



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



Département des Sciences de la Matière

N° de série :.....

Mémoire de fin d'études

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

Filière : Chimie.

Spécialité : Chimie Analytique et Environnement

Intitulé:

*Modélisation de la température de fusion
d'une série d'alcool.*

Réalisé par :

-Ben hbessa Ahlam.

Dirigé par : KERTIOU NOUREDDINE

Présente le : 30 /06/2019

Membres de jury :

-Bouakkadia Amel MCB.

-Samiei Salima .

Année universitaire 2018/2019

Dédicace

*Je dédie ce modeste travail A mes très chers parents et ma
grande famille.*

*A tous les enseignants de département des Sciences de la
Matière à l'université de kenchela*

A mes très chers frères.

*A mes amis, NABILA, CHAHRA ZAD, RABAB, NADIA
SAMIA, FATMA*

Tous mes amis.

A tous qui m'aiment et que j'aime.

A vous.

Remerciements

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

*J'exprime mes remerciements à mon encadreur Dr: **KERTIOU NOUREDDINE** pour la grande bienveillance qu'elle m'a accordé et pour leur efficaces et précieuses aides, tout le long d'achèvement de ce mémoire.*

Je remercie également tous les membres du jury d'avoir accepté de jurer ce travail.

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

Un remerciement A toutes et à tous qui, de loin ou de près, ont contribué à la réalisation de ce mémoire.

Merci....

**Modélisation de la
température de fusion
d'une série d'alcool.**

SOMMAIRE

Dédicaces et remerciements	-
Liste des figures	A
Liste des tableaux	B
Liste des équations	C
Symboles et abréviations	D
Introduction Générale	1
Partie Théorique.	
I. Collecte des données	3
II. Optimisation de la géométrie moléculaire et génération du descripteur moléculaire.	4
II.1 .Préparation de base des données	4
II.1.1.Logiciels « ChemDraw 7 »	4
II.1.2. Logiciels « HyperChem »	4
II.2. Récupération et stabilisation des molécules	5
II.2.1. Stabilisation structure des molécules (minimisation de l'énergie)	5
II.2.2 .Mécanique moléculaire	5
II.2.3. Récupération des fichiers	6
II .3. Le Logiciel « Dragon »	6
III . Les relations structure-propriété/activité quantitatives	7
III.1.Introduction	7
14524III.2. Généralités sur la modélisation QSPR/QSAR	7

III.3. Principe des méthodes QSPR/QSAR	8
III.4. Méthodologie générale d'une étude QSPR/QSAR	9
III.4.1. Bases de données	10
III.4.2. Méthodes d'analyse des données	11
III.4.3. Descripteurs moléculaires	11
III.4.3.1. Définition moléculaire.	11
III.4.3.2. Les descripteurs moléculaires	11
III.4.3.3. Exemple : calcul d'ATS KW	12
III.4.3.4. Importance des descripteurs	14
IV. Diagramme montre les étapes de prédiction	15
V. L'objectif de la prédiction	15
VI. Méthodes de régressions linéaires et multilinéaire	17
VI.1. Aperçu général	17
VI.2. Evaluation préliminaire des données	17
VI.3. Régression linéaire multiple	18
VI.4. Méthodes de sélection des descripteurs	19
VI.4.1. Régression pas à pas	19
VI.4.2. Algorithme génétique	21
VI.5. Paramètres d'évaluation de la qualité de l'ajustement	21
VI.6. Facteur d'inflation de la variance [FIV]	22
VI.7. Test de randomisation	22
VI.8. Validation externe	23

VII .Applications des méthodes QSPR/QSAR	24
VII. Les alcools.	25
VII.1. Définition Alcool	25
VII.2. Etude chimique	25
VII.2.1 Nomenclature	25
VII.2.2. Classification	26
VII.2.3. Utilisation	27
VII.2.4. Propriétés physico-chimiques	27
IX. La température de fusion	29
Partie Application	
I. Calcul du modèle	32
I .1. Choix du modèle	32
I .2. Choix du famille de descripteurs	33
I .3.Evaluation préliminaire des données	34
I.4. Calcul des corrélations entre les différents descripteurs	36
I.5.Analyse de régression	36
I .6.Diagramme de Williams	38
I.7. Evaluation de la qualité de l'ajustement	39
I.8.Tests de randomisations	40
Conclusion générale	41
Références Bibliographiques	42
Résumé	-

LISTE DES FIGURES

N°	FIGURES	PAGES
Figure-01	Représentation des molécules par Chem Draw	04
Figure-02	applique le Field Force à Composé	06
Figure-03	Procédure d'obtention et de validation d'un modèle QSPR/QSAR	10
Figure-04	Représentation des blocs des descripteurs moléculaire.	14
Figure-05	Diagramme de prédiction par QSPR	15
Figure-06	le cycle de prédiction	16
Figure-07	Diagramme de notre travail	16
Figure-08	Tf (Température de fusion)	30
Figure-09	Test de normalité pour la propriété (Tf) et les trois descripteurs respectivement : (X5), (EEig09x) et (MEcc).	35
Figure-10	Diagramme de Williams des composés.	38
Figure-11	Graphe des valeurs prédites Tf en fonction des valeurs observées.	39
Figure-12	Test de randomisation associé au modèle QSPR.	40

LISTE DES TABLEAUX

N°	TABLEAUX	PAGES
Tableau-01	Nomenclature et valeurs de propriété des alcools étudiés	03
Tableau-02	Quelques blocks des descripteurs calculés par logiciel dragon .	12
Tableau-03	Les composés étudiés et leurs solubilités	31
Tableau-04	Classes et signification des descripteurs	33
Tableau-05	Valeurs des descripteurs moléculaires sélectionnés	34
Tableau-06	Paramètres de la normalité des données.	36
Tableau-07	Matrice de corrélation.	36
Tableau-08	Analyse de régression	37
Tableau-09	Les paramètres statistiques pour l'ensemble de calibration (17composés) et celui de validation (6 composés).	37

LISTE DES EQUATIONS

N°	EQUATION	PAGE
(01)	Propriété = f(Descripteurs)	09
(02)	$ATSkw = \sum_{i=1}^{nSK=1} \sum_{j>i} W_i \cdot W_j \cdot \delta_j$	12
(03)	$y_i = \beta_0 + \sum_{j=1}^p \beta_j X_{ij} + \xi_i$	18
(04)	E(e)=0	18
(05)	E(e _i , e _j) = 0 si i ≠ j	19
(06)	n - p - 1 ≥ 10	19
(07)	$R^2 = 1 - \frac{SCE}{SCT} = 1 - \frac{\sum_1^n (y_i - \hat{y}_i)^2}{\sum_1^n (y_i - \bar{y})^2}$	22
(08)	$SDEP = \sqrt{\frac{1}{n} \sum_1^n (y_i - \hat{y}_{(i)})^2} = \sqrt{\frac{PRESS}{n}}$	22
(09)	$SDEC = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$	23
(10)	$SDEP_{ext} = \sqrt{\frac{\sum_{i=1}^{n_{ext}} (y_i - \hat{y}_{(i)})^2}{n_{ext}}}$	23
(11)	$Q_{100}^2 = \frac{SCT - PRESS}{SCT}$	23
(12)	$Q_{ext}^2 = 1 - \frac{\sum_{i=1}^{n_{ext}} (y_i - \hat{y}_i)^2 / n_{ext}}{\sum_{i=1}^n (y_i - \bar{y})^2 / n}$	24
(13)	Tf = 1307 (±189.2) + 115 (± 8.523) X5 -37.3 (±10.59) EEig09x - 1145(±191.9) MEcc.....(13)	38

Symboles et abréviations

AG:	Algorithme génétique (Genetic Algorithm).
e_i :	Résidu différence entre les valeurs observée (y_i) et estimée (\hat{y}_i).
F :	Statistique de Fisher.
H :	Matrice de projection, ou matrice chapeau.
h_{ii} :	Éléments diagonaux de la matrice chapeau.
LOO:	Cross-validation by leave-one-out: Validation croisée par omission d'une observation
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).
Pp :	méthode pas à pas
pc :	Probabilité de croisement.
PLS:	Moindres carrés partiels.
pM:	Probabilité de mutation.
PRESS :	Somme des carrés des erreurs de prédiction.
Q ² LOO :	Coefficient de prédiction.
QSPR :	Quantitative Structure/ Property Relationships. (Relations Structur /Propriétés quantitatives.)
R ² :	Coefficient de détermination.
s :	Erreur standard.
SCE :	Somme des carrés des écarts.
SCT :	Somme des carrés totale.
SDEC:	Standard Deviation Error in Calculation: Déviation standard l'erreur calculée .
SDEP:	Standard Déviation Error of Prediction: Déviation standard de l'erreur de prédiction .
SDEPext:	External Standard Deviation Error of Prediction: Déviation standard de l'erreur de prédiction externe.

Tf:	Température de fusion.
X :	Matrice des valeurs observées des variables explicatives.
X ' :	Matrice transposée de X .
y :	Vecteur de dimension n.
yi :	Valeur observée.
^yi :	Valeur estimée.
α :	Niveau de confiance.

Introduction générale :

Les alcools sont des molécules organiques ressemblants aux alcanes sur les quels se trouve greffé un groupement hydroxyle –OH

Ces composés organiques ont des propriétés particulières

On appelle changement d'état toutes les transitions de phases. Les trois phases les plus courantes sont : la phase solide, la phase liquide et la phase gazeuse.

Les variations de température et/ou de pression et/ou de volume permettent de passer d'un état à un autre, en d'autres termes, de changer d'état.

Plus la chaîne carbonée d'un alcool est longue et plus les interactions de Van der Waals ou liaisons intermoléculaires sont fortes. Cela signifie qu'il faut apporter plus d'énergie thermique pour vaincre ces interactions et provoquer un changement d'état.

La densité et la température de fusion croissent régulièrement avec la masse molaire.

Par conséquent, les températures de fusion d'un alcool sont d'autant plus élevées que leur chaîne carbonée est longue.

Les méthodes de contributions de groupes [01], peuvent être appliquées pour la prédiction des propriétés de composés homologues. En général, ces méthodes sont inefficaces quand on extrapole à des nombres d'atomes de carbone élevés [02]. Les équations théoriques, qui présentent un comportement asymptotique correct, fournissent des résultats médiocres concernant les composés à petits nombres d'atomes de carbone.

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

Les techniques les plus courantes pour établir des modèles QSPR utilisent l'analyse de régression (régression multilinéaire : MLR; projection des structures latentes par les moindres carrés partiels : PLS), les réseaux neuronaux RNA, et les méthodes de classification [03].

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

Introduction générale

régresseurs (descripteurs moléculaires), on peut chercher seulement un ensemble réduit de régressors 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 [04].

Nous avons appliqué la méthode MLR pour modéliser la température de fusion, de 23 alcools industriellement importants dans la perspective du génie chimique. En utilisant les algorithmes génétiques AG pour choisir le bon modèle.

Ce mémoire est constitué de deux parties:

- ✓ A la suite d'une introduction générale, la première partie est une étude consacrée à différentes méthodes de traitement QSPR. ; Aussi étude quelques propriétés d'alcool, et influence de température de fusion

- ✓ La deuxième partie présente la partie application

Le mémoire est enfin complété par une conclusion général.

I- Collecte des données

La propriété considérée (Température de fusion) est l'une des grandeurs physiques importantes dans la chimie analytique.

Les données utilisées dans ce travail, prélevées dans [tableau01], concernent 23 alcools ; elles sont réunies dans le tableau - 01.

- On a choisis aléatoirement 06 alcools [pour la validation externe].
- Les 17 alcools [pour calibration qui sert à la construction du modèle].

Tableau - 01 Nomenclature et valeurs de propriété des alcools étudiés

N°	Composé	CAS	T _{fp} , K
01	Methanol	67-56-1	175.49
02	Ethanol	64-17-5	159.05
03	1-propanol	71-23-8	147.00
04	2-propanol*	67-63-0	183.65
05	1-butanol*	71-36-3	183.35
06	1-pentanol	71-41-0	194.25
07	2-methyl-2-butanol	75-85-4	264.40
08	3-methyl-1-butanol	123-51-3	155.95
09	Cyclohexanol	108-93-0	297.65
10	1-hexanol	111-27-3	229.20
11	2-hexanol*	626-93-7	223.00
12	2-methyl-2-pentanol	590-36-3	171.15
13	1-heptanol*	111-70-6	239.20
14	1-octanol	111-87-5	257.65
15	2-octanol	123-96-6	241.15
16	2-ethyl-1-hexanol	104-76-7	203.20
17	1-nonanol	143-08-8	268.15
18	1-decanol*	112-30-1	280.05
19	1-dodecanol	112-53-8	297.10
20	1-hexadecanol	4485-13-6	322.45
21	1-heptadecanol*	1454-85-9	327.00
22	1-octadecanol	112-92-5	331.00
23	1-eicosanol	629-96-9	339.00

* composés de validation externe.

II- Optimisation de la géométrie moléculaire et génération du descripteur moléculaire.

II-1 -Préparation de la base des données

II-1-1-Logiciels « ChemDraw 7 »

ChemDraw Pro [05] 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.

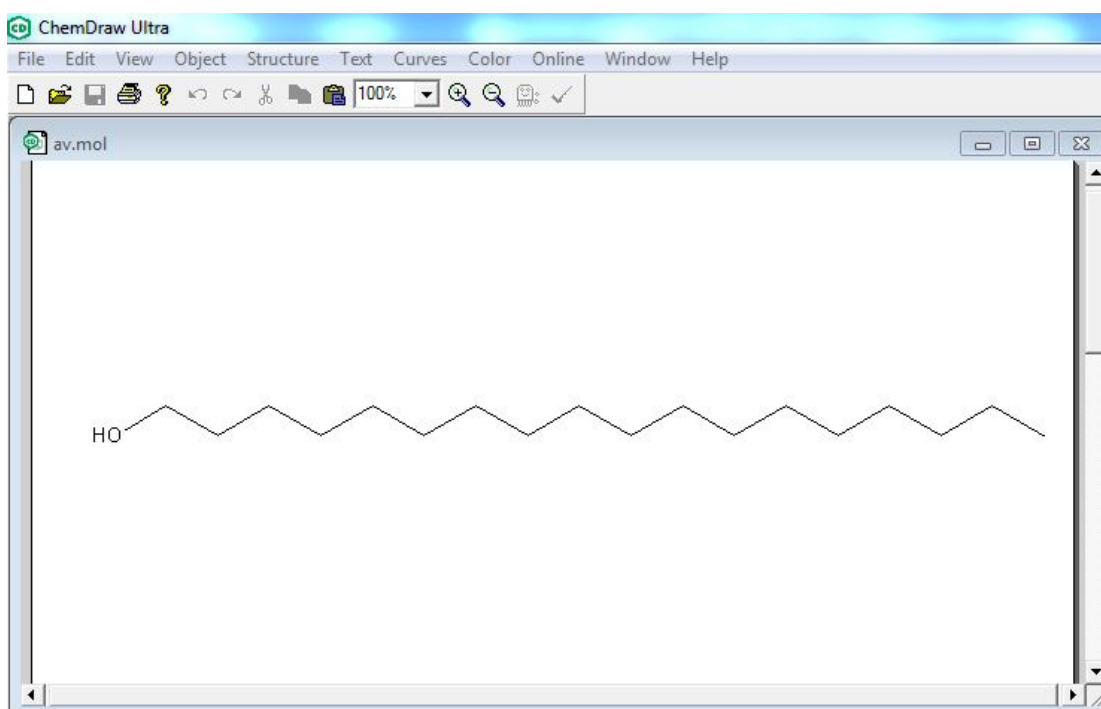


Figure 01 : Représentation des molécules par Chem Draw

Composé : 1-octadecanol

II-1-2-Logiciels « HyperChem »

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

➤ **Fonctionnalités:**

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

HyperChem est utilisé dans cette étude pour construire et optimiser les molécules, chaque molécule est enregistrée comme un fichier d'extension "Hin" après l'optimisation. Nous avons utilisé la méthode semi empirique MM+ pour l'optimisation. On a 23 molécules donc on obtient autant de fichiers Hin, en suite on va calculer les descripteurs moléculaires à partir de ces fichiers par logiciel DRAGON pour chaque molécule [06].

II-2-Récupération et stabilisation des molécules

II-2-1- Stabilisation structure des molécules (minimisation de l'énergie)

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

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 [07].

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

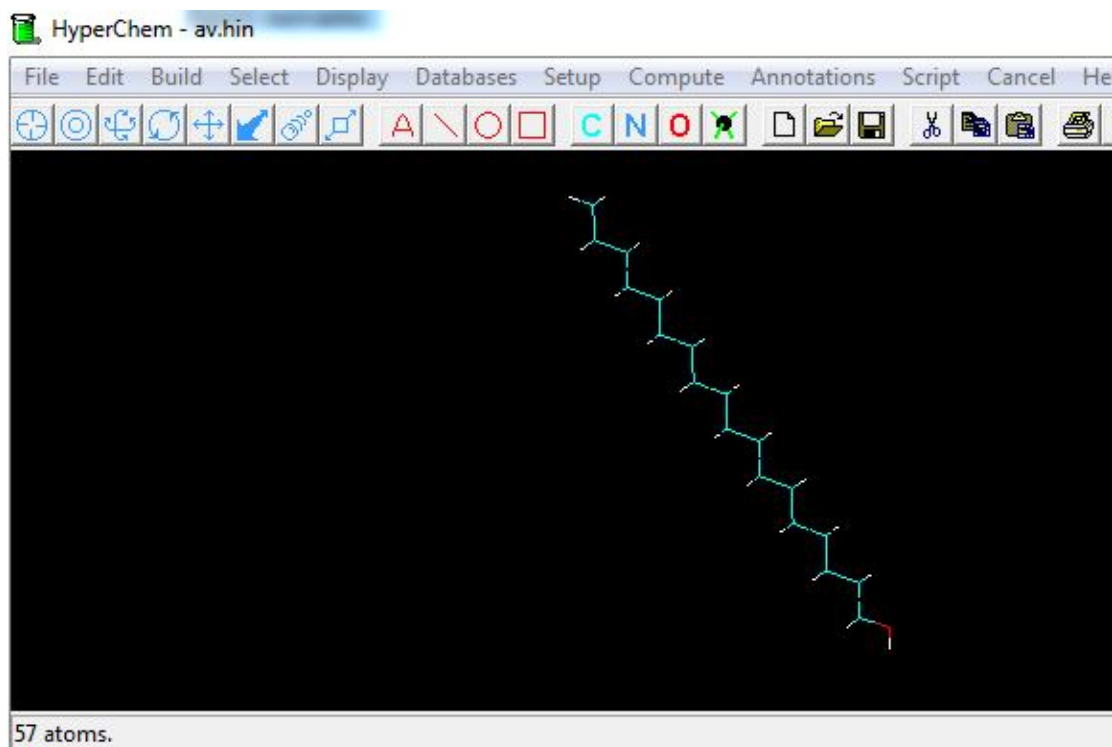


Figure 02 : applique le Field Force à Composé : -1octadecanol

II-2-3- Récupération des fichiers

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

Dans le calcul de descripteurs moléculaires, les structures chimiques des composés optimisés sont nécessaires. Les structures chimiques de tous les 23 composés dans notre jeu de données, ont été établies dans le logiciel HyperChem, et pré-optimisés en utilisant le champ avant MM + mécanique.

II-3- 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 [08].

III- Les relations structure-propriété/activité quantitatives

III-1-Introduction

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

Ces méthodes s'appuient sur le principe que les propriétés physico-chimiques et les activités biologiques des molécules dépendent fortement de leurs structures chimiques.

III-2-Généralités sur la modélisation QSPR/QSAR

Une relation QSPR/QSAR est un modèle ou une formule mathématique qui permet de relier, d'une manière quantitative, la structure d'une molécule à une propriété ou à une activité donnée. Les méthodes QSPR/QSAR sont de plus en plus utilisées, du fait de la croissance des moyens de calculs (machines, logiciels,...).

Récemment, on assiste à la mise en place d'un nouveau règlement REACH [09] (Registration, Evaluation, Authorisation and Restriction of Chemicals) qui recommande l'utilisation des méthodes alternatives pour limiter le recours à l'expérimentation.

En fait, les premiers travaux QSPR/QSAR remontent au 19^{ème} siècle. En effet, dès 1868, Crum-Brown et Fraser [10] ont postulé l'existence de relations entre les activités physiologiques et les structures chimiques en reliant les variations de l'activité biologique à des modifications structurales. Cependant, à cette époque, les structures moléculaires n'étaient pas encore connues

Partie Théorique

Une avancée importante vers les modèles QSPR/QSAR proprement dits a été réalisée grâce au développement des équations de Hammett [11] dans lesquelles les constantes σ caractérisent de manière quantitative les vitesses de réactions pour les composés organiques.

$$\text{Log } k/k_0 = \rho \sigma$$

où k et k_0 sont les constantes respectives de la réaction étudiée et de celle d'une référence et ρ une constante de réaction dépendant du type de réaction [11,12].

Les premiers travaux utilisant la méthodologie QSPR/QSAR telle qu'employée actuellement sont dus à Hansch [13] et Free et Wilson [14].

D'un côté, Hansch a proposé des modèles reliant directement l'activité biologique des composés avec les propriétés hydrophobes, électroniques et stériques à l'échelle moléculaire.

D'un autre côté, Free et Wilson ont développé des modèles empiriques, dits de contributions de groupes, pour l'étude de l'activité biologique.

Au cours de ces dernières décennies, l'utilisation des méthodes QSPR/QSAR n'a pas cessé de progresser.

Elles sont même devenues indispensables en chimie pharmaceutique, en toxicologie et pour la conception de médicaments [15-17].

III-3- Principe des méthodes QSPR/QSAR

Le principe des méthodes QSPR/QSAR est d'établir une relation mathématique reliant de manière quantitative des propriétés moléculaires, appelées descripteurs, avec une observable macroscopique (activité biologique, toxicité, propriété physico-chimique, etc.), pour une série de composés chimiques similaires à l'aide de méthodes d'analyses de données.

La forme générale d'un tel modèle est la suivante :

$$\text{Propriété/ Activité} = f(D_1, D_2, \dots, D_n, \dots) \dots\dots\dots (1)$$

D_1, D_2, \dots, D_n sont des descripteurs des structures moléculaires.

L'objectif d'une telle méthode est d'analyser les données structurales afin de détecter les facteurs déterminants pour la propriété /activité mesurée.

Pour ce faire, différents types d'outils statistiques peuvent être employés :

- Régressions linéaires simples et multiples [18] (voir chapitre 2),
- Régressions aux moindres carrées partielles (PLS) [19],
- Arbres de décision [20],
- Réseaux de neurones [21-23],
- Algorithmes génétiques [24].
- Vecteurs Machines [23].

Une fois cette relation est établie et validée, elle peut alors être employée pour la prédiction de la propriété /activité de nouvelles molécules, pour lesquelles les valeurs expérimentales ne sont pas disponibles. De tels modèles peuvent être également utilisés pour mieux comprendre les mécanismes et les modes d'action.

III-4- Méthodologie générale d'une étude QSPR/QSAR

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

- a-** Constituer une base de données à partir des mesures expérimentales fiables de la propriété ou de l'activité de chaque composé.
- b-** Sélectionner les descripteurs en relation avec la propriété ou l'activité étudiée.
- c-** Diviser cette base de données, aléatoirement, en une série d'apprentissage (training set) qui contient généralement les 2/3 de la base de données et une série de test (test set) constituée par le 1/3 restant
- d-** Etablir des modèles mathématiques en utilisant la série d'apprentissage.
- e-** Caractériser les modèles élaborés par leurs indices de validation internes et vérifier leur robustesse par un test de hasardisation (Randomisation) de la variable dépendante Y (réponse).
- f-** Valider les modèles élaborés en utilisant la série de test et calculer leurs paramètres statistiques de validation externe.
- g-** Elaborer le domaine d'applicabilité du modèle retenu.

- h- Explorer et exploiter les modèles validés pour comprendre les mécanismes et les modes d'action.

Remarque: La stratégie adoptée dans les programmes CODESSA et CORAL est différente de celle présentée ci-dessus. Ces programmes utilisent toutes les informations contenues dans la structure chimique et le meilleur modèle obtenu est celui qui possède les meilleures données statistiques. Cependant, bien que ces modèles aient un bon pouvoir prédictif, ils ne sont pas toujours utiles pour interpréter et expliquer les mécanismes et les modes d'action.

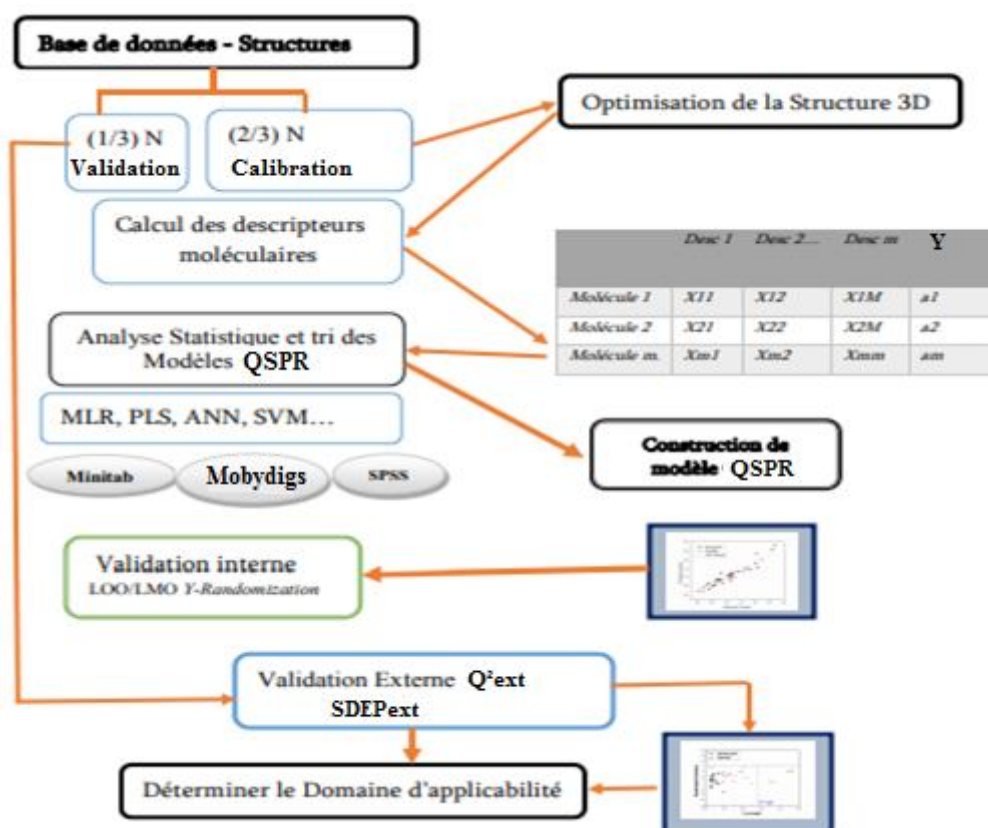


Figure03 : Procédure d'obtention et de validation d'un modèle QSPR/QSAR

III-4.1. Bases des données

Un modèle QSPR/QSAR est très dépendant des données expérimentales. Le choix de la base de données est décisif dans le développement de tel modèle.

Une base de données doit être composée de données expérimentales aussi fiables que possible obtenues en suivant un protocole unique puisque les erreurs sur celles-ci se

propageront dans le modèle final. Plusieurs bases de données sont disponibles sur le net, on peut citer :

- ECHA: <http://echa.europa.eu/fr>
- Oracdatabase: <http://oracdatabase.com/list-of-foods/search/>
- TOXNET...

III-4-2- Méthodes d'analyse des données

Une méthode d'analyse des données est nécessaire pour élaborer un modèle QSPR/QSAR.

Il existe plusieurs méthodes pour construire un modèle et analyser les données statistiques de ce dernier, ces méthodes sont disponibles dans des logiciels tels que, Excel, Origin Microcal, Minitab, Statistica, SPSS, R,... La méthode utilisée dans ce travail est la méthode de Régression Linéaire (SLR/MLR) implémentée sur Minitab et qui est présentée en détail dans la partie 3 de ce chapitre.

III-4-3-Descripteurs moléculaires

III-4-3-1-Définition d'un descripteur : c'est l'information contenue dans la structure d'une molécule.

III-4.3.2. Groupe des descripteurs moléculaires

En effet, les descripteurs moléculaires sont basés sur différentes théories, tel que le quantum chimique, la théorie de l'information, la théorie du graphe, et ainsi de suite, et également utilisés pour modeler plusieurs propriétés de produits chimiques dans les différents domaines scientifiques tel que la toxicologie, la chimie analytique, la chimie physique...etc. [25].

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)

Partie Théorique

Tableau 02 : Quelques blocks des descripteurs calculés par logiciel dragon [25]

Classe	Sous- classe
Descripteurs constitutionnels	- Dénombrement des atomes ou des liaisons. - Descripteurs basés sur les masses atomiques.
Descripteurs Topologiques	- Indices topologiques (connectivité). - Descripteurs théoriques d'information. - Descripteurs topo-chimiques.
Descripteurs Géométriques	- 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	- Charges atomiques partielles. - Moments électriques moléculaires - Polarisabilités moléculaires. - Descripteurs du champ électrique moléculaire.
Descripteurs liés aux orbitales Moléculaires	- Energie des OM frontières - Ordres de liaison - Indices de réactivité de Fukui.
Descripteurs température Dépendants	- Fonctions thermodynamiques. - Descripteurs facteurs de Boltzmann pondérés.
Descripteurs de solvation	- 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	- Descripteurs topographiques. - Descripteurs électro-topologiques. - Descripteurs de la charge partielle de l'aire de la surface.

III-4 .3.3- Exemple : calcul de ATSK_w

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

$$ATSk_w = \sum_{i=1}^{n_{SK}=1} \sum_{j>i} W_i \cdot W_j \cdot \delta_j \quad (2)$$

- n_{SK} est le nombre d'atomes de molécule d'non-Hydrogène
- δ_{ij} est le delta de Kronecker ($\delta_{ij} = 1$ si $d_{ij} = k$, zéro autrement, d_{ij} étant la distance topologique entre deux atomes considérés).

Les blocks des descripteurs moléculaires calculés par Dragon sont représentés dans la figure suivante :

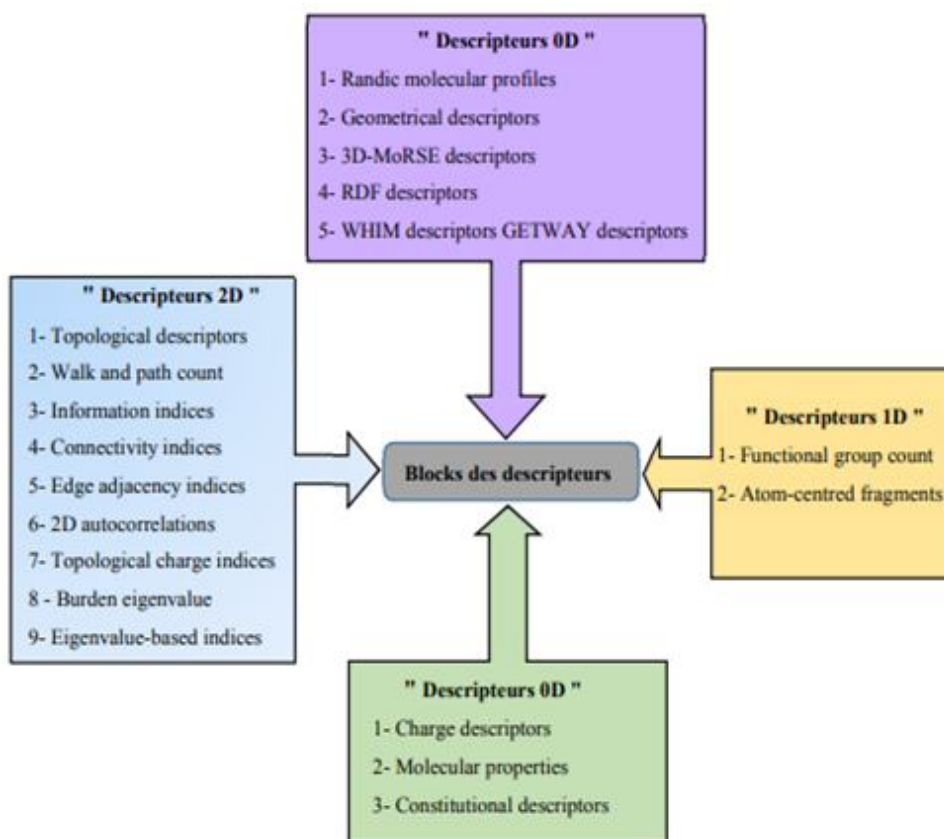


Figure- 04 : Représentation des blocs des descripteurs moléculaire.

III-4 .3.4-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 :

- 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 [26].

IV-Diagramme montre les étapes de prédiction

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

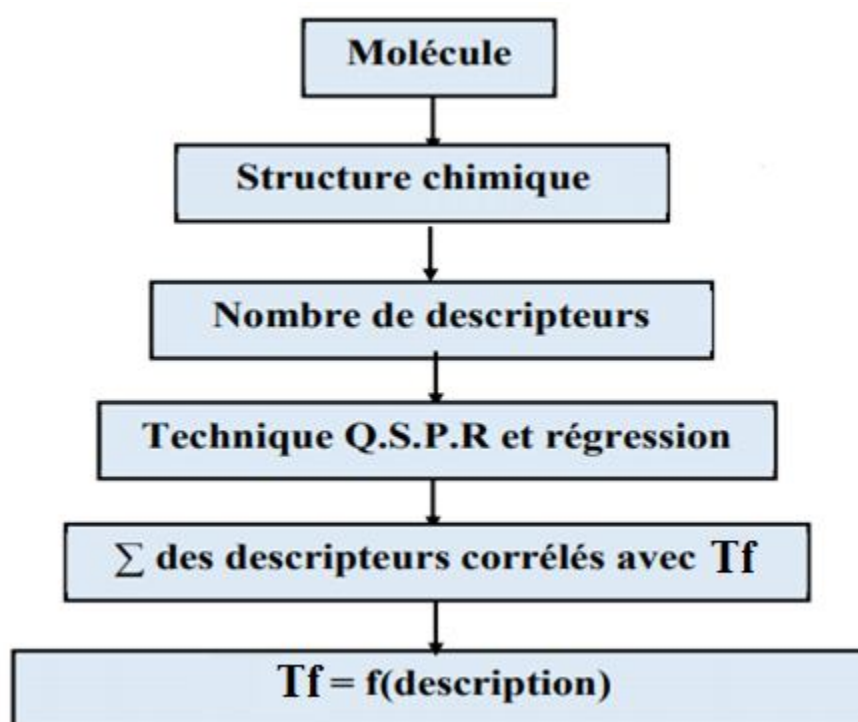


Figure 05 : Diagramme de prédiction par QSPR

V- 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.

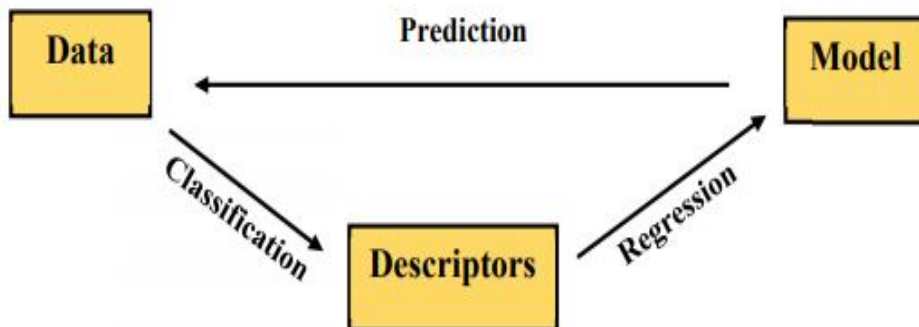


Figure06 : le cycle de prédiction

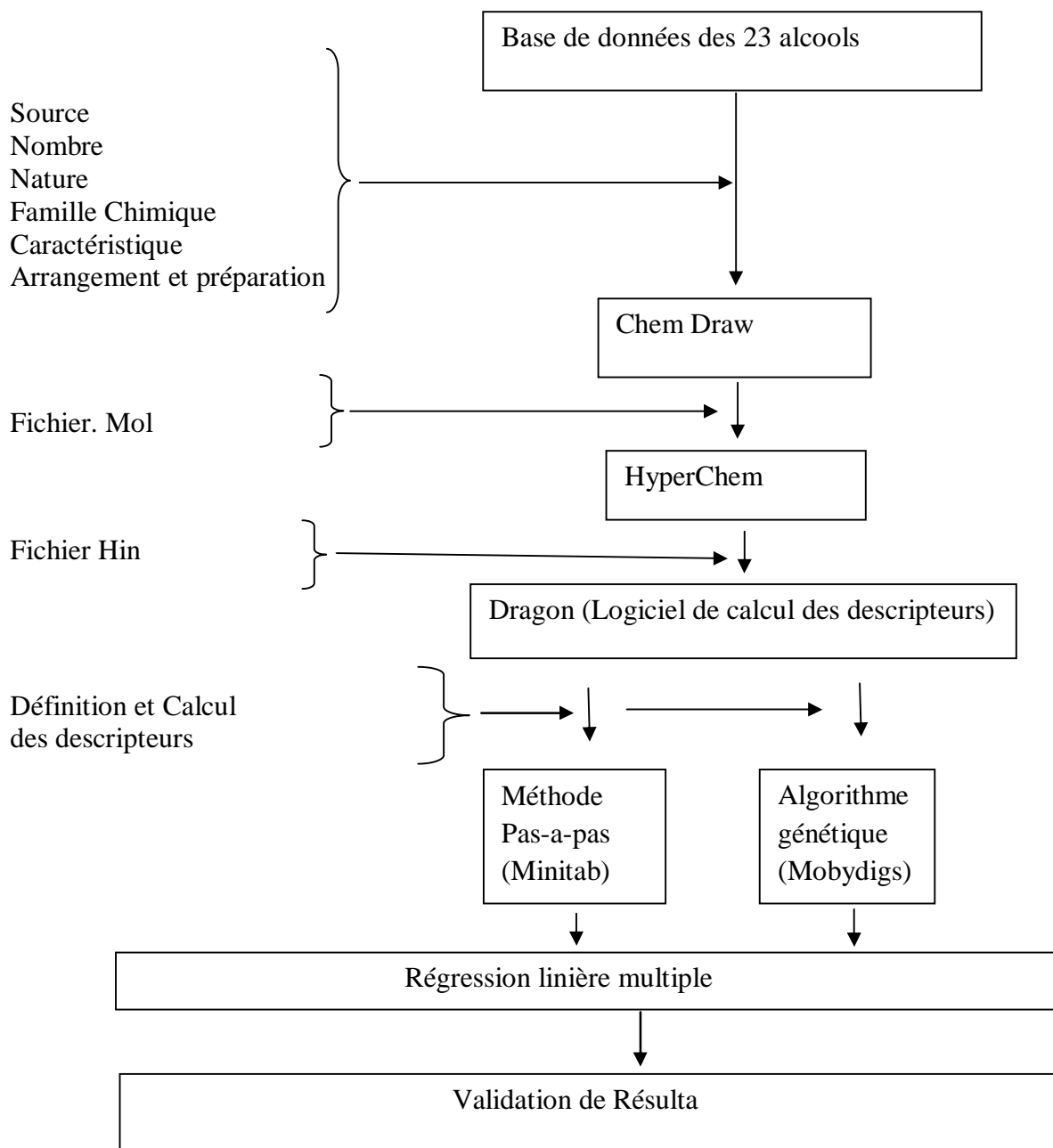


Figure07 : Diagramme de notre travail

VI- Méthodes de régressions linéaires et multilinéaire

VI-1 Aperçu général

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

- a) Un nombre maximal de données expérimentales (de préférence toutes) doivent être caractérisées par des valeurs de descripteurs originaux complémentaires.
- b) Les valeurs des descripteurs doivent être obtenues de la même source et, de préférence, mesurées selon le même protocole expérimental ou calculées en utilisant le même logiciel.
- c) Le nombre de descripteurs dans les modèles de régression multiples doit être minimisé, sans perte d'information, ce que mettent en évidence les critères statistiques (valeurs des tests t et F...).
- d) Dans les modèles MLR, les descripteurs utilisés doivent être statistiquement orthogonaux.
- e) Pourvu que les autres critères soient similaires, la nature physique ou chimique du descripteur sélectionné doit être la plus proche de la propriété ou du phénomène étudié.

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

VI-2- Evaluation préliminaire des données

Avant d'entamer le développement effectif des équations de régression QSPR, il est hautement recommandé d'examiner la qualité statistique des données de départ, à la

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

On distingue habituellement dans un tel pré- traitement des données les analyses uni variées des analyses bi-variées [27-32].

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.

VI-3- Régression linéaire multiple

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

$$y_i = \beta_0 + \sum_{j=1}^p \beta_j X_{ij} + \varepsilon_i \dots \dots \dots (3)$$

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

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

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

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

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

c) Les résidus sont indépendants entre eux :

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

Leurs covariances sont nulles.

d) 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 [33] tel que:

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

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

VI-4- Méthodes de sélection des descripteurs

VI-4-1- Régression pas à pas

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

A - Méthode rétrograde (backwards election)

Cette méthode consiste à construire un modèle de régression complet (intégrant toutes les variables explicatives), et à en retirer une par une les variables dont le t partiel est non significatif (en commençant par celle qui explique le moins de variation).

Inconvénient : une fois qu'une variable a été retirée, elle ne peut plus être réintroduite dans le modèle, même si, à la suite du retrait d'autres variables, elle redevenait significative. Cette approche est néanmoins assez libérale (elle a tendance à garder un nombre plus élevé de variables dans le modèle final que les autres approches ci-dessous).

B - Méthode progressive (forwards election)

Approche inverse de la précédente : elle sélectionne d'abord la variable explicative la plus corrélée à la variable dépendante. Ensuite, elle sélectionne, parmi celles qui restent, la variable explicative dont la corrélation partielle est la plus élevée (en gardant constantes là où les variables déjà retenues). Et ainsi de suite tant qu'il reste des variables candidates dont le coefficient de corrélation partiel est significatif. Inconvénient : lorsqu'une variable est entrée dans le modèle, aucune procédure ne contrôle si sa corrélation partielle reste significative après l'ajout d'une ou de plusieurs autres variables. Cette technique est en général plus conservatrice que la précédente, ayant tendance à sélectionner un modèle plus restreint (moins de variables explicatives) que la sélection rétrograde.

Des simulations récentes [34] montrent que même la sélection progressive, la plus conservatrice des trois variantes, est trop libérale, c'est-à-dire qu'elle laisse souvent entrer au moins une variable non significative dans le modèle. C'est la raison pour laquelle nous proposons désormais d'appliquer un double critère d'arrêt à la sélection pas à pas (plus spécifiquement à la sélection progressive) :

- * Le niveau α habituel

- * Le R^2_{aj} du modèle comprenant toutes les variables candidates.

Pour ce deuxième critère, on calcule tout d'abord le R^2_{aj} global d'une régression multiple comprenant toutes les variables explicatives candidates. Ensuite, durant la procédure de sélection, on arrête la sélection lorsque le niveau α présélectionné ou le R^2_{aj} global est atteint.

Cette procédure garantit une erreur de type I correcte et réduit fortement le nombre de variables explicatives introduites à tort dans le modèle. C'est la méthode de sélection que nous préconisons.

VI-4.2-Algorithmes génétiques

La modélisation de processus génétiques a initié le développement des algorithmes génétiques, qui peuvent être exploités dans une grande variété de problèmes d'optimisation [35].

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 ont été réalisées par pas-à-pas en utilisant le logiciel de calcul statistique MINITAB version 16.2.0 [36] ; et par algorithme génétique, dans la version MOBY DIGS de Todeschini [37].

VI-5 – Paramètres d'évaluation de la qualité de l'ajustement

Deux paramètres sont couramment utilisés :

- **Le coefficient de détermination multiple**

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

$$R^2 = 1 - \frac{SCE}{SCT} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \dots\dots\dots(7)$$

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

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

$$EQMP = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_{(i)})^2} = \sqrt{\frac{PRESS}{n}} \dots\dots\dots(8)$$

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

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

VI-7- Test de randomisation

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

VI-8 - Validation externe

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

- Ensemble d'estimation (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 R2 et Q2 seules, qui constituent de bons tests uniquement pour des données réparties régulièrement.

$$EQMC = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \dots\dots\dots(9)$$

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

La validation croisée par « leave – one - out » (LOO) [38] consiste à recalculer le modèle sur (n-1) observations, et à utiliser le modèle ainsi obtenu pour calculer la grandeur d'intérêt du composé écarté, notée $\hat{y}(i)$. On répète le procédé pour chacune des grandeurs d'intérêt. La somme des carrés des erreurs de prédiction, désignée par le symbole PRESS (équation. (11)), est une mesure de la dispersion des estimations. On l'utilise pour définir le coefficient de prédiction :

$$Q_{Lo0}^2 = \frac{SCT - PRESS}{SCT} \dots\dots\dots(11)$$

Contrairement à R2 qui augmente avec le nombre de paramètres du modèle, le facteur Q²LOO 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 de $Q^2_{LOO} > 0,5$ est considérée comme satisfaisante, et une valeur supérieure à 0,9 est excellente [39].

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 n'est pas suffisante, et peut conduire à une surestimation de la capacité prédictive du modèle lorsqu'il est appliqué à des composés réellement externes.

Evidemment, on peut être amené à écarter 2, 3 ou un plus grand nombre d'éléments à la fois, ce qui conduit aux procédures LMO (leave – many- out). Cependant, ces procédures ne sont que rarement rapportées avec les résultats QSPR courants, et n'ont pas été pleinement exploitées dans le présent travail.

Dans le cas où on a suffisamment de données qui n'ont pas servi dans la création du modèle ou après collecte de nouvelles, on peut ou on doit procéder à la validation de

Ce dernier, c'est la validation externe. La statistique se rapportant à ce procédé, notée Q^2_{ext} , est calculée comme suit :

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

Pour une grande valeur de Q^2_{LOO} , une valeur élevée de Q^2_{ext} permet de présager d'une bonne capacité prédictive du modèle. Dans notre cas : $n=17$ et $n_{ext}=6$.

VII- Applications des méthodes QSPR/QSAR

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

- **Propriétés physico-chimiques** : point d'ébullition, Point de fusion, densité, indice de réfraction, température critique, viscosité, solubilité, pression de vapeur, tension superficielle, Coefficients de partition : eau/octanol, air/eau, huile/air, lait/plasma [40]...

Partie Théorique

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

- **Autres propriétés/activités** :

- Prédiction de la toxicité aquatique des composés chimiques vis-à-vis des espèces environnementales.

- Toxicité des nanoparticules

- Toxicité des pesticides et des colorants

- Propriétés inhibitrices de corrosion

- Concentration micellaire critique

- Prédiction de plusieurs propriétés dangereuses telle que l'explosibilité et l'inflammabilité de certaines familles de molécules chimiques.

- Conception des médicaments et de nombreux autres produits tels que les agents tensio-actifs, parfums, les colorants et les produits de la chimie fine.

VIII -les alcools.

VIII-1-Définition

En chimie organique, un alcool est un composé organique dont l'un des carbones (celui-ci étant tétraédrique) est lié à un groupement hydroxyle (-OH). L'éthanol (ou alcool éthylique) entrant dans la composition des boissons alcoolisées est un cas particulier d'alcool, mais tous les alcools ne sont pas propres à la consommation. En particulier, le méthanol est toxique et mortel à haute dose . [41]

VIII-2- Etude chimique

VIII-2-1- Nomenclature

Lorsque l'alcool est la fonction principale, il suffit de remplacer la voyelle terminale « e » de l'alcane correspondant par le suffixe -ol et d'indiquer le numéro de l'atome de carbone où le groupe hydroxyle est fixé, bien que, parfois, lorsqu'elle n'est pas nécessaire à la description, cette dernière information soit omise.

Si elle n'est pas la fonction principale, il faut ajouter le préfixe hydroxy- précédé du numéro de l'atome de carbone où le groupe est fixé.

Pour la base conjuguée de l'alcool, l'ion alcoolate, il suffit de remplacer la voyelle terminale « e » par le suffixe -olate. (Ne pas confondre avec le suffixe -oate caractéristique du carboxylate, base conjuguée de l'acide carboxylique.) [41]

Exemples :

- ✓ éthanol :
- ✓ 2-butanol :
- ✓ acide 3-hydroxy-propanoïque :

VIII-2-2- Classification

De manière générique, un alcool contient donc la séquence : R – OH

Où R est un radical organique variable, souvent un alkyle.

Selon la nature du carbone portant le groupement alcool, on distingue :

- les alcools primaires, dont le carbone comportant le groupement hydroxyle est lié à au moins deux atomes d'hydrogène et un radical organique R :
- les alcools secondaires, dont le carbone comportant le groupement hydroxyle est lié à un atome d'hydrogène et deux radicaux organiques R et R' :
- les alcools tertiaires, dont le carbone comportant le groupement hydroxyle est lié à trois radicaux organiques R, R' et R'' :
- les phénols, sont parfois considérés comme des alcools particuliers dont le groupement hydroxyle est lié à un carbone d'un cycle benzénique. Leur réactivité étant tellement différente de celle des autres alcools (ici le carbone portant le groupement -OH n'est pas tétraédrique), les phénols sont généralement classés en dehors de la famille des alcools. [41]

Il existe aussi un groupe considéré parfois comme un cas particulier d'alcools appelés énols. Il s'agit d'une molécule dans laquelle la fonction hydroxyle est attachée sur un carbone d'une double liaison C=C (ici encore le carbone portant le groupement -

OH n'est pas tétraédrique). Il s'agit en fait d'une forme tautomère d'un aldéhyde ou d'une cétone. La forme majoritaire est généralement l'aldéhyde ou la cétone, et non l'énol, sauf cas particuliers où la forme énol est stabilisée par mésomérie comme par exemple les phénols.

VIII-2-3-Utilisation

Les alcools sont utilisés dans l'industrie chimique comme :

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

VIII-2-4-Propriétés physico-chimiques

2-4-1-Aspect

Les alcools de faible masse moléculaire se présentent à température ambiante comme des liquides incolores ; les alcools plus lourds comme des solides blanchâtres.

2-4-2- Polarité et présence de liaisons hydrogène

Le groupe hydroxyle rend généralement la molécule d'alcool polaire. Cela est dû à sa géométrie (coudée, de type AX₂E₂), et aux électronégativités respectives du carbone, de l'oxygène et de l'hydrogène ($\delta(O) > \delta(C) > \delta(H)$). Ces groupes peuvent former des liaisons hydrogène entre eux ou avec d'autres composés (ce qui explique leur solubilité dans l'eau et dans les autres alcools).

2-4-3- Point d'ébullition

Le point d'ébullition est élevé chez les alcools :

- en raison du groupement hydroxyle qui permet les liaisons hydrogène ;
- en raison de la chaîne carbonée qui subit des forces de van der Waals.

Aussi, le point d'ébullition des alcools est-il d'autant plus élevé que :

Le nombre de fonctions alcools est grand : un diol a une température d'ébullition supérieure à celle de l'alcool simple équivalent, qui lui-même a une température d'ébullition supérieure à l'hydrocarbure correspondant. Par exemple, parmi les alcools dérivés de l'iso propane, le glycérol (propan-1, 2,3-triol) bout à 290 °C, le propylène glycol (propan-1,2-diol) à 187 °C et le 1- propanol à 97 °C ;

La chaîne carbonée est longue : parmi les alcools linéaires, le méthanol bout à 65 °C, l'éthanol à 78 °C, le 1- propanol à 97 °C, le 1- butanol à 118 °C, le 1- pentanol à 138 °C et le 1- hexanol à 157 °C ;

La chaîne carbonée est linéaire, par maximisation de la surface de la molécule susceptible de subir les forces de van der Waals. Par exemple, parmi les pentanols, le 2, 2,1-diméthyl-propanol à 102 °C, le 2,1-méthyl-butanol à 131 °C et le 1- pentanol bout à 138 °C.

2-4-4- Solubilité

La solubilité dans l'eau des alcools dépend des deux mêmes facteurs que précédemment, mais qui sont ici antagonistes :

La chaîne carbonée, hydrophobe, tend à rendre la molécule non soluble ;

Le groupement hydroxyle, hydrophile (grâce à ses liaisons hydrogène) tend à rendre la molécule soluble.

Ainsi, les alcools sont d'autant plus solubles dans l'eau que :

Partie Théorique

La chaîne carbonée est courte : le méthanol, l'éthanol et le 1-propanol sont solubles en toutes proportions dans l'eau, le 1-butanol a une solubilité de 77 g·L⁻¹ à 20 °C, le 1-pentanol de 22 g·L⁻¹, le 1-hexanol de 5,9 g·L⁻¹, le 1-heptanol de 2 g·L⁻¹ et les alcools plus lourds sont pratiquement insolubles ;

Le nombre de fonctions alcools est élevé. Par exemple, les butanes diols sont solubles en toutes proportions tandis que le 1-butanol a une solubilité de 77 g·L⁻¹ ;

La chaîne carbonée est ramifiée : parmi les pentanols, le 2, 2,1-diméthyl-propanol a une solubilité de 102 g·L⁻¹, le 2-1-méthyle-butanol de 100 g·L⁻¹ et le 1-pentanol de 22 g·L⁻¹.

Les alcools de faible masse moléculaire sont généralement solubles dans les solvants organiques comme l'acétone, ou l'éther. [41]

IX- La température de fusion :

➤ La température de fusion :

La chaleur peut permettre aux zones organisées (lamelles, sphérolites) de se former dans un polymère par exemple, mais trop de chaleur peut aussi les détruire. Si nous continuons à chauffer notre matériau au dessus de T_c, nous atteindrons une autre transition thermique, appelée fusion. Lorsque nous atteignons la température de fusion du matériau, ou T_f, les zones organisées commencent à se disloquer. Lorsque nous atteignons la T_f. Il y a une chaleur latente de fusion au même titre qu'il y a une chaleur latente de cristallisation. Pour que les zones organisées du matériau fondent, elles doivent absorber une certaine quantité de chaleur. La fusion est une transition de premier ordre. Cela signifie que lorsque vous atteignez la température de fusion, la température du matériau ne montera pas avant que toutes les zones organisées n'aient fondues. Cela signifie que le petit appareil de chauffage sous le creuset " échantillon " se trouve dans l'obligation de fournir beaucoup de chaleur au matériau pour, d'une part fondre les zones organisées et d'autre part maintenir l'élévation continue de température au même taux que celle de le cret " référence ". Cet apport de chaleur supplémentaire, nécessaire à la fusion, correspond, sur une mesure par DSC, au creux schématisé ci-dessous : [42]

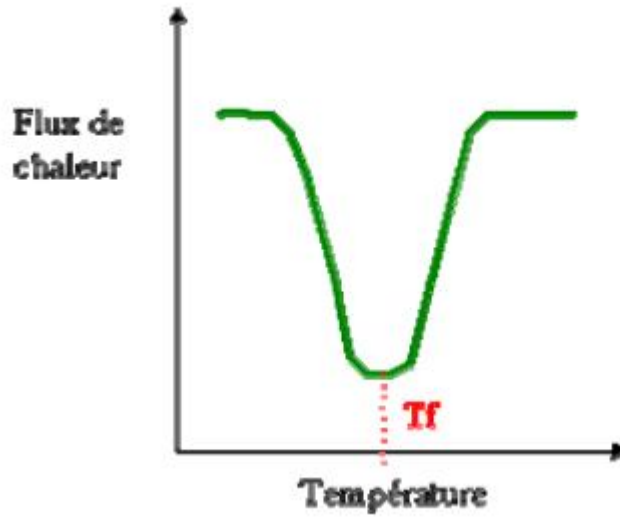


Figure 08: Tf (Température de fusion)

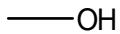
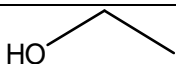
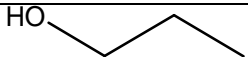
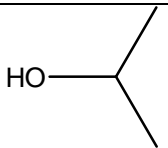
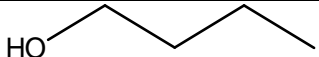
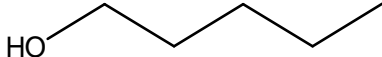
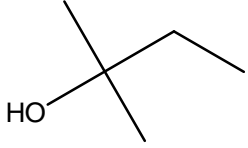
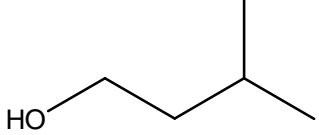
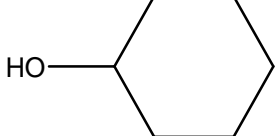
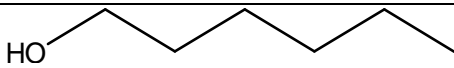
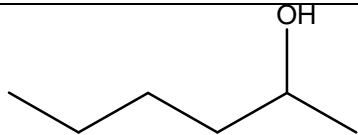
Nous pouvons mesurer la chaleur latente de fusion en mesurant la surface de ce creux. Et bien sûr, nous prendrons la température au fond du creux comme température de fusion du matériau Tf. Comme nous devons apporter de l'énergie au matériau pour obtenir sa fusion, nous sommes dans le cas d'une transition, dite " endothermique

Partie Application

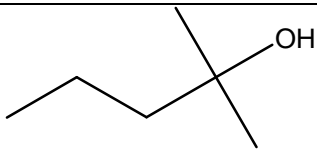
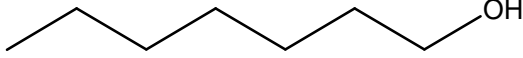
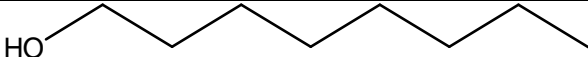
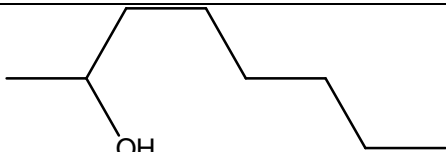
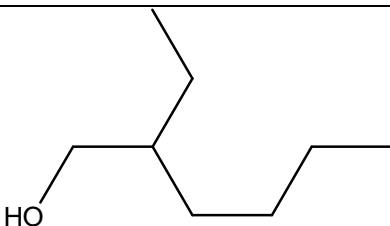
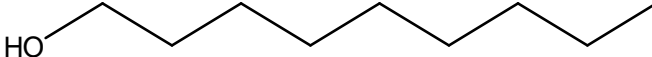
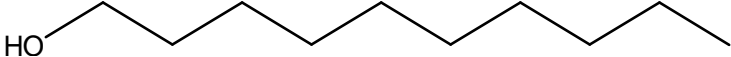
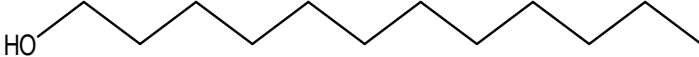


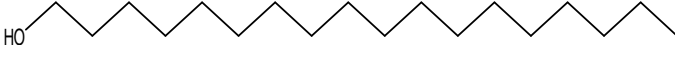
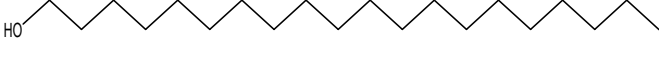
Dans la partie application, nous avons déterminé la modélisation de la température de fusion d'une série d'alcool. L'ensemble des données est divisé aléatoirement en deux sous ensembles, le premier contient 17 composés pour le calibrage et le second contient 6 composés pour la validation externe.

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

Tableau 03: les composés étudiés et leurs solubilités

N°	Formule	Structure	T _{fp} , K
1	CH ₄ O		175.49
2	C ₂ H ₆ O		159.05
3	C ₃ H ₈ O		147.00
4	C ₃ H ₈ O*		183.65
5	C ₄ H ₁₀ O*		183.35
6	C ₅ H ₁₂ O		194.25
7	C ₅ H ₁₂ O		264.40
8	C ₅ H ₁₂ O		155.95
9	C ₆ H ₁₂ O		297.65
10	C ₆ H ₁₄ O		229.20
11	C ₆ H ₁₄ O*		223.00

Partie Application

N°	Formule	Structure	Tfp, K
12	C ₆ H ₁₄ O		171.15
13	C ₇ H ₁₆ O*		239.20
14	C ₈ H ₁₈ O		257.65
15	C ₈ H ₁₈ O		241.15
16	C ₈ H ₁₈ O		203.20
17	C ₉ H ₂₀ O		268.15
18	C ₁₀ H ₂₂ O*		280.05
19	C ₁₂ H ₂₆ O		297.10
20	C ₁₆ H ₃₄ O		322.45
21	C ₁₇ H ₃₆ O*		327.00
22	C ₁₈ H ₃₈ O		331.00
23	C ₂₀ H ₄₂ O		339.00

* composés de validation externe

I-Calcul du modèle :

I-1- Choix du modèle :

Les molécules sont dessinées par le logiciel ChemDraw (ChemDraw ultra 7.0) puis elles sont optimisées en utilisant le logiciel HyperChem (HyperchemTM Release 6.03 for Windows, 2000). Les descripteurs moléculaires ont été calculés à l'aide du logiciel informatique Dragon (R Todeschini et al. 2005), plus de 1600 descripteurs sont calculés. D'après l'algorithme génétique dans la version MobyDigs (B Ballabio et al, 2004), plusieurs modèles sont obtenues pour chaque jeu de groupe; le choix a été opté

Partie Application

pour le modèle qui conduit aux meilleurs statistiques des 100 modèles générés par algorithmes génétiques. Sa validation est effectuée par 6 composés, alors que sa calibration par 17

I- 2- Choix d'une famille de descripteurs :

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

Tableau 04 : Classes et signification des descripteurs.

Descripteur	Classe	Signification
X5	Connectivity descriptors	The connectivity indices X_k , k being the order, are derived from the simple vertex degrees δ . The connectivity index X_5 is usually known as the <i>Randic connectivity index</i> .
EEig09x	Edge adjacency indices	The edge adjacency matrix is derived from the H-depleted molecular graph and encodes the connectivity between graph edges. It is a square symmetric matrix of dimension $B \times B$, where B is the number of bonds between non-hydrogen atom pairs. The entries of the matrix equal one if the considered bonds are adjacent and zero otherwise.
MEcc	Geometrical descriptors	The molecular eccentricity (MEcc) is a shape descriptor calculated from the eigenvalues λ of the molecular inertia matrix

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

Nous rapporterons les résultats obtenus pour 23 alcools sont présentées dans le tableau 05 au dessous:

Tableau-05 : Valeurs des descripteurs moléculaires sélectionnés

No.	composé	Tf	X5	EEig09x	MEcc
1	CH4O	175.49	0.000	0.000	0.997
2	C2H6O	159.05	0.000	0.000	0.998
3	C3H8O	147.00	0.000	0.000	0.999
4	C3H8O*	183.65	0.000	0.000	0.991
5	C4H10O*	183.35	0.000	0.000	1.000
6	C5H12O	194.25	0.250	0.000	1.000
7	C5H12O	264.40	0.000	0.000	0.915
8	C5H12O	155.95	0.000	0.000	0.999
9	C6H12O	297.65	0.901	0.000	0.969
10	C6H14O	229.20	0.354	0.000	1.000
11	C6H14O*	223.00	0.408	0.000	1.000
12	C6H14O	171.15	0.000	0.000	0.980
13	C7H16O*	239.20	0.479	0.000	1.000
14	C8H18O	157.65	0.604	0.000	1.000
15	C8H18O	241.15	0.568	0.000	1.000
16	C8H18O	203.20	0.577	0.000	0.999
17	C9H20O	268.15	0.729	-0.902	1.000
18	C10H22O*	280.05	0.854	-0.683	1.000
19	C12H26O	297.10	1.104	-0.136	1.000
20	C16H34O	322.45	1.604	0.815	1.000
21	C17H36O*	327.00	1.729	1.000	1.000
22	C18H38O	331.00	1.854	1.165	1.000
23	C20H42O	339.00	2.104	1.445	1.000

* composés de validation externe.

I-3- Evaluation préliminaire des données

Avant de procéder au développement effectif des équations de régression QSPR, nous avons vérifié la qualité statistique des données initiales (variable dépendante et descripteurs).

Partie Application

Quelques résultats partiels seront présentés à titre d'illustration, en exploitant les données qui apparaissent dans le tableau 1.

Les diagrammes de probabilité obtenus à partir de ces données figures 2, montrent que les variables considérées se distribuent toujours selon la loi normale, puisque les R obtenus sont systématiquement supérieurs aux R critique donnés par les tables.

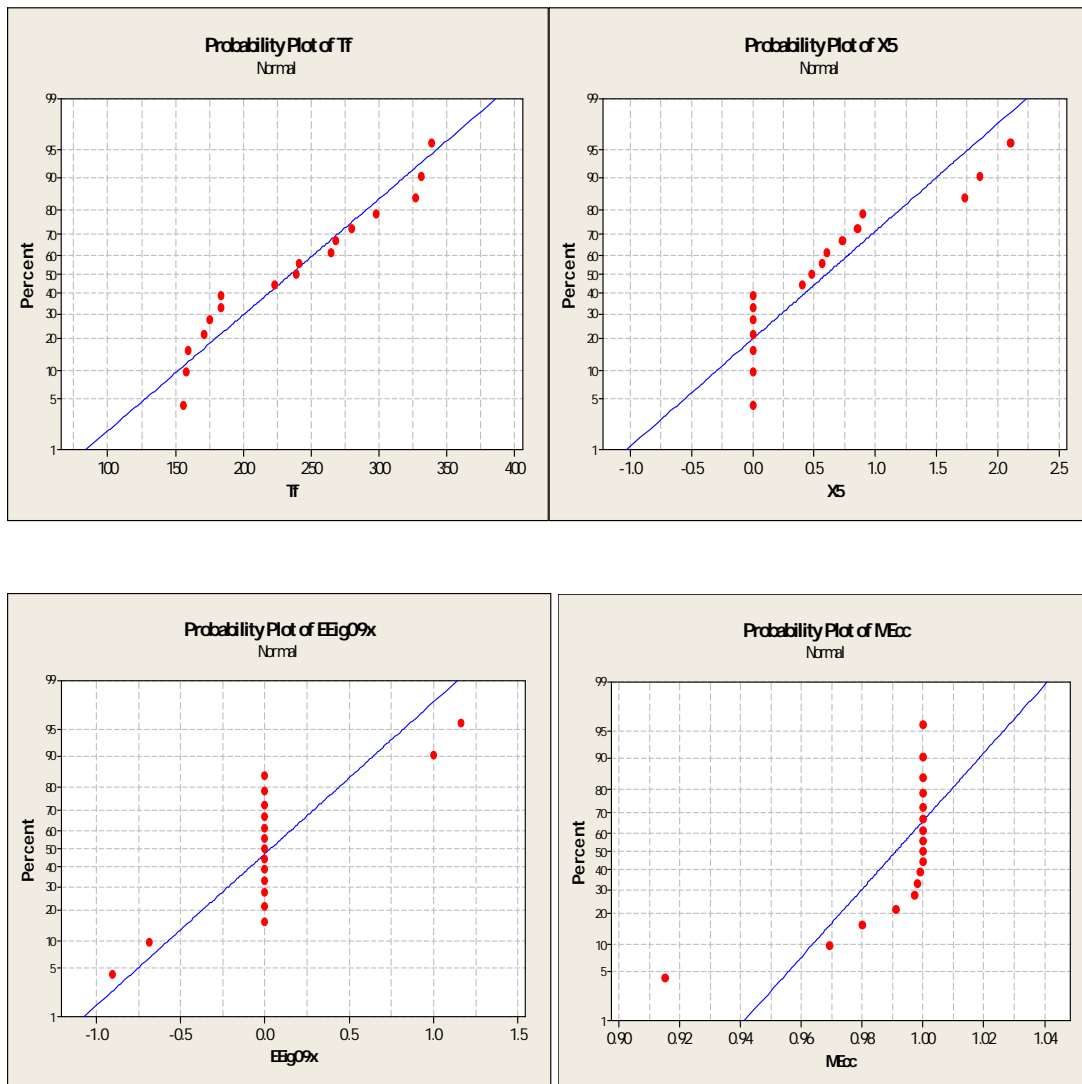


Figure 09: Test de normalité pour la propriété (Tf) et les trois descripteurs respectivement : (X5), (EEig09x) et (MEcc).

Partie Application

Le tableau suivant permet de juger la distribution des données.

Tableau 06: Paramètres de la normalité des données.

	Tf	Descripteurs		
		X5	EEig09x	MEcc
R	0,964	0,970	0,990	0,960
P	0,100	0,100	0,100	0,010
R _C		0,950		
α		0,005		

D'après les figures et 09 le tableau 06; la distribution des données est normale.

I-4- Calcul des corrélations entre les différents descripteurs

Le coefficient de corrélation, r , de Bravais-Pearson a servi pour mettre en évidence les relations possibles entre les différents descripteurs des 17 composés, la matrice de corrélation obtenue à l'aide de la commande "corrélation" du logiciel MINITAB, montre que les descripteurs sont entre eux plus ou moins corrélés. Les couples des descripteurs qui présentent des valeurs de $r > 0.90$, sont très fortement corrélés et apportent la même information, ce qui fait qu'ils ne peuvent apparaître dans une même équation de régression et ce n'est le cas dans notre cas.

Tableau 07: Matrice de corrélations.

	Tf	X5	EEig09x
X5	0.889		
EEig09x	0.513	0.714	
MEcc	-0.095	0.262	0.110

I-5 -Analyse de régression

Les valeurs des paramètres statistiques (tableau 09) montrent que les trois descripteurs (X5; EEig09x; MEcc) (tableau04) permettent de corrélérer Tf des 17 alcools.

En effet, la valeur du coefficient de détermination (R^2) signifie que 95,21% de la variabilité de la Tf peut être expliquée par ces trois descripteurs, alors que la racine de l'erreur quadratique moyenne de prédiction est de l'ordre de (SDEP= 18,65) ; en outre ce modèle est significatif avec une valeur du paramètre de Fisher : (F=86,12).

Partie Application

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

L'équation de régression est :

$$Tf = 1307 (\pm 189.2) + 115 (\pm 8.523) X5 - 37.3 (\pm 10.59) EEig09x - 1145 (\pm 191.9) MEcc \dots (13)$$

Tableau 08 : Analyse de régression.

Régresseur	Coef	SE Coef	T	P	VIF
Constante	1306.9	189.2	6.91	0.000	
X5	114.55	8.52	13.44	0.000	2.189
EEig09x	-37.27	10.59	-3.52	0.004	2.064
MEcc	-1144.6	191.9	-5.96	0.000	1.088

Les paramètres statistiques de cette validation (Q^2_{ext} , $SDEP_{ext}$) ainsi que les valeurs- des statistiques pour l'ensemble de calibration sont regroupés dans le tableau-8.

Tableau 09: Les paramètres statistiques pour l'ensemble de calibration (17composés) et celui de validation (6 composés).

Les paramétrer statistique	
Size	3
R2	95.21
Q2	91.28
Q2ext	94.58
R2adj	94.1
SDEC	13.828
SDEP	18.652
SDEPext	14.714
Kx	40.02
Kxy	49.77
F	86.1223
SE	15.8131
PRESS	5914.211

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

I-6-Diagramme de Williams :

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

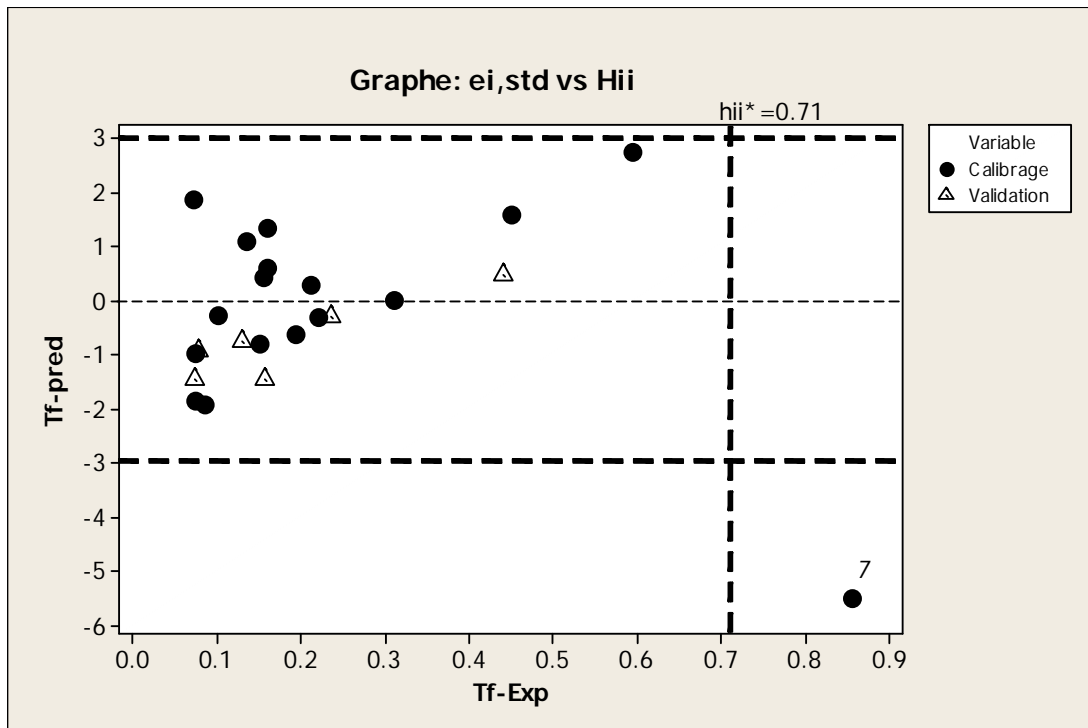


Figure 10: Diagramme de Williams des composés.

Comme le montre la figure, les valeurs de $(e_{i\text{ std}})$ sont comprises entre les bornes ± 3 ,

$h_{ii}^* = \frac{3 \times (p + 1)}{n} = 0,71$ symbolisée par la droite parallèle à l'axe des $e_{i\text{ std}}$ à l'exception

du composé 7 qui présente à la fois un point Lévis et une observation aberrante il s'agit du composé : 2-methyl-2-butanol

I-7- Evaluation de la qualité de l'ajustement :

Nous avons deux paramètres utilisés pour évaluer la qualité de l'ajustement; la valeur du coefficient de détermination $R^2 = 95,21\%$ qui explique très bien la variabilité de Tf en fonction des descripteurs choisis; la racine de l'erreur quadratique moyenne de prédiction $SDEP = 18,65$, dont la petite valeur indique un modèle très hautement significatif, que justifie la grande valeur du paramètre de Fisher : $F = 86,12$.

La qualité de l'ajustement a été vérifiée en procédant à une validation croisée par « Leave –one –out », La figure 11, qui reproduit les valeurs prédites Tf en fonction de celles observées, fait ressortir une faible dispersion caractéristique d'un bon ajustement, d'ailleurs confirmé par la grande valeur de $Q^2 = (87,74\%)$.

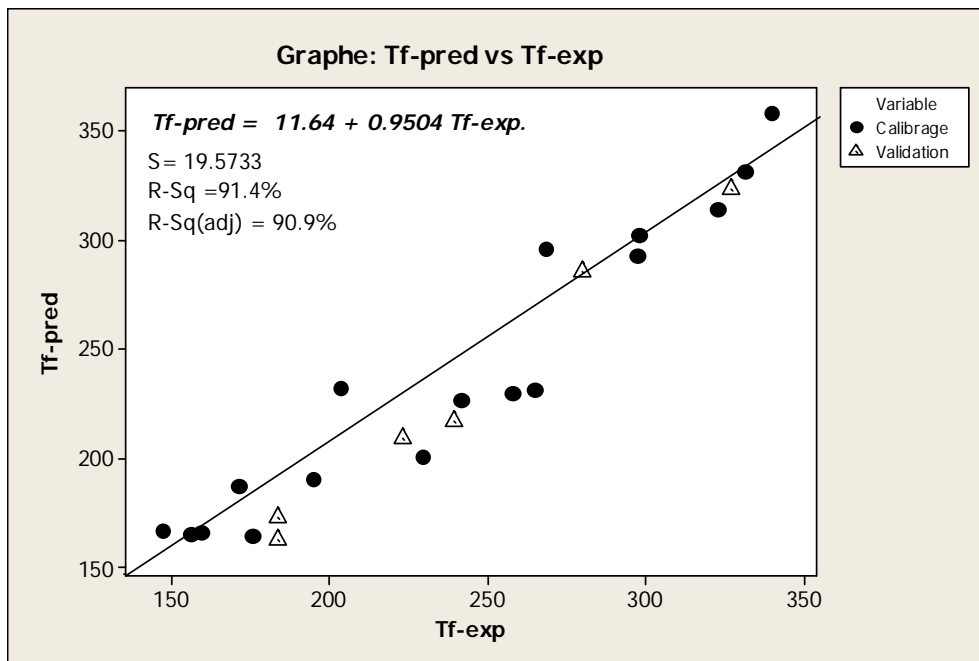


Figure 11 : Graphe des valeurs prédites Tf en fonction des valeurs observées.

I-8-Tests de randomisations :

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

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

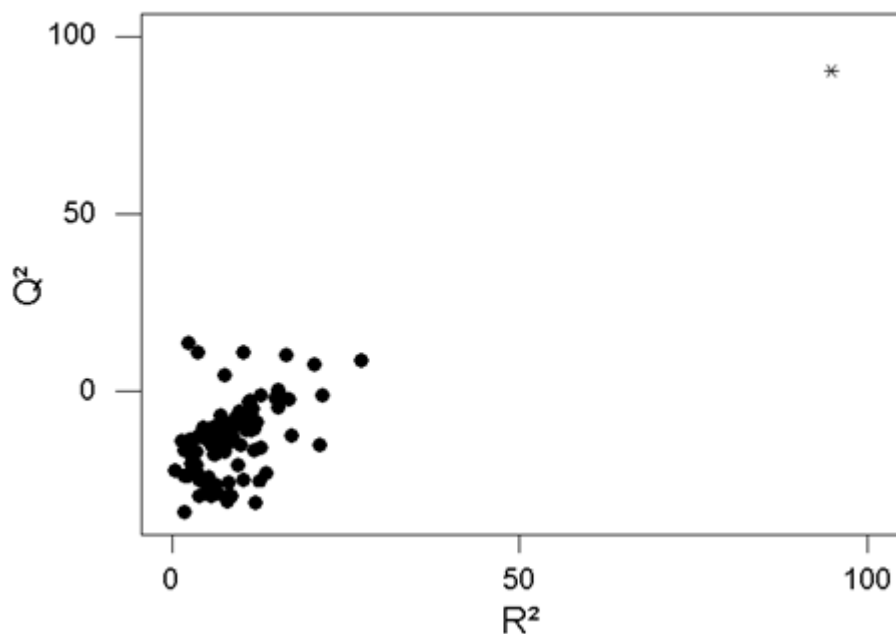


Figure 12: Test de randomisation associé au modèle QSPR.

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

Conclusion générale

Nous avons appliqué la méthodologie QSPR pour relier la propriété (Température de fusion) d'une série d'alcool, à des descripteurs moléculaires théoriques reflétant certaines particularités.

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

Les 23 données de base ont été éclatées aléatoirement en deux ensembles disjoints :

- un ensemble principal de 17 composés utilisés pour le calcul et, éventuellement, les essais du modèle ;
- un ensemble de 06 composés pour la prédiction externe.

Les statistiques ($R^2 > 90\%$; $RMSE < 20$; $Q^2 > 90\%$) calculées établissent la pertinence du modèle MLR développé.

L'analyse des résidus a permis de détecter une observation influente et même temps aberrante.

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

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

Ainsi, la température de fusion (T_f) peut être prédite à partir de la structure moléculaire des alcools en utilisant la modélisation linéaire MLR.

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

Références Bibliographiques

- [01]: J.J. Marano, G. D. Holder, *Ind. Eng. Res* , 1997, 36.
- [02]: H. Li, H. Higashi, K. Tamura, *Fluid Phase Equilib*, 2006, 239.
- [03]: R. C. Reid, J. M. Prausnitz, B. E. Poling. *The Properties of Gases and Liquids*, 4th ed; Mc Graw Hill. New York, 1987.
- [04]: K. A. M. Gasem, C. H. Ross, R. L. Robinson Jr. , *Can. J. Chem. Eng.* , 1993, 77.
- [05]: ChemDraw Release 7.0.1 Chemical Structure Drawing Standard, 2002.
- [06]: HyperchemTM Release 7.5 for windows, Molecular Modelling system, 2000.
- [07]: T. Thomas-Danguin. *Intensité olfactive des composés purs et de mélanges: application au masquage des odeurs*, Université Claude Bernard, Lyon, 1997, 224.
- [08]: R. Todeschini, V. Consonni, M. Pawan, DRAGON, Software for the calculation of Molecular Descriptors. Release 5.3 for Windows, Milano, 2005.
- [09]: N. Margossian, *Le règlement REACH - La réglementation européenne sur les produits chimiques*, Dunod / L'usine Nouvelle, Paris, 2007.
- [10]: A. Crum Brown, T.R. Fraser, on the connection between chemical constitution and physiological action. Part 1. On the physiological action of salts of the ammonium bases, derived from strychnia, brucia, thebia, codeia, morphia and nicotia. *Trans. Roy. Soc. Edinburgh*, 1868, 25, 151-203.
- [11]: L.P. Hammett, The effect of structure upon the reactions of organic compounds. Benzene derivatives. *J. Am. Chem. Soc*, 1937, 59, 96-103.
- [12]: A.C. Brown, and T.R. Fraser, On the connection between chemical constitution and Physiological Action; with special reference to the physiological action of the salts of the ammonium bases derived from Strychnia, Brucia, Thebaia, Codeia, Morphia, and Nicotia, *J. Anat. Physiol*, 1868, 2, 224-242.
- [13]: C. Hansch, A. Leo, D. Hoekmann, *Exploring QSAR: hydrophobic, electronic and steric constants*. Washington, DC: Am. Chem. Soc, 1995.

Références Bibliographiques

- [14]: S.M. Free, J.W. Wilson, A mathematical contribution to structure–activity studies. *J. Med. Chem.*, 1964, 7, 395-399.
- [15]: M. Grover, B. Singh, M. Bakshi, S. Singh, Quantitative structure- Property relationships in pharmaceutical research Part2. *Pharm. Sci. Tech. Today*, 2000, 3, 50-57.
- [16]: M. Grover, B. Singh, M. Bakshi, S. Singh, Quantitative structure–property relationships in pharmaceutical research–Part 1 *Pharm. Sci. Tech. Today*, 2000, 3, 28-35.
- [17]: A.R.D. Katritzky, C. Fara, R.O. Petrukhin, D.B. Tatham, U. Maran, A. Lomaka, M.Karelson, The present utility and future Potential for medicinal chemistry of QSAR/QSPR with whole molecule descriptors, *Curr. Top. Med. Chem.*, 2002, 2, 1333-1356.
- [18]: J. Ghasemi, S. Saaidpour, S.D. Brown, QSPR study for estimation of acidity constants of some aromatic acids derivatives using multiple linear regression (MLR) analysis. *J. Mol. Struct. (Theochem)*. 2007, 805, 27-32.
- [19]: P. Geladi, B.R. Kowalski, Partial Least Squares Regression: a Tutorial, *Anal. Chim. Acta.* 1986, 185, 1-17.
- [20]: C. A. Rice-Evans, N. J. Miller, G. Paganga. *Free Radic. Biol. Med.*, 1996, 20- 933.
- [21]: P.Cos, L. Ying, M. Calomme, J. H. Hu, K. Cimanga, B. Van Poel, L. Pieters, A. J. Vlietinck, D. Vanden Berghe. *J. Nat. Prod.*, 1998, 61- 71.
- [22]: P. Cristina, S. Ilonka, T. Bartek, *Rev. Génie Ind.*, 2009, 4- 25.
- [23]: R. Apak, S. Gorinstein, *Pure Appl. Chem.*, 2013, 85- 957.
- [24]: E. Rozoy, M. Araya-Farias, S. Simard, D. Kitts, J. Lessard. *Food Chem* , 2013, 138-1982.
- [25] : P, Le Cloirec, 2002, Introduction au traitement de l'air, Les techniques de l'ingénieur *Traité environnement* ,1-8.

Références Bibliographiques

- [26] : K, Saadi .Contribution á l'étude de la Relation structure chimique- odeur Utilisation de la technique Random Forest (Application à la famille des pyrazines), Mémoire de Magister. UNIVERSITÉ KASDI Merbah Ourgla, 2009,30.
- [27]: A. R. Katritzky, V. S. Lobanov, M. Karelson. CODESSA Reference Manual. University of Florida, Gainesville, 1994.
- [28]: V. Y. Nalimov. The Application of Mathematical Statistics to Chemical Analysis, Addison- Wesley, Reading, MA ,1962.
- [29]: R. Calcutt, R. Body. Statistics for Analytical Chemists. Champman & Hall, New York, 1983.
- [30]: J. C. Miller, J. N. Miller. Statistics for Analytical Chemistry. Ellis Horwood, New York, 1988.
- [31] : P. C. Meier, R. E. Zund. Statistical Methods in Analytical Chemistry. Wiley, New York ,1993.
- [32] : P. Dagnélie. Statistique Théorique Et Appliquée. Tomes 1 et 2. De Boeck & Larcier s. à, 1998.
- [33] : R. Tomassone, E. Lesquoy, C. Miller. La régression : nouveaux regards sur une ancienne méthode statistique. Masson, INRA ,1983.
- [34] : F. G., P. Legendre & D, blanchet. Borcard, 2008. Forward selection of explanatory variables. Ecology 89(9): 2623-2632.
- [35]: L. Chambers. Practical Handbook of Genetic Algorithms. Lewis Publishing ,1995.
- [36]: MINITAB Release 16.2.0.0 for Microsoft langage pack 2.
- [37] : R. Todeschini. MOBY DIGS Software for Multilinear Regression Analysis and Variable Subset Selection by Genetic Algorithm.Release for Windows. Milano Srl.
- [38] : N.R Draper, H. Smith, Applied Regression Analysis, Third Edition, Wiley series in Probability and Statistics, New York, 1998.

Références Bibliographiques

[39] : L. Eriksson, J. Jaworska, A.P. Worth, M.T.D. Perspective, 111(10), 2003, 1361-1375.

[40]:A. R. Katritzky, M. Kuanar, S. Slavov, and C. D. Hall, Quantitative Correlation of Physical and Chemical Properties with Chemical Structure: Utility for Prediction, Chem. Rev. 2010, 110, 5714–5789.

[41] : Les méfaits de l'alcool dans l'apprentissage chez les jeunes par gaspard muzama funzi, université paris V-D-U 2012.

[42] : mémoire Etude des propriétés physico-chimiques des liquides ioniques dans le but de leur utilisation dans un concentrateur solaire à Tlemcen le 09 Avril 2013

Introduction

Générale

Partie Théorique

Partie

Application

. Conclusion Générale

Résumé :

Un modèle QSPR a été développé pour la prédiction de la température de fusion. Les données, concernant 23 alcools, ont été séparées en deux sous-ensembles disjoints comprenant respectivement 17 éléments pour le calcul et le test (éventuel) du modèle, et 06 éléments pour sa validation externe.

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

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

Mots-clés:

Alcools – Température de fusion – Descripteurs moléculaires théoriques – Modèle QSPR.

Abstract

One model QSPR was developed for the prediction of the melting point. The data, concerning 23 alcohol's, were separate in two disjoined subsets including/understanding respectively 17 elements for calculates and the possible test of the models, and 06 elements for its external validation.

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

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

Key words:

Alcohol's - melting point – molecular descriptors theoretical – QSPR Model.

ملخص :

تم تطوير نموذج بطريقة ال-QSPR لتنبؤ درجة الذوبان

المعطيات الخاصة بـ 23 كحول تم تقسيمها الى مجموعتين الاولى تحتوي على 17 عنصر لحساب وتجريب النموذج اما الثانية تحتوي على 16 عنصر ل تصديق الخارجي للنموذج . المواصفات الجزيئية النظرية تم حسابها باستعمال برمجيات النمذجة الجزيئية المتوفرة في السوق. حجم النموذج تم تحديده عن طريق اخذ القيمة الملائمة لـ R^2 اما تحديد النماذج تم بطريقة Algorithmه Génétique. قيم المعالم الاحصائية R^2 , Q^2 , SDEC, SDEP, SDEPext (المتحصل عليها تؤكد تعلق النموذج المطور).

الكلمات الدالة :

كحول-درجة الذوبان- مواصفات جزيئية نظرية – نموذج QSPR