



MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITÉ «Abbès LAGHROUR» DE KHENCHELA
FACULTÉ DES SCIENCES ET DE LA TECHNOLOGIE
Département Sciences de la matière

N° de série :

Mémoire de fin d'études

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

Spécialité : Chimie analytique

Option : chimie analytique et environnement

MODÉLISATION DE LA TEMPÉRATURE DE FUSION D'UNE SÉRIE D'ÉTHERS

Réalisé par :

- BENDJEDDOU Rayen
- BENDJEDDOU Kawther

Dirigé par : KERTIOU Noureddine

Membres de jury :

SAMAAI Salima	MCB	Présidente
BOUAAKADIA Amel	MCA	Examinatrice

Année Universitaire 2020-2021



Remerciements




*Tout d'abord, nous tenons à remercier **Dieu**, de nous avoir donné la santé, la volonté et la patience pour mener à terme notre formation et pouvoir réaliser ce travail.*

*Nous tenons à remercier notre encadreur **Dr KERTIOU Noureddine** pour son encadrement, son aide, ses précieux conseils avisés dans ce travail.*

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

Enfin nous tenons à remercier tous nos professeurs qui nous en enseignés durant notre cursus.



DEDICACES



Je dédie ce travail à :

A ma très chère maman

A l'âme de mon père

A l'âme de mon frère « Hazem »

A mes sœurs et mes frères

A mes meilleurs Amis

Moufida et Djaafer

Kawther

DEDICACES



Je dédie ce travail à :

A mon très cher papa

A ma très chère maman

A mes frères

A mon époux Hichem

A mes meilleurs Amis

Rayen

SOMMAIRE

Dédicaces et remerciements

Liste des tableaux.....A

Liste des figures.....B

Symboles et abréviations.....C

INTRODUCTION GÉNÉRALE.....2

PARTIE THEORIQUE

I-GENERALITE5

I-1-Introduction 5

I-2- Les relations structure-propriétés quantitatives «QSPR» 5

I-2-1-Définition de la méthode QSPR..... 5

I-2-2-Principe de la méthode QSPR 6

I-2-3-Les méthodes mathématiques utilisés par le model QSPR 7

I-2-4-La relation structure- propriété quantitative (QSPR).....8

II- COLLECTE DES DONNEES.....8

III- OPTIMISATION DE LA GEOMETRIE MOLECULAIRE ET GENERATION

DESDESCRIPTEURS MOLECULAIRES 11

III-1-Préparation de base des données 11

III-1-1-Logiciels « ChemDraw» 11

III-1-2-Récupération et stabilisation les molécules de fichier Hin..... 13

III-2- Calcul des descripteurs moléculaires 15

III-2-1- Le Logiciel « Dragon » 15

III-2-2-Descripteurs moléculaires..... 16

III-3-Importance des descripteurs 18

III-4- Diagramme montre les étapes de prédiction..... 18

IV-METHODES UTILISEES POUR LE DEVELOPPEMENT DE MODELES

<i>QSAR/QSPR</i>	22
IV-1- Introduction	22
IV-2- Méthodes de régressions linéaires et multilinéaires	23
IV-2-1 Aperçu général	23
IV-2-2- Evaluation préliminaire des données	24
IV-2-3- Régression linéaire multiple	25
IV-3- Méthodes de sélection des descripteurs	26
IV-4- Paramètres d'évaluation de la qualité de l'ajustement	26
IV-5- Facteur d'inflation de la variance [FIV]	27
IV -6- Test de randomisation	28
IV -7- Validation externe	28

PARTIE APPLICATION

V-1-Conditions expérimentales	31
V-2-Choix de la taille du modèle	31
V-3-Sélection des descripteurs	32
V-4-Calcul des corrélations entre les différents descripteurs	36
V-5-Equations de régressions	37
V-6-Analyse de régression	37
V-7-Vérification de la qualité de l'ajustement	42
V-8-Test de randomisation	43
V-9-Validation externe	44
CONCLUSION GENERALE	47
REFERENCES BIBLIOGRAPHIQUES	49
ANNEXE	52

LISTE DES TABLEAUX

TABLEAU	TITRE	PAGE
1	Nomenclature et valeurs de la propriété étudiée	8
2	Quelques blocks des descripteurs calculés par logiciel dragon	17
3	Valeurs des descripteurs moléculaires sélectionnés	32
4	Classes et significations des descripteurs	35
5	Corrélation entre Tf et les descripteurs	36
6	Paramètres de régression	37
7	Valeurs des Tf expérimentales, calculées, e_i , $e_{i\text{std}}$ et h_{ii} pour l'ensemble de calibration	38
8	Valeurs des paramètres statistiques pour l'ensemble de calibration	40
9	Valeurs des Tf expérimentales, calculées, e_i , $e_{i\text{std}}$ et h_{ii} pour l'ensemble de validation	44
10	Valeurs des Q^2_{ext} et $SDEP_{\text{ext}}$	45

LISTE DES FIGURES

FIGURE	TITRE	PAGE
1	Principe de la méthode QSPR	7
2	Représentation des molécules par le ChemDraw	12
3	Logiciel Hyperchem	14
4	Logiciel « Dragon »	15
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 du nombre de descripteurs k	31
9	Diagramme de Williams	41
10	Les 02 composés influents	42
11	Graphe des valeurs Tf calculées en fonction des valeurs expérimentales	42
12	Test de randomisation associé au modèle QSPR	43

SYMBOLES ET ABBREVIATIONS

- 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.
- Q²_{LOO}:** Coefficient de prédiction.
- Q²_{ext} :** Coefficient de prédiction externe.
- QSAR:** Quantitative Structure/ Activity Relationships. Relations Structure/Activités Quantitatives).
- QSPR:** Quantitative Structure/ Propriety Relationships. Relations Structure/Propriétés Quantitatives).
- R²:** 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
GENERALE



Les éthers sont des composés organiques qui sont largement utilisés comme solvants en raison de leur bonne solubilité à la fois dans l'eau et dans les solvants organiques. Parmi leurs applications générales, il faut mentionner les encres et les colorants, les vernis, les peintures et les agents de nettoyage à sec et de nettoyage des vitres. L'industrie des semi-conducteurs fait aussi beaucoup appel à ces composés en tant que solvants et agents de nettoyage.

La détermination expérimentale des propriétés physiques comme les températures d'ébullition, de fusion et critique peut s'avérer dangereuse avec ces composés toxiques, ou se traduire par une lourde charge, économiquement inacceptable.

Les études de simulation qui permettent une précision souvent meilleure que la répétabilité des mesures inter-laboratoires constituent une alternative intéressante.

En raison de cette importance, il serait utile de développer des modèles prédictifs, qui peuvent être employés pour prévoir ces propriétés pour de nouveaux composés sans expérimentation. Il y a beaucoup des travaux précédents qui ont visé à prévoir ces propriétés. Jusqu'ici, la méthode la plus prometteuse est le QSPR, qui emploie des descripteurs dérivés seulement de la structure moléculaire.

Dans toutes les techniques QSPR, les applications pratiques nécessitent au départ un ensemble de composés de propriétés connues. Cet ensemble, dit de calibration, est utilisé dans une première étape pour le calcul d'un modèle prédictif. Ce qui conduit à des équations mathématiques simples associant, aussi bien que possible, les descripteurs (ou régresseurs) et les propriétés mesurées (observations ou encore variables dépendantes). Si l'ensemble de calibration constitue un échantillon représentatif de la population, on admet alors que l'introduction de nouveaux éléments dont la valeur de la variable dépendante est inconnue (qu'on désigne encore par données de prédiction), n'affectera pas la stabilité du modèle, et que l'on peut ainsi estimer ces valeurs manquantes avec suffisamment de confiance.

Dans ce travail, nous nous sommes intéressés à la température de fusion d'une série d'éthers en utilisant les algorithmes génétiques pour la sélection d'un ensemble limité de variables explicatives ainsi que la régression multilinéaire (MLR) pour le calcul du modèle.

Nous présentons dans cette étude en plus d'une introduction et une conclusion générale, deux parties:

La première partie est consacrée 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).

La Partie Application, nous présentons et discutons le modèle calculé ainsi que les résultats obtenus.



PARTIE
THEORIQUE



I-GENERALITE

I-1-Introduction

Les Modèles QSPR et QSAR consistent en un ensemble de méthodes statistiques permettant de relier les propriétés physiques et d'activité d'un produit chimique à sa structure chimique. Des applications importantes peuvent être trouvées aujourd'hui dans de nombreux domaines de la chimie, de la biologie et de l'ingénierie. Le problème central dans QSPR et QSAR est de convertir les structures chimiques en descripteurs moléculaires qui sont pertinents pour une certaine propriété physique ou bio-activité (Diudea, 2000).

Diverses propriétés physico- chimiques peuvent être soumises à des études QSPR / QSAR, mais l'application principale est la découverte de médicaments - bien que leurs capacités de modélisation prédictive se soient étendues également à d'autres domaines, par exemple la catalyse (Maldonado et Rothenberg, 2009).

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 «OSPR»

I-2-1-Définition de la méthode OSPR

Le QSPR (Quantitative Structure-Property Relationship) 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

Les méthodes QSPR reposent sur le principe suivant : des molécules ayant les mêmes propriétés sont proches dans l'espace chimique. L'espace chimique (Figure 01) est un espace à n dimensions qui correspondent à des variables décrivant la structure moléculaire. Ce sont ces variables, appelées descripteurs, qui vont être reliées à la propriété étudiée par l'intermédiaire de différentes méthodes statistiques. La propriété (Y) est exprimée comme une fonction de la structure : $Y = f(X)$

$$Y = f(\text{structure}) = f(\text{descripteurs}) \quad (01)$$

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 linéaires multiples (RLM), 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 [3].

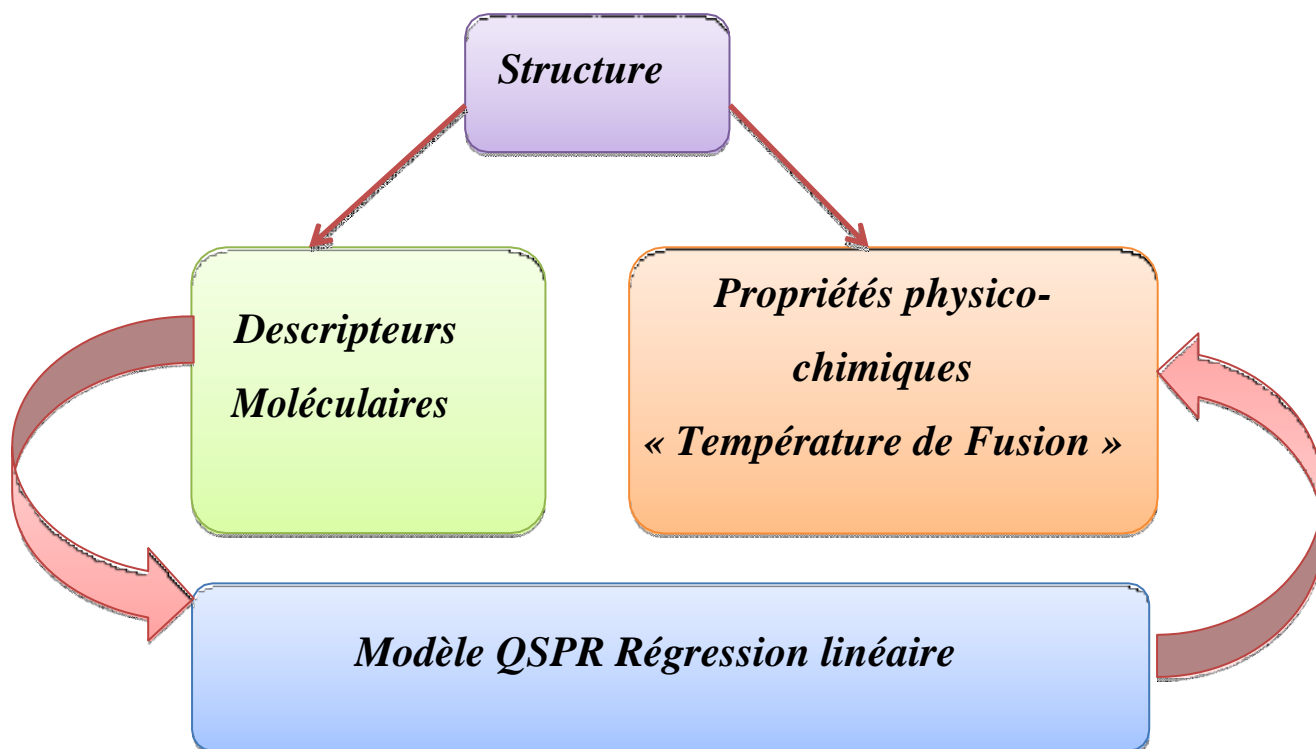


Figure 01- Principe de la méthode QSPR

I-2-3- Les méthodes mathématiques utilisés par le model QSPR

Pour les méthodes utilisées en QSPR sont deux types:

a) **Linéaires**

- Régression linéaire simple
- Régression linéaire multiple MLR.
- Régressions aux moindres carrés partielles (PLS).

b) **Non linéaires**

- Réseau de neurones artificiel RNA.
- SVM.
- Arbres de décision.

I-2-4- La relation structure- propriété quantitative (OSPR)

Les relations de propriété de structure quantitative (QSPR) sont des modèles prédictifs permettant de calculer les propriétés de composés chimiques à partir de leurs seules structures moléculaires. Ils reposent sur une relation mathématique entre la propriété et des descripteurs de la structure moléculaire des composés ciblés.

Si ces modèles ont été largement utilisés pour la prédiction d'activités biologiques en toxicologie ou en pharmacologie depuis de nombreuses années, ils trouvent aujourd'hui de plus en plus d'applications pour la prédiction de propriétés physico-chimiques. L'essor de ces approches s'est même récemment accentué avec la mise en place du règlement européen REACH qui recommande l'emploi de tels modèles, une fois validés, pour l'acquisition des données nécessaires à l'enregistrement des substances chimiques [4].

II- COLLECTE DES DONNEES

La propriété considérée (Température de fusion) est l'une des grandeurs physiques importantes. Les données utilisées dans ce travail, concernent 80 composés, [5] sont réunies dans le tableau (01) : On a choisi aléatoirement 20 éthers [pour la validation externe] ; les 20 derniers éthers du tableau (01), et le reste (60 éthers) pour la calibration ou à la construction du modèle.

Tableau 01- Nomenclature et valeurs de la propriété étudiée

N°	Composé	Tf
1	Diméthyleéther	159,05
2	Diéthyléther	156,86
3	Méthyl t-butyléther	164,4
4	Di-isopropyléther	187,6
5	Butyléthyléther	149
6	Di-n-butyléther	177,8
7	2-Methylfuran	181,7
8	Tétrahydrofuran	164,56
9	Tetrahydropyran	223,9
10	Epichlorohydrin	247

Tableau 1 : (Suite)

N°	Composé	Tf
11	Chloromethylméthyléther	169,5
12	Bis(chlorométhyl)éther	231,5
13	Bis(2-chloroéthyl)éther	221,1
14	Bis(2-chloroisopropyl)éther	176
15	2-Chloroéthyl vinyléther	203
16	Bis(2-chloroéthoxy)méthane	240,7
17	Anisole	235,87
18	2-Chloroanisole	246,2
19	3-Chloroanisole	241
20	4-Chloroanisole	255
21	2,3-Dichloroanisole	305
22	2,6-Dichloroanisole	283
23	2, 4,6-Trichloroanisole	334,5
24	2, 3, 4,5-Tetrachloroanisole	361
25	2, 3, 5,6-Tetrachloroanisole	357
26	4,5-Dichloroveratrole	356
27	3, 4,5-Trichloroveratrole	339
28	Tetrachloroveratrole	363
29	Benzyléthyléther	275,5
30	Styreneoxide	237,4
31	Diphenyléther	299,87
32	4-Chlorodiphenyl éther	265
33	2,6-Dichlorodiphenyl éther	312
34	2, 4,4'-Trichlorodiphenyl éther	313
35	2, 4,5-Trichlorodiphenyl éther	334
36	2,2', 4,4'-Tetrachlorodiphenyl éther	342
37	2, 4,4',5-Tetrachlorodiphenyl éther	335
38	3,3', 4,4'-Tetrachlorodiphenyl éther	343
39	2,2', 4,4',5-Pentachlorodiphenyl éther	298

Tableau-1 (suite)

N°	Composé	Tf
41	2, 3,3', 4,4' - Pentachlorodiphenyléther	338
42	2,2',3,4,4',5'-Hexachlorodiphenyl éther	342
43	2,2',3,4,4',6'-Hexachlorodiphenyl éther	394
44	2,2',4,4',5,5'-Hexachlorodiphenyl éther	388
45	2,2',4,4',5,6'-Hexachlorodiphenyl éther	369
46	2,3',4,4',5,5'-Hexachlorodiphenyl éther	378
47	2,2',3,4,4',5,5'-Heptachlorodiphenyl éther	362
48	2,2',3,4,4',6,6'-Heptachlorodiphenyl éther	415
49	2,2', 3,3', 4,4', 5,5'-Octachlorodiphenyl éther	399
50	2,2', 3,3', 4,4',6,6'-Octachlorodiphenyl éther	399
51	4-Bromodiphenyl éther	291
52	4,4'-Dibromodiphenyl éther	330,7
53	2, 4,4'-Tribromodiphenyl éther	337,25
54	3,3', 4,4'-Tetrabromodiphenyl éther	371
55	2,2', 3, 4,4'-Pentabromodiphenyl éther	393
56	2,2', 4,4',5-Pentabromodiphenyl éther	365,5
57	2,2', 4,4',6-Pentabromodiphenyl éther	383
58	2,2',4,4',5,5'-Hexabromodiphenyl éther	434,5
59	2,2',4,4',5,6'-Hexabromodiphenyl ether	405
60	2,2', 3, 4, 5,5',6-Heptabromodiphenyl éther	445
61	Di-n-propyléther	158,2
62	1,2 propylèneoxide	161,1
63	Furan	187,39
64	1,4-Dioxane	284,95
65	2, 3,4-Trichloroanisole	343
66	Veratrole	295,5
67	Phenetole	243,57
68	2-Chlorodiphenyl éther	318
69	2,4-Dichlorodiphenyl éther	273

Tableau-1 suite et fin

N°	Composé	Tf
70	2,2', 3,4,4'-Pentachlorodiphenyl éther	339
71	3,3', 4,4',5-Pentachlorodiphenyl éther	343
72	2,2',3,3',4,4'-Hexachlorodiphenyl ether	368
73	2,2', 3, 4,4',5-Hexachlorodiphenyl éther	353
74	2,2',3,4,4',5,6'-Heptachlorodiphenyl ether	409
75	2,2', 3,3', 4,4', 5,6'-Octachlorodiphenyl éther	422
76	2,2', 3,3', 4,4', 5,5',6-Nonachlorodiphenyl éther	450
77	Decachlorodiphenyléther	404
78	2,2', 4,4'-Tetrabromodiphenyl éther	357
79	2,3,3',4,4',5'-Hexabromodiphenyl éther	409
80	Decabromodiphenyléther	575,5

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

III-1 -Préparation de base des données

III-1-1-Logiciels « ChemDraw»

ChemDraw fournit des chimistes avec un ensemble d'outils riche, faciles à utiliser pour créer des publications prêt, dessins scientifiquement significatifs de molécules et de réactions [6]. Essayant de concevoir et dessiner de nouvelles molécules en utilisant les différents outils disponibles dans ChemDraw, et de les visualiser, et plus important encore, les enregistrer dans différents formats. ChemDraw est très pratique pour les réactions d'écriture à l'aide de produits chimiques. Vous devriez être en mesure de réaliser des structures dans différents formats disponibles [7]. **EX**: 2,4,6-Trichloroanisole

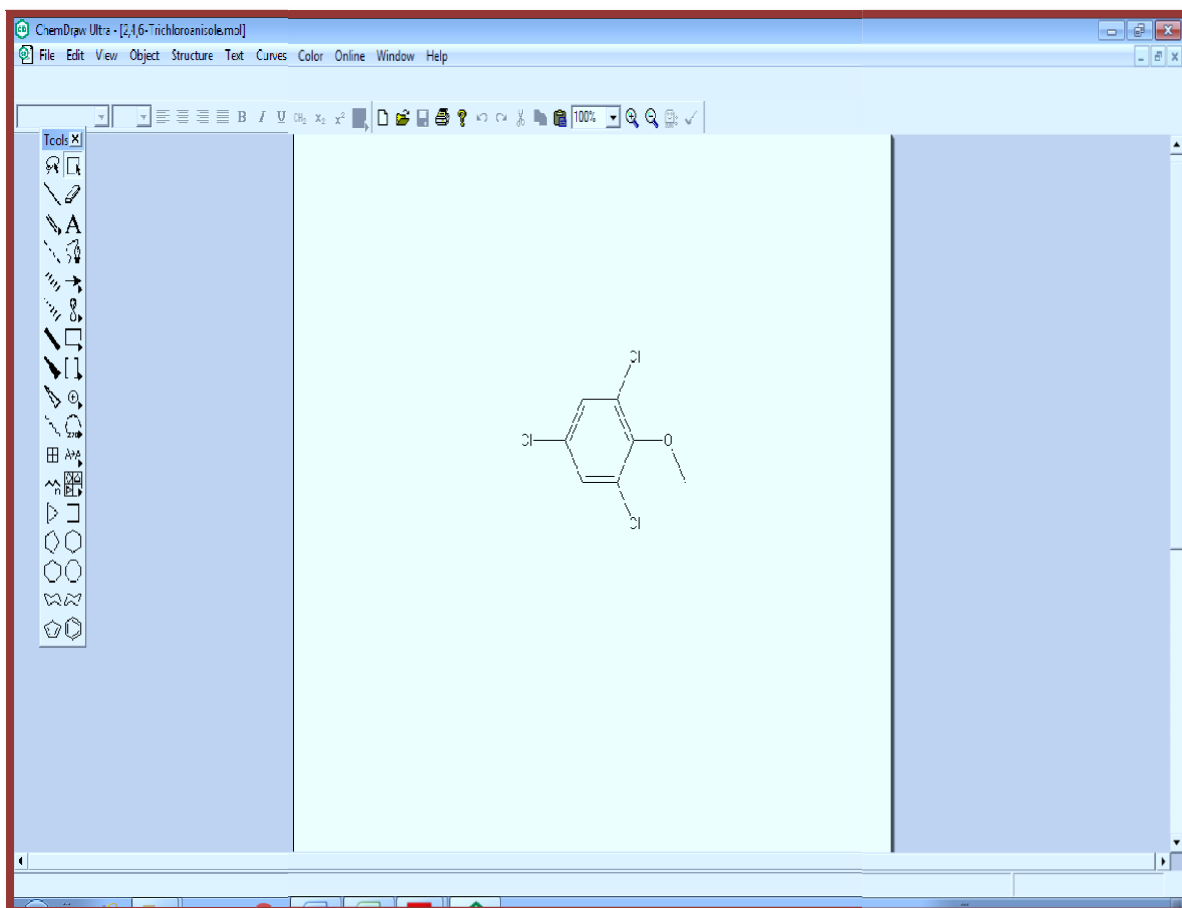


Figure 02- Représentation des molécules par le ChemDraw

III-1-2-Logiciels « HyperChem 8 »

HyperChem [8] 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 103 molécules donc on obtient 103 fichiers Hin, en suite on calcule les descripteurs moléculaires à partir de ces fichiers par le logiciel Dragon [9].

III-1-3-Récupération et stabilisation les molécules de fichier Hin

III-1-3-1- Stabilisation structure des molécules (minimisation de l'énergie)

Pour stabilise forme de structure de chaque molécule ou minimisation de l'énergie on utilise le HyperChem, 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 [10].

III-1-3-2- 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 [11].

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 concerta à méthode MM+ de caractéristiques Field Force présente dans figure(3) suivante :

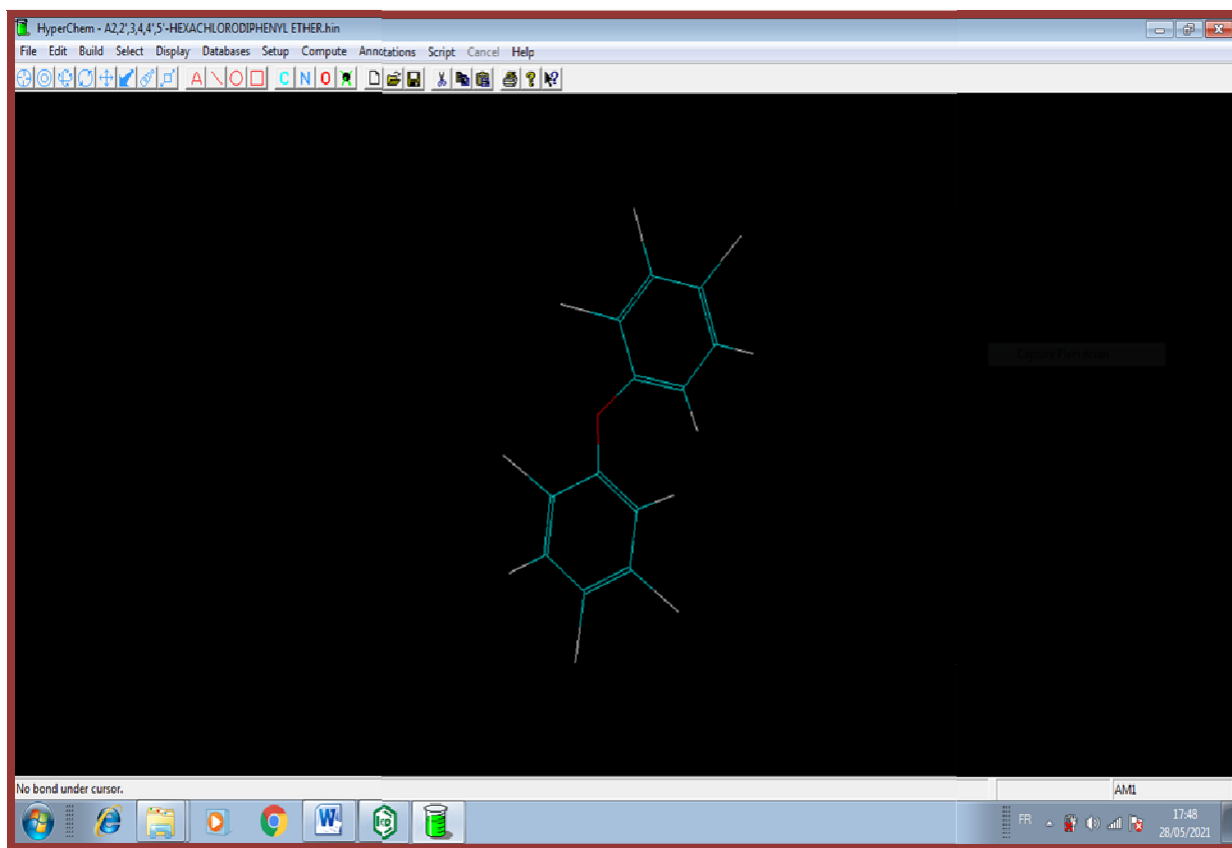


Figure 03- Logiciel Hyperchem

III-1-3-3- Récupération des fichiers HyperChem HIN

Une fois que vous avez construit une structure en HyperChem, vous pouvez l'enregistrer pour une utilisation ultérieure. C'est une bonne idée puisque cela vous fait gagner du temps si vous voulez revoir votre structure à une date ultérieure. Pourquoi construire deux fois ?! Vous pouvez le faire en allant dans Fichier et Enregistrer. Vous devez 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[10].

III-2- 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 structurales. Dans la prochaine étape pour tous les 65 composés, descripteurs moléculaires ont été calculés par le logiciel Dragon. Dragon peut calculer plus de 1600 descripteurs moléculaires pour chaque structure dans notre jeu des données [12].

III-2-1- 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 [12].



Figure 04- Le Logiciel « Dragon »

III-2-2-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 structurales. Ces grandeurs sont appelées descripteurs autrement dit c'est un Codage de la structure en nombre.

III-2-2-1-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 T_b des descripteurs moléculaires ont été proposés dérivé de théories différentes et approches avec le but de prédire biologique et propriétés physico-chimiques des molécules [13]. Il y a plusieurs de descripteurs :

- Descripteurs constitutionnels
- Descripteurs topologiques
- Descripteurs géométriques
- Descripteurs électrostatiques
- Descripteurs quantiques
- Descripteurs thermodynamiques

III-2-2-2- 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. [14]. 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- 02 Quelques blocks des descripteurs calculés par logiciel dragon

Classe	Sous-classe
Descripteurs constitutionnels	<ul style="list-style-type: none"> - Dénombrement des atomes ou des liaisons. - Descripteurs basés sur les masses atomiques.
Descripteurs topologiques	<ul style="list-style-type: none"> - Indices topologiques (connectivité). - Descripteurs théoriques d'information. - Descripteurs topo-chimiques.
Descripteurs géométriques	<ul style="list-style-type: none"> - Descripteurs liés à la distance. - Descripteurs liés à l'aire de la surface. - Descripteurs liés au volume. - Descripteurs du champ stérique moléculaire.
Descripteurs liés à la distribution de charge	<ul style="list-style-type: none"> - Charges atomiques partielles. - Moments électriques moléculaires - Polarisabilités moléculaires. - Descripteurs du champ électrique moléculaire.
Descripteurs liés aux orbitales Moléculaires	<ul style="list-style-type: none"> - Energie des OM frontières - Ordres de liaison - Indices de réactivité de Fukui.
Descripteurs température Dépendants	<ul style="list-style-type: none"> - Fonctions thermodynamiques. - Descripteurs facteurs de Boltzmann pondérés.
Descripteurs de solvation	<ul style="list-style-type: none"> - Energie électrostatique de solvation. - Energie de dispersion de solvation. - Enthalpie libre de formation de cavité. - Descripteurs de liaison hydrogène. - Entropie de solvation. - Descripteurs d'énergie de solvation linéaire théorique.
Descripteurs mixtes	<ul style="list-style-type: none"> - 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, ils peuvent être obtenus quand les molécules sont transformées dans une représentation moléculaire qui autorise quelque traitement mathématique. Les descripteurs moléculaires sont très importants pour :

- ❖ Indiquons la description de la configuration de la molécule à étudier.
- ❖ Décrivons 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 [15].

III-4- Diagramme montre les étapes de prédiction

Les premiers essais de modélisation QSPR développés 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 [16].

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

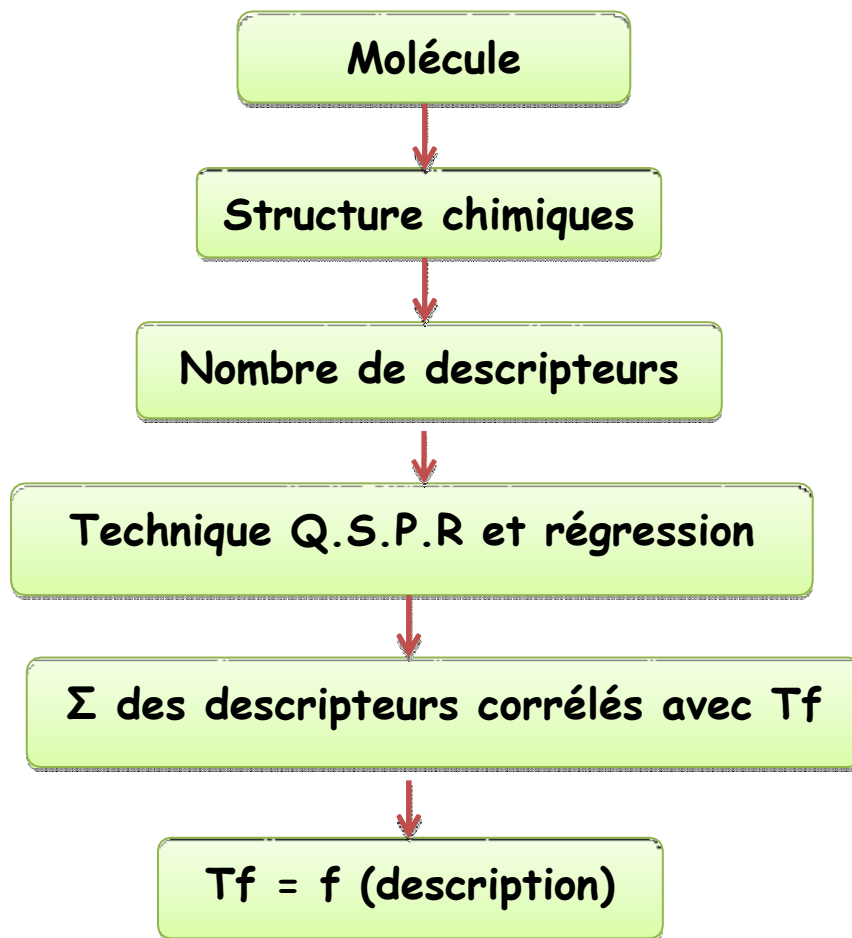


Figure 05- Diagramme de prédiction par QSPR

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.

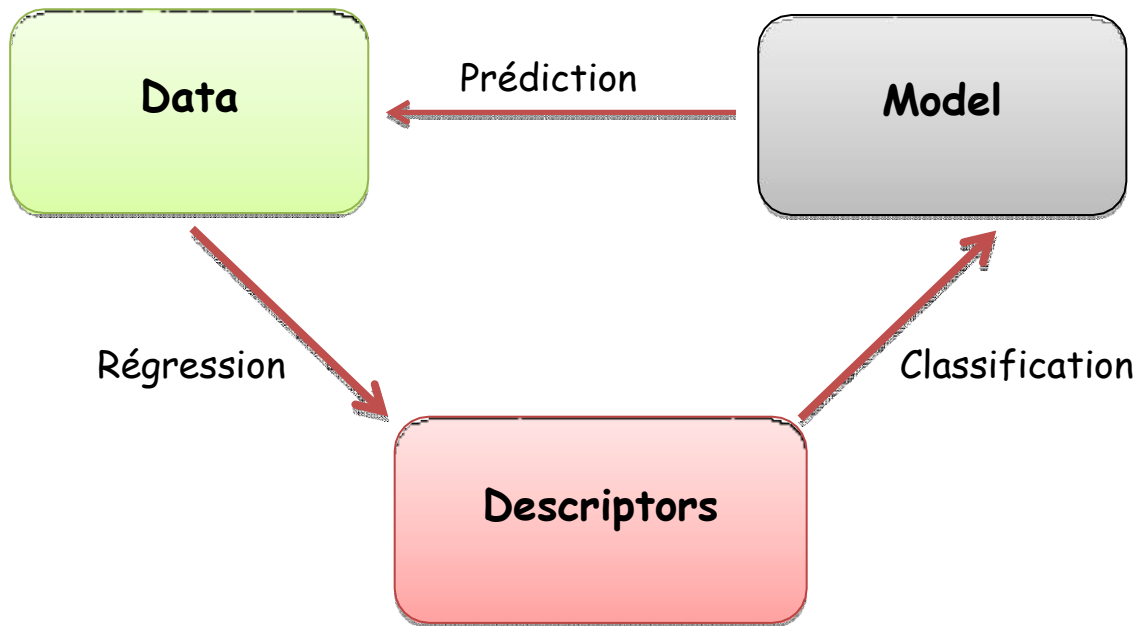


Figure 06 - Le cycle de prédiction

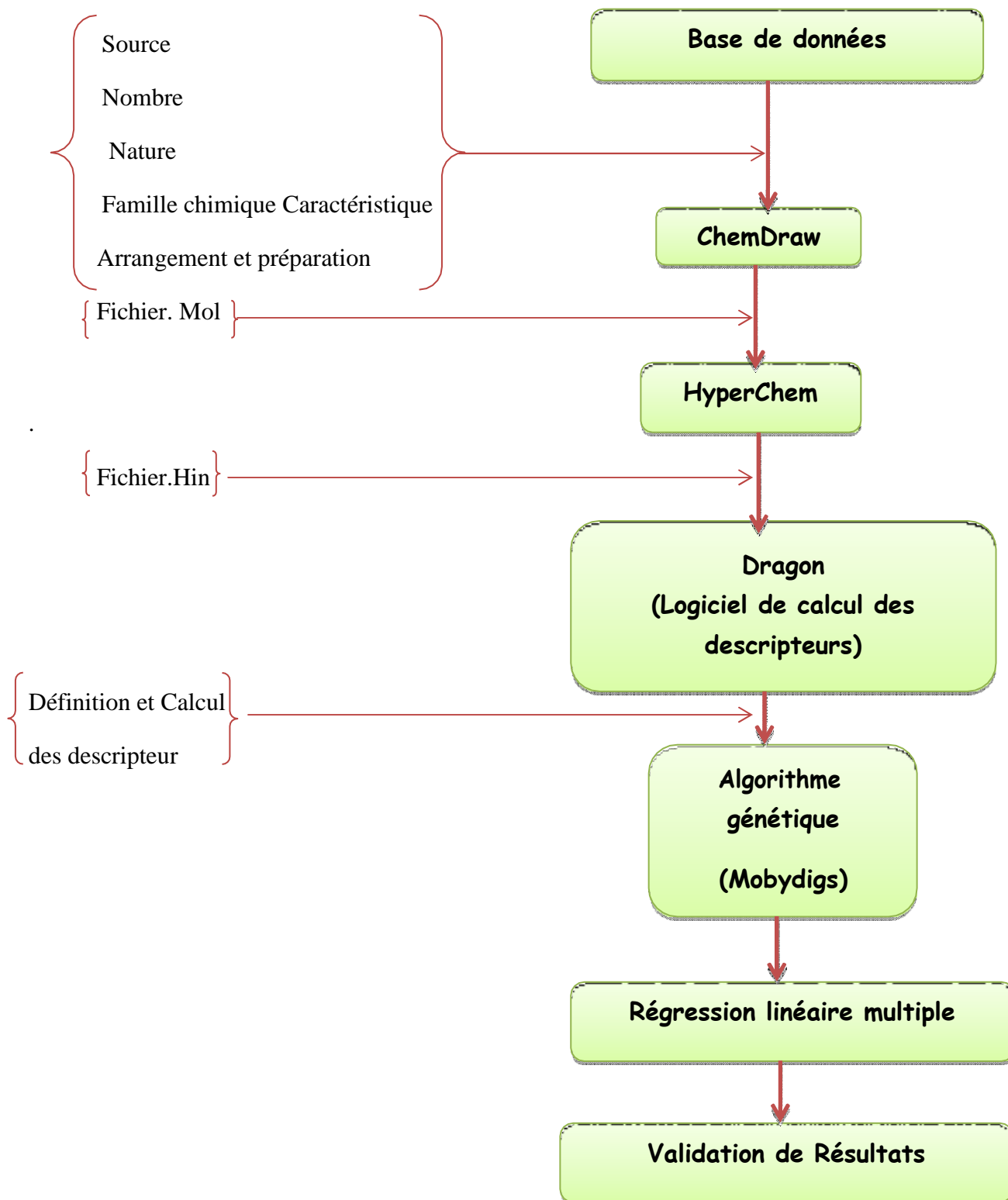


Figure07 -Diagramme de notre travail

Le diagramme précédant illustré les étapes du travail et les techniques quant utilisé dans la procédure de prédiction de la propriété étudiée (Tf).

IV- METHODES UTILISEES POUR LE DEVELOPPEMENT DE MODELES 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 facile [17]. Tout d'abord, un très grand nombre (>1600) 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 multi colinéarité 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 principaux. Une alternative aux méthodes très classiques de régression linéaire multiple (MLR) et d'analyse en composantes principales (ACP) est la technique de régression par les moindres carrés partiels (MCP ou PLS) [18-22].

IV -2- Méthodes de régressions linéaires et multilinéaires

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 :

- Un nombre maximal de données expérimentales (de préférence toutes) doivent être caractérisées par des valeurs de descripteurs originaux complémentaires.
- 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.
- 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...).

- Dans les modèles MLR, les descripteurs utilisés doivent être statistiquement orthogonaux.
- 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 [12].

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 corrélérer (variable dépendante) et les descripteurs utilisés dans la corrélation (variables indépendantes).

On distingue habituellement dans un tel pré- traitement des données les analyses univariées des analyses bi-variées [23,28].

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 -2-3- Régression linéaire multiple

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

$$Y_i = Q_0 + \sum_{j=1}^p \beta_j X_{ij} + \varepsilon_i \quad (02)$$

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 β_j sont calculés en utilisant la méthode des moindres carrés ordinaires. Les variables aléatoires ξ 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 :

- Les résidus (E) ont une espérance mathématique nulle : $E(e) = 0$
- Le modèle choisi est correct (aucune variable explicative n'a été omise).
- Les résidus sont indépendants entre eux :

$$E(e_i, e_j) = 0 \quad \text{Si } i \neq j \quad (03)$$

Leurs covariances sont nulles.

- Les résidus ont tous même variance σ^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 [29] tel que :

$$\mathbf{n - p - 1 \geq 10} \quad (04)$$

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

IV -3- Méthodes de sélection des descripteurs

❖ 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 [30]. 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 les sélections des descripteurs en utilisant le logiciel de calcul statistique MINITAB version 16.2.0 [31] ; et par algorithme génétique, dans la version MOBY DIGS de Todeschini [32].

IV-4- 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 de fusion) "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.

$$R^2 = 1 - \frac{SCE}{SCT} = 1 - \frac{\sum_1^n (y_i - \hat{y}_i)^2}{\sum_1^n (y_i - \bar{y})^2} \quad (05)$$

où \hat{y}_i est la valeur estimée du paramètre physique, et \bar{y} la moyenne des valeurs expérimentales.

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

$$SDEP = \sqrt{\frac{1}{n} \sum_1^n (y_i - \hat{y}_{(i)})^2} = \sqrt{\frac{PRESS}{n}} \quad (06)$$

IV -5-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.[12]

IV -6- 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).

IV -7- 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 (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

$$SDEC = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (07)$$

$$SDEP_{\text{ext}} = \sqrt{\frac{\sum_{i=1}^{n_{\text{ext}}} (y_i - \hat{y}_{(i)})^2}{n_{\text{ext}}}} \quad (08)$$

La validation croisée par « leave – one - out » (LOO) [33] 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, est une mesure de la dispersion des estimations. On l'utilise pour définir le coefficient de prédiction :

$$Q_{\text{LOO}}^2 = \frac{\text{SCT} - \text{PRESS}}{\text{SCT}} \quad (09)$$

Contrairement à R^2 au coefficient qui augmente avec le nombre de paramètres du modèle, le facteur Q_{LOO}^2 affiche une courbe avec maximum (ou avec palier) obtenu pour un certain nombre de descripteurs, puis décroît de façon monotone. Ce fait confère une grande importance au coefficient Q_{LOO}^2 . Une valeur $Q_{\text{LOO}}^2 > 0.5$ est considérée comme satisfaisante, et une valeur supérieure à 0,9 est excellente [34].

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

Evidemment, on peut être amené à écarter 2, 3 ou un plus grand nombre d'éléments à la fois, ce qui conduit aux procédures LMO (leave – many- out). 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_{ext}^2 , est calculée comme suit :

$$Q_{\text{ext}}^2 = 1 - \frac{\sum_{i=1}^{n_{\text{ext}}} (y_i - \hat{y}_i)^2 / n_{\text{ext}}}{\sum_{i=1}^n (y_i - \bar{y})^2 / n} \quad (10)$$

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



PARTIE

APPLICATION



Nous traiterons dans ce travail la propriété physique envisagée (T_f), en considérant les mêmes ensembles d'estimation et de validation. Rappelant que les 80 composés ont été éclatés aléatoirement en deux sous ensemble comportant 60 pour la calibration et 20 pour la validation. Les résultats ainsi obtenus seront discutés.

V-1-Conditions expérimentales

Nous avons commencés à utiliser les algorithmes génétiques en utilisant le logiciel Mobydigs [33], ce dernier nous permet de calculer une centaine de modèles de différentes dimensions, parmi ces modèles nous avons choisi celui porte le moins de descripteurs et obéi aux conditions expérimentales telles que :

La corrélation entre les descripteurs, le FIV et les grandes valeurs de R^2 , Q^2 et Q^2_{ext} .

V-2-Choix de la taille du modèle

En traçant la courbe R^2 en fonction de nombre de descripteurs k (figure suivante) on peut dire que notre modèle optimal est celui à 5 descripteurs parce qu'après cela l'amélioration est faible lorsqu'on ajoute d'autre descripteurs.

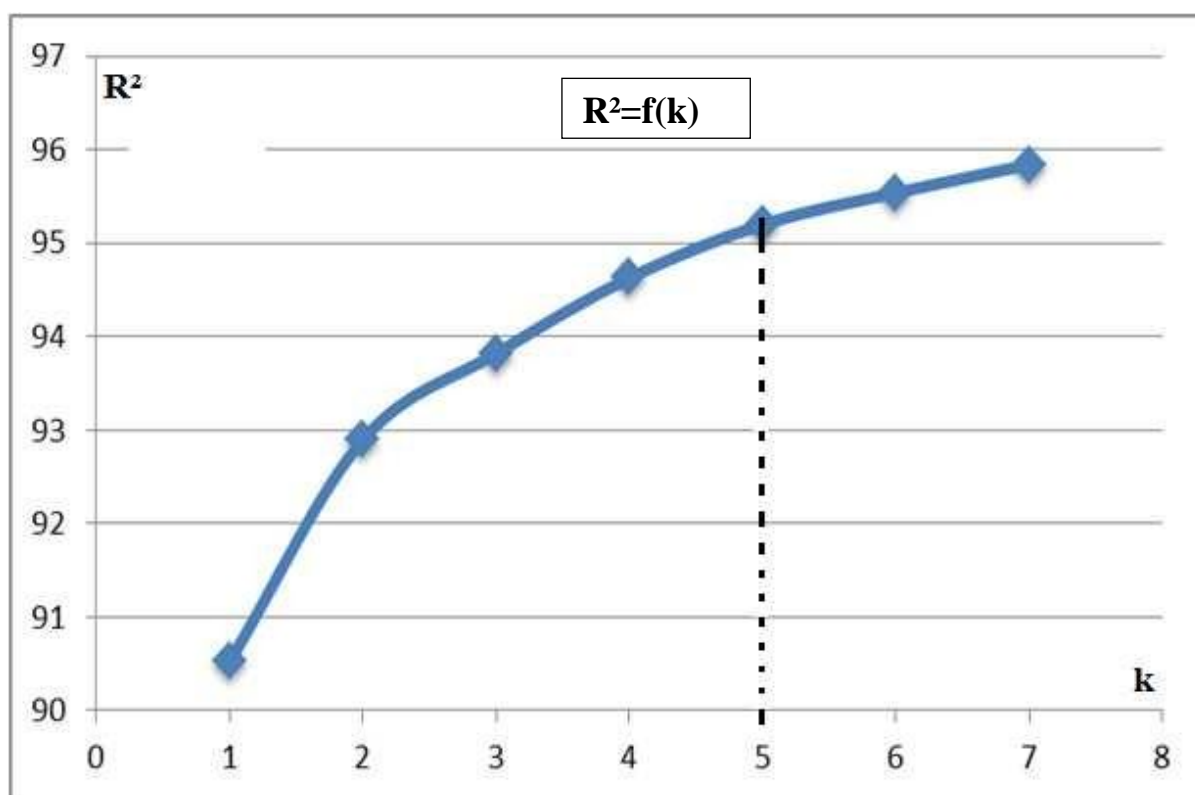


Figure 08 - Variation de R^2 en fonction du nombre de descripteurs k

V-3-Sélection des descripteurs

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

Parmi les modèles obtenus pour la taille optée, on a sélectionné les 5 descripteurs qui donnent les meilleurs résultats.

Nous rapporterons leurs valeurs, classes et quelques significations dans les tableaux (03) et (04)

Tableau -03 Valeurs des descripteurs moléculaires sélectionnés

N°	Composé	Me	PW5	HATS1m	HATS2p	Kv
1	Diméthyléther	1	0	0,062	0,265	0,649
2	Diéthyléther	0,98	0	0,029	0,157	0,74
3	Méthyl t-butyléther	0,98	0	0,036	0,129	0,263
4	Di-isopropyléther	0,98	0	0,032	0,127	0,495
5	Butylethyléther	0,98	0,043	0,026	0,124	0,842
6	Di-n-butyléther	0,97	0,05	0,022	0,106	0,896
7	2-Méthylfuran	1	0,023	0,062	0,194	0,608
8	Tetrahydrofuran	0,99	0	0,074	0,232	0,332
9	Tetrahydropyran	0,98	0,063	0,07	0,213	0,331
10	Epichlorohydrin	1,03	0	0,197	0,386	0,659
11	Chlorométhylméthyléther	1,03	0	0,149	0,447	0,739
12	Bis(chlorométhyl)éther	1,07	0	0,232	0,589	0,723
13	Bis(2-chloroéthyl)éther	1,03	0,043	0,154	0,261	0,837
14	Bis(2-chloroisopropyl)éther	1,01	0	0,059	0,209	0,406
15	2-Chloroéthyl vinyléther	1,01	0,033	0,092	0,233	0,829
16	Bis(2-chloroéthoxy)méthane	1,03	0,05	0,13	0,204	0,885
17	Anisole	0,99	0,068	0,057	0,186	0,602
18	2-Chloroanisole	1,01	0,063	0,116	0,204	0,486
19	3-Chloroanisole	1,01	0,074	0,119	0,219	0,614
20	4-Chloroanisole	1,01	0,078	0,102	0,217	0,672
21	2,3-Dichloroanisole	1,03	0,066	0,162	0,228	0,512
22	2,6-Dichloroanisole	1,03	0,059	0,118	0,209	0,488
23	2,4,6-Trichloroanisole	1,05	0,066	0,161	0,234	0,503
24	2,3,4,5-Tetrachloroanisole	1,07	0,08	0,253	0,266	0,491
25	2,3,5,6-Tetrachloroanisole	1,07	0,081	0,216	0,248	0,491
26	4,5-Dichloroveratrole	1,04	0,095	0,147	0,172	0,485
27	3,4,5-Trichloroveratrole	1,05	0,091	0,153	0,176	0,486
28	Tetrachloroveratrole	1,07	0,095	0,473	0,844	0,445
29	Benzylethyl éther	0,98	0,083	0,055	0,145	0,744

Tableau- 03 (Suite)

N°	Composé	Me	PW5	HATS1m	HATS2p	Kv
30	Styreneoxide	0,99	0,068	0,117	0,196	0,605
31	Diphenyl éther	0,99	0,09	0,049	0,095	0,77
32	4-Chlorodiphenyl éther	1	0,092	0,08	0,114	0,797
33	2,6-Dichlorodiphenyl éther	1,02	0,091	0,084	0,112	0,693
34	2, 4,4'-Trichlorodiphenyl éther	1,03	0,093	0,14	0,142	0,772
35	2, 4,5-Trichlorodiphenyl éther	1,03	0,094	0,148	0,139	0,714
36	2,2', 4,4'-Tetrachlorodiphenyl éther	1,05	0,092	0,164	0,151	0,732
37	2, 4,4',5-Tetrachlorodiphenyl éther	1,05	0,095	0,176	0,156	0,739
38	3,3', 4,4'-Tetrachlorodiphenyl éther	1,05	0,084	0,183	0,167	0,8
39	2,2', 4,4',5-Pentachlorodiphenyl éther	1,06	0,094	0,197	0,164	0,705
40	2,2',4,4',6-Pentachlorodiphenyl éther	1,06	0,091	0,16	0,153	0,721
41	2,3,3',4,4' - Pentachlorodiphenyl éther	1,06	0,083	0,204	0,174	0,766
42	2,2',3,4,4',5'-Hexachlorodiphenyl ether	1,07	0,088	0,223	0,177	0,714
43	2,2',3,4,4',6'-Hexachlorodiphenyl ether	1,07	0,086	0,19	0,167	0,723
44	2,2',4,4',5,5'-Hexachlorodiphenyl ether	1,06	0,084	0,489	0,243	0,709
45	2,2',4,4',5,6'-Hexachlorodiphenyl ether	1,07	0,093	0,192	0,166	0,699
46	2,3',4,4',5,5'-Hexachlorodiphenyl ether	1,07	0,096	0,225	0,175	0,678
47	2,2',3,4,4',5,5'-Heptachlorodiphenyl ether	1,09	0,09	0,252	0,188	0,684
48	2,2',3,4,4',6,6'-Heptachlorodiphenyl ether	1,09	0,09	0,192	0,169	0,704
49	2,2', 3,3', 4,4', 5,5'-Octachlorodiphenyl éther	1,1	0,086	0,278	0,201	0,686
50	2,2', 3,3',4,4',6,6'-Octachlorodiphenyl éther	1,1	0,09	0,221	0,181	0,696
51	4-Bromodiphenyl éther	1	0,092	0,124	0,126	0,811
52	4,4'-Dibromodiphenyl éther	1,01	0,093	0,194	0,156	0,836
53	2, 4,4'-Tribromodiphenyl éther	1,02	0,093	0,272	0,174	0,771
54	3,3', 4,4'-Tetrabromodiphenyl éther	1,03	0,084	0,368	0,212	0,808

Tableau -03 (Suite et fin)

N°	Composé	Me	PW5	HATS1m	HATS2p	Kv
55	2,2', 3, 4,4'-Pentabromodiphenyl éther	1,04	0,086	0,39	0,211	0,721
56	2,2', 4,4',5-Pentabromodiphenyl éther	1,04	0,094	0,401	0,208	0,677
57	2,2',4,4',6-Pentabromodiphenyl éther	1,04	0,091	0,319	0,192	0,693
58	2,2',4,4',5,5'-Hexabromodiphenyl éther	1,05	0,096	0,464	0,225	0,64
59	2,2',4,4',5,6'-Hexabromodiphenyl ether	1,05	0,093	0,389	0,211	0,665
60	2,2', 3, 4, 5,5',6-Heptabromodiphenyl éther	1,06	0,095	0,459	0,227	0,605
61	Di-n-propyléther	0,98	0,043	0,026	0,132	0,857
62	1,2 propyleneoxide	1	0	0,119	0,264	0,552
63	Furan	1,01	0	0,086	0,138	0,5
64	1,4-Dioxane	1,01	0,063	0,103	0,229	0,354
65	2, 3,4-Trichloroanisole	1,05	0,07	0,204	0,253	0,543
66	Veratrole	1	0,077	0,052	0,134	0,48
67	Phenetole	0,99	0,078	0,047	0,16	0,726
68	2-Chlorodiphenyl éther	1	0,09	0,08	0,107	0,718
69	2,4-Dichlorodiphenyl éther	1,02	0,092	0,111	0,125	0,747
70	2,2', 3,4,4'-Pentachlorodiphenyl éther	1,06	0,086	0,193	0,166	0,736
71	3,3',4,4',5-Pentachlorodiphenyl éther	1,06	0,081	0,217	0,179	0,771
72	2,2',3,3',4,4'-Hexachlorodiphenyl ether	1,07	0,081	0,222	0,18	0,736
73	2,2', 3, 4,4',5-Hexachlorodiphenyl éther	1,07	0,089	0,226	0,178	0,705
74	2,2',3,4,4',5,6'-Heptachlorodiphenyl ether	1,09	0,088	0,222	0,179	0,697
75	2,2', 3,3', 4,4', 5,6'-Octachlorodiphenyl éther	1,1	0,088	0,249	0,192	0,697
76	2,2', 3,3', 4,4', 5,5',6-Nonachlorodiphenyl éther	1,12	0,088	0,275	0,202	0,682
77	Decachlorodiphenyléther	1,13	0,09	0,274	0,203	0,679
78	2,2', 4,4'-Tetrabromodiphenyl ether	1,03	0,092	0,328	0,189	0,713
79	2,3,3',4,4',5'-Hexabromodiphenyl ether	1,05	0,08	0,485	0,24	0,731
80	Decabromodiphenyléther	1,09	0,09	1,112	0,507	0,579

❖ Les 20 derniers composés (61-80) sont destinés à la validation externe.

Tableau- 04 Classes et significations des descripteurs

Descripteur	Classe	Signification
Me	Constitutional descriptors	Constitutional descriptors are the most simple and commonly used descriptors, reflecting the molecular composition of a compound without any information about its molecular geometry. The mean atomic van der Waals volume (scaled on Carbon atom).
PW5	Topological descriptors	are calculated by summing the ratios of the atomic path count over the atomic walk count of the same order k and then dividing by the number of non-H atoms (nSK). Since path/walk count ratio is independent of molecular size, these descriptors can be considered as shape descriptors. DRAGON calculates path/walk shape indices from order 2 up to 5; the index of first order is not provided as the counts of the paths and walks of length one are equal and, therefore, the corresponding molecular index always equals one for all molecules.
Kv	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.
HATS1m	GETAWAY descriptors	are molecular descriptors for structure-property correlations, but they can also be used as molecular profiles suitable for similarity/diversity analysis studies. These descriptors, as based on spatial autocorrelation, encode information on structural fragments and therefore seem to be particularly suitable for describing differences in congeneric series of molecules. Differently from the Moreau-Broto autocorrelations, GETAWAYs are geometrical descriptors encoding information on the effective position of substituents and fragments in the molecular space. Moreover, they are independent of molecule alignment and, at some extent, account also for information on molecular size and shape as well as for specific atomic properties.
HATS2p		

V-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 60 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.

Tableau-05 Corrélation entre Tf et les descripteurs

Correlations: Tf; Me; PW5; HATS1m; HATS2p; Kv					
	Tf	Me	PW5	HATS1m	HATS2p
Me	0,770 0,000				
PW5	0,824 0,000	0,480 0,000			
HATS1m	0,770 0,000	0,648 0,000	0,482 0,000		
HATS2p	-0,014 0,914	0,267 0,039	-0,233 0,073	0,394 0,002	
Kv	0,120 0,362	0,073 0,579	0,257 0,047	0,064 0,628	-0,212 0,103

Cell Contents: Pearson correlation
P-Value

Cette matrice nous permet de voir que les descripteurs Me, PW5 et HATS1m sont bien corrélés avec la propriété, par contre les deux autres sont moins corrélés mais ils portent un complément pour le modèle.

Egalement les 3 descripteurs ne sont pas corrélés entre eux à l'exception des deux autres qui présentent un $p > 0.05$.

V-5- Equations de régressions

L'équation et les paramètres de régression sont illustrés comme suite :

$$T_f = - 629 + 838 Me + 1066 PW5 + 280 HATS1m - 134 HATS2p - 48,4Kv \quad (11)$$

Tableau-06 paramètres de régression

Predictor	Coef	SE Coef	T	P	VIF
Constant	-628,8	110,2	-5,71	0,000	
Me	837,5	111,5	7,51	0,000	1,900
PW5	1065,7	117,7	9,05	0,000	2,023
HATS1m	279,76	35,09	7,97	0,000	2,365
HATS2p	-134,07	31,76	-4,22	0,000	1,726
Kv	-48,42	20,46	-2,37	0,022	1,101

V-6-Analyse de régression

Les statistiques calculées établissent la pertinence du modèle. En effet, la valeur de R^2 signifie que 99,80 % de la variabilité de T_f est expliquée par les 5 descripteurs sélectionnés.

Selon les valeurs du test t ($|t|$), on peut classer les descripteurs sélectionnés dans ce modèle d'après leur contribution qui se présente dans l'ordre: $PW5 > HATS1m > Me$. Les valeurs des VIF (< 5) suggèrent que ces descripteurs sont faiblement corrélés les uns avec les autres. Ainsi, le modèle peut être considéré comme une équation de régression optimale.

Tableau-07 Valeurs des T_f expérimentales, calculées, e_i , e_{std} et h_{ii} pour l'ensemble de calibration

N°	composé	$T_{f_{Exp}}$	$T_{f_{Pred}}$	e_i	e_{Std}	h_{ii}
1	Dimethylether	159,05	159,12	0,07	0	0,092
2	Diethylether	156,86	141,28	-13,65	-0,79	0,124
3	Methyl t-butylether	164,4	174,38	7,62	0,54	0,237
4	Di-isopropylether	187,6	155,64	-27,67	-1,62	0,134
5	Butylethylether	149	191,71	38,68	2,12	0,094
6	Di-n-butylether	177,8	186,58	7,64	0,44	0,13
7	2-Methylfuran	181,7	195,87	13,43	0,69	0,052
8	Tetrahydrofuran	164,56	175,61	9,31	0,57	0,157
9	Tetrahydropyran	223,9	236,26	10,21	0,64	0,174
10	Epichlorohydrin	247	199,14	-41,7	-2,42	0,129
11	Chloromethylmethylether	169,5	181,97	10,32	0,65	0,172
12	Bis(chloromethyl)ether	231,5	212,49	-13,22	-1,08	0,305
13	Bis(2-chloroethyl)ether	221,1	249,22	26,14	1,38	0,07
14	Bis(2-chloroisopropyl)ether	176	187,46	9,93	0,58	0,134
15	2-Chloroethyl vinylether	203	206,93	3,63	0,19	0,078
16	Bis(2-chloroethoxy)methane	240,7	254,39	12,6	0,67	0,08
17	Anisole	235,87	234,59	-1,19	-0,06	0,068
18	2-Chloroanisole	246,2	266,92	19,61	1,01	0,054
19	3-Chloroanisole	241	271,45	29,16	1,47	0,042
20	4-Chloroanisole	255	267,79	12,13	0,62	0,052
21	2,3-Dichloroanisole	305	293,73	-10,85	-0,54	0,037
22	2,6-Dichloroanisole	283	277,84	-4,91	-0,25	0,049
23	2,4,6-Trichloroanisole	334,5	308,94	-24,25	-1,24	0,051
24	2,3,4,5-Tetrachloroanisole	361	364,16	2,95	0,15	0,066
25	2,3,5,6-Tetrachloroanisole	357	357,08	0,07	0	0,072
26	4,5-Dichloroveratrole	356	336,49	-17,96	-0,96	0,08
27	3,4,5-Trichloroveratrole	339	343,61	4,25	0,23	0,077
28	Tetrachloroveratrole	363	374,05	3,21	0,97	0,71
29	Benzylethylether	275,5	236,29	-35,15	-1,96	0,103
30	Styreneoxide	237,4	250,75	12,57	0,65	0,058
31	Diphenylether	299,87	255,82	-39,92	-2,19	0,094
32	4-Chlorodiphenyl ether	265	276,17	10,27	0,55	0,08
33	2,6-Dichlorodiphenyl ether	312	296,55	-14,62	-0,75	0,054
34	2,4,4'-Trichlorodiphenyl ether	313	315,83	2,71	0,14	0,043
35	2,4,5-Trichlorodiphenyl ether	334	321,77	-11,78	-0,59	0,037
36	2,2',4,4'-Tetrachlorodiphenyl ether	342	338,71	-3,16	-0,16	0,039

Tableau- 07 (suite et fin)

N°	composé	Tf _{Exp}	Tf _{Pred}	e _i	e _{Std}	h _{ii}
37	2,4,4',5-Tetrachlorodiphenyl ether	335	344,75	9,38	0,47	0,038
38	3,3',4,4'-Tetrachlorodiphenyl ether	343	329,64	-12,81	-0,64	0,041
39	2,2',4,4',5-Pentachlorodiphenyl ether	298	360,62	60,14	3,02	0,04
40	2,2',4,4',6-Pentachlorodiphenyl ether	319	346,68	26,29	1,34	0,05
41	2,3,3',4,4' - Pentachlorodiphenylether	338	344,34	6,08	0,31	0,041
42	2,2',3,4,4',5'-Hexachlorodiphenyl Ether	342	366,35	23,22	1,18	0,046
43	2,2',3,4,4',6'-Hexachlorodiphenyl ether	394	352,48	-39,24	-2,02	0,055
44	2,2',4,4',5,5'-Hexachlorodiphenyl ether	388	425,14	30,39	1,94	0,182
45	2,2',4,4',5,6'-Hexachlorodiphenyl ether	369	363,79	-4,93	-0,25	0,054
46	2,3',4,4',5,5'-Hexachlorodiphenyl ether	378	376,23	-1,69	-0,09	0,047
47	2,2',3,4,4',5,5'-Heptachlorodiphenyl ether	362	394,79	30,19	1,61	0,079
48	2,2',3,4,4',6,6'-Heptachlorodiphenyl ether	415	372,77	-38,02	-2,1	0,1
49	2,2',3,3',4,4',5,5'-Octachlorodiphenyl ether	399	402,05	2,73	0,15	0,103
50	2,2',3,3',4,4',6,6'-Octachlorodiphenyl ether	399	391,36	-6,75	-0,38	0,116
51	4-Bromodiphenyl ether	291	284,83	-5,7	-0,3	0,075
52	4,4'-Dibromodiphenyl Ether	330,7	307,45	-21,61	-1,14	0,07
53	2,4,4'-Tribromodiphenyl ether	337,25	340,2	2,77	0,14	0,063
54	3,3',4,4'-Tetrabromodiphenyl ether	371	357,21	-12,23	-0,69	0,113
55	2,2',3,4,4'-Pentabromodiphenyl ether	393	378,24	-13,22	-0,74	0,104
56	2,2',4,4',5-Pentabromodiphenyl ether	365,5	397,44	28,41	1,6	0,11
57	2,2',4,4',6-Pentabromodiphenyl ether	383	368,37	-13,85	-0,71	0,053
58	2,2',4,4',5,5'-Hexabromodiphenyl Ether	434,5	419,14	-12,94	-0,79	0,158
59	2,2',4,4',5,6'-Hexabromodiphenyl ether	405	397,38	-6,95	-0,38	0,088
60	2,2',3,4,5,5',6-Heptabromodiphenyl ether	445	426,17	-16,1	-0,96	0,145

L'analyse des résidus fait ressortir, des erreurs absolues moyennes égales à **15.62**.

La colonne 07 (tableau08) donne les valeurs de h_{ii} , $i^{\text{ème}}$ 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

$$h^* = \frac{3p}{n} = 3 \times 6 / 60 = 0.3 \quad (12)$$

On constate que tous les h_{ii} sont inférieures à cette valeur critique à l'exception des composés 27 et 80.

Toutes les valeurs des paramètres statistiques de calibration sont regroupées dans le tableau 09.

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

N	60
R²	93,85
Q²	92,56
F	164,68
S	21,1639
SDEC	20,0778
SDEP	22,0817

Pour la robustesse du modèle est assurés par la valeur de $Q^2_{LOO} > 99$ alors que les valeurs de l'erreur quadratique moyenne de prédiction et de calcul sont petites et proches ; en plus ce modèle est significatif avec une grande valeur du paramètre de Fisher : ($F = 164,68$).

Le domaine d'application a été discuté à l'aide du diagramme de Williams (figure 9) qui représente les résidus de prédiction standardisés en fonction des valeurs des leviers (h_i).

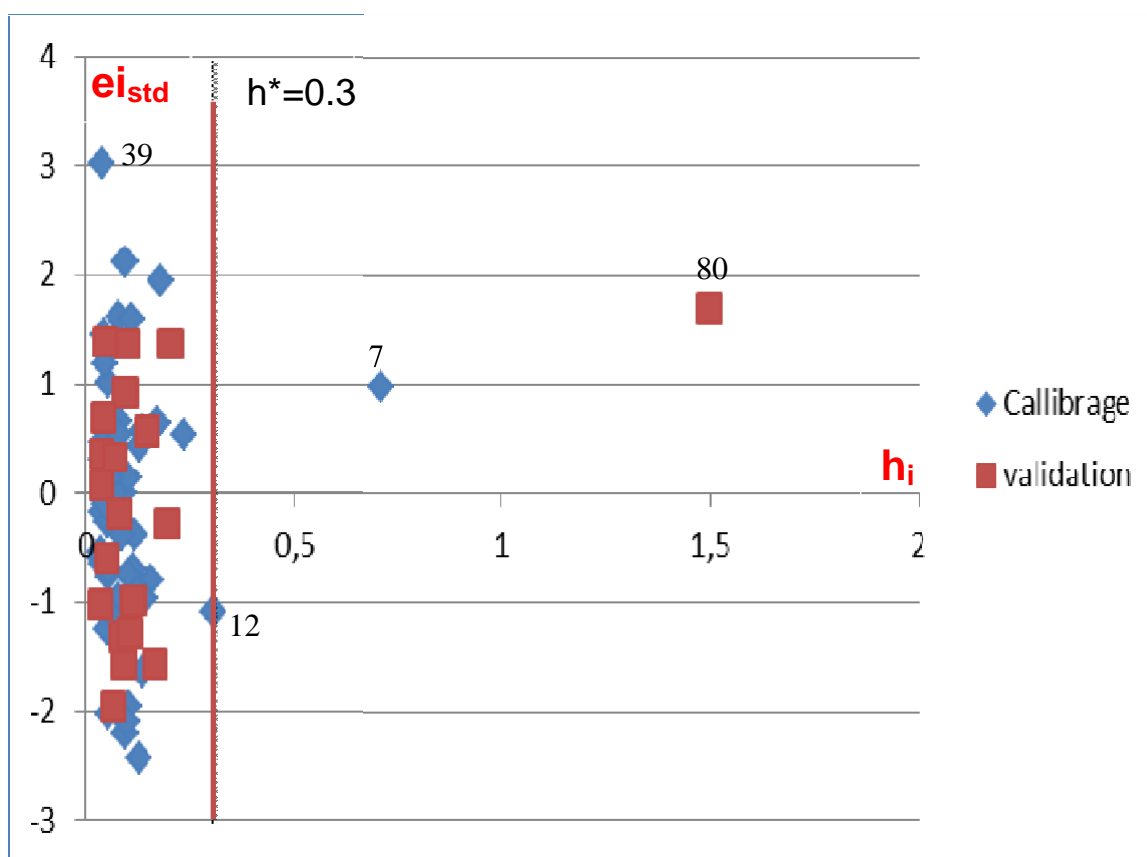
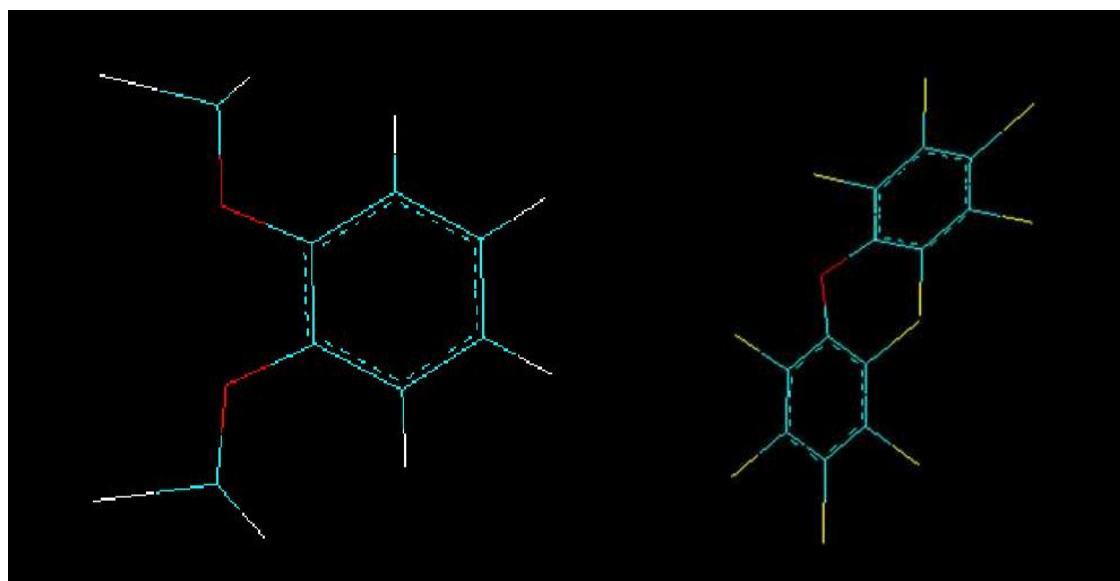


Figure 09- Diagramme de Williams

Le diagramme de Williams fait ressortir deux points influents ce sont les composés 80 de l'ensemble de validation et 27 de l'ensemble de calibration, également il faut signaler qu'il y a deux composés se trouvent juste à la limite il s'agit du composé 12 ($h_i = 0.305$) et le composé 39 avec un résidu standardisé égal à +3.02.



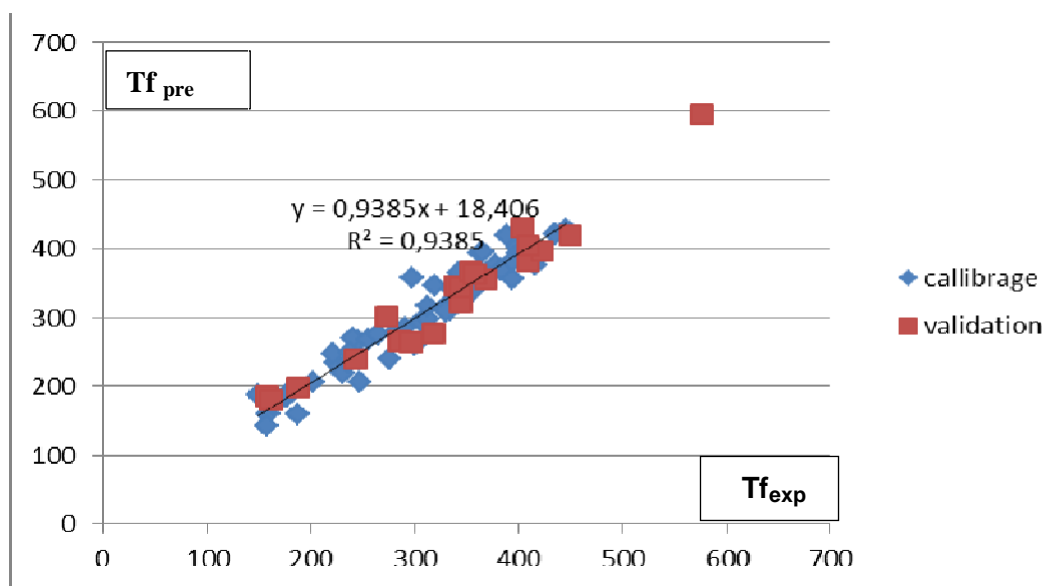
(Tetrachloroveratrole)

(Decabromodiphenylether)

Figure 10- Les 02 composés influents

V-7- Vérification de la qualité de l'ajustement :

La qualité de l'ajustement a été vérifiée en représentant les valeurs calculées de Tf avec notre modèle en fonction des celles observées ou expérimentales (colonnes (3) et (4) tableau 08). La figure 12 montre un bon ajustement traduit par une faible dispersion autour de la droite de régression.

*Figure 11- Graphe des valeurs Tf calculées en fonction des valeurs expérimentales.*

V-8-Test de randomisation :

Dans le but d'établir que le modèle obtenu n'est pas dû au hasard ou à une sur spécification, nous avons appliqué le test de randomisation. Ainsi 100 nouveaux vecteurs de la température de fusion ont été générés par permutation des positions des composantes du vecteur réel. La figure 12 qui représente le graphe des coefficients statistiques Q^2 et R^2 permet de comparer les résultats obtenus pour les modèles randomisés (carrés orange) au modèle réel de départ (carré vert).

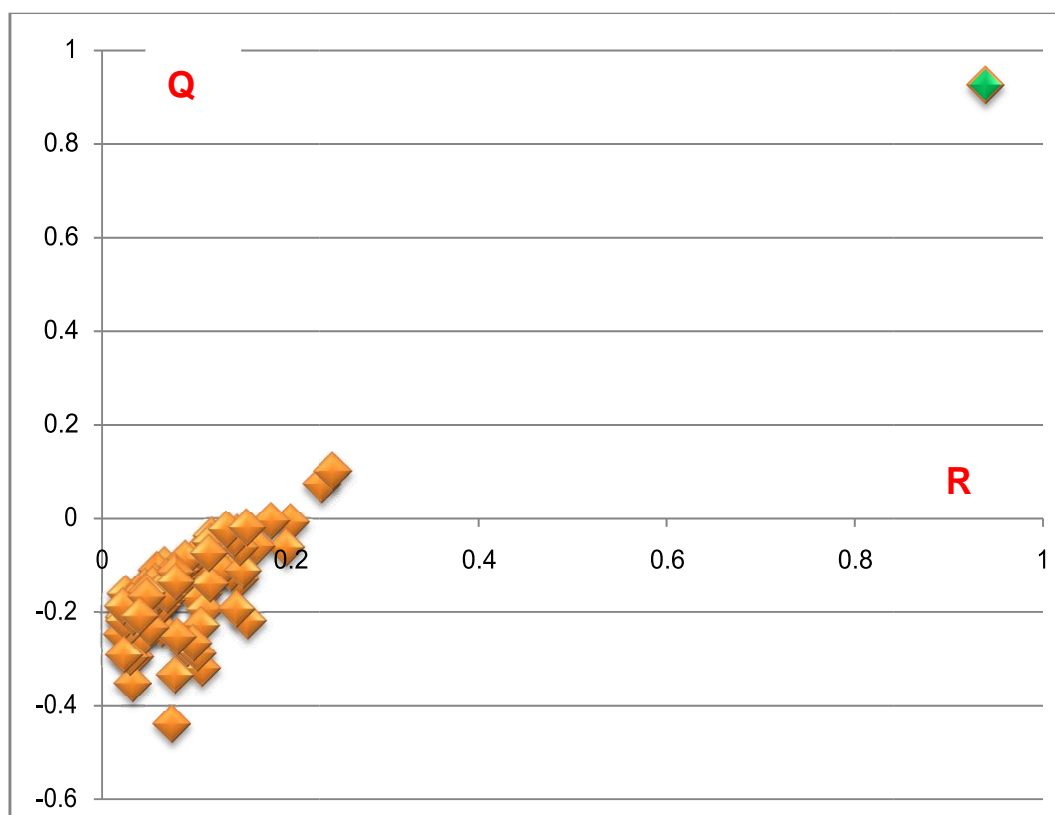


Figure 12- Test de randomisation associé au modèle QSPR

Les carrés orange représentent la température de fusion ordonnée de façon aléatoire, et le carré vert 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/température réelle a été très bien établie.

V-9-Validation externe

Pour généraliser le modèle choisi on procède à une validation externe sur les 20 composés choisis aléatoirement et qui ne font pas partie de l'ensemble d'essai. Les résultats obtenus (tableau 10) 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 de fusion prédite et celles observée avec une dispersion autre de la droite d'ajustement (figure 12).

Tableau -09 Valeurs des T_f expérimentales, calculées, e_i , $e_{i\text{std}}$ et h_{ii} pour l'ensemble de validation

N°	Composé	$T_{f\text{Exp}}$	$T_{f\text{pred}}$	e_i	$e_{i\text{std}}$	h_{ii}
61	Di-n-propylether	158,2	185,88	27,68	1,38	0,099
62	1,2 propyleneoxide	161,1	179,89	18,79	0,93	0,096
63	Furan	187,39	198,45	11,06	0,57	0,147
64	1,4-Dioxane	284,95	265,21	-19,74	-0,99	0,114
65	2, 3,4-Trichloroanisole	343	322,06	-20,94	-1,01	0,035
66	Veratrole	295,5	264,12	-31,38	-1,56	0,092
67	Phenetole	243,57	240,02	-3,55	-0,18	0,08
68	2-Chlorodiphenyl ether	318	277,91	-40,09	-1,96	0,066
69	2,4-Dichlorodiphenyl ether	273	301,64	28,64	1,39	0,048
70	2,2', 3, 4,4'-Pentachlorodiphenyl ether	339	346,73	7,73	0,37	0,039
71	3,3', 4,4',5-Pentachlorodiphenyl éther	343	344,67	1,67	0,08	0,041
72	2,2',3,3',4,4'-Hexachlorodiphenyl éther	368	356,01	-11,99	-0,58	0,05
73	2,2', 3, 4,4',5-Hexachlorodiphenyl éther	353	367,42	14,42	0,7	0,045
74	2,2',3,4,4',5,6'-Heptachlorodiphenyl ether	409	382,24	-26,76	-1,32	0,087
75	2,2', 3,3', 4,4', 5,6'-Octachlorodiphenyl éther	422	396,43	-25,57	-1,28	0,106
76	2,2', 3,3', 4,4', 5,5',6-Nonachlorodiphenyl éther	450	419,84	-30,16	-1,56	0,165
77	Decachlorodiphenyl éther	404	430,07	26,07	1,38	0,204
78	2,2', 4,4'-Tetrabromodiphenyl éther	357	363,79	6,79	0,33	0,07
79	2,3,3',4,4',5'-Hexabromodiphenyl ether	409	403,97	-5,03	-0,27	0,195
80	Decabromodiphenyl éther	575,5	595,09	19,59	1,7	1,497

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	20
Q^2_{ext}	92,88
$SDEP_{ext}$	22,0817

La valeur de Q^2_{ext} est proche ou même supérieure à celle de la prédiction interne ce qui confirme que le modèle a une bonne capacité prédictive, également pour le $SDEP_{ext}$ qui a une valeur proche à celle de SDEP.



CONCLUSION
GENERALE



CONCLUSION GENERALE

La méthodologie QSPR a été utilisée pour relier la température de fusion, d'un mélange de composés d'éthers comportant des cycles aromatiques et des chaînes ouvertes, à des descripteurs moléculaires théoriques calculés à l'aide de logiciels spécialisés.

Les 80 composés ont été divisés en deux groupes comprenant 60 pour la calibration et le reste pour la validation externe.

La taille du modèle a été fixée à 5 descripteurs par la valeur optimale de R^2 et la sélection des variables explicatives a été réalisée par algorithme génétique dans le logiciel MOBY DIGS en maximisant la valeurs du coefficient de prédiction Q^2_{LOO} .

Les statistiques obtenues nous a permet de s'assurer de la qualité d'ajustement, de la robustesse interne et externe, des capacités prédictives et de la possibilité d'extension suffisante le modèle choisi.

Ainsi la température de fusion peut être prédite à partir de la structure ou/et la géométrie de la molécule par un modèle linéaire.

Enfin on doit investiguer la ou les causes possibles des aberrations relevées pour notre modèle, si c'est un problème d'optimisation alors il faut chercher d'autres méthodes ou ce sont des erreurs d'expérimentateur.

Ce travail peut être étendu à un nombre plus important de composés et le choix ou l'éclatement des données en deux ensembles disjoints (calibration et validation) pourrait se faire d'une manière plus réfléchie, en utilisant d'autres méthodes non-linéaires à savoir les réseaux de neurones artificiels ou le SVM.



RÉFÉRENCES

BIBLIOGRAPHIQUES



REFERENCES BIBLIOGRAPHIQUES

N°	REFERENCES
1	A.R. Katritzky., D.C. Fara., R.O. Petrukhin, (2002), top. Med. Chemm 1333-1356.
2	C. Hansch, 1969, a quantitative approach to biochemical structure activity relationships. Accounts of chemicalresearch, 2, 232-239.
3	R. Leardi , 2001 Chemometr. 15, 559-569.
4	V. Prana, P. Rotureau, G. Fayet, D. André, S. Hub, P. Vicot, L. Rao, C. Adamo, predction of the thermal decomposition of organic peroxides by validated QSPR models, journal of hazardous materials, 2014, 276, 216-224.
5	L. Fisbbein, 1973. Chromatography of environmental hazards, Vol. II. Elsevier,Amsterdam
6	Searching scientific literature directly with Chem Draw v14 Chemistry World .29 July 2014.
7	Allen M. P., TildesleyD.J.,Computer stimulation of liquids .Oxford .1987.
8	Hyperchem™ Release 7.5 for windows, molecular modelling system, 2000.
9	R. Todeschini, V. Consonni, M. Pawan, DRAGON, Software for the calculation of molecular descriptors. Release 5.3 for windows, Milano, 2005.
10	B .Hoggas, Ch.Amamiche , modélisation de la température d'ébullition des alcanes en utilisant une approche QSPR, mémoire de master (L.M.D), chimie analytique et environnement ,2016 :p8.
11	Thomas-Danguin T. (1997).Intensité olfactive des composés purs et demélanges:application au masquage des odeurs, Université ClaudeBernard,Lyon,p224.
12	S .Remache ,W.Redah , Etude QSRR de la rétention chromtographique des HAP, mémoire de master (L.M.D), chimie analytique et environnement , 2018 :p 20.
13	Dragon_ Aide blocs des descripteurs.
14	Le Cloirec P. (2002). Introduction au traitement de l'air, Les techniques de l'ingénieur traité environnement (G 1700): 1-8.
15	S. Khaled. 2009.Contribution l'étude de la relation structure chimique- odeur utilisation de la technique random forest (application à la famille des pyrazines), mémoire de magister. universite kasdi merbah ourgla.p30.
16	AI ACCESS, 91940, Les ulis, Franonce.
17	M. karelson,2000. Molecular descriptors in QSAR/QSPR. Wiley- Interscience, p. 385
18	B. Kowalski, R. Gerlach, H.Wold,1982. Systems under indirect observation (K. Jorenskog et H. Wold, eds.), North Holland, Amsterdam, 191-206.

- 19 L. Erikson, E. Johannson, N. Kettaneh- Wold,2001.Multi and megavariate data analysis- principles and applications. umetricsacademy, ume.
- 20 S. Wold, A. Ruhe, H.Wold, W. Dunn,1984. SIAMJ. Sci. Stat. Comput., 5, 735.
- 21 S. Wold,1984. Chemometrics: mathematics and statistics in chemistry. reidel, dordrecht, the netherlands.
- 22 P. Gelada, B. R. Kowalski, Anal. 1986. Chim. Acta, 185, 1.
- 23 A. R. Katritzky, V. S. Lobanov, M. Karelson,1994. Codessa reference manual. university of florida, gainesville.
- 24 V. Y. Nalimov, 1962. The application of mathematical statistics to chemical analysis, addison- wesley, reading, ma.
- 25 R. Calcutt, R. Body,1983. Statistics for analytical chemists. champman& Hall, New York
- 26 J. C. Miller, J. N. Miller,1988. Statistics for analytical chemistry. Ellis Horwood, New York .
- 27 P. C. Meier, R. E. Zund,1993. Statistical methods in analytical chemistry. Wiley, New York .
- 28 P. Dagnélie,1998. Statistique théorique et appliquée. Tomes 1 et 2. De Boeck &Larcier s. a.
- 29 R. Tomassone, E. Lesquoy, C. Miller,1983. La régression : nouveaux regards sur une ancienne méthode statistique. masson, INRA .variables. ecology 89(9): 2623-2632.
- 30 L. Chambers. Practical handbook of genetic algorithms. Lewis Publishing (1995).
- 31 MINITAB Release 16.2.0.0 for microsoft langagepack 2.
- 32 R. Todeschini. MOBY DIGS software for multilinear regression analysis and variable subset selection by genetic algorithm.release for windows. Milano Srl.
- 33 n.r draper, h. smith, applied regression analysis, third edition, wiley series in probability and statistics, New York, 1998.
- 34 L. Eriksson, J. Jaworska, A.P. Worth, M.T.D. Perspective, 111(10), (2003), 1361-1375.

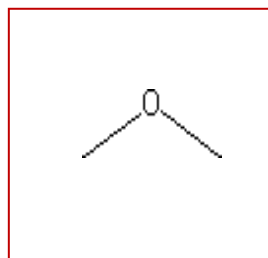


ANNEXE

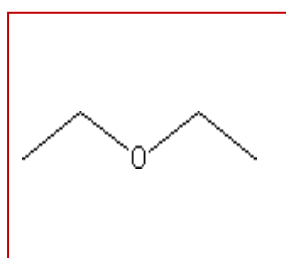


ANNEXE

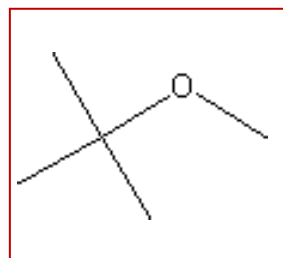
Les structures chimiques des 80 éthers étudiés :



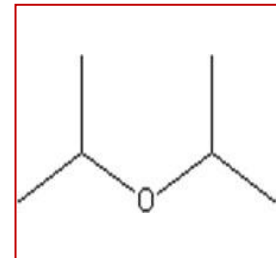
Diméthyle éther



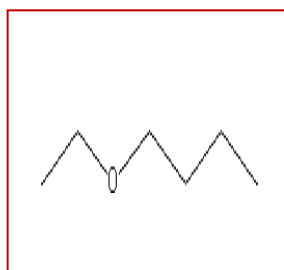
Diéthyl éther



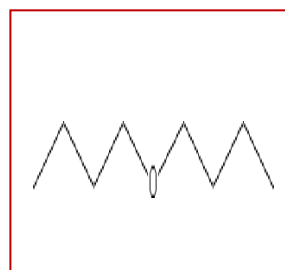
Methyl t-butyl ether



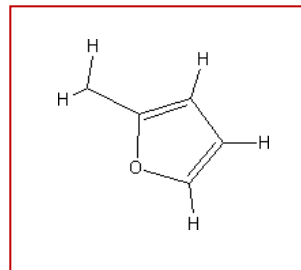
Di-isopropyl ether



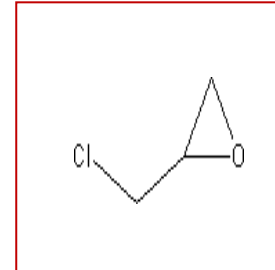
Butyl ethyl ether



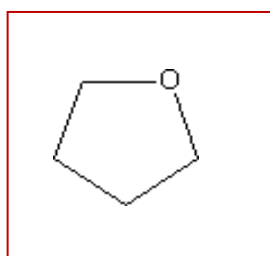
Di-n-butyl ether



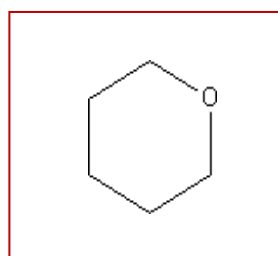
2-Methylfuran



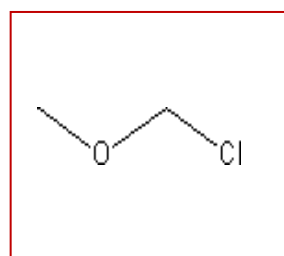
Epichlorohydrin



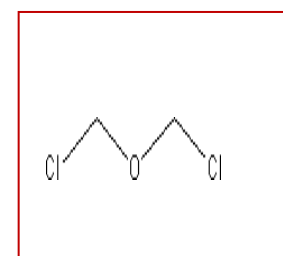
Tetrahydrofuran



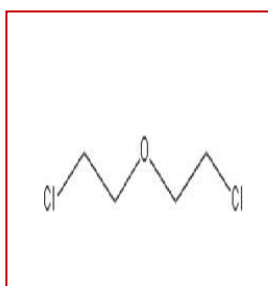
Tetrahydropyran



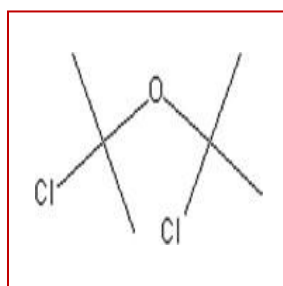
Chloromethyl methyl ether



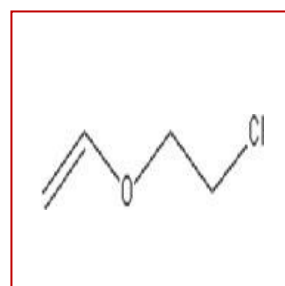
Bis(chloromethyl)ether



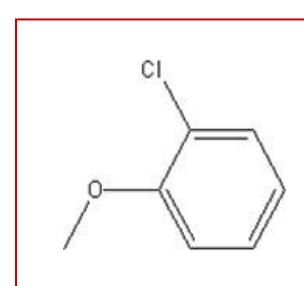
Bis(2-chloroethyl)ether



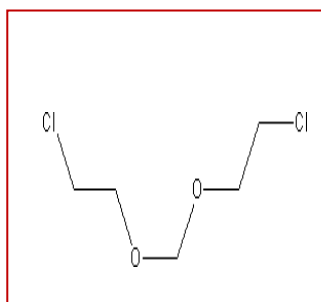
Bis(2-chloroisopropyl)ether



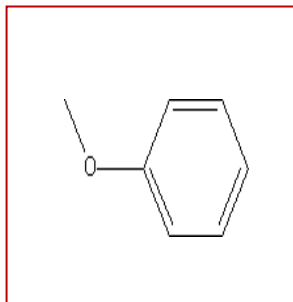
2-Chloroethyl vinyl ether



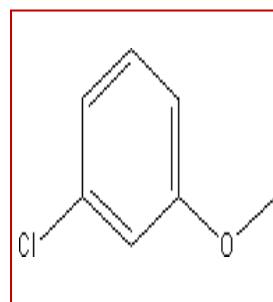
2-Chloroanisole



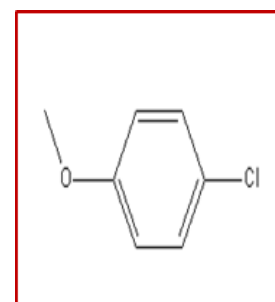
Bis(2-chloroethoxy)methane



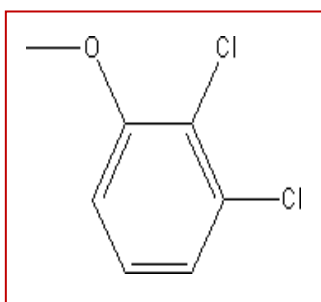
Anisole



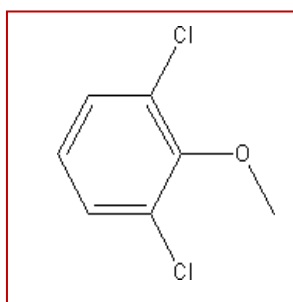
3-Chloroanisole



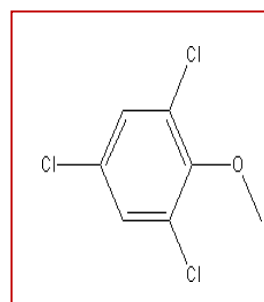
4-Chloroanisole



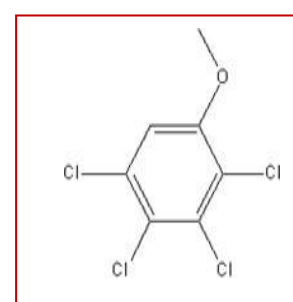
2,3-Dichloroanisole



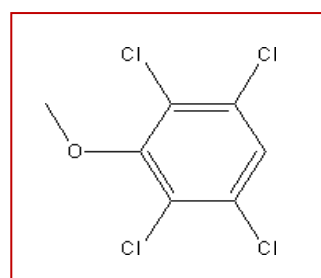
2,6-Dichloroanisole



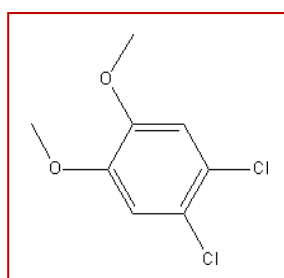
2,4,6-Trichloroanisole



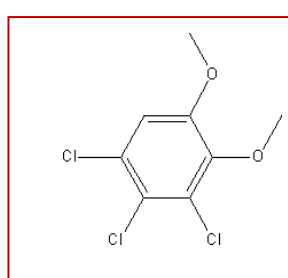
2,3,4,5-Tetrachloroanisole



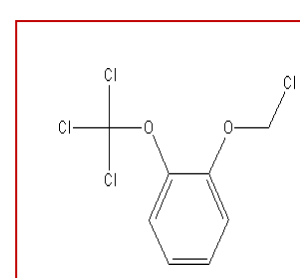
2,3,5,6-Tetrachloroanisole



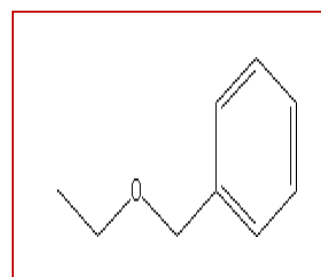
4,5-Dichloroveratrole



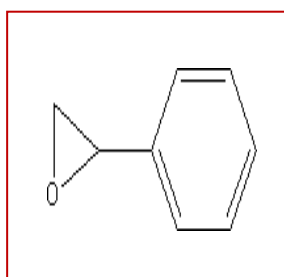
3,4,5-Trichloroveratrole



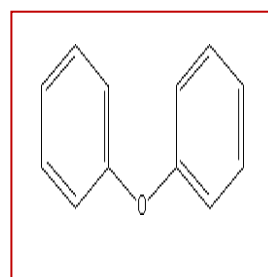
Tetrachloroveratrole



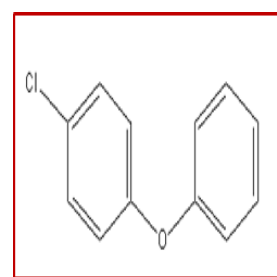
Benzyl ethyl ether



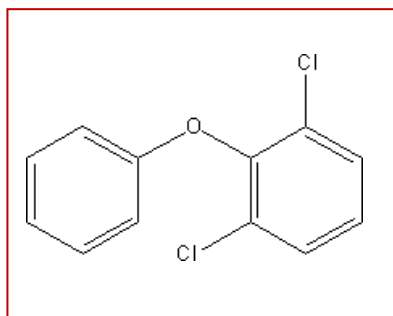
Styrene oxide



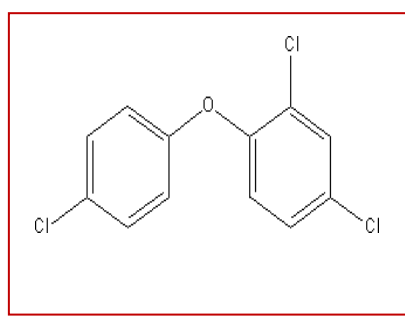
Diphenyl ether



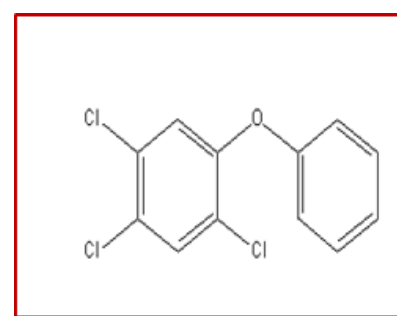
4-Chlorodiphenyl ether



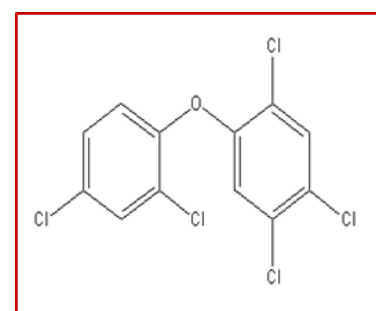
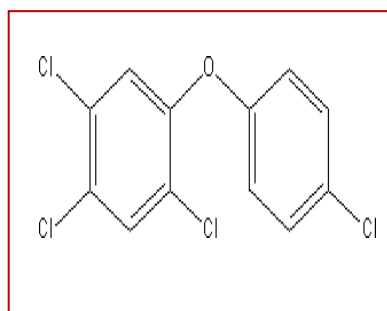
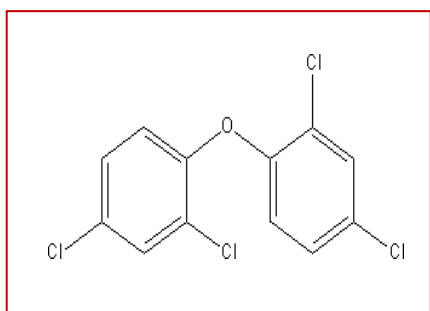
2,6-Dichlorodiphenyl ether



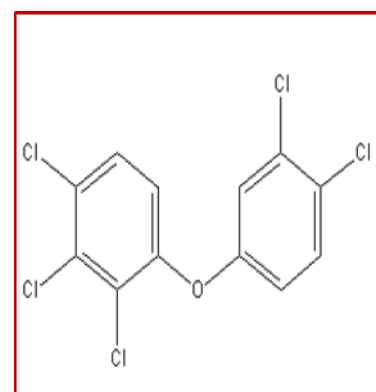
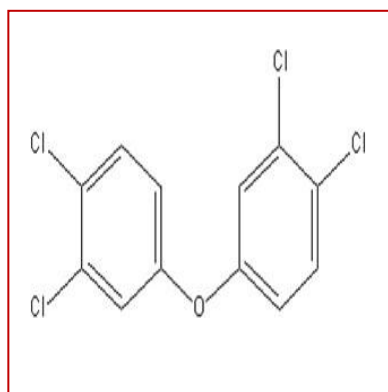
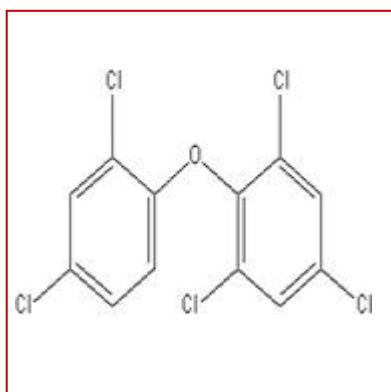
2,4,4'-Trichlorodiphenyl



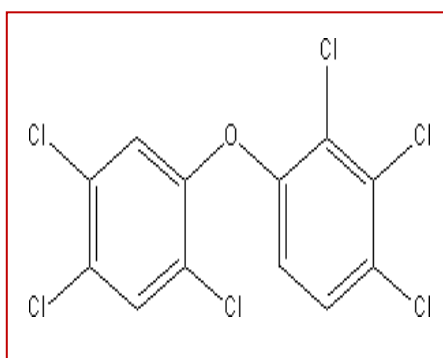
2,4,5-Trichlorodiphenyl ether



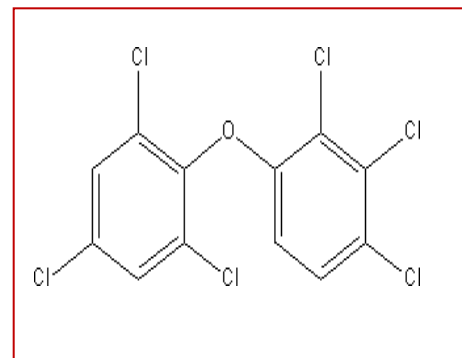
2,2',4,4'-Tetrachlorodiphenyl ether 2,4,4',5-Tetrachlorodiphenyl ether 2,2',4,4',5- Pentachlorodiphenyl ether



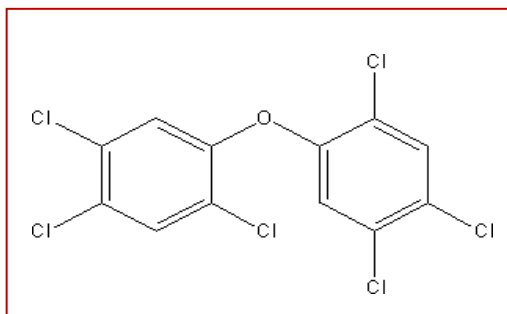
2,2',4,4',6-Pentachlorodiphenyl ether 3,3',4,4'-Tetrachlorodiphenyl ether 2,3,3',4,4' - Pentachlorodiphenyl ether



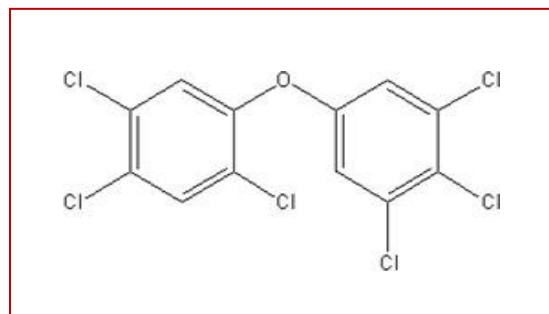
2,2',3,4,4',5'-Hexachlorodiphenyl ether



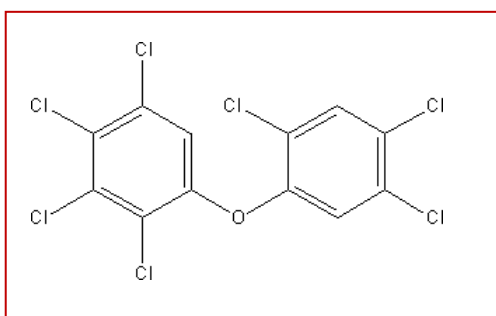
2,2',3,4,4',6'-Hexachlorodiphenyl ether



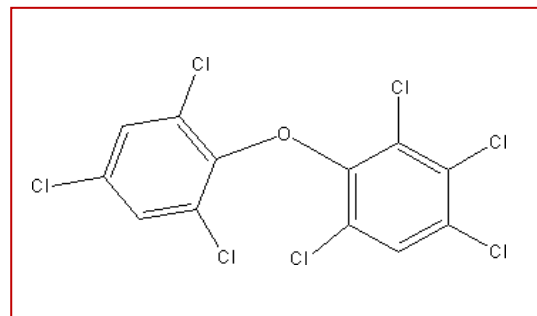
2,2',4,4',5,5'-Hexachlorodiphenyl ether



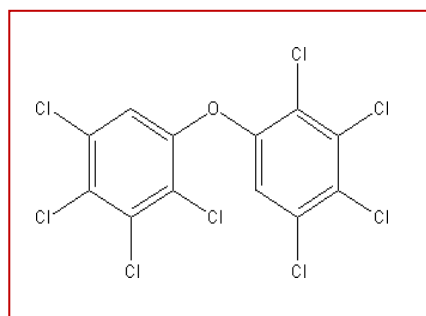
2,3',4,4',5,5'-Hexachlorodiphenyl ether



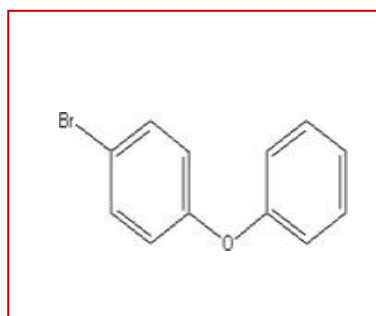
2,2',3,4,4',5,5'-Heptachlorodiphenyl ether



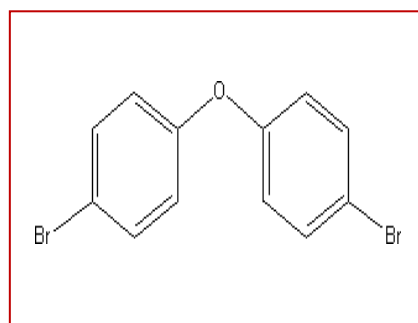
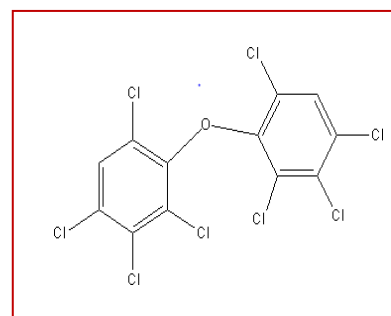
2,2',3,4,4',6,6'-Heptachlorodiphenyl ether



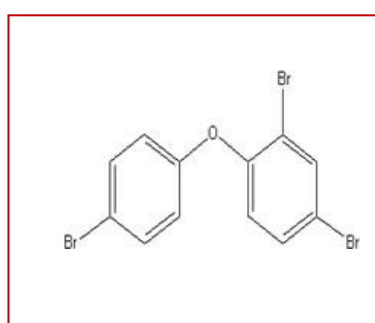
2,2',3,3',4,4',5,5'-Octachlorodiphenyl ether



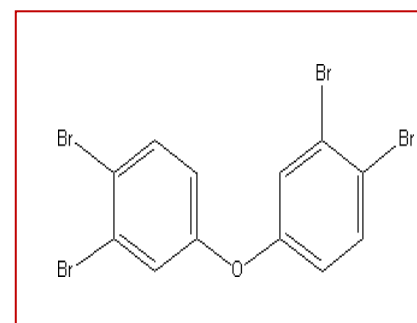
2,2',3,3',4,4',6,6'-Octachlorodiphenyl ether



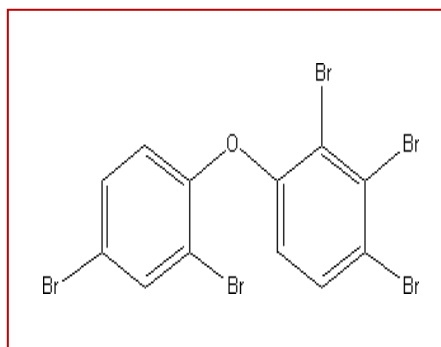
4,4'-Dibromodiphenyl Ether



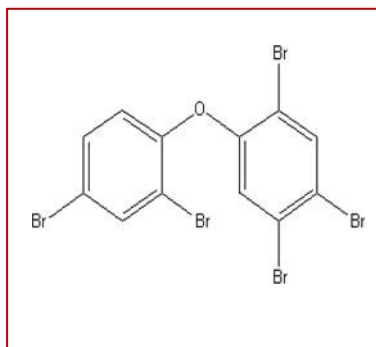
2,4,4'-Tribromodiphenyl ether



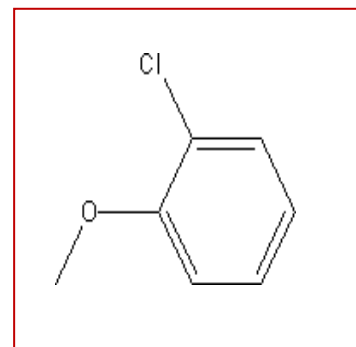
3,3',4,4'-Tetrabromodiphenyl ether



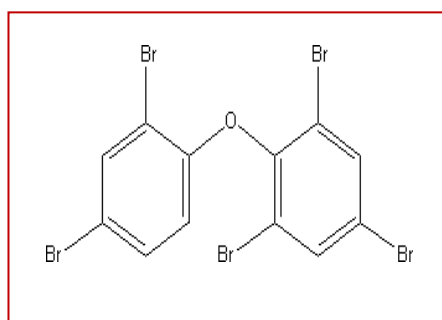
2,2',3,4,4'-Pentabromodiphenyl ether



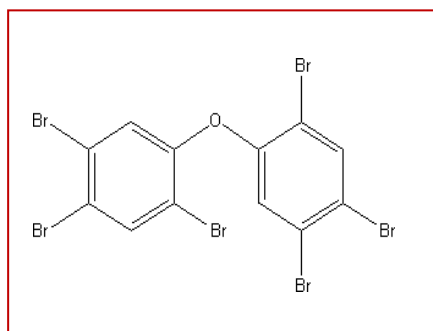
2,2',4,4',5-Pentabromodiphenyl ether



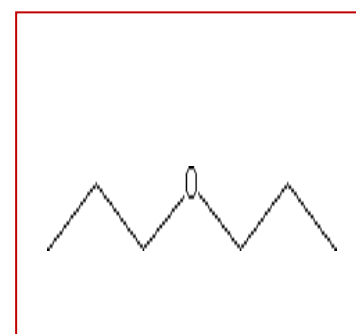
2-Chloroanisole



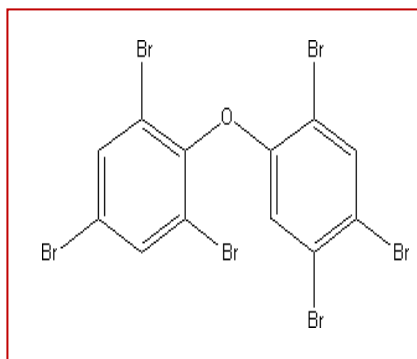
2,2',4,4',5-Pentabromodiphenylether



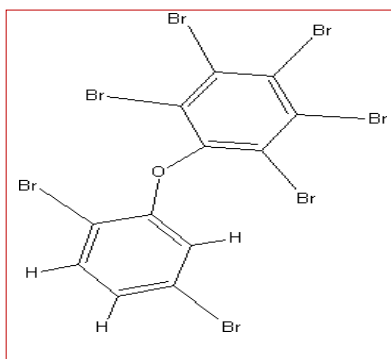
2,2',4,4',5,5'-Hexabromodiphenylether



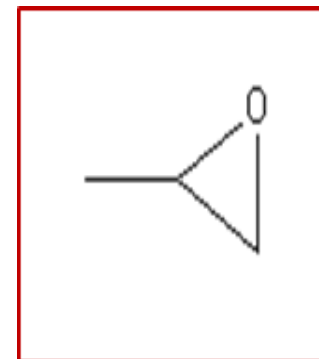
Di-n-propyl ether



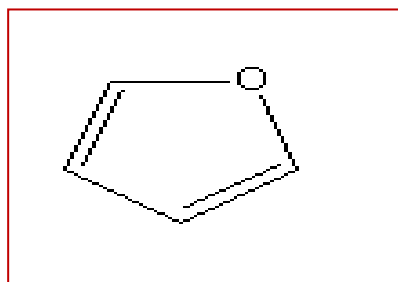
2,2',4,4',5,6'-Hexabromodiphenyl ether



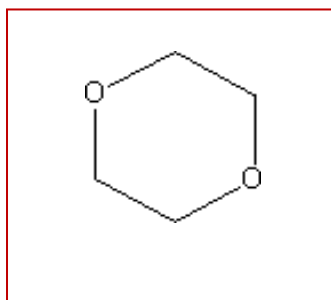
2',3,4,5,5',6'-Heptabromodiphenyl ether



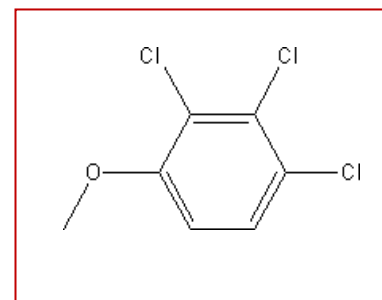
1,2 propylene oxide



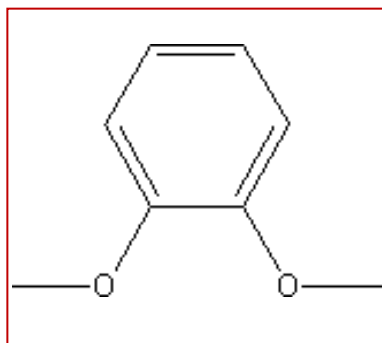
Furan



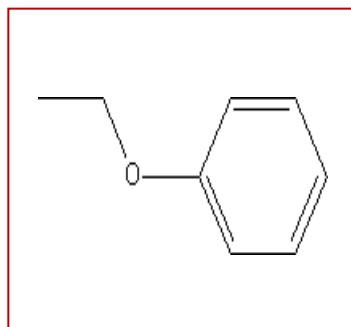
1,4-Dioxane



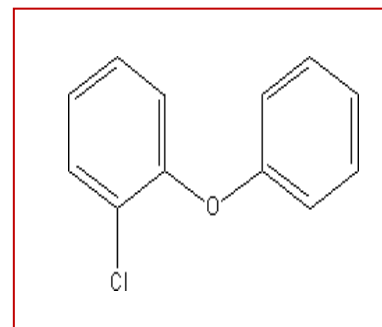
2,3,4-Trichloroanisole



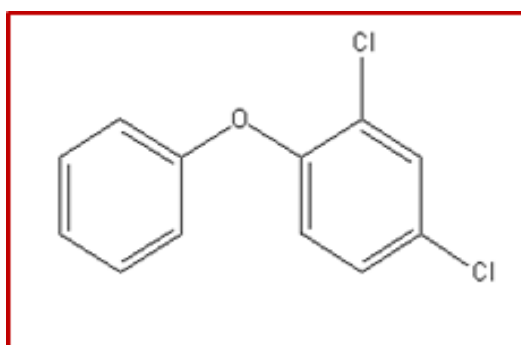
Veratrol



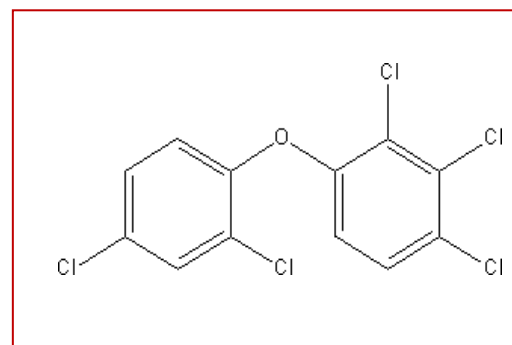
Phenetole



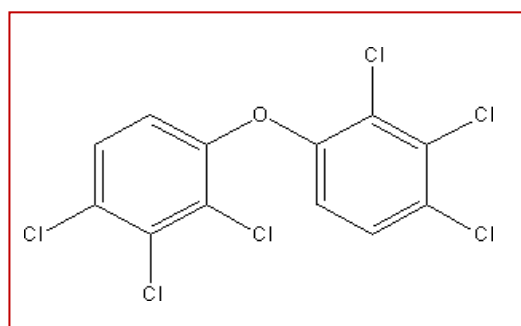
2-Chlorodiphenyl ether



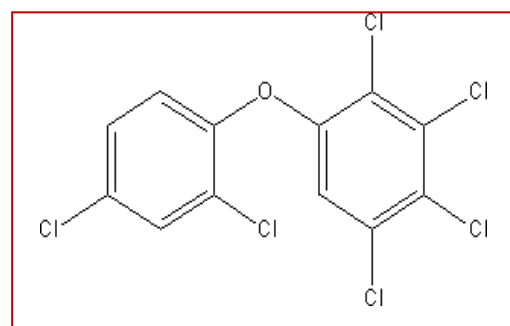
2,4-Dichlorodiphenyl ether



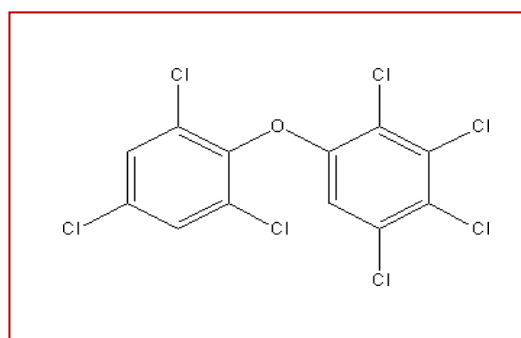
2,2',3,4,4'-Pentachlorodiphenyl ether



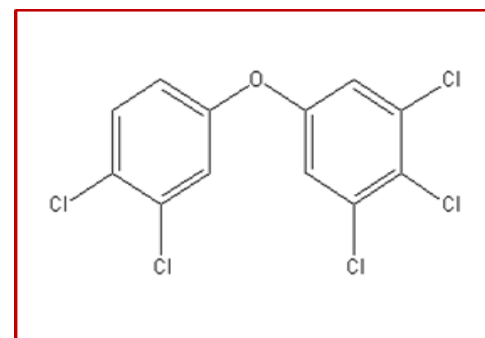
2,2',3,3',4,4'-Hexachlorodiphenyl ether



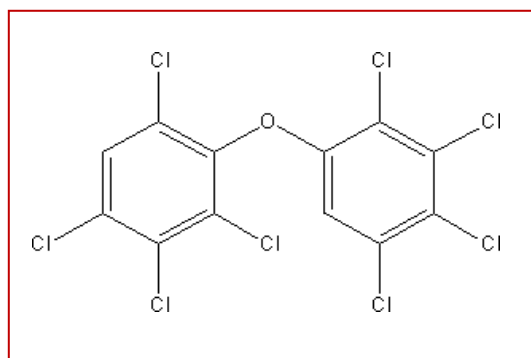
2,2',3,4,4',5-Hexachlorodiphenyl Ether



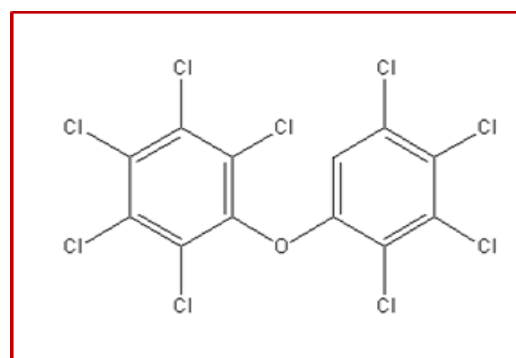
2,2',3,4,4',5,6'-Heptachlorodiphenyl ether



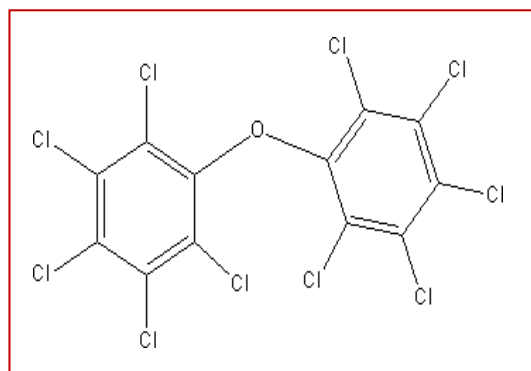
3,3',4,4',5-Pentachlorodiphenyl ether



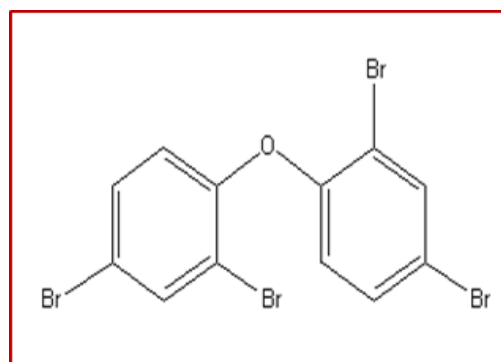
2,2',3,3',4,4',5,6'-Octachlorodiphenyl Ether



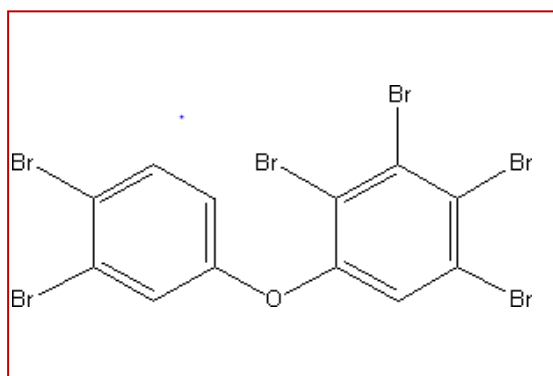
2,2',3,3',4,4',5,5',6-Nonachlorodiphenyl ether



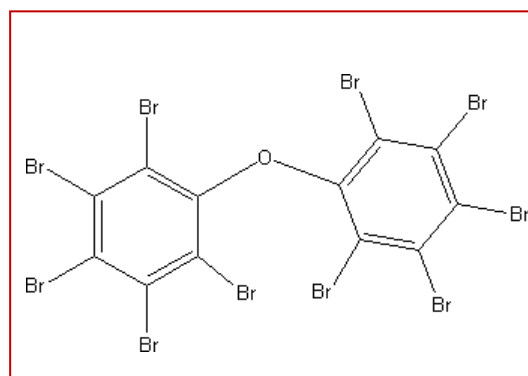
Decachlorodiphenyl ether



2,2',4,4'-Tetrabromodiphenyl ether



2,3,3',4,4',5'-Hexabromodiphenyl ether



Decabromodiphenyl ether

Résumé

Un modèle QSPR a été développé pour la prédiction de la température de fusion d'une série d'éthers. Les 80 données, ont été séparées en deux sous-ensembles disjoints comprenant respectivement 60 éléments pour le calcul et le test (éventuel) du modèle, et 20 éléments pour sa validation externe.

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

Les valeurs des paramètres statistiques (R^2 , Q^2 , SDEC, SDEP, $SDEP_{ext}$) obtenues attestent de la pertinence du modèle développé.

Mots-clés:

Ethers – Température de fusion – Descripteurs moléculaires théoriques – Algorithme génétique.

Abstract

A QSPR model has been developed for the prediction of the boiling temperature of a series of ethers. The 80 objects were separated into two disjoint subsets comprising respectively 60 elements for the calculation and the test of the model, and 20 elements for its external validation.

Theoretical molecular descriptors were calculated using commercial molecular modeling software. The size of the model was determined by optimizing the value of R^2 , and the selection of descriptors carried out by genetic algorithm.

The values of the statistical parameters (R^2 , Q^2 , SDEC, SDEP, $SDEP_{ext}$) obtained attest to the relevance of the developed model.

Key words:

Ethers - melting point - Theoretical molecular descriptors - Genetic algorithm.

ملخص

تم تطوير نموذج بطريقة ال QSPR لتنبؤ درجة الذوبان لمجموعة من الاثيريات. ال 80 مركب تم تقسيمها الى مجموعتين الاولى تحتوي على 60 عنصر لحساب وتجريب النموذج اما الثانية تحتوي على 20 عنصر لتصديق الخارجي للنموذج. المواصفات الجزيئية النظرية تم حسابها باستعمال برمجيات النمذجة الجزيئية المتوفرة في السوق حجم النموذج تم تحديده عن طريق القيمة المثالية ل R^2 ، أما إختيار المواصفات عن طريق الخوارزمية المورثية. قيم المعالم الاحصائية ($R^2, Q^2, SDEC, SDEP, SDEP_{EXT}$) المتحصل عليها تؤكد تعلق النموذج المطورة.

الكلمات الدالة:

اثيرات – درجة الذوبان- مواصفات جزيئية نظرية- الخوارزمية المورثية.