MINISTERE DE L'ENSIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE « Abbès LAGHROUR » DE KHENCHELA
FACULTE DES SCIENCES ET TECHNOLOGIE

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 :

PRÉDICTION DE LA TEMPÉRATURE D'ÉBULLITION D'UNE SÉRIE D'ÉTHERS

Réalisé par :

- **Bir Imen**
- **Dib Amel**

Dirigé par :

- D.KERTIOU.Noureddine

Membres de jury :

SAMAI Salima	*MCB*	*Présidente*
BOUAAKADIA Amel	*MCA*	*Examinatrice*

Promotion : 2020/2021.

Dédicaces

Je dédie cet humble travail:

À la flamme qui illumine ma vie, qui a tant souffert pour me voir arriver à mon dut,

À ma chère Mère **DALOULA**

À mon Père **CHAABANE** qui me guide vers le pas de la réussite et savoir

Aucun dédicace ne saurait exprimer l'amour, l'estime, le dévouement et le respect que j'ai toujours pour vous. Ce travail est le fruit de tes sacrifices que tu as consentis pour mon éducation et ma formation le long de ces années.

À mes frères NADJM EDDINE et ZAKARIA

À mes *sœurs* ZINEB et KHAOULA

À mes grands-pères et ma grand-mère

À mes oncles et mes tantes À mon binôme AMEL

À toutes mes chères amies, AMEL,ARIDJ, AYA, SALEMA ; ASMA, KHAOULA, MANAR

HADIL, RAYANE, MALAK,

RITADJE , AYMEN ,SALAH, ZINNE EDDINE.

À toutes mes camarades surtout les étudiants de Chimie analytique et Chimie général

Merci

…Imen

Dédicaces

Je dédie ce travail à :

A me très chère mère ZAHRA

Quoi que je fasse ou que je dise, je ne saurai point te remercier comme il se doit. Ton affection me couvre, ta bienveillance me guide et la présence à mes côtés a toujours été me source de force pour affronter les différents obstacles.

A mon très cher père SÉFI

Tu as toujours été à mes côtes pour me soutenir et m'encourager. Que ce travail traduit ma gratitude.

A mon très cher frère BILEL

Ma cousine FATIMA

A toute ma famille

A mes tantes et mes oncles

A mon binôme IMEN

A tous mes chères amis IMEN, ARIDJ, ASMA, KHAOULA, ILHEM, RAYEN, RAHIMA, JIHEN, AHLEM, NASSIMA, SARA, AYMEN, ZINE EDDIN

A tous mes camarades surtout les étudiants de chimie analytique

Amel

Remerciements

Nous remercions avant, après, et à l'infini **ALLAH** de nous entourer de ses grâce et clair de ses lumière, et nous avoir donné la patience, le courage, pour établir ce modeste travail

Nos sincères remerciements à notre cher encadreur **KERTIOU Noureddine** pour avoir accepté de

diriger ce travail de Master. Nous le remercions vivement pour le temps qu'il a consacré, ces explications et sa patience.

Nous désirons vivement exprimer nos extrêmes gratitudes et nos remerciements à **Mme SAMAI Salima** d'avoir accepté de présider ce jury et **Mme Bouakkadia Amel** d'avoir accepté cette invitation à ce jury et d'examiné ce modeste travail.

Sans oublier de remercier tous nos professeurs qui nous avoir

aidés durant la réalisation de ce travail et toutes les personnes qui par leurs paroles, leurs écrits, leurs conseils.

Nous remercions notre famille surtout nos très chers parents pour le soutien financier, moral

Enfin, nous remercions nos amis et camarades des promotions précédentes,

dans les meilleurs moments comme dans les pires.

m e r c i

Sommaire

Liste des Tableaux……..i

Liste des Figures……ii

Symboles et Abréviations……iii

INTRODUCTION GÉNÉRALE……2

PARTIE THEORIQUE

GENERALITES……5

1-Introduction 5

I- Les relations structure-propriété/activité quantitatives…….5

I-1.Définition de la méthode QSPR 5

I-2. Principe de la méthode QSPR.....6

I-3. Les méthodes mathématiques utilisés par le modèle QSPR.....7

I-4. La relation structure-rétention quantitative (QSPR) 7

I-5. COLLECTE DES DONNEES 8

II- Optimisation de la géométrie moléculaire et génération des descripteurs moléculaires……11

II-1. Préparation de base des données :.....11

II-1.1 Logiciels « ChemDraw 7 ».....11

II-1.2 Logiciels « HyperChem 8 » : 12

II-2. Récupération et stabilisation des molécules de fichier Hin 12

II-2.1 Stabilisation structure des molécules (minimisation de l'énergie).....12

II-2.2 Mécanique Moléculaire 12

II-2.3 Récupération des fichiers HyperChem HIN.....13

III- Calcul des descripteurs moléculaires…….14

III-1. Le Logiciel « Dragon » 14

III-2. Descripteurs moléculaires 15

III-2.1 Définition d'un descripteur : 15

III-2.2 Groupe des descripteurs moléculaires 15

III-2.2.1 Les descripteurs empiriques 15

III-2.2.2 Les descripteurs moléculaires théoriques 16

III-2.2.3 Les descripteurs géométriques 16

III-3. Importance des descripteurs.....18

III-4. Diagramme montre les étapes de prédiction 19

III-5. L'objectif de la prédiction.....20
IV- Méthodes utilisées pour le développement de modèles QSAR/QSPR…….21
IV-1. Introduction.....21
IV-2. Méthodes de régressions linéaire et multilinéaire. 23
IV-2.1 Aperçu général.....23
IV-2.2 Evaluation préliminaire des données.....24
III-2.3 Régression linéaire multiple 24
IV-3. Paramètres d'évaluation de la qualité de l'ajustement.....27
IV-4. Validation externe 28

PARTIE APPLICATION

1-Sélection des descripteurs.....31
2-Calcul des corrélations entre les différents descripteurs.....31
3- Calcul de l'équation de régression 32
4- Choix du modèle : 32
5-Matrice de corrélation 33
6-Equation de régression 36
7- Analyse de régression 37
8-Diagramme de Williams : 39
9-Vérification de la qualité de l'ajustement : 40
10-Test de randomisation 41
11 -Validation externe.....42
Conclusion Générale…….…..46
Références Bibliographiques…….48

LISTE DES TABLEAUX

TABLEAU	TITRE	PAGE
1	Nomenclature et valeurs de la propriété étudiée des éthers	08
2	Quelques blocks des descripteurs calculés par logiciel dragon	17
5	Valeurs des descripteurs moléculaires sélectionnés	34
4	Classes et significations des descripteurs	33
3	Corrélations Tb avec les descripteurs	33
6	Analyse de régressions ; Tb avec les 5 descripteurs	36
7	Valeurs des Tb expérimentales, calculées, ecal, eistd et hii pour l'ensemble de calibration	37
8	Valeurs des paramètres statistiques pour l'ensemble de calibration	39
9	Valeurs des Tb expérimentales, calculées, ei, eistd et hii pour l'ensemble de validation	43
10	Valeurs des Q^2ext et SDEPext	43

LISTE DES FIGURES

Figure	Titre	page
1	Modèle de l'étude de relation structure activité	07
2	Représentation des molécules par le ChemDraw	11
3	Le logiciel Hyperchem	13
4	Le logiciel DRAGON	14
5	Diagramme de prédiction par QSPR	19
6	Le cycle de prédiction	20
7	Diagramme de notre travail	21
8	Variation de R^2 en fonction de nombre de descripteurs k	32
9	Diagramme de Williams	39
10	Les 4 composés influents	40
11	Graphe des valeurs calculées en fonction des valeurs observées pour l'ensemble de calibration	41
12	Graphe des valeurs prédites en fonction des valeurs observées pour l'ensemble validation	41
13	Test de randomisation associé au modèle QSPR	42

SYMBOLES ET ABREVIATIONS

ACP : Analyse en composantes principales.

AG: Algorithme génétique (Genetic Algorithm).

ei: Résidu différence entre les valeurs observées (y_i) et estimée ($\hat{y}_i$).

F: Statistique de Fisher.

FIV : Facteur d'inflation de la variance.

H : Matrice de projection, ou matrice chapeau.

hii : Eléments diagonaux de la matrice chapeau.

k : Nombre de descripteurs.

LMO : leave – many- out.

LOO: leave – one – out.

MCO : Les moindres carrés ordinaires

MCP: les moindres carrés partiels.

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).

PLS: Moindres carrés partiels.

PRESS: Somme des carrés des erreurs de prédiction.

Q^2_{LOO}: Coefficient de prédiction.

Q^2ext : Coefficient de prédiction externe.

QSAR: Quantitative Structure/ Activity Relationships. Relations Structure/Activités Quantitatives).

QSPR: Quantitative Structure/ Propriety Relationships. Relations Structure/Propriétés Quantitatives).

R^2: Coefficient de détermination.

RMSE : Root Mean Squared Error.

RNA : Réseau de neurones artificiel.

S : Erreur standard.

SCE: Somme des carrés des écarts.

SCT: Somme des carrés totale.

SDEC: Standard Deviation Error in Calculation : Déviation standard de l'erreur calculée.

SDEP: Standard Deviation Error of Prediction : Déviation standard de l'erreur de prédiction.

$SDEP_{ext}$: External Standard Deviation Error of Prediction: Déviation standard de l'erreur de prédiction externe.

Tf : Température de diffusion.

t: t de Student.

X : Matrice des valeurs observées des variables explicatives.

X' : Matrice transposée de X .

y: Vecteur de dimension n.

yi : Valeur observée.

ŷi: Valeur estimée.

Introduction

Les éthers sont largement utilisés comme solvants des résines, des laques, des peintures, des vernis, des colorants et des encres, comme composants des pâtes colorantes, des nettoyants, des savons liquides, des produits cosmétiques et des fluides hydrauliques. Les éthers de propylène-glycol et de butylène-glycol sont de bons dispersants et d'excellents solvants des laques, peintures, résines, colorants, huiles et graisses.

La température d'ébullition d'un composé reflète les interactions entre molécules du liquide, ainsi que la différence entre les fonctions de partition moléculaires internes dans le liquide et le gaz obtenue à l'ébullition.

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.

Comme, dans la pratique, il n'est pas toujours certain de disposer de données expérimentales fiables, le recours aux méthodes d'estimations, qui permettent un gain de temps et d'argent, est incontournable.

Les relations quantitatives stucture/propriété (QSPR) très utilisées depuis la fin des années 1980, constituent des modèles mathématiques pour l'approximation des relations souvent complexes, entre la structure caractérisée par des nombres (appelés descripteurs moléculaires) et les propriétés des composés.

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 la méthode des algorithmes génétiques et la régression multilinéaire : MLR , pour modéliser la température d'ébullition de 80 éthers . 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 ainsi que les résultats obtenus.

Partie théorique

GENERALITES

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- Les relations structure-propriété/activité quantitatives

I-1.Définition de la méthode QSPR

La relation quantitative structure propriété (QSPR) est le processus par lequel la structure chimique est liée quantitativement avec une propriété physicochimique bien définie, telle que les températures d'ébullition, de fusion ou la pression critique [2].

Les grands axes de développement d'un modèle QSPR peuvent être décrits comme suit :

- Choisir des descripteurs adaptés au problème structure – propriété.
- Définir une relation qui les corrèle à la propriété en question.
- Etablir des critères de performance et de validation qui aideront au choix des meilleurs modèles et estimer des incertitudes de prédiction [3].

I-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 physicochimique,…, etc.), pour une série de composés chimiques similaires à l'aide de méthodes d'analyses de données. Aussi la forme générale de modèle est :

$$\textbf{Propriété = f(Descripteurs)} \qquad (01)$$

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.

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 [5].

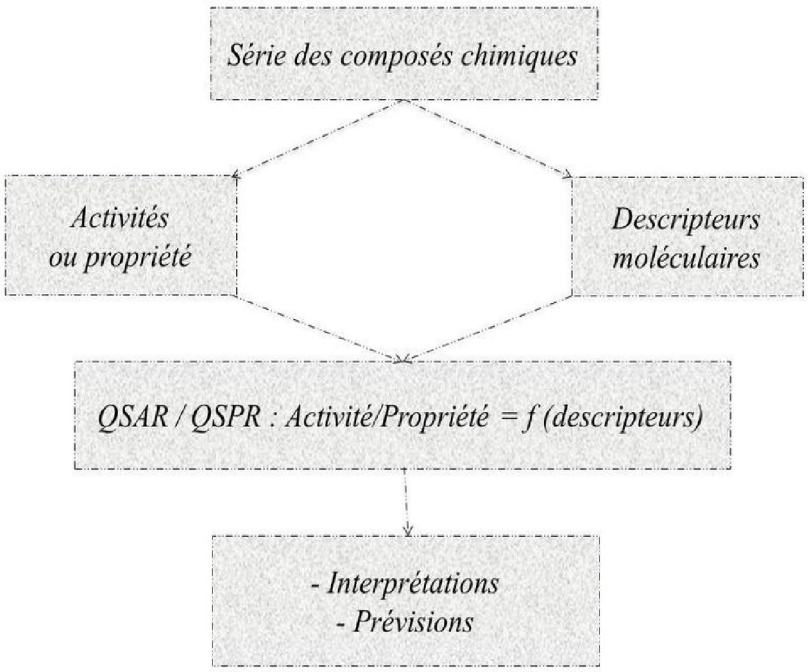


FIGURE-01 : Modèle de l'étude de relation structure activité

I-3. Les méthodes mathématiques utilisés par le modèle QSPR

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

I-4. La relation structure-rétention quantitative (QSPR)

Au cours des décennies passées, les Relations Quantitatives Structure- Activité/ Propriété (QSAR) sont devenues un puissant outil théorique, alternatif à la mécanique quantique, pour la description et la prédiction des propriétés des systèmes moléculaires complexes dans différents environnements. L'approche QSAR procède de l'hypothèse d'une correspondance univoque entre n'importe quelle propriété physique, affinité chimique, ou activité biologique d'un composé chimique et sa structure moléculaire [6]. Cette dernière peut être représentée par la composition chimique, la connectivité des atomes, la surface d'énergie potentielle, et la fonction d'onde électronique d'un composé. Différents descripteurs

moléculaires physico- chimiques reflétant la structure peuvent être déterminés empiriquement ou en utilisant des méthodes théoriques et computationnelles de différentes complexités. Il est à souligner que la connaissance de la constitution chimique exacte et/ou de la structure moléculaire tridimensionnelle des composés chimiques étudiés est un pré- requis à l'application de l'approche QSAR. Le succès de l'approche QSAR dépend de façon critique de la définition précise et de l'utilisation appropriée des descripteurs moléculaires. On distingue, arbitrairement, les descripteurs moléculaires empiriques des descripteurs moléculaires théoriques.

I-5. COLLECTE DES DONNEES

La propriété considérée (température d'ébullition) est industriellement importante dans la perspective du génie chimique. Les données utilisées dans ce travail, prélevées dans [7], concernent les 65éthers sont réunies dans le tableau 01.

Tableau- 01 Nomenclature et valeurs de propriété des éthers étudiés.

N	Composé	Tb
1	Dimethylether	248.31
2	Diethylether	307.59
3	Methyl t-butylether	328
4	Di-n-propylether	363.08
5	Di-isopropylether	359.4
6	2-Methylfuran	337.7
7	Tetrahydrofuran	338
8	Tetrahydropyran	361
9	1,4-Dioxane	374.5
10	Epichlorohydrin	391
11	Chloromethylmethylether	332.5
12	Bis(chloromethyl)ether	379
13	Bis(2-chloroisopropyl)ether	460

Tableau - 01 (suite)

N	Composé	Tb
14	2-Chloroethyl vinylether	381
15	Bis(2-chloroethoxy)methane	488
16	Anisole	426.7
17	2-Chloroanisole	471.5
18	3-Chloroanisole	466.5
19	4-Chloroanisole	470.2
20	2,4,6-Trichloroanisole	514
21	2,3,4,5-Tetrachloroanisole	622.8
22	3,4,5-Trichloroveratrole	563
23	Tetrachloroveratrole	659.5
24	Phenetole	442.81
25	Benzylethylether	458.6
26	Styreneoxide	467.1
27	Diphenylether	531
28	2-Chlorodiphenyl ether	540
29	4-Chlorodiphenyl ether	557.5
30	2,4-Dichlorodiphenyl ether	273
31	2,4,5-Trichlorodiphenyl ether	598
32	2,4,4',5-Tetrachlorodiphenyl ether	627
33	2,2',4,4',5-Pentachlorodiphenyl ether	638
34	3,3',4,4',5-Pentachlorodiphenyl ether	675
35	2,2',3,4,4',6'-Hexachlorodiphenyl ether	652
36	2,2',4,4',5,5'-Hexachlorodiphenyl ether	663
37	2,2',4,4',5,6'-Hexachlorodiphenyl ether	649
38	2,3',4,4',5,5'-Hexachlorodiphenyl ether	682
39	2,2',3,4,4',6,6'-Heptachlorodiphenyl ether	660

Tableau-01 (suite et fin)

N	Composé	Tb
40	2,2',3,3',4,4',5,6'-Octachlorodiphenyl ether	701
41	2,2',3,3',4,4',5,5',6-Nonachlorodiphenyl ether	716
42	Decachlorodiphenylether	731
43	4,4'-Dibromodiphenyl Ether	612
44	2,4,4'-Tribromodiphenyl ether	643
45	2,2',4,4'-Tetrabromodiphenyl ether	669
46	2,2',4,4',6-Pentabromodiphenyl ether	707
47	2,2',4,4',5,5'-Hexabromodiphenyl ether	726
48	2,2',4,4',5,6'-Hexabromodiphenyl ether	744
49	Decabromodiphenylether	803
50	Butylethylether	365.3
51	Di-n-butylether	413.28
52	1,2 propyleneoxide	308
53	Furan	304.5
54	Bis(2-chloroethyl)ether	451.5
55	2,3-Dichloroanisole	389
56	2,6-Dichloroanisole	390
57	2,3,4-Trichloroanisole	534
58	2,3,5,6-Tetrachloroanisole	559.9
59	Veratrole	479
60	4,5-Dichloroveratrole	530
61	2,2',4,4'-Tetrachlorodiphenyl ether	611
62	3,3',4,4'-Tetrachlorodiphenyl ether	648
63	4-Bromodiphenyl ether	578
64	2,2',4,4',5-Pentabromodiphenyl ether	689
65	2,3,3',4,4',5'-Hexabromodiphenyl ether	749

- De 1 à 49 les composés de calibration.
- De 50 à 65 les composés de validation.

II- Optimisation de la géométrie moléculaire et génération des descripteurs moléculaires

II-1. Préparation de base des données :

II-1.1 Logiciels « ChemDraw 7 »

ChemDraw Pro [8] 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.

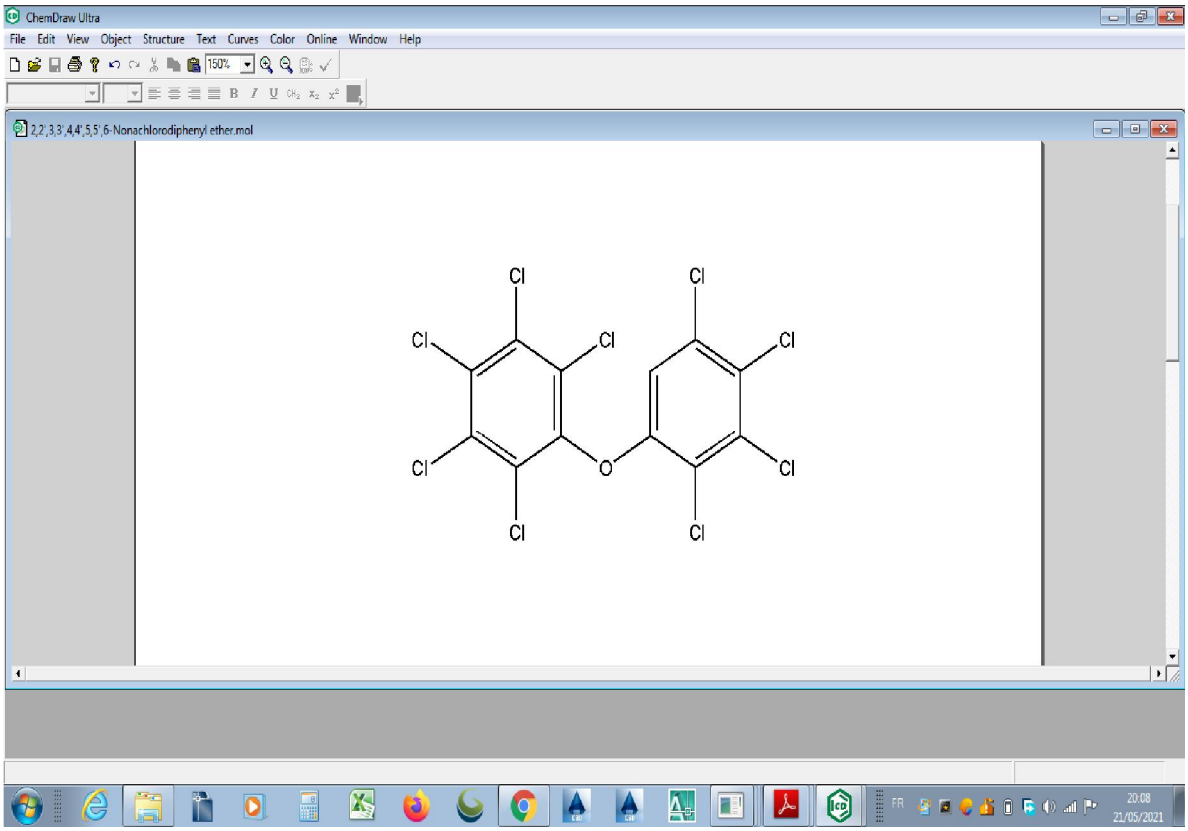


FIGURE- 02 *Représentation des molécules par le ChemDraw*

II-1.2 Logiciels « HyperChem 8 » :

HyperChem [9] 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 vous permet de faire réellement de la modélisation : il possède plus de méthodes de calculs (Mécanique Moléculaire, semi-empirique et ab- initio) pour que vous puissiez calculer plus de propriétés.

HyperChem est utilisé dans cette étude pour construire et optimiser les molécules, chaque molécule est enregistrée comme un fichier nommé "Hin" après l'optimisation.

Nous avons utilisé la méthode semi empirique MM+ pour l'optimisation. On a 65 molécules donc on obtient 65 fichiers Hin, en suite on calcule les descripteurs moléculaires à partir de ces fichiers par le logiciel Dragon [10].

II-2. Récupération et stabilisation des molécules de fichier Hin

II-2.1 Stabilisation structure des molécules (minimisation de l'énergie)

Pour stabiliser la forme de structure de chaque molécule ou minimisation de l'énergie on utilise le HyperChem, qui pouvez effectuer une minimisation de l'énergie (ou de la géométrie d'optimisation) d'une molécule 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 [11].

II-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

cautionnement [12]. HyperChem comprend quatre haut mécanique moléculaire des champs de force : de nouvelles implémentations de techniques élaborées et publiées par des groupes de recherche respectés mais nous dans ce travail on concerte à méthode MM+ de caractéristiques Field Force présente dans figure suivante :

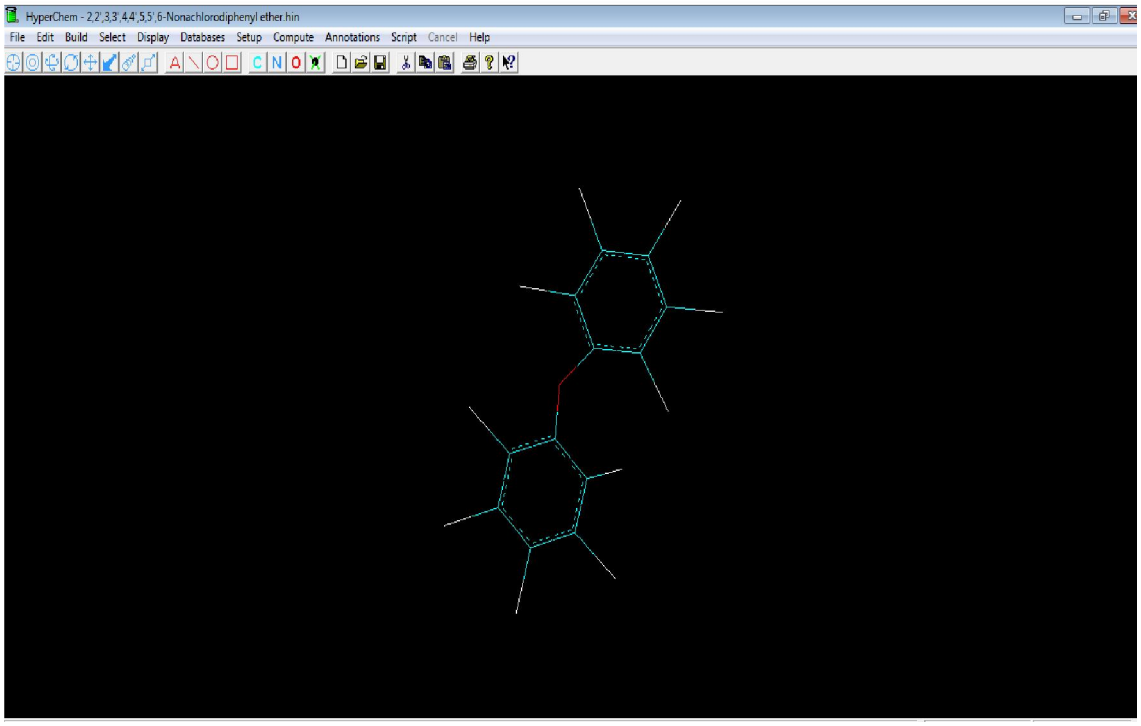


Figure- 03 : Le logiciel Hyperchem

II-2.3 Récupération des fichiers HyperChem HIN

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 construire deux fois ?! Nous pouvons le faire en allant dans Fichier et Enregistrer. Nous devons lui donner un nom comme Hin. Le fichier peut être rappelé pour la visualisation et la manipulation à une date ultérieure. Enregistrez-le dans le dossier public dans le répertoire approprié. Dans le calcul de descripteurs moléculaires, les structures chimiques des composés optimisés sont nécessaires. Les structures chimiques de tous les 65 composés dans notre jeu de données, ont

été établies dans le logiciel HyperChem, et pré optimisés en utilisant le champ avant MM+ mécanique [13].

III- Calcul des descripteurs moléculaires

Afin d'exploiter au maximum les informations contenues dans les structures moléculaires, celles-ci sont traduites en une série de grandeurs (en général scalaires) qui quantifient leurs caractéristiques physico-chimiques et structurelles. Dans la prochaine étape pour tous les composés, les descripteurs moléculaires ont été calculés par le logiciel dragon. Qui peut calculer plus de 1600 descripteurs moléculaires pour chaque structure dans notre jeu des données [14].

III-1. Le Logiciel « Dragon »

Dragon 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 la figure 4 [15].

Figure- 04 : Le Logiciel DRAGON

III-2. Descripteurs moléculaires

III-2.1 Définition d'un descripteur :

Le descripteur moléculaire est le résultat final d'une procédure logique et mathématique qui transforme l'information chimique chiffrée dans une représentation symbolique d'une molécule dans un nombre utile ou le résultat de quelque expérience standard.

Les descripteurs moléculaires sont les traits communs les plus considérables de structure moléculaire qui peut être utilisée pour développer la « Relation Structure – propriété ». Dans notre cas, la propriété est le point d'ébullition Tb des descripteurs moléculaires ont été proposés dérivé de théories différentes et approches avec le but de prédire biologique et propriétés physico-chimiques des molécules [16].

III-2.2 Groupe des descripteurs moléculaires

En effet, les descripteurs moléculaires sont basés sur plusieurs théories différentes, tel que quantum chimie, théorie de l'information, chimie organique, théorie du graphique, et ainsi de suite, et est utilisé pour modeler des plusieurs propriétés différentes de produits chimiques dans les champs du scientifique tel que toxicologie, chimie analytique, chimie physique…etc [17].

Actuellement, il est possible de calculer plus de 1600 descripteurs moléculaires qui sont représentés dans le tableau 2 qui peuvent être classés en 20 classes (blocs) logiques.

III-2.2.1 Les descripteurs empiriques

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

Les descripteurs structurels les plus répandus ont été définis pour quantifier les propriétés d'induction, l'effet mésomère ou de résonance, et les effets stériques des composés chimiques. Les descripteurs de solvatation reflètent les interactions du soluté avec la massedu solvant environnant (effets de solvant macroscopiques ou non spécifiques), et les liaisons

spécifiques, souvent des liaisons hydrogène entre le soluté et les molécules individuelles de solvant (effets de solvant spécifiques ou microscopiques) [18].

III-2.2.2 Les descripteurs moléculaires théoriques

Ils peuvent, conventionnellement, être répartis en un certain nombre de classes, selon leur complexité ou leur méthode de calcul. Les descripteurs théoriques les plus simples sont des descripteurs constitutionnels qui peuvent être construits à partir de l'information sur la composition chimique du composé considéré.

Les nombres, absolus et relatifs, des différents types d'atomes et de liaisons chimiques, la masse molaire, et le nombre de différents cycles dans le composé représentent quelques descripteurs constitutionnels typiques [19].

III-2.2.3 Les descripteurs géométriques

Ils sont obtenus à partir de la structure tridimensionnelle des molécules définie par les coordonnées des noyaux atomiques et la grosseur de la molécule représentée, par exemple, par le rayon atomique de Van der Waals.

Les molécules de la plupart des composés chimiques possèdent une certaine flexibilité conformation elle et les surfaces de potentiels moléculaires respectives possèdent de multiples minima locaux. Selon la structure de la molécule, le nombre de ces minima peut être très grand et, par conséquent, il est plutôt difficile de trouver le minimum d'énergie global pour des conditions expérimentales établies.

En effet, les descripteurs moléculaires sont basés sur plusieurs théories différentes, tel que quantum chimie, théorie de l'information, chimie organique, théorie du graphique, et ainsi de suite, et est utilisé pour modeler des plusieurs propriétés différentes de produits chimiques dans les champs du scientifique tel que toxicologie, chimie analytique, chimie physique…etc [17].

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) logiques.

Tableau- 2: Quelques blocks des descripteurs calculés par logiciel dragon

Classe	**Sous- classe**
Descripteurs constitutionnels	- Dénombrement des atomes ou des liaisons. - Descripteurs basés sur les masses atomiques.
Descripteurs topologiques	- Indices topologiques (connectivité, Winner, Bala ban…). - Descripteurs théoriques d'information. - Descripteurs topo chimiques.
Descripteurs géométriques	- Descripteurs liés à la distance. - Descripteurs lies à l'aire de la surface. - Descripteurs liés au volume. - Descripteurs du champ stérique moléculaire.
Descripteurs liés à la distribution de charge	- Charges atomiques partielles. - Moments électriques moléculaires - Polarisabilités moléculaires. - Descripteurs du champ électrique moléculaires
Descripteurs liés aux orbitales moléculaires	- Energie des OM frontières - Ordres de liaison - Indices de réactivité de Fukui
Descripteurs température	- Fonctions thermodynamiques
Dépendants	- Descripteurs facteurs de Boltzmann pondérés.

Tableau -2 (suite et fin)

Classe	**Sous- classe**
Descripteurs de solvatation	- Energie électrostatique de solvatation. - Energie de dispersion de solvatation. - Enthalpie libre de formation de cavité. - Descripteurs de liaison hydrogène. - Entropie de solvatation. - Descripteurs d'énergie de solvatation linéaire théorique.
Descripteurs mixtes	- Descripteurs topographiques. - Descripteurs électro topologiques. - Descripteurs de la charge partielle de l'aire de la surface.

III-3. 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 :

- ✓ Indiquer la description de la configuration de la molécule à étudier.
- ✓ Décrire 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 [20].

III-4. Diagramme montre les étapes de prédiction

Les premiers essais de modélisation QSPR développé par Wiener, 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 [21]

La prédiction des propriétés s'effectué par plusieurs étapes intéressantes illustrant dans le diagramme suivant :

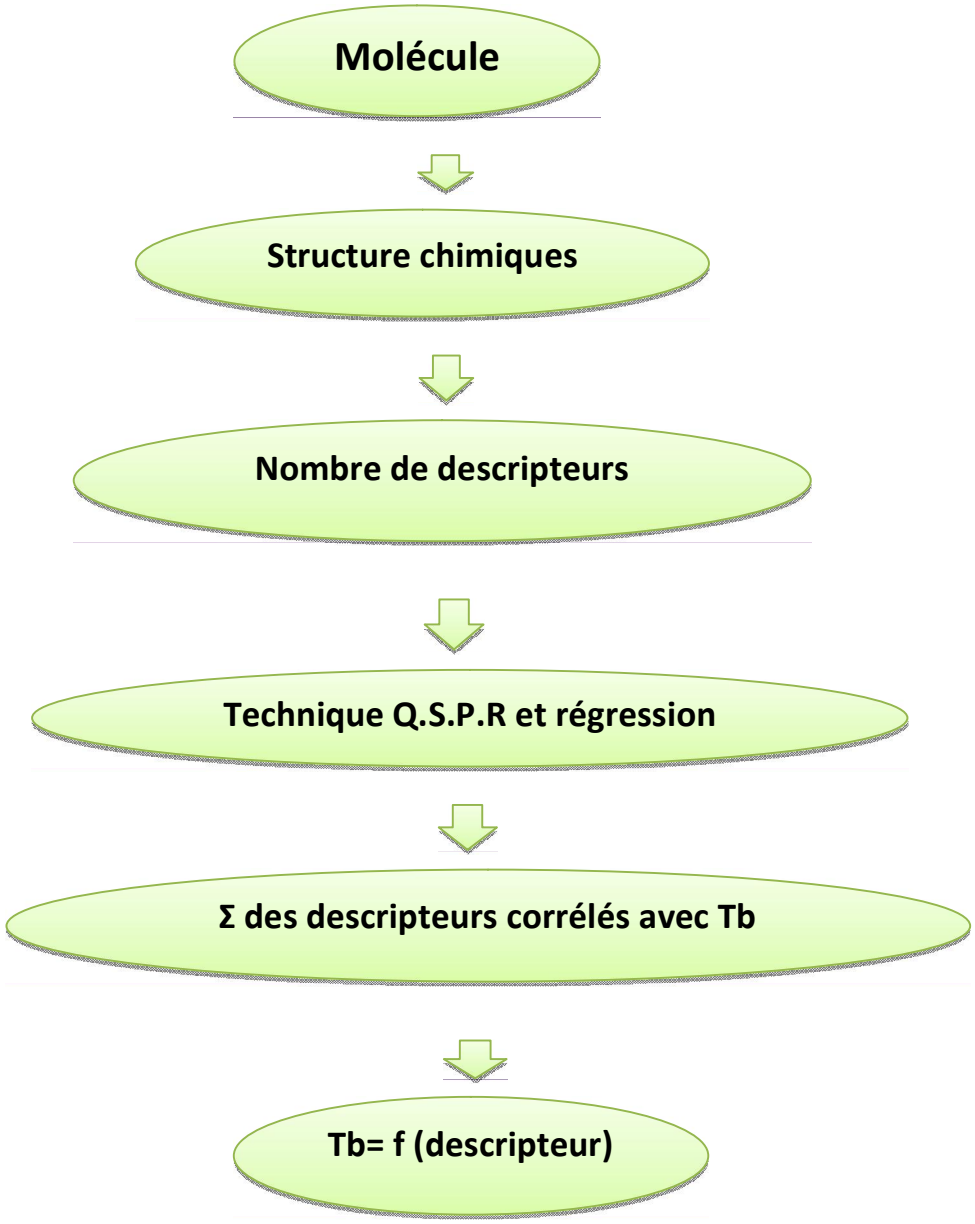


Figure- 05 Diagramme de prédiction par QSPR.

III-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.

Le cycle de prédiction est représenté dans la ci-dessous

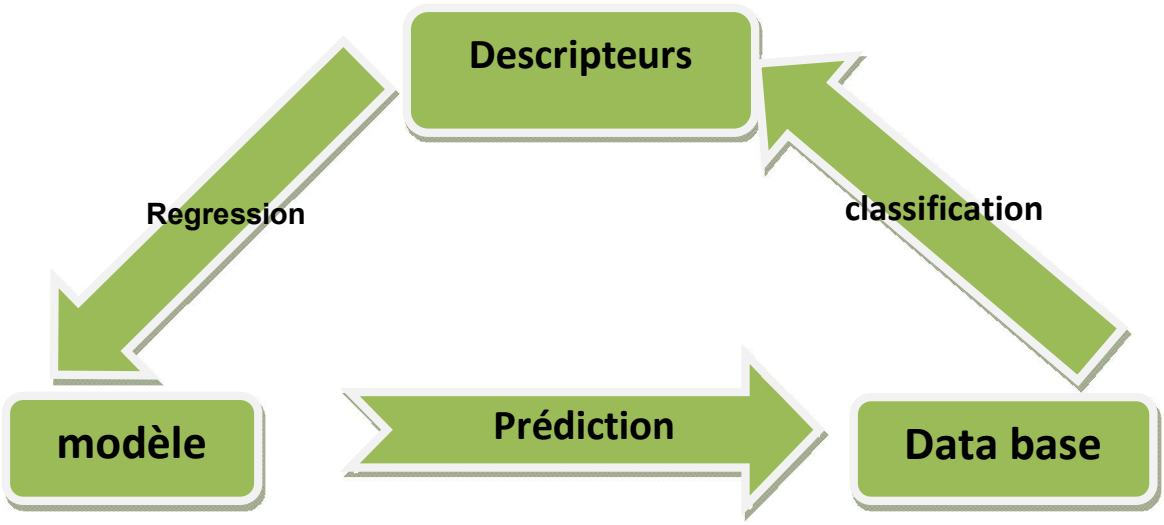


Figure- 06 Le cycle de prédiction

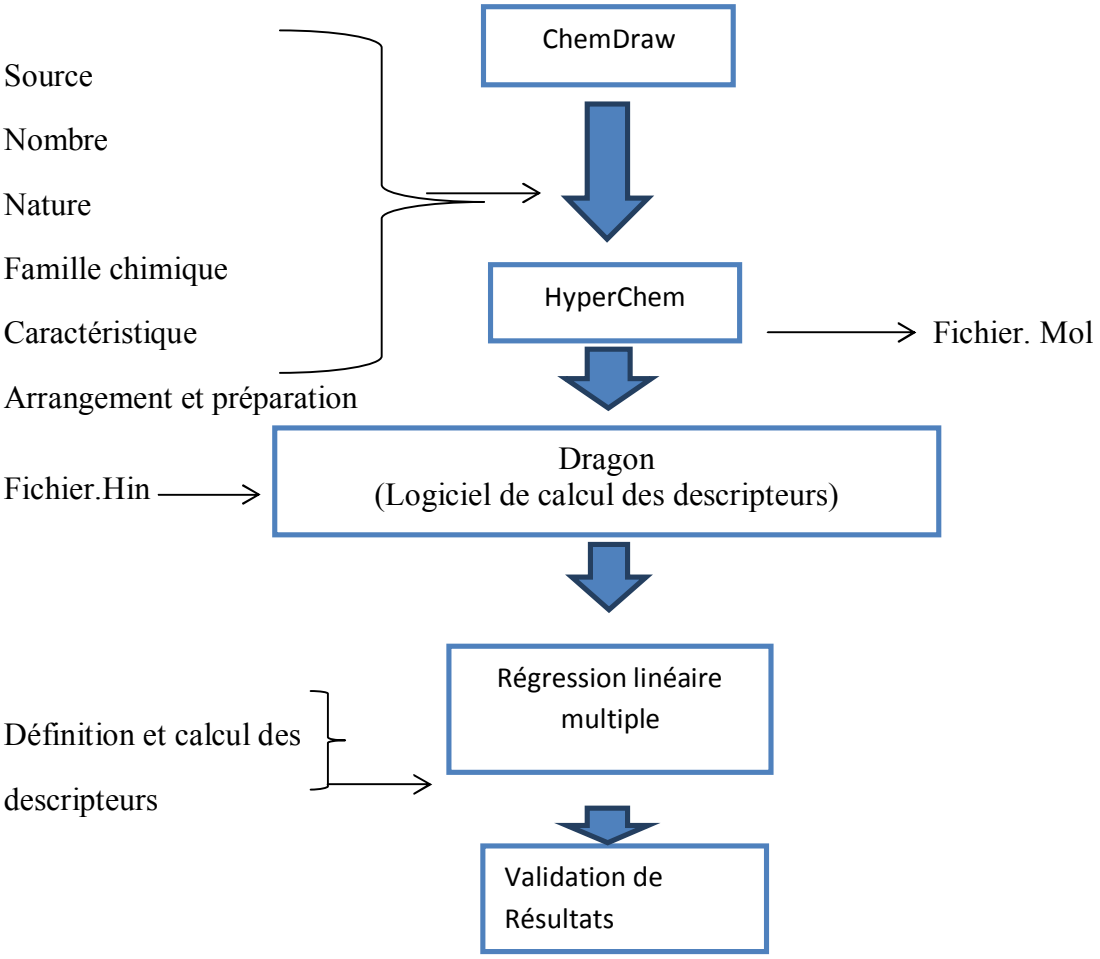


Figure- 07 : Diagramme de notre travail

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

IV-1. Introduction

L'application pratique des gammes des descripteurs moléculaires dans le développement de modèles QSAR/QSPR n'est pas une tâche aisée [22]. Tout d'abord, un très grand nombre (>3000) de descripteurs moléculaires, de différentes complexités et de conceptions diverses ont été imaginés et proposés au cours des (60 dernières) années. Ensuite, pendant ce temps, aucune règle stricte n'a été établie, ni même proposée, pour la sélection de descripteurs adaptés parmi la myriade de descripteurs disponibles. Ce choix a souvent été basé sur l'intuition chimique des chercheurs, ou en se pliant à la tradition.

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

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

En dernier ressort, les modèles QSAR/QSPR peuvent être développés selon des modèles mathématiques différents, généralement en relation avec l'analyse statistique multivariée. Le premier modèle, et le plus largement utilisé, consiste en une équation (multi) linéaire obtenue par régression des données expérimentales en fonction d'un ensemble de descripteurs pré-sélectionnés (ou d'un seul), en utilisant la méthode des moindres carrés ordinaires (MCO). Dans quelques cas, les modèles physiques ou chimiques connus du phénomène étudié laissent prévoir certaines formes mathématiques non linéaires (exponentielles ou logarithmiques) de la dépendance entre les données expérimentales et les descripteurs moléculaires. Les modèles QSAR/QSPR peuvent alors être établis à l'aide de la technique de régression par les moindres carrés non linéaires. D'autres modèles ont été développés en utilisant l'analyse factorielle ou l'analyse en composantes principales. L'intérêt de ces méthodes est qu'elles évacuent le problème de multicolinéarité inhérent aux méthodes de régression linéaires. Cependant, l'interprétation des équations QSAR/QSPR est alors entravée par la nature formelle des facteurs ou des composantes principales. Une alternative aux méthodes très classiques de

régression linéaire multiple (RLM) et d'analyse en composantes principales (ACP) est la technique de régression par les moindres carrés partiels (MCP ou PLS) [23-27].

On a également appliqué les méthodes modernes de l'intelligence artificielle au développement de modèles QSAR/QSPR [28-32]. Ces méthodes comprennent: les réseaux de neurones (RNA), les algorithmes génétiques (GA), et d'autres méthodes globales d'optimisation.

Nous présenterons dans ce qui suit une courte vue d'ensemble des différentes méthodes mathématiques utilisées pour développer nos modèles.

IV-2. Méthodes de régressions linéaire et multilinéaire.

IV-2.1 Aperçu général

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

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

En réalité, il est difficile de se conformer pratiquement aux 5 principes énoncés.
Cependant, la négligence de plusieurs d'entre eux peut conduire à des équations inutiles sans aucun pouvoir prédictif sinon très limité [33].

IV-2.2 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 corréler (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 [34-39].

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.

III-2.3 Régression linéaire multiple

L'exemple développé à partir de deux variables permet de comprendre la logique de la théorie de la régression mais il ne peut être généralisé de la sorte aux régressions multiples. Le système à deux équations à deux inconnus présenté se résolvait facilement comme on l'a vu.
Les équations se compliquent avec plusieurs régresseurs, deux méthodes distinctes permettent de résoudre les équations. La première repose sur la connaissance des coefficients de

corrélation linéaire simple de toutes les paires de variables entre elles, de la moyenne arithmétique et des écarts-types de toutes les variables. La seconde repose sur des calculs matriciels.

Pour n observations, on peut écrire le modèle de régression linéaire multiple sous la forme :

$$y = \beta 0 + \sum_{j=i}^{p} \beta j \, Xij + \xi \qquad \text{pour } i = 1 \ldots\ldots n, \tag{02}$$

où les y_i, x_{i1}, x_{i2}..., x_{ip} : sont des données respectivement relatives aux variables Y, X1..., Xp.

Les estimateurs βj sont calculés en utilisant la méthode des moindres carrés ordinaires. Les variables aléatoires ξi représentent les termes d'erreur non observables du modèle. On peut estimer ces erreurs par les résidus ordinaires ei, différence entre les valeurs observées yi et les valeurs estimées ŷi

Le but de la régression est d'expliquer une variable Y à l'aide d'une variable plusieurs variables (X1, ...,Xq). La variable Y est appelée variable dépendante, ou variable à expliquer et les variables Xj (j=1,...,q) sont appelées variables indépendantes, ou variables explicatives.

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

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

$$\mathbf{E\,(e) = 0} \tag{03}$$

c) Les résidus sont indépendants entre eux :

$$\mathbf{E\,(ei, ej) = 0 \qquad si\ i \neq j} \tag{04}$$

Leurs covariances sont nulles.

d) Les résidus ont tous même variance δ^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 [40] tel que :

$$\mathbf{n - p - 1 \geq 10} \qquad (05)$$

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

Algorithme génétique

Les algorithmes génétiques, métaphores biologiques inspirées des mécanismes de l'évolution darwinienne et de la génétique, et utilisés comme outils d'optimisation ou de recherche combinatoire, nous examinerons leur terminologie de base, leurs principes, les opérateurs participants à l'exploration de l'espace de recherche tout en mettant l'accent sur leur utilisés, leur conception et leur mode de fonctionnement, ces algorithmes sont souvent associés à un processus stochastique, pour éviter qu'un mauvais choix de départ ou des biais systématiques ne les pénalise, aussi , nous exposerons d'une manière générale les méthodes les plus utilisées pour chaque opérateur génétique à savoir sélection, mutation et croisement, nous concluons par la difficulté qui réside dans le choix Des différents paramètres d'un algorithme génétique [41].

IV-3. Paramètres d'évaluation de la qualité de l'ajustement

Deux paramètres sont couramment utilisés :

IV-3.1 Le coefficient de détermination multiple :

- Le coefficient de détermination multiple

$$R^2 = 1 - \frac{SCE}{SCT} = 1 - \frac{\sum_1^n (Y_{i-\hat{Y}_i})^2}{\sum_1^n (Y_i - \overline{y_2})^2} \quad (06)$$

où $\hat{Y}_I$ est la valeur estimée du paramètre physique, et $\overline{y}$ la moyenne des valeurs observé

- La racine de l'erreur quadratique moyenne de prédiction (désignée également par SDEP ; CF infra) :

$$SDEP = \sqrt{\frac{1}{N}\sum_1^N (Y_I - \widehat{\boldsymbol{Y}_{(\iota)}})^2} \quad (07)$$

IV-3.2 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 prédicteurs restants (présence de multi colinéarité parmi les prédicteurs). 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 prédicteurs (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 prédicteurs sert souvent d'indicateur de multi colinéarité importance, Si le FIV > 5-10 la qualité de l'estimation des coefficients de régression est faible [42].

IV-3.3 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 obtenu (considéré comme vecteur expérimental réel) un modèle QSPR, selon la méthode habituelle. Ce procédé est répété est répété plusieurs fois (100 dans notre cas)[43].

IV-4. 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 Square d'Error), calculée sur différents ensembles :

- Ensemble d'estimation (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.

$$SDEC = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \widehat{y})^2}{n}} \quad (08)$$

$$SDEP_{ext} = \sqrt{\frac{\sum_{i=1}^{next}(y_i - \widehat{y})^2}{next}} \quad (09)$$

La validation croisée par « leave – one - out » (LOO) [35] 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 ŷ(i). On répète le procédé pour chacune des grandeurs d'intérêt. La somme des carrés des erreurs de prédiction, désignée par le symbole PRESS, est une mesure de la dispersion des estimations. On l'utilise pour définir le coefficient de prédiction :

$$Q^2_{LOO} = \frac{SCT - PRESS}{SCT} \quad (10)$$

Contrairement à R^2 au coefficient qui augmente avec le nombre de paramètres du modèle, le facteur Q^2_{LOO} 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^2_{LOO} .Une valeur $Q^2_{LOO} > 0.5$ est considérée comme satisfaisante, et une valeur supérieure à 0,9 est excellente [36]. Si de petites valeurs de Q^2_{LOO} 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 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^2ext, est calculée comme suit [44]:

$$Q^2_{ext} = \frac{\sum_{I=1}^{n\,ext}(y_i - \widehat{y_i})^2 / next}{\sum_{i=1}^{n}(y_i - \widehat{y})^2 / n} \qquad (11)$$

Partie application

1-Sélection des descripteurs

Plusieurs classes de descripteurs moléculaires ont été calculées pour la molécule entière : Descripteurs relatifs à la construction moléculaire, et descripteurs topologiques. Nous rapporterons quelques résultats obtenus pour toutes les molécules (calibration et validation).
Le modèle a été calculé par algorithme génétique dont le logiciel mobydigs. Le niveau de signification a été fixé à 0.05 tant pour l'introduction que pour l'expulsion des variables.

Nous traiterons dans ce travail la propriété physicochimique envisagée (Tb), en considérant les mêmes ensembles d'estimation et de validation. Rappelant que les 65 composés ont été éclatés aléatoirement en deux sous ensemble comportant 49 pour la calibration et 16 pour la validation.

Egalement, les modèles ont été calculés pour les différentes dimensions par algorithme génétique en utilisant le logiciel mobydigs.

Dans les études QSPR, cinq composés, à la limite, doivent être associées à chaque variable explicative

Le nombre de degrés de liberté final (eq :05) doit être au moins égal à 10 cette condition est bien vérifiée pour les différents modèles étant donner que le nombre de composés est égal à 49 ce que nous permettent d'aller jusqu'au 10 variables.

Une indépendance globale acceptable des descripteurs sera vérifiée lorsque les facteurs d'inflation de la variance (FIV) calculés pour chacun d'eux sont inférieurs à 5.

Parmi les modèles optimaux générés, celui qui fournit la valeur maxima le pour les paramètres statistiques Q^2, R^2 et Q^2ext tout en vérifiant la condition : FIV < 5 comporte les 5 descripteurs calculés par le logiciel DRAGON, dont les symboles, les classes et les significations sont réunis dans le tableau 3

2-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 49 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.

3- Calcul de l'équation de régression

Après avoir éliminé les descripteurs qui n'obéissent pas à la condition précédente, nous avons utilisés la méthode des algorithmes génétiques pour le choix du modèle.

Le logiciel Mobydigs nous permet de calculer 100 modèles de différentes dimensions avec 10 modèles de chaque taille.

En traçant le graphe des R^2 en fonction de nombre de descripteurs ; on voit que le bon modèle qui possède le moins de descripteurs est celui à 5, parce qu'après ce dernier les valeurs des R^2 augmentent d'un pas faible.

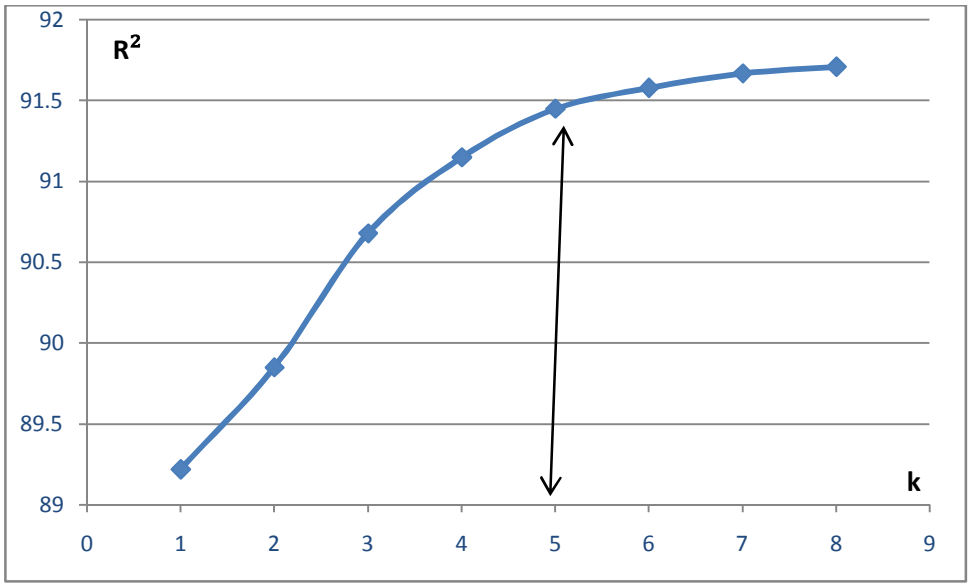


Figure- 08 Variation de R^2 en fonction de nombre de descripteurs k

4- Choix du modèle :

Parmi les modèles à 5 descripteurs on doit choisir celui regroupe des descripteurs non corrélés entre eux en calculant la matrice de corrélation.

5-Matrice de corrélation

Tableau- 03 *Corrélations Tb avec les 5 descripteurs*

	Tb	ATS1m	E2s	R4p+	HATS2v
ATS1m	0.945 0.000				
E2s	0.549 0.000	0.552 0.000			
R4p+	0.482 0.000	0.565 0.000	0.693 0.000		
HATS2v	0.148 0.304	0.108 0.456	0.386 0.006	0.516 0.000	
Mor04p	-0.034 0.815	-0.040 0.782	-0.113 0.433	0.251 0.079	0.158 0.274

Cette matrice nous permette de voir que les 3 premiers descripteurs sont bien corrélés avec la propriété, par contre les deux autres sont moins corrélés mais ils portent une information complémentaire pour le modèle portant ils sont corrélés avec les autres avec un p> 0.05

Les classes, les définitions et les valeurs des descripteurs entrant dans le modèle obtenu sont reniées dans les tableaux (04) et (05).

Tableau- 04 Classes et significations des descripteurs

Descripteur	Classe	Signification
ATS1m	2D autocorrelation indices	The **Broto-Moreau autocorrelation** ATS*kw*, *w* being the atomic property used to weight the molecular graph and *k* the lag, is evaluated by considering separately all the contributions of each different path length (*lag*) in the molecular graph, as collected in the topological distance matrix. In other words, the total spatial autocorrelation at *lagk* ATS*kw* is obtained by summing all the products *w*i×*w*j of all the pairs of atoms *i* and *j*
E2s	WHIM descriptors	WHIM descriptors (*Weighted Holistic Invariant Molecular descriptors*) are geometrical descriptors based on statistical indices calculated on the projections of the atoms along principal axes
R4p+ HATS2v	GETAWAY descriptors	The GETAWAY (GEometry, Topology, and Atom-Weights AssemblY) descriptors have recently been proposed as chemical structure descriptors derived from a new representation of molecular structure, the Molecular Influence Matrix (MIM), denoted by H and defined as the following
Mor04p	3D-MoRSE descriptors	3D-MoRSE (3D-Molecule Representation of Structures based on Electron diffraction) descriptors are based on the idea of obtaining information from the 3D atomic coordinates by the transform used in electron diffraction studies for preparing theoretical scattering curves

Tableau- 05 Valeurs des descripteurs moléculaires sélectionnés

N	Composé	ATS1m	E2s	R4p+	HATS2v	Mor04p
1	Dimethylether	1.299	0.361	0.027	0.19	-0.217
2	Diethylether	1.734	0.042	0.023	0.11	-0.31
3	Methyl t-butylether	1.897	0.134	0.021	0.1	-0.043
4	Di-n-propylether	2.037	0.039	0.013	0.096	-0.059
5	Di-isopropylether	2.037	0.135	0.028	0.096	0.051
6	2-Methylfuran	2.037	0.16	0.028	0.145	-0.741
7	Tetrahydrofuran	1.897	0.131	0.014	0.184	-0.485
8	Tetrahydropyran	2.037	0.138	0.018	0.17	-0.293
9	1,4-Dioxane	2.12	0.104	0.018	0.189	-0.328
10	Epichlorohydrin	2.154	0.353	0.039	0.287	-0.295
11	Chloromethylmethylether	1.89	0.084	0.052	0.312	-0.173
12	Bis(chloromethyl)ether	2.258	1.087	0.16	0.404	-0.171
13	Bis(2-chloroisopropyl)ether	2.608	0.194	0.078	0.164	0.635
14	2-Chloroethyl vinylether	2.154	0.015	0.026	0.16	-0.199
15	Bis(2-chloroethoxy)methane	2.656	0.025	0.015	0.142	-0.187
16	Anisole	2.268	0.11	0.023	0.139	-0.357
17	2-Chloroanisole	2.535	0.466	0.051	0.153	-0.062
18	3-Chloroanisole	2.535	0.186	0.026	0.165	-0.114
19	4-Chloroanisole	2.535	0.081	0.023	0.164	-0.091
20	2,4,6-Trichloroanisole	2.919	0.513	0.073	0.177	0.098
21	2,3,4,5-Tetrachloroanisole	3.067	0.702	0.078	0.203	0.323
22	3,4,5-Trichloroveratrole	3.053	0.248	0.048	0.131	0.122
23	Tetrachloroveratrole	3.184	0.379	0.052	0.565	0.378
24	Phenetole	2.367	0.102	0.018	0.123	-0.248
25	Benzylethylether	2.457	0.092	0.018	0.117	-0.332
26	Styreneoxide	2.457	0.098	0.026	0.163	-0.654
27	Diphenylether	2.751	0.117	0.014	0.082	-0.78
28	2-Chlorodiphenyl ether	2.924	0.271	0.033	0.092	-0.548
29	4-Chlorodiphenyl ether	2.924	0.103	0.014	0.097	-0.521
30	2,4-Dichlorodiphenyl ether	3.071	0.286	0.05	0.107	-0.362
31	2,4,5-Trichlorodiphenyl ether	3.199	0.395	0.049	0.118	-0.337
32	2,4,4',5-Tetrachlorodiphenyl ether	3.313	0.396	0.048	0.131	0.018
33	2,2',4,4',5-Pentachlorodiphenyl ether	3.415	0.501	0.047	0.137	0.168
34	3,3',4,4',5-Pentachlorodiphenyl ether	3.415	0.372	0.055	0.15	0.289
35	2,2',3,4,4',6'-Hexachlorodiphenyl ether	3.508	0.456	0.043	0.14	0.14
36	2,2',4,4',5,5'-Hexachlorodiphenyl ether	4.131	0.427	0.103	0.199	0.199
37	2,2',4,4',5,6'-Hexachlorodiphenyl ether	3.508	0.526	0.044	0.139	0.139
38	2,3',4,4',5,5'-Hexachlorodiphenyl ether	3.508	0.625	0.046	0.146	0.146
39	2,2',3,4,4',6,6'-Heptachlorodiphenyl ether	3.593	0.53	0.04	0.141	0.141
40	2,2',3,3',4,4',5,6'Octachlorodi-phenylether	3.671	0.619	0.051	0.16	0.16
41	2,2',3,3',4,4',5,5',6-Nonachlorodiphenyl ether	3.743	0.694	0.051	0.168	0.168
42	Decachlorodiphenylether	3.811	0.724	0.05	0.169	0.169

Tableau- 05 suite et fin

N	MolID	ATS1m	E2s	R4p+	HATS2v	Mor04p
43	4,4'-Dibromodiphenyl Ether	3.366	0.11	0.02	0.131	0.131
44	2,4,4'-Tribromodiphenyl ether	3.573	0.23	0.098	0.145	0.145
45	2,2',4,4'-Tetrabromodiphenyl ether	3.744	0.338	0.094	0.156	0.156
46	2,2',4,4',6-Pentabromodiphenyl ether	3.89	0.366	0.087	0.158	0.158
47	2,2',4,4',5,5'-Hexabromodiphenyl ether	4.018	0.491	0.089	0.185	0.185
48	,2',4,4',5,6'-Hexabromodiphenyl ether	4.018	0.43	0.085	0.173	0.173
49	Decabromodiphenylether	4.409	0.552	0.137	0.414	0.414
50	Butylethylether	2.037	0.041	0.013	0.089	0.089
51	Di-n-butylether	2.268	0.041	0.01	0.078	0.078
52	1,2 propyleneoxide	1.734	0.121	0.02	0.211	0.211
53	Furan	1.897	0.084	0.013	0.122	0.122
54	Bis(2-chloroethyl)ether	2.448	0.033	0.022	0.178	-0.049
55	2,3-Dichloroanisole	2.745	0.33	0.049	0.173	0.173
56	2,6-Dichloroanisole	2.745	0.17	0.064	0.157	0.157
57	2,3,4-Trichloroanisole	2.919	0.527	0.074	0.193	0.193
58	2,3,5,6-Tetrachloroanisole	3.067	0.334	0.079	0.188	0.188
59	Veratrole	2.512	0.241	0.024	0.096	0.096
60	4,5-Dichloroveratrole	2.903	0.161	0.022	0.127	0.127
61	2,2',4,4'-Tetrachlorodiphenyl ether	3.313	0.407	0.048	0.128	0.128
62	3,3',4,4'-Tetrachlorodiphenyl ether	3.313	0.256	0.02	0.14	0.14
63	4-Bromodiphenyl ether	3.105	0.107	0.02	0.107	0.107
64	2,2',4,4',5-Pentabromodiphenyl ether	3.89	0.412	0.091	0.172	0.172
65	2,3,3',4,4',5'-Hexabromodiphenyl ether	4.018	0.394	0.104	0.197	0.197

- De 1 à 49 les composés de calibration.
- De 50 à 65 les composés de validation externe.

6-Equation de régression

$$Tb = -46.3 + 198\ ATS1m + 87.6\ E2s - 1146\ R4p+ + 180\ HATS2v + 19.2\ Mor04p \quad (12)$$

Tableau- 06 : Analyse de régressions ; Tb avec les 5 descripteurs

Predictor	Coef	SE Coef	T	P	VIF
Constant	-46.27	31.35	-1.48	0.147	
ATS1m	198.39	11.50	17.25	0.000	1.747
E2s	87.56	44.02	1.99	0.053	2.448
R4p+	-1145.6	359.6	-3.19	0.003	3.218
HATS2v	180.50	90.94	1.98	0.053	1.486
Mor04p	19.16	14.17	1.35	0.183	1.308

7- Analyse de régression

Les valeurs des paramètres statistiques montrent que les cinq descripteurs **ATS1m; E2s; R4p+; HATS2v; Mor04p** (**tableau-06**) permettent de corréler la température d'ébullition des 49 éthers (R^2> 99%), confirment par les valeurs de VIF < 5.

Tableau- 07 Valeurs des Tb expérimentales, calculées, ecal, eistd et hii pour l'ensemble de calibration

N	Object	Y Exp.	Y-Calc	Err.Calc.	Hat	Eistd
1	Dimethylether	248.31	242.18	-6.13	0.17	-0.18
2	Diethylether	307.59	288.89	-18.7	0.091	-0.47
3	Methyl t-butylether	328	334.88	6.88	0.073	0.17
4	Di-n-propylether	363.08	362.47	-0.61	0.066	-0.01
5	Di-isopropylether	359.4	355.8	-3.6	0.071	-0.09
6	2-Methylfuran	337.7	351.7	14	0.077	0.34
7	Tetrahydrofuran	338	349.36	11.36	0.07	0.27
8	Tetrahydropyran	361	374.31	13.31	0.052	0.31
9	1,4-Dioxane	374.5	390.57	16.07	0.057	0.38
10	Epichlorohydrin	391	413.41	22.41	0.095	0.56
11	Chloromethylmethylether	332.5	329.43	-3.07	0.166	-0.09
12	Bis(chloromethyl)ether	379	383.26	4.26	0.615	0.38
13	Bis(2-chloroisopropyl)ether	460	440.45	-19.55	0.128	-0.52
14	2-Chloroethyl vinylether	381	377.58	-3.42	0.064	-0.08
15	Bis(2-chloroethoxy)methane	488	487.64	-0.36	0.064	-0.01
16	Anisole	426.7	405.14	-21.56	0.043	-0.5
17	2-Chloroanisole	471.5	465.4	-6.1	0.065	-0.15
18	3-Chloroanisole	466.5	470.68	4.18	0.035	0.1
19	4-Chloroanisole	470.2	465.18	-5.02	0.048	-0.12
20	2,4,6-Trichloroanisole	514	528.79	13.9	0.06	0.33
21	2,3,4,5-Tetrachloroanisole	622.8	569.71	-45.7	0.139	-1.23
22	3,4,5-Trichloroveratrole	563	551.61	-10.94	0.039	-0.25
32	Tetrachloroveratrole	659.5	686.5	8.77	0.675	1.02
24	Phenetole	442.81	428.4	-13.81	0.042	-0.32
25	Benzylethylether	458.6	442.6	-15.31	0.043	-0.35
26	Styreneoxide	467.1	434.8	-30.29	0.062	-0.72
27	Diphenylether	531	490.73	-37.49	0.069	-0.9
28	2-Chlorodiphenyl ether	540	525.19	-14.2	0.042	-0.33
29	4-Chlorodiphenyl ether	557.5	532.74	-23.23	0.062	-0.55
30	2,4-Dichlorodiphenyl ether	273	554.4	270.08	0.04	6.19
31	2,4,5-Trichlorodiphenyl ether	598	581.02	-16.37	0.036	-0.37
32	2,4,4',5-Tetrachlorodiphenyl ether	627	613.98	-12.38	0.05	-0.29
33	2,2',4,4',5-Pentachlorodiphenyl ether	638	650.39	11.16	0.1	0.28

Tableau-07 suite et fin

N	Object	Y Exp.	Y-Calc	Err.Calc.	Hat	Eistd
34	3,3',4,4',5-Pentachlorodiphenyl ether	675	630.56	-41.64	0.063	-0.99
35	2,2',3,4,4',6'-Hexachlorodiphenyl ether	652	652.67	0.63	0.059	0.01
36	2,2',4,4',5,5'-Hexachlorodiphenyl ether	663	732.52	59.26	0.148	1.62
37	2,2',4,4',5,6'-Hexachlorodiphenyl ether	649	652.6	3.3	0.084	0.08
38	2,3',4,4',5,5'-Hexachlorodiphenyl ether	682	681.22	-0.65	0.17	-0.02
39	2,2',3,4,4',6,6'-Heptachlorodiphenyl ether	660	653.41	-4.91	0.255	-0.16
40	2,2',3,3',4,4',5,6'-'Octachlorodi-phenyl ether	701	697.27	-3.38	0.095	-0.08
41	2,2',3,3',4,4',5,5',6-Nonachlorodiphenyl ether	716	714.32	-1.47	0.128	-0.04
42	Decachlorodiphenylether	731	707.41	-17.87	0.242	-0.58
43	4,4'-Dibromodiphenyl éther	612	636.01	21.42	0.108	0.55
44	2,4,4'-Tribromodiphenyl éther	643	596.98	-37.67	0.181	-1.1
45	2,2',4,4'-Tetrabromodiphenyl ether	669	659.69	-7.89	0.153	-0.22
46	2,2',4,4',6-Pentabromodiphenyl ether	707	666.57	-34.36	0.15	-0.95
47	2,2',4,4',5,5'-Hexabromodiphenyl ether	726	751.79	20.27	0.214	0.63
48	2,2',4,4',5,6'-Hexabromodiphenyl ether	744	707.84	-32.34	0.106	-0.82
49	Decabromodiphenylether	803	786.04	-11.3	0.334	-0.45

L'analyse des résidus fait ressortir, des erreurs absolues moyennes égales à 20.46

La colonne 06 (tableau 07) donne les valeurs de hii, ième terme diagonal de la matrice de projection : H=X(X'X)-1 X' ou : X est la matrice des valeurs observées des variables explicatives et X' sa transposée, ces valeurs sont utiles pour le calcul des résidus caractéristiques. La valeur critique pour déterminer les points leviers correspond à $h^* = 0.36$

On constate que tous les hii sont inférieures à cette valeur critique 0.36, à l'exception de ceux des composés 1 et 65.

Toutes les valeurs des paramètres statistiques sont regroupées dans le **tableau(08)**

Tableau- 08 Valeurs des paramètres statistiques pour l'ensemble de calibration

n	49
R^2	91.29
Q^2	90.1
F	90.11
S	46.369
SDEC	43.4374
SDEP	46.3012

Pour la robustesse du modèle est assurés par la valeur de LOO Q^2> 99% alors que les valeurs de l'erreur quadratique moyenne de prédiction et de calcul sont petites.

8-Diagramme de Williams :

On a représenté, sur la même figure 09; pour les deux ensembles (calibration et validation); Le domaine d'application a été discuté à l'aide de Williams qui représente les résidus de prédiction standardisés en fonction des valeurs des leviers hii

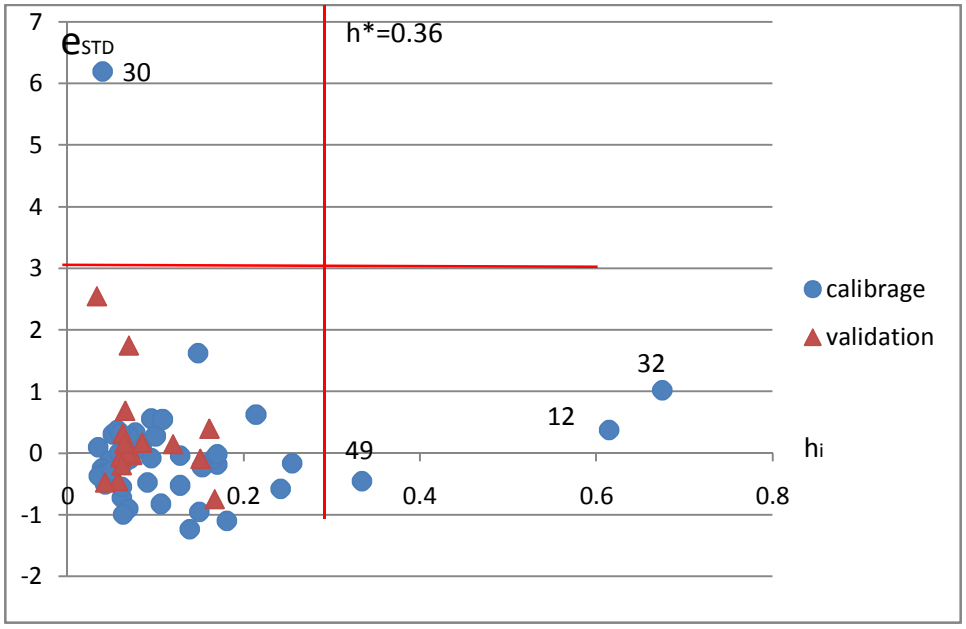
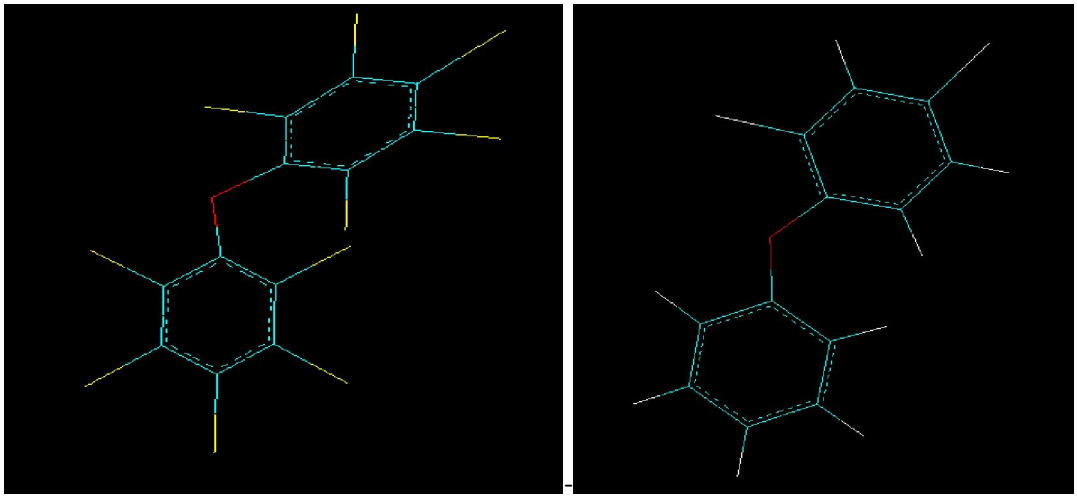


Figure- 09 Diagramme de Williams

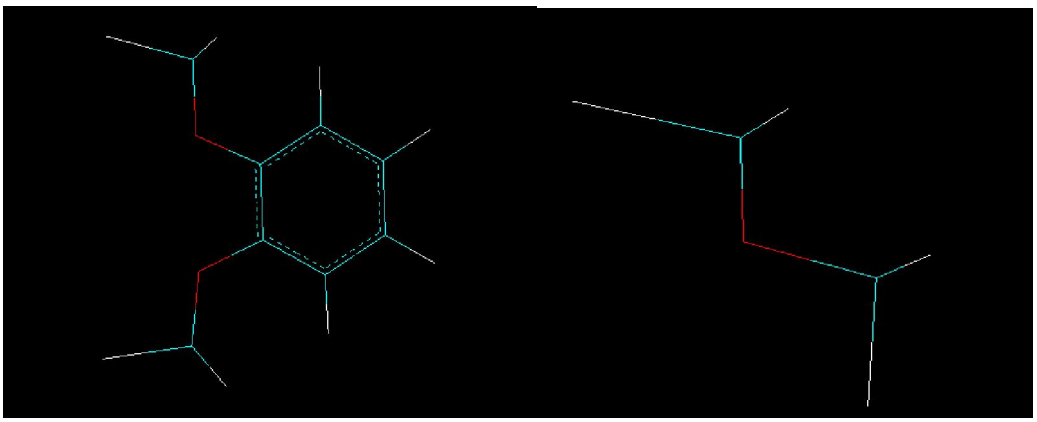
Le domaine d'application représenté par le diagramme de Williams fait ressortir trois points influents qui ont des valeurs supérieures à 0.36, il s'agit des composés (bis(chloromethyl)ether), (Tetrachloroveratrole), et (2,4-Dichlorodiphenyl ether)

En plus d'une observation aberrante c'est le (Decabromodiphenylether) avec une erreur standardisée supérieure à +3.

Les quatre points cités précédemment appartiennent à l'ensemble de calibration.



Decabromodiphenylether 2,4-Dichlorodiphenyl ether



Tetrachloroveratrole Bis (chloromethyl)ether

Figure- 09 Les 4 composés influents

9-Vérification de la qualité de l'ajustement :

La qualité de l'ajustement a été vérifiée en représentant les valeurs calculées de Tb en fonction des celles observées ou expérimentales (colonne 3 et 4 ; Tableau 07). La figue 08 montre un bon ajustement traduit par une faible dispersion autour de la droite de régression

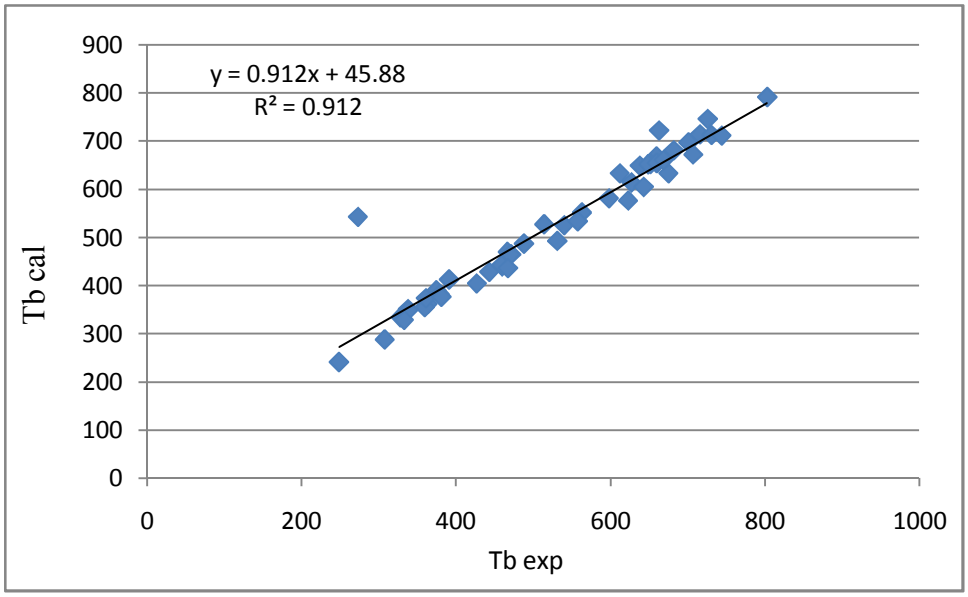


Figure- 10 Graphe des valeurs calculées en fonction des valeurs observées pour l'ensemble de calibration

10-Test de randomisation

Dans le but d'établir que le modèle obtenu n'est pas dû au hasard ou à une sur spécification, nous avons appliqué le test de randomisation. Ainsi 100 nouveaux vecteurs de l'indice de rétention ont été générés par permutation des positions des composantes du vecteur réel.

La figure 16 qui représente le graphe des coefficients statistiquesQ_2 et R_2 permet de comparer les résultats obtenus pour les modèles randomisés (carrés bleus) au modèle réel de départ (carré rouge brique).

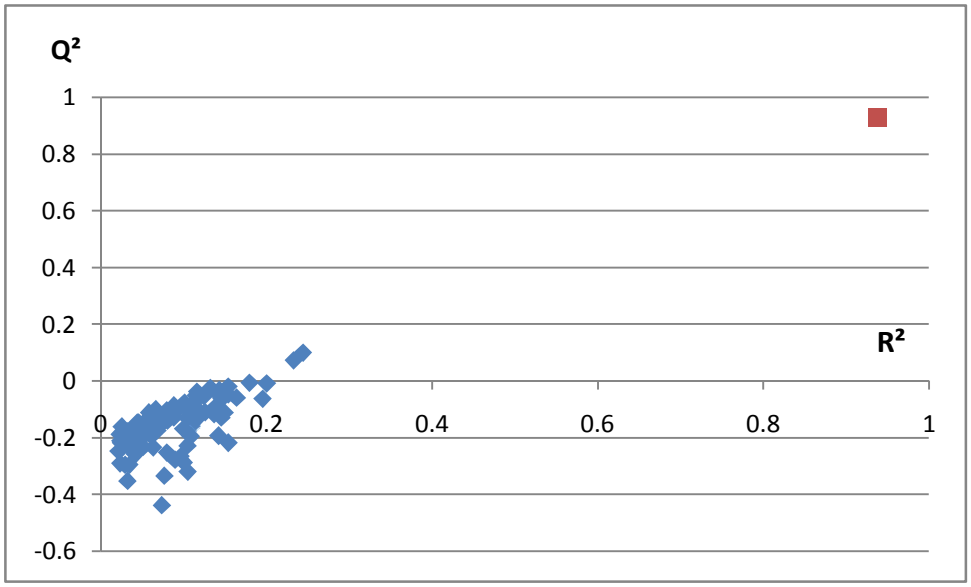


Figure- 12 Test de randomisation associé au modèle QSPR

Les carrés bleus représentent l'indice de rétention ordonnée de façon aléatoire, etle carré rouge brique correspond au modèle réel.

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

11 -Validation externe

Pour généraliser le modèle choisi on procède à une validation externe sur les 16 composés choisis aléatoirement et qui ne font pas partie de l'ensemble d'essai. Les résultats obtenus (tableau 09) montrent que les valeurs prédites sont très proches des valeurs observées, ce qui confirme que le modèle choisi décrit de façon excellente la relation température d'ébullition observée.

Tableau- 09 Valeurs des Tb expérimentales, prédites, ei, eistd et hii pour l'ensemble de validation

N	Object	Y Exp.	Y-Pred	Err.Pred.	Hat	Std.Err.Pred.
50	Butylethylether	365.3	363.92	-1.38	0.074	-0.03
51	Di-n-butylether	413.28	419.43	6.15	0.12	0.14
52	1,2 propyleneoxide	308	315.13	7.13	0.085	0.16
53	Furan	304.5	334.96	30.46	0.066	0.68
54	Bis(2-chloroethyl)ether	451.5	448.2	-3.3	0.061	-0.07
55	2,3-Dichloroanisole	389	504.67	115.67	0.034	2.54
56	2,6-Dichloroanisole	390	467.97	77.97	0.07	1.74
57	2,3,4-Trichloroanisole	534	533.84	-0.16	0.069	0
58	2,3,5,6-Tetrachloroanisole	559.9	539.14	-20.76	0.058	-0.46
59	Veratrole	479	457.01	-21.99	0.043	-0.48
60	4,5-Dichloroveratrole	530	543.83	13.83	0.063	0.31
61	2,2',4,4'-Tetrachlorodiphenyl ether	611	617.89	6.89	0.065	0.15
62	3,3',4,4'-Tetrachlorodiphenyl ether	648	643.57	-4.43	0.151	-0.1
63	4-Bromodiphenyl ether	578	569.02	-8.98	0.062	-0.2
64	2,2',4,4',5-Pentabromodiphenyl ether	689	705.68	16.68	0.161	0.39
65	2,3,3',4,4',5'-Hexabromodiphenyl ether	749	717.12	-31.88	0.167	-0.75

La qualité d'ajustement illustrée par la représentation graphique des valeurs prédites en fonction de celles observées (figure 13) conduit également à un bon ajustement avec un R^2=92.5%

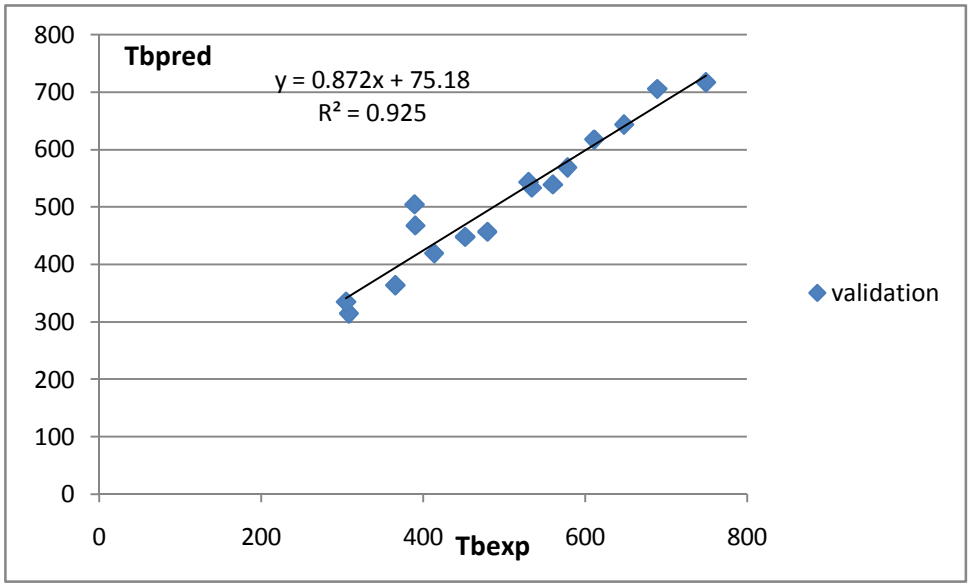


Figure- 13 Graphe des valeurs prédites en fonction des valeurs observées pour l'ensemble validation

Toutes les valeurs des paramètres statistiques de validation sont regroupées dans le tableau ci-dessous :

Tableau 10 : Valeurs des Q^2_{ext} et $SDEP_{ext}$

n	16
Q^2**ext**	90.1
SDEP$_{ext}$	46.3012

Conclusion Générale

Nous avons appliqué la méthodologie QSPR pour relier la propriété (température d'ébullition) d'une série d'éthers; à des descripteurs moléculaires théoriques reflétant certaines particularités, à chaînes ouvertes ou portant des cycles aromatiques dont plusieurs isomères.

Un modèle QSPR comportant 5 descripteurs a été établi en utilisant l'analyse de régression multilinéaire et la sélection des variables explicatives a été réalisée par algorithme génétique, dans la version MOBYDIGS de TODESCHINI, en maximisant Q^2_{L00}.

Les 65 données de base ont été éclatées aléatoirement en deux ensembles disjoints, invariants :

- un ensemble principal de 49 composés utilisés pour le calcul et, éventuellement, les essais du modèle ;
- un ensemble de 16 composés pour la prédiction externe.

Les statistiques (R^2 ;Q^2 ; SDEC ; SDEP) calculées établissent la pertinence du modèle développé, malgré la présence des corrélations entre les descripteurs avec des valeurs de $p>0.05$.

L'analyse des résidus a permis de détecter 4 observations aberrantes de l'ensemble de calibration.

A chaque fois, la qualité de l'ajustement a été vérifiée en procédant à une validation croisée par "leave – one - out". Les grandes valeurs de Q^2 obtenues (> 90 %) reproduisent pratiquement celles du coefficient de détermination multiple correspondant, ce qui fait ressortir la qualité de l'ajustement pour notre modèle obtenu.

Le test de randomisation montre, que seul le vecteur réel des observations conduit à des valeurs élevées des statistiques R^2 et Q^2, ce qui prouve que les modèles obtenus ne sont pas aléatoires.

Les valeurs RMSE sont faibles et 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 de chaque modèle.

Enfin, d'autres méthodes (réseaux de neurones, 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 testées.

Références Bibliographiques

[1]. A.R. Katritzky., D.C. Fara., R.O. Petrukhin, Top. Med. Chemm.2002,1333-1356.

[2]. O C. Hansen. Quantitative structure-activity relationships (QSAR) and pesticides: Ministry of the Environment, Environmental Protection Agency; 2004.

[3]. AC. Brown, Fraser TR. V.−On the connection between chemical constitution and physiological action. Part. I.−On the physiological action of the salts of the ammonium bases, derived from strychnia, brucia, thebaia, codeia, morphia, and nicotia. Earth and Environmental Science Transactions of The Royal Society of Edinburgh. 1868;25(1):151-203.

[4]. Free SM, Wilson JW. A mathematical contribution to structure-activity studies. Journal of Medicinal Chemistry. 1964;7(4):395-9.

[5]. R. Leardi , Chemometr. 2001,15, 559-569.

[6]. M.Karelson. Molecular descriptors in QSAR/QSPR. Wiley- Inter science. 2000, 385

[7]. D. Mackay W.Y. Shiu M. Kuo-Ching S. Chi Lee Physical-Chemical Properties and Environmental Fate for Organic Chemicals in 2006

[8]. ChemDraw Release 7.0.1 Chemical Structure Drawing Standard, 2002.

[9]. HyperchemTM Release 7.5 for windows, Molecular Modelling system, 2000.

[10]. R. Todeschini, V. Consonni, M. Pawan, DRAGON, Software for the calculation of Molecular Descriptors. Release 5.3 for Windows, Milano, 2005.

[11]. R. Sirid. R. Wahiba Etude. QSRR de la rétention chromatographique des HAP. 2018, 19.

[12]. T. Thomas-Danguin, Intensité olfactive des composés purs et demélanges: application au masquage des odeurs, Université Claude Bernard, Lyon. 1997, 224.

[13]. R. Sirid. R. Wahiba Etude. QSRR de la rétention chromatographique des HAP. 2018, 20.

[14]. H. Bachir A. Chokri. Modélisation de la température d'ébullition des alcanes en utilisant une approche QSPR. 2016, 9-10.

[15]. H. Bachir A. Chokri modélisation de la température d'ébullition des alcanes en utilisant une approche QSPR. 2016, 10.

[16]. Dragon_ Aide blocs des descripteurs.

[17]. P. Le Cloirec,. Introduction au traitement de l'air, Les techniques de l'ingénieur traité environnement (G 1700), 2002, 1-8.

[18**].** B. Mahrez, Estimation statistique des températures de fusion pour quelques hydrocarbures aliphatiques. 2016, 30.