



MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR  
ET DE LA RECHERCHE SCIENTIFIQUE  
UNIVERSITÉ «Abbés LAGHROUR» DE KHENCHELA  
FACULTÉ DES SCIENCES ET TECHNOLOGIE



Département des Sciences de la Matière

N° de série :.....

## Mémoire de fin d'études

*Pour l'obtention du diplôme de Master (L.M.D)*

**Filière : Chimie.**

**Spécialité : Chimie Analytique et Environnement**

**Intitulé:**

***Etude QSPR de la température d'ébullition  
des anilines***

*Réalisé par :*

*-OUCIF Hafida*

*-LOUCIF Sara*

*Dirigé par : KERTIOU Noureddine*

*Membres de jury :*

*HAKKAR Farida*

*MAA*

*U.A.L. Khenchela*

*Présidente*

*KERTIOU Noureddine*

*MCB*

*U.A.L. Khenchela*

*Encadreur*

*BOUAKKADIA Amel*

*MCB*

*U.A.L. Khenchela*

*Examinatrice*

Année universitaire 2019/2020

## ***Dédicaces***

*Et je dédie ce modeste travail à :*

*Mon père et ma mère; qui mon donné tous qu'ils peuvent.*

*A mes frères et mes sœurs,*

*A toute ma famille,*

*-A mes amis :(Sara- Manel – Souhila-...),*

*à tous mes amis de promotion.*

*A tous les enseignants de département des Sciences de la*

*Matière a l'université de kenchela*

*Merci beaucoup.*

***O. Hafida***

## ***Dédicaces***

*Et je dédie ce modeste travail à :*

*Mon père et ma mère; qui mon donné tous qu'ils peuvent.*

*A mes frères et mes sœurs,*

*A toute ma famille,*

*-A mes amis :(Hafida - Manel – Souhila-...),*

*à tous mes amis de promotion.*

*A tous les enseignants de département des Sciences de la*

*Matière a l'université de kenchela*

*Merci beaucoup.*

***L. sara***

# Remerciements

*Avant tout, nous tenons à remercier«Allah» le tous puissant, pour nous avoir donné la force et la patience.*

*Nous tenons à exprimer nos profondes gratitudees à :  
M' l'encadreur : KERTIOU Noureddine pour avoir dirigé ce mémoire,  
avec un encadrement sérieux et efficace.*

*Nous remercie également tous les membres du jury d'avoir accepté d'examiner ce travail.*

*Et enfin je remercié l'ensemble des enseignants et collègues de notre  
promotion.*

# SOMMAIRE

<b>TITRE</b>	<b>PAGE</b>
LISTE DES FIGURES.....	I
LISTE DES TABLEAUX .....	II
Symboles et abréviations.....	III
<b>Introduction Générale</b>	
Introduction générale .....	02
<b>Partie Théorique</b>	
I- GENERALITE.....	05
I-1-Introduction.....	05
I -2- Les relations structure-propriétés quantitatifs <QSPR>.....	05
I -2-1-Définition de la méthode QSPR.....	05
I-2-2- Principe de la méthode QSPR.....	05
I-2-3- Méthodologie générale d'une étude QSPR/QSAR.....	06
I-2-4- Les méthodes mathématiques utilisés par le model QSPR.....	07
I-3- Les descripteurs moléculaires.....	08
I-3 -1- Définition d'un descripteur .....	08
I-3-2- Groupe des descripteurs moléculaires.....	08
I-3-3- Exemple : calcul de ATSKw.....	10
I-3-4 Bloc des descripteurs.....	10
I-3-5-Le choix des descripteurs.....	11
I.3.6. Importance des descripteurs.....	12
I.4. Diagramme montre les étapes de prédiction.....	12
I-5- L'objectif de la prédiction.....	13
II- COLLECTE DES DONNEES.....	13

III-OPTIMISATION DE LA GEOMETRIE MOLECULAIRE ET GENERATION DES DESCRIPTEURS MOLECULAIRES.....	16
III-1 -Préparation de base des données.....	16
III-1-1-Logiciels « ChemDraw 7 ».....	16
III-1-2-Logiciels « HyperChem ».....	16
III-2-Récupération et stabilisation des molécules.....	17
III-2-1- Stabilisation structure des molécules (minimisation de l'énergie).....	17
III-2-2- Mécanique moléculaire.....	17
III-2-3- Récupération des fichiers.....	18
III-3- Le Logiciel « Dragon ».....	18
IV-Méthodes de régressions linéaires et multilinéaire.....	19
IV-1 Aperçu général.....	19
IV-2- Evaluation préliminaire des données.....	20
IV-3- Régression linéaire multiple.....	20
IV-4- Méthodes de sélection des descripteurs.....	22
IV-4-1- Régression pas à pas.....	22
IV-4.2-Algorithmes génétiques.....	22
IV-5 – Paramètres d'évaluation de la qualité de l'ajustement.....	22
• Le coefficient de détermination multiple.....	22
• La racine de l'erreur quadratique moyenne de prédiction .....	23
VI-6-Facteur d'inflation de la variance [FIV].....	23
VI-7- Test de randomisation.....	23
VI-8 - Validation externe.....	24
V- Généralités sur les anilines .....	25
V-1- Aniline.....	25
V-2- Historique.....	25
V-3- Production de l'aniline.....	26

V-4- Présence de l'aniline dans l'environnement.....	26
V-5- Utilisations.....	27
V-6- Propriétés Chimiques.....	27
V-7- Propriétés Physiques.....	27
V-8- Toxicité.....	28
IV-Température d'ébullition (Teb).....	29

## Partie Application

I-Calcul du modèle.....	31
I-1- Choix du modèle .....	31
I-2- Choix de la taille du modèle.....	31
I- 3- Choix d'une famille de descripteurs.....	32
I-4- Calcul des corrélations entre les différents descripteurs .....	34
I-5 Analyse de régression.....	35
II-Diagramme de Williams.....	36
III- Evaluation de la qualité de l'ajustement.....	37
VI- Test de randomisation.....	38
VII- Validation externe.....	39

## Conclusion Générale

Conclusion générale.....	41
--------------------------	----

## REFERENCES BIBLIOGRAPHIQUES

Références Bibliographiques.....	43
ANNEXE.....	48
RESUMES.....	52

## LISTE DES FIGURES

N°	FIGURE	PAGE
Figure 01	Principe de la méthode QSPR	06
Figure 02	Procédure d'obtention et de validation d'un modèle QSPR/QSAR	07
Figure 03	Représentation des blocs des descripteurs moléculaire.	11
Figure 04	Diagramme de prédiction par QSPR	13
Figure 05	Représentation des molécules par Chem Draw	16
Figure 06	Application de la méthode Field Force	18
Figure 07	Interface principale du logiciel DRAGON	19
Figure 08	Structure de l'aniline	25
Figure 09	Variation de $R^2$ en fonction de la taille des modèles k	31
Figure 10	Diagramme de Williams des composés.	37
Figure 11	Graphe des valeurs prédites $T_f$ en fonction des valeurs observées.	38
Figure 12	Test de randomisation associé au modèle QSPR.	39

## LISTE DES TABLEAUX

<b>N°</b>	<b>TABLEAU</b>	<b>PAGE</b>
Tableau 01	Quelques blocks des descripteurs calculés par logiciel dragon.	08
Tableau 02	Nomenclature et valeurs de propriété des anilines étudiées	14
Tableau 03	Principales caractéristiques physiques	28
Tableau 04	Classes et signification des descripteurs	32
Tableau 05	Valeurs des descripteurs moléculaires sélectionnés	33
Tableau 06	Matrice de corrélations.	35
Tableau 07	Analyse de régression	35
Tableau 08	Les paramètres statistiques pour l'ensemble de calibration (35 composés) et celui de validation (15 composés).	36

## Symboles et abréviations

AG:	Algorithme génétique (Genetic Algorithm).
$e_i$ :	Résidu différence entre les valeurs observée ( $y_i$ ) et estimée ( $\hat{y}_i$ ).
F :	Statistique de Fisher.
$h_{ii}$ :	Eléments diagonaux de la matrice chapeau.
LOO:	Validation croisée par omission d'une observation (leave-one-out)
MM+ :	Mécanique Moléculaire
MLR:	Régression linéaire multiple.
n:	Dimension de la population (échantillon).
n-p :	Nombre de degrés de liberté.
p :	Nombre de descripteurs en comptant la constante (Nombre de paramètres).
PCR :	Régression en composantes principales.
PLS:	Moindres carrés partiels.
PRESS :	Somme des carrés des erreurs de prédiction.
Q <sup>2</sup> LOO :	Coefficient de prédiction.
QSAR :	Relations Activité /Propriétés quantitatives (Quantitative Activity/ Propriety Relationships)
QSPR	Relations Structure /Propriétés quantitatives (Relationships Quantitative Structure/ Propriety)
R <sup>2</sup> :	Coefficient de détermination.
REACH :	Le règlement européen (Registration, Evaluation, Authorisation and restriction of Chemicals)
RNA :	Réseaux de neurone artificiel
s :	Erreur standard.
SCE :	Somme des carrés des écarts.
SCT :	Somme des carrés totale.
SDEC:	Déviati on standard de l'erreur calculée. (Standard Deviation Error in Calculation)
SDEP:	Déviati on standard de l'erreur de prédiction (Standard Déviati on Error of Prediction)
SDEPext:	Deviation standard de l'erreur de prediction externe (External Standard Deviation Error of Prediction:
T <sub>eb</sub> :	Température d'ébullition

$y$ :	Vecteur de dimension $n$ .
$y_i$ :	Valeur observée.
$\hat{y}_i$ :	Valeur estimée.
$\alpha$ :	Niveau de confiance.

**Introduction**

**Générale**

## **Introduction générale :**

Les anilines sont parmi les composés industriellement importantes dans la perspective du génie chimique, car sont utilisées dans la fabrication des colorants, le caoutchouc, et les herbicides, mais d'autre part sont toxique pour la santé par contact avec la peau ou par ingestion. La détermination expérimentale systématique au laboratoire des données concernant ces substances chimiques se traduit par une lourde charge, économiquement inacceptable, pour les industriels et les autorités de régulation, et dépasse de loin, de toutes les façons, les capacités de recherche disponibles, dans ce concept le recours aux méthodes d'estimations, qui permettent un gain de temps et d'argent, est incontournable.

La température d'ébullition peut être mesurée facilement, et sa prédiction reste d'une portée limitée. Cependant, c'est la propriété physico-chimique la plus utilisée dans les exercices de modélisation.

Les relations structure/propriétés quantitatives, désignées par l'abréviation QSPR (Quantitative Structure/Property Relationships), très utilisées depuis une vingtaine d'années, constituent des modèles mathématiques pour l'approximation des relations, souvent complexes, entre la structure caractérisée par des descripteurs moléculaires, et les propriétés physico-chimiques des composés.

Les techniques les plus courantes pour établir des modèles QSPR utilisent l'analyse de régression (régression multilinéaire : MLR (utilisée dans ce travail); projection des structures latentes par les moindres carrés partiels : PLS ), les réseaux neuronaux RNA, et les méthodes de classification.

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 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.

Nous avons appliqué une méthode hybride algorithme génétique / régression multilinéaire : AG/MLR , pour modéliser la température d'ébullition d'une série d'anilines.

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: régression multilinéaire, traitement statistique pour l'évaluation de la qualité de l'ajustement (robustesse des modèles ; validation externe).

Dans la partie application, nous présentons et discutons le modèle calculé : AG/MLR, pour modéliser la température d'ébullition; Discussion des résultats.

# **Partie**

# **Théorique**

## **I- GENERALITE**

### **I-1-Introduction**

Les relations quantitatives structure-activité/propriété (QSPR/QSAR) sont de plus en plus utilisées, du fait de la croissance des moyens de calculs. Très récemment, la mise en place du nouveau règlement européen REACH, qui recommande leur utilisation pour limiter le recours à l'expérience, donne un nouvel essor au développement de tels modèles prédictifs.

Dans les dernières années, l'utilisation des méthodes QSAR n'a cessé de progresser. Elle est même devenue indispensable en chimie pharmaceutique et pour la conception de médicaments. Leur développement dans une gamme plus large d'applications, leur ouvre d'ailleurs de grandes perspectives (ex : points d'ébullition, températures critiques, pressions critiques, densité...etc.).

Il s'agit de présenter ici le principe des modèles QSPR ainsi que ceux des différents outils employés pour leur mise en place et leur évaluation : bases de données expérimentales, descripteurs, outils d'analyse de données [1].

### **I -2- Les relations structure-propriétés quantitatives <QSPR>**

#### **I -2-1-Définition de la méthode QSPR**

Le QSPR (**Q**uantitative **S**tructure-**P**roperty **R**elationships) est le procédé par lequel des liens quantitatifs sont établis entre la structure moléculaire d'un ensemble de composés avec une propriété physico-chimique. Les grandes phases de développement d'un modèle QSPR peuvent être décrites comme suit :

- Choisir des descripteurs adaptés au problème structure-propriété,
- Exploiter les valeurs des descripteurs comme variables, afin de définir une relation qui les corrèle à la propriété en question, à l'aide de machines d'apprentissage. C'est la fouille de données.
- Établir des critères de performance et de validation qui aideront au choix des meilleurs modèles pour le problème posé et estimer des incertitudes de prédiction [2].

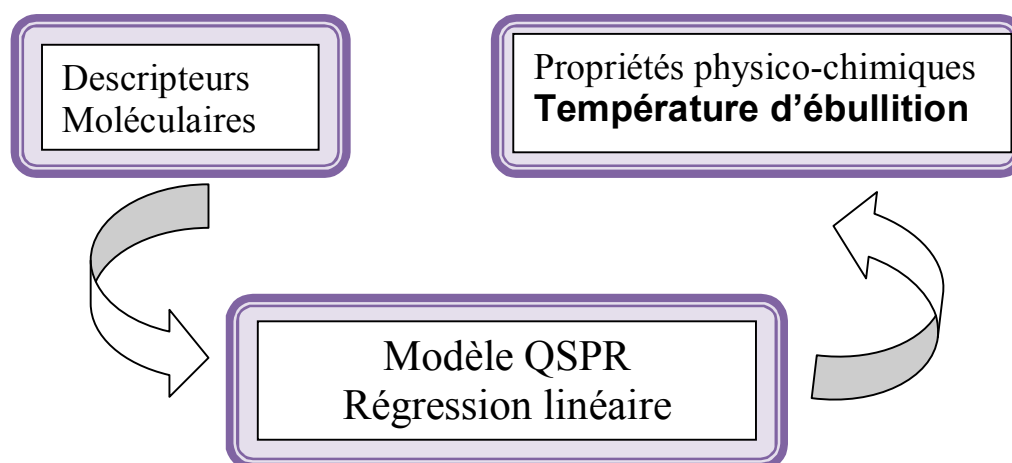
#### **I-2-2- Principe de la méthode QSPR**

Le principe des méthodes QSPR est, comme leur nom l'indique, de mettre en place une relation mathématique reliant de manière quantitative des propriétés moléculaires aussi bien électroniques que géométriques, appelées descripteurs, avec une observable macroscopique

(activité biologique, toxicité, propriété physico-chimique, etc.), pour une série de composés chimiques similaires à l'aide de méthodes d'analyses de données. Aussi la forme générale de modèle est:

$$\text{Propriété} = f(\text{Descripteurs}) \quad (01)$$

L'objectif d'une telle méthode est donc d'analyser les données structurales afin de détecter les facteurs déterminants pour la propriété mesurée. Pour ce faire, différents types d'outils peuvent être employés : régressions multilinéaires (MLR), régressions aux moindres carrés partiels (PLS), arbres de décision, réseaux de neurones, algorithmes génétiques. Une fois cette relation mise en place et validée sur un jeu validation, elle peut alors être employée pour la prédiction de la propriété de nouvelles molécules, pour lesquelles la valeur expérimentale n'est pas disponible, voire pour des molécules encore non synthétisées. De tels modèles peuvent également, dans certains cas, être utilisés pour mieux appréhender les phénomènes moléculaires mis en jeu dans la propriété d'intérêt [3].



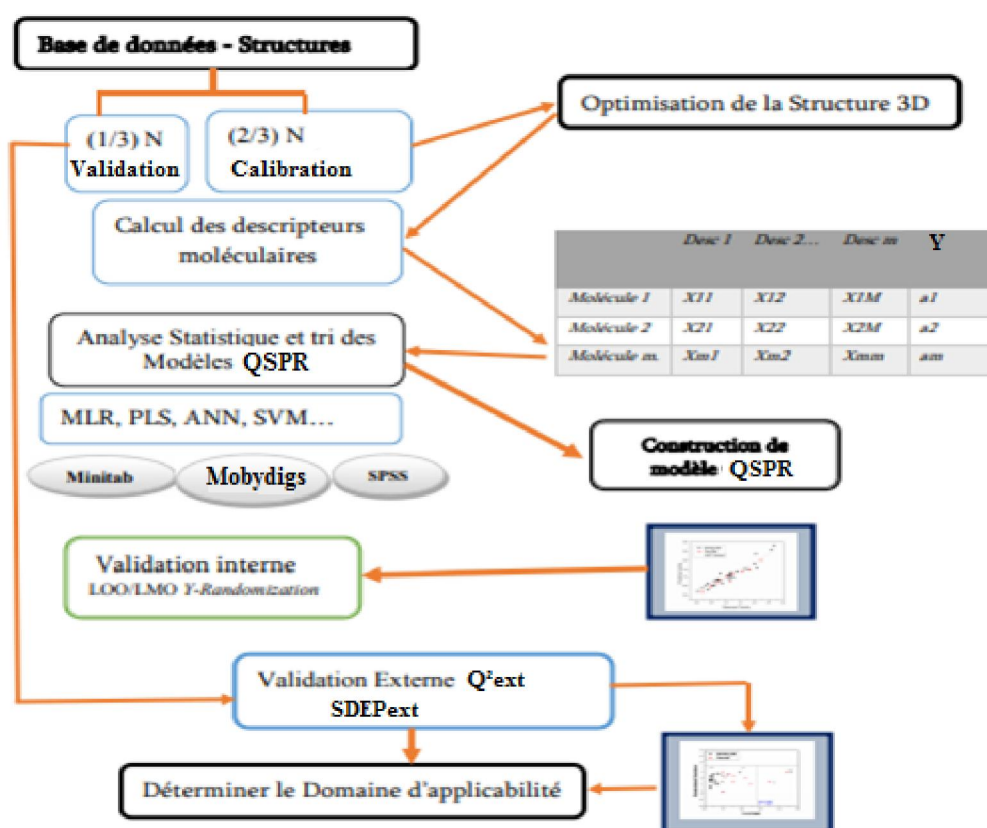
**Figure 01** Principe de la méthode QSPR

### I-2-3- Méthodologie générale d'une étude QSPR/QSAR [4]

La méthodologie générale d'une étude QSAR/QSPR est la suivante :

- a- Constituer une base de données à partir des mesures expérimentales fiables de la propriété ou de l'activité de chaque composé.
- b- Diviser cette base de données, aléatoirement, en une série d'apprentissage (training 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

- c- Sélectionner les descripteurs en relation avec la propriété ou l'activité étudiée.
- d- Etablir des modèles mathématiques en utilisant la série d'apprentissage.
- e- Caractériser les modèles élaborés par leurs indices de validation internes et vérifier leur robustesse par un test de hasardisation (Randomisation) de la variable dépendante Y (réponse).
- f- Valider les modèles élaborés en utilisant la série de test et calculer leurs paramètres statistiques de validation externe.
- g- Elaborer le domaine d'applicabilité du modèle retenu.
- h- Explorer et exploiter les modèles validés pour comprendre les mécanismes et les modes d'action.



**Figure 02** : Procédure d'obtention et de validation d'un modèle QSPR/QSAR

#### I-2-4- Les méthodes mathématiques utilisés par le model QSPR

- Régression linéaire
- Régression linéaire multiple (RLM)
- Régression en composantes principales(PCR)
- Régression moindre carré (PLS)

### **I-3- Les descripteurs moléculaires:**

#### **I-3 -1- Définition d'un descripteur :**

Un descripteur est un paramètre numérique propre à une structure chimique donnée, également c'est un résultat d'un procédé mathématique ou d'une certaine expérience normalisée. Ces descripteurs représentent un moyen de corréliser une structure chimique et une valeur de propriété physico-chimique ou biologique. Les descripteurs sont finalement des intermédiaires numériques se substituant à la molécule elle-même [5]. Ces descripteurs peuvent être utilisés pour évaluer l'influence de la structure moléculaire ou les relations propriétés-structure, aussi pour l'analyse de symétrie et la projection des bases de données des molécules [6].

#### **I-3-2- Groupe des descripteurs moléculaires**

En effet, les descripteurs moléculaires sont basés sur différentes théories, tel que le quantum chimique, la théorie de l'information, la théorie du graphe, et ainsi de suite, et également utilisés pour modéliser plusieurs propriétés de produits chimiques dans les différents domaines scientifiques tel que la toxicologie, la chimie analytique, la chimie physique...etc. [7].

Actuellement, il est possible de calculer plus de 1600 descripteurs moléculaires qui sont représentés dans le tableau suivant qui peuvent être classés en 20 classes (blocs)

**Tableau 01 :** Quelques blocks des descripteurs calculés par logiciel dragon [7]

<b>Classe</b>	<b>Sous- classe</b>
Descripteurs Constitutionnels	- Dénombrement des atomes ou des liaisons. - Descripteurs basés sur les masses atomiques.
Descripteurs Topologiques	- Indices topologiques (connectivité). - Descripteurs théoriques d'information. - Descripteurs topo-chimiques.

**Tableau 01 : Suite et fin**

<b>Classe</b>	<b>Sous- classe</b>
Descripteurs Géométriques	<ul style="list-style-type: none"><li>- Descripteurs liés à la distance.</li><li>- Descripteurs liés à l'aire de la surface.</li><li>- Descripteurs liés au volume.</li><li>- Descripteurs du champ stérique moléculaire.</li></ul>
Descripteurs liés à la distribution de charge	<ul style="list-style-type: none"><li>- Charges atomiques partielles.</li><li>- Moments électriques moléculaires</li><li>- Polarisabilités moléculaires.</li><li>- Descripteurs du champ électrique moléculaire.</li></ul>
Descripteurs liés aux orbitales Moléculaires	<ul style="list-style-type: none"><li>- Energie des OM frontières</li><li>- Ordres de liaison</li><li>- Indices de réactivité de Fukui.</li></ul>
Descripteurs température Dépendants	<ul style="list-style-type: none"><li>- Fonctions thermodynamiques.</li><li>- Descripteurs facteurs de Boltzmann pondérés.</li></ul>
Descripteurs de Solvation	<ul style="list-style-type: none"><li>- Energie électrostatique de solvation.</li><li>- Energie de dispersion de solvation.</li><li>- Enthalpie libre de formation de cavité.</li><li>- Descripteurs de liaison hydrogène.</li><li>- Entropie de solvation.</li><li>- Descripteurs d'énergie de solvation linéaire théorique.</li></ul>
Descripteurs mixtes	<ul style="list-style-type: none"><li>- Descripteurs topographiques.</li><li>- Descripteurs électro-topologiques.</li><li>- Descripteurs de la charge partielle de l'aire de la surface.</li></ul>

### I-3-3- Exemple : calcul de ATSKw

Broto-Moreau auto corrélation L'ATS kW, **W** étant la propriété atomique pesant le graphique moléculaire et k le retard, est évalué en considérant séparément toutes les contributions de chaque longueur de chemin différente (retard) dans le graphique moléculaire, comme rassemblé en matrice topologique de distance. En d'autres termes, tout la auto corrélation spatial au lag k ATSkw est obtenu en additionnant tout le  $\times W_i$  des produits  $W_j$  de toutes les paires d'atomes i et j, pour lesquels la distance topologique égale le retard comme:

$$ATSkw = \sum_{i=1}^{nSK=1} \sum_{j>i} W_i \cdot W_j \cdot \delta_j \quad (02)$$

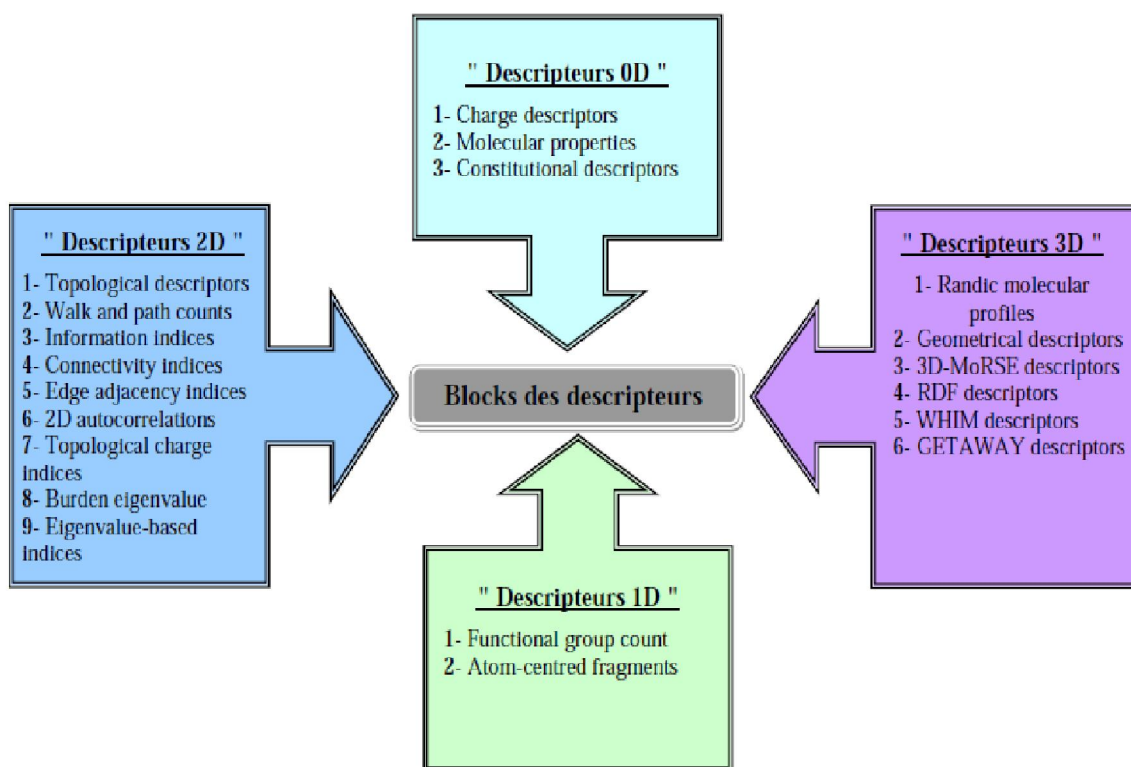
- nSK est le nombre d'atomes de molécule d'non-Hydrogène
- **dij** est le delta de Kronecker (**dij** =1 si **dij**= k, zéro autrement, **dij** étant la distance topologique entre deux atomes considérés).

### I-3-4 Bloc des descripteurs [8] :

Les descripteurs moléculaires calculés par **DRAGON** sont divisés en 20 blocs logiques. Il existe quatre types de descripteurs moléculaires 0D, 1D, 2D, 3D.

- Descripteurs moléculaires 0D, 1D : il y a 5 bloc, contient des nombre d'atomes, des groupements fonctionnels, des propriétés des molécules, et les charges.
- Descripteurs moléculaires 2D : il y a 9 bloc, possède des descripteurs simples et autres dérivées des algorithmes appliqués à une représentation topologique.
- Descripteurs moléculaires 3D: il y a 6 bloc, contiens des descripteurs moléculaires dérivés d'une représentation géométrique appelé aussi descripteurs géométriques.

Les blocks des descripteurs moléculaires calculés par DRAGON sont représentés dans la figure suivant :



**Figure 03:** Représentation des blocs des descripteurs moléculaires [8].

**I-3-5-Le choix des descripteurs :** pour avoir le bon choix de descripteurs moléculaires on doit respecter les points suivants.

- 1- Les descripteurs doivent avoir une relation directe, et une description suffisante à la structure moléculaire des composés chimiques à étudier.
- 2- Les descripteurs doivent avoir des liens avec les propriétés chimiques et physiques et l'activité des composés chimiques.
- 3- La simplicité des descripteurs.
- 4- La représentation descriptive de la totalité de la structure moléculaire des composés.
- 5- La diversité (variétés) de ces descripteurs.
- 6- la sélection exacte et la distinction entre les descripteurs décisifs (principaux), et d'écarts (secondaires).
- 7- La simplicité et l'efficacité de ces descripteurs, puisque la complexité des provoque une perturbation sur les résultats d'estimation [9].
- 8- Il n'est pas nécessaire d'avoir des relations parfaites, entre les variables (descripteurs) prédictives, Parce qu'elles causent des difficultés sur la régression linéaire (confusion parfaite d'effets et perturbation dans l'interprétation des résultats) [10].

### **I.3.6. Importance des descripteurs :**

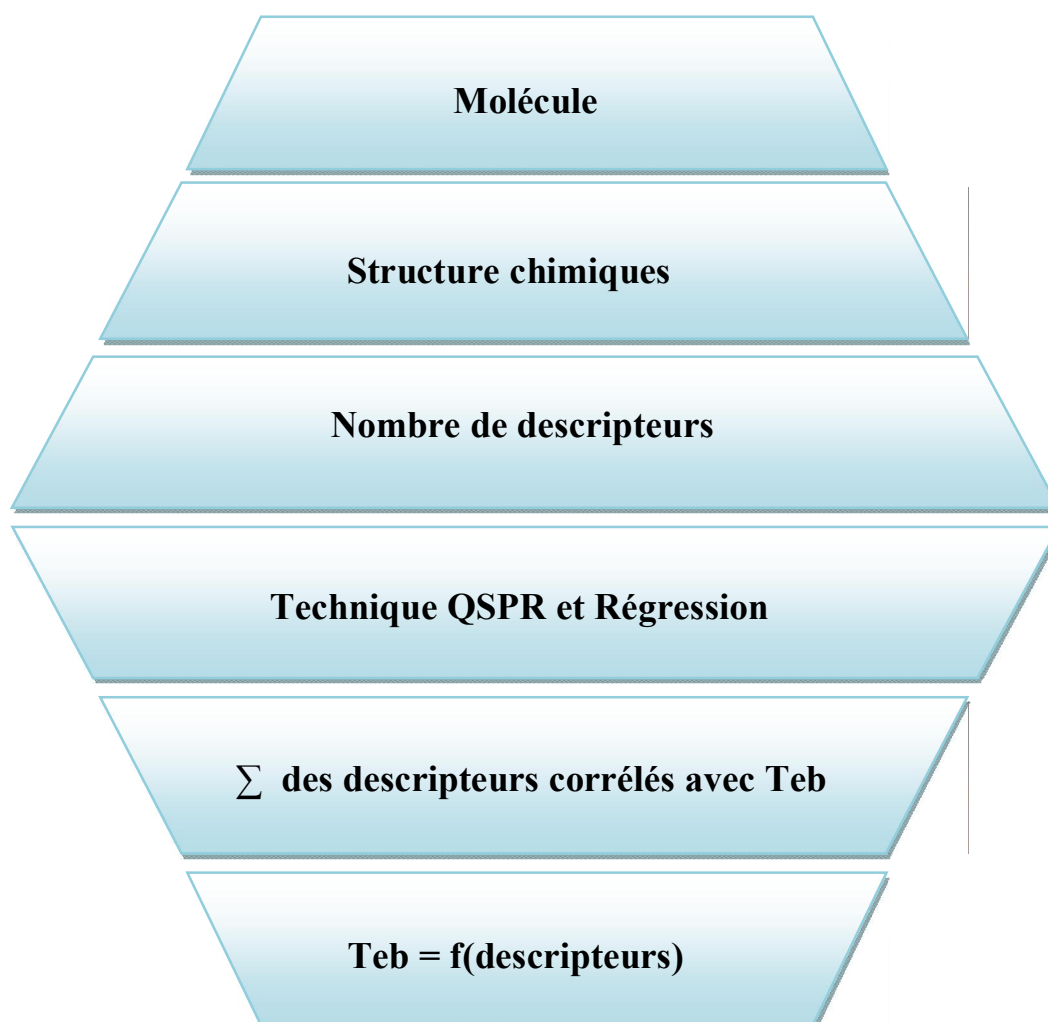
Les descripteurs moléculaires jouent un rôle fondamental en chimie, sciences pharmaceutiques, la protection de l'environnement, recherche de la santé et contrôle de qualité, être obtenu quand les molécules sont transformées dans une représentation moléculaire qui autorise quelque traitement mathématique. Les descripteurs moléculaires sont très importants pour :

- Indiquent la description de la configuration de la molécule à étudiée.
- Décrivent tous les paramètres descriptifs de la molécule.

Les descripteurs moléculaires sont utilisés pour, une connaissance de statistiques, chimio métriques, et les principes des approches QSAR/QSPR sont nécessaires en plus de la connaissance spécifique du problème [8].

### **I.4. Diagramme montre les étapes de prédiction**

Les premiers essais de modélisation QSPR développés par Wiener, reliant point d'ébullition et structure moléculaire QSAR. Depuis, l'essor de nouvelles techniques de modélisation par apprentissage, linéaires d'abord, puis non linéaires, ont permis la mise en place de nombreuses méthodes ; elles reposent pour la plupart sur la recherche d'une relation entre un ensemble de nombres réels, descripteurs de la molécule, et la propriété ou l'activité que l'on souhaite prédire. La prédiction des propriétés s'effectue par plusieurs étapes intéressantes illustrant dans le diagramme suivant :



**Figure 04:** *Diagramme de prédiction par QSPR.*

### **I-5- L'objectif de la prédiction**

L'objectif principal est d'établir un modèle de prédiction de propriété en se basant sur la structure moléculaire et plus précisément les descripteurs moléculaires.

Ce modèle permet, en suite la classification des composées subi à la prédiction.

### **II- COLLECTE DES DONNEES**

La propriété considérée (Température d'ébullition) est l'un des grandeurs physiques important dans la chimie analytique.

Les données utilisées dans ce travail, prélevées dans [tableau02], concernent 50 anilines ; elles sont réunies dans le tableau - 02.

On a choisis aléatoirement 15 anilines (soit 30%) pour la validation externe [marqués par un astérisque].

**Tableau 02** : Nomenclature et valeurs de propriété des anilines étudiées [11]

N°	Composé	N° de CAS	Teb(K)
1	2-Fluoroaniline	348-54-9	448
2	Aniline	62-53-3	457
3	3-Fluoroaniline	372-19-0	461
4	4-Méthylaniline	106-49-0	473
5	2-Méthylaniline	95-53-4	473.3
6	3-Méthylaniline	108-44-1	476.3
7	2-Chloroaniline	95-51-2	481.8
8	2-Ethylaniline	578-54-1	482.65
9	3-Ethylaniline	587-02-0	487
10	2,4-diméthylaniline	95-68-1	487
11	2,5-diméthylaniline	95-78-3	487
12	2,6-diméthylaniline	87-62-7	488
13	2-chloro-6-méthylaniline	87-63-8	488
14	4-Ethylaniline	589-16-2	490.5
15	2,6-dichloroaniline	608-31-1	492.9
16	2-chloro-4-méthylaniline	615-65-6	493
17	3,5-diméthylaniline	108-69-0	493.5
18	2-Isopropylaniline	643-28-7	494
19	4-Propylaniline	2696-84-6	498
20	4-Isopropylaniline	99-88-7	498
21	2-Propylaniline	1821-39-2	499
22	3,4-diméthylaniline	95-64-7	501
23	2-Bromoaniline	615-36-1	502
24	3-Chloroaniline	108-42-9	503.5
25	5-chloro-2-méthylaniline	95-79-4	512
26	2,6-dichloro-3-méthylaniline	64063-37-2	514.1
27	4-chloro-2-méthylaniline	95-69-2	517
28	2,4-dichloroaniline	554-00-7	518
29	3-chloro-2-méthylaniline	87-60-5	518

**Tableau 02 : Suite et fin**

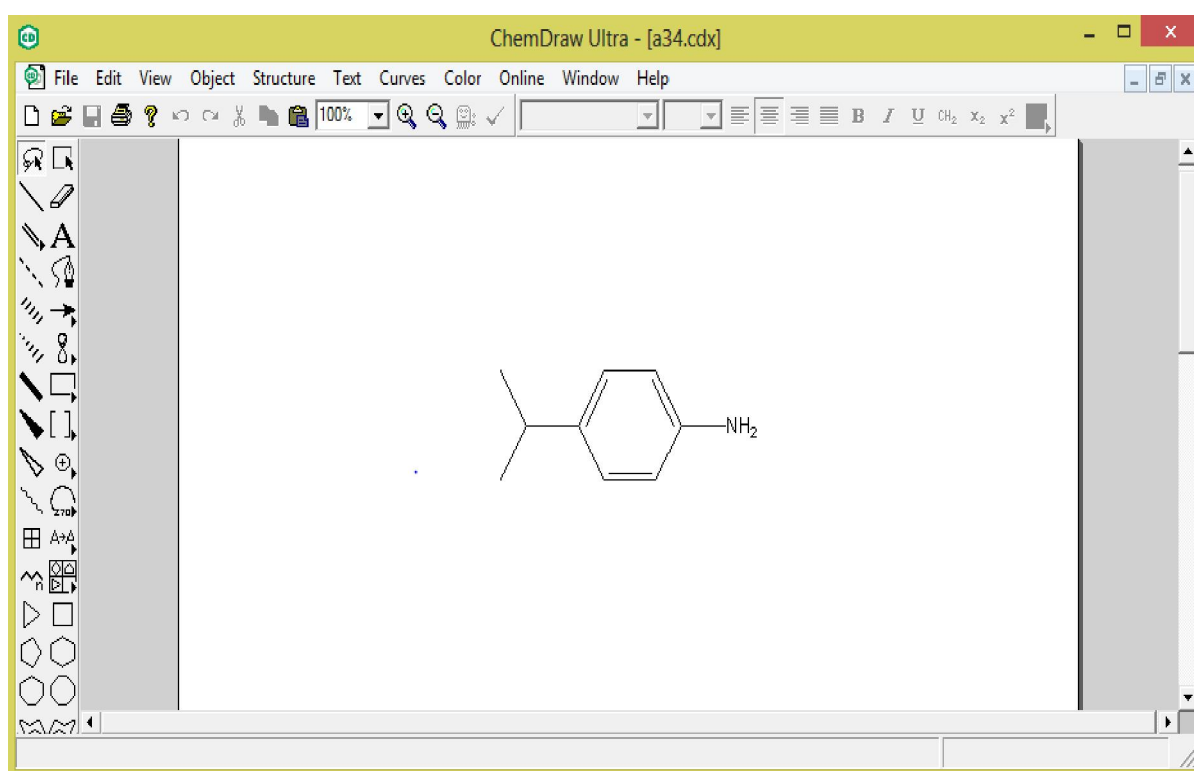
N°	Composé	N° de CAS	Teb(K)
30	2,3-Dichloroaniline	608-27-5	523
31	3-Bromoaniline	591-19-5	524
32	2,5-dichloroaniline	95-82-9	524
33	4-Butylaniline	104-13-2	534
34	3,5-dichloroaniline	626-43-7	534
35	2-Iodoaniline	615-43-0	535
36	2,4,6-trichloroaniline	634-93-5	535
37	4-Pentylaniline	33228-44-3	539.5
38	2-iodo-5-méthylaniline	13194-69-9	539.5
39	2,4,5-trichloroaniline	636-30-6	543
40	3-Iodoaniline	626-01-7	544.6
41	3,4-dichloroaniline	89059-40-5	545
42	3-chloro-4-méthylaniline	33240-95-8	545
43	4-Hexylaniline	33228-45-4	555
44	2,3,4-trichloroaniline	634-67-3	565
45	2-Phénylaniline	90-41-5	572
46	4-Phénylaniline	92-67-1	575
47	4-Octylaniline	16245-79-7	583
48	3,4,5-trichloroaniline	634-91-3	587.8
49	3-Phénylaniline	2243-47-2	614.7
50	2-chloro-5-methyl aniline	95-81-8	502

### III- OPTIMISATION DE LA GEOMETRIE MOLECULAIRE ET GENERATION DES DESCRIPTEURS MOLECULAIRES

#### III-1 -Préparation de base des données

##### III-1-1-Logiciels « ChemDraw 7 »

ChemDraw Pro [12] est le progiciel de dessin de structures chimiques le plus utilisé au monde grâce à ses fonctions d'édition avancées. Il permet d'interroger des bases de données chimiques et de dessiner la structure plane d'un composé conforme à la nomenclature de l'IUPAC.



**Figure 05 :** Représentation des molécules par Chem Draw

**Composé :** 4-isopropyl aniline

##### III-1-2-Logiciels « HyperChem »

HyperChem [13] rassemble dans une même interface un ensemble d'outils dédiés à la modélisation moléculaire, qui est connu pour sa qualité, flexibilité, et facilité d'usage.

## □ **Fonctionnalités:**

HyperChem est le logiciel qui nous permet de faire réellement de la modélisation : il possède plus de méthodes de calculs (mécanique moléculaire, semi-empirique et ab-initio) pour que nous puissions calculer plus de propriétés.

HyperChem est utilisé dans cette étude pour construire et optimiser les molécules, chaque molécule est enregistrée comme un fichier d'extension "Hin" après l'optimisation. Nous avons utilisés la méthode semi empirique MM+ pour l'optimisation. On a 50 molécules donc on obtient autant de fichiers Hin.

### **III-2-Récupération et stabilisation des molécules**

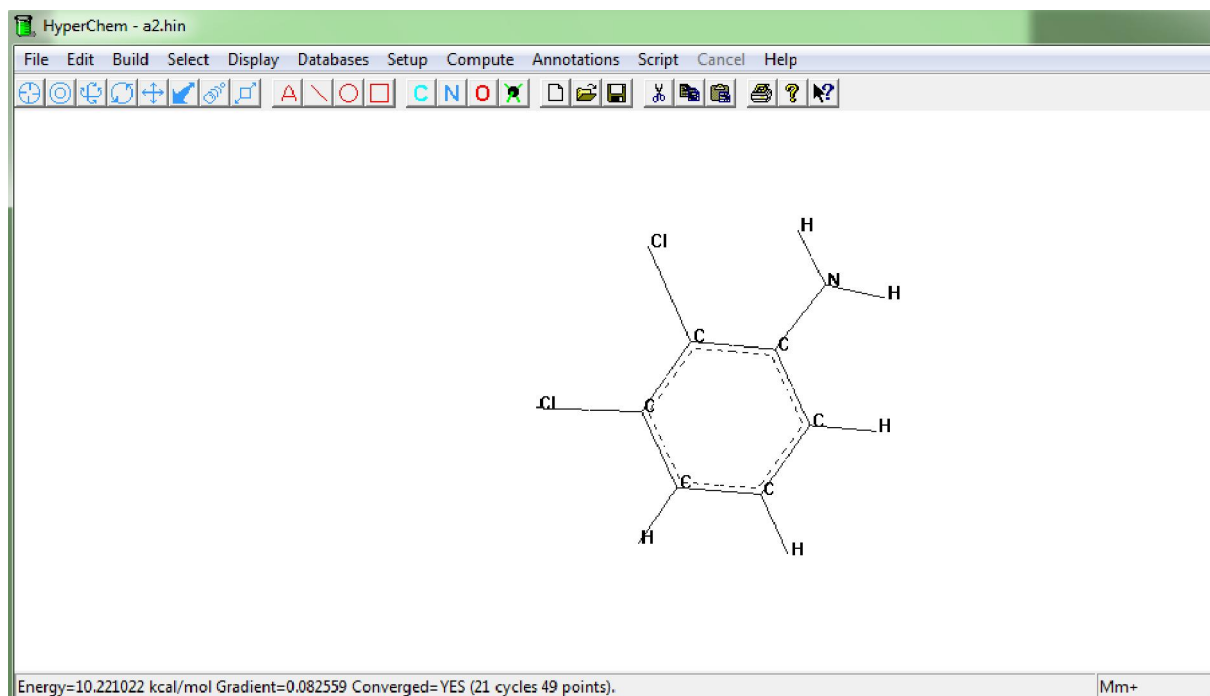
#### **III-2-1- Stabilisation structure des molécules (minimisation de l'énergie)**

Pour stabiliser la structure de chaque molécule ou la minimisation de l'énergie (ou de la géométrie d'optimisation), on utilise l'HyperChem, qui peut effectuer cette opération en utilisant une variété de méthodes de calcul. Les deux mécanismes moléculaires et les méthodes semi-empiriques sont disponibles. Minimisation de l'énergie modifie la géométrie ou la forme d'une molécule d'abaisser l'énergie potentielle de la molécule et pour donner une conformation plus stable.

#### **III-2-2- Mécanique moléculaire**

Champs de force mécanique moléculaire utiliser les équations de la mécanique classique pour décrire les surfaces d'énergie potentielle et des propriétés physiques des molécules. Une molécule est décrite comme une collection d'atomes qui interagissent les uns avec les autres par de simples fonctions analytiques. Cette description est appelée un champ de force. Une composante d'un champ de force est l'énergie résultant de la compression et l'étirement d'un caoutchouc [14].

HyperChem comprend quatre haut-mécanique moléculaire des champs de force : de nouvelles implémentations de techniques élaborées et publiées par des groupes de chercheurs. Mais dans ce travail on concentre à la méthode MM+ de caractéristiques Field Force présente dans la figure suivante :



**Figure 06** Application de la méthode Field Force au composé : 2,3-Dichloroaniline

### III-2-3- Récupération des fichiers

Une fois que nous avons construit une structure en HyperChem, nous pouvons l'enregistrer pour une utilisation ultérieure. C'est une bonne idée puisque cela nous fait gagner du temps si nous voulons revoir notre structure à une date ultérieure. Pourquoi la construire deux fois?! Nous pouvons le faire en allant dans Fichier et Enregistrer. Nous devons lui donner un nom comme d'extension Hin.

Dans le calcul de descripteurs moléculaires, les structures chimiques des composés optimisés sont nécessaires. Les structures chimiques de tous les 50 composés dans notre jeu de données, ont été établies dans le logiciel HyperChem, et pré-optimisés en utilisant la méthode semi empirique MM+.

### III-3- Le Logiciel « Dragon »

Dragon [15] est une application pour le calcul des descripteurs moléculaires. Ces descripteurs peuvent être utilisés pour évaluer l'influence de la structure moléculaire ou les relations propriétés-structure, aussi pour l'analyse de symétrie et la projection des bases de données des molécules

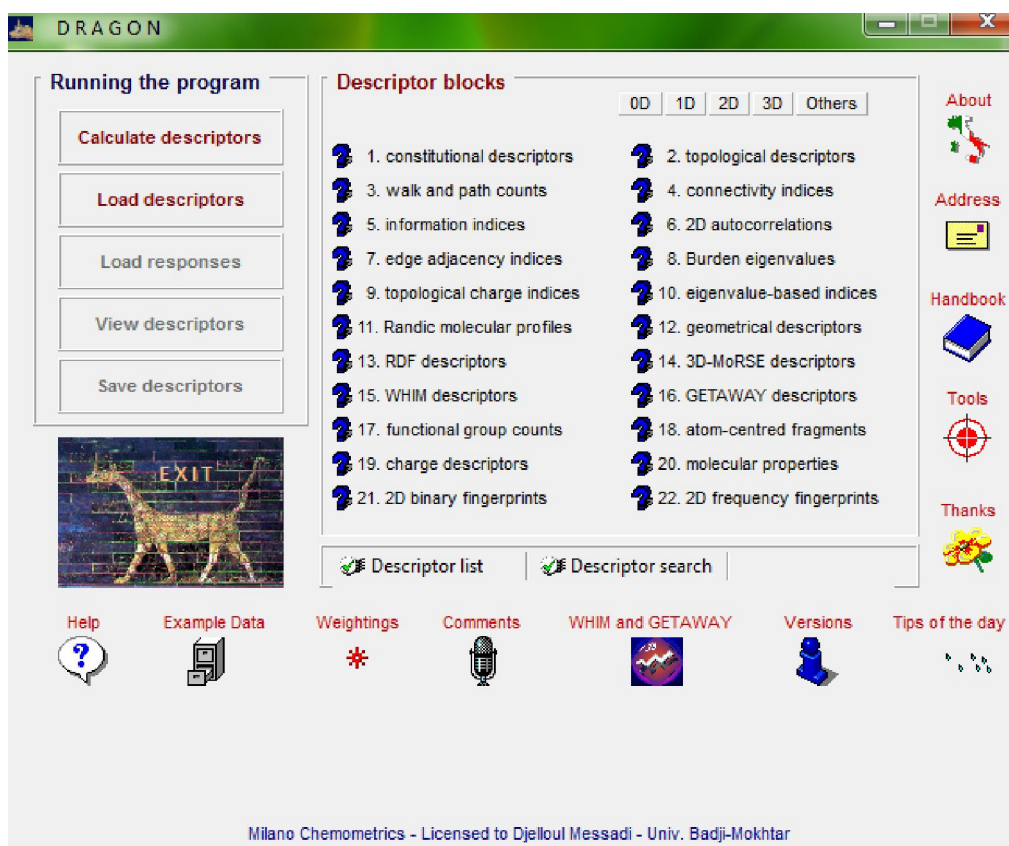


Figure 07 : Interface principale du logiciel DRAGON [15]

## IV-Méthodes de régressions linéaires et multilinéaires

### IV-1 Aperçu général

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

- a) Un nombre maximal de données expérimentales (de préférence toutes) doivent être caractérisées par des valeurs de descripteurs originaux complémentaires.
- b) Les valeurs des descripteurs doivent être obtenues de la même source et, de préférence, mesurées selon le même protocole expérimental ou calculées en utilisant le même logiciel.
- c) Le nombre de descripteurs dans les modèles de régression multiples doit être minimisé, sans perte d'information, ce que mettent en évidence les critères statistiques (valeurs des tests t et F...).

d) Dans les modèles MLR, les descripteurs utilisés doivent être statistiquement orthogonaux.

e) Pourvu que les autres critères soient similaires, la nature physique ou chimique du descripteur sélectionné doit être la plus proche de la propriété ou du phénomène étudié.

En réalité, il est difficile de se conformer pratiquement aux 5 principes énoncés.

Cependant, la négligence de plusieurs d'entre eux peut conduire à des équations inutiles sans aucun pouvoir prédictif sinon très limité.

#### **IV-2- Evaluation préliminaire des données**

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

On distingue habituellement dans un tel pré- traitement des données les analyses uni variées des analyses bi-variées [16-21].

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

Pour un ensemble de descripteurs différents, il est nécessaire d'effectuer une analyse des données bi-variée, c'est-à-dire de calculer le coefficient de corrélation linéaire R entre chacune des paires de l'ensemble des descripteurs. Si R est statistiquement significatif ( $R > 0,9$ ), ces deux descripteurs ne peuvent être utilisés simultanément lors de l'analyse par MLR.

#### **IV-3- Régression linéaire multiple**

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

$$y_i = \beta_0 + \sum_{j=1}^p \beta_j X_{ij} + \xi \quad (03)$$

Où les  $y_i, x_{i1}, x_{i2}, \dots, x_{ip}$  : sont des données respectivement relatives aux variables Y,  $X_1, \dots, X_p$ . Les estimateurs  $\beta_j$  sont calculés en utilisant la méthode des moindres carrés

ordinaires. Les variables aléatoires  $\xi_i$  représentent les termes d'erreur non observables du modèle. On peut estimer ces erreurs par les résidus ordinaires  $e_i$ , différence entre les valeurs observées  $y_i$  et les valeurs estimées  $\hat{y}_i$ .

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

a) Les résidus (E) ont une espérance mathématique nulle :

$$E(e) = 0 \dots\dots\dots ; (04)$$

b) Le modèle choisi est correct (aucune variable explicative n'a été omise).

c) Les résidus sont indépendants entre eux :

$$E(e_i, e_j) = 0 \quad \text{si } i \neq j \dots\dots\dots ; (05)$$

Leurs covariances sont nulles.

d) Les résidus ont tous même variance  $\sigma^2$  (propriété d'homoscédasticité). Par ailleurs, l'emploi de tests statistiques pour analyser la variation expliquée par la régression conduit à admettre que :

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

L'analyse des résidus présente un intérêt à plusieurs égards.

Elle permet en effet de vérifier, a posteriori, la validité du modèle utilisé, en ce qui concerne, d'une part la forme de celui-ci (linéarité ou non linéarité de la relation, par exemple) et d'autre part, certaines hypothèses plus spécifiques, telles que l'égalité des variances résiduelles, la normalité des résidus ou l'absence d'auto-corrélation. Pour minimiser l'influence des erreurs de détermination des valeurs explicatives (ou régresseurs) sur la précision des résultats de la régression 5 données (variables dépendantes, ou encore observations) doivent, à la limite, être associées à chaque variable explicative. Le nombre de degrés de liberté final ( $n-p-1$ ) doit être [22] tel que:

$$n - p - 1 \geq 10 \dots\dots\dots ; (06)$$

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

#### **IV-4- Méthodes de sélection des descripteurs**

##### **IV-4-1- Régression pas à pas**

Une technique est parfois employée pour sélectionner un nombre réduit de variables qui explique pourtant une quantité raisonnable de variation. Il existe plusieurs variantes de cette régression dite "pas à pas" (stepwise régression en anglais).

##### **IV-4.2- Algorithme génétique**

La modélisation de processus génétiques a initié le développement des algorithmes génétiques, qui peuvent être exploités dans une grande variété de problèmes d'optimisation [23].

Dans un algorithme génétique adapté à l'optimisation, une solution potentielle est considérée comme un individu dans une population. La valeur de la fonction de coût associée à une solution mesure « l'adaptation » de l'individu associé à son environnement. Un algorithme génétique simule l'évolution, sur plusieurs générations, d'une population initiale dont les individus sont mal adaptés au moyen d'opérateurs génétiques de reproduction et de mutation. Après un certain nombre de générations, la population est constituée d'individus bien adaptés, autrement dit des solutions supposées « bonnes » au problème d'optimisation.

Dans ce travail la sélection des descripteurs a été réalisée par algorithme génétique, dans la version MOBY DIGS de Todeschini [24] et le calcul des paramètres statistiques par le logiciel MINITAB version 16.2.0 [25] .

#### **IV-5 – Paramètres d'évaluation de la qualité de l'ajustement**

Deux paramètres sont couramment utilisés :

- **Le coefficient de détermination multiple**

Afin de se faire une idée sur la qualité de l'ajustement ainsi réalisé nous avons calculé le coefficient de détermination,  $R^2$ , qui exprime la part de la variation dépendante

Y (=température d'ébullition) "expliquée" ou "justifiée" par la régression.

Ce paramètre correspond au carré du coefficient de corrélation, il est compris entre 0 et 1 et s'exprime toujours en pour cent.

Si la valeur de  $R^2$  est proche de 1 ou 100% ; nous avons donc un ajustement d'excellente qualité ; par contre si la valeur de  $R^2$  est faible et proche de 0 ou 0% l'ajustement est mauvais.

$$R^2 = 1 - \frac{SCE}{SCT} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \dots\dots\dots(07)$$

Où  $\hat{y}$  est la valeur estimée du paramètre physique, et  $\bar{y}$  la moyenne des valeurs observées.

• **La racine de l'erreur quadratique moyenne de prédiction (désignée également par SDEP) :**

$$EQMP = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} = \sqrt{\frac{PRESS}{n}} \dots\dots\dots(08)$$

**VI-6-Facteur d'inflation de la variance [FIV]**

Le facteur d'inflation de la variance sert à détecter si descripteur présente une association linéaire forte avec les prédicateurs restants (présence de multi colinéarité parmi les prédicateurs). Le facteur d'inflation de la variance donne une mesure de l'accroissement de la variance d'un coefficient de régression estimé s'il existe une corrélation entre prédicateurs (multi colinéarité). FIV =1 indique qu'il n'y a pas de relations, si non FIV est supérieur à 1 le facteur FIV le plus grand parmi tous les prédicateurs sert souvent d'indicateur de multi colinéarité importance, Si le FIV > 5-10 la qualité de l'estimation des coefficients de régression est faible.

**VI-7- Test de randomisation**

Ce test permet de mettre en évidence des corrélations dues au hasard. Il consiste à générer un vecteur « propriété considérée » par permutation aléatoire des composantes du vecteur réel. On calcule alors sur le vecteur obtenu (considéré comme vecteur expérimental réel) un modèle QSPR, selon la méthode habituelle. Ce procédé est répété plusieurs fois (100 dans notre cas).

## VI-8 - Validation externe

Il est intéressant, pour juger de la qualité du modèle, de considérer la racine de l'écart quadratique moyen (RMSE, pour Root Mean Squared Error), calculée sur différents ensembles :

- Ensemble d'estimation ou de calibration (appelée SDEC)
- Ensemble de validation croisée (appelée également SDEP)
- Ensemble de prédiction externe (désignée par SDEPext).

Ces valeurs RMSE sont mieux adaptées, pour juger de la qualité d'un modèle que les valeurs de  $R^2$  et  $Q^2$  seules, qui constituent de bons tests uniquement pour des données réparties régulièrement.

$$EQMC = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \dots\dots\dots(09)$$

$$EQMP\ ext = \sqrt{\frac{\sum_{i=1}^{n\ ext} (y_i - \hat{y}^{(i)})^2}{n\ ext}} \dots\dots\dots(10)$$

La validation croisée par « leave – one - out » (LOO) [26] consiste à recalculer le modèle sur (n-1) observations, et à utiliser le modèle ainsi obtenu pour calculer la grandeur d'intérêt du composé écarté, notée  $\hat{y}(i)$ . On répète le procédé pour chacune des grandeurs d'intérêt.

La somme des carrés des erreurs de prédiction, désignée par le symbole PRESS (équation. (11)), est une mesure de la dispersion des estimations. On l'utilise pour définir le coefficient de prédiction :

$$Q^2LOO = \frac{SCT - PRESS}{SCT} \dots\dots\dots(11)$$

Contrairement à  $R^2$  qui augmente avec le nombre de paramètres du modèle, le facteur  $Q^2LOO$  affiche une courbe avec maximum (ou avec palier) obtenu pour un certain nombre de descripteurs, puis décroît de façon monotone. Ce fait confère une grande importance au coefficient  $Q^2LOO$ . Une valeur de  $Q^2LOO > 0,5$  est considérée comme satisfaisante, et une valeur supérieure à 0,9 est excellente [27].

Si de petites valeurs de  $Q^2LOO$  indiquent des modèles peu robustes, caractérisés par de faibles capacités prédictives internes, le contraire n'est pas nécessairement vrai.

En fait, si une forte valeur de  $Q^2_{LOO}$  est une condition nécessaire de robustesse et d'une possible capacité prédictive élevée d'un modèle, cette condition seule n'est pas suffisante, et peut conduire à une surestimation de la capacité prédictive du modèle lorsqu'il est appliqué à des composés réellement externes.

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

Cependant, ces procédures ne sont que rarement rapportées avec les résultats QSPR courants, et n'ont pas été pleinement exploitées dans le présent travail.

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

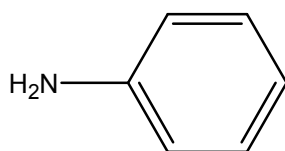
$$Q^2_{ext} = 1 - \frac{\sum_{i=1}^{n_{ext}} (y_i - \hat{y}_i)^2 / n_{ext}}{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / n} \dots\dots\dots(12)$$

Pour une grande valeur de  $Q^2_{LOO}$ , une valeur élevée de  $Q^2_{ext}$  permet de présager d'une bonne capacité prédictive du modèle. Dans notre cas :  $n=35$  et  $n_{ext} = 15$ .

## V- Généralités sur les anilines

### V-1- Aniline [28]

L'aniline, connue également sous les noms de: amino-benzène, phénylamine ou benzène amine, est un composé organique aromatique de formule chimique  $C_6H_5NH_2$ .



**Figure 08** : Structure de l'aniline

### V-2- Historique [29]

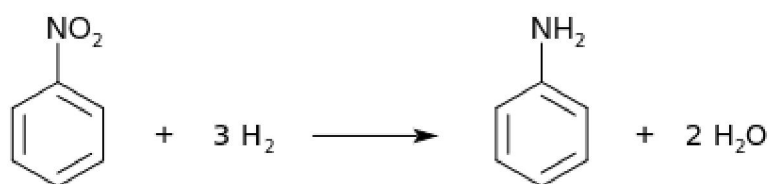
L'aniline a été isolée pour la première fois en 1826 par Otto Unverdorben par distillation de l'indigo. Celui-ci la baptisa crystalline. En 1834, Friedrich Runge parvint à isoler du goudron de houille, une substance qui, une fois traitée par du chlorure de chaux prend une couleur

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.

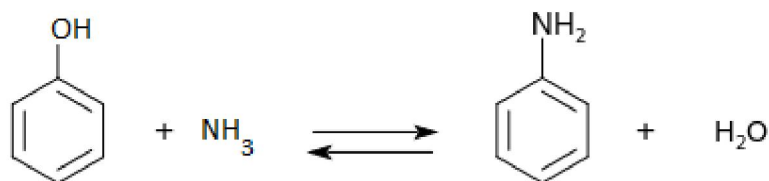
### V-3- Production de l'aniline [30]

L'aniline peut être fabriquée à partir de deux procédés:

- Réduction ou hydrogénation du nitrobenzène selon la réaction suivante:



- Substitution amine du phénol selon la réaction suivante :



### V-4- Présence de l'aniline dans l'environnement [31]

Il n'existe pas de source naturelle d'aniline. L'aniline peut pénétrer dans l'environnement (les eaux résiduaires et les flux d'air rejetés) à partir de sources anthropogènes au cours de n'importe quel stade de la production, du stockage, du transport, de l'utilisation et de l'élimination.

La réduction du nitrobenzène, la décomposition des déchets organiques divers (colorants, caoutchoucs, pesticides, plastiques et peintures), les usines de combustion du bois et les incinérateurs de déchets sont aussi responsables de la présence de l'aniline dans l'environnement [31].

### **V-5- Utilisations [32, 33]**

L'aniline est une matière première utilisée en synthèse organique pour la fabrication de nombreux produits, tels que :

- Isocyanates, intermédiaires de synthèse des polyuréthanes ;
- Accélérateurs de vulcanisation dans l'industrie du caoutchouc ;
- Matières colorantes ;
- Produits phytosanitaires (herbicides, fongicides) et pharmaceutiques.

### **V-6- Propriétés Chimiques [33 - 35]**

Dans les conditions normales de température et de pression, l'aniline est un produit stable. Elle 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, etc., mais avec des oxydants puissants, la réaction peut être violente. Il 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.

La combustion de l'aniline dégage des fumées toxiques (oxyde d'azote et oxyde de carbone).

### **V-7- Propriétés Physiques [32- 35]**

L'aniline est un liquide huileux, incolore quand il est fraîchement distillé, mais ayant une 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 avec la plupart des solvants organiques.

**Tableau 03: Principales caractéristiques physiques**

<b>NOM DE LA SUBSTANCE</b>	<b>DÉTAILS</b>	
<b>Aniline</b>	N° CAS	62-53-3
	Etat Physique	Liquide
	Masse molaire	93,13g/mol
	Point de fusion	- 6,2 °C
	Point d'ébullition	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
	Viscosité	4,4 mPa.s
	Température d'auto-inflammation	615 à 630 °C
	Limites d'explosivité ou limite inférieure d'inflammabilité (en volume % dans l'air)	1,2 à 1,3%
	limite supérieure :	11%

### **V-8- Toxicité**

L'aniline est très toxique car elle est hémolytique (destruction des globules rouges) pour les organismes aquatiques, et dangereux pour l'environnement, et doit être manipulée avec précaution. Une exposition à des concentrations élevées peut être mortelle. Chauffée à plus de 190 °C, elle se décompose en produisant des fumées toxiques (contenant notamment de l'ammoniaque) [36].

Certains organismes classent l'aniline dans la liste des substances cancérogènes, mais sans preuves suffisantes. La valeur limite d'exposition professionnelle est fixée en France à 2 ppm. La dose létale a été évaluée à 250 mg·kg<sup>-1</sup> par voie orale chez le rat [36].

L'aniline est absorbée de façon prépondérante par voie pulmonaire (environ 90% de la quantité inhalée) et par voie cutanée, mais aussi par voie digestive. Transporté par le sang, l'aniline est métabolisée dans le foie ensuite éliminée dans l'urine sous forme de 4-aminophénol.[37]

#### **IV- Température d'ébullition (Teb)**

La température d'ébullition est la température à laquelle les phases liquide et gazeuse d'une substance pure sont en équilibre à une pression donnée, c'est la température à laquelle la substance change d'état, du liquide au gaz à une pression donnée. Le point d'ébullition normal est le point d'ébullition à la pression atmosphérique normale ( $1,013 \cdot 10^5$  kPa). En termes d'interactions intermoléculaires, le point d'ébullition représente la température à laquelle les molécules possèdent l'énergie thermique suffisante pour surmonter les attractions intermoléculaires liant les molécules dans le liquide.

La température d'ébullition d'un composé pur augmente avec la taille, la ramification de la molécule, et avec la présence des liaisons hydrogènes et des interactions dipôle-dipôle.

La température d'ébullition est importante pour la caractérisation et l'identification du composé. Elle fournit également une indication de la volatilité d'un composé. D'autres propriétés physiques, telles que la température critique [38], le point d'éclair [38], et l'enthalpie de vaporisation [39], peuvent être prédits ou estimés à partir des points d'ébullition. Le besoin de données fiables pour l'optimisation des processus industriels, le développement des modèles QSPR fiables pour l'estimation des points d'ébullition normaux pour les composés qui ne sont pas encore synthétisés est devenu important.

# **Partie**

# **Application**

Dans la partie application, nous avons déterminé la modélisation de la température d'ébullition d'une série d'anilines. L'ensemble des données est divisé aléatoirement en deux sous ensembles, le premier contient 35 composés (soit 70 %) pour la calibration et le second contient 15 composés pour la validation externe.

Le meilleur modèle obtenu sera validé par l'approche hybride algorithme génétique et régression multilinéaire AG/RLM.

## I-Calcul du modèle :

### I-1- Choix du modèle :

Les molécules sont dessinées par le logiciel ChemDraw (ChemDraw ultra 7.0) (ANNEXE) puis elles sont optimisées en utilisant le logiciel HyperChem (HyperchemTM Release 6.03). Les descripteurs moléculaires ont été calculés à l'aide du logiciel informatique Dragon, plus de 1600 descripteurs sont calculés. D'après l'algorithme génétique dans la version MobyDigs, plusieurs modèles sont obtenus pour chaque jeu de groupe; le choix a été opté pour le modèle qui conduit aux meilleurs statistiques des 100 modèles générés par algorithmes génétiques. Sa validation est effectuée par 15 composés, alors que sa calibration par 35

### I-2- Choix de la taille du modèle:

Un autre facteur très important à tenir compte c'est la taille du modèle pour cela on doit chercher un modèle qui contient le moins de descripteurs et donne les meilleurs résultats.

En traçant le graphe de  $R^2$  en fonction de nombre de descripteurs  $k$  (figure 09) ; on voit que le bon modèle qui possède le moins de descripteurs est celui à 4, parce qu'après ce modèle les valeurs de  $R^2$  augmentent d'un pas faible.

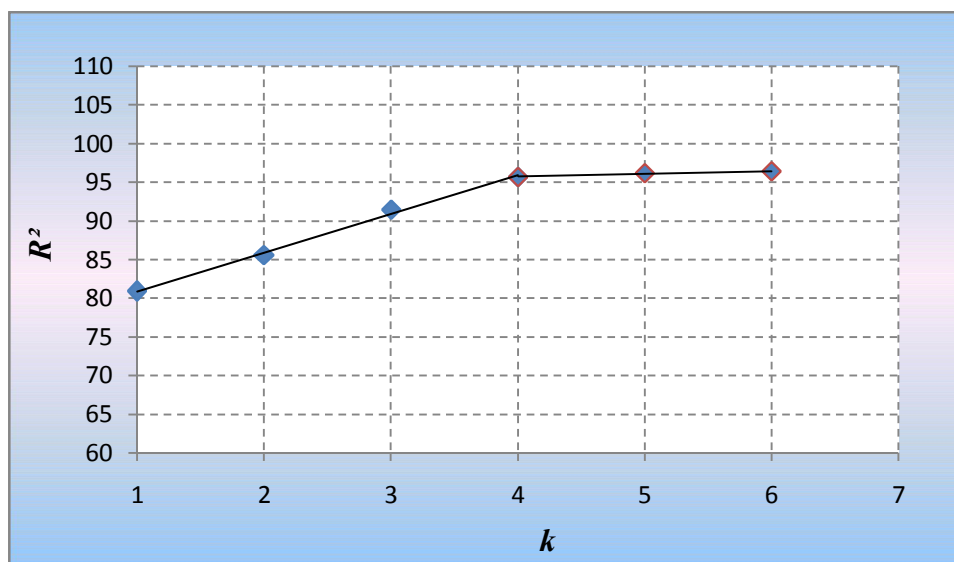


Figure 09 Variation de  $R^2$  en fonction de la taille des modèles  $k$

### I- 3- Choix d'une famille de descripteurs :

Les descripteurs entrant dans le modèle obtenus par algorithme génétique, leurs classes et de brèves définitions sont donnés dans le tableau 04

**Tableau 04** : Classes et signification des descripteurs.

Descripteur	Classe	Signification
MATS3p	2D autocorrelation indices	The <b>Moran autocorrelation</b> $MATS_{kw}$ , $w$ being the atomic property used to weight the molecular graph and $k$ the lag, is calculated by applying the Moran coefficient to the molecular graph
RCI	Geometrical descriptors	The Jug RC index (RCI) is an aromaticity index based on the idea of ring current whose magnitude is determined by its weakest link in the ring [K. Jug, Croat.Chem.Acta 1984, 57, 941-953]. The weakest link is calculated as the maximum bond distance of aromatic bonds (inversely proportional to the minimum total bond order).
HTu	GETAWAY descriptors	The other set of GETAWAY descriptors, shown in the table below, is based on the spatial autocorrelation formulas, weighting the molecule atoms by physico-chemical properties together with 3D information encoded by the elements of the molecular influence matrix H and influence/distance matrix R.
AMR	Molecular properties	The molar refractivity (AMR) is calculated according to the Ghose-Crippen model, based on a group contribution method [A.K.Ghose and G.M.Crippen, J.Comput. Sci. 1987, 27, 21]. Each atom in every structure is classified into one of the 120 atom types. The list of the atom types with the corresponding molar refractivity contributions is given under Atom-centred fragments. The AMR estimates are provided only for compounds having C, H, O, N, S, Se, P, B, Si, and halogens.

Plusieurs classes de descripteurs moléculaires ont été calculées pour la molécule entière :  
Descripteurs relatifs a la construction moléculaire ; et descripteur topologiques.

Nous rapporterons les résultats obtenus pour les 50 anilines qui sont présentées dans le tableau 05 au dessous:

**Tableau 05 :** Valeurs des descripteurs moléculaires sélectionnés par AG

N	Composé	Teb	MATS3p	RCI	HTu	AMR
1	Aniline	457	-0.3	1.418	8.209	30.758
2	5-chloro-2-méthylaniline	512	-0.035	1.425	10.391	40.604
3	4-Propylaniline	498	-0.111	1.418	15.43	45.002
4*	4-Phénylaniline	575	-0.052	1.417	13.716	55.895
5	4-Pentylaniline	539.5	-0.182	1.417	19.342	54.543
6	4-Octylaniline	583	-0.196	1.418	25.439	68.269
7	4-Méthylaniline	473	-0.184	1.418	10.544	35.8
8	4-Isopropylaniline	498	-0.091	1.415	15.28	44.949
9	4-Hexylaniline	555	-0.083	1.418	21.473	58.805
10*	4-Ethylaniline	490.5	-0.125	1.419	13.258	40.401
11	4-chloro-2-méthylaniline	517	-0.035	1.425	10.369	40.604
12	4-Butylaniline	534	-0.23	1.418	17.549	50.125
13	3-Phénylaniline	614.7	-0.052	1.416	13.512	55.895
14	3-Méthylaniline	476.3	-0.184	1.418	10.511	35.8
15	3-Iodoaniline	544.6	-0.113	1.419	8.043	43.167
16	3-Fluoroaniline	461	-0.385	1.419	8.177	30.975
17	3-Ethylaniline	487	-0.125	1.418	13.036	40.401
18	3-Chloroaniline	503.5	-0.015	1.419	8.114	35.563
19	3-chloro-4-méthylaniline	545	0.008	1.418	10.497	40.604
20	3-chloro-2-méthylaniline	518	-0.015	1.428	10.436	40.604
21*	3-Bromoaniline	524	-0.031	1.418	8.077	38.381
22	3,5-diméthylaniline	493.5	-0.125	1.416	13.078	40.841
23*	3,5-dichloroaniline	534	-0.005	1.418	8.033	40.368
24*	3,4-diméthylaniline	501	-0.1	1.417	13.07	40.841
25	3,4-dichloroaniline	545	0.179	1.42	8.056	40.368
26	3,4,5-trichloroaniline	587.8	0.188	1.417	7.983	45.173
27*	2-Propylaniline	499	-0.136	1.416	15.537	45.418
28	2-Phénylaniline	572	-0.108	1.432	13.708	55.895

\* composés de validation externe.

**Tableau 05 : Suite et fin**

N	Composé	Teb	MATS3p	RCI	HTu	AMR
29	2-Isopropylaniline	494	-0.167	1.423	15.33	44.949
30	2-Iodoaniline	535	-0.369	1.424	8.076	43.167
31*	2-iodo-5-méthylaniline	539.5	-0.303	1.423	10.55	48.208
32	2-Fluoroaniline	448	-0.051	1.432	8.189	30.975
34	2-Ethylaniline	482.65	-0.213	1.427	12.98	40.401
35*	2-Chloroaniline	481.8	-0.482	1.421	8.139	35.563
36*	2-chloro-6-méthylaniline	488	-0.414	1.425	10.475	40.604
37*	2-chloro-5-methyl aniline	502	-0.43	1.42	10.588	40.604
38*	2-chloro-4-méthylaniline	493	-0.43	1.421	10.435	40.604
39*	2-Bromoaniline	502	-0.464	1.423	8.106	38.381
40	2,6-dichlorol-3-méthylaniline	514.1	-0.505	1.424	10.487	45.409
41	2,6-dichloroaniline	492.9	-0.571	1.424	8.085	40.368
42*	2,5-diméthylaniline	487	-0.213	1.423	12.736	40.841
43*	2,5-dichloroaniline	524	-0.316	1.422	8.063	40.368
44	2,4-diméthylaniline	487	-0.213	1.424	12.926	40.841
45	2,4-dichloroaniline	518	-0.316	1.422	8.081	40.368
46	2,4,6-trichloroaniline	535	-0.456	1.425	8.031	45.173
47	2,4,5-trichloroaniline	543	-0.134	1.423	8.007	45.173
48	2,3-Dichloroaniline	523	-0.121	1.423	8.069	40.368
49	2,3,4-trichloroaniline	565	-0.025	1.422	8.018	45.173
50*	2-Methylaniline	473.3	-0.286	1.424	10.452	35.8

\* composés de validation externe.

#### **I-4 Calcul des corrélations entre les différents descripteurs**

Le coefficient de corrélation,  $r$ , de Bravais-Pearson a servi pour mettre en évidence les relations possibles entre les différents descripteurs des 35 composés, la matrice de corrélation obtenue à l'aide de la commande "corrélation" du logiciel MINITAB, montre que les descripteurs sont entre eux plus ou moins corrélés. Les couples des descripteurs qui présentent des valeurs de  $r > 0.90$ , sont très fortement corrélés et apportent la même information, ce qui fait qu'ils ne peuvent apparaître dans une même équation de régression et ce n'est le cas dans notre cas.

**Tableau 06:** Matrice de corrélations.

	Teb (k)	MATS3p	RCI	HTu
MATS3p	0.351			
RCI	-0.172	-0.172		
HTu	0.246	0.034	-0.279	
AMR	0.77	0.074	-0.161	0.759

### I-5 Analyse de régression

Les valeurs des paramètres statistiques (tableau 8) montrent que les quatre descripteurs (MATS3p; RCI; HTu; AMR) (tableau.1) permettent de corrélérer Teb des 35 anilins.

En effet, la valeur du coefficient de détermination ( $R^2$ ) signifie que 95,89% de la variabilité de la Teb peut être expliquée par ces trois descripteurs, alors que la racine de l'erreur quadratique moyenne de prédiction est de l'ordre de (SDEP= 8,90) ; en outre ce modèle est significatif avec une valeur du paramètre de Fisher : (F=175,03).

- **L'équation de régression ainsi établie est reproduite ci-après :**

L'équation de régression est :

$$\mathbf{Teb = 2079(\pm 487.1) + 59.28 (\pm 8.819)MATS3p - 1234.6 (\pm 342.5) RCI - 7.62(\pm 0.534) HTu + 6.75 (\pm 0.283)AMR} \quad \mathbf{(13)}$$

**Tableau 07 :** Analyse de régression.

Régresseur	Coef	SE Coef	T	P	VIF
Constante	2078.9	487.1	4.27	0	
MATS3p	59.289	8.819	6.72	0	1.039
RCI	-1234.6	342.5	-3.6	0.001	1.126
HTu	-7.6195	0.534	-14.27	0	2.525
AMR	6.7502	0.2832	23.84	0	2.396

D'après l'analyse de régression (tableau 07), toutes les valeurs de VIF sont inférieures à 5 ce qui prouve que les descripteurs sont indépendants, par contre les valeurs de contribution T ne sont pas homogènes.

Les paramètres statistiques de cette validation ( $Q^2_{ext}$ ,  $SDEP_{ext}$ ) ainsi que les valeurs des statistiques pour l'ensemble de calibration sont regroupés dans le tableau 08.

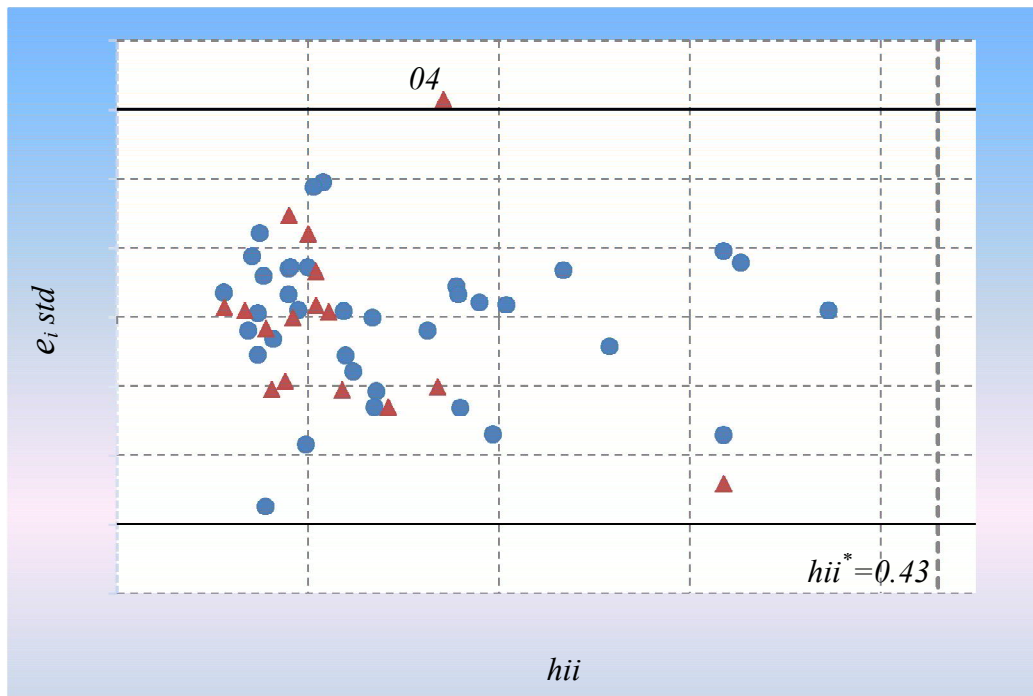
**Tableau 08:** Les paramètres statistiques pour l'ensemble de calibration (35composés) et celui de validation (15 composés).

<b>Les paramètres statistiques</b>	
Taille	04
R <sup>2</sup>	95.89
Q <sup>2</sup>	94.63
Q <sup>2</sup> ext	94.54
R <sup>2</sup> adj	95.34
SDEC	7.79
SDEP	8.90
SDEPext	8.97
Kx	32.25
Kxy	36.92
F	175.03
SE	8.41
PRESS	2770.61

Les résultats représentés dans le tableau 8 prouvent que la relation entre la variable dépendante (Teb) et les variables indépendantes (descripteurs sélectionnés par algorithme génétique) est linéaire.

## **II-Diagramme de Williams :**

On a représenté, sur la même figure; pour les deux ensembles (calibration et validation); les valeurs de  $e_{i\text{std}}$  et celle de  $h_i$ .



**Figure 10:** Diagramme de Williams des composés.

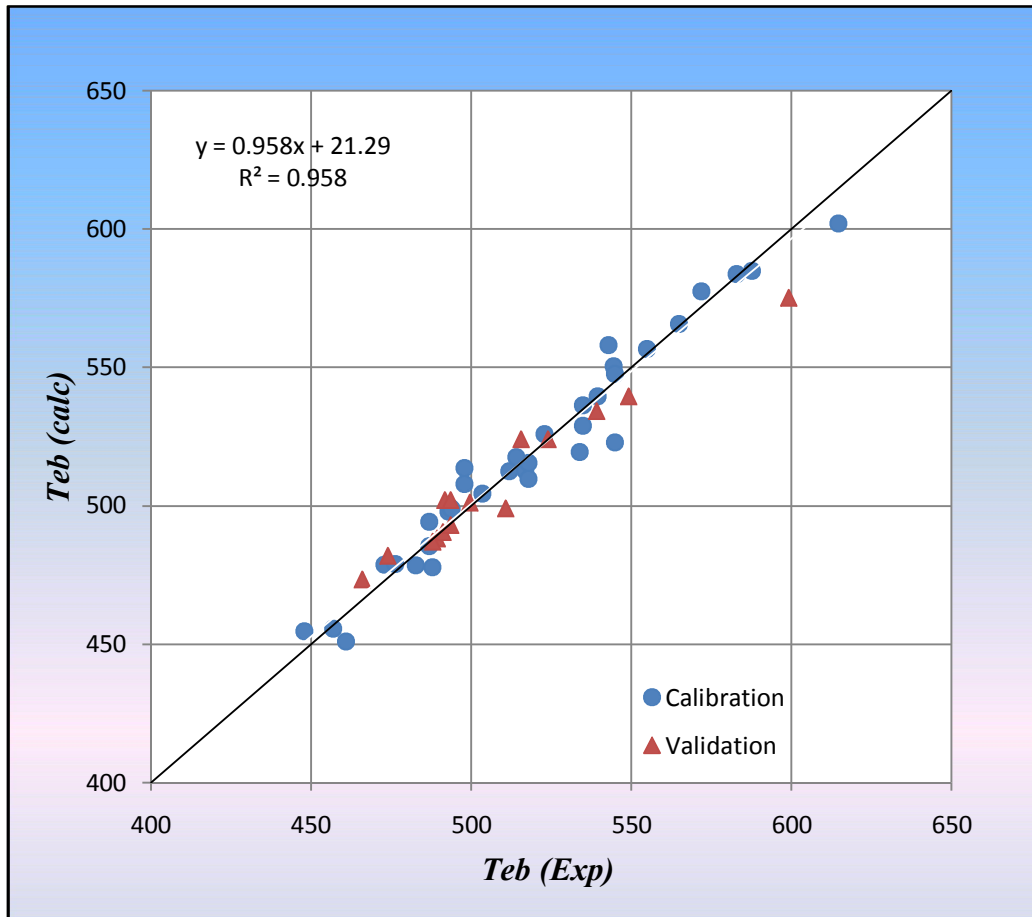
Comme le montre la figure, les valeurs de  $(e_i \text{ std})$  sont comprises entre les bornes  $\pm 3$  à l'exception du composé 04 qui présente un point aberrant il s'agit du composé :

4-Phénylaniline, l'absence des points de levier pour un  $h_{ii}^* = \frac{3(k+1)}{n} = 0,43$  symbolisée par la droite parallèle à l'axe des  $e_i \text{ std}$ .

### III- Evaluation de la qualité de l'ajustement :

Nous avons deux paramètres utilisés pour évaluer la qualité de l'ajustement; la valeur du coefficient de détermination  $R^2 = 95,89\%$  qui explique très bien la variabilité de  $T_{eb}$  en fonction des descripteurs choisis; la racine de l'erreur quadratique moyenne de prédiction  $SDEP = 8,90$ , dont la petite valeur indique un modèle très hautement significatif, que justifie la grande valeur du paramètre de Fisher :  $F = 175,03$ .

La qualité de l'ajustement a été vérifiée en procédant à une validation croisée par "Leave – one – out", La figure 11, qui reproduit les valeurs prédites  $T_{eb}$  en fonction de celles observées, fait ressortir une faible dispersion caractéristique d'un bon ajustement, d'ailleurs confirmé par la grande valeur de  $Q^2 = (94,63\%)$ .

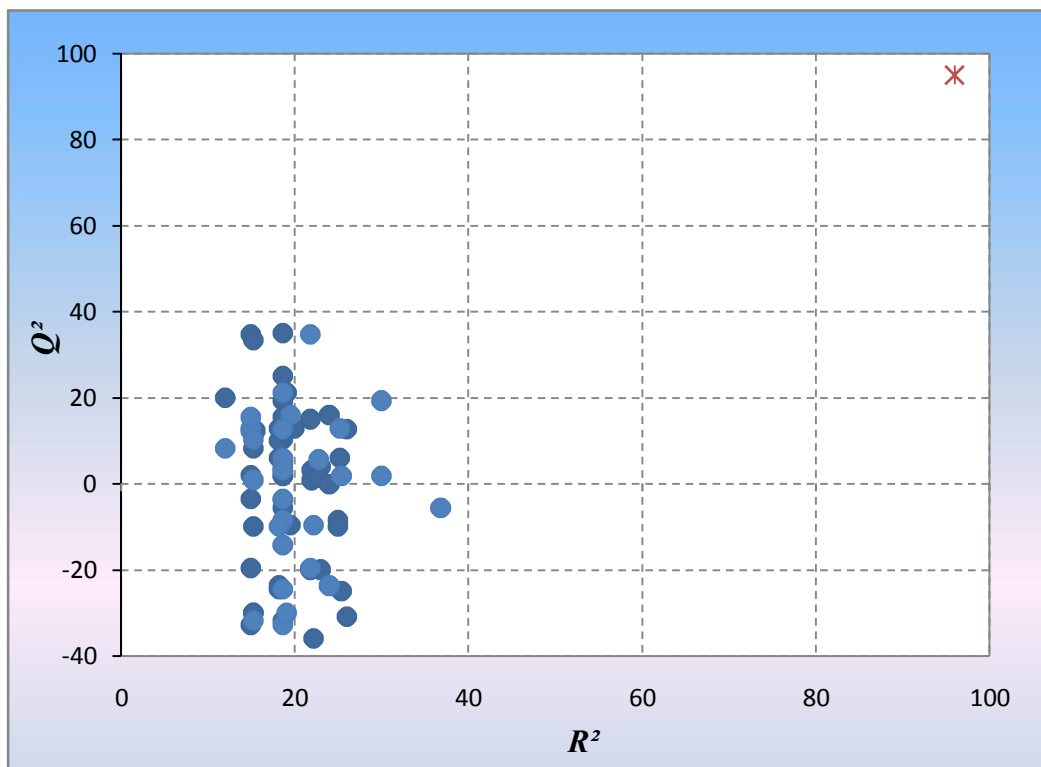


**Figure 11 :** Graphe des valeurs prédites  $Teb$  en fonction des valeurs observées.

## **VI- Test de randomisations :**

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,

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



**Figure 12:** Test de randomisation associé au modèle QSPR.

Il est clair que les statistiques obtenues pour les vecteurs modifiés des températures d'ébullition sont plus petits que celles du modèle QSPR réel, et pour la majeure partie on obtient un  $Q^2 < 0$ , Ceci permet d'assurer qu'une relation structure/ quantitative réelle a été établie.

### VII- Validation externe

Pour généraliser le modèle choisi on procède à une validation externe sur les 15 composés choisis aléatoirement et qui ne font pas partie de l'ensemble d'essai (tableau 05). Les résultats obtenus (  $Q^2_{\text{ext}} = 94.54\%$  ;  $SDEP_{\text{ext}} = 8.97$  ) montrent que les valeurs prédites sont très proches des valeurs observées (figure 11). On remarque une faible dispersion des points autour de la 1<sup>ère</sup> bissectrice, ce qui confirme que le modèle choisi décrit de façon adéquate la relation entre la température d'ébullition prédite et observée.

# **Conclusion**

## **Générale**

## Conclusion générale

Nous avons appliqué la méthodologie QSRR pour relier la propriété (température d'ébullition) d'une série hétérogène d'anilines, à des descripteurs moléculaires théoriques reflétant certaines particularités.

Le modèle QSPR a été établi en utilisant l'analyse de régression multilinéaire.

Les 50 données de base ont été éclatées aléatoirement en deux ensembles disjoints :

- un ensemble principal de 35 composés utilisés pour le calcul et, éventuellement, les essais du modèle ;

- un ensemble de 15 composés pour la prédiction externe.

La taille du modèle ( $p=4$ ) est fixée par la valeur optimale de  $R^2$  en fonction de nombre de descripteurs. La sélection des variables explicatives a été réalisée par les algorithmes génétiques.

Les statistiques ( $R^2 > 95\%$  ;  $RMSE < 09$  ;  $Q^2 > 94\%$ ) calculées établissent la pertinence du modèle MLR développé.

L'analyse des résidus a permis de détecter une observation aberrante et l'absence des observations influentes.

La qualité de l'ajustement a été vérifiée en traçant le graphe des valeurs calculées (prédites) en fonction de celles observées, un bon ajustement était observé cela traduit par une faible dispersion.

Les valeurs RMSE sont proches les unes des autres, ce qui permet, tout à la fois, de s'assurer de la bonne capacité prédictive et de la possibilité d'extension suffisante du modèle obtenu.

Ainsi, la température d'ébullition peut être prédite à partir de la structure moléculaire des anilines en utilisant la modélisation linéaire MLR.

Enfin, d'autres méthodes (RNA, SVM ..... ) qui peuvent s'avérer plus avantageuses en ce qui concerne la précision et l'interprétation des modèles, et du point de vue de la capacité de généralisation, peuvent être tester.

**REFERENCES**

**BIBLIOGRAPHIQUES**

## Références bibliographiques

- [1] A.R. Katritzky., D.C.Fara., R.O. Petrukhin, Top. Med. Chem, 2002 ,1333-1356.
- [2] CHansch, a quantitative approach to biochemical structure activity relationships. Accounts of chemical research, 1969, 2, 232-239.
- [3] R. Leardi, Chemometr. 2001 ,15, 559-569.
- [4] A.Z. Dudek, T. Arodz, J. Gàlvez. Computational methods in developing quantitative structure-activityrelations (QSARs), A review, Combinatorial Chemistry and High Through put Screening, 2006, 9, 213–228.
- [5] IUPAC., Pure and Applied Chemistry. 2002 ,74, 1823–1830.
- [6] Neely., Partition Coefficients to Measure Bioconcentration Potential of Organic Chemicals in Fish. Environ. Sci. Technol. 1974, 8, 1113-1115.
- [7] P, Le Cloirec, Introduction au traitement de l'air, Les techniques de l'ingénieur Traité environnement ,2002 ,1-8.
- [8] K. Saadi. Contribution à l'étude de la Relation structure chimique- odeur Utilisation de la technique Random Forest (Application à la famille des pyrazines ), Mémoire de Magister. UNIVERSITÉ KASDI Merbah Ouragla ,2009, 30.
- [9] B. Souyei, contribution à la prédiction de coefficient de partage Octanol /eau par la technique QSPR. 2007, 64 -68.
- [10] D. Chessel, J. Thioulouse, S. Dray, Co-inertia Analysis and the Linking of Ecological Tables. Ecology. 2003, 84 ,3078-3089.
- [11] R. C. Reid, J. M. Prausnitz, B. E. Poling. The Properties of Gases and Liquids, 4 th ed. ;McGraw Hill. New York, 1987.
- [12] ChemDraw Release 7.0.1 Chemical Structure Drawing Standard, 2002.
- [13] Hyperchem™ Release 7.5 for windows, Molecular Modelling system, 2000.
- [14] T. Thomas-Danguin. Intensité olfactive des composés purs et demélanges : application au masquage des odeurs, Université Claude Bernard, Lyon,1997,224.

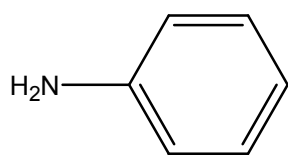
- [15] R. Todeschini, V. Consonni, M. Pawan, DRAGON, Software for the calculation of Molecular Descriptors. Release 5.3 for Windows, Milano, 2005.
- [16] A. R. Katritzky, V. S. Lobanov, M. Karelson. CODESSA Reference Manual. University of Florida, Gainesville, 1994.
- [17] V. Y. Nalimov. The Application of Mathematical Statistics to Chemical Analysis, Addison- Wesley, Reading, MA ,1962.
- [18] R. Calcutt, R. Body. Statistics for Analytical Chemists. Chapman & Hall, New York, 1983.
- [19] J. C. Miller, J. N. Miller. Statistics for Analytical Chemistry. Ellis Horwood, New York, 1988.
- [20] P. C. Meier, R. E. Zund. Statistical Methods in Analytical Chemistry. Wiley, New York , 1993.
- [21] P. Dagnélie. Statistique Théorique Et Appliquée. Tomes 1 et 2. De Boeck & Larcier s. à, 1998.
- [22] R. Tomassone, E. Lesquoy, C. Miller. La régression : nouveaux regards sur une ancienne méthode statistique. Masson, INRA ,1983.
- [23] L. Chambers. Practical Handbook of Genetic Algorithms. Lewis Publishing, 1995.
- [24] R. Todeschini. MOBY DIGS Software for Multilinear Regression Analysis and Variable Subset Selection by Genetic Algorithm. Release for Windows. Milano Srl.
- [25] MINITAB Release 16.2.0.0 for Microsoft language pack 2.
- [26] N.R Draper, H. Smith, Applied Regression Analysis, Third Edition, Wiley series in Probability and Statistics, New York, 1998.
- [27] L. Eriksson, J. Jaworska, A.P. Worth, M.T.D. Perspective, 111(10), 2003, 1361-1375.
- [28] A. Boudellioua, Modification d'électrode par électro polymérisation de l'aniline : Application à l'électro analyse, mémoire de master, UNIVERSITE BADJI MOKHTAR-ANNABA, 2018.

- [29] A.Ouasdi, F.Ferhat, Modélisation et Optimisation de la Complexation par la  $\beta$ -Cyclodextrine d'un polluant organique modèle : l'Aniline, mémoire de master, Université A. MIRA – BEJAIA, 2013.
- [30] F. J. Marques Neves, Modelling and Optimization of Large-Scale Processes: Application to the liquid-phase aniline production the liquid-phase aniline production", Thèse de Doctorat, Université de Coimbra, 2007.
- [31] F. Afoufou, N. Guesbaya, S. Achour, Effet de la minéralisation des eaux naturelles sur l'élimination de composés organiques aromatiques par coagulation floculation, Courrier du Savoir – N°08, Juin (2007), 75-81.
- [32] Aniline. European Union Risk Assessment Report. Vol. 50. European Chemical Bureau; 2004 ([http:// ecb.jrc.it](http://ecb.jrc.it)).
- [33] Aniline. In : HSDB. NLM, 2005 ([http:// toxnet.nlm.nih.gov](http://toxnet.nlm.nih.gov)).
- [34] M. R. Nabid and A. A. Entezami, Comparative study on the enzymatic polymerization of N-substituted aniline derivatives .Polym. Adv. Technol. in press 2004; DOI: 10.1002/pat.582.
- [35] N. Gospodinova, L. Terlemezyan, conducting polymers prepared by oxidative polymerization: polyaniline, Prog. Polym.Sci., Vol. 23, 1998, 1443–1484.
- [36] E. Bombard, B. Herbold, Génotoxique activities of aniline and its metabolites and their relationship to the carcinogenicity of aniline in the spleen of rats, Critical Reviews in Toxicology 35, 2005, 783-835.
- [37] H. Ma, J. Wang, Z. Sherif. Abdel-Rahman, T.K. Hazra, P. J. Boor, M.F. Khan, Induction of NEIL1 and NEIL2 DNA glycosylases in aniline-induced splenic toxicity. Toxicology and Applied Pharmacology 251, 2011, 1–7.
- [38] Fisher et al., Boson localization and the super fluid-insulator transition. M. Fisher, P. Weichman, G. Grinstein, et D. Fisher. Phys. Rev. B, 40, 1989, 546.
- [39] K, Satyanarayana, M.C. Kakati: Correlation of Flash Points, Fire and Materials 15, str 1991, 97-100.

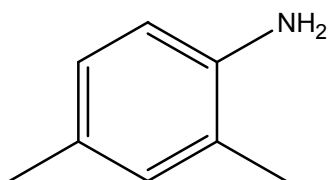
[40] C.Y.Okada, M. Rechsteiner. Introduction of macromolecules into cultured mammalian cells by osmotic lysis of pinocytic vesicles. *Cell*, 29(1),1982 , 33-41.  
[https://doi.org/10.1016/0092-8674\(82\)90087-3](https://doi.org/10.1016/0092-8674(82)90087-3)

# **ANNEXE**

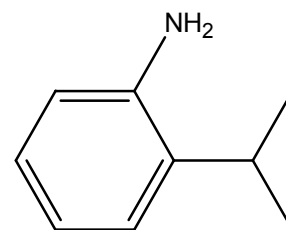
## ANNEXE



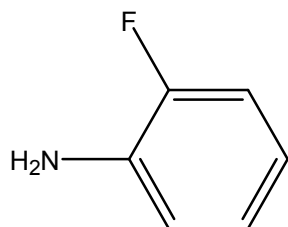
Aniline



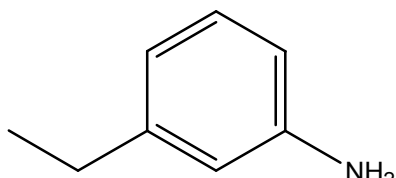
2,4-dimethylaniline



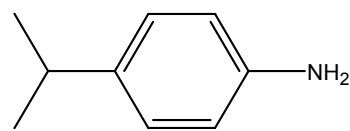
2-Isopropylaniline



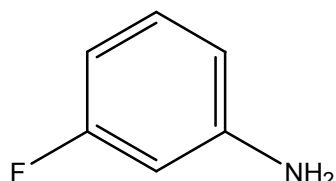
2-Fluoroaniline



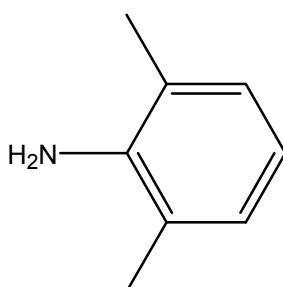
3-Ethylaniline



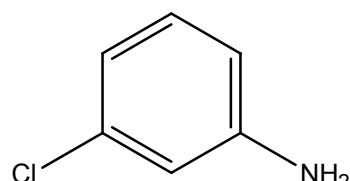
4-Isopropylaniline



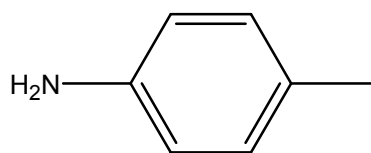
3-Fluoroaniline



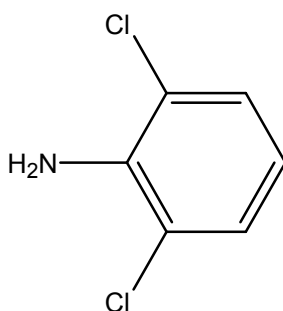
2,6-dimethylaniline



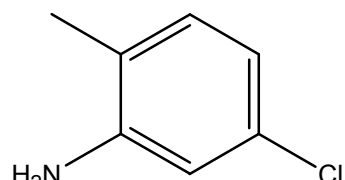
3-Chloroaniline



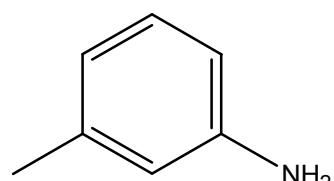
4-Methylaniline



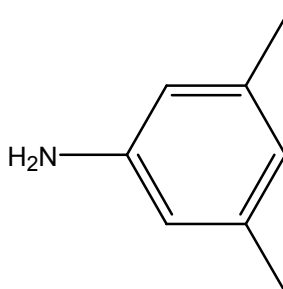
2,6-dichloroaniline



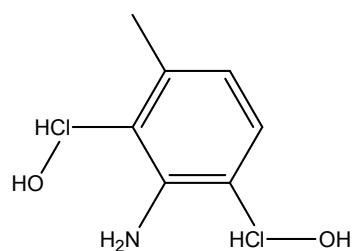
5-chloro-2-methylaniline



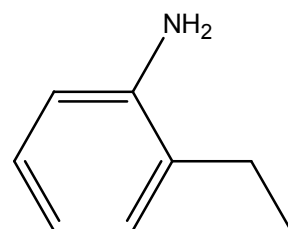
3-Methylaniline



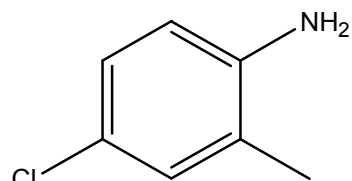
3,5-dimethylaniline



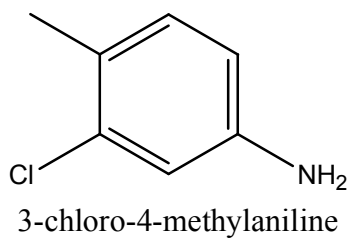
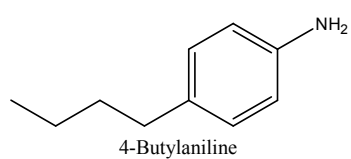
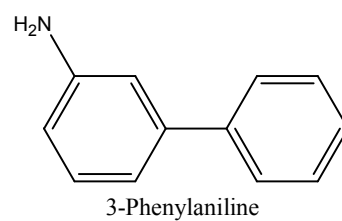
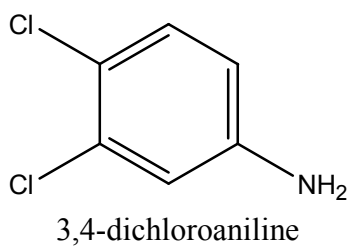
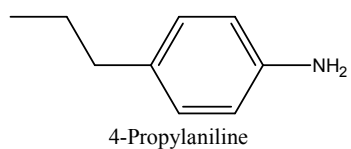
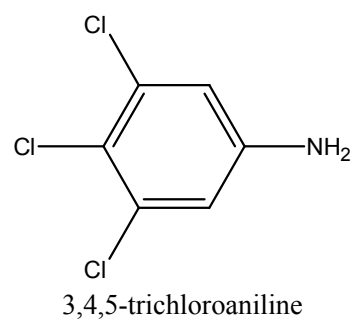
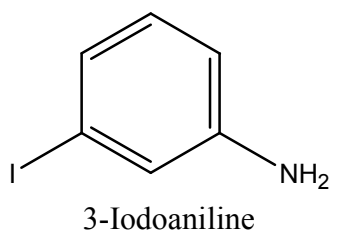
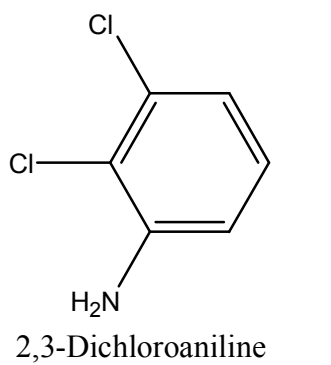
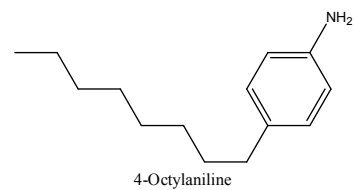
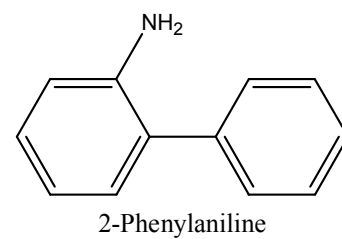
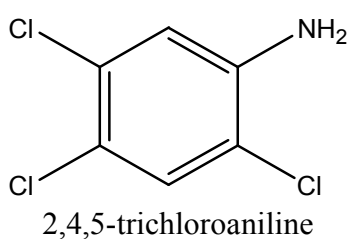
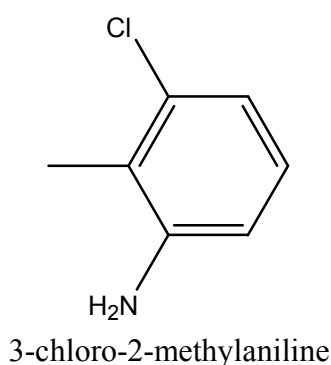
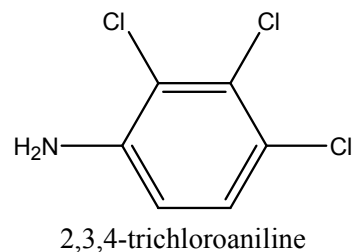
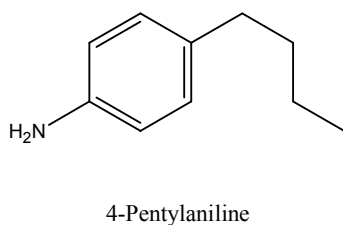
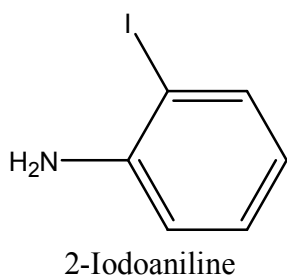
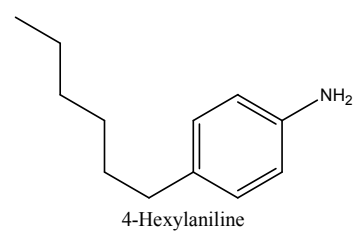
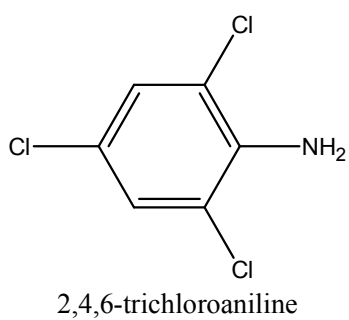
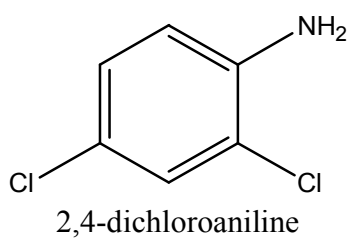
2,6-dichloro-3-methylaniline

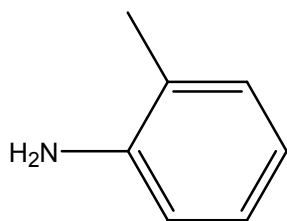


2-Ethylaniline

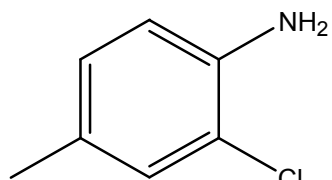


4-chloro-2-methylaniline

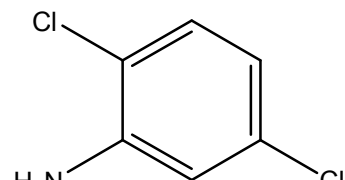




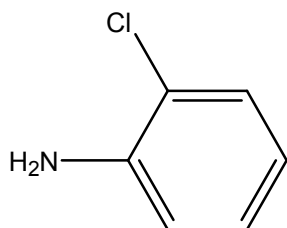
2-Methylaniline



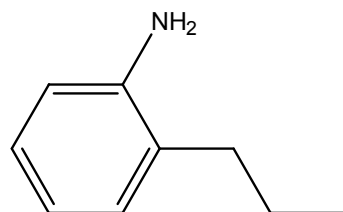
2-chloro-4-methylaniline



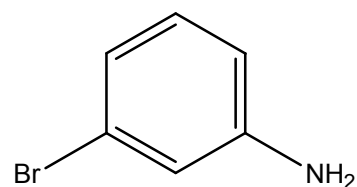
2,5-dichloroaniline



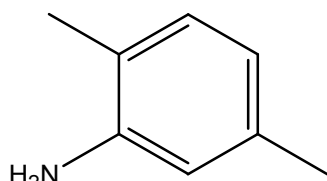
2-Chloroaniline



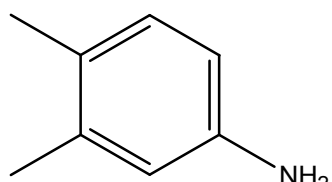
2-Propylaniline



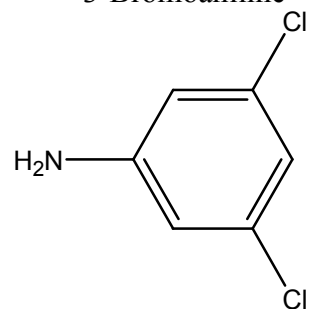
3-Bromoaniline



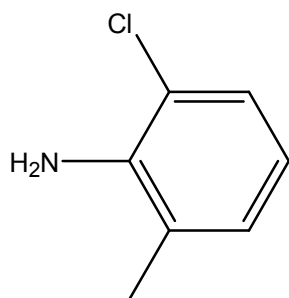
2,5-dimethylaniline



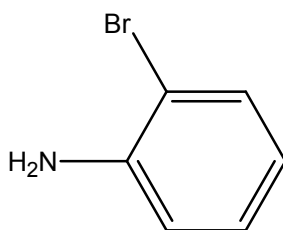
3,4-dimethylaniline



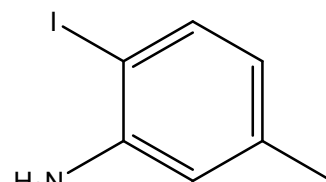
3,5-dichloroaniline



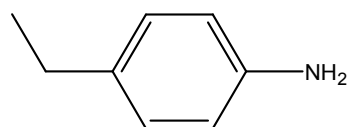
2-chloro-6-methylaniline



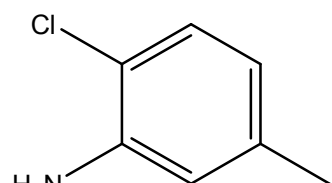
2-Bromoaniline



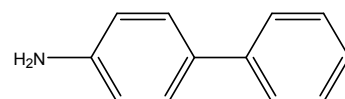
2-iodo-5-methylaniline



4-Ethylaniline



2-chloro-5-methyl aniline



4-Phenylaniline

# RESUMES

## Résumé :

Un modèle QSPR a été développé pour la prédiction de la température d'ébullition. Les données, concernant 50 anilines, ont été séparées en deux sous-ensembles disjoints comprenant respectivement 35 éléments pour le calcul et le test (éventuel) du modèle, et 15 éléments pour sa validation externe.

Des descripteurs moléculaires théoriques ont été calculés en utilisant des logiciels de modélisation moléculaire du commerce DRAGON. La taille du modèle a été déterminée en prenant la valeur optimale du  $R^2$  et la sélection des descripteurs réalisée par algorithme génétique.

Les valeurs des paramètres statistiques ( $R^2$ ,  $Q^2$ , SDEC, SDEP, SDEPext) obtenues attestent de la pertinence des modèles développés.

## Mots-clés:

Anilines – Température d'ébullition – Descripteurs moléculaires théoriques – Modèle QSPR.

## Abstract

One model QSPR was developed for the prediction of the melting point. The data, concerning 50 aniline's, were separate in two disjoined subsets including/understanding respectively 35 elements for calculates and the possible test of the models, and 15 elements for its external validation.

Theoretical molecular descriptors were calculated by using software of molecular modeling of the trade DRAGON. Size of the model at summer determined by taking the optimal value of  $R^2$  and using genetic algorithm method's.

Values of the statistical parameter ( $R^2$ ,  $Q^2$ , SDEC, SDEP, SDEPext) obtained attest relevance of the model developed.

## Key words:

Aniline's - Boiling point – molecular descriptors theoretical – QSPR Model.

## ملخص : تم تطوير نموذج بطريفة الـ QSPR لتنبؤ درجة الغليان

المعطيات الخاصة بـ 50 انيلين تم تقسيمها الى مجموعتين الاولى تحتوي على 35 عنصر لحساب وتجريب النموذج اما الثانية تحتوي على 15 عنصر ل تصديق الخارجي للنموذج. المواصفات الجزيئية النظرية تم حسابها باستعمال برمجيات النمذجة الجزيئية المتوفرة في السوق. حجم النموذج تم تحديده عن طريق اخذ القيمة الملائمة لـ  $R^2$  اما تحديد النماذج تم بطريفة Algorithmes Génétique. قيم المعالم الاحصائية ( $R^2$ ,  $Q^2$ , SDEC, SDEP, SDEPext) المتحصل عليها تؤكد تعلق النموذج المطور.

الكلمات الدالة : انيلين-درجة الغليان- مواصفات جزيئية نظرية – نموذج QSPR