



**Faculté des Sciences  
Département de Chimie**

# MÉMOIRE

Présenté pour l'obtention du diplôme MAGISTER

*Par M<sup>elle</sup> ZIANI NADIA*

*DES en Chimie*

**Option : Chimie et environnement**

## THÈME

***ETUDE QSAR, à l'aide de descripteurs moléculaires théoriques 3D, de la toxicité d'alcools et d'amines aliphatique. comparaison avec l'approche log P***

***Devant le jury :***

<b>PRESIDENTE :</b>	<b>M<sup>me</sup>. FERTIKH Nadia</b>	<b>MC</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>. DJALLAL Ahmed</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>
<b>INVITEE :</b>	<b>M<sup>lle</sup>. TOUHAMI Imen</b>	<b>MAA</b>	<b>EPST-Annaba</b>

**Année 2010**

## *Remerciements*

*La présente étude a été réalisée au laboratoire de Sécurité Environnementale et Alimentaire de l'Université d'ANNABA sous la direction de monsieur le **Pr. Djelloul MESSADI**. Aussi, je me permets de lui exprimer ma profonde reconnaissance, pour le bienveillant intérêt qu'il a accordé à cette étude.*

**Je désire vivement exprimer mon extrême gratitude et mes remerciements à M<sup>r</sup>. KHATMI Djameledine et tous les membres du jury.**

1. **M<sup>me</sup> :FERTIKH Nadia**
2. **M<sup>r</sup>. DJALLAL Ahmed**
3. **M<sup>lle</sup>. TOUHAMI Imen**

*Enfin, je ne saurais oublier mes collègues de laboratoire et également tous ceux qui par leur présence ou par leur aide m'ont permis de mener à bien ce travail.*

## *Dédicace*

Je dédie ce modeste travail à :

- ♥ Mon père
- ♥ Ma mère
- ♥ Mes frères et mes sœurs :(Ridha, Mohamed, Abdel Waheb, Abdel malek, Souraya et Ahlem).
- ♥ La mémoire de ma grand mère et mon grand père.
- ♥ Ma famille.
- ♥ Mes amis.
- ♥ Enfin, toute l'équipe du labo 34.

*ZIANI NADIA*

إن إدخال المقاييس الجزيئية النظرية ثلاثية 1988  
تأثير الستيروئيدوميرية أو الإنوتيوميرية  
(CIC50; CL50) والأمنيات الخطية يكون مرتبط بالواصف  
الهدن (ADDD) : / الذي يحسب بواسطة  
الوحيد لبنية المركبات الكيميائية بعد تحسين هندستهم.

التشخيص الإحصائي ( $R^2, R^2_{adj}, Q^2, Q^2_{ext}, EQMC, EQMP, EQMP_{ext}, F, S$ )  
عليها (Log P; ADDD) سيتم فيما بينها.

مفاتيح الكلمات :  
- الأمنيات الخطية- المنهجية QSAR - رباعي هيمينابيروفورم

## ABSTRACT

The introduction from 1988 of molecular parameters theoretical three-dimensional (3D) takes into account the influence of different conformers, enantiomers or stereoisomers.

The toxicity (**CIC50; LC50**) alcohols and aliphatic amines is linked to a geometric descriptor (ADDD: middle distance / distance degree) calculated using only the structure of chemical compounds after optimizing their geometry.

The diagnostic statistics ( $R^2$ ,  $R^2_{adj}$ ,  $Q^2$ ,  $Q^2_{ext}$ , EQMC, EQMP, EQMP<sub>ext</sub>, F, S) models obtained by both approaches (Log P; ADDD) will be compared.

**Key words:** aliphatic Alcohols- aliphatic amines – Toxicity, QSAR Methodology- Tetrahymena pyriformes.

## RESUME

L'introduction, depuis 1988, des paramètres moléculaires théoriques tridimensionnels (3D) permet de tenir compte de l'influence des différents conformères, stéréo-isomères ou énantiomères.

La toxicité (**CIC50** ; **CL50**) d'alcools et d'amines aliphatiques sera reliée à un descripteur géométrique (ADDD : Ordre distance /distance moyenne.) ,calculé en utilisant uniquement la structure des composés chimiques après optimisation de leur géométrie .

Les diagnostics statistiques ( $R^2$ ,  $R^2_{adj}$ ,  $Q^2$ ,  $Q^2_{ext}$ , EQMC, EQMP,  $EQMP_{ext}$ , F, S) des modèles obtenus par les deux approches (Log P; ADDD) seront comparés .

### Mots clés :

Les alcools, les amines aliphatiques – Toxicité- Méthodologie QSAR- Tetrahymena pyriformes.

	<b>Titre</b>	<b>Page</b>
Tableau I	Toxicité ( <b>CIC50</b> ) pour les alcools aliphatiques et les amines.	12
Tableau II	Toxicité ( <b>CL50</b> ) pour les amines.	13
Tableau III	Descripteurs moléculaires intervenant dans la modélisation de <b>pCIC50</b> .	18
Tableau IV	Valeurs des <b>pCIC50</b> observées ( <b>pCIC50<sub>obs</sub></b> ) et calculées ( <b>pCIC 50<sub>cal</sub></b> ), des résidus ordinaires ( $e_i$ ) et standardisés ( $e_{i\text{ std}}$ ) ainsi que des leviers ( $h_{ij}$ ).	22
Tableau V	Quelques caractéristiques des éléments de l'ensemble de validation externe .	23
Tableau VI	Valeurs des <b>pCIC 50</b> observées ( <b>pCIC 50<sub>obs</sub></b> ) et calculées ( <b>pCIC 50<sub>cal</sub></b> ), des résidus ordinaires ( $e_i$ ) et standardisés ( $e_{i\text{ std}}$ ) ainsi que des leviers ( $h_{ij}$ ).	27
Tableau VII	Quelques caractéristiques des éléments de l'ensemble de validation externe.	29
Tableau VIII	Valeurs des <b>pCL50</b> observées ( <b>pCL50<sub>obs</sub></b> ) et calculées ( <b>pCL50<sub>cal</sub></b> ), des résidus ordinaires ( $e_i$ ) et des prédictions standardisés ( $e_{i\text{ std}}$ ) ainsi que des leviers ( $h_{ij}$ ).	33
Tableau IX	Valeurs des <b>pCL50</b> observées ( <b>pCL50<sub>obs</sub></b> ) et calculées ( <b>pCL50<sub>cal</sub></b> ), des résidus ordinaires ( $e_i$ ) et de prédiction standardisés ( $e_{i\text{ std}}$ ) ainsi que des leviers ( $h_{ij}$ ).	37

	<b>Titre</b>	<b>Page</b>
Figure 1	Valeurs des coefficients de partages calculés (Log P) en fonction des expérimentales (Log Kow).	20
Figure 2	Diagramme de Williams pour les éléments des ensembles de calibration(20) et de validation (10)	24
Figure3	Droites d'ajustement des deux ensembles.	25
Figure 4	Test de randomisation associé au modèle QSPR.	26
Figure 5	Diagramme de Williams pour le Modèle 2	29
Figure 6	Droites d'ajustement pour le Modèle 2	30
Figure 7	Test de randomisation pour le Modèle 2	31
Figure 8	Diagramme de Williams pour tout les composés	34
Figure 9	Droite d'ajustement des <b>pCL50</b> observés en fonction des <b>pCL50</b> calculés	35
Figure 10	Test de randomisation	36
Figure 11	Diagramme de Williams pour tout les composés	38
Figure 12	Droite d'ajustement des <b>pCL50</b> observés en fonction des <b>pCL50</b> calculés.	39
Figure 13	Test de randomisation	40



## SOMMAIRE

REMERCIEMENTS	
DEDICACE	
RESUMES	I-II -III
LISTE DES TABLEAUX	IV
LISTE DES FIGURES	V
SYMBOLES ET ABREVIATIONS	VI
INTRODUCTION GENERALE	2
<b>PARTIE GENERALITES</b>	
<b>I-LES ALCOOLS</b>	6
I-1 Utilisation des alcools	6
I-2 Classification des alcools	6
I-3 Propriétés des alcools	7
I-4 Toxicité des alcools	7
<b>II-LES AMINES</b>	8
II-1 Utilisation des amines	8
II-2 Classification des amines	8
II-3 Propriétés des amines	9
II-4 Toxicité des amines	9
III-1 – COLLECTE DES DONNEES	10
III-2- CALCUL ET SELECTION DES DESCRIPTEURS MOLECULAIRES.	13
III-3-REGRESSION LINEAIRE SIMPLE	14
III-4- DEVELOPPEMENT ET VALIDATION DU MODELE.	15
<b>PARTIE EXPERIMENTALE</b>	
<b>I- MODELISATION DE LA CONCENTRATION D'INHIBITION DE LA CROISSANCE</b>	18
<b>MODELE 1: <math>pCIC50=f(\log P)</math></b>	21
I-1-1- Equation et analyse de régression	21
I-1-2- Analyse des résidus et diagnostics d'influence	21
I-1-3- Validation externe	23
I-1-4- Diagramme de Williams	24
I-1-5- Qualité de l'ajustement	24
<b>MODELE 2: <math>pCIC50=f(ADDD)</math></b>	26
I-2-1- Equation et analyse de régression	26
I-2-2- Analyse des résidus et diagnostics d'influence	27
I-2-3- Validation externe	28
I-2-3- Diagramme de Williams	29
I-2-5- Qualité de l'ajustement	30
<b>II- MODELISATION DE LA DOSE LA DOSE LETALE 50</b>	32
<b>MODELE 1: <math>pCL50=f(\log P)</math></b>	32
II-1-1- Equation et analyse de régression	32
II-1-2 Analyse des résidus et diagnostics d'influence	33
II-1-3 Le diagramme de Williams	34

II -1- 4	Qualité de l'ajustement	35
II -1- 5	Test de randomisation	36
	<b>MODELE 2: <math>pCL50 = f(ADDD)</math></b>	36
II-2-1	Equation et analyse de régression	36
II-2-2-	Analyse des résidus et diagnostics d'influence	37
II-2-3	Le diagramme de Williams	38
II-2-d	- Qualité de l'ajustement	39
II-2-5	Test de randomisation	40
	<b>CONCLUSION GENERALE</b>	42
	<b>REFERENCES BIBIOGRAPHIQUES</b>	44
	<b>ANNEXES</b>	47

## INTRODUCTION GENERALE

De très nombreux produits chimiques, ne sont pas caractérisés par leurs propriétés physico-chimiques ou leurs activités biologiques ,utiles pour contrôler leur évolution dans l'environnement ou leurs toxicités potentielles. Les raisons en sont :

- Leur grand nombre.
- Les Coûts élevés des procédés expérimentaux de mesure.
- La non disponibilité des appareillages ou des réactifs nécessaires à leur analyse ; leurs complexités et la dangerosité qu'ils peuvent présenter.

L'élaboration de relations quantitatives structure-activité (QSAR) est devenue cruciale dans le développement des substances chimiques organiques, que ce soit dans le domaine de la pharmacie ou dans celui de l'agrochimie ; les logarithmes des coefficients de partage ( $\log P$ ) ont trouvé une grande utilisation dans la conception de médicaments ou de pesticides. Ils entrent ainsi fréquemment dans l'élaboration des QSAR. De plus, le coefficient de partage *n*-octanol/eau joue un rôle important dans les stades précoces de l'évaluation des risques engendrés par un produit chimique sur l'environnement.

Des logiciels informatiques spécialisés permettent le calcul de plus de 3000 descripteurs moléculaires appartenant à différentes classes. Plutôt que de rechercher à expliquer la variable dépendante (propriété physique ou biologique considérée) par tous les régresseurs (descripteurs moléculaires), on peut chercher seulement un ensemble réduit de régresseurs qui donne une reconstitution aussi satisfaisante de la variable à expliquer. Parmi les stratégies mises en œuvre pour la sélection d'un ensemble limité de variables explicatives, on peut citer : les méthodes de pas à pas, ainsi que les algorithmes évolutifs et génétiques.

Dans ce travail, nous nous sommes intéressés à la toxicité des alcools aliphatiques et des amines caractérisée par la concentration d'inhibition 50 % de la croissance (**CIC50**) et la dose létale (**CL50**), qui seront reliées à un descripteur géométrique (ADDD :Degré distance /distance moyenne) ,calculé en utilisant uniquement la structure des composés chimiques après optimisation de leur géométrie .

Les diagnostics statistiques ( $R^2$ ,  $R^2_{adj}$ ,  $Q^2$ ,  $Q^2_{ext}$ , EQMC, EQMP,  $EQMP_{ext}$ , F, S ) des modèles obtenus par les deux approches (Log P; ADDD) seront comparés .

Notre mémoire comporte en plus de la bibliographie, d'une introduction et d'une conclusion générale, deux grandes parties :

Dans la Partie Généralités, nous avons développé tout ce qui a trait au prétraitement des molécules (introduction des molécules, optimisation de leur géométrie) en vue du calcul des descripteurs moléculaires à l'aide de différents logiciels de modélisation moléculaire. Nous y avons également développé les connaissances théoriques de base utilisées tout au long de ce travail: algorithmes génétiques, régression linéaire simple, traitement statistique pour l'évaluation de la qualité de l'ajustement (robustesse des modèles; détection des observations aberrantes; test de randomisation; validation externe).

Dans la partie expérimentale, nous présentons et discutons les modèles calculés.

Les alcools et les amines sont des composés organiques qui sont largement utilisés dans le milieu industriel et dans la vie quotidienne. Que ce soit professionnellement ou à la maison, ces deux familles de produits sont incontournables. Les alcools sont trouvés à l'état pur ou en mélange dans des préparations spécifiques. Ils sont utilisés comme diluants des encres d'imprimerie, des résines, des vernis, des peintures et des colles à moquette. Ils sont aussi largement utilisés comme excipients pour les produits pharmaceutiques ou cosmétiques ou comme milieu réactionnel dans l'industrie chimique [1]. Les amines sont utilisées comme intermédiaires chimiques pour la synthèse de produits pharmaceutiques, cosmétiques, détergents.... Elles sont aussi utilisées comme solvants et inhibiteurs de corrosion [2].

Les alcools ont des effets néfastes bien connus, entraînant notamment des incoordinations motrices ou une excitation intellectuelle. Les alcools liquides et leurs vapeurs sont irritants pour la peau, les yeux et les muqueuses en cas de contact prolongé ou répété. L'inhalation accidentelle d'une grande quantité de vapeurs d'alcool peut conduire à des syndromes ébrioux ou narcotiques avec nausées, malaises, vomissements et maux de tête [1]. Quant aux effets indésirables des amines, les problèmes de santé pouvant se développer chez les travailleurs surexposés sont divers, allant de l'irritation cutanée au cancer [3].

Le transport, la distribution, l'accumulation et l'absorption des xénobiotiques (médicaments ou toxiques), pour certains leur fixation sur les protéines plasmatiques, leur passage à travers les membranes, leur entrée dans les cellules, les interactions enzyme-inhibiteur ou ligand-récepteur de nature hydrophobe, l'activité pharmacologique et pharmacocinétique des médicaments, la toxicité des médicaments ou des contaminants,..., les propriétés liées à la formulation telles que solubilité, sont autant d'exemples dans lesquels la lipophilie des molécules constitue un descripteur physico-chimique de la première importance. La lipophilie d'une molécule est mesurée classiquement par son aptitude à se distribuer dans un système biphasique soit liquide-liquide soit solide-liquide. Le coefficient de partage dans le système *n*-octanol/eau est connu depuis longtemps comme étant un des paramètres physico-chimiques quantitatifs qui est le mieux corrélé à l'activité des molécules organiques [4].

### I-LES ALCOOLS :

En chimie organique, un **alcool** est un composé organique dont l'un des carbones (celui ci étant tétragonal) est lié à un groupement hydroxyle (-OH). L'éthanol (ou alcool éthylique) entrant dans la composition des boissons alcoolisées est un cas particulier d'alcool, mais tous les alcools ne sont pas propres à la consommation. En particulier, le méthanol est toxique et mortel à haute dose.

#### I-1 Utilisation

Les alcools sont utilisés dans l'industrie chimique comme:

- Solvants : l'éthanol, peu toxique, est utilisé dans les parfums et les médicaments ;
- Combustibles : le méthanol et l'éthanol peuvent remplacer l'essence et le fioul car leur combustion ne produit pas de fumées toxiques ;
- Réactifs : les polyuréthanes, les esters ou les alcènes peuvent être synthétisés à partir des alcools ;
- Antigels : la basse température de solidification de certains alcools comme le méthanol et l'éthylène glycol en font de bons antigels .,

#### I-2 Classification des alcools

De manière générique, un alcool contient donc la séquence R - OH où R est un radical organique variable, souvent un alkyle. Selon la nature du carbone portant le groupement alcool, on distingue :

Les **alcools primaires**, dont le carbone comportant le groupement hydroxyle est lié à au moins deux atomes d'hydrogène et **un** radical organique R.

Les **alcools secondaires**, dont le carbone comportant le groupement hydroxyle est lié à un atome d'hydrogène et **deux** radicaux organiques R et R'.

Les **alcools tertiaires**, dont le carbone comportant le groupement hydroxyle est lié à **trois** radicaux organiques R, R' et R".,

- Il existe aussi un groupe considéré parfois comme un cas particulier d'alcools appelés **énols**. Il s'agit d'une molécule dans laquelle la fonction hydroxyle est attachée sur un carbone d'une double liaison C=C (ici encore le carbone portant le groupement -OH n'est pas tétragonal). Il s'agit en fait d'une forme tautomère d'un aldéhyde ou d'une cétone. La forme majoritaire est généralement l'aldéhyde ou la cétone, et non l'énol, sauf cas particuliers où la forme énol est stabilisée par mésomérie comme par exemple les phénols.

### I-3 Propriétés des alcools

Les alcools sont des composés amphotères, c'est-à-dire qu'ils sont à la fois acide et base. En d'autres termes, ils peuvent être protonnés par action d'un acide ou déprotonnés par action d'une base. Dans le cas de cette déprotonation il sera possible de faire une *O*-alkylation, et donc d'obtenir un éther. Dans le cas de la protonation, on fait un ion oxonium qui conduira à un carbocation qui pourra subir un réarrangement de façon à donner le carbocation le plus stable.

### I-4 Toxicité des alcools

De tous les alcools, le plus toxique est le méthanol dans la mesure où il exerce une action sélective au niveau du nerf optique, pouvant provoquer la cécité ou la mort. Les effets néfastes de l'absorption d'éthanol sont aussi bien connus, l'alcoolémie entraînant notamment des incoordinations motrices ou une excitation intellectuelle. De manière générale, les manifestations d'une intoxication modérée se traduiront par des maux de tête, des troubles digestifs et un syndrome ébrieux.

Les alcools liquides et leurs vapeurs sont irritants pour la peau, les yeux et les muqueuses en cas de contact prolongé ou répété. L'alcool furfurylique, plus agressif que les autres alcools, peut provoquer des larmoiements à de très faibles expositions (15 ppm) et des irritations respiratoires.

L'inhalation accidentelle d'une grande quantité de vapeurs d'alcool peut conduire à des syndromes ébrieux ou narcotiques avec nausées, malaises, vomissements et maux de tête.

**II-LES AMINES :**

Les amines sont des composés azotés qui dérivent formellement de l'ammoniac  $\text{NH}_3$  par remplacement d'un ou plusieurs atomes d'hydrogène par des groupes carbonés. Le nombre  $n$  des atomes d'hydrogène liés à l'azote, définit la classe de l'amine. Leur découverte est due au chimiste français Wurtz en 1849.

$n$	2	1	0
Classe	primaire	secondaire	tertiaire

**II-1 Utilisation:**

Les amines sont utilisées en grande quantité dans des procédés industriels variés, comme un intermédiaire réactionnel, dans des mélanges ou comme solvant. On les emploie surtout dans l'industrie chimique, dans celle des polymères et du caoutchouc, en agriculture comme pesticides, de même que dans la composition de peintures, d'adhésifs et de textiles, ainsi que dans l'industrie pharmaceutique [2].

**II-2 Classification des amines :**

La fonction amine recouvre un ensemble très étendu de composés. On distingue plusieurs séries :

- Acyclique : l'atome d'azote est relié à un ou plusieurs groupes alkyles. Exemple : la méthylamine  $\text{CH}_3\text{NH}_2$ .
- Alicyclique : l'atome d'azote est lié à un cycle non aromatique. Exemple : la cyclohexylamine.
- Hétérocyclique : l'atome d'azote est engagé dans un cycle qui peut être ou non aromatique.
- Aromatique : l'atome d'azote est lié à un cycle aromatique. Exemple : la phénylamine ou aniline.

Les amines aromatiques sont utilisées dans de nombreuses industries. La population générale peut être également exposée à ces amines, en raison du tabagisme ou lors d'apports alimentaires. Les voies de contamination peuvent être cutanée, respiratoire et digestive. Les effets toxiques aigus trouvés peuvent être une méthémoglobinémie (aniline), une anémie



hémolytique, une hépatite (MDA), une rhabdomyolyse avec insuffisance rénale chronique, une cardiomyopathie ou une atteinte oculaire.

Les effets chroniques peuvent être une irritation, une sensibilisation cutanée ou respiratoire .

### **II-3 Propriétés des amines**

La présence de l'atome d'azote est la cause des propriétés des amines. Cet atome présente un doublet non liant, ce qui donne aux amines un caractère basique et nucléophile. Dans le cas d'amine primaire et secondaire, la liaison N-H peut se rompre, ce qui leur donne un (faible) caractère acide. Les amines sont volatiles, ont une odeur forte et sont hydrosolubles [3].

### **II-4 Toxicité des amines**

Les amines sont un produit irritant et corrosif pour la peau, les yeux, les voies respiratoires et digestives. La gravité des symptômes peut varier selon les conditions d'exposition (durée de contact, concentration du produit, etc.).

L'exposition aux brouillards peut causer une irritation des yeux, de la peau et des voies respiratoires. L'exposition à de fortes concentrations peut provoquer l'œdème pulmonaire. Les symptômes de l'œdème pulmonaire (principalement toux et difficultés respiratoires) se manifestent souvent après un délai pouvant aller jusqu'à 48 heures. L'effort physique peut aggraver ces symptômes. Le repos et la surveillance médicale sont par conséquent essentiels[5].

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.

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 [6].

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 (**CIC50**), dont le logarithme de l'inverse, soit  $\text{pCIC50} = \log(\text{CIC50})^{-1}$ , servira d'indicateur de toxicité.

La dose létale **50** ou **DL50** (**LD50** en anglais pour *Léthale Dose 50*) ou **CL50** (concentration létale **50**) est un indicateur quantitatif de la toxicité d'une substance. Cette notion s'applique également aux irradiations.

Cet indicateur mesure la dose de substance causant la mort de **50 %** d'une population animale donnée (souvent des souris ou des rats) dans des conditions d'expérimentation précises. C'est la masse de substance nécessaire pour tuer **50 %** des animaux dans un lot. Elle s'exprime en milligrammes de matière active par kilogramme d'animal. Plus ce chiffre est petit, plus la substance est toxique. Cette dose n'est valable que pour une espèce précise (le plus souvent le rat) et un mode d'introduction précis dans l'organisme (ingestion, inhalation, application cutanée). Notons toutefois que la **CL50** peut varier, parfois fortement, en fonction du solvant utilisé ainsi qu'en fonction du sexe de l'animal. Ces chiffres ne sont pas directement extrapolables à l'homme.

### III – COLLECTE DES DONNEES

Les tests de toxicité ont été réalisés 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
Protéose 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 ml par fiole), quantité qui n'altère pas la reproduction de *Tétrahymena*.

La **CL50** est le plus souvent exprimée pour une ingestion orale chez le rat. La **CL50** mesurée par application dermale chez le lapin donne une information complémentaire sur le danger d'une substance. La valeur par ingestion est le plus souvent plus faible que la valeur dermale.

Ces mesures ne donnent toutefois qu'une idée partielle de la toxicité d'une substance, à laquelle il faut rattacher l'existence ou non d'un antidote. Ainsi, le parathion éthyle, malgré une **CL50** orale aiguë de 3,6 mg/kg, est relativement moins dangereux que le paraquat (**CL50** de 236 mg/kg) car il existe un antidote efficace contre le premier mais aucun contre le second.

Il est commun d'affirmer que les pesticides ayant une **CL50** orale inférieure à 50 mg/kg constituent des substances chimiques extrêmement toxiques pour l'homme.

Dans la pratique, il est également important de connaître la **CL50** du produit commercial utilisé. Celle-ci est calculée à partir de la concentration de la matière active dans la formulation au moyen de la formule suivante : Toxicité de la formulation (en mg ma/kg) = (**CL50** matière active x 100) / % matière active de la formulation

Nous avons réuni respectivement dans les tableaux I et II les toxicités relatives pCIC50 et p**CL50** d'intérêt .

**Tableau I** –Toxicité (CIC50) pour les alcools aliphatiques et les amines.

N°	Composés	Numéro de CAS	pCIC50
1	Méthanol	67-56-1	-2.77
2	Ethanol	64-17-5	-2.41
3	propan-1-ol	71-23-8	-1.84
4	Butan-1-ol	71-36-3	-1.52
5	pentan-1-ol	71-41-0	-1.12
6	hexan-1-ol	111-27-3	-0.47
7	heptan-1-ol	111-70-6	0.02
8	octan-1-ol	111-87-5	0.5
9	nonan-1-ol	143-08-8	0.77
10	decan-1-ol	112-30-1	1.1
11	undecan-1-ol	112-42-5	1.87
12	dodecan-1-ol	112-53-8	2.07
13	tridecan-1-ol	112-70-9	2.28
14	propan-2-ol	67-63-0	-1.99
15	pentan-2-ol	6032-29-7	-1.25
16	pentan-3-ol	584-02-1	-1.33
17	2-méthyl-1-butanol	137-32-6	-1.13
18	3-méthyl-1-butanol	123-51-6	-1.13
19	3-méthyl-2-butanol	598-75-4	-1.08
20	(tert) pentanol	75-85-4	-1.27
21	(neo) pentanol	75-84-3	-0.96
22	1-propylamine	107-10-8	-0.85
23	1-butylamine	109-73-9	-0.7
24	1-amylamine	110-58-7	-0.61
25	1-hexylamine	111-26-2	-0.34
26	1-heptylamine	111-68-2	0.1
27	1-octylamine	111-86-4	0.51
28	1-nonylamine	112-20-9	1.59
29	1-décylamine	2016-57-0	1.95
30	1-undécylamine	7307-55-3	2.26

**Tableau I1** –Toxicité (CL50) pour les amines.

N°	Composés	Numéro de CAS	pCL50
1	Propylamine	107-10-8	-0.717
2	Butylamine	109-73-9	-0.564
3	Sec-butylamine	33966-50-6	-0.575
4	Amylamine	110-58-7	-0.308
5	Hexylamine	111-26-2	0.252
6	Octylamine	111-86-4	1.396
7	Heptylamine	111-68-2	0.723
8	1-Méthylheptylamine	693-16-3	1.403
9	Nonylamine	112-20-9	1.822
10	Decylamine	2016-57-1	2.184
11	Undecylamine	7307-55-3	2.912
12	Tridecylamine	2869-34-3	3.487
13	Benzylamine	100-46-9	0.021
14	N,N-Diméthylbenzylamine	103-83-3	0.554
15	Aniline	62-53-3	-0.088
16	4-Toluidine	106-49-0	-0.143
17	4-Ethylaniline	589-16-2	0.220
18	4-Butylaniline	104-13-2	1.165
19	4-Octylaniline	16245-79-7	3.233
20	4-Décylaniline	37529-30-9	3.576
21	N-Méthylaniline	100-61-8	0.030
22	N,N-diméthylaniline	121-69-7	0.190
23	N,N-diéthylaniline	91-66-7	0.959

#### IV- CALCUL ET SELECTION DES DESCRIPTEURS MOLECULAIRES.

Nous avons utilisé le logiciel de modélisation moléculaire Hyperchem™ 6.03 [7] pour représenter les molécules puis, à l'aide de la méthode semi-empirique AM1 (Austin Model), 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 0,001 kcal/mol. Les géométries ainsi optimisées ont été transférées dans le logiciel informatique DRAGON version 5.3 [8] pour le calcul de 1664 descripteurs appartenant à différentes classes. Les descripteurs d'un même groupe, à valeur constante (écarts types inférieurs à 0,0001), et ceux hautement corrélés ( $R > 0,95$ ) ont été exclus.

En opérant sur les données de calibration, des sous-ensembles de descripteurs significatifs ont été sélectionnés par algorithme génétique (GA – VSS), dans la version MOBYDIGS de Todeschini [9] en maximisant le coefficient de prédiction  $Q^2_{\text{LOO}}$  et éventuellement le coefficient de prédiction externe  $Q^2_{\text{ext}}$ .

### **V-REGRESSION LINEAIRE SIMPLE.**

La plus simple et la plus populaire des techniques de régression.

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

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

Comme toutes les techniques prédictives, elle a deux objectifs :

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

La Régression Linéaire Simple traite de la question suivante :

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

Dans un premier temps, la RLS cherche à matérialiser le fait que les points expérimentaux sont approximativement alignés. Elle le fait en identifiant la "meilleure droite" passant au travers du nuage de points. Cette droite, dite "Droite des Moindres Carrés" (DMC) sera caractérisée par une pente  $b$  et une ordonnée à l'origine  $a$ . Ces grandeurs seront les deux premiers paramètres (ou coefficients) du modèle de la RLS.

## VI- DEVELOPPEMENT ET VALIDATION DU MODELE.

L'analyse de régression linéaire a été réalisée avec le logiciel MOBYDIGS [9] en utilisant la méthode des moindres carrés ordinaires.

La qualité de l'ajustement a été évaluée par le coefficient de détermination,  $R^2$ , et l'écart quadratique moyen calculé sur l'ensemble de calibration :

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

$y_i$  et  $\hat{y}_i$  étant les valeurs mesurées et calculées de la variable dépendante.

Les techniques de validation croisée ont été exploitées pour l'évaluation de la prédiction interne ( $Q_{LMO}^2$ ; bootstrap), et de la robustesse ( $Q_{LOO}^2$ ) du modèle.

La validation croisée par "leave-one-out" (LOO) [10] consiste à recalculer le modèle sur (n-1) objets, et utiliser le modèle ainsi obtenu pour prédire la valeur de la variable dépendante du composé écarté. Le procédé est répété pour chacun des objets de l'ensemble de calibration. La somme des carrés des erreurs de prédiction (désignée par l'acronyme PRESS pour Predictive Residual Sum of Squares) est une mesure de la dispersion des estimations. On l'utilise pour définir le coefficient de prédiction ( $Q_{LOO}^2$ ), et l'écart quadratique moyen de prédiction (ou EQMP) :

$$Q_{LOO}^2 = 1 - \frac{PRESS}{SCT} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_{i/i})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

$$EQMP = \sqrt{\frac{1}{n} PRESS} \quad (3)$$

$y_{i/i}$  désignant la réponse du i-ème objet estimée en utilisant un modèle obtenu sans faire intervenir cet i-ème objet, et  $\bar{y}$  la valeur moyenne des n observations; la sommation porte sur l'ensemble des composés de calibration; SCT est la somme des carrés totale.

Une valeur de  $Q_{LOO}^2 > 0,5$  est, en général, considérée comme satisfaisante et une valeur  $Q_{LOO}^2 > 0,9$  est excellente [11].

En fait, si une forte valeur de  $Q^2$  est une condition nécessaire d'une possible capacité prédictive élevée d'un modèle, cette condition seule n'est pas suffisante.

Pour éviter une surestimation de la capacité prédictive du modèle nous avons également appliqué la procédure "leave-more-out" (LMO), en excluant 50 % des objets à chaque étape ( $Q_{LMO/50}^2$ ). La procédure est répétée 5000 fois dans le présent travail.

Dans la technique de validation par bootstrap on simule de nouveaux échantillons de taille (n), par tirages aléatoires avec remise. De cette façon l'ensemble de calibration, qui conserve sa taille initiale (n), se compose, en général, d'objets répétés, l'ensemble d'évaluation rassemblant les objets exclus [12]. Le modèle est calculé sur l'ensemble de calibration et les réponses prédites pour l'ensemble d'évaluation. Tous les carrés des différences entre valeurs prédites et réelles des objets de l'ensemble d'évaluation sont collectés dans le PRESS. Cette procédure de construction des ensembles de calibration et d'évaluation est répétée plusieurs milliers de fois (5000 dans cette étude), les PRESS sont additionnés, et une capacité de prédiction moyenne calculée [13].

L'application du modèle, calculé sur l'ensemble de calibration, aux 10 composés de l'ensemble de validation, permet de vérifier de manière fiable la capacité prédictive du modèle obtenu.

L'équation ( 4 ) permet le calcul de  $Q_{ext}^2$  :

$$Q_{ext}^2 = 1 - \frac{\sum_{i=1}^{n_{ext}} (\hat{y}_{i/i} - y_i)^2 / n_{ext}}{\sum_{i=1}^{n_{tr}} (y_i - \bar{y}_{tr})^2 / n_{tr}} = 1 - \frac{PRESS / n_{ext}}{SCT / n_{tr}} \quad (4)$$

L'indice (ext) se rapportant aux objets de l'ensemble de validation externe (ou à ceux de l'ensemble d'évaluation obtenu par bootstrap), et l'indice tr à ceux de l'ensemble de calibration (training set).

$$EQMP_{ext} = \sqrt{\frac{1}{n_{ext}} \sum_{i=1}^{n_{ext}} (y_i - \hat{y}_i)^2} \quad (5)$$

La somme portant sur les objets de l'ensemble test ( $n_{ext}$ ).



## I - MODELISATION DE LA CONCENTRATION D'INHIBITION DE LA CROISSANCE 50:

Afin de procéder à une modélisation complète, nous devons valider le ou les modèle (s) obtenu (s). On a donc scindé notre ensemble de 30 observations en deux sous-groupes, un de calibration (à 20 composés) et un autre (à 10 composés) pour la validation statistique externe.

Le fractionnement ainsi fait, on a élaboré notre modèle sur l'ensemble de calibration. Puis on s'est assuré de sa validité sur le second ensemble.

En utilisant l'approche algorithme génétique dans la version MobyDigs, on a optimisé la fonction  $Q^2_{LOO}$  pour le choix de descripteurs significatifs .

On a choisi de développer et de comparer les deux modèles suivants (Tableau III) :

1-  $pCIC\ 50 = f(\log P)$

2-  $pCIC\ 50 = f(ADDD)$

**Tableau III** - Descripteurs moléculaires intervenant dans la modélisation de **pCIC50**

	Descripteur	Classe	Signification
Modèle 1	log P	Descripteur empirique	Coefficient de partage octanol/eau
Modèle 2	ADDD	Descripteur géométrique	Ordre distance /distance moyen.

Ce tableau comporte une brève définition des descripteurs ainsi que les blocs ou classes aux quels ils appartiennent.

On peut aussi définir [14-15] ADDD comme étant un descripteur géométrique dérivé de la matrice Distance/Distance notée **DD**, qui est une matrice carrée symétrique  $A \times A$ , où  $A$  est le nombre d'atomes non hydrogène dans la molécule.

Les éléments de cette matrice sont les quotients des éléments correspondants de la matrice de géométrie moléculaire  $\mathbf{G}$  et de la matrice des distances topologiques  $\mathbf{D}$ ; donc chaque élément est défini par l'équation :

$$[DD]_{ij} = \frac{[G]_{ij}}{[D]_{ij}} = \frac{r_{ij}}{d_{ij}} \quad (6)$$

Où  $r_{ij}$  et  $d_{ij}$  sont respectivement les distances géométriques et topologiques.

Les sommes des lignes de cette matrice contiennent des informations sur la compacité de la molécule; en fait, dans des structures compactes, ces sommes tendent à être relativement petites lorsque les distances interatomiques sont petites, quand les distances topologiques augmentent lorsque la taille de la structure augmente. Par conséquent, la moyenne de la somme des lignes est un invariant moléculaire nommé degré distance par distance moyen, ou ADDD pour (average distance/distance degree) c'est-à-dire :

$$ADDD = \frac{1}{A} \sum_{i=1}^A \sum_{j=1}^A \frac{r_{ij}}{d_{ij}} \quad (7)$$

Le deuxième descripteur utilisé est l'hydrophobicité des composés organiques qui est exprimée par le coefficient de partage octanol-eau ( $K_{ow}$ ) qui estime la solubilité dans deux phases : aqueuse et organique ; le *n*-octanol étant communément utilisé pour représenter cette dernière. L'équation suivante résume cette brève définition :

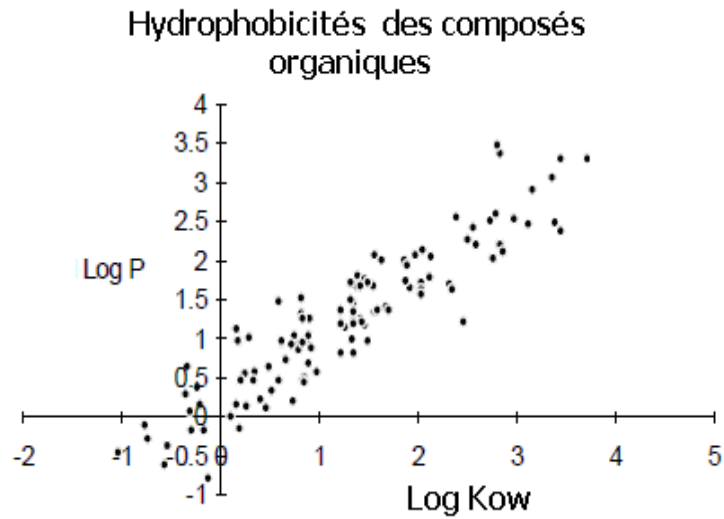
$$K_{ow} = \frac{C_{org}}{C_{aq}} \quad (8)$$

Les valeurs du coefficient de partage ayant une grande amplitude de variation, elles sont habituellement exprimées sous formes logarithmiques [16].

Le  $\log K_{ow}$  est essentiel pour comprendre les mécanismes de transport et de distribution des produits chimiques dans l'environnement, par exemple le mécanisme impliquant l'absorption des substances chimiques par transport à travers une membrane biologique ce qui est, par définition, la cause de leurs plus au moins grandes toxicités.

Dans notre étude nous symboliserons par  $\log P$ , en accord avec le logiciel utilisé pour son calcul, le coefficient de partage. Le module des propriétés QSAR du logiciel de modélisation moléculaire HyperChem offre un calcul rapide du  $\log P$  en utilisant les paramètres atomiques définis par Ghose, Pritchett et Crippen [17] pour un échantillon de

composés organiques ; ces paramètres ont été par la suite recalculés par Ghose et ses associés [18]. Le graphe suivant permet de juger l'exactitude de cette méthode.



**Figure 1** : Valeurs des coefficients de partage calculés (*Log P*) en fonction des valeurs expérimentales (*Log Kow*).

**Modèle 1 : pCIC 50 =f(logP)****I-1-1- Equation et analyse de régression :**

Le modèle obtenu a pour équation :

$$\mathbf{pCIC50} = - 2.06(\pm 0.09778) + 0.803(\pm 0.03538) \log P \quad (9)$$

Les paramètres statistiques de ce modèle sont réunis ci après :

$n_{tr}$	$n_{valid}$	$R^2$	$Q^2$	$Q^2_{boot}$	$R^2_{adj}$	EQMC	EQMP	F	S
20	10	96.62	95.73	95.06	96.43	0.268	0.301	514.83	0.282

$Q^2_{Imo10\%}$	$Q^2_{Imo20\%}$	$Q^2_{Imo30\%}$	$Q^2_{Imo40\%}$	$Q^2_{Imo50\%}$
95.67	96.00	95.75	96.26	95.05

La concentration inhibitrice 50% de la croissance pour les 20 composés utilisés pour la calibration (élaboration du modèle) est bien corrélée avec le descripteur, d'où la grande valeur du coefficient de détermination  $R^2$ . Notre modèle a de très bonnes capacités prédictives confirmées par la valeur de  $Q^2$  qui est élevé. La statistique de Fisher montre que notre modèle est très significatif. Les écarts quadratiques (EQMP/C) sont faibles et proches, la stabilité interne est vérifiée car les valeurs de  $Q^2_{loo}$ ,  $Q^2_{Imo}$ , ainsi que celle de la validation par bootstrap sont très proches.

**I-1-2- Analyse des résidus et diagnostics d'influences :**

Les valeurs des résidus, ordinaires et de prédiction standardisés, ainsi que les valeurs de  $h_{ii}$  sont présentées dans le tableau **IV**, dont les colonnes 3 et 4 reproduisent les valeurs expérimentales et calculées de la concentration inhibitrice 50% de la croissance des composés considérés.

**Tableau IV** : Valeurs des **pCIC 50** observés (**pCIC 50<sub>obs</sub>**) et calculés (**pCIC 50<sub>cal</sub>**), des résidus ordinaires ( $e_i$ ) et standardisés ( $e_{istd}$ ) ainsi que des leviers ( $h_{ii}$ ).

N°	Composé	<b>pCIC 50<sub>obs</sub></b>	<b>pCIC 50<sub>cal</sub></b>	$e_i$	$e_{i\ std}$	$h_{ii}$
1	methanol	-2.77	-2.6768	0.0932	0.4448	0.18
2	ethanol	-2.41	-2.3075	0.1025	0.4566	0.142
3	propan-1-ol	-1.84	-1.858	-0.018	-0.0752	0.104
4	pentan-1-ol	-1.12	-0.8064	0.3136	1.2085	0.055
5	hexan-1-ol	-0.47	-0.4291	0.0409	0.1564	0.05
6	heptan-1-ol	0.02	0.0044	-0.0156	-0.0601	0.053
7	nonan-1-ol	0.77	0.9034	0.1334	0.5436	0.089
8	decan-1-ol	1.1	1.3369	0.2369	1.0173	0.121
9	dodecan-1-ol	2.07	2.0594	-0.0106	-0.052	0.193
10	tridecan-1-ol	2.28	2.4928	0.2128	1.158	0.249
11	propan-2-ol	-1.99	-2.0185	-0.0285	-0.1217	0.117
12	2-methyl-1-butanol	-1.13	-0.9188	0.2112	0.8175	0.057
13	3-methyl-1-butanol	-1.13	-0.9188	0.2112	0.8175	0.057
14	3-methyl-2-butanol	-1.08	-1.0312	0.0488	0.19	0.061
15	(tert)pentanol	-1.27	-1.0874	0.1826	0.7128	0.063
16	1-propylamine	-0.85	-1.6734	-0.8234	-3.3688	0.092
17	1-hexylamine	-0.34	-0.405	-0.065	-0.2488	0.05
18	1-heptylamine	0.1	0.0044	-0.0956	-0.3677	0.053
19	1-octylamine	0.51	0.3816	-0.1284	-0.5017	0.064
20	1-undecylamine	2.26	1.658	-0.602	-2.719	0.15

**I-1-3- Validation externe :**

Pour vérifier les capacités prédictives de notre modèle on a eu recours à sa validation sur un ensemble prévu à cet effet et choisi dès le départ. Cet ensemble de validation, qui n'a pas servi à l'élaboration du modèle, est constitué des composés numérotés de 21 à 30 (tableau V).

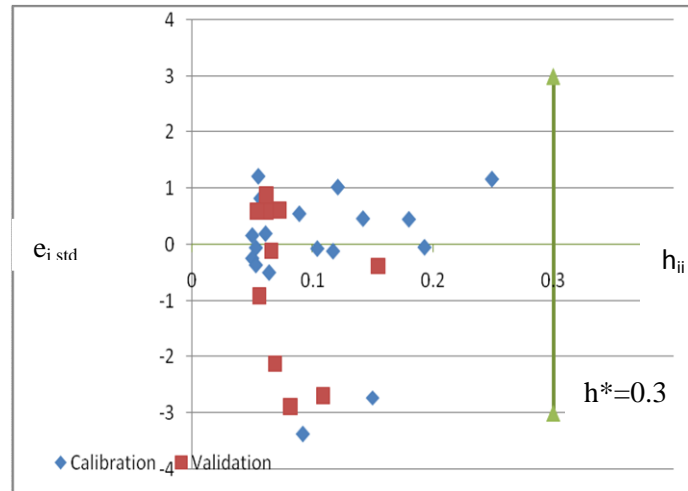
**Tableau V :** Quelques caractéristiques des éléments de l'ensemble de validation externe

N°	Composé	pCIC50 <sub>obs</sub>	pCIC50 <sub>pred</sub>	e <sub>i std</sub>	h <sub>ij</sub>
21	butan-1-ol	-1.52	-1.3872	0.5281	0.074
22	octan-1-ol	0.5	0.3337	-0.6568	0.061
23	undecan-1-ol	1.87	1.6773	-0.798	0.146
24	pentan-2-ol	-1.25	-1.0087	0.9522	0.06
25	pentan-3-ol	-1.33	-1.0059	1.2786	0.059
26	(neo)pentanol	-0.96	-1.0332	-0.2887	0.06
27	1-butylamine	-0.7	-1.2665	-2.2458	0.069
28	1-amylamine	-0.61	-0.8417	-0.912	0.055
29	1-nonylamine	1.59	0.9114	-2.718	0.087
30	1-decylamine	1.95	1.3619	-2.3967	0.119

La valeur de  $Q^2_{\text{ext}} = 92.1\%$  nous renseigne sur capacité du modèle à prédire des valeurs qui n'ont pas servi à le générer. L'écart quadratique moyen de prédiction externe  $EQMP_{\text{ext}} = 0.41$  est faible, gage aussi d'une bonne aptitude à la prédiction.

**I-1-4- Diagramme de Williams :**

On a représenté, sur la même figure 1 pour les deux ensembles (calibration et validation); les valeurs de  $e_{i\text{std}}$  (tableau IV colonne 6 et tableau V colonne 5) de  $h_{ii}$  (tableau IV colonne 7 et tableau V colonne 6).



**Figure2:** Diagramme de Williams pour les éléments des ensembles de calibration(20) et de validation(10)

Les valeurs des leviers étant toutes inférieures à la valeur critique  $h^* = 0.3$  il n'y a pas de point influent ; tous les résidus de prédiction standardisés sont inférieurs à 3 unités d'écart type (3) à l'exception du point «16».

(1-propylamine) dans l'ensemble de calibration qui est un point aberrant.

**I-1-5- Qualité de l'ajustement :**

La figure 2 représente les deux droites d'ajustement pour l'ensemble de calibration ( $pCIC50_{\text{obs}}$  en fonction de  $pCIC50_{\text{cal}}$  définie par l'équation (9) et celle pour l'ensemble de validation ( $pCIC50_{\text{obs}}$  en fonction de  $pCIC50_{\text{pred}}$  définie par l'équation (10).

L'équation de régression est :

Pour l'ensemble de validation :

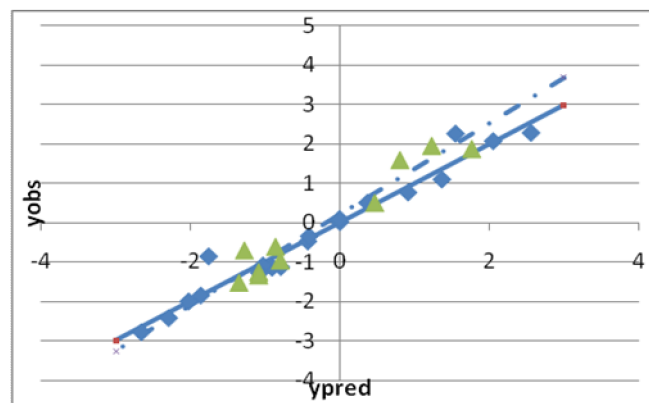
$$pCIC50_{obs} = 0.207105 + 1.15668 pCIC50_{pred} \quad (10)$$

$$S = 0.3671095 \quad R\text{-carré} = 93.8\% \quad R\text{-carré(ajust)} = 93.1\%$$

Pour l'ensemble de calibration

$$pCIC50_{obs} = -0.0000002 + 1.00000 pCIC50_{calc} \quad (11)$$

$$S = 0.282359 \quad R\text{-carré} = 96.6\% \quad R\text{-carré(ajust)} = 96.4\%$$



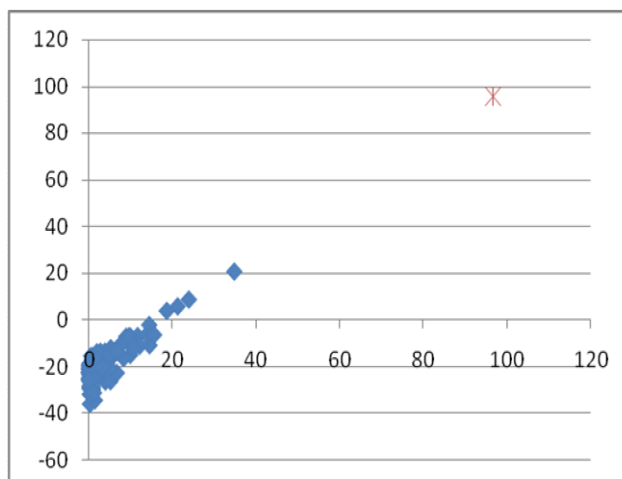
*Figure 3 : Droites d'ajustement des deux ensembles .*

On remarque, relativement, une faible dispersion autour des droites d'ajustement (Calibration ou validation) ce qui traduit la faiblesse des erreurs lors du calcul (Calibration) et de la prédiction (Validation).

On déduit donc, de la figure précédente, qu'on a un bon ajustement confirmé par les valeurs du coefficient de régression  $R^2$  élevé, pour les deux ensembles.



La validité du modèle a été éprouvée par le test de randomisation de **pCIC50** (Figure 3 ).



**Figure 4 :** Test de randomisation associé au modèle *QSPR* .

Les losanges noircis représentent les concentrations inhibitrices 50% de la croissance ordonnées de façon aléatoire, et l'étoile correspond au modèle réel.

Il est clair que les statistiques obtenues pour le **pCIC50** modifié sont plus petites que celles du modèle réel, et pour la majeure partie on obtient un  $Q^2 < 0$ . Ceci permet d'assurer qu'une relation structure /activité réelle a été établie .

## Modèle 2 : **pCIC50 = f (ADDD)**

### I-2-1- Equation et analyse de régression :

Le modèle obtenu par RLS est résumé par l'équation de régression suivante :

$$\mathbf{pCIC50} = - 3.33(\pm 0.1342) + 0.138(\pm 0.005602) \mathbf{ADDD} \quad (12)$$

les paramètres statistiques de ce modèle sont réunis ci-après :

<b>n<sub>tr</sub></b>	<b>n<sub>valid</sub></b>	<b>Q<sup>2</sup></b>	<b>R<sup>2</sup></b>	<b>Q<sup>2</sup><sub>Boot</sub></b>	<b>R<sup>2</sup><sub>adj</sub></b>	<b>EQMC</b>	<b>EQMP</b>	<b>F</b>	<b>s</b>
20	10	96.24	97.11	95.67	96.94	0.248	0.283	603.82	0.261

Suite de la page précédente.

$Q^2_{lmo10\%}$	$Q^2_{lmo20\%}$	$Q^2_{lmo30\%}$	$Q^2_{lmo40\%}$	$Q^2_{lmo50\%}$
97.20	97.40	97.21	97.56	96.75

Le modèle calculé sur les 20 observations que compte le groupe de calibration corrèle très bien la concentration inhibitrice 50% de la croissance à cette variable, puisque la valeur de  $R^2$  est grande. La valeur de  $Q^2$  l'est aussi et montre les capacités prédictives internes du modèle. La valeur élevée de la statistique de Fisher indique un modèle très hautement significatif. Les écarts quadratiques (EQMP/C) sont faibles et proches, la stabilité interne du modèle est vérifiée car les  $Q^2_{Boot}$ ,  $Q_{LMO}$  et  $Q_{LOO}$  sont très proches.

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

Les valeurs de **pCIC50** calculées et expérimentales, les résidus ordinaires et standardisés ainsi que les valeurs des leviers pour chaque composé sont rassemblés dans les colonnes de 3 à 7 du tableau VI.

**Tableau VI:** Valeurs des **pCIC 50** observés (**pCIC 50<sub>obs</sub>**) et calculés (**pCIC 50<sub>cal</sub>**), des résidus ordinaires ( $e_i$ ) et de prédiction standardisés ( $e_{i\text{std}}$ ) ainsi que des leviers ( $h_{ii}$ ).

N°	Composé	<b>pCIC50<sub>obs</sub></b>	<b>pCIC50<sub>cal</sub></b>	$e_i$	$e_{i\text{std}}$	$h_{ii}$
1	methanol	-2.77	-2.65	0.12	0.6147	0.177
2	ethanol	-2.41	-2.2172	0.1928	0.9141	0.133
3	propan-1-ol	-1.84	-1.8056	0.0344	0.1541	0.1
4	pentan-1-ol	-1.12	-0.9665	0.1535	0.643	0.059
5	hexan-1-ol	-0.47	-0.5383	-0.0683	-0.2826	0.051
6	heptan-1-ol	0.02	-0.1053	-0.1253	-0.519	0.052

Tableau VI: Suite

N°	Composé	pCIC50 <sub>obs</sub>	pCIC50 <sub>cal</sub>	e <sub>i</sub>	e <sub>i std</sub>	h <sub>ii</sub>
7	nonan-1-ol	0.77	0.7768	0.0068	0.0294	0.082
8	decan-1-ol	1.1	1.2254	0.1254	0.5724	0.111
9	dodecan-1-ol	2.07	2.1336	0.0636	0.3406	0.201
10	tridecan-1-ol	2.28	2.5929	0.3129	1.8882	0.262
11	propan-2-ol	-1.99	-1.8209	0.1691	0.7595	0.101
12	2-methyl-1-butanol	-1.13	-1.0194	0.1106	0.4646	0.06
13	3-methyl-1-butanol	-1.13	-1.0162	0.1138	0.4779	0.06
14	3-methyl-2-butanol	-1.08	-1.0453	0.0347	0.1461	0.061
15	(tert)pentanol	-1.27	-1.0428	0.2272	0.9556	0.061
16	1-propylamine	-0.85	-1.6859	-0.8359	-3.6982	0.092
17	1-hexylamine	-0.34	-0.4115	-0.0715	-0.2956	0.05
18	1-heptylamine	0.1	0.0244	-0.0756	-0.3142	0.054
19	1-octylamine	0.51	0.4661	-0.0439	-0.1863	0.067
20	1-undecylamine	2.26	1.8159	-0.4441	-2.228	0.165

**I-2-3- Validation externe :**

L'ensemble des données écartées avant le calcul du modèle nous a servi à valider ce dernier, afin d'évaluer sa capacité à prédire d'éventuelles nouvelles valeurs. Les composés de validation sont numérotés de 21 à 30 dans le tableau **VII**.

Les valeurs des erreurs standards de prédiction  $e_{istd}$ , les valeurs des leviers ainsi que les valeurs prédites  $pCIC50_{pred}$  et observées  $pCIC50_{obs}$  de la concentration inhibitrice 50% de la croissance  $y$  sont aussi rapportées.

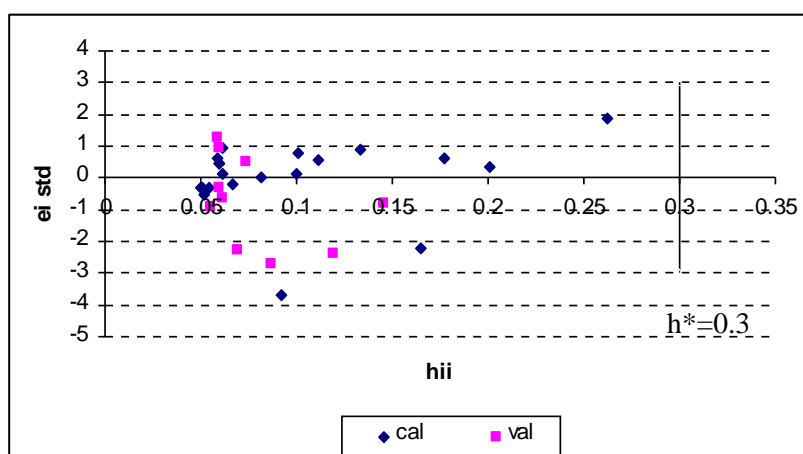
**Tableau VII :** Quelques caractéristiques des éléments de l'ensemble de validation externe.

N°	Composé	pCIC50 <sub>obs</sub>	pCIC50 <sub>pred</sub>	e <sub>i std</sub>	h <sub>ii</sub>
21	1-butanol	-1.52	-1.3872	0.5281	0.074
22	1-octanol	0.5	0.3337	-0.6568	0.061
23	1-undecanol	1.87	1.6773	-0.798	0.146
24	2-pentanol	-1.25	-1.0087	0.9522	0.06
25	3-pentanol	-1.33	-1.0059	1.2786	0.059
26	(neo)pentanol	-0.96	-1.0332	-0.2887	0.06
27	1-butylamine	-0.7	-1.2665	-2.2458	0.069
28	1-amylamine	-0.61	-0.8417	-0.912	0.055
29	1-nonylamine	1.59	0.9114	-2.718	0.087
30	1-decylamine	1.95	1.3619	-2.3967	0.119

Les valeurs introduites pour la validation du modèle ont été prédites sans trop d'erreur (valeur de l'écart quadratique moyen de prédiction externe faible EQMP<sub>ext</sub> = 0.378). Cette constatation est confirmée par la grande valeurs de Q<sup>2</sup><sub>ext</sub>=93.26% qui est une preuve de l'aptitude du modèle à la prédiction.

**I-2-4- Diagramme de Williams :**

La figure (4) 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 5:** Diagramme de Williams pour le Modèle 2

Le diagramme de Williams fait ressortir un point aberrant « 16 » (1-propylamine) dans l'ensemble de calibration et une absence de points influents.

**I- 2-5-Qualité de l'ajustement :**

La qualité de l'ajustement peut être appréciée dans la figure (5) par les deux droites d'ajustements dont les équations sont :

- Pour l'ensemble de validation

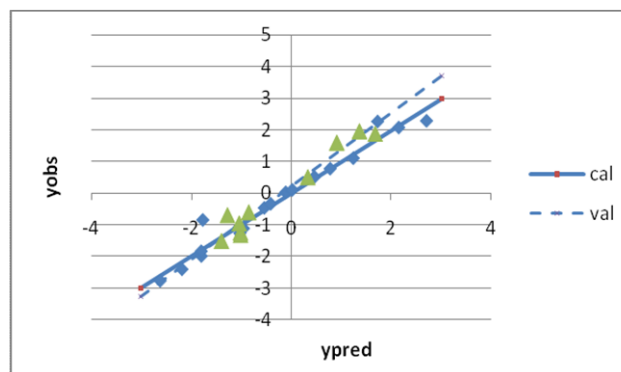
$$pCIC50_{obs} = 0.215680 + 1.15844 pCIC50_{pred} \quad (13)$$

S = 0.315493                  R-carré = 95.4%                  R-carré (ajust) = 94.9%

- Pour l'ensemble de calibration

$$pCIC50_{Exp.} = -0.0000141 + 0.999989 pCIC50_{Calc} \quad (14)$$

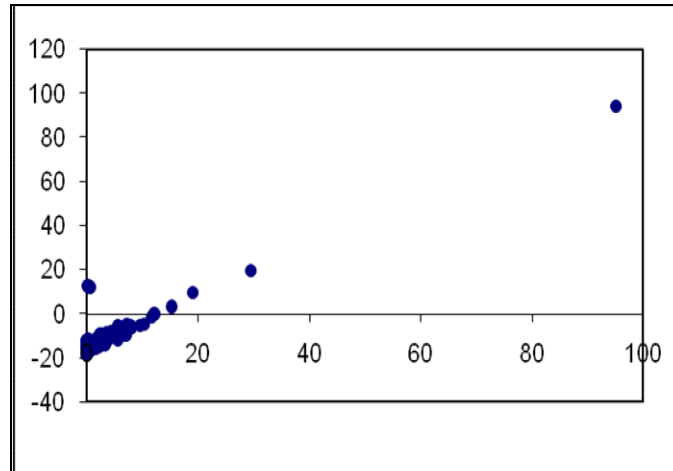
S = 0.261376                  R-carré = 97.1 %                  R-carré(ajust) = 96.9 %



**Figure 6:** Droites d'ajustement : cas du modèle 2

La faible dispersion des points autour des droites montre que les valeurs prédites (pour l'ensemble de validation) et calculées (pour l'ensemble de calibration) sont en adéquation avec les valeurs expérimentales. Les valeurs de R<sup>2</sup> montrent la bonne qualité d'ajustement que présente notre modèle.

Par le test de randomisation (figure 6), nous nous sommes assuré qu'une relation structure-activité réelle a été établie par notre modèle.



*Figure7: Test de randomisation :cas du modèle 2*

Les 100 modèles pour lesquels nous avons randomisé les valeurs de la concentration inhibitrice 50% de la croissance.ont des valeurs ou faibles ou négatives pour  $Q^2$ .

Seul le modèle réel offre des valeurs élevées pour les deux statistiques représentées dans la figure 6.

## II- MODELISATION DE LA CONCENTRATION LEALE 50 :

Vu le nombre réduit des données pour ce groupe nous n'avons pas cherché à le décomposer en un ensemble de calibration (pour le calcul du modèle) et un ensemble de test (pour la validation statistique externe).

Parmi les modèles obtenus par algorithme génétique, notre choix s'est porté sur les deux premiers caractérisés par les valeurs élevées de  $Q^2$  et  $R^2$ .

On a développé et comparé les deux modèles suivants :

$$1- pCL50 = f(\log P)$$

$$2- pCL50 = f(ADDD)$$

### Modèle 1 : $pCL50 = f(\log P)$

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

Le modèle obtenu a pour équation :

$$pCL50 = - 1.20 (\pm 0.1077) + 0.804 (\pm 0.03455) \log P \quad (15)$$

Les paramètres statistiques sont

$R^2$	$Q^2$	$Q^2_{boot}$	$R^2_{adj}$	SDEC	SDEP	F	S
96.26	95.5	94.93	96.09	0.255	0.279	541.1333	0.2664

$Q^2_{lmo10\%}$	$Q^2_{lmo20\%}$	$Q^2_{lmo30\%}$	$Q^2_{lmo40\%}$	$Q^2_{lmo50\%}$
95.67	95.67	95.67	95.67	95.67

$R^2$  et  $R^2_{adj}$  sont grands et reflètent le bon ajustement du modèle, la différence entre eux lui confère une certaine stabilité.

La stabilité interne est vérifiée car les valeurs de  $Q^2_{LOO}$ ,  $Q^2_{LMO}$  ainsi que celle de la validation par bootstrap sont très proches.

**II-1-2 Analyse des résidus et diagnostics d'influence :**

Le logiciel MOBYDIGS utilisé pour le calcul du modèle nous aide à déterminer les valeurs des résidus caractéristiques réunis dans le tableau **VIII**.

**Tableau VIII:** Valeurs des **pCL50** observées (**pCL50<sub>obs</sub>**) et calculées (**pCL50<sub>cal</sub>**), des résidus ordinaires (**e<sub>i</sub>**) et de prédiction standardisés (**e<sub>istd</sub>**) ainsi que des leviers (**h<sub>ij</sub>**).

N°	Composé	pCL50 <sub>obs</sub>	pCL50 <sub>cal</sub>	e <sub>i</sub>	e <sub>i std</sub>	h <sub>ij</sub>
1	Propylamine	-0.717	-0.815	-0.098	-0.4489	0.124
2	Butylamine	-0.564	-0.4213	0.1427	0.6194	0.092
3	Sec-butylamine	-0.575	-0.6382	-0.0632	-0.2821	0.109
4	Amylamine	-0.308	-0.0034	0.3046	1.2686	0.067
5	Hexylamine	0.252	0.4547	0.2027	0.8213	0.05
6	Octylamine	1.396	1.2502	-0.1458	-0.5871	0.046
7	Heptylamine	0.723	0.8645	0.1415	0.568	0.044
8	1-Méthylheptylamine	1.403	1.0654	-0.3376	-1.3554	0.044
9	Nonylamine	1.822	1.6681	-0.1539	-0.6309	0.057
10	Decylamine	2.184	2.094	-0.09	-0.3814	0.078
11	Undecylamine	2.912	2.5199	-0.3921	-1.7472	0.108
12	Tridecylamine	3.487	3.3637	-0.1233	-0.6417	0.196
13	Benzylamine	0.021	-0.3248	-0.3458	-1.4843	0.085
14	N,N-Diméthylbenzylamine	0.554	0.3663	-0.1877	-0.7637	0.052
15	Aniline	-0.088	-0.4775	-0.3895	-1.7016	0.096
16	4-Toluidine	-0.143	-0.0837	0.0593	0.2484	0.071
17	4-Ethylaniline	0.22	0.3743	0.1543	0.6275	0.052
18	4-Butylaniline	1.165	1.3306	0.1656	0.6685	0.047
19	4-Octylaniline	3.233	3.0342	-0.1988	-0.9642	0.157
20	4-Décylaniline	3.576	3.878	0.302	1.8084	0.268
21	N-Méthylaniline	0.03	0.1172	0.0872	0.3597	0.061
22	N-N-diméthylaniline	0.19	0.6556	0.4656	1.8746	0.046
23	N,N-diéthylaniline	0.959	1.4592	0.5002	2.0289	0.05

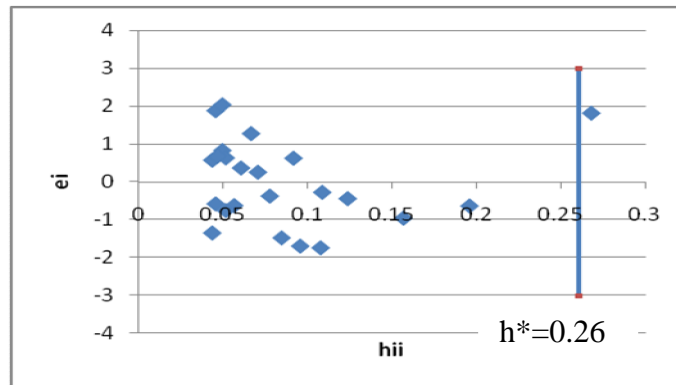


Tous les résidus standardisés  $e_{i\text{std}}$  de la colonne (6) sont inférieurs à 3 unités d'écart type (3 ).

La colonne (7) donne les valeurs de  $h_{ii}$ ,  $i^{\text{ème}}$  terme diagonal de la matrice de projection :  $\underline{H} = \underline{X} (\underline{X}' \underline{X})^{-1} \underline{X}'$  où  $\underline{X}$  est la matrice des valeurs observées des variables explicatives et  $\underline{X}'$  sa transposée. La valeur critique pour déterminer les points leviers

correspond à  $h^* = \frac{3p}{n} = 3 \cdot 2 / 23 = 0,26$ . On constate que tous les  $h_i$  sont inférieurs à cette valeur critique 0,26 à l'exception du composé 21 (N-Méthylaniline ).

### II -1-3 Le diagramme de Williams



**Figure 8 :** Diagramme de Williams pour tout les composés.

Le diagramme précédent ( $e_{i\text{std}}$  en fonction de  $h_{ii}$ ) de la figure (7) présente clairement l'observation influente qui est : le point 21(N-Méthylaniline ), qui possède un bras de levier important.

### II -1- 4 Qualité de l'ajustement

La qualité de l'ajustement se vérifie sur la figure (8), qui reproduit les valeurs observées **pCL50** en fonction de celles calculées. Elle fait ressortir une faible dispersion caractéristique d'un bon ajustement, d'ailleurs confirmé par la grande valeur de  $R^2$  (= 95.5%).

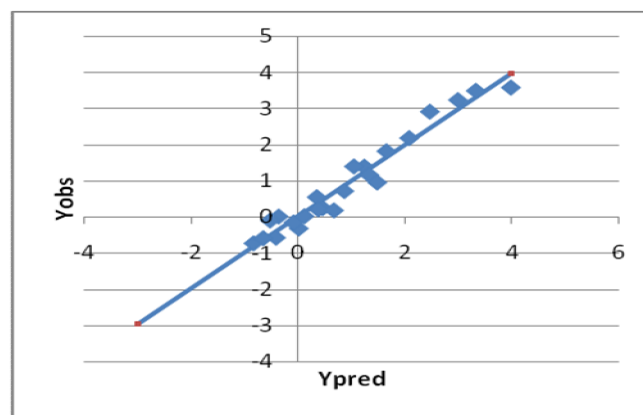
L'équation de régression est

$$\mathbf{pCL50}_{\text{obs}} = 0.0090396 + 0.991354 \mathbf{pCL50}_{\text{calc}} \quad (16)$$

$$S = 0.292162$$

$$R\text{-carré} = 95.5\%$$

$$R\text{-carré(ajust)} = 95.3\%$$



**Figure 9 :** Droite d'ajustement des **pCL50** observés en fonction des **pCL50** calculés.

### II -1- 5 Test de randomisation :

Dans le but d'établir que le modèle obtenu n'est pas dû au hasard ou à une sur-spécification, nous avons appliqué le test de randomisation.

Ainsi 100 nouveaux vecteurs de **la concentration létale 50** ont été générés par permutation des positions des composantes du vecteur réel:

$$y = (y_1, y_2, \dots, y_{27}) \xrightarrow{RND} y_{RND} = (y_8, y_5, \dots, y_2)$$

et utilisés comme sources d'observations pour des modèles QSAR dans les conditions optimales établies .

La figure (9) qui représente le graphe des coefficients statistiques  $Q^2$  et  $R^2$  permet de comparer les résultats obtenus pour les modèles randomisés au modèle réel de départ.

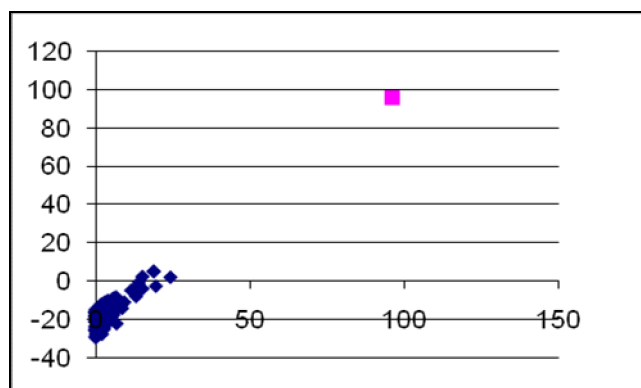


Figure 10: Test de randomisation

Il est clair que les statistiques obtenues pour les vecteurs modifiés de la concentration létale 50 sont plus petites que celles du modèle QSPR réel. Le modèle proposé n'est donc pas aléatoire.

Modèle 2 :  $pCL50 = f(ADDD)$

### II-2-1 Equation et analyse de régression :

L'équation du modèle calculé est la suivante :

$$pCL50 = - 2.41(\pm 0.1221) + 0.138(\pm 0.004665) ADDD \quad (17)$$

Les paramètres statistiques sont reproduits ci après :

$R^2$	$Q^2$	$Q^2_{boot}$	$R^2_{adi}$	EQMC	EQMP	F	S
97.64	97.09	96.65	97.53	0.202	0.225	870.61	0.212

$Q^2_{lmo10\%}$	$Q^2_{lmo20\%}$	$Q^2_{lmo30\%}$	$Q^2_{lmo40\%}$	$Q^2_{lmo50\%}$
97.20	97.20	97.20	97.20	97.20

Les paramètres statistiques reproduits ci-dessus montrent une bonne explication de la variabilité de  $pCL50$ , par le descripteur choisi, de l'ordre de 97.64% pour le coefficient de détermination, et une robustesse du modèle due à la grande valeur de  $Q^2$  (97.09%) .

Les valeurs des écarts quadratiques moyens de prédiction (EQMP) et de calcul (EQMC) sont proches et faibles. La grande valeur du paramètre de Fisher montre un modèle hautement significatif.

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

Le calcul effectué à l'aide du logiciel MobyDigs entre autres des résidus ordinaires,  $e_i$ , et de prédiction standardisés, que nous noterons  $e_{i\text{ std}}$ , donne les résultats qui apparaissent respectivement dans les colonnes 4 et 5 du tableau IV.

**Tableau IX :** Valeurs des **pCL50** observés (**pCL50<sub>obs</sub>**) et calculés (**pCL50<sub>cal</sub>**), des résidus ordinaires ( $e_i$ ) et de prédiction standardisés ( $e_{i\text{ std}}$ ) ainsi que des leviers ( $h_{ii}$ ).

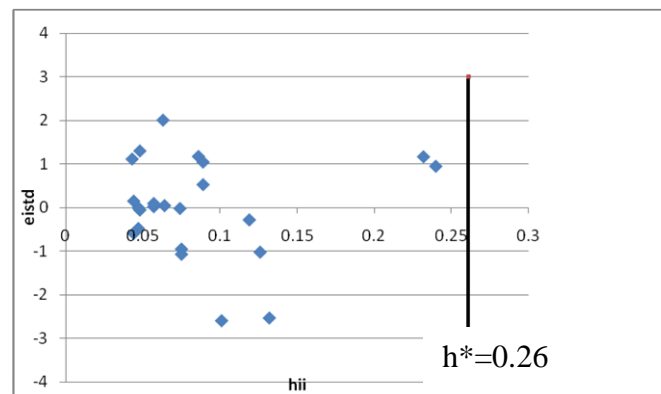
Composé	<b>pCL50<sub>obs</sub></b>	<b>pCL50<sub>cal</sub></b>	$e_i$	$e_{i\text{ std}}$	$h_{ii}$
Propylamine	-0.717	-0.7662	-0.0492	-0.2812	0.119
Butylamine	-0.564	-0.3468	0.2172	1.1756	0.086
Sec-butylamine	-0.575	-0.3823	0.1927	1.047	0.089
Amylamine	-0.308	0.078	0.386	2.0109	0.063
Hexylamine	0.252	0.5081	0.2561	1.3041	0.048
Octylamine	1.396	1.3857	-0.0103	-0.0525	0.048
Heptylamine	0.723	0.944	0.221	1.1168	0.043
1-Méthylheptylamine	1.403	1.3086	-0.0944	-0.4796	0.047
Nonylamine	1.822	1.831	0.009	0.0468	0.064
Decylamine	2.184	2.2815	0.0975	0.5302	0.089
Undecylamine	2.912	2.7354	-0.1766	-1.0213	0.126
Tridecylamine	3.487	3.6535	0.1665	1.169	0.232
Benzylamine	0.021	-0.158	-0.179	-0.9506	0.075
N,N-Diméthylbenzylamine	0.554	0.5534	-0.0006	-0.0032	0.047
Aniline	-0.088	-0.5559	-0.4679	-2.5963	0.101
4-Toluidine	-0.143	-0.1464	-0.0034	-0.0181	0.074
4-Ethylaniline	0.22	0.2245	0.0045	0.0234	0.057
4-Butylaniline	1.165	1.0485	-0.1165	-0.5892	0.044
4-Octylaniline	3.233	2.7994	-0.4336	-2.5337	0.132
4-Décylaniline	3.576	3.7094	0.1334	0.951	0.24
N-Méthylaniline	0.03	-0.171	-0.201	-1.0691	0.075
N-N-diméthylaniline	0.19	0.2088	0.0188	0.0974	0.057
N,N-diéthylaniline	0.959	0.989	0.03	0.1516	0.044

Les valeurs des résidus de prédiction standardisés (colonne 5) sont toutes inférieures à 3 unités d'écart type (3 ).

Les valeurs de  $h_{ii}$ , reproduites dans la colonne 5, étant toutes inférieures à la valeur critique  $h_{ii}^* = \frac{3p}{n} = 3 \cdot 2/23 = 0,26$  ne mettent pas en évidence de point levier.

### II-2-3 Le diagramme de Williams

La figure 10 qui représente la variation des résidus de prédiction standardisés  $e_{istd}$ , en fonction des leviers  $h_{ii}$  des composés ne laisse pas apparaître de points aberrants ni influents.



**Figure 11** : *Diagramme de Williams pour tout les composés .*

## II-2-4 - Qualité de l'ajustement :

La qualité de l'ajustement à été vérifiée en représentant les valeurs observées ou expérimentales  $pCL50_{obs}$  (colonne2 Tableau IX) en fonction des valeurs calculées  $pCL50_{cal}$  (colonne 3 Tableau IX) à l'aide de notre modèle. La figure 11 montre une faible dispersion autour de la droite d'ajustement .

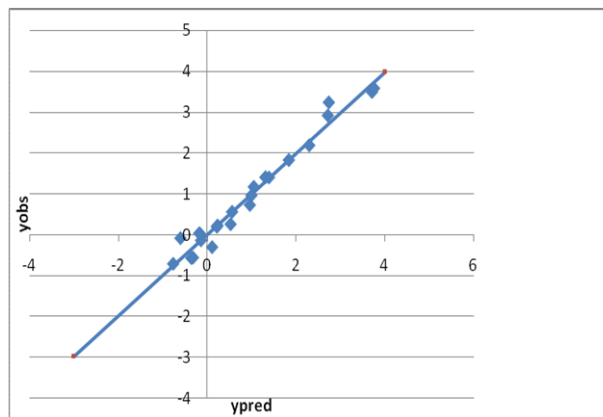
L'équation de régression est :

$$pCL50_{obs}=0.0061639+0.993426 pCL50_{calc} \quad (18)$$

S=0.234964

R-carré=97.1%

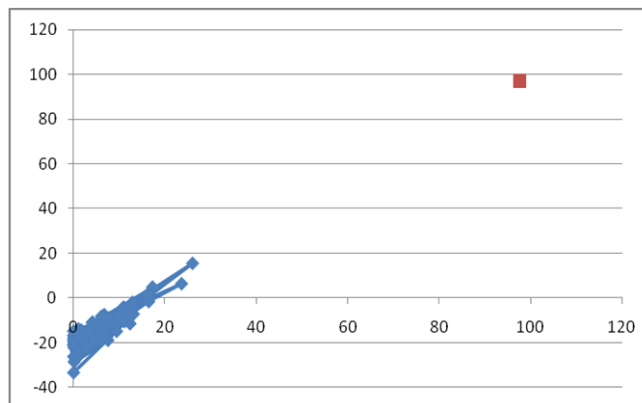
R-carré (ajust)=97.0%



**Figure 12:** Droite d'ajustement des  $pCL50$  observés en fonction des  $pCL50$  calculés.

**II-2-5 Test de randomisation :**

Afin de nous assurer que notre modèle n'est pas dû au hasard on a procédé au test de randomisation de  $y$ . La figure 12 fait apparaître clairement que les valeurs de  $R^2$ , pour les modèles dont on a modifié ‘**la concentration létale 50**’, sont très inférieures à celle du modèle réel, de plus on retrouve des valeurs négatives de  $Q^2$  pour les modèles randomisés. Ceci nous confirme que notre modèle relie réellement **la concentration létale 50** à la variable explicative choisie et qu'il n'est pas fortuit.



*Figure 13: Test de randomisation*

Dans ce travail on a essayé de modéliser en utilisant le méthodologie QSAR par une approche autre que logP (coefficient de partage octanol /eau )la concentration inhibitrice de la croissance **CIC50** et la concentration létale (**CL50**).

Les modèles obtenus par régression linéaire simple avec ADDD (descripteur géométrique) comme variable explicative pour les deux variables à expliquer **pCIC50** et **pCL50** sont comme le montre les statistiques des tableaux (1) et (2), très robustes et avec de bonnes capacités prédictives interne ( $Q^2_{LOO,LMO,BOOT}$ ), ils sont aussi caractérisés par un bon ajustement.

Dans le cas de la concentration inhibitrice, de la croissance, les composés externes (10 composés) on été prédits sans trop d'erreurs car on a une faible valeur du EQMP ext.

### Statistique des modèles obtenus

Tableau (1) :cas de **pCL50**

		Log p	ADDD
<b>R<sup>2</sup></b>	<b>1</b>	<b>96.26</b>	<b>97.64</b>
<b>R<sup>2</sup><sub>aju</sub></b>	<b>2</b>	<b>96.09</b>	<b>97.53</b>
<b>Q<sup>2</sup></b>	<b>3</b>	<b>95.5</b>	<b>97.09</b>
<b>Q<sup>2</sup><sub>lmo 10</sub></b>	<b>4</b>	<b>95.67</b>	<b>97.20</b>
<b>Q<sup>2</sup><sub>lmo 20</sub></b>	<b>5</b>	<b>96.00</b>	<b>97.39</b>
<b>Q<sup>2</sup><sub>lmo 30</sub></b>	<b>6</b>	<b>95.73</b>	<b>97.22</b>
<b>Q<sup>2</sup><sub>lmo 40</sub></b>	<b>7</b>	<b>96.26</b>	<b>97.57</b>
<b>Q<sup>2</sup><sub>lmo 50</sub></b>	<b>8</b>	<b>95.04</b>	<b>96.75</b>
<b>Q<sup>2</sup><sub>boot</sub></b>	<b>9</b>	<b>94.93</b>	<b>96.65</b>

Tableau (2) :cas de **pCIC50**

		Log p	ADDD
<b>R<sup>2</sup></b>	<b>1</b>	<b>96.62</b>	<b>97.11</b>
<b>R<sup>2</sup><sub>aju</sub></b>	<b>2</b>	<b>96.43</b>	<b>96.94</b>
<b>Q<sup>2</sup></b>	<b>3</b>	<b>95.73</b>	<b>96.24</b>
<b>Q<sup>2</sup><sub>lmo 10</sub></b>	<b>4</b>	<b>95.67</b>	<b>97.20</b>
<b>Q<sup>2</sup><sub>lmo 20</sub></b>	<b>5</b>	<b>96.00</b>	<b>97.40</b>
<b>Q<sup>2</sup><sub>lmo 30</sub></b>	<b>6</b>	<b>95.75</b>	<b>97.21</b>
<b>Q<sup>2</sup><sub>lmo 40</sub></b>	<b>7</b>	<b>96.26</b>	<b>97.56</b>
<b>Q<sup>2</sup><sub>lmo 50</sub></b>	<b>8</b>	<b>95.05</b>	<b>96.75</b>
<b>Q<sup>2</sup><sub>boot</sub></b>	<b>9</b>	<b>95.06</b>	<b>95.67</b>

Notre approche s'avère être meilleur pour les deux propriétés modélisées. la représentation graphique des résultats des tableaux ; figure (1) pour **pCL50** et figure (2) pour **pCIC50** abonde dans le sens de cette conclusion

### Comparaison graphique des modèles :

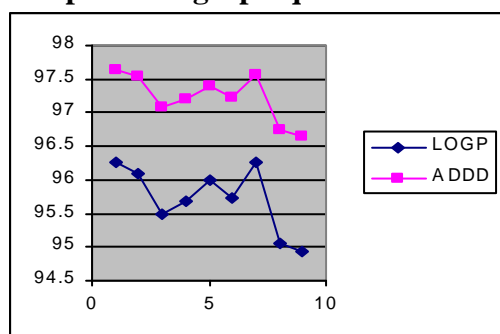


Figure (1) : Cas de **pCL50**

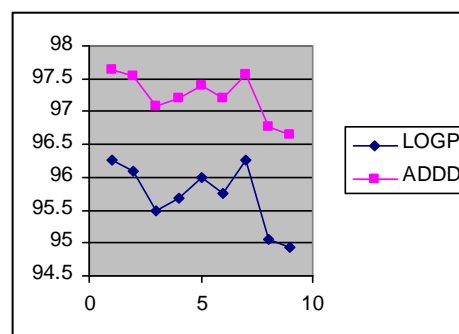


Figure (2) : cas de **pCIC50**

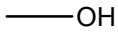
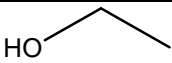
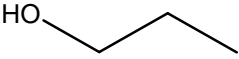
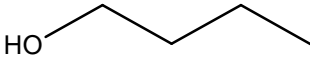
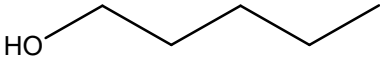
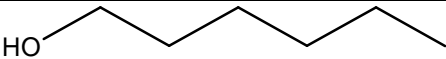
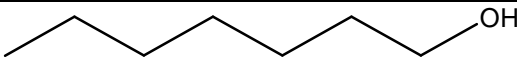
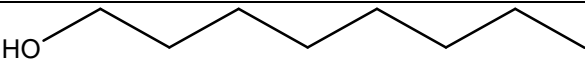
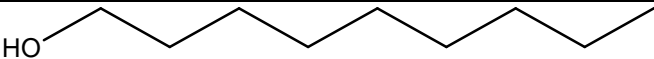
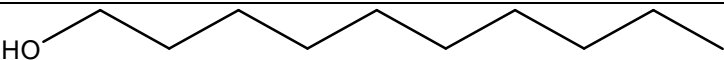
Notons ici que dans le cas de la concentration létale nous n'avons pas procédé à la validation statistique externe faute de données expérimentales suffisantes .



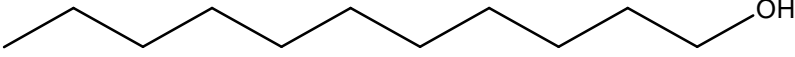
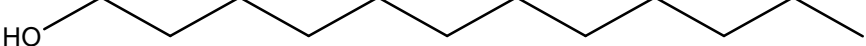

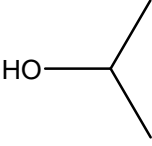
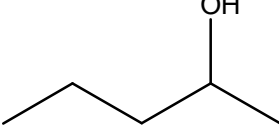
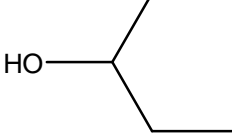
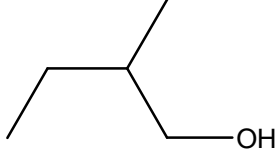
- [1] C. Boust, Institut national de recherche et sécurité .1<sup>ère</sup> édition avril 2004, page 1-6, [www.inrs.fr](http://www.inrs.fr).
- [2] R. R. Lauwerys, V. Haufroid, P. Huet, D. Lison Toxicologie industrielle et intoxications professionnelles, 2007, page 643-644.
- [3] les amines aliphatiques, Encyclopædia Universalis France S.A 2005 et 2009, <http://www.Encyclopædia.com>.
- [4] A. Carpy, Importance de la lipophilie en modélisation moléculaire, *analysis magazine*, 1999, 1, 18 et 21.
- [5] National Institute for Occupational Safety and Health, *RTECS (Registry of toxic effects of chemical substances)*, Hamilton, Ont.: Canadian Centre for Occupational Health and Safety (CD-ROM) <http://ccinfoweb.ccohs.ca/rtecs/search.html> .
- [6] G. Persoone, D. Dive, *Ecotox. Environ. Saf*, 1978, 2, 105-144.
- [7] Hyperchem<sup>TM</sup> Rebase 6.03 for Windows, Molecular Modeling system (2000)
- [8] R. Todeschini, V. Consonni, M. Pavan. DRAGON, Software for the Calculation of Molecular Descriptors. Release 5.3 for Windows, Milano (2005)
- [9] R. Todeschini, D. Ballabio, V. Consonni, A. Mauri, M. Pavan. MOBYDIGS Software for Multilinear Regression Analysis and variable Subset Selection by Genetic Algorithm. Release I.1 for Windows, Milano (2009)
- [10] D. M. Allen. The relationship between variable selection and data augmentation and a method for prediction. *Technometrics* 16: 125-127 (1974)
- [11] L. Erikson, J. Jaworska, A. Worth, M. Cromin, R.M. McDowell, P. Gramatica. Methods for reliability, uncertainty assessment, and applicability evaluations of regression based and classification QSPRs. *Environmental Health Perspective*. 111 (10) : 1361-1375 (2003)
- [12] \* B. Efron, R.J. Tibshirani. An introduction to the Bootstrap. Chapman & Hall (1993) - \*\* R. Wehrens, H. Putter, L.M.C. Buydens. The Bootstrap. A Tutorial. *Chem. Int. Lab. Syst.* 54 : 35-52 (2000)

- [13] R. Todeschini, V. Consonni. Handbook of Molecular Descriptors. R. Mannhold, H. Kubinyi, H. Timmerman eds. Wiley – VCH Verlag GmbH, Weinheim (2000).
- [14] M. Randic, A.F. Kleiner, L.M. DeAlba, *J.Chem.Inf.Comput.Sci.* 1994, 34, 277-286.
- [15] M. Randic, G. Krilov, *Int.J.Quant.Chem.* 1999, 75, 1017-1026.
- [16] H. Kubinyi, *QSAR: Hansch Analysis and Related Approaches*, VCH, Weinheim, 1993
- [17] Ghose, Pritchett and Crippen [*J. Comput. Chem.*, **9**, 80 (1988)]
- [18] V.N . Viswanadhan, A. K. Ghose, G. N. Revankar, R. K. Robins, [*J. Chem. Inf. Comput. Sci.*, 29, 163 (1989)]

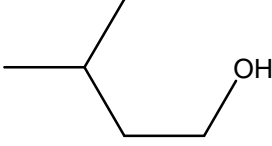
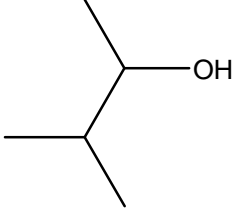
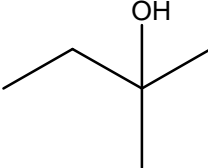
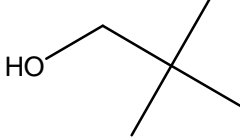
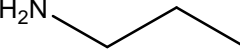
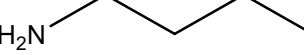
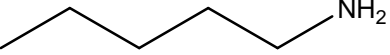
## Annexe1: Structures et numéro de CAS des composés utilisés (les alcools et les amines) .

N°	Composés (IUPAC)	Structure
1	Méthanol	
2	Ethanol	
3	propan-1-ol	
4	butan-1-ol	
5	pentan-1-ol	
6	hexan-1-ol	
7	heptan-1-ol	
8	octan-1-ol	
9	nonan-1-ol	
10	décan-1-ol	

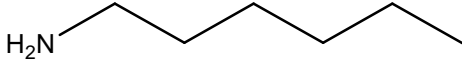
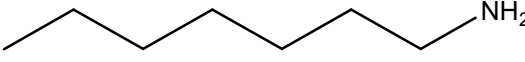
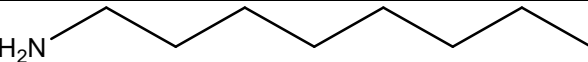
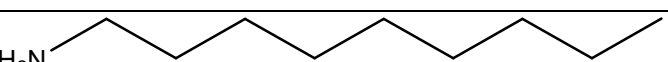
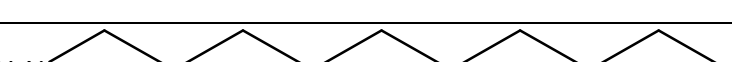
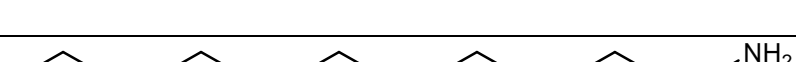
## Annexe1 :suite

N°	Composés (IUPAC)	Structure
11	undécan-1-ol	
12	dodecan-1-ol	
13	tridecan-1-ol	
14	propan-2-ol	
15	pentan-2-ol	
16	pentan-3-ol	
17	2-méthyl-1-butanol	

## Annexe1 :suite

N°	Composés (IUPAC)	Structure
18	3-méthyl-1-butanol	
19	3-méthyl-2-butanol	
20	(tert) pentanol	
21	(neo) pentanol	
22	1-propylamine	
23	1-butylamine	
24	1-amylamine	

## Annexe1 :suite

N°	Composés (IUPAC)	Structure
25	1-hexylamine	
26	1-heptylamine	
27	1-octylamine	
28	1-nonylamine	
29	1-décylamine	
30	1-undecylamine	

**Tableau I-a** : Toxicité relative, hydrophobicité et valeurs de descripteur pour les alcools aliphatiques et des amines.

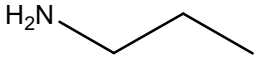
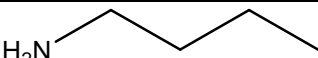
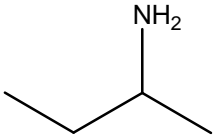
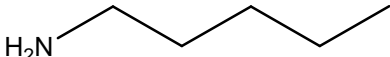
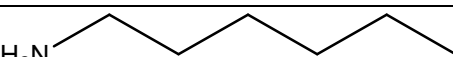
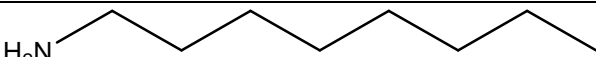
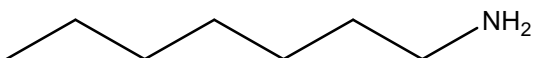
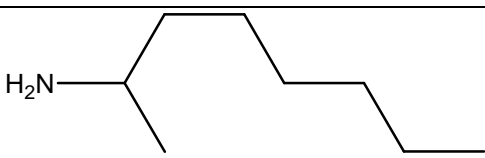
N°	Composés	pCIC50	ADDD	Log P
1	Méthanol	-2.77	4.964	-0.77
2	Ethanol	-2.41	8.108	-0.31
3	propan-1-ol	-1.84	11.098	0.25
4	butan-1-ol	-1.52	14.138	0.88
5	pentan-1-ol	-1.12	17.194	1.56
6	hexan-1-ol	-0.47	20.305	2.03
7	heptan-1-ol	0.02	23.451	2.57
8	octan-1-ol	0.5	26.640	3.15
9	nonan-1-ol	0.77	29.859	3.69
10	decan-1-ol	1.1	33.118	4.23
11	undecan-1-ol	1.87	36.401	4.77
12	dodecanol	2.07	39.716	5.13
13	tridecan-1-ol	2.28	43.053	5.67
14	propan-2-ol	-1.99	10.987	0.05
15	pentan-2-ol	-1.25	16.888	1.21
16	pentan-3-ol	-1.33	16.908	1.21
17	2-méthyl-1-butanol	-1.13	16.810	1.42
18	3-méthyl-1-butanol	-1.13	16.833	1.42
19	3-méthyl-2-butanol	-1.08	16.622	1.28
20	(tert) pentanol	-1.27	16.640	1.21
21	(neo) pentanol	-0.96	16.710	1.57
22	1-propylamine	-0.85	11.968	0.48
23	1-butylamine	-0.7	15.015	0.97

Tableau I-a :suite

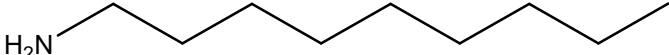
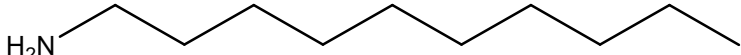
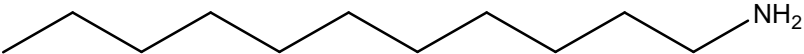
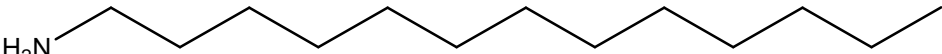
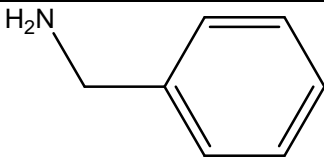
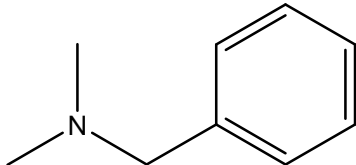
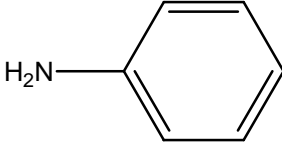
24	1-amylamine	-0.61	18.101	1.49
25	1-hexylamine	-0.34	21.226	2.06
26	1-heptylamine	0.1	24.393	2.57
27	1-octylamine	0.51	27.602	3.04
28	1-nonylamine	1.59	30.837	3.57
29	1-décylamine	1.95	34.110	4.10
30	1-undécylamine	2.26	37.408	4.63



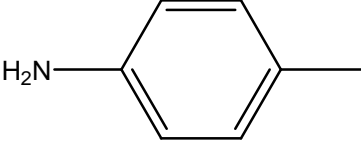
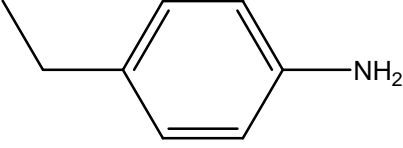
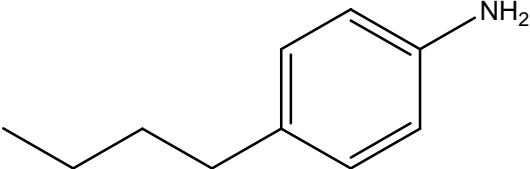
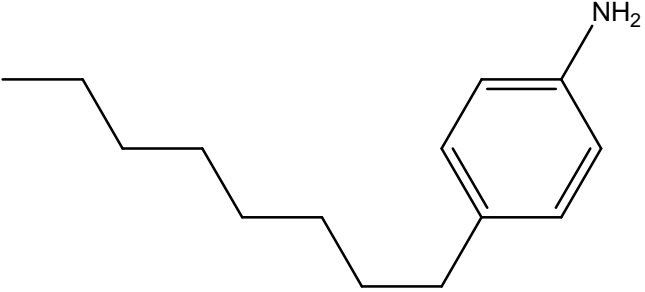
## Annexe2: Structures et numéro de CAS des composés utilisés (les amines).

N°	Composés (IUPAC)	Structure
1	Propylamine	 Propylamine
2	Butylamine	 Butylamine
3	sec-butylamine	 sec-butylamine
4	Amylamine	 Amylamine
5	Hexylamine	 Hexylamine
6	Octylamine	 Octylamine
7	Heptylamine	 Heptylamine
8	1-Méthylheptylamine	 1-Méthylheptylamine

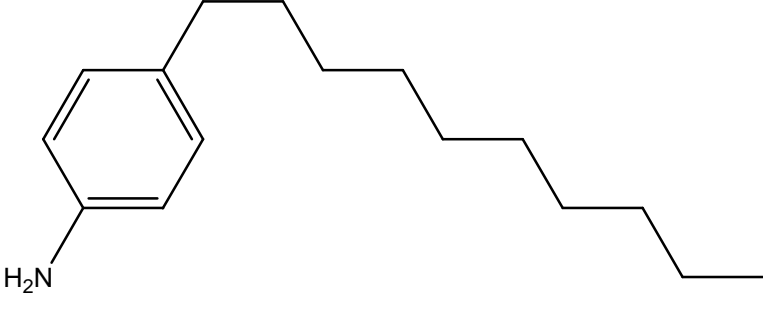
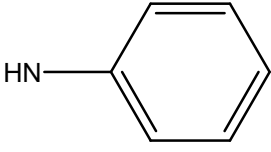
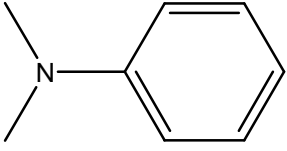
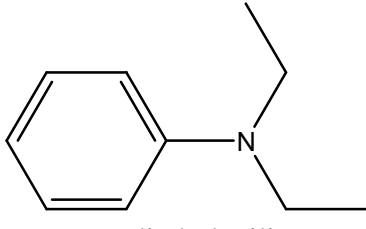
## Annexe2 :suite

N°	Composés (IUPAC)	Structure
9	Nonylamine	 <p>Nonylamine</p>
10	Decylamine	 <p>Decylamine</p>
11	Undecylamine	 <p>Undecylamine</p>
12	Tridecylamine	 <p>Tridecylamine</p>
13	Benzylamine	 <p>Benzylamine</p>
14	N,N-Diméthylbenzylamine	 <p>N,N-Dimethylbenzylamine</p>
15	Aniline	 <p>Aniline</p>

## Annexe2 :suite

N°	Composés (IUPAC)	Structure
16	4-Toluidine	 <p>4-Toluidine</p>
17	4-Ethylaniline	 <p>4-Ethylaniline</p>
18	4-Butylaniline	 <p>4-Butylaniline</p>
19	4-Octylaniline	 <p>4-Octylaniline</p>

## Annexe2 :suite

N°	Composés (IUPAC)	Structure
20	4-Décylaniline	 <p><chem>Nc1ccc(cc1)CCCCCCCCC</chem></p>
21	N-Méthylaniline	 <p><chem>CNc1ccccc1</chem> N-Methylaniline</p>
22	N-N-diméthylaniline	 <p><chem>CN(C)c1ccccc1</chem> N-N-dimethylaniline</p>
23	N,N-diéthylaniline	 <p><chem>CCN(CC)c1ccccc1</chem> N,N-diethylaniline</p>

**Tableau I1-1** –Toxicité (CL50) et valeurs de descripteur pour les amines.

N°	Composés	pCL50	ADDD	Log P
1	Propylamine	-0.717	11.968	0.48
2	Butylamine	-0.564	15.015	0.97
3	Sec-butylamine	-0.575	14.757	0.70
4	Amylamine	-0.308	18.101	1.49
5	Hexylamine	0.252	21.226	2.06
6	Octylamine	1.396	27.602	3.05
7	Heptylamine	0.723	24.393	2.57
8	1-Méthylheptylamine	1.403	27.042	2.82
9	Nonylamine	1.822	30.837	3.57
10	Decylamine	2.184	34.110	4.10
11	Undecylamine	2.912	37.408	4.63
12	Tridecylamine	3.487	44.078	5.68
13	Benzylamine	0.021	16.387	1.09
14	N,N-Diméthylbenzylamine	0.554	21.555	1.95
15	Aniline	-0.088	13.496	0.90
16	4-Toluidine	-0.143	16.471	1.39
17	4-Ethylaniline	0.220	19.166	1.96
18	4-Butylaniline	1.165	25.152	3.15
19	4-Octylaniline	3.233	37.873	5.27
20	4-Décylaniline	3.576	44.484	6.32
21	N-Méthylaniline	0.030	16.292	1.64
22	N-N-diméthylaniline	0.190	19.052	2.31
23	N,N-diéthylaniline	0.959	24.720	3.31