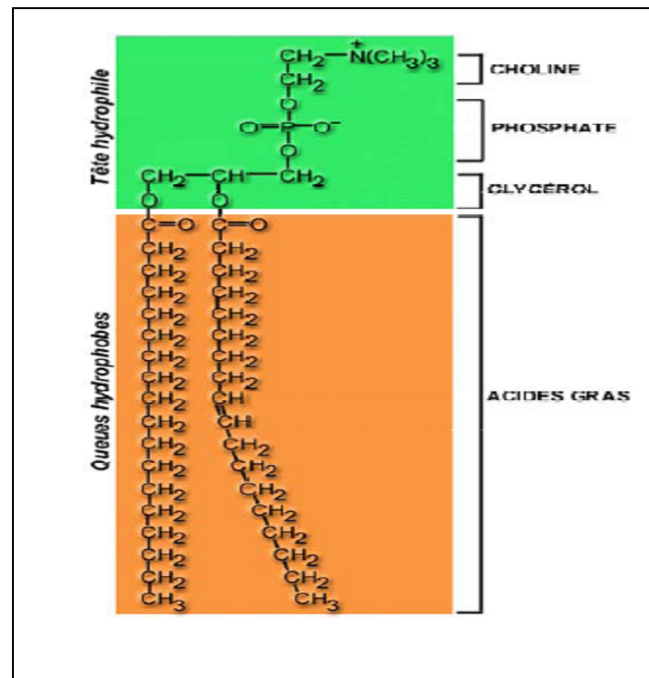
- Les transporteurs d'électrons et de protons formant la chaîne respiratoire, qui permettent la phosphorylation oxydative.

La plus célèbre et la plus évidente de ses molécules est l'ATPase de la membrane interne.

### III- Les membranes :

L'analyse chimique des membranes plasmiques isolées à partir d'hématies montre qu'elles sont constituées de lipides et de protéines. Cette composition a été confirmée pour l'ensemble des membranes biologiques. La masse (en poids sec) de ses deux constituants se répartit en 60 % de protéines et 40 % de lipides. Le rapport des masses protéines/lipides est en général voisin de 1 pour les membranes qui sont plus riches en protéines (jusqu'à 80 % en masse pour la membrane interne des mitochondries).

Les lipides membranaires sont essentiellement des phospholipides ils représentent environ 55 % des lipides membranaires, L'exemple le plus classique de phospholipide est la *phosphatidylcholine* (figure 1)



**Figure-28:** Modèle moléculaire de la phosphatidylcholine

Cette molécule est un ester bâti sur le glycérol (trialcool). Elle comporte deux chaînes aliphatiques issues des acides gras engagés dans l'estérification de deux groupements alcool du glycérol. Le troisième groupement alcool est estérifié par l'acide phosphorique (sous forme d'un groupement phosphate, ionisé, donc chargé, au pH cellulaire). L'ensemble forme un groupement phosphatidyl.

Le groupement phosphatidyl est ici lié à la choline, groupement chargé positivement.

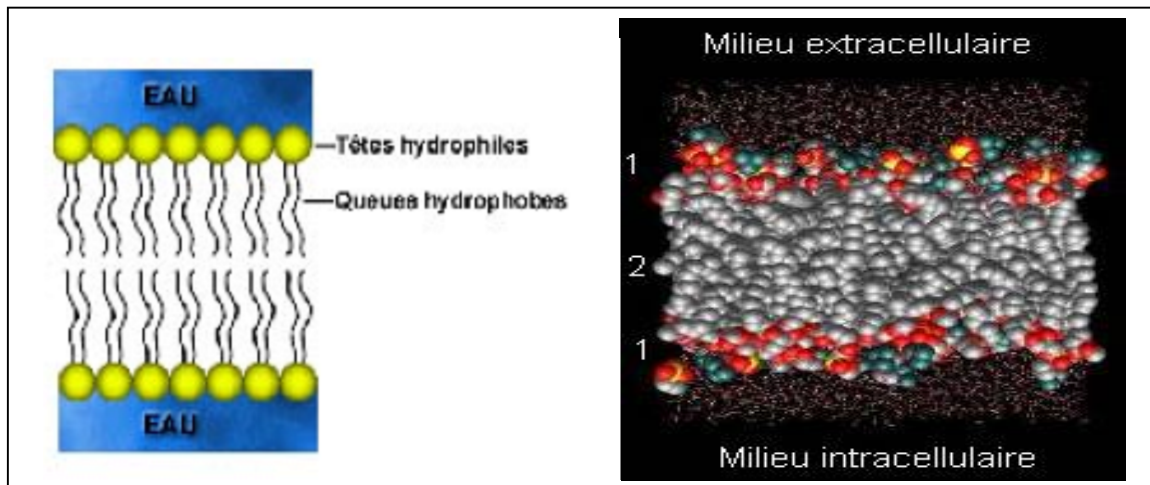
Comme le groupement phosphate est chargé négativement au pH cellulaire, l'ensemble de la molécule est neutre.

Cette molécule possède la particularité de présenter deux zones distinctes dans l'espace :

- Les chaînes aliphatiques sont, comme tous les groupements constitués de carbone et d'hydrogène, fortement hydrophobe ;
- La tête de la phosphatidylcholine, avec ses groupements chargés ou polaires, est fortement hydrophile.

La molécule est donc à la fois hydrophile et hydrophobe : on dit qu'elle est *amphiphile*

Tous les lipides membranaires sont amphiphiles. Or le milieu cellulaire, que ce soit hors ou dans la cellule, est un milieu aqueux; Les lipides membranaires interagissent donc, dans le milieu aqueux de la cellule, de façon à rassembler leurs pôles hydrophiles et à présenter au milieu leurs pôles hydrophiles : ils forment des bicouches lipidiques, maintenues par « interactions hydrophobes » (figure-3). D'autres interactions faibles stabilisent l'ensemble : liaisons ioniques entre les zones polaires des lipides, interactions de van der Waals entre les queues hydrophobes.



**Figure-29:** Structure de la membrane (Têtes hydrophiles, Queues hydrophobes).

Les protéines membranaires sont extrêmement variées, ce sont des éléments essentiels de la diversité et de la spécificité des membranes.

L'hydrophobie des molécules conditionne en partie leur transport membranaire. Le passage d'une molécule hydrophobe à travers une membrane biologique n'est guère problématique : elle est liposoluble, et franchit sans encombre la bicouche lipidique. C'est le cas des gaz respiratoires, par exemple.

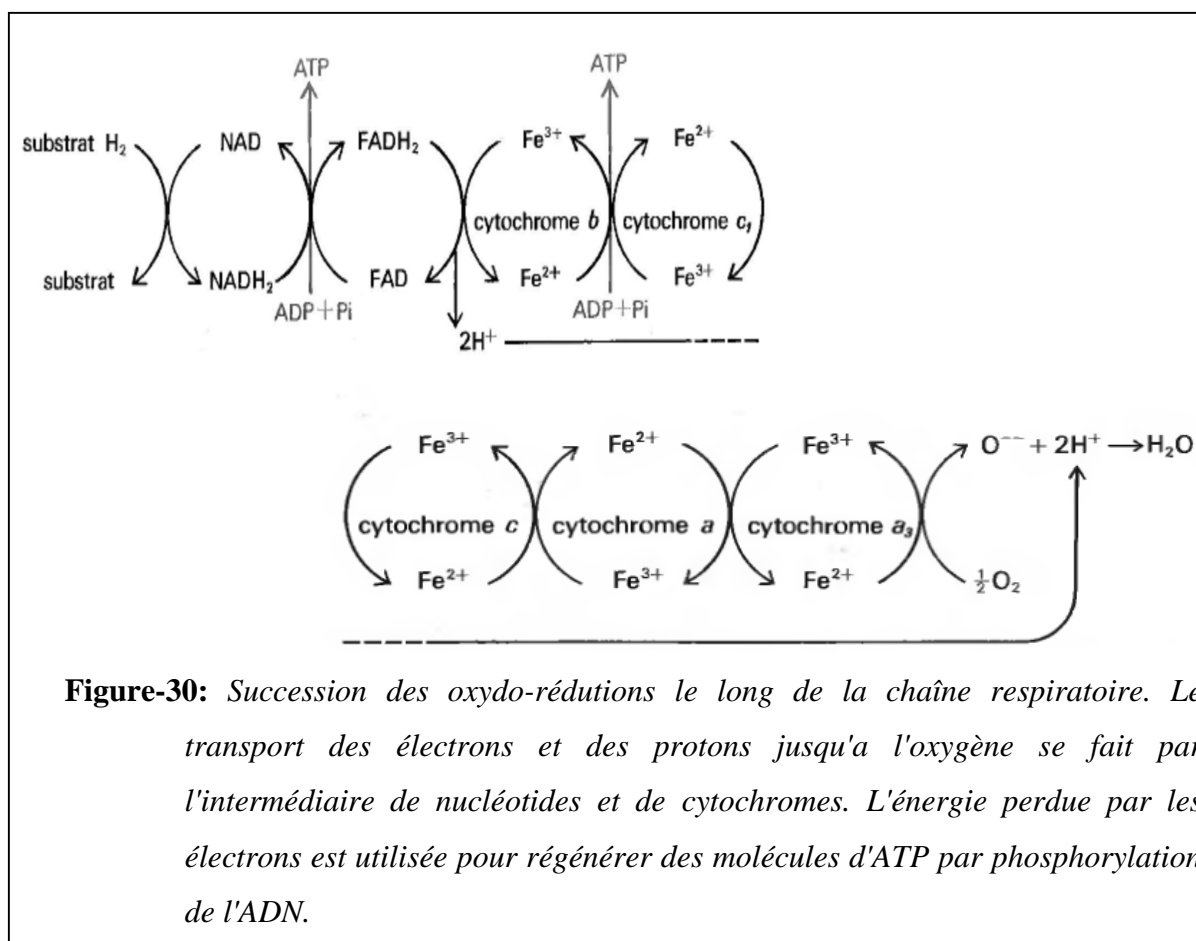
#### IV- La chaîne respiratoire :

Au niveau des mitochondries certains métabolites sont oxydés, et l'énergie dégagée par l'oxydation sert en partie à régénérer de l'ATP par phosphorylation de l'ATP. L'ensemble des réactions d'oxydation et de phosphorylation constitue la phosphorylation oxydative.

En fait, au niveau des mitochondries les oxydations se font par déshydrogénation.

L'hydrogène enlevé au combustible, ou substrat, est capté par des accepteurs successifs dont le dernier est l'oxygène. On remarque lors de l'étude de diverses voies métaboliques du glucose, des processus de déshydrogénation dans lesquels l'accepteur final d'hydrogène est un composé organique, ces processus constituent la fermentation. Dans les mitochondries, où l'accepteur final est l'oxygène, on parle de *respiration*.

Les électrons perdus par le substrat, lors de la déshydrogénation, sont transportés jusqu'à l'oxygène par une série de corps appelés transporteurs d'électrons, puis l'oxygène activé par ces électrons devient accepteur de protons et il se forme de l'eau. Les premiers accepteurs qui captent l'hydrogène sont des nucléotides: NAD, NADP et FAD. Les transporteurs d'électrons sont des cytochromes, protéines à groupement porphyrinique voisin de l'hème de l'hémoglobine; les cytochromes renfermant un atome de fer, leur passage de l'état réduit à l'état oxydé est lié au changement de valence du métal. Divers types de cytochromes ont été isolés à partir des mitochondries, on les désigne par les lettres a, b, c. Les cytochromes mitochondriaux a et b sont insolubles dans l'eau et ils ne sont actifs qu'associés à des phospholipides, le cytochrome c est par contre très soluble dans l'eau. La succession des oxydo-réductions permet ainsi le transport de l'hydrogène enlevé au substrat jusqu'à l'oxygène (figure 30).



**Figure-30:** Succession des oxydo-réductions le long de la chaîne respiratoire. Le transport des électrons et des protons jusqu'à l'oxygène se fait par l'intermédiaire de nucléotides et de cytochromes. L'énergie perdue par les électrons est utilisée pour régénérer des molécules d'ATP par phosphorylation de l'ADN.

Le transport des électrons et des protons met ainsi en jeu des protéines et des nucléotides. Certains de ces nucléotides, comme la FAD, sont en fait des coenzymes de déshydrogénases à coenzymes flaviniques appelées flavoprotéines.

Elles possèdent une couleur jaune; ce sont elles qui sont responsables de la teinte jaune caractéristique des culots de mitochondries isolées. D'autre part les recherches récentes laissent à penser qu'il entre également dans cet ensemble de transporteurs une quinone, l'ubiquinone ou coenzyme Q, qui peut être réduite ou oxydée, et se place sans doute entre la FAD et le cytochrome b. L'ensemble des corps qui participent au transport de l'hydrogène jusqu'à l'oxygène constitue la chaîne respiratoire, mais ses divers composants ne sont pas encore connus. Certains constituants sont hydrosolubles, d'autres sont insolubles dans l'eau et fortement liés à des lipides. L'analyse de la chaîne est malaisée car il est difficile d'en séparer les divers composants.

Au cours des recherches, on a d'abord isolé des fragments de la chaîne et l'on a pu montrer ensuite qu'il correspondaient à plusieurs transporteurs. Ainsi, on a préparé une partie soluble appelée  $NADH_2$ -Cytochrom c réductase qui catalyse la réduction du cytochrome c par la  $NADH_2$ ; ce fragment contient en fait une  $NADH_2$ -déshydrogénase, qui a une FAD

comme coenzyme, le coenzyme Q, les cytochromes *b* et *c<sub>1</sub>*. De même on a isolé une fraction insoluble, la cytochrome- oxydase qui catalyse la réoxydation du cytochrome *c* réduit: le cytochrome *a* et le cytochrome *a<sub>3</sub>*, elle contient de plus des atomes de cuivre dont on ne sait ni à quels corps ils s'associent, ni le rôle qu'ils peuvent jouer.

Au cours du transport de l'hydrogène jusqu'à l'oxygène, l'énergie perdue par les électrons permet la régénération de molécules d'ATP, par phosphorylation de molécules d'ADP. Ces phosphorylations se font a trois niveaux de la chaîne respiratoire: lors de l'oxydation par la FAD de la NADH<sub>2</sub> (ou de la NADPH<sub>2</sub>, la NADP était également un accepteur d'hydrogène comme la NAD); lors de l'oxydation du cytochrome *b* par le cytochrome *c<sub>1</sub>*; enfin lors de l'oxydation de cytochrome *b* par cytochrome *c<sub>3</sub>* lors de l'oxydation du cytochrome *a<sub>3</sub>*. Lors de l'hydrogène est enlevé de la NADH<sub>2</sub> ou de la NADPH<sub>2</sub>, trois molécules d'ATP se régènèrent lorsqu'il est enlevé a la FADH<sub>2</sub> deux molécule seulement se régènèrent (figure 4). Les mécanismes de couplage de la phosphorylation de l'ATP avec les trois réactions d'oxydation, ne sont pas encore élucidés; ils mettent en jeu des enzymes encore inconnues [97].

#### **IV-1 Inhibiteurs et découplants de la Chaîne Respiratoire Mitochondriale (CRM):**

##### **IV-1-1 Inhibiteurs:**

Ce sont des corps chimiques qui par leurs liaisons avec l'enzyme modifient la vitesse de la réaction enzymatique, soit en l'accélérant (Activateur), ou bien en la ralentissant (Inhibiteur). Ils y a d'autres molécules qui peuvent intervenir pour ralentir les réactions de la chaîne respiratoire mitochondriale, ce sont les inhibiteurs de la chaîne respiratoire mitochondriale. Compte tenu de l'importance vitale de cette voie métabolique pour notre organisme, la plupart de ces inhibiteurs sont des poisons dangereux [98].

##### **IV-1-2 Découplants :**

Corps chimique qui empêche le transfert de l'énergie entre les enzymes d'oxydo-reduction et l'ATP, produit d'une enzyme de phosphorylation dans une voie métabolique de phosphorylation dans une voie métabolique de respiration ou de fermentation.

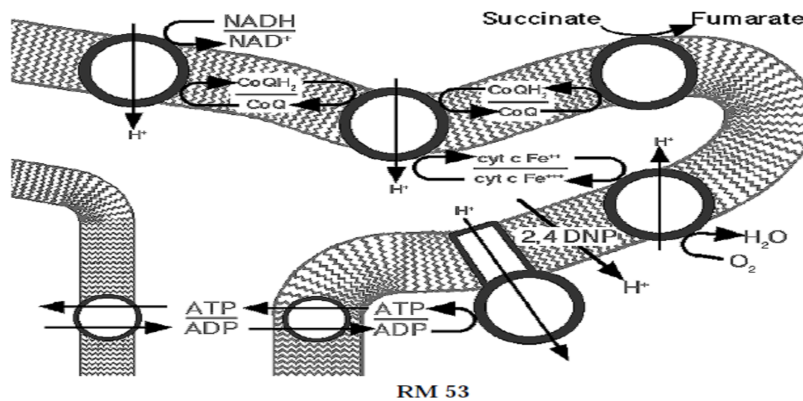
- La notion de couplage entre la réaction de phosphorylation de l'ATP et les oxydoréductions de la chaîne respiratoire mitochondriale, permet de définir un nouveau type de régulateurs de cette voie métabolique, les découplants.

- Ces composés suppriment la transmission de l'énergie entre les oxydoréductases et l'ATPase.

Ils diminuent le gradient de protons transmembranaire ; ils favorisent l'hydrolyse de l'ATP par le complexe  $F_0-F_1$  ; ils stoppent l'effet indirect de l'oligomycine sur le reste de la chaîne respiratoire ; ils activent enfin les oxydations cellulaires, l'utilisation du NADH et de l'Oxygène [98].

#### IV-1-2 Exemple de découplant de la CRM : le 2,4-DiNitroPhénol

Examinons à titre d'exemple l'effet du 2,4 dinitrophénol (2,4 DNP) sur la chaîne respiratoire mitochondriale, ce découplant est dissous dans les lipides de la membrane interne de la mitochondrie qu'il rend perméable aux protons. Cette fuite de protons à travers la membrane, entraîne la diminution du gradient de protons transmembranaire. Par suite les oxydoréductases de la chaîne respiratoire accélèrent leur pompage de protons et donc les oxydations couplées. C'est pourquoi les découplants sont des activateurs de la respiration et des oxydations cellulaires.



Le pompage des protons lorsque le gradient est abaissé consomme moins d'énergie ; il y a donc davantage d'énergie libérée dans le milieu sous forme de chaleur.

Quant à l'ATPase dont la réaction est réversible, ne disposant plus du gradient suffisant pour fournir l'énergie dont elle a besoin pour la phosphorylation de l'ADP, elle catalysera la réaction dans l'autre sens, hydrolysant l'ATP en ADP et en phosphate en utilisant le reste de l'énergie pour pomper des protons de la matrice vers l'espace intermembranaire et libérer de la chaleur dans le milieu. Lorsque le gradient devient plus faible encore, une des sous-unités du complexe  $F_0-F_1$  inhibe cette activité d'hydrolyse de l'ATP restant.

- 1- G. Aurélie, Une nouvelle méthode d'apprentissage de données structurées: applications à l'aide à la découverte de médicaments. Université Pierre et Marie Curie Paris 6 (2008).
- 2- S.E. Manahan, Toxicological chemistry and biochemistry, 3rd ed. CRC Press (2003).
- 3- J. E. Morrison, D. Matthews, R. Washington, P.V. Fennessey, M. Harrison, Phenol motor blocks in children: plasma concentrations and cardiac dysrhythmias. *Anesthesiology*, 75, 359-362 (1991).
- 4- M.A. Warner, J.V. Harper, Cardiac dysrhythmias associated with chemical peeling with phenol. *Anesthesiology*, 62, 366-367(1985).
- 5- R.E. Gosselin, R.P.Smith, H.C. Hodge, Phenol in *Clinical Toxicology of Commercial Products*, 5th ed., Williams & Wilkins, Baltimore, III-344–III-348 (1984).
- 6- N.Staimer, S.J. Gee, B.D. Hammock, Development of a class-selective enzyme immunoassay for urinary phenolic glucuronides, *Anal. Chim. Acta*, 441, 27–36 (2001).
- 7- ECOTOX User Guide: ECOTOXicology Database System Version 4.0 [en ligne]. U.S. Environmental Protection Agency (EPA), 2007. Disponible sur : <http://cfpub.epa.gov/ecotox/>.
- 8- T. Öberg, A QSAR for baseline toxicity: validation, domain of application, and prediction. *Chemical Research in Toxicology*, 17 (12), 1630-1637 (2004).
- 9- R.G. Clements, *et al.* Estimating toxicity of industrial chemicals to aquatic organisms using structure activity relationships, Office of Pollution Preventions and Toxics, US Environmental Protection Agency, Washington, DC (1988).
- 10- M. Pintore, *et al*, Predicting toxicity against the fathead minnow by adaptative fuzzy partition. *QSAR & Combinatorial Science*, 22, 210-219 (2003).
- 11- K.L.E. Kaiser, et S.P. Niculescu, Using probabilistic neural networks to model the toxicity of chemicals to the fathead minnow (*Pimephales promelas*) : a study based on 865 compounds. *Chemosphere*, 38 (14), 3231-3245 (1999).
- 12- A. Crum-Brown, T. Frazer, On the connection between chemical constitution and physiological action. *Transactions of the Royal Society of Edinburgh*, 25,151-203 (1868-69).
- 13- M. C. Richet, *Compt. Rend. Soc. Biol. (Paris)* 45, 775-776 (1893)
- 14- C. Hansch, T. Fujita, , *J. Am. Chem. Soc.* 86, 1616-1626 (1964)
- 15- Jr. Free, S. M, J. W. Wilson,. *Med. Chem.* 7, 395-399 (1964)
- 16- C. Hansch, *Ace. Chem. Res.* 2, 232-239 (1969)
- 17- C. Hansch , J. M. Clayton, *J. Pharm. Sci.* 62, 1-21 (1973)
- 18- T.Fujita, T.Ban, , *J. Med. Chem.*44, 148-152 (1971)

- 19- M. Karelson. Molecular descriptors in QSAR/QSPR. Wiley- Interscience, 385 (2000).
- 20- B. Kowalski, R. Gerlach, H. Wold. Systems under Indirect Observation (K. Jöreskog , H. Wold, eds.), North Holland, Amsterdam, 191-206 (1982).
- 21- L. Eriksson, E. Vohannson, N. Kettaneh- Wold. Multi and Megavariate Data Analysis- Principles and Applications. Umetricsacademy, Umeå (2001).
- 22- S. Wold, A. Ruhe, H.Wold, W. Dunn. SIAMJ. Sci. Stat. Comput., 5, 735 (1984).
- 23- S. Wold. Chemometrics: Mathematics and Statistics in Chemistry. Reidel, Dordrecht, The Netherlands (1984).
- 24- P. Gelada, B. R. Kowalski, Anal. Chim. Acta, 185, 1 (1986).
- 25- A. Höskuldsson, J. Chemometrics, 2, 211 (1988).
- 26- J. A. Burns, G. M. Whiteside. Chem. Rev., 93, 2583 (1993).
- 27- L. S. Anker, P. C. Jurs. Anal. Chem., 64, 1157 (1992).
- 28- T. Aoyama, Y. Suzuki, H. Ichikawa. J. Med. Chem., 33, 2583 (1990).
- 29- T. A. Andrea, J. Med. Chem., 34, 2824 (1991).
- 30- P. C. Jurs, Computer Software Applications in Chemistry. Second Edition, J. Wiely (1996).
- 31- A. R. Katritzky, V. S. Lobanov, M. Karelson. CODESSA Reference Manual. University of Florida, Gainesville (1994).
- 32- V. Y. Nalimov. The Application of Mathematical Statistics to Chemical Analysis, Addison- Wesley, Reading, MA (1962).
- 33- R. Calcutt, R. Body. Statistics for Analytical Chemists. Champman & Hall, New York (1983).
- 34- J. C. Miller, J. N. Miller. Statistics for Analytical Chemistry. Ellis Horwood, New York (1988).
- 35- P. C. Meier, R. E. Zund. Statistical Methods in Analytical Chemistry. Wiley, New York (1993).
- 36- P. Dagnélie. Statistique Théorique Et Appliquée. Tomes 1 et 2. De Boeck & Larcier s. a. (1998).
- 37- R. Tomassone, E. Lesquoy, C. Miller. La régression : nouveaux regards sur une ancienne méthode statistique. Masson, INRA (1983).
- 38- R. Wehrens, H. Putter, L. M. C. Buydens. Chemom. Int. Lab. Syst., 54, 35- 52 (2000).
- 39- L. Eriksson, J. Jaworska, A. P. Worth, M. T. D. Cronin, R. M. Mc Dowell, P. Gramatica. Methods for Reliability and Uncertainty Assessment and for Applicability

- Evaluations of Classification and Regression Based QSARs. *Environmental Health Perspectives* 111, 1361-1375 (2003).
- 40- N. R. Draper, H. Smith. *Applied Regression Analysis*, 3<sup>rd</sup> ed; Wiley Series in Probability and Statistics (1998).
- 41- R.D. Snee, Validation of Regression Models: Methods and Examples *Technometrics*, 19, 415-428 (1977).
- 42- Matlab Version 7.0.0.19920 (Release 14) The Language of Technical Computing The MathWorks, Inc (2004).
- 43- Hyperchem<sup>TM</sup> Release 6.03 for windows, Molecular Modeling System (2000).
- 44- R. Todeschini, V. Consonni, M. Pavan. DRAGON, Software for the Calculation of Molecular Descriptors. Release 5.3 for windows, Milano (2005).
- 45- CambridgeSoft, ChemBioDraw/CLOGP trail Version 12.0, CambridgeSoft Corporation 100 Cambridge Park Drive, USA.
- 46- SRC KOWWIN software, SRC-LOGKOW Version 1.67, Syracuse Research Corporation, Syracuse, USA (Sept 2008). <http://www.syrres.com/eSc/kowwin.htm>.
- 47- G. Persoone, D. Dive, *Ecotox. Environ. Saf*, 2, 105-144 (1978).
- 48- T.W. Schultz, M. Cajina-Quezada, J.N. Dumont, *Arch. Environ. Contam. Toxicol*, 9, 591-598 (1980).
- 49- T.W. Schultz, M. Cajina-Quezada, 1982. *Arch. Environ. Contam. Toxicol*. 11, 353-361.
- 50- P.J. Taylor, Hydrophobic Properties of Drugs. In *Quantitative Drug Design. Vol. 4* (Ramsden, C.A., ed.), Pergamon Press, Oxford (UK), 241-294 (1990).
- 51- F. A. Carrupt, B. Testa, P. Gaillard, Computational Approaches to Lipophilicity: Methods and Applications. In *Reviews in Computational Chemistry, Vol. 11* Wiley-VCH, New York (NY), 241-315 (1997).
- 52- <http://www.Organic-Chemistry.org/prog/peo/>
- 53- ACD/LogP, v. 10, Advanced Chemistry Development, Inc., Toronto, ON, Canada, [www.acdlabs.com](http://www.acdlabs.com), (2007).
- 54- V.K. Gobar, Reliable assessment of logP of compounds of pharmaceutical relevance. *SAR and QSAR in Environmental Research*, 10, 371-380 (1999).
- 55- J.Devillers, *et al.* Simulating lipophilicity of organic molecules with a backpropagation neural network. *Journal of Pharmaceutical Sciences*, 87 (9), 1086-1090 (1998).
- 56- T. Fujita, J. Iwasa, C. Hansch, *J. Am. Chem. SOC*, 86, 5175-5180 (1964).
- 57- C. G. Nys, R. F. Rekker, *Chim. Ther*, 8, 521-535 (1973).
- 58- R. F. Rekkerh, M. Dekort, *Eur.J. Med. Chem.* 14, 479-488 (1979).
- 59- G. G. Nus, R. F. REKKERC, *him. Ther*, 9, 361-375 (1974).

- 60- R. F. Rekker, The hydrophobic fragmental constant. Its derivation and application. A means of characterizing membrane systems. Pharmacochem. Library Vol.1, Elsevier, Amsterdam, (1977).
- 61- P. Broto, G. Moreau, C. Vanduycke, Eur. J. Med. Chem, 19, 71-78 (1984).
- 62- A. J. Leo, P. Y. C. Jow, C. Silipo, C. Hansch, J. Med. Chem, 18, 865-868 (1975).
- 63- A. Hansch, J. LEO, Substituent Constants for Correlation Analysis in Chemistry and Biology. Wiley, New York, (1979).
- 64- A. J. Leo, J. Pharm. Sci, 76, 166- 168 (1987).
- 65- A. J. Leo, Methods Enzymol., 202(1991),
- 66- A. J. Leo, Chem. Rev, 93, 1281- 1306 (1993).
- 67- J. T. Chou, P. C. Jurs, J. Chem, Inf Comput. Sci, 19, 172-178 (1979).
- 68- W. M. Meylanp, H. Howard, J. Pharm. Sci., 84, 83-92 (1995).
- 69- A. K. Ghose, M. CRIPPEN, J. Comp. Chem, 7, 565-577 (1986).
- 70- A. K. Ghose, M. Crippen, J. Chem. Inf. Comp. Sci, 27, 21-35 (1987).
- 71- A. K. Ghose, P. Ritchettg, M. Crippen, J. Comp. Chem, 9, 80- 90 (1988).
- 72- V. N. Viswanadhaan, A.K. Ghose, Inf. Comput. Sci, 29, 163-172 (1989).
- 73- Moriguchi, S. Hirono, Q. Liu, I. Nakagome, Y. Matsushita, Chem. Pharm. Bull, 40, 127-130 (1992).
- 74- T.W. Schultz, G.W. Holcombe, G.L. Phillips Relationships of quantitative- structure activity to comparative toxicity of selected phenols in the *Phimephales promelas* and *Tetrahymena pyriformis* test systems. Ecotoxicol Environ, Saf, 12, 146-153 (1986).
- 75- T.W. Schultz, D.T. Lin, S.K. Wesley, QSAR for monosubstituted phenols and the polar narcosis mechanism of toxicity, Qual Assur Good Pract Reg Law, 1, 132-143 (1992).
- 76- G.D. Veith, S.J. Broderius, Structure-toxicity relationships for industrial chemicals causing type (II) narcosis syndrome. In: Kaiser KLE (ed) QSAR in environmental toxicology-II. D Reidel Publishing Co, Dordrecht, The Netherlands, 385 (1987).
- 77- T.W. Schultz, M. Cajina-Quezada, Structure-activity relationships for mono alkylated or halogenated phenols, Toxicol Lett, 37, 121-130 (1987).
- 78- J. Saarikoski, M. Viluksela, Relations between physico-chemical properties of phenols and their toxicity and accumulation in fish, Ecotoxicol Environ Saf, 6, 501-512 (1982).
- 79- T.W. Schultz Nonpolar narcosis, A review of the mechanism of action for baseline toxicity. In: Cowgill UM, Williams LR (eds) Aquatic toxicology and environmental fate. Twelfth symposium, ASTM STP-1027, Amer Soc Test Mater, Philadelphia, PA, 104-109 (1989).

- 80- T.W. Schultz, The use of the ionization constant (pKa) in selecting models of toxicity of phenols, *Ecotoxicol Environ Saf*, 14, 178-183(1987).
- 81- T.W. Schultz, Relative toxicity of para-substituted phenols: LogKow and pK, dependent structure-activity relationships, *Bull Environ Contam Toxicol*, 38, 994-999 (1987).
- 82- M. T. Cronin, T.W. Schultz, Development of Quantitative Structure- Activity Relationships for the Toxicity of Aromatic Compounds to *Tetrahymena pyriformis*: Comparative Assessment of the Methodologies, *Chem. Res. Toxicol*, 14, 1284-1295 (2001).
- 83- G.D. Veith, S.J. Broderius, Rules for distinguishing toxicants that cause type I and type II narcosis syndromes. *Environ Health Perspect*, 87, 207-211(1990), S.H. Yalkowsky, S.C. Valvani Solubility and partitioning I: Solubility of nonelectrolytes in water. *J Pharm Sci*, 69, 912-922 (1980).
- 84- M.J. Kamlet, R.M. Doherty, G.D. Veith, R.W. Taft, M.H. Abraham, Solubility properties in polymers and biological media. 7. An analysis of toxicant properties that influence inhibition of bioluminescence in *Photobacterium phosphoreum* (the microtox test). *Environ Sci Technol*, 20, 690-695 (1986).
- 85- H. Terada, The interaction of highly active uncouplers with mitochondria, *Biochim Biophys Acta*, 639, 225-242 (1981).
- 86- M. Cajina-Quezada, T.W. Schultz, Structure-toxicity relationships for selected weak acid respiratory uncouplers, *Aquat Toxicol*, 17, 239-252 (1990).
- 87- J.S. Jaworska, T.W. Schultz, Mechanism-based comparison of acute toxicity elicited by industrial organic chemicals in procaryotic and eucaryotic systems, *Ecotoxicol Environ Saf*, 29, 200-2 (1994).
- 88- H. Terada, Uncouplers of oxidative phosphorylation. *Environ Health Perspect*, 87, 213-218 (1990).
- 89- S.G.A. McLanghlin, J.P. Dilger, Transport of protons across membranes by weak acids, *Physiol Rev*, 60, 825-862 13 (1980).
- 90- S. Balaza, E. Sturdik, E. Durkova, M. Antalík, P. Sulo, Quantitative structure-activity relationship of carbonylcyanide phenylhydrazones as uncouplers of mitochondrial oxidative phosphorylation, *Biochim Biophys Acta*, 851, 93-98 (1986).
- 91- E. Sturdik, S. Balaza, E. Durcova, M. Sturdicova, P. Sulo, M. Antahík, V. Mikes, V. Dadák, Model-based relation between physicochemical properties uptake and uncoupling effect of carbonylcyanide phenylhydrazones on oxidative phosphorylation at cellular level, *Quant Struct Act Relat*, 7, 221-225 (1988).

- 92- T.W Schultz, D.T Lin, T.S .Wilke, L.M. Arnold Quantitative structure-activity relationships for *Tetrahymena pyriformis* population growth endpoint: A mechanism of action approach. In: W. Karcher, J. Devillers (eds) Practical applications of quantitative structure-activity relationships (QSAR) in environmental chemistry and toxicology. Kluwer Academic Publishers, Dordrecht, The Netherlands, 241-262 (1990)
- 93- K. Pirselova, S. Balaz, T. W. Schultz, Model-Based QSAR for Ionizable Compounds: Toxicity of Phenols Against *Tetrahymena pyriformis*, *Toxicol, Environ. Contam. Toxicol*, 30, 170-177 (1996).
- 94- R. Todeschini, D. Ballabio, V. Consonni, A. Mauri, M. Pavan. Moby Digs Software for Multilinear Regression Analysis and Variable Subset Selection by Genetic Algorithm. Release 1.1 for Windows, Milano (2009).
- 95- S. Spycher, E. Pellegrini, J. Gasteiger, Use of Structure Descriptors To Discriminate between Modes of Toxic Action of Phenols, *J. Chem. Inf. Model*, 45, 200-208 (2005).
- 96- Y. Bassaglia, Biologie cellulaire, J. F. Divernois, Moloine (2001).
- 97- P. Favard, Biologie et Physiologie Cellulaire, Hermann, Collection Méthodes Paris, 109-114 (1966).
- 98- A. Raisonier, Respiration Mitochondriale, Université de Paris, 61-101, (2005).

# REERENCES BIBLIOGRAPHIQUES