



République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur
et de la Recherche Scientifique



Faculté de la Technologie

Mémoire de Fin d'Étude

En vue de l'obtention du diplôme de

MASTER ACADEMIQUE

Domaine : Technologies

Filière : Génie des Procédés

Spécialité : Génie Chimique

Présenté par :

Yagoub Youcef

Lakrouche Mohammed Seghir

Neguia Seif Eddine

Thème

Application de la méthodologie QSAR à une série des
composés organiques

soutenue le

17/5/2021

Devant le Jury :

Dr : Sarouti Abdelghani

Président

Université d'El Oued.

Dr: Zegoud Laid

Examineur

Université d'El Oued.

Dr : Menaceur Souheila

Rapporteur

Université d'El Oued

2020/2021

Remerciements

En premier lieu, nous remercions le bon Dieu de nous avoir donné la force et la patience nécessaire pour achever ce travail de mémoire.

Au terme de ce travail nous tenons à remercier en premier lieu, notre promiseuse pour son aide précieuse, pour ses remarques pertinentes et surtout pour le temps qu'elle nous a accordée et ceci malgré ses nombreuses charges.

Nous tenons à remercier les membres de jury qui ont bien voulu Nous faire l'honneur de participer à ce jury.

Nous remercier également tous les enseignants du département Génie des Procédés pour leur contribution à cette formation.

Nous remercier pour tout a fait Merci à tous ceux qui nous a contribués de près ou de loin à l'élaboration de ce mémoire.

Dédicaces

À mes chers parents, que Dieu vous garde auprès de moi.

À mes chers frères et sœurs.

À mes fidèles amis.

Yagoub Youcef

À mes chers parents, que Dieu vous garde auprès de moi.

À mes chers frères et sœurs.

À mes fidèles amis.

Lakrouche Mohammed Seghir

À mes chers parents, que Dieu vous garde auprès de moi.

À mes chers frères et sœurs.

À mes fidèles amis.

Neguia Seif Eddine

Résumé

Dans ce travail, nous avons obtenu une relation mathématique entre la structure des composés aniliques exprimée par des propriétés chimiques et la caractéristique d'inhibition de la croissance des cellules logiGc50 pour obtenir une équation de prédiction de cette propriété. Cette étude est réalisée sur 48 composés aniliques en utilisant la technique QSAR et la méthode de régression linéaire multiple (MLR) et utilise programme minitab auxquels les resultats étaient comme suit :

Les resultats obtenus ont donné lieu à deux modèles :

- ✓ Le modèle (1) est construit avec un seul descripteur et le modèle (2) avec deux descripteurs ensemble.

Modèle (1) : ($R=0,5988$) ($S=0,4214$) ($F=102,5$).

Modèle (2) : ($R=0.8903$)($S=0.222760$)($F=182.66$).

Aussi, nous avons conclu que la propriété d'inhibition de la croissance des cellules logiGc50 varie proportionnellement avec le coefficient de partage $\log(P)$.

Mots-clés — QSAR, logiGc50, équation de régression linéaire (MLR), $\log(P)$, composés organiques

ملخص

في هذا العمل، حصلنا على علاقة رياضية بين بنية المركبات الأنيلية المعبر عنها بالخصائص الكيميائية وخاصة تثبيط نمو الخلايا logiGc50 للحصول على معادلة تنبؤ لهذه المركبات. حيث ان النموذج الاول يبنى بخاصية واحدة والنموذج الثاني بخاصيتان معًا. تم إجراء هذه التجربة على مركبات الأنيلية (48 مركبًا) باستخدام تقنية QSAR وطريقة الانحدار الخطي المتعدد (MLR) واستخدام برنامج. mintab حيث كانت النتائج المتحصل عليها كالآتي:

النموذج الاول : ($R = 0,5988$) ($F = 102.5$) ($S = 0.4214$)

النموذج الثاني : ($R = 0.8903$) ($S = 0.222760$) ($F = 182.66$)

بالإضافة إلى دور معامل التقسيم اللوغاريتمي (P) في هذه الدراسة، فإن زيادة هذا المعامل يزيد من السمية وخطرها على الخلايا.

الكلمات المفتاحية – QASR ; logiGc50 , معادلة الانحدار الخطي (MLR) ، $\log(P)$, المركبات العضوية.

Table des matières

Remerciements.....	I
Dédicaces	II
Résumé	III
ملخص.....	IV
Table des matières	V
Liste des figures.....	VIII
Liste des tableaux	IX
Liste des Abréviations.....	V
Introduction générale.....	1
Chapitre I : Méthode QSAR	3
I.1 Introduction	3
I.2 Aperçu historique.....	3
I.3 Définition	4
I.4 Principe	4
I.5 Objectifs	5
I.6 Méthodologie générale d'une étude QSAR	6
I.6.1 Elaborer le modèle.....	6
I.6.2 Validation du modèle.....	9
I.7 Régression linéaire Multiple (MLR).....	9
I.7.1 Estimation des paramètres statistiques du modèle	9
I.7.2 Ecart type des coefficients de régression.....	12
I.7.3 Confiance à accorder aux résultats.....	12
I.7.4 Indicateurs de qualité d'une régression linéaire	14
I.7.5 Critères de comparaison de modèle.....	15
I.8 Applications des méthodes QSAR/QSPR.....	16
I.9 Logiciels utilisés dans nos études RQSA.....	16
Chapitre II : Aniline	Error! Bookmark not defined.
II.1 Introduction	18
II.1.1 Historique :	19

II.1.2	Famille Alanaline :	19
II.1.3	Buts	Error! Bookmark not defined.
II.1.4	EFFETS SUR La santé	22
II.2	Caractéristiques	23
II.2.1	Utilisations	23
II.2.2	Propriétés physiques.....	23
II.2.3	Propriétés chimiques	24
II.2.4	Récipient de stockage	25
II.3	La toxicologie	25
II.3.1	Notions et définitions	26
II.3.2	Toxicité des produits chimiques organiques.....	27
II.3.3	Les approches QSAR pour l'étude de la toxicité	28
Chapitre III	: Application sur série d'aniline	30
III.1	Introduction	30
III.2	Méthodologie.....	30
III.2.1	Base de données:	30
III.2.2	II.2Pourquoi nous intéressons-nous aux composés organiques (composés carbonés)? ...	31
III.2.3	Calcul des propriétés.....	33
III.2.4	Analyse statistique	35
III.3	Les résultats de la première étape (Modèles QSAR à un seul paramètre)	38
III.3.1	Analyse des résultats de la première phase	40
III.3.2	Analyse des résidus Les résultats de la première étape	40
III.3.3	Discutez des résultats de la première étape.....	40
III.4	Les résultats de la deuxième étape (Modèles QSAR à un deux paramètres)	41
III.4.1	Analyse des résultats de la deuxième étape.....	44
III.4.2	Résultats de l'analyse des résidus de la deuxième étape	44
III.4.3	Discutez des résultats de la deuxième étape.....	44
III.5	Validation les résultats.....	45
III.5.1	Le premier modèle (MW).....	45
III.6	Coefficient de partage log (P)	48
III.6.1	Analyse des résultats	48
III.6.2	Discuter des résultats.....	49
Conclusion générale		50

Bibliographie	51
----------------------------	-----------

Liste des figures

Figure I.1 : Modèle de l'étude QSAR	5
Figure II.1 : Formule de l'aniline.....	19
Figure II.2 : La préparation de 2,4,6-trichloroaniline	20
Figure II.3 : 2-chloro aniline	20
Figure II.4: Le milieu et les différents éléments pouvant affecter l'organisme vivant	26
Figure II.5 : Les différentes espèces aquatiques utilisées pour l'étude de la toxicité aquatique ..	28
Figure III.1 : Forme générale d'aniline	31
Figure III.2 : Programme ChemOffice Ultra	33
Figure III.3 : Modèle phénolique utilisant hyperchem	34
Figure III.4 : Programme DRAGON	34
Figure III.5 : Correlation entre les valeurs.....	35
Figure III.6 : Courbe graphique de l'équation de régression linéaire.....	38
Figure III.7 : Courbe graphique Le pourcentage d'erreur pour chaque point	39
Figure III.8 : Courbe graphique L'erreur en fonction de la valeur révisée	39
Figure III.9 : Courbe graphique Le pourcentage d'erreur pour chaque point (deuxième étape)...	42
Figure III.10 : Courbe graphique L'erreur en fonction de la valeur révisée (deuxième étape).....	43
Figure III.11 : Courbe graphique Le pourcentage d'erreur pour chaque point (deuxième étape).	43

Liste des tableaux

Tableau I.1 : Exemples de descripteurs électroniques de la chimie quantique.....	7
Tableau II.1: À 25 °C et 101 kPa, 1 ppm = 3,87 mg/m ³	24
Tableau II.2: Données de toxicité pour les produits chimiques de haute production	27
Tableau III.1 : Un ensemble de propriétés pour les composés aniliques.....	31
Tableau III.2 : Le degré de force de la relation du coefficient de corrélation	37
Tableau III.3 : ajustements et diagnostics pour les observations inhabituelles	39
Tableau III.4 : Tableau des valeurs des composés moins liés au modèle.....	40
Tableau III.5 : Tableau des valeurs des composés plus liés au modèle.....	41
Tableau III.6 : Résultats des caractéristiques étudiées (deuxième étape)	41
Tableau III.7 : ajustements et diagnostics pour les observations inhabituelles (deuxième étape)	43
Tableau III.8 : Tableau des valeurs des composés moins liés au modèle (deuxième étape)	44
Tableau III.9 : Tableau des valeurs des composés plus liés au modèle (deuxième étape)	45
Tableau III.10 : Valeurs des paramètres pour la vérification extérieure de MW . Error! Bookmark not defined.	
Tableau III.11 : Valeurs des paramètres pour la vérification extérieure de (DP03, R3v+) ... Error! Bookmark not defined.	
Tableau III.12 : Tableau des résultats pour coefficient de partage log (P) .. Error! Bookmark not defined.	

Liste des Abréviations

DP02 profil moléculaire no. 02 (Le profil moléculaire rance DPk)

DP03 profil moléculaire no. 03 (Le profil moléculaire rance DPk)

HOMO Highest Occupied Molecular Orbital

LMO Leave Many Out

LOO Leave One Out

MLR Multiple Linear Regression

Mp Melting point .Point de fusion

Mv grammes de soluté par mL de solution x 100 et cela serait égal à grammes de soluté par 100 mL de solution.

MV Volume Moléculaire.

MW Masse moléculaire

nF n = nombre de moles d'électrons F = constante de Faraday, 96500 coulombs / mol

QSAR Quantitative Structure-Activity Relationships

QSPR Quantitative Structure-Property Relationships

R Coefficient de Corrélation.

R_{adj} Coefficient de Détermination Ajusté.

RDF080m Fonction de distribution radiale - 8,5 / pondérée par les masses atomiques

R3v+ Autocorrélation maximale R3v + R du décalage 3 / pondérée par le volume de van der Waals

SP05 Hybridation sp⁵» signifie que les orbitales atomiques que les atomes de carbone dans le cyclopropane utilisent pour se lier les uns aux autres sont 1/6 s et 5/6 p. Il y en a deux, donc ensemble, ils utilisent 1/3 de l'orbitale s et 5/3 des orbitales p (sur 3).

SLR Simple Linear Regression

Ss Solid Solution

SSE Sum of Squares Error

SSR Sum of Squares Regression

Introduction générale

Le développement technologique et industriel rapide au cours des dernières décennies vise à accroître le bien-être des humanités. Cependant, elle n'est pas sans conséquences pour la santé humaine et l'environnement. C'est à cause des produits chimiques différents types de développement, ce qui provoque la fuite de substances toxiques dans l'environnement.

Aujourd'hui, l'intérêt des chercheurs dans le domaine de la chimie s'est accru pour l'étude de la propriété chimique du facteur de toxicité (inhibant la croissance cellulaire logiGc50), qui est l'effet de certaines substances sur les cellules, car ces substances agissent pour arrêter la croissance des cellules. L'un des domaines d'étude les plus importants pour ces laboratoires est l'industrie pharmaceutique et des pesticides.

Les programmes de modélisation moléculaire sont parmi les programmes les plus importants qui contribuent à l'imagerie du composé chimique sous la forme de tridimensionnalité et donne les propriétés des molécules, et dans cette étude, nous avons utilisé un ensemble de logiciels, qui est tridimensionnalité et nous donne les propriétés des molécules, et dans cette étude, nous avons utilisé un ensemble de logiciels, qui est (HyperChem) et (Dragon) pour la modélisation des molécules et le calcul des spécifications, et le programme mathématique (minitab) .

Le but de ce travail est d'obtenir un modèle mathématique qui explique la toxicité des composés étudiés en fonction de leurs propriétés chimiques. Ce modèle nous permet de calculer la toxicité de composés inconnus sans avoir besoin de tests en laboratoire

Cette étude a été réalisée sur une étape théorique composée de deux chapitres. Le premier chapitre traite des généralités de la méthode QSAR et le deuxième chapitre nous avons parlé des composés organiques, notamment les composés aniliques, qui sont à la base de l'expérience , Et la partie pratique dans l'aspect comment travailler et obtenir des résultats, puis analyser et discuter les résultats et validation. et l'impact du coefficient de partage log(P) et son rôle dans l'augmentation du risque de toxicité .

Chapitre I

Méthode QSAR

I.1 Introduction

La connaissance des activités et des propriétés est d'une importance essentielle pour pouvoir utiliser et classer les composés chimiques. La caractérisation expérimentale complète est difficile, voire impossible, pour des raisons de coût, de temps, de dangerosité de certains essais ou d'éthique (limitations des essais sur les animaux) [1]. L'utilisation des méthodes alternatives à l'expérience est devenue plus qu'indispensable. Parmi ces méthodes, on trouve les méthodes de modélisation moléculaire qui permettent de justifier les données expérimentales disponibles et prédire les propriétés/activités pour des composés nouveaux ou des composés pour lesquels les données expérimentales ne sont pas disponibles. Parmi ces méthodes de modélisation les plus utilisées, on peut citer les méthodes QSAR (Quantitative Structure-Activity Relationship) et QSPR (Quantitative Structure-Property Relationships). Ces méthodes s'appuient sur le principe que les propriétés physico-chimiques et les activités biologiques des molécules dépendent fortement de leurs structures chimiques [2].

I.2 Aperçu historique

En 1868, Crum-Brown et Fraser ont publié une équation qui est considérable pour être la première formulation d'une relation quantitative structure-activité, dans leurs recherches sur différents alcaloïdes. Par conséquent, ils ont supposé que « l'activité physiologique » φ devait être fonction de la structure chimique C (Eq. 1) [3]

$$\varphi = f(C) \quad (1)$$

Richet a découvert en 1893 que la toxicité des composés organiques suit inversement leur solubilité dans l'eau (il a conclu "plus ils sont solubles, moins ils sont toxiques") [4]. Une telle relation correspond à l'équation (Eq. 2), où $\Delta\varphi$ traduit les différences dans les valeurs d'activité biologique, causées par des changements correspondants dans les propriétés chimiques et surtout physicochimiques : ΔC .

$$\Delta\varphi = f(\Delta C) \quad (2)$$

Le développement de la technique QSAR a subi plusieurs transformations mais l'année 1964 est considérée comme le début des méthodes QSAR modernes où Hansch et Fujita ont établi les premières corrélations entre les propriétés physico-chimiques (log P, pKa, paramètres stériques et électroniques) et l'activité biologique (activité enzymatique, pharmacologique) [5].

La méthodologie QSAR s'est rapidement développée à partir du 1988, en raison de l'introduction du paramètre moléculaire tridimensionnel, qui expliquaient l'influence de différents conformères, stéréoisomères ou énantiomères (modèles 3D QSAR). Le premier modèle publié possédant ces caractéristiques était l'analyse comparative du champ moléculaire (CoMFA "Comparative Molecular Field Analysis") proposé par Cramer et al, qui est actuellement l'une des méthodologies QSAR les plus utilisées. [6].

I.3 Définition

Par définition, QSAR (relation quantitative structure-activité) est une corrélation mathématique entre une activité biologique ou une propriété moléculaire spécifiée et une ou plusieurs propriétés structurelles physicochimiques et/ou moléculaires, appelées descripteurs puisqu'elles "décrivent" l'activité ou la propriété examinée [7]. Le terme générique QSAR peut être exprimé comme :

$$\text{Activité biologique} = f(\text{paramètres})$$

I.4 Principe

Le principe d'une étude QSAR (Fig I.1), consiste à trouver une relation mathématique reliant de manière quantitative une activité biologique, mesurée pour une série de composés similaires dans les mêmes conditions expérimentales, avec des descripteurs moléculaires à l'aide des méthodes statistiques. Ces études permettent d'analyser les données structurales afin de détecter les facteurs déterminants pour l'activité étudiée. Pour ce faire, différents types de méthodes statistiques peuvent être employées, les plus répandues sont : la méthode de moindres carrées partielles, la régression linéaire simple et multiple et les réseaux de neurones artificiels. La modélisation QSAR comporte généralement trois étapes [8]:

- Recueillir ou, si possible, concevoir un ensemble d'information sur la structure et les propriétés des produits chimiques.
- Choisir des descripteurs capables de relier correctement la structure chimique à l'activité biologique.
- Appliquer des méthodes statistiques qui corrélient les changements de structure avec les changements dans l'activité biologique.

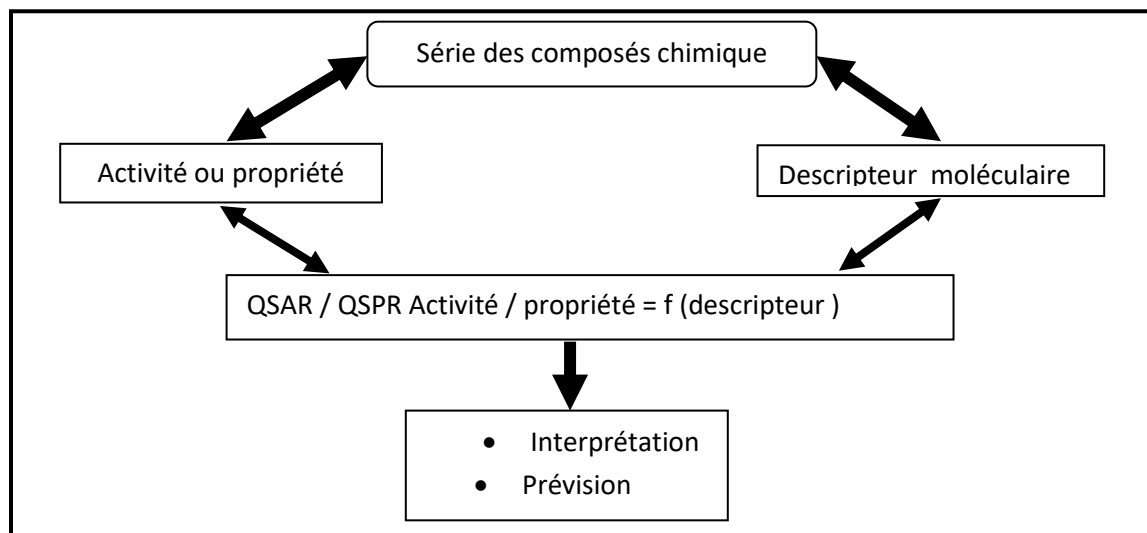


Figure I.1 : Modèle de l'étude QSAR

L'expression mathématique obtenue peut alors être utilisée comme moyen prédictif de l'activité étudiée pour de nouvelles molécules ou des molécules pour lesquelles les données expérimentales ne sont pas disponibles

I.5 Objectifs

Les objectifs pratiques d'une étude QSAR sont nombreux et ces techniques sont largement utilisées dans de nombreuses situations. Le but des études QSAR comprend donc ce qui suit [9] :

- Prédire l'activité biologique et les propriétés physico-chimiques par des moyens rationnels.
- Comprendre les mécanismes d'action au sein d'une série de produits chimiques.

- Économies sur les coûts de développement de produits.
- Les prévisions pourraient réduire la nécessité de tests sur les animaux longs et coûteux.

I.6 Méthodologie générale d'une étude QSAR

Pour aboutir à un modèle QSAR explicatif et prédictif on doit [10] :

- Constituer une base de données à partir des mesures expérimentales fiables de l'activité de chaque composé.
- Sélectionner les descripteurs en relation avec l'activité étudiée.
- Diviser cette base de données, aléatoirement, en une série d'apprentissage (Trai-ning set) qui contient généralement les 2/3 de la base de données et une série de test (Test set) constituée par le 1/3 restant.
- Etablir des modèles mathématiques en utilisant la série d'apprentissage.
- Valider les modèles élaborés en utilisant la série de test et calculer leurs paramètres statistiques de validation externe.
- Elaborer le domaine d'applicabilité du modèle retenu.

On résume ces opérations dans trois étapes : Elaborer le modèle, Validation du modèle et Détermination du domaine d'applicabilité du modèle.

I.6.1 Elaborer le modèle

I.6.1.1 Base de données

Le choix de la base de données expérimentale initiale est une étape critique pour le développement des modèles QSAR, les données expérimentales doivent être obtenues en suivant le même protocole [11].

I.6.1.2 Descripteur moléculaire:

Un descripteur moléculaire est toute propriété (une valeur numérique) propre à une structure chimique donnée. Ces valeurs peuvent être obtenues expérimentalement ou calculées à partir de la structure de la molécule. Les descripteurs moléculaires jouent un rôle fondamental dans les études de la relation quantitative structure activité. Ils sont utilisés en tant que variables indépendantes pour prédire une variable dépendante (activité) [12].

Historiquement, deux grands schémas pour la classification des descripteurs moléculaires ont été établis : l'un en fonction de leur origine (constitutionnel, topologique, géométrique, quantique, thermodynamique...), et un autre sur leur dimensionnalité (1D, 2D, 3D) [13].

I.6.1.2.1 Descripteurs théoriques

Définition d'un descripteur : c'est l'information contenue dans la structure d'une molécule.

Les descripteurs 1D : sont remportés directement à partir de la formule brute de la molécule. Exemples : le nombre d'atomes et la masse moléculaire,...etc

Les descripteurs 2D : la formule développée d'une molécule peut nous offrir ce type de descripteurs. On distingue :

- *Les indices 2D constitutionnels* : ce type d'indices caractérise les différents composants de la molécule. Il s'agit par exemple du nombre de liaisons simples ou multiples, du nombre de cycles...
- *Les indices 2D topologiques* : Ils donnent des informations sur la taille, la forme globale et les ramifications d'une structure. Exemples: indice de Wiener [14], indice de Randić [15], indice de connectivité de valence de Kier-Hall [16], indice de Balaban [17], ...etc

Les descripteurs 3D : sont estimés à partir des positions relatives des atomes dans l'espace, et décrivent des caractéristiques plus complexes. Leurs calculs nécessitent donc de connaître la géométrie 3D de la molécule.

- *Les descripteurs 3D géométriques* : les plus importants sont le volume moléculaire, la surface accessible au solvant, le moment principal d'inertie.
- *Les descripteurs 3D électroniques* : permettent de quantifier les différents types d'interactions inter- et intramoléculaires, de grande influence sur l'activité biologique des molécules. Le calcul de la plupart de ces descripteurs nécessite la recherche de la géométrie pour laquelle l'énergie est minimale, et fait souvent appel à la chimie quantique.

Exemples:

Tableau I.1 : Exemples de descripteurs électroniques de la chimie quantique

Descripteur	Définition
-------------	------------

ELUMO	Energie de l'orbitale la plus basse vacante
EHOM	Energie de l'orbitale la plus haute occupée
μ	Potentiel chimique
η	Dureté chimique
σ	Mollesse chimique
ω	Indice d'électrophile globale
AEI	Indice de l'énergie d'activation
MSD	Densité de Spin de Mulliken
ESOMO	Energie de l'orbitale occupée par un électron

I.6.1.2.2 Descripteurs thermodynamiques

Les descripteurs thermodynamiques sont calculés sur la base de la fonction de partition totale Q de la molécule. Exemples : l'enthalpie de formation, l'enthalpie libre...

I.6.1.2.3 Descripteurs empiriques (physico-chimique)

Les descripteurs physicochimiques, (ou indices physicochimiques) certains d'entre eux reflètent la composition moléculaire du composé (le nombre et le type d'atomes et de liaisons présents dans la molécule, le nombre de cycle, les propriétés donneur/accepteur de liaison H, cation, anion, etc....) [18]. D'autres représentent le caractère hydrophile ou lipophile de la molécule généralement évalué à partir du coefficient de partage Octanol/eau représenté par le log P [19].

- **Coefficient de partage** : Le logP, aussi appelé log Kow, est une mesure de la solubilité différentielle de composés chimiques dans deux solvants (coefficient de partage octanol/eau). Le logP est égal au logarithme du rapport des concentrations de la substance étudiée dans l'octanol et dans l'eau. $\log P = \log(C_{\text{oct}}/C_{\text{eau}})$ Le logP rend compte de la tendance d'une molécule à s'accumuler dans les membranes biologiques des organismes vivants. S'il est élevé, le risque de bioaccumulation est important. Le partage d'une molécule entre une phase aqueuse et une phase lipidique conditionne en partie ses propriétés biologiques telles que le transport, le passage
- **Mesure et estimation de log p** : La méthode la plus classique et la plus fiable des méthodes de détermination du log P est la méthode dite du « flacon agité » [20]. Ce descripteur peut être obtenu également par les méthodes d'estimation en l'occurrence : méthode de Hansch [21], méthode Rekker [22], méthode de Ghose et Viswanathan [23].

Plusieurs logiciels sont utilisés pour l'estimation de logP tels qu'ACD, chemdraw, Chemaxon, Chempider (online), Hyperchem

I.6.1.3 Méthodes d'analyse des données

Une méthode d'analyse des données est nécessaire pour élaborer un modèle QSPR/QSAR. Il existe plusieurs méthodes pour construire un modèle et analyser les données statistiques de ce dernier, ces méthodes sont disponibles dans des logiciels tels que, Excel, Origin Microcal, Minitab, Statistica, SPSS, R,...

I.6.1.4 Régression linéaire Multiple (MLR)

Elle repose sur l'hypothèse qu'il existe une relation linéaire entre une variable dépendante (à expliquer) Y (ici, l'activité) et une série de p variables indépendantes (explicatives) Xi (ici, les descripteurs).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon \quad (3)$$

Cette équation est linéaire par rapport aux paramètres (coefficients de régression) $\beta_0, \beta_1, \dots, \beta_p$. La détermination de l'équation (3) se fait alors à partir d'une base de données de n échantillons (observations) pour laquelle à la fois les variables indépendantes et la variable dépendante sont connues.

Il s'agit donc de considérer un système d'équations qui peut être donné sous la forme matricielle suivante [30]:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \dots \\ Y_n \end{pmatrix} X = \begin{pmatrix} 1 & X_{11} & \dots & X_{1p} \\ 1 & & & \\ 1 & & & \\ \dots & & & \\ 1 & X_{n1} & \dots & X_{np} \end{pmatrix} \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \dots \\ \beta_p \end{pmatrix} \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \dots \\ \varepsilon_n \end{pmatrix}$$

I.7 Validation du modèle

Une fois développé, le modèle doit être interprété en analysant tous les paramètres statistiques de ce modèle (Chapitre III), sa qualité doit être aussi étudiée, cette qualité est vérifiée

par ce que l'on appelle validation. Sa robustesse, c'est-à-dire l'influence des composés de la série d'apprentissage sur le modèle, est estimée par des méthodes de validation interne. Afin d'estimer son pouvoir prédictif, des données expérimentales supplémentaires sont nécessaires afin de déterminer la capacité du modèle à prédire ces valeurs c'est ce que l'on appelle validation externe. Enfin, il est important de savoir quel type de molécules utilisées avec quel modèle. On parle alors de domaine d'applicabilité.

I.7.1 Validation interne

Dans le passé, la validation interne d'un modèle QSPR/QSAR a été réalisée en utilisant la validation croisée LOO (leave-one out) ou LMO (leave-many out) qui est quantifiée par le coefficient R_{2cv} . Ce processus consiste à extraire un certain nombre k de molécules du jeu initial à N molécules et à construire un nouveau modèle avec les $(N-k)$ molécules restantes à l'aide des descripteurs choisis (seules les constantes de la régression changent). Ce processus est ensuite réitéré pour retirer et prédire les valeurs de toutes les molécules de la série d'apprentissage. En fonction du nombre de molécules retirées à chaque itération, on parlera de Leave-One-Out (LOO) ou de Leave-Many-Out (LMO) selon qu'une ou plusieurs molécules est (sont) retirée(s) [24]. Dans ces dernières années, d'autres méthodes sont utilisées pour faire la validation interne, tel que la hasardisation de la réponse (Y-Randomization).

- **Y-Randomisation :** Afin de s'assurer qu'un modèle QSAR est fiable, les tests d'Y-Randomization [25] sont une des techniques les plus employées. En effet, il n'est pas rare d'obtenir des corrélations fortuites (ou « chance corrélation »), c'est-à-dire un modèle affichant de bons résultats statistiques (R_2 , SD) pour l'apprentissage, mais impliquant des descripteurs qui ne sont pas reliés à la propriété/activité modélisée. Ces modèles aléatoires peuvent être détectés par la procédure Y-randomization. Elle consiste à mélanger aléatoirement les propriétés/activités expérimentales pour la série d'apprentissage en utilisant les mêmes descripteurs, de nouveaux modèles sont obtenus. Ces derniers doivent avoir des performances très faibles.

Cependant, la validation interne est insuffisante pour étudier le pouvoir prédictif d'un modèle. Pour cette raison la validation externe du modèle est devenue une norme et une partie obligatoire dans la modélisation QSAR [26] [27].

I.7.2 Validation externe

Cette méthode consiste à prédire la propriété/activité d'une série de molécules appelée généralement série de test qui ne sont pas dans la série de développement du modèle, cette validation est caractérisée par les paramètres $R^2(\text{test})$ $R^2_{cv}(\text{test})$. Récemment plusieurs études [28] [29] ont montré l'insuffisance des paramètres R_2 , R_{2CV} pour vérifier le pouvoir prédictif des modèles QSAR. Par conséquent, d'autres paramètres doivent être vérifiés pour cet objectif. Ces paramètres sont connus sous le nom « critères de validation externe » ou souvent appelés « critères de Trophsa » [28].

– **Critères de validation Externe (série de test)**

- $R^2 > 0.7$ (critère 1)
- $R^2_{cv} > 0.6$ (critère 2)
- $\frac{R^2 - R_0^2}{R^2} < 0.1$ et $0.85 \leq K \leq 1.15$ (critère 3)
- $|R^2 - R_0^2| \leq 0.3$ (critère 5)
- $\frac{R^2 - \hat{R}_0^2}{R^2} < 0.1$ et $0.85 \leq \hat{K} \leq 1.15$ (critère4)

* R_0^2 coefficient de corrélation entre les valeurs prédites et expérimentales pour la série de test .

k : est la constante de la droite (à l'origine) de corrélation (valeurs prédites en fonction des valeurs expérimentales)

$R_0'^2$ coefficient de corrélation entre les valeurs expérimentales et prédites pour la série de test .

I.8 Estimation des paramètres statistiques du modèle

Dans le cas d'un modèle à p variables, le critère des moindres carrés s'écrit:

$$S(\beta_0 \dots \beta_p) = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n \varepsilon^2 = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{1i} - \dots \beta_p X_{pi})^2$$

Les valeurs des β qui minimisent ce critère seront les solutions $\beta_0, \beta_1, \dots, \beta_p$ du système linéaire de $(p+1)$ équations à $(p+1)$ inconnues obtenues comme suit :

$$\hat{Y} = X\hat{\beta} = XB \quad \beta = (X'X)^{-1}X'Y \quad (4)$$

Avec

X' : la matrice transposée de la matrice des variables explicatives

$X (X'X)^{-1}$: la matrice inverse de la matrice

$(X'X) Y$: vecteur des valeurs de la variable à expliquer

I.8.1 Ecarts type des coefficients de régression

On peut calculer les écart-type des estimateurs (les coefficients de régression) en utilisant la relation $\sigma^2 (X'X)^{-1}$ avec σ^2 la moyenne des carrés des erreurs ($SSE/n-1-p$) La diagonale de cette matrice nous donnera les écarts type des estimateurs (erreurs sur les coefficients de régression).

I.8.2 Confiance à accorder aux résultats

On cherche à tester la validité d'hypothèses concernant les (vrais) paramètres du modèle c'est-à-dire confirmer l'existence du modèle et la contribution de chaque descripteur dans l'explication de la réponse Y .

I. 8.3 Test de la signification globale de la régression

– Test de Fisher-Snedecor

Ce test a surtout un intérêt dans le cadre de la régression multiple, c'est-à-dire avec p régresseurs $X_1, X_2, \dots, X_p$:

Ce test permet de connaître l'apport global de l'ensemble des variables $X_1, \dots, X_p$ à la détermination de Y .

On teste l'hypothèse nulle : $H_0 : \beta_1 = \dots \beta_p = 0$ contre H_a : il existe au moins un β_j parmi $\beta_1 \dots \beta_p$ non égal à 0.

$$F = \frac{MS_{\text{error}}}{MS_{\text{model}}} \quad (5)$$

avec

$$MS_{\text{model}} = \frac{SS_{\text{Model}}}{p} \quad (6)$$

et

$$MS_{\text{error}} = \frac{SS_{\text{error}}}{n-p-1} \quad (7)$$

Si H_0 est vraie et sous réserve des suppositions suivantes :

Ce rapport F est une valeur observée d'une variable à p et $n-p-1$ degrés de liberté (p : nombre de variables explicatives, n : nombre d'observations).

La statistique F suit une loi de Fisher-Snedecor

$$F = \frac{MS_{\text{model}}}{MS_{\text{error}}} \approx F(p, n - p - 1)$$

Si la quantité F observée dépasse seuil, on rejette l'hypothèse H_0 (au niveau $\alpha=5\%$) et dans le cas contraire, on conserve H_0 .

Si $F_{\text{observé}} \geq F_{1-\alpha}(p, n - p - 1)$

Alors : $H_0 : \beta_1=0 \dots \beta_p=0$ doit être rejetée au niveau α .

Cependant, la plupart des logiciels ne demandent pas de fixer a priori un niveau α pour effectuer le test : ils donnent le niveau à partir duquel notre décision aurait changé : c'est la notion de p -value.

Si $p\text{-value} \leq \alpha$ alors on rejette l'hypothèse nulle

– **Test de signification de chaque paramètre (chaque descripteur) t-Student**

Pour voir la contribution de chaque paramètre dans l'explication de la variable dépendante Y on utilise la statistique « t » définie auparavant en régression simple.

On calcule t-test pour chaque paramètre β_i en utilisant la formule ci-dessous

$$t - \text{observé} = \frac{\hat{\beta}_i}{s(\hat{\beta}_i)} \quad (8)$$

Avec $s(\hat{\beta}_i)$ est l'erreur sur le paramètre β_i

En pratique si on choisit le risque $\alpha=5\%$ et si n est grand ($n>30$) pour approcher la loi de Student par la loi normale la valeur critique de t-test oscille le plus souvent autour de 2 (1,96).

I.8.4 Indicateurs de qualité d'une régression linéaire

I.8.4.1 Coefficient de détermination R^2

Pour évaluer la précision avec laquelle la dépendance trouvée décrit la variance de la variable dépendante, le coefficient de détermination (R^2) est utilisé. Un modèle, que l'on qualifie de bon, possède des estimations proches des vraies valeurs de Y. Les deux quantités SCE totale (SST) et SCE modèle (SSR) sont des sommes des carrés donc toujours positives ou nulles. Le rapport de ces deux quantités nous donne le coefficient de détermination R^2 :

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (9)$$

Il est compris entre 0 et 1. Plus il est proche de 1 et plus la régression permet d'expliquer une grande partie de la variance totale de la variable à expliquer.

I.8.4.2 Déviation standard (SD)

La fiabilité de la prédiction de la variable dépendante peut être évaluée également par la valeur de l'erreur type d'estimation « s » ou déviation standard (SD).

$$S^2 = MSE = \frac{\sum_i^n (Y_i - \hat{Y}_i)^2}{n-p-1} = \frac{SSE}{n-P-1} \quad (10)$$

Root Mean Square Error= SD

$$\text{Root MSE} = \text{SD} = s \quad (11)$$

L'estimation de l'erreur-type appelé aussi déviation standard SD est une mesure de la dispersion des valeurs observées de la variable dépendante sur la droite de régression (de surface). Les petites valeurs de SD signifient un bon ajustement statistique du modèle et une forte fiabilité de la prédiction.

I.8.4.3 Sélection de variables et choix du modèle dans le cas d'une régression multiple

Parmi l'ensemble des p variables disponibles, toutes n'ont pas nécessairement un intérêt dans la modélisation de Y , et il peut alors être néfaste de les utiliser. Nous sommes alors en présence de différents modèles possibles parmi lesquels il faut faire un choix.

I.8.5 Critères de comparaison de modèle

I.8.5.1 Coefficient de détermination R^2

Il n'est pas recommandé d'utiliser R^2 pour comparer des modèles avec un nombre différent de descripteurs, le coefficient R^2 nous dira toujours de choisir le modèle avec le plus grand nombre de descripteurs car son R^2 sera plus important (on projette sur un espace plus grand), même si les variables sont sans effets sur la réponse Y . Par conséquent, un autre paramètre statistique peut être utilisé en régression multiple, appelé $R^2_{\text{ajusté}}$.

I.8.5.2 Coefficient de détermination ajusté $R^2_{\text{ajusté}}$

Ce coefficient est utilisé en régression multiple par ce qu'il tient compte du nombre de paramètres du modèle.

$$R^2_{adj} = 1 - \frac{(n - \text{intercept})(1 - R^2)}{n - p} \quad (12)$$

Ou bien

$$R^2_{adj} = \frac{R^2(n - 1) - p}{n - p - 1}$$

Cette formule indique notamment que R^2_{adj} est toujours inférieur à R^2 , et ceci d'autant plus que le modèle contient un grand nombre de prédictors (descripteurs).

I.9 Applications des méthodes QSAR/QSPR

Les applications des méthodes QSPR/QSAR sont très nombreuses, elles touchent tous les domaines où la structure chimique intervient, entre autre on peut citer [31]:

Propriétés physico-chimiques : Point d'ébullition, Point de fusion, densité, indice de réfraction, température critique, viscosité, solubilité, pression de vapeur, tension superficielle, Coefficients de partition : eau/octanol, air/eau, huile/air, lait/plasma...

Activités biologiques : Anti VIH, Anti malaria, Anti Diabète, Anti Cancer, Anti-oxydante, Anti inflammatoire...

Autres propriétés/activités : Prédiction de la toxicité aquatique des composés chimiques vis-à-vis des espèces environnementales, Toxicité des nanoparticules, Toxicité des pesticides et des colorants, Propriétés inhibitrices de corrosion, Concentration micellaire critique, Prédiction de plusieurs propriétés dangereuses telle que l'explosibilité et l'inflammabilité de certaines familles de molécules chimiques, Conception des médicaments et de nombreux autres produits tels que les agents tensio-actifs, parfums, les colorants et les produits de la chimie fine ...

I.10 Logiciels utilisés dans nos études RQSA

Il existe plusieurs logiciels libres ou commerciaux disponibles dans les études QSAR. Ceux-ci comprennent des logiciels spécialisés pour dessiner les structures chimiques, générant des structures 3D, le calcul des descripteurs moléculaires et le développement de modèles QSAR. Les logiciels utilisés dans nos travaux sont :

Les structures 3D des molécules ont été générées par HyperChem.

Les descripteurs ont été calculés par DRAGON et HyperChem.

L'analyse descriptive et la validation des modèles ont été faites par MINITAB.

Les modèles QSAR ont été générés en utilisant les logiciels suivants :

- Pour la régression linéaire multiple, la régression non linéaire multiple et la régression des moindres carrés partiels, nous avons utilisé programme minitab

Chapitre II

composés organiques d'aniline

II.1 Introduction

En 1826, Otto Unverdorben isole une nouvelle espèce chimique par distillation de l'indigo, espèce également obtenue en 1834 par Friedlieb Runge à partir des goudrons de houille. Cette nouvelle molécule est à nouveau préparée quelques années plus tard, en 1841, par Carl Fritzsche en faisant réagir de l'indigo avec de la potasse. L'indigo étant extrait de la plante *Indigofera anil*, le nom d'aniline sera donné à ce nouveau composé organique. En 1856, William Perkin réalise la première synthèse d'un colorant, la mauvéine, à partir d'aniline. En 1865, est fondée, en Allemagne, la « Badische Anilin & Sodafabrik », qui deviendra la BASF. L'aniline est alors largement utilisée à l'échelle industrielle pour la production de colorants, comme la fuchsine. Ces colorants seront rapidement supplantés par un nouveau type de molécules : les colorants azoïques, obtenus à partir d'aniline par des réactions de diazotation.

Aujourd'hui, l'aniline reste un composé organique de première importance. Elle est préparée par réduction du nitrobenzène ou par action d'ammoniac sur le chlorobenzène. La plus grosse partie du tonnage produit est utilisée dans la chimie des polymères, en particulier pour la préparation du monomère diisocyanate intervenant dans la synthèse du polyuréthane. Les colorants noirs à base d'aniline sont encore employés pour la teinture des cuirs, dans les encres d'imprimerie ou pour le marquage du linge. L'aniline reste utilisée dans le procédé de vulcanisation des caoutchoucs, ainsi que dans le domaine pharmaceutique pour la conception d'agents bactéricides tels que les dérivés de sulfanilamides. Depuis quelques années, l'aniline connaît un renouveau dans le domaine des matériaux innovants. En effet, en polymérisant l'aniline, on obtient un polymère conducteur. L'aniline, dont la formule est donnée figure II.1 est largement exploitée car elle possède une réactivité très riche, de par la présence simultanée d'un cycle benzénique et d'une fonction amine au sein de sa molécule.

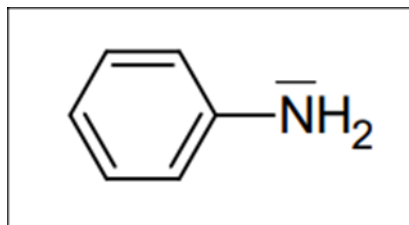


Figure II.II1 : Formule de l'aniline

Les différentes parties du problème sont totalement indépendantes et, dans chaque partie, de nombreuses questions sont également indépendantes. Chaque réponse doit être correctement rédigée et justifiée par des arguments précis et concis.

II.1.1 Historique :

L'aniline a été isolée pour la première fois en 1826 par Otto Un verdorben par distillation de l'indigo. Celui-ci la baptisa crystalline. En 1834, Friedrich Runge parvint à isoler du goudron de houille une substance qui une fois traitée par du chlorure de chaux prend une couleur bleue. Il la baptisa kyanol ou cyanol. En 1841, C. J. Fritzsche obtint une substance huileuse en traitant de l'indigo avec de la potasse, qu'il baptisa aniline, d'après le nom d'une plante produisant de l'indigo, Indigofera anil. Le mot « anil » est lui-même issu des termes sanskrits nīla, bleu profond, et nīlā plante d'indigo. A peu près en même temps, N.N. Zinin découvrit que la réduction du nitrobenzène permet d'obtenir une base qu'il baptisa benzidam. Par la suite, August Wilhelm von Hofmann étudia ces différentes substances et démontra en 1855 qu'elles sont identiques. La première utilisation d'aniline à l'échelle industrielle concerna la fabrication de la mauvéine, un colorant violet découvert en 1856 par William Henry Perkin.

II.1.2 Famille Alanaline :

L'alanine est un des 20 acides aminés principalement retrouvés dans les chaînes peptidiques des protéines. Pour la synthétiser, on fait agir le malonate d'éthyle sur l'acétate de sodium. L'anion obtenu réagit sur le bromoéthane en milieu basique puis une saponification suivie d'un passage en milieu acide. Puis on fait réagir le dibrome en présence de phosphore. Enfin l'action de l'ammoniac en excès suivie d'une hydrolyse acide permet d'obtenir l'alanine. La β -alanine est l'acide 3-aminopropanoïque ou acide β -aminopropanoïque, un isomère de position de l'alanine.

Dans cette molécule, le groupe amine est en position β (3) du groupe carboxyle, alors qu'il est en α sur l'alanine. Certains composés appartiennent à la famille des analines :

- 2-4-6-trichloro aniline :

La 2,4,6-Trichloroaniline est un composé chimique de formule $C_6H_4Cl_3N$. Il est utile comme intermédiaire dans les réactions chimiques.

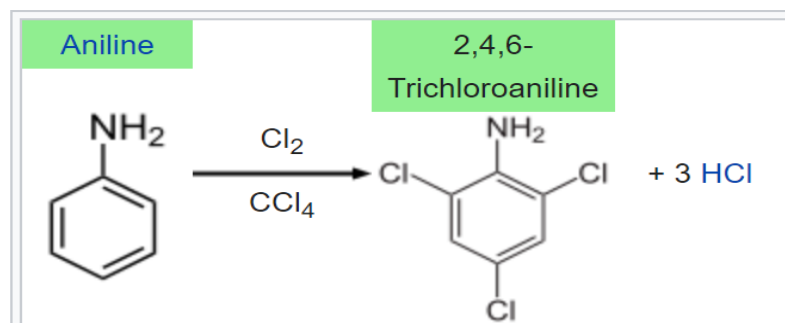


Figure III.2 : La préparation de 2,4,6-trichloroaniline

- 2-chloro aniline :

La 2-chloroaniline est largement utilisée dans les industries des polymères, du caoutchouc, des produits pharmaceutiques et des colorants. La détermination de la 2-chloroaniline dans les eaux usées industrielles polymères a été examinée par HPLC avec microdialyse en ligne. La dégradation de la 2-chloroaniline dans l'eau par ozonation ainsi que par photolyse (lumière UV de 254 nm) et radiolyse (rayons γ) en présence d'ozone a été étudiée.

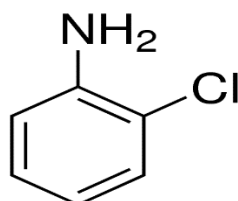


Figure II. 3 : 2-chloro aniline

Propriétés réaction chimique des Aniline

Est une Aniline substance toxique avec des effets possibles cancérigène (Phrase de risque R40). Il est un liquide huileux avec une odeur caractéristique et l'apparence incolore, cependant, au fil du temps, il produit des impuretés et oxyde résineux de couleur rouge-brun. Il brûle

facilement avec une grande flamme et fumée. Du point de vue chimique est une base faible. Les amines aromatiques telles que l'aniline sont généralement moins de base d'amines aliphatiques, ceci est parce que le paire électron de l'atome d'azote est partiellement partagé avec le cycle aromatique à résonance et il est donc pas disponible pour être vendus à des espèces chimiques aigre.

- Aniline réagit avec des acides forts pour donner des sels anilinium $C_6H_5-NH_3^+$ que, généralement, pour le chauffage décomposent libération ammoniacque; réagit également avec le des halogénures d'acyle, comme le chlorure d'acétyl CH_3COCl , ou des acides carboxyliques dans déshydrater environnement pour donner amides. Les amides sont souvent appelés aniline anilides; $CH_3-CO-NH-C_6H_5$, par exemple, il est l'acétanilide.
- Pour la réaction avec halogénure alkyle aniline former des amines secondaires et tertiaires; en faisant bouillir avec sulfure de carbone il donne la solfocarbaniide (Difeniltiourea, $CS (NHC_6H_5)_2$) Qu'il peut être décomposé en C phényle thiocyanate C_6H_5CNS et trifenilguanidine $C_6H_5N = C (NHC_6H_5)$.
- comme le phénol aniline aussi facilement des réactions subit substitution électrophile sur le noyau aromatique en position par et jardin potager. Pour la réaction avec 'acide sulfurique à $180^\circ C$ donne l'acide sulfanilique p- $C_6H_4-SO_3H$ dont les amides sont la classe de médicaments sulfamides, antibactérien couramment utilisé au début de XX siècle.
- L'aniline et ses dérivés substitués sur le noyau aromatique réagit facilement avec l'acide nitreux pour donner la sels diazonium; à travers le groupe $-NH_2$ aniline peut être converti en d'autres des groupes fonctionnels tels que $-OH$, $-CN$ ou un halogène ou peuvent être synthétisés composés colorants, lesdits colorants diazoïques ».
- avec 'acroléine aniline réagit conformément à la synthèse de Skraup comme suit:
- $CH-CH=CH_2 + C_6H_5-NH_2 \rightarrow C_6H_5-NH-CH_2-CH=O$
- L'aniline est aisément oxydé à la fois sur le groupe $-NH_2$ à la fois sur le noyau aromatique; l'oxydation de l'aniline en milieu alcalin produit l'azobenzène; avec 'acide arsénique vous obtenez le colorant aniline pourpre, avec l'acide chromique Il est transformé en correspondant quinone, tandis que pour l'oxydation avec chlorés en

présence de sels métalliques, en particulier de vanadium, vous obtenez le colorant d'aniline noir.

- L'oxydation avec acide chlorhydrique et chlorate de potassium donne le chloranile; oxydation avec permanganate de potassium en solution neutre il se transforme en nitrobenzène, dans la solution alcaline à oxyde azobenzène, de l'ammoniac et l'acide oxalique, en solution acide, il se transforme en noir d'aniline.
- La réaction avec l'acide hypochloreux donne 4-aminophénol et 4-amino-diphénylamine.

II.1.3 L'importance commerciale de l'aniline

La valeur commerciale de l'aniline est liée à sa polyvalence comme intermédiaire dans la synthèse chimique des produits industriels tels que les médicaments, les auxiliaires et colorants.

À partir de 1858 aniline est la matière première utilisée dans la production de centaines de substances colorantes, y compris magenta, safranine, l'indigo et bien d'autres. aniline est produit industriellement par la réduction du nitrobenzène avec du fer et de l'acide chlorhydrique, puis le purifier pour distillation dans un courant de vapeur d'eau.

II.1.4 EFFETS SUR La santé

L'aniline est l'un des premiers cancérigène reconnu pour l'homme. Ludwig Rehn, en 1895, a rejoint l'incidence de l'exposition au cancer de la vessie des travailleurs industriels allemands de colorants artificiels.[3] Causes en particulier Les tumeurs de la vessie et même les reins, les cancers de la peau, le foie et le sang. La molécule est pas cancérigène lui-même, mais elle le devient à la suite de son métabolisme foie par des procédés d'hydroxylation de travailler enzymes dépendantes cytochrome P450 (chaîne oxydative de microsomes). Les principaux dérivés de l'aniline sont le para-amino-phénol et de 2,4-dioxy-aniline; à leur tour, ces intermédiaires sont tenus de subir des réactions oxydoréduction à se transformer en leur respective quinones: Para-iminoquinone et 2-hydroxy-para-iminoquinone.

Ce sont ces molécules, ces derniers (électrophiles) hautement réactifs qui réagissent facilement avec des résidus spécifiques de protéines et d'acides nucléiques en établissant une liaison de type covalent. Dans le cas des protéines si elle vient à ces enzymes peuvent être

inactivés en permanence et doivent être dégradés. Dans le cas des acides nucléiques, en particulier le ADN, interactions covalentes se produisent avec les bases guanine et adénine, Les réactions qui conduisent les deux reticulations entre brins d'ADN à la fois à la rupture de la même. En fonction des gènes à l'intérieur les touchés et le siège sera en mesure d'avoir des altérations variables dans le phénotype cellulaire, par exemple des altérations des gènes suppresseurs de tumeur peut rendre la cellule plus sensible à la transformation néoplasique.

II.2 Caractéristiques

II.2.1 Utilisations

- isocyanates, intermédiaires de synthèse des polyuréthannes.
- accélérateurs de vulcanisation dans l'industrie du caoutchouc.
- matières colorantes.
- produits phytosanitaires (herbicides, fongicides) et pharmaceutiques.

II.2.2 Propriétés physiques

L'aniline est un liquide huileux, incolore quand il est fraîchement distillé mais ayant tendance à brunir sous l'action de la lumière et de l'air. Elle possède une odeur âcre caractéristique d'amine, détectable à de très faibles concentrations (de l'ordre de 0,5 ppm). Elle est légèrement soluble dans l'eau (3,5 % à 25 °C) et est miscible à la plupart des solvants organiques.

Tableau II.1: Valeurs physicochimiques pour l'aniline .

Nom substance	Détails	
Aniline	N ° CAS	62-53-3
	Etat physique	Liquid
	Masse molaire	93.13
	Points de fusion	-6.2 ° C
	Points de d'ébullitions	184 ° C
	Densité	1.022
	Densité gaz/Vapeur	3.3
	Pression de vapeur	0.4 hPa à 20°C
	Point d'éclair	70 à 76 °C (selon les Sources) en coupelle fermée
	Viscosité	4.4 mPa.s
	Température d'auto-	615 à 630 °C (selon les sources)
	Limites d'explosivité ou d'inflammabilité (en volume % dans l'air)	Limite inférieure: 1.2 à 1.3 % Limite supérieure: 11 %
	Coefficient de partage n-octanol/eau (log pow)	0.9

II.2.3 Propriétés chimiques

Dans les conditions normales de température et de pression, l'aniline est un produit stable. L'aniline est une base faible qui réagit avec les acides forts en formant des sels d'anilinium très solubles dans l'eau. Elle réagit également avec les halogénures d'acyle pour former des amides.

L'oxydation de l'aniline peut être réalisée par un grand nombre de composants tels le peroxyde d'hydrogène, le chlorate de sodium ou de potassium, l'acide chromique, le permanganate de potassium, l'acide hypochloreux... Avec des oxydants puissants, la réaction peut être violente.

L'aniline peut s'enflammer et réagir violemment avec de nombreux composés dans les cas suivants :

- production d'un composé combustible et explosif avec l'acide perchlorique et le formaldéhyde.
- réaction violente avec le trichloronitrométhane.

- inflammation spontanée avec l'acide nitrique fumant, les peroxydes de sodium et de potassium.

Les métaux alcalins et alcalino-terreux réagissent avec le groupement amine pour donner des anilides métalliques avec dégagement d'hydrogène. La combustion de l'aniline dégage des fumées toxiques (oxydes d'azote et oxyde de carbone).

II.2.4 Récipient de stockage

L'aniline est généralement stockée dans des récipients en acier. Le verre est également utilisable pour de petites quantités. Dans ce cas, les récipients seront protégés par une enveloppe métallique plus résistante, convenablement ajustée.

II.3 La toxicologie

L'organisme humain ainsi que tous les autres organismes vivants sont en relation avec leur milieu par un ensemble d'échanges qui contribuent à maintenir un équilibre dynamique. Par exemple, la respiration permet d'absorber l'oxygène de l'air et d'y rejeter du dioxyde de carbone. Quoi que nous fassions, le milieu nous influence et nous l'influons. Ce principe d'action-réaction signifie que toute action a des conséquences. Le milieu ne constitue cependant pas un tout homogène, mais plutôt un ensemble composé de nombreux éléments, comprenant les produits chimiques qui peuvent affecter la santé des organismes vivants et l'équilibre de l'écosystème. (Figure II.4).

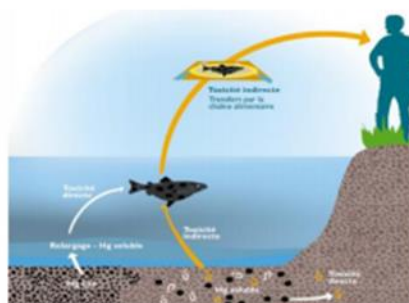
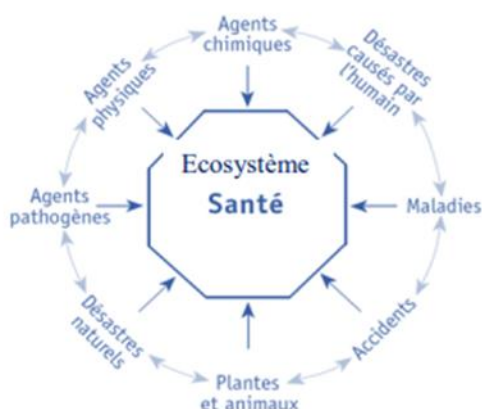


Figure II.4: Le milieu et les différents éléments pouvant affecter l'organisme vivant et l'écosystème

Chaque année, l'industrie met des centaines de nouveaux produits sur le marché, venant ainsi accroître le nombre de ceux qui existent déjà. Il est important de connaître l'innocuité (qualité de ce qui n'est pas nuisible) ou la nocivité (caractère de ce qui est nuisible) des produits chimiques pour bien en saisir les effets sur notre santé et sur notre environnement.

II.3.1 Notions et définitions

II.3.1.1 Définition de la toxicologie

La toxicologie est depuis longtemps reconnue comme étant la science des poisons. Elle étudie les effets nocifs des substances chimiques sur les organismes vivants.

II.3.1.2 Définition de l'écotoxicologie

L'écotoxicologie étudie les impacts des agents polluants sur la structure et le fonctionnement des écosystèmes. Les effets d'un agent polluant dépendent de plusieurs facteurs, comme par exemple l'évolution du polluant dans le milieu, le mode et la voie d'administration du polluant.

II.3.1.3 Définition d'un toxique (poison)

Un poison, ou toxique, est une substance capable de perturber le fonctionnement normal d'un organisme vivant. Il peut être de source naturelle (ex. : poussières, pollen) ou artificielle (ex. : urée-formaldéhyde), ou de nature chimique (ex. : acétone, benzène, anthrax...) ou biologique (ex. : aflatoxines, anthrax).

II.3.1.4 Définition de la toxicité

il s'agit de la capacité inhérente à une substance chimique de produire des effets nocifs chez un organisme vivant et qui en font une substance dangereuse. On parle alors d'un effet toxique, cet effet peut être local qui survient au point de contact, ou bien systémique qui survient à un endroit éloigné du point de contact initial.

II.3.1.5 Définition de la dose

La dose est la quantité d'une substance à laquelle un organisme vivant est exposé.

II.3.1.6 Types de toxicité

- Toxicité aiguë: est définie comme celle qui résulte de l'exposition unique et massive à un produit chimique entraînant des dommages corporels pouvant conduire à la mort .

Elle introduit la notion de dose « absorbée » (par ingestion, inhalation ou contact cutané)

et se mesure par la **IGC50** .

- Toxicité chronique : est le résultat de l'exposition prolongée à plus ou moins faible dose à un xénobiotique toxique dont les effets néfastes ne se feront sentir que quelques mois à quelques années voire des dizaines d'années plus tard.
- Toxicité sub-aigüe: correspond à un stade d'exposition intermédiaire de l'ordre de trois mois.

II.3.2 Toxicité des produits chimiques organiques

Sur 100 000 produits chimiques libérés dans l'environnement, moins de 1 - 5% ont des données de toxicité disponibles. Même pour les produits chimiques de volume de production élevé ou HPVCs (les substances chimiques produites en quantités >1000 tonnes par an dans l'UE ou > environ 442 tonnes par an aux Etats-Unis) il y'a un manque d'informations concernant leur toxicité, comme le montre le Tableau II.2.

Tableau II.2: Données de toxicité pour les produits chimiques de haute production

	Union européennes (UE)	Etats-Unis (USA)
Nombre	2465	2863
Données complètes de toxicité	3 %	7 %
Données partielles de toxicité	43 %	50 %
Pas de données de toxicité	54 %	43 %

Avec une préoccupation croissante sur l'environnement et la santé humaine, les gouvernements et les organismes de réglementation à travers le monde cherchant à évaluer les risques éco-toxicologiques posés par la libération des substances chimiques. Ils ont proposé que **30000** produits chimiques existants soient testés sur les animaux pour une gamme des effets toxiques. Ce serait évidemment une tâche très coûteuse, impliquant l'utilisation des milliers d'animaux, par conséquent le recours aux méthodes alternatives est devenu d'une grande importance.

II.3.3 Les approches QSAR pour l'étude de la toxicité

QSAR pour la toxicité remonte au 19ème siècle. En 1863, A.F.A. Crois à l'Université de Strasbourg a observé que la toxicité des alcools à des mammifères augmente lorsque la solubilité des alcools dans l'eau diminue. Dans les années 1890, Hans Horst Meyer de l'Université de Marburg et Charles Ernest Overton de l'Université de Zurich, ont noté que la toxicité des composés organiques était tributaire de la lipophilie. Au début des années 1960 Corwin Hansch a proposé un modèle mathématique pour corréler l'activité biologique et la structure chimique, cette date est considérée comme étant la naissance des méthodes QSAR. Depuis, l'utilisation des QSAR en toxicologie n'a pas cessé d'évoluer.

Des modèles QSAR sont maintenant mis au point en utilisant une variété d'approches, de méthodes d'analyse de données et de paramètres. Un grand nombre d'études QSAR de toxicité et en particulier la toxicité aiguë ont été publiées dans la littérature. La plupart des données de toxicité pour l'environnement ont été obtenues en utilisant des espèces aquatiques en l'occurrence les poissons, les Daphnies, les protozoaires « *Tetrahymena Pyriformis* », *Vibrio fischeri*, les algues ...

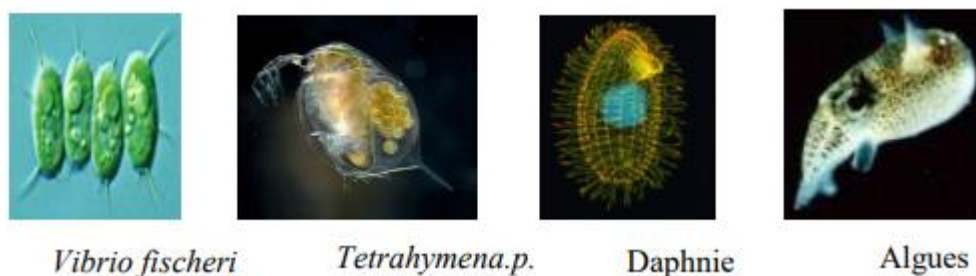


Figure II.5 : Les différentes espèces aquatiques utilisées pour l'étude de la toxicité aquatique

Cronin et Dearden ont examiné la littérature concernant la modélisation

QSAR de la toxicité aquatique. Plusieurs modes d'action ont été identifiés chez les espèces aquatiques, à savoir narcose non polaire, narcose polaire, découplage de la phosphorylation oxydative, irritation de la membrane respiratoire, inhibition d'acétylcholinestérase, saisie de système nerveux central, inhibition de photosynthèse, et l'alkylation. Cependant, ceux-ci sont généralement plus largement regroupés comme: narcose non polaire, narcose polaire, réactifs sélectifs (électrophiles), et les molécules à mécanismes d'action spécifiques. Verhaar et al. ont développé un système sur la base de la présence de groupes fonctionnels pour classer les produits chimiques dans ces quatre groupes .

Chapitre III

Application sur série d'aniline

III.1 Introduction

L'amplification spectaculaire des nouveaux composés par l'industrie chimique en général et en particulier par l'agrochimie, la pétrochimie et la pharmaco-chimie est accompagnée d'une augmentation de la charge toxique dans l'environnement. Pour cette raison, le développement d'outils capables d'évaluer les effets dangereux sur les espèces vivantes devrait recevoir une attention particulière. Les méthodes Quantitatives structure-propriété/activité (QSPR/QSAR) sont parmi les outils les plus pratiques en chimie physique. Ces méthodes sont basées sur l'axiome que la variance dans les propriétés physico-chimiques et les activités des composés chimiques est déterminée par la variance dans leurs structures moléculaires. L'expression de la toxicité chimique est une combinaison de pénétration dans ou à travers des membranes biologiques et de l'interaction de la substance toxique avec le site d'action. Ce principe est représenté mathématiquement par le QSAR générique suivant : $(\text{toxicité})^{-1} = A (\log \text{ de pénétration}) + B (\log \text{ de l'interaction}) + C \dots\dots\dots(1)$

III.2 Méthodologie

III.2.1 Base de données:

Un groupe de composés chimiques (48 éléments) de nature aniline, nous essayons de déterminer la raison de les appeler logiGc50 (inhibition de la croissance cellulaire).

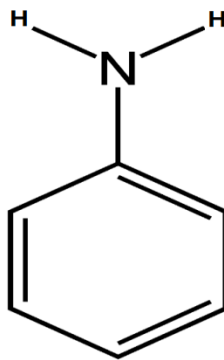


Figure III.1 : Forme générale d'aniline

III.2.2 Pourquoi nous intéressons-nous aux composés organiques (composés carbonés)?

Notre intérêt pour les composés organiques est dû aux raisons suivantes:

1. Ces composés abondent, car le nombre de composés organiques connus à ce jour est d'environ 3 millions Composés. Cela est dû à la capacité des atomes de carbone à se lier les uns aux autres par des liaisons fortes. Les chaînes moléculaires se présentent sous différentes formes et tailles et sont également capables de se lier à d'autres atomes tels que l'oxygène Azote, halogènes et autres.
2. L'importance des composés de carbone (organiques) pour l'homme, car ils occupent une place importante dans la vie. Par exemple, le logement, les vêtements, les médicaments, les détergents, les pesticides et les outils cosmétiques Composés organiques.
3. Les composés organiques se distinguent par des propriétés physiques et chimiques spéciales qui les distinguent des autres composés.
4. Les réactions ne se produisent dans les composés chimiques que dans une certaine fraction. Pendant que l'autre partie n'entre pas dans l'interaction.

Tableau III.1 : Un ensemble de propriétés pour les composés aniliques

N o.	MolID	logiGc50	SP05	R3v+	DP03	DP02	RDF080m	Ss	Mv	Log(P)	nC
1	2-3-4-trichloro aniline	1.35	4.228	0.087	4.095	3.852	0.000	27.00	0.78	3.33	6
2	2-3-chloro aniline	1.02	3.640	0.087	3.732	3.583	0.000	23.22	0.73	2.82	6
3	2-4-5-trichloro aniline	1.30	4.421	0.086	4.198	3.907	0.000	27.00	0.78	3.45	6

4	2-4-6-trichloro aniline	1.01	4.342	0.039	4.167	3.892	0.000	27.00	0.78	3.52	6
5	2-4-dichloro aniline	0.56	3.924	0.040	3.885	3.661	0.000	23.22	0.73	2.78	6
6	2-4-dimethyl aniline	-0.30	3.708	0.022	3.780	3.596	0.000	19.00	0.60	1.84	8
7	2-5-dichloro aniline	0.58	3.982	0.039	3.897	3.664	0.000	23.22	0.73	2.75	6
8	2-5-dimethyl aniline	-0.35	3.737	0.023	3.787	3.599	0.000	19.00	0.60	1.83	8
9	2-6-dichloro aniline	0.33	3.702	0.040	3.751	3.590	0.000	23.22	0.73	2.76	6
10	2-6-dichloro-3-methyl aniline	0.69	4.200	0.041	4.068	3.830	0.000	24.89	0.69	2.9	7
11	2-6-dimethyl aniline	-0.43	3.472	0.024	3.640	3.522	0.000	19.00	0.60	1.84	8
12	2-bromo aniline	0.46	3.175	0.056	3.406	3.324	0.000	18.08	0.71	2.11	6
13	2-chloro aniline	-0.09	3.061	0.041	3.350	3.291	0.000	19.44	0.68	1.9	6
14	2-chloro-4-methyl aniline	0.24	3.799	0.039	3.827	3.627	0.000	21.11	0.65	1.9	7
15	2-chloro-5-methyl aniline	0.20	3.855	0.037	3.840	3.631	0.000	21.11	0.65	2.58	7
16	2-chloro-6-methyl aniline	0.12	3.595	0.037	3.699	3.558	0.000	21.11	0.65	2.2	7
17	2-ethyl aniline	-0.25	3.752	0.023	3.770	3.595	0.000	18.83	0.60	1.74	8
18	2-fluoro aniline	-0.31	2.811	0.025	3.226	3.215	0.000	23.33	0.64	1.26	6
19	2-iodo aniline	0.35	3.295	0.070	3.463	3.357	0.000	17.45	0.73	2.32	6
20	2-isopropyl aniline	0.10	3.918	0.024	3.978	3.782	0.000	20.67	0.59	2.12	9
21	2-methyl aniline	-0.55	2.930	0.024	3.287	3.252	0.000	17.33	0.61	1.66	7
22	2-phenyl aniline	0.78	5.490	0.019	4.956	4.486	0.006	27.00	0.67	2.84	12
23	2-propyl aniline	0.06	4.758	0.023	4.368	3.996	0.025	20.33	0.59	2.3	9
24	3-4-5-trichloro aniline	1.51	4.287	0.088	4.147	3.884	0.000	27.00	0.78	3.32	6
25	3-4-dichloro aniline	1.14	3.879	0.089	3.869	3.655	0.000	23.22	0.73	2.69	6
26	3-4-dimethyl aniline	-0.29	3.687	0.029	3.771	3.593	0.000	19.00	0.60	1.84	8
27	3-5-dichloro aniline	0.71	3.897	0.038	3.907	3.682	0.000	23.22	0.73	2.9	6
28	3-5-dimethyl aniline	-0.37	3.680	0.017	3.801	3.617	0.000	19.00	0.60	2	8
29	3-bromo aniline	0.52	3.381	0.058	3.544	3.396	0.000	18.08	0.71	2.1	6
30	3-chloro aniline	0.09	3.262	0.042	3.487	3.363	0.000	19.44	0.68	1.88	6
31	3-chloro-2-methylaniline	0.45	3.558	0.044	3.685	3.552	0.000	21.11	0.65	2.58	7
32	3-chloro-4-methyl aniline	0.45	3.775	0.049	3.818	3.623	0.000	21.11	0.65	2.27	7
33	3-ethyl aniline	-0.12	3.877	0.021	3.904	3.675	0.000	18.83	0.60	2.1	8
34	3-fluoro aniline	0.04	2.986	0.025	3.355	3.285	0.000	23.33	0.64	1.3	6
35	3-iodo aniline	0.61	3.512	0.071	3.607	3.432	0.000	17.45	0.73	2.86	6
36	3-methyl aniline	-0.43	3.117	0.026	3.418	3.322	0.000	17.33	0.61	2.80	7
37	3-phenyl aniline	0.86	5.659	0.021	5.104	4.573	0.015	27.00	0.67	3.3	12
38	4-butyl aniline	1.05	5.909	0.022	5.152	4.510	1.099	21.83	0.58	3.05	10
39	4-chloro-2-methyl aniline	0.35	3.839	0.033	3.837	3.629	0.000	21.11	0.65	2.27	7
40	4-ethyl aniline	0.04	4.203	0.021	4.048	3.742	0.006	18.83	0.60	1.96	8
41	4hexyl aniline	2.04	6.638	0.023	5.736	4.972	2.784	24.83	0.57	4.21	12
42	4-isopropyl aniline	0.21	4.390	0.024	4.291	3.951	0.017	20.67	0.59	2.49	9
43	4-methyl aniline	-0.02	3.319	0.021	3.503	3.359	0.000	17.33	0.61	1.39	7
44	4-octyl aniline	2.34	7.434	0.022	6.291	5.390	3.538	27.83	0.57	5	14
45	4-pentyl aniline	1.67	6.335	0.022	5.486	4.770	3.079	23.33	0.58	3.39	11

46	4-phenyl aniline	0.95	5.871	0.022	5.215	4.630	0.420	27.00	0.67	2.86	12
47	4-propyl aniline	0.49	5.072	0.024	4.598	4.123	0.318	20.33	0.59	2.4	9
48	5--chloro-2-methyl aniline	051	3.87	0.03	3.845	3.632	0	21.11	0.65	2.58	7

III.2.3 Calcul des propriétés

- **ChemOffice Ultra** ChemDraw Ultra fournit à l'utilisateur de précieuses ressources supplémentaires. En plus de ChemDraw Pro, le package comprend ChemFinder Std, ChemInfo Std et ClipArt Pro, permettant un accès facile aux bases de données, aux informations sur des milliers de composés et aux illustrations professionnelles. à l'aide de ce programme, nous dessinons la structure moléculaire des composés aniliques .

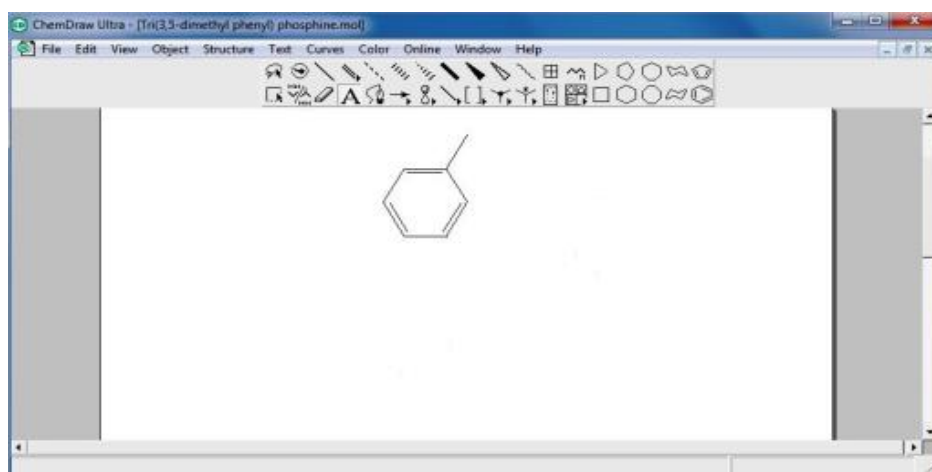


Figure III.2 : Programme ChemOffice Ultra

- **HyperChem** est un environnement de modélisation moléculaire sophistiqué qui est connu pour sa qualité, sa flexibilité et sa facilité d'utilisation. Unification de la visualisation et de l'animation 3D avec des calculs de chimie quantique, de la mécanique moléculaire et de la dynamique , en utilisant ce programme, nous traçons tous les composés aniliques pour obtenir leur structure 3D , et le calcul de l'énergie de stabilité par des méthodes semi-expérimentales PM3, RM1 et mécaniques moléculaires MM+.

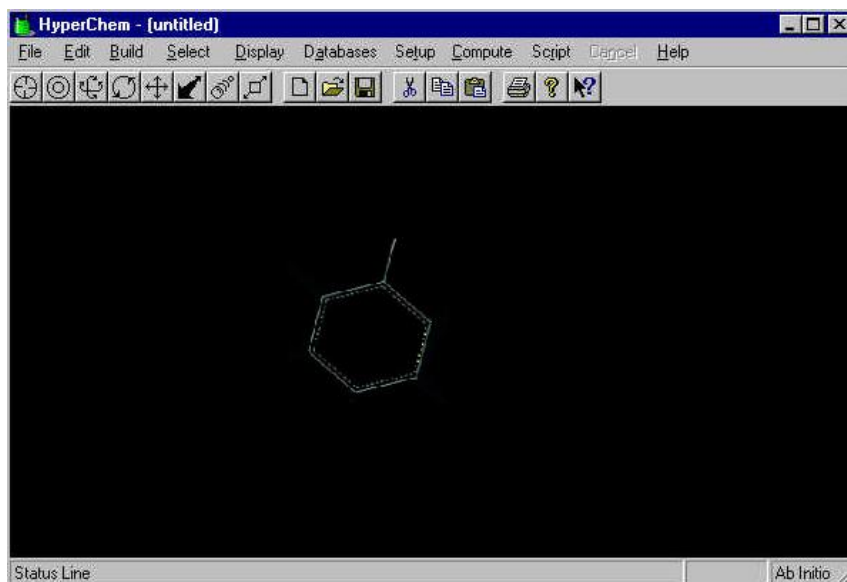


Figure III.3 : Modèle phénolique utilisant hyperchem

- **DRAGON** en utilisant le programme Dragon, nous pouvons obtenir les propriétés des composés aniliques après les avoir modélisés en HyperChem et nous obtenons les résultats affichés dans Tableau III.1.



Figure III.4 : Programme DRAGON

- **MINITAB** En utilisant ce programme nous appliquons la méthode QSAR pour obtenir une instruction de régression linéaire.
 - Avant l'application, on supprimer les colonnes aux trois quarts de zéros
 - Si modèles QSAR à un deux paramètre:

Avant de calculer, nous nous assurons que les variables ne sont pas liées les unes aux autres, nous utilisons Minitab .Les variables sont utilisées R faible. un cas spécial, pour les variables en termes de logiGc50 nous prenons la valeur la plus élevée de R .

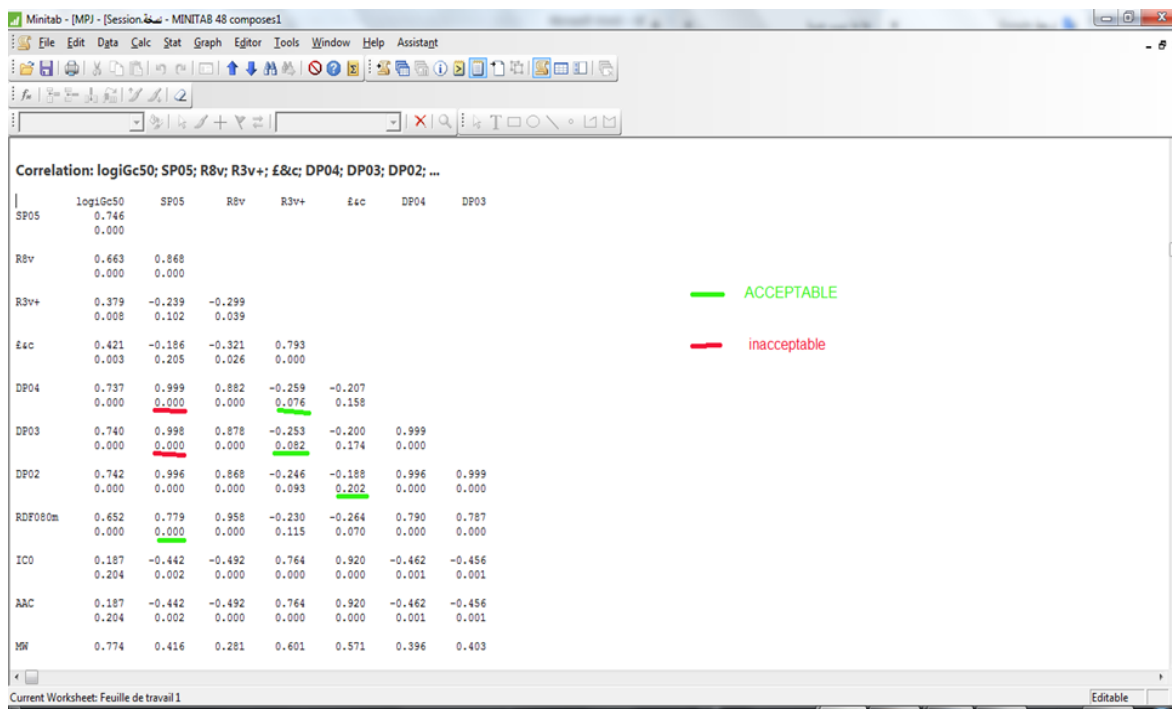


Figure III.5 : Correlation entre les valeurs

III.2.4 Analyse statistique

Des modèles Structure-toxicité ont été générés en utilisant la méthode de régression multilinéaire (MLR) de MINITAB version 17, les valeurs log (1/IGC50) indiquées en millimolaire, comme variable dépendante. Propriétés (tableau 2) sont utilisées comme variables indépendantes. Les modèles sont évalués par la valeur de R² (coefficient de détermination), le R²-ajusté, la valeur SD (racine du carré moyen des erreurs) et la valeur F (statistique Fischer). Le nombre d'observations N est également à noter.

Plusieurs modèles QSAR linéaires impliquant un, deux et trois (dans ce travail, nous utilisons deux variables au maximum) descripteurs sont élaborés et les plus fortes corrélations multivariées sont identifiées par l'option "Best sub-set" du programme MINTAB.

III.2.4.1 Écart type (SD)

Nous définissons l'écart-type d'un échantillon de taille n tiré d'une population comme la racine carrée de la variance de cette Les données et donc l'écart type des données $x_1, x_2, \dots, x_n$, dont la moyenne est arithmétique, et notée Avec le symbole S^2 , l'écart type peut être calculé par l'équation suivante

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \quad (13)$$

$\bar{X}$: Moyenne arithmétique. X_i : Le centre de la classe i , qui est égal à (la limite supérieure de la classe + la limite inférieure de la classe) / 2 . N : Nombre de catégories

III.2.4.2 Écarts absolus moyens (AAD)

Il mesure l'écart des valeurs par rapport à leur moyenne arithmétique, quel que soit le signe de cet écart, tel qu'il est calculé. Il est indiqué par le symbole σ et la moyenne des écarts absolus peut être calculée à l'aide de l'équation suivante

$$\sigma = \frac{\sum_{i=1}^n |X_i - \bar{X}|}{n} \quad (14)$$

III.2.4.3 Test statistique

Pour mesurer le moral du modèle, le test (F) est utilisé pour tester le il existe une relation significative entre la variable dépendante (Y_j) et les variables expliquées $X_1, X_2, \dots, X_K$.

$$F = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 / K}{\sum_{i=1}^n (y_i - \bar{y})^2 / (n - k - 1)} \quad (15)$$

y_i : La valeur estimée de la variable dépendante.

y_i : de la variable dépendante.

y : le vecteur de dimension $.n$

n: le nombre total.

k: nombre naturel variable

III.2.4.4 2-Coefficient de détermination (R)

Appelé aussi coefficient de corrélation, il est défini par :

$$R^2 = \frac{\text{variation due à la régression}}{\text{variation totale}} \quad (16)$$

Lorsque l'échantillon est d'étendue assez faible, il est nécessaire d'apporter une correction pour l'erreur systématique. L'estimation de la force de la liaison par le coefficient de corrélation multiple est d'autant plus exagérée que le nombre de degrés de liberté de l'échantillon $f = N - l$. La formule de correction est :

$$\bar{R}^2 = R^2 - (1 - R^2) \frac{l-1}{N-l} \quad (17)$$

$(N-l)$ le nombre de degrés de liberté (étant le nombre de coefficients significatifs) à la variance de reproductibilité.

III.2.4.5 Coefficient de corrélation linéaire (r)

Le coefficient de corrélation est une mesure quantitative relative utilisée pour déterminer le type et la force de la relation entre deux variables.

Tableau III.2 : Le degré de force de la relation du coefficient de corrélation

Valeur du coefficient de corrélation	Force de corrélation
+1	Complète (Expulser)
0.70 à 0.89	Fort (Expulser)
0.50 à 0.69	Faible (Expulser)
0	Inexistant
-0.50 à -0.69	Faible (Sens inverse)
-0.70 à -0.89	fort (Sens inverse)
-1	Complète (Sens inverse)

III.3 Les résultats de la première étape (Modèles QSAR à un seul paramètre)

Dans ce travail, l'effet des propriétés de 48 composés aniliques a été étudié pour inhiber la croissance de mes cellules (l'étude comprenait 39 caractéristiques) une méthode a été utilisée QSAR et les résultats ont été obtenus à l'aide d'un programme . ce travail a été réalisé en deux étapes , La première étape consiste à étudier chaque caractéristique seule dans son effet sur la croissance cellulaire et la deuxième étape consiste à étudier chacune des deux caractéristiques ensemble .

En fonction des résultats obtenus à ce stade :

- Notez qu'il y a une valeur significative dans ($R\text{-sq}=0.9588$) au niveau de la propriété MV et La valeur la plus basse en ($S= 0.421413$)
- Équation de régression linéaire pour un meilleur résultat (MV) :

$$\log iGc50 = -2.072 + 0.01678 MW \quad (1)$$

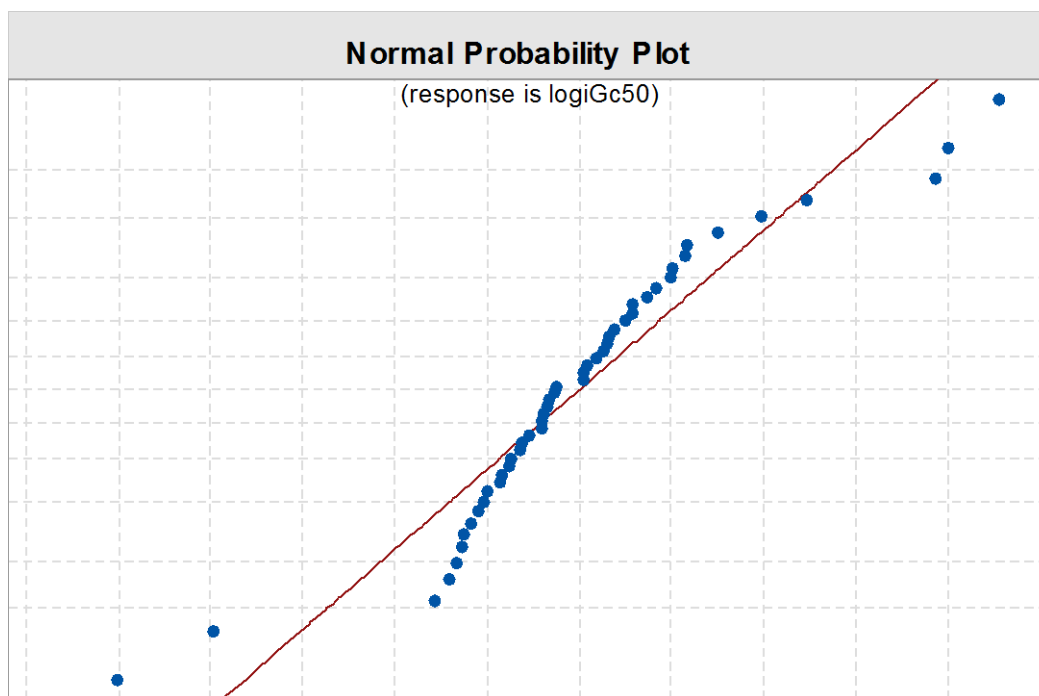


Figure III.6 : Courbe graphique de l'équation de régression linéaire

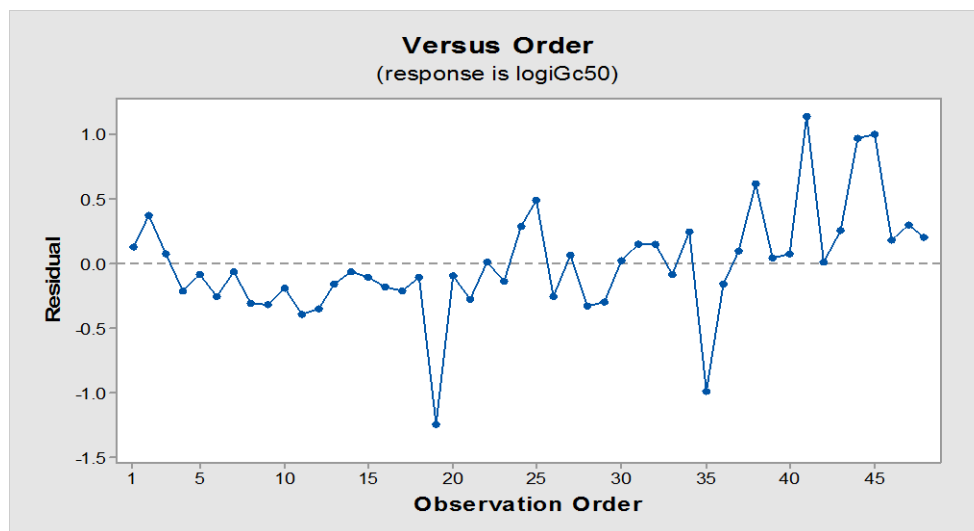


Figure III.7 : Courbe graphique Le pourcentage d'erreur pour chaque point

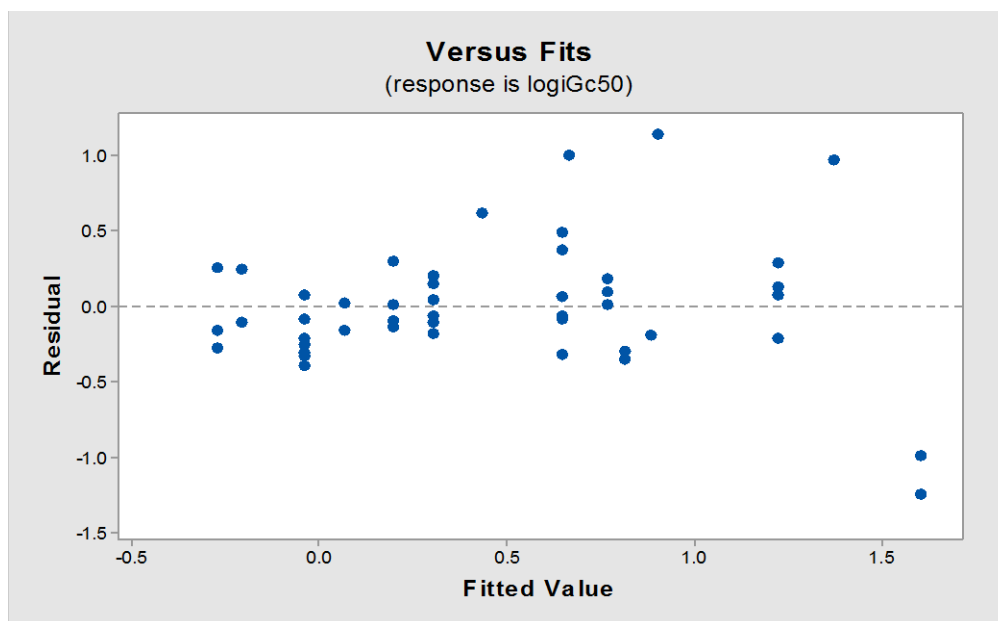


Figure III.8 : Courbe graphique L'erreur en fonction de la valeur révisée

Tableau III.3 : ajustements et diagnostics pour les observations inhabituelles

logiGc50	Resid	Std Resid
0.350	-1.253	-3.19 R X
0.610	-0.993	-2.53 R X
2.040	1.137	2.75R
2.340	0.966	2.40 R
1.670	1.002	2.41 R

R grosse erreur

X Une erreur inhabituelle

III.3.1 Analyse des résultats de la première phase

Le coefficient de détermination(R) montre la précision des résultats du modèle choisi. Plus le coefficient de détermination était significatif, plus les résultats étaient précis.

Dans ce cas, le meilleur facteur de détermination est obtenu $R=0.5988$ Ainsi, la variation en pourcentage de la variable modélisée = 59.88 % .

III.3.2 Analyse des résidus Les résultats de la première étape

$$\text{Résidu} = 1 - R = 1 - 0.5988 = 0.4012 = 40.12 \%$$

Ce rapport concerne les résidus présentant un groupe de composés organiques auxquels ce modèle ne s'applique pas. Et il a besoin de plusieurs autres explications (son effet sur la croissance cellulaire dépend d'autres caractéristiques).

III.3.3 Discutez des résultats de la première étape

Dans cette étape, nous nous concentrerons sur les composés les moins corrélés avec le modèle et ceux qui sont les plus étroitement liés.

III.3.3.1 Les composés les moins corrélés avec le modèle

Tableau III.4 : Tableau des valeurs des composés moins liés au modèle

Composé organique	logiGc50	Résid
2-iodo aniline	0.350	-1.253
3-iodo aniline	0.610	-0.993
4hexyl aniline	2.040	1.137
4-octyl aniline	2.340	0.966

Ces quatre composés organiques sont les moins compatibles avec le modèle à variable unique. Avec un faible facteur d'identification et un groupe de composés en dehors du modèle d'étude, un modèle différent avec deux variables est nécessaire.

III.3.3.2 Les composés les plus corrélés avec le modèle

Tableau III.5 : Tableau des valeurs des composés plus liés au modèle

Composé organique	logiGc50	Résid
2-4-dichloro aniline	0.56	0.08621
2-phenyl aniline	0.78	0.01265
3-chloro aniline	0.09	0.02165
4-chloro-2-methyl aniline	0.35	0.04624

A partir du graphique de logiGc50 en termes de propriété MW pour les composés chimiques, on note un pourcentage de compatibilité avec le modèle (0.5988) et les composés les plus compatibles, qui sont expliqués par l'équation de régression linéaire du tableau 6.

III.3.3.3 Une comparaison entre les composés les plus compatibles et les moins compatibles:

À partir du tableau 6 nous constatons que les composés aniliques les plus compatibles avec l'équation de régression linéaire de la variable MW sont des composés contenant du chlore et le moins toxique, et à partir du tableau 5 les composés aniliques les moins liés à la forme sont des composés contenant de l'iode et présentant une toxicité élevée .

III.4 Les résultats de la deuxième étape (Modèles QSAR à un deux paramètres):

Tableau III.6 : Résultats des caractéristiques étudiées (deuxième étape)

la caractéristique	S	R-sq	R-sq(adj)	R-sq(pred)	T	σ	F
Se , MW	0.343651	73.90%	72.74%	68.27%	-9.52	0.000	63.70
DP02, Mp	0.269086	84.00%	83.29%	80.82%	-13.65	0.000	118.10
RDF080m, Mv	0.296885	80.52%	79.65%	77.57%	-8.68	0.000	93.00
DP03, R3v+	0.222760	89.03%	88.55%	87.55%	-16.74	0.000	182.66
SP05, Ss	0.37293	69.26%	67.89%	64.19%	-7.68	0.000	50.70
AMW, nSK	0.317295	77.75%	76.76%	72.98%	-11.17	0.000	78.62
AAC, nF	0.644612	8.16%	4.08%	0.00%	-0.69	0.147	2.00

Après avoir essayé toutes les variantes possibles compatibles avec les conditions expérimentales En fonction des résultats obtenus à ce stade :

- Notez qu'il y a une valeur significative dans ($R\text{-sq}=0.8903$) au niveau de la propriété (DP03, R3v+) et La valeur la plus basse en ($S= 0.222760$)
- Équation de régression linéaire pour un meilleur résultat (DP03, R3v+) :

$$\log iGc50 = -3.837 + 0.8856 \text{ DP03} + 18.83 \text{ R3v+} \quad (2)$$

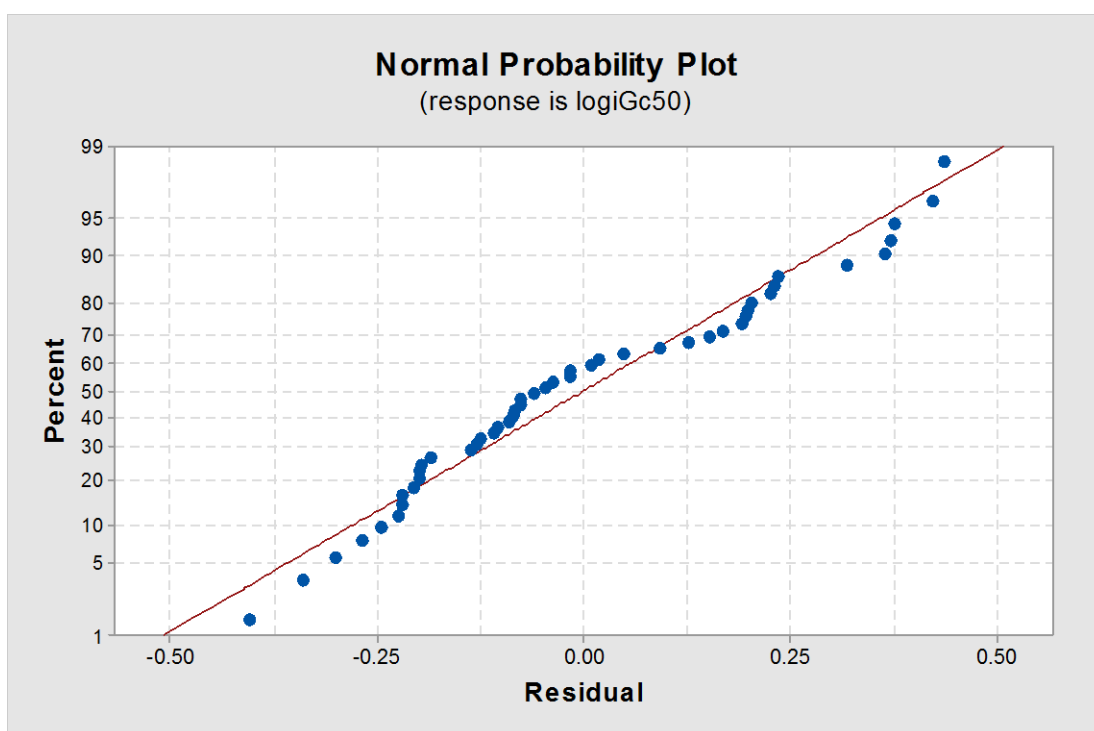


Figure III.9 : Courbe graphique Le pourcentage d'erreur pour chaque point (deuxième étape)

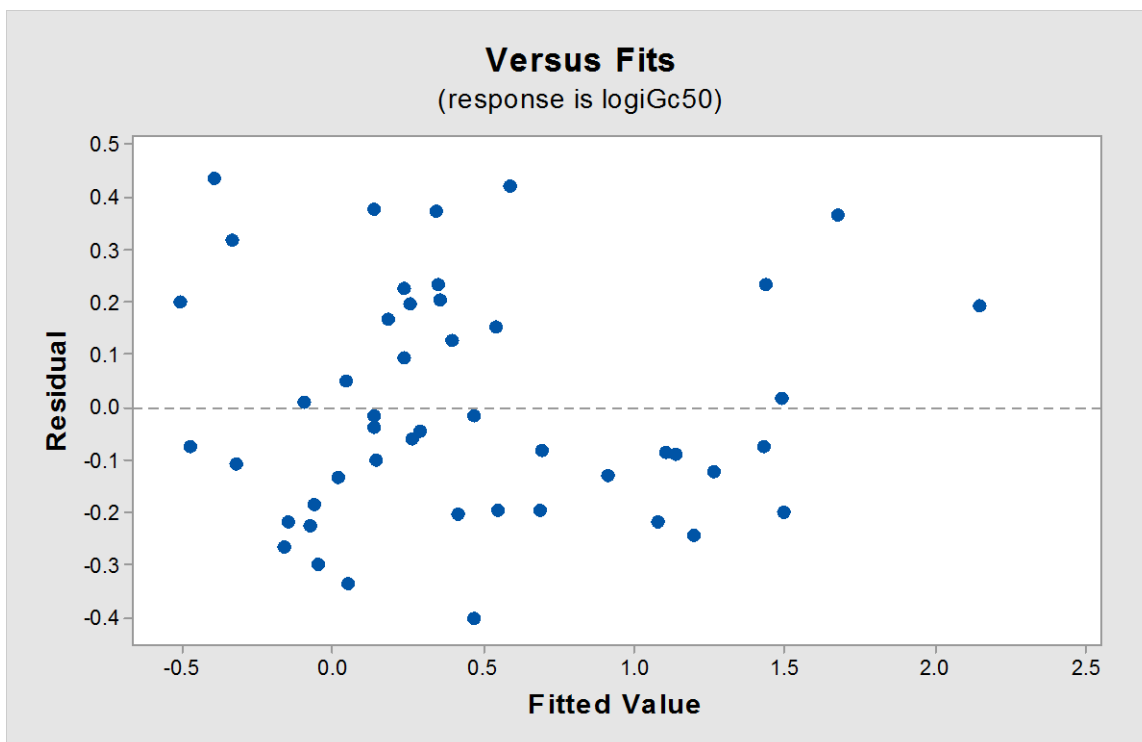


Figure III.10 : Courbe graphique L'erreur en fonction de la valeur révisée (deuxième étape)

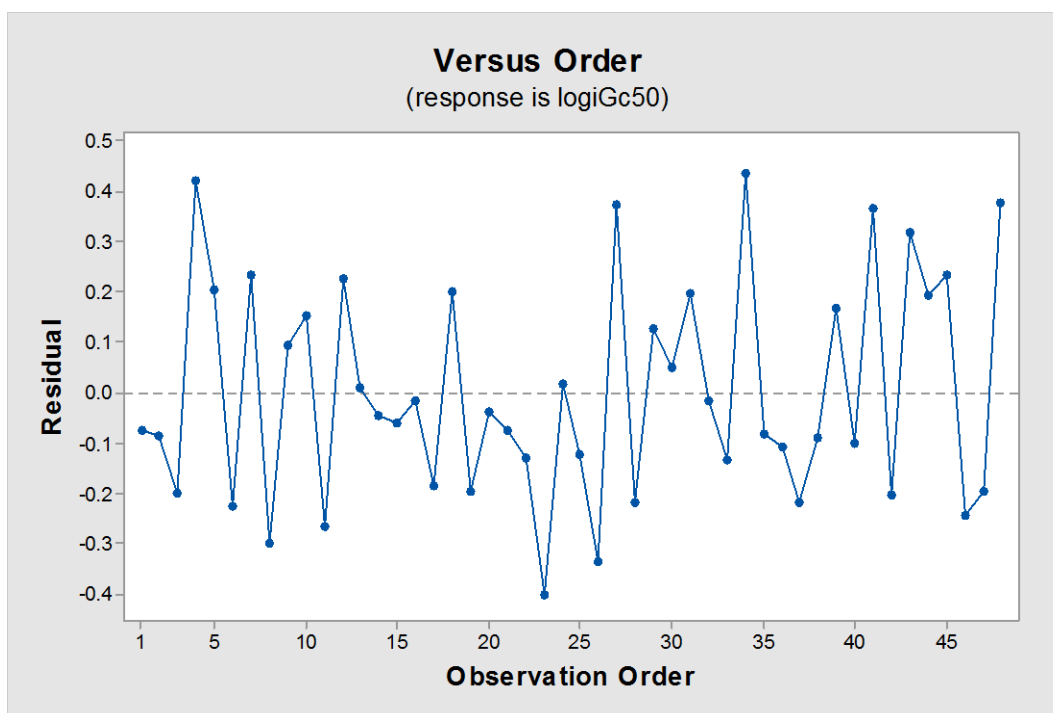


Figure III.11 : Courbe graphique Le pourcentage d'erreur pour chaque point (deuxième étape)

Tableau III.7 : ajustements et diagnostics pour les observations inhabituelles (deuxième étape)

logiGc50	Resid	Std Resid
0.0400	0.4354	2.02 R
2.3400	0.1918	1.01 X

III.4.1 Analyse des résultats de la deuxième étape

Le facteur de détermination est l'un des paramètres les plus courants décrivant la précision des résultats (plus le facteur de détermination 1 est proche, plus les résultats sont précis)

Dans ce cas le meilleur résultat obtenu : $R = 0.8903$ et $S = 0.222760$

III.4.2 Résultats de l'analyse des résidus de la deuxième étape

$$\text{résidus} = 1 - 0.8903 = 0.1097$$

Ce rapport est consacré au groupe de composés organiques qui ne s'appliquent pas à ce modèle, ce qui représente un ratio de 0,1097. Et il a besoin de nombreuses explications (son effet sur la croissance cellulaire dépend d'autres caractéristiques).

III.4.3 Discutez des résultats de la deuxième étape

Dans cette étape, nous nous concentrerons sur les composés les moins corrélés avec le modèle et ceux qui sont les plus étroitement liés.

III.4.3.1 Les composés les moins corrélés avec le modèle (deuxième étape)

Tableau 0III.8 : Tableau des valeurs des composés moins liés au modèle (deuxième étape)

Composés organiques	logiGc50	Resid
3-fluoro aniline	0.04	2.02 R
4-octyl aniline	2.34	1.01 X

A partir de ce tableau, il est montré que ces deux composés ne correspondent pas au modèle étudié, car leur toxicité ne correspond pas aux caractéristiques étudiées (Le taux d'erreur est élevé)

III.4.3.2 Les composés les plus corrélés avec le modèle de la deuxième étape

Tableau III.9 : Tableau des valeurs des composés plus liés au modèle (deuxième étape)

Composés organiques	logiGc50	Resid
2-chloro aniline	-0.09	0.008566
2-chloro-6-methyl aniline	0.12	-0.015189
3-4-5-trichloro aniline	1.51	0.017891
3-chloro-4-methyl aniline	0.45	0.016498
2-isopropyl aniline	0.1	-0.037511

III.4.3.3 Une comparaison entre les composés les plus compatibles et les moins compatibles (deuxième étape)

À partir du tableau III.4 nous constatons que les composés aniliques les plus compatibles avec l'équation de régression linéaire de la variable (DP03, R3v+) sont des composés contenant du chlore et le moins toxique, et à partir du tableau III.3.1 composés aniliques les moins liés à la forme sont des composés contenant octyl et présentant une toxicité élevée.

III.5 Validation les résultats

Ce point se concentre sur la mesure dans laquelle le modèle correspond aux résultats, autrement dit, nous allons répéter les calculs et les comparer avec les résultats pour déterminer l'étendue de la convergence et la précision du modèle

III.5.1 Le premier modèle (MW)

*Validation intérieur:

Un modèle QSAR peut être obtenu par le hasard, pour cette raison, il faut toujours s'assurer de sa stabilité. pour la vérification interne, nous avons utilisé 32 des composés étudiés (2-3-4-trichloro aniline, 2-3-chloro aniline, 2-4-5-trichloro aniline, 2-4-6-trichloro aniline, 2-4-dichloro aniline, 2-4-dimethyl aniline, 2-5-dichloro aniline, 2-5-dimethyl aniline, 2-6-dichloro aniline, 2-6-dichloro-3-methyl aniline, 2-6-dimethyl aniline, 2-bromo aniline, 2-chloro aniline, 2-chloro-4-methyl aniline, 2-chloro-5-methyl aniline, 2-chloro-6-methyl aniline, 2-ethyl aniline, 2-fluoro

aniline.2-iodo aniline.2-isopropyl aniline.2-methyl aniline.2-phenyl aniline.2-propyl aniline.3-4-5-trichloro aniline.3-4-dichloro aniline.3-4-dimethyl aniline.3-5-dichloro aniline.3-5-dimethyl aniline.3-bromo aniline.3-chloro aniline.3-chloro-2-methyl aniline.3-chloro-4-methyl aniline)

. **LOO Leave one out** : La procédure LOO (mettre une molécule à parts) a donné une valeur moyenne $R_{adj}^2 = 62.99\%$ Cette valeur est loin de la valeur de $R^2 = 72.16\%$ ce résultat montre que notre modèle QSAR est sensible à ce processus de mise à jour, le modèle explique la toxicité de certains composés chimiques uniquement. Cette preuve est actuellement suffisante pour validation interne.

***Validation extérieur:**

Pour la vérification externe, nous avons utilisé 16 composés aniliques (3-ethyl aniline.3-fluoro aniline.3-iodo aniline.3-methyl aniline.3-phenyl aniline.4-butyl aniline.4-chloro-2-methyl aniline.4-ethyl aniline.4hexyl aniline.4-isopropyl aniline.4-methyl aniline.4-octyl aniline.4-pentyl aniline.4-phenyl aniline.4-propyl aniline.5-chloro-2-methyl aniline)

Dans cette étape, nous faisons ce qui suit :

•Comparaison des paramètres statistiques

Tableau III.10 : Valeurs des paramètres pour la vérification extérieur de MW

Les variables	la valeur
s	0.4481
R^2	0.58
R^2_{adj}	0.5793
Q^2	0.0780
PRESS	14.7574
F	19.28
d Durbin-w	1.47
SDEC	0.44
SDEP	0.64

• Vérification des critères de Tropsha

$$R_0^2 = 0.6299$$

$$\frac{R^2 - R_0^2}{R^2} = 0.000536 \quad , \quad |R^2 - R_0^2| = 0.8 > 0.3$$

En plus de la vérification interne, qui a donné des résultats faibles. La vérification externe n'a pas donné de bons résultats. Par conséquent, ce modèle n'est généralement pas applicable, expliquez simplement quelques cas particuliers

III.5.2 Validation la deuxième modèle (DP03, R3v+) :

On fait validation de la même manière dans le premier modèle

*Validation intérieur :

• **LOO Leave one out** : La procédure LOO (mettre une molécule à parts) a donné une valeur moyenne $R_{adj}^2 = 85.25\%$ la valeur est très proche de la valeur de $R^2 = 87.41\%$ ce résultat montre que notre modèle QSAR est n'est pas sensible à ce processus de mise à jour, le modèle explique la toxicité composés chimiques. Cette preuve est actuellement suffisante pour validation interne.

*Validation extérieur :

- **Comparaison des paramètres statistiques**

Tableau III.11 : Valeurs des paramètres pour la vérification extérieure de (DP03, R3v+)

Les variables	la valeur
s	0.3064
R ²	0.89
R ² adj	0.8857
Q ²	0.7002
PRESS	4.7984
F	67.920
d Durbin-w	1.93
SDEC	0.30
SDEP	0.37

- **Vérification des critères de Tropsha**

$$R_0^2 = 0.8902$$

$$\frac{R^2 - R_0^2}{R^2} = 0.0161 < 0.1 \quad , \quad |R^2 - R_0^2| = 0.28 < 0.3$$

Tous les critères de Tropsha sont vérifiés Sur la base des résultats obtenus lors de la vérification externe et vérification interne. Les résultats de la validation externe montrent que le meilleur modèle est hautement prédictif (DP03, R3v+) et peut être utilisé pour mesurer la toxicité de nouveaux composés aniliques.

III.6 Coefficient de partage log (P)

En plus des résultats de toxicité des composés étudiés. Coefficient de partage il est très important de connaître l'étendue de la solubilité et de la perméabilité de ces composés dans les solutions organiques et minérales (en général, les solutions minérales que nous utilisons de l'eau). Coefficient de partage est Connu comme la loi :

$$\log (P) = \text{Log}(\text{Coct}/\text{Ceau}) \quad (3)$$

Coct : Concentration d'octanol (Le milieu organique)

Ceau : Concentration de la solution minérale (l'eau)

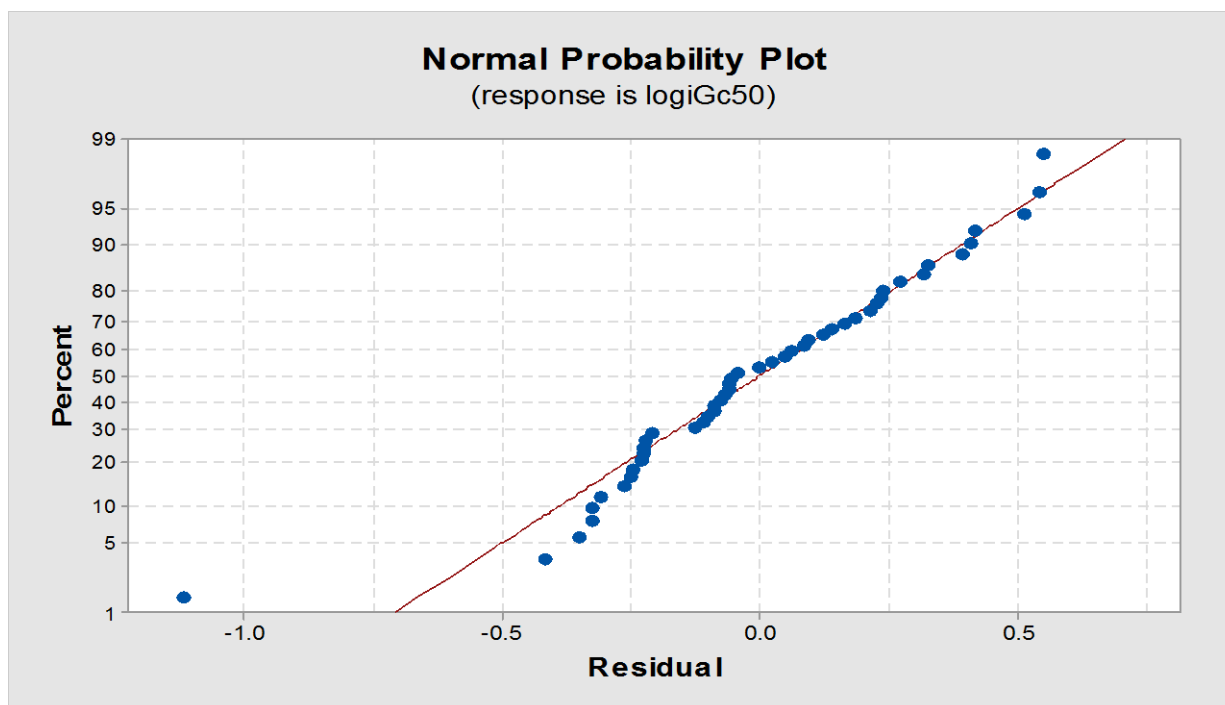


Figure III.12 : Courbe graphique Pour log iGc50 en termes de log(P)

III.6.1 Analyse des résultats

Le groupe aniline contient du chlore avec un module de partage faible à élevé et le groupe contenant du méthyle a un facteur de partition moyen d'après les différents résultats, la 4-octyl-aniline a un coefficient de partage très élevé .

En plus des résultats du coefficient de partage, nous construisons une équation de régression linéaire multiple pour $\log iGc50 = a + b \text{ Log}(P)$.

Avec A et b Constantes

Nous avons obtenu de bons résultats

S	R-sq	R-sq(adj)	R-sq(pred)	F	σ
0.306965	78.71%	78.25%	76.93%	170.08	0.000

$$\log iGc50 = -1.549 + 0.7982 \text{ Log}(P) \quad (18)$$

- **Analyse des résidus**

$$\text{résidus} = 1 - 0.7871 = 0.2129 = 21.29\%$$

III.6.2 Discuter des résultats

Les valeurs du coefficient de partage changent avec la stéréoisomérie des composés organiques (Bien que le nombre d'atomes dans le composé soit égal) et toute addition d'un atome au composé affecte directement les valeurs des coefficients de partage.

Pour les résultats de l'équation de régression linéaire multiple il y a trois composés organiques avec une erreur élevée 3-méthyl aniline (-3.68 R) et 4hexyl aniline (0.80 X) et 4-octylaniline (-0.39 X), de plus, nous constatons que les composés sont plus compatibles avec l'équation de régression linéaire multiple sont des composés contenant du chlore, phenyl ou éthyle.

Conclusion générale

Après avoir étudié et effectué les calculs des modèles (1) et (2), nous avons obtenu la relation suivante :

$\text{LogiGc50} = a + \sum (b_j \times x_j)$, où ce modèle était meilleur parmi les autres auxquels ses résultats sont les suivants ($R = 0,5988$) ($S = 0,4214$) ($F = 102,5$) pour modèle (1) et ($R=0.8903$) ($S=0.222760$) ($F=182.66$) pour le deuxième modèle.

Aussi, nous avons noté que le coefficient de partage affecte grandement les cellules à travers la quantité adsorbée, et avec l'effet d'autres propriétés, il pourrait atteindre un stade dangereux. De plus, et afin d'obtenir un modèle QSAR, nous avons utilisé la méthode de régression multiple où nous avons obtenu de bons résultats : $S=0,306965$ $R=78,71$ $F=170.08$

Sur la base de ces résultats, on peut conclure ce qui suit :

- L'étude par la technique QSAR/QSPR en utilisant la méthode (MLR) est une solution fiable, car elle permet aux chimistes de trouver des valeurs de Log(P)_w aux composés inconnus, notamment des composés chimiques dangereux.
- La capacité prédictive dépend du type et du nombre de composés chimiques dans la base de données. Et pour améliorer les résultats, on peut procéder à :
 - ✓ Collecter le plus grand nombre des composés à étudier,
 - ✓ Effectuer une optimisation complète des composés afin d'atteindre le degré maximum de stabilité.
 - ✓ Réaliser une modélisation de composés organiques n'ayant pas fait l'objet d'étude auparavant.

Bibliographie

- [1] B. K. ERRAHOUÏ née, Etude des relations quantitatives structure–toxicité des composés chimiques à l’aide des descripteurs moléculaires. « Modélisation QSAR », UNIVERSITE ABOU BEKR BELKAÏD DE TLEMCEN, 2015.
- [2] O. B. S. H. A. A. L. K. C. S. M. Mabrouk Hamadache, « A Quantitative Structure Activity Relationship for acute oral toxicity of pesticides on rats: Validation, Domain of Application and Prediction,» *Journal of Hazardous Materials, Elsevier*, pp. 28-40, 2016.
- [3] H. Kubinyi, QSAR: Hansch analysis and related approaches. Page: 4.
- [4] H. Kubinyi, 3D QSAR in Drug Design: Volume 1: Theory Methods and Applications, Ludwigshafen, Germany.
- [5] S. Chtita, Modélisation de molécules organiques hétérocycliques biologiquement actives par des méthodes QSAR/QSPR. Recherche de nouveaux médicaments, Maroc: Uni-versité Moulay Ismail – Présidence, 2017.
- [6] R. Carbó-Dorca, D. Robert, L. Amat, X. Gironés et E. Besalú, Molecular Quantum Similarity in QSAR and Drug Design..
- [7] K. Roy, «Advances in QSAR Modeling: Applications in Pharmaceutical, Chemical, Food, Agricultural and Environmental sciences,» pp. 57-58.
- [8] K. A. Benazzouz Hicham, Relation Structure Activité : Etude Qualitative et Quantitative et Développement de Recherche sur les Coumarines, Université Abou Bekr Belkaïd Tlemcen.
- [9] T. P. J. L. M. T. C. Mark T. D. Cronin, Recent Advances in QSAR Studies: Methods and Applications, Page: 4.
- [10] K. A. Benazzouz Hicham, Relation Structure Activité : Etude Qualitative et Quantitative et Développement de Recherche sur les Coumarines., Université Abou Bekr Belkaïd Tlemcen.
- [11] A. Fortuné, «Techniques de Modélisation Moléculaire appliquées à l'Etude et à l'Optimisation de Molécules Immunogènes et de Modulateurs de la Chimiorésistance,» *Université Joseph-Fourier-Grenoble I*, , 2006.
- [12] E. H. H. O. L.M. Arterburn, « Distribution, interconversion, and dose response of n– 3 fatty acids in humans,» *The American journal of clinical nutrition*, 83 , pp. 1467S-1476S, 2006.
- [13] M. C. T. S. J. M. S. Enoch, « An evaluation of global QSAR models for the prediction of the toxicity of phenols to *Tetrahymena pyriformis*,» *Chemosphere*, 71, pp. 1225-1232, 2008.
- [14] H. W. J. A. C. Soc., 69, 17, 1947.
- [15] M. R. J. A. C. Soc., «,» 97, 6609, 1975.
- [16] L. H. W. M. J. P. S. L.B. Kier, 64,1971, 1975.

- [17] A. B. C. P. Lett., 89, 399, 1982.
- [18] B. Efron, «Estimating the error rate of a prediction rule: improvement on cross- validation,» *Journal of the American statistical association*, vol. 78, pp. 316-331, 1983.
- [19] M. Efroymson, « Multiple regression analysis,» *Mathematical methods for digital computers*, pp. 191-203, 1960.
- [20] «B. Lee, F.M. Richards, The interpretation of protein structures: estimation of static accessibility,» *Journal of molecular biology*, vol. 55, pp. 379-IN374, 1971.
- [21] P. Rogers, «Genetics and biochemistry of Clostridium relevant to development of fermentation processes,» *Advances in applied microbiology, Elsevier*, pp. 1-60, 1986.
- [22] T. N. D. I. L. T. G. J. W. L. J.B. Ball, « Computer-aided mapping of the. beta.-adrenoceptor I: Explanation for effect of para substitution on blocking activity at the. beta.-1-adrenoceptor,» *Journal of medicinal chemistry*, vol. 35 , pp. 4676-4682, 1992.
- [23] A. G. G. Crippen, « Atomic physicochemical parameters for three-dimensional- structure-directed quantitative structure-activity relationships. 2. Modeling dispersive and hydrophobic interactions,» *J Chem Inf Comput Sci*, vol. 27 , pp. 21-35, 1987.
- [24] H. Z. T. O. A. G. A. T. L. Zhang, «QSAR Modeling of the Blood-Brain Barrier Permeability for Diverse Organic Compounds,» *Pharm. Res.*, vol. 25, pp. 1902-1914, 2008.
- [25] P. J. L. He, « Assessing the reliability of a QSAR model's predictions,» *J. Mol. Graph. Model*, vol. 23 , pp. 503-523, 2005.
- [26] P. G. V. G. A. Tropsha, «The importance of being Earnest: Validation is the absolute essential for successful application and interpretation of QSPR models,» *QSAR Comb.Sci*, vol. 22, pp. 69-77, 2003.
- [27] S. P. I. M. K. R. P.P. Roy, «On Two Novel Parameters for Validation of Predictive QSAR models,» *Molecules*, vol. 14, pp. 1660-1701, 2009.
- [28] A. T. A. Golbraikh, « Beware of q2! J. Mol. Graph. Model.,» vol. 20, p. 269–276, 2002 .
- [29] P. H. D. M. Y. E. M. A. G. H. Z. a. A. T. M. Martin, « Does Rational Selection of Training and Test Sets Improve the Outcome of QSAR Modeling?,» *J. Chem. Inf. Model* , vol. 52, p. 2570–2578, 2012.
- [30] O. FADIA, ETUDE QSAR DE LA TOXICITE D'UNE SERIE DE, TLEMCEN: UNIVERSITE ABOU-BEKR BELKAID , 2019.
- [31] M. K. S. S. a. C. D. H. A. R. Katritzky, «Quantitative Correlation of Physical and Chemical Properties with Chemical Structure: Utility for Prediction,» *Chem. Rev*, vol. 110, p. 5714–5789, 2010.