

République Algérienne Démocratique et Populaire

**Ministère de l'Enseignement Supérieur et de la
Recherche Scientifique**



CENTRE UNIVERSITAIRE D'EL-OUED

INSTITUT DES SCIENCES ET TECHNOLOGIE



Mémoire de fin d'étude

Présenté pour l'obtention du diplôme de

LICENCE ACADEMIQUE

Domaine : Sciences et technologie

Filière : Génie de procédés

Spécialité : Génie de procédés

Thème

**Etude Q.S.P.R du Point D'ébullition Pour Les
Composées Organiques (Alcanes) en Utilisant les
Descripteurs Moléculaires.**

Presenté par :

HASSI ZOHRA

LECHELAH SIHAM.

Encadré par:

M^{er}.SOUYEI Belgacem

Discuté par:

M^{me}. ZOUARI Ahmed Rachida

2011-2012

Remerciements

En premier lieu, nous tenons à remercier notre DIEU, notre Créateur pour m'avoir donné la force pour accomplir ce travail.

Nous tenons à notifier un remerciement spécial à tous nos professeurs .

Nous tenons à remercier Mr. SOUYEI BELGACEM notre promoteur pour son grand soutien et ses conseils considérables.

Nous remercions également tous les professeurs de CENTRE UNIVERSITAIRE D'EL-OUED et surtout les professeurs de département de génie des procédés.

Qu'ils trouvent ici toute notre gratitude et toute notre reconnaissance pour leur aide, leur encadrement dont j'ai pu bénéficier durant toute notre scolarité.

Nous désirons aussi, exprimer nos essentiels remerciements :

, Mr. LAOUINI Salaheddine , tous les professeurs de génie des procédés.

Enfin nos remerciements vont à tous ceux qui ont contribué de près ou de loin pour l'aboutissement de ce travail.

Liste des Figures

Figure	Titre	page
figure .I .1	Le point d'ébullition	4
Figure .I. 3	Influence du point d'ébullition sur d'autres propriétés physico-chimiques.	6
Figure. I .4	Outil de prédiction des propriétés physico-chimique	9
Figure. II.1	Montre la meilleure droite de régression linéaire Simple	14
Figure. II.1	Montre la meilleure droite de régression linéaire Simple	14
Figure. II.2	Montre la réalisation de valeur ϵ_i et la droite de régression	14
Figure. III.1	Les étapes de travail (Diagramme de travail)	20
Figure .III.2	relation <i>descripteurs moléculaires Enter QSPR et point d'ébullition</i>	21
Figure. III.3	décrit les étapes de la préparation de la base de données	22
Figure. III.4	résulta graphique de XLSTAT (regression)	23

Liste des Tableau

Tableaux	Titre	Page
Tableau III.1	Montre les résultats de calcul des descripteurs	20

SOMMAIRE

<i>Introduction générale</i>	1
------------------------------	---

Chapitre I : Définition et détermination expérimentale du point d'ébullition

I-1-Définition du point d'ébullition	4
I-1-1- Définition générale	4
I-1-2- Définition du point d'ébullition	4
I-1-3le point d'ébullition normal	5
I-1-4- La température de saturation	5
I.1.5. Point d'ébullition comme une propriété de référence d'un composé pur	6
I-2-Influence du point d'ébullition sur d'autres propriétés physico-chimiques	7
I.3.Les méthodes (QSAR/QSPR)	7
I.3.1. Définition de QSAR / QSPR	7
I.3.2. Méthode quantitative structure-propriété relation (QSPR)	8
I.4. Diagramme montre les étapes de prédiction	10
I.5.L'objectif de la prédiction	11
I.6. Les descripteurs moléculaires	11
I.6.1. La sélection des descripteurs moléculaire	10
I.6.2. Mapping (plan)	12

Chapitre II : La Régression Simple et Multiple

II-1-La régression Linéaire Simple (RLS)	14
II-2 Régression Linéaire Multiple (RLM)	16
II-2-1 Principes de la Régression Linéaire Multiple	17
II-3 Valider un modèle	17

Chapitre III : Diagramme de travail et la base des données

III-1- Les étapes de travail	19
III.1. 1. <u>Modélisation</u>	19
III.2. Base des données	20
III.2.1 <u>source</u>	20
III-2-2 <u>nature</u>	20
III.3. Génération des descripteurs moléculaires	20
III .3.1. Propriétés moléculaires	21
III .4. Organigramme	21
III .4.1. La préparation la base des données	22
III.4.2. définition Logiciel DRAGON	22
III.4.3. Travail par Logiciel DRAGON	23
III .5. Logiciel XLSTAT	23
III.5.1. définition Logiciel XLSTAT	28
III .5. 2. Travail par Logiciel XLSTAT	28
III-4-3- Logiciel XLSTAT	28
III .6.Déterminer les éléments au-delà du domaine de confiance	28

Chapitre IV : Résultats et discussion	
IV-1- les résultats du bloc topologique charge indices	31
IV-1-1-Equation du modèle	31
IV-1-2-Coefficients d'ajustement	31
IV-1-3-Les paramètres du modèle	31
IV-1-4-Analyse des résultats	32
IV-1-5- discussion	32
IV-2- les résultats du bloc connectivite index	32
IV-2-1-Equation du modèle	32
IV-2-2- Les paramètres du modèle	33
IV-2-3- Coefficients d'ajustement	33
IV-2-4-Analyse des résultats	34
IV-2-5- discussion	34
IV-3-Analyse globale	35
IV-4-Discussion générale	35
Conclusion générale	38
Références bibliographique	39

Nomenclature

Symboles

log P

QSPR

BP

y_p

RLS

MLS

TPSAL

Définitions

coefficient de partage

méthodes quantitatives relation structure-activité

le point d'ébullition

la propriété cible

régression linéaire simple

modèle linéaire simple

topologique surface polaires

Introduction générale

INTRODUCTION :

Aujourd'hui, l'industrie chimique joue un rôle majeur et divers dans notre vie, ce rôle se diffère suivant la nature des composés chimiques utilisés. Il existe plusieurs variétés de composés chimiques classifiés comme produits dangereux (Toxiques), qui ont un impact sérieux sur la santé de l'homme et de l'environnement.

D'après tout cela, on comprend qu'il y a une nécessité urgente et sérieuse qui demande et exige le traitement avec une grande prudence et exactitude de ces composés chimiques. Cela ne peut se faire qu'après avoir élaboré une manière pour l'étude et le calcul avec précision des propriétés physico-chimiques qui ont une influence sur l'environnement comme la volatilité, la solubilité, **la température d'ébullition**, la pression de vapeur, le coefficient de partage et l'adsorption. C'est pour cette raison que nous avons proposé ce travail qui permettra de simuler les propriétés précitées sans toutefois faire des manipulations aux laboratoires.

La propriété choisie pour l'étude est **la température d'ébullition**.

Premièrement et facilement à mesurer en tant que propriété d'un Composé, le point d'ébullition normal (NBP) est utilisé pour caractériser et d'identifier un nouveau composé, et d'estimer d'autres propriétés physiques, telles que les températures critiques, les pressions de vapeur, les enthalpies de vaporisation, etc [1-2]. Il est bien connu que NBP est déterminé par les interactions intermoléculaires dans le liquide et par la différence de fonction de partition moléculaire interne entre les phases gazeuse et liquide au point d'ébullition. Par conséquent, le NBP dépend indirectement de la structure chimique des molécules. Il n'est pas surprenant que de nombreuses méthodes ont été développées pour estimer la NBP d'un composé de la structure [1].

Le modèle de la relation quantitative structure-propriété(QSPR) a été tout à fait largement rapporté dans la littérature pour prédire le point d'ébullition [3-4].

Un point clé dans les études QSPR est la caractérisation simple et précise des caractéristiques structurales qui sont associées à la propriété observée, connue sous le nom des descripteurs moléculaires [4,5]. La plupart des descripteurs topologiques sont caractérisés par fixe des valeurs numériques, qui sont indépendantes de la propriété considérée [6]. Basé sur l'idée que d'un paramètre variable subit un changement lors de l'analyse de régression, Randić proposé l'indice de connectivité variable [7,8], qui présente amélioré les résultats de la régression. Dans ces descripteurs, l' poids optimaux qui représente les contributions relatives à atomes individuels sont optimisées au cours de la procédure de modélisation mieux pour répondre à la propriété modélisée.

Introduction générale

Le but de ce travail, est concentré sur le développement d'un modèle mathématique basé sur les données expérimentales disponibles et en fonction de la structure moléculaire, permettant ainsi d'estimer, d'étudier et d'identifier la propriété physico-chimiques notamment **la température d'ébullition**, qui est un paramètre clé pour l'évaluation de l'impact des composés chimiques sur l'environnement.

Ce travail a été réalisé dans le laboratoire de valorisation et promotion des ressources sahariennes pour la modélisation des produits et procédés assistés par ordinateur.

Pour accomplir notre travail, ce mémoire sera divisé en quatre chapitres:

Le premier chapitre, est consacré à des généralités sur les produits chimiques toxiques, les définitions des termes techniques utilisées à savoir de BP, QSAR/QSPR.

Dans le deuxième chapitre, nous développerons un rappel théorique sur la **Régression simple, régression multiple et la validation**, ceci est très important dans le cadre des développements des relations mathématiques à partir des données.

Dans le troisième chapitre nous décrivons le formalisme de la méthodologie choisie, le diagramme (les étapes) de travail effectué.

Le quatrième chapitre sera consacré à l'interprétation des résultats de nos calculs sur la prédiction du point d'ébullition (**BP**).

I- Définition et détermination du point d'ébullition

Introduction:

La température plus élevée que peut atteindre un corps avant de s'évaporer, sous forme gazeuse, librement, se nomme le *point d'ébullition*. Cette température se calcule à la pression atmosphériques de 1 atmosphère (101,3 kPa).[10] . Le point d'ébullition d'un corps pur est défini par un couple de valeurs dans son diagramme de phase : la **température d'ébullition** et la **pression d'ébullition**. Dans le diagramme des phases, ce point est situé sur la ligne séparant la phase gazeuse de la phase liquide. Il est donc caractérisé par les grandeurs physiques pression et température lors de la transition entre l'état liquide et l'état gazeux du corps pur en question Pour l'eau, le point d'ébullition est de 100 °C. Le changement d'état est soumis à une chaleur de vaporisation (égale environ à 2 250 J/g pour l'eau).Il est possible de définir de façon équivalente le point d'ébullition comme étant la température à laquelle la pression de vapeur du liquide est de 1 atmosphère.

I.1.Définition du point d'ébullition:

I.1.1. Définition générale :

Le point d'ébullition réunit donc les conditions sous les quelles il peut y avoir une transition de phase de l'état liquide à l'état gazeux d'un corps pur, ce qu'on qualifie d'ébullition ou d'évaporation. De surcroît, le point d'ébullition joue le rôle inverse du point de condensation. Dans le cas du corps pur, ces deux points sont confondus. Lorsqu'on porte à ébullition un mélange de corps purs, le comportement diffère et l'on observe alors une zone d'ébullition au lieu d'un seul point. Si la transition de la phase liquide à la phase gazeuse se fait au-dessous du point d'ébullition, on parle d'évaporation.[10] Dans les tables de thermodynamique, les températures d'ébullitions sont données dans l'état standard, c'est-à-dire à 1 013,25 kPa. Ce point d'ébullition est alors appelé **point d'ébullition standard**, et la température **température d'ébullition standard**. Le terme "point d'ébullition" est souvent employé pour désigner la température d'ébullition standard dans le langage courant, bien qu'il réduise le couple de valeurs à une seule valeur, ce qui est un abus de langage. Une application courante de l'interdépendance entre température d'ébullition et pression d'ébullition est l'autocuiseur. C'est grâce à une augmentation de la pression (couramment de l'ordre du bar) que l'on peut faire passer la température d'ébullition de l'eau de 100 °C à environ 120 °C. Ces deux températures correspondent bien à des températures d'ébullition. Cependant, seule la valeur de 100 °C est une valeur prise dans l'état standard, et par la température standard d'ébullition de l'eau. C'est pourquoi la confusion des deux dénominations est impropre, pas du tout naturelle et doit être évitée.[10]

I.1.2. le point d'ébullition:

Le **point d'ébullition** d'une substance est la température à laquelle la pression de vapeur du liquide est égale à la pression de l'environnement entourant le liquide. Un liquide dans un vide a un point d'ébullition inférieur lorsque ce liquide est à la pression atmosphérique. Un liquide à haute pression a un point d'ébullition plus élevé

que lorsque le liquide est à la pression atmosphérique. En d'autres termes, le point d'ébullition d'un liquide varie en fonction de la pression environnante de l'environnement [9].

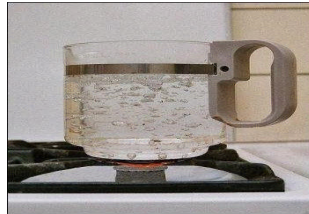


figure I-1: le point d'ébullition de l'eau.

I.1.3.le point d'ébullition normal :

Le point d'ébullition atmosphérique ou (le **point de pression d'ébullition atmosphérique**) d'un liquide est le cas particulier où la pression de vapeur du liquide est égale à la pression atmosphérique. A cette température, la pression de vapeur du liquide devient suffisante pour surmonter la pression atmosphérique et permettre aux bulles de vapeur pour former à l'intérieur de la masse du liquide. Le **point d'ébullition normal** est présent (en 1982) définie par l'IUPAC que la température à laquelle se produit d'ébullition sous une pression de 1 bar. La chaleur de vaporisation est la quantité d'énergie nécessaire pour convertir ou de vaporiser un liquide saturé (c'est à dire un liquide à son point d'ébullition) en vapeur . Liquides peuvent changer à une vapeur à des températures inférieures à leur point d'ébullition à travers le processus de L'évaporation. L'évaporation est un phénomène de surface dans lequel des molécules situées près du bord du liquide. [9]

I.1.4.La température de saturation:

signifie *le point d'ébullition* . La température de saturation est la température pour une pression de saturation correspondant à laquelle un liquide bout dans sa phase vapeur . Le liquide peut être dite pour être saturé avec de l'énergie thermique. Aucune addition des résultats de l'énergie thermique dans une transition de phase [1]. Si la pression dans un système reste constante (isobare), une vapeur à la température de saturation commence à se condenser dans sa phase liquide sous forme d'énergie thermique est enlevé. De même, un liquide à température de saturation et de la pression à bouillir.

Ainsi, le point d'ébullition est dépendante de la pression. Habituellement, les points d'ébullition sont publiés à l'égard de la pression atmosphérique (101,325 kilo pascals ou 1atm). A des altitudes plus élevées, où la pression atmosphérique est beaucoup plus faible, le point d'ébullition est également plus faible. Les augmentations du point d'ébullition avec une pression accrue à le point critique , où le gaz et les propriétés liquides deviennent identiques. Le point d'ébullition ne peut pas être augmenté au-delà du point critique est atteint. Le point d'ébullition ne peut pas être réduite au-dessous du point triple. De même, le point d'ébullition diminue avec la pression jusqu'à ce que le point triple .Si la chaleur de vaporisation et la pression de vapeur

d'un liquide à une certaine température est connue, le point d'ébullition normal peut être calculée en utilisant l' équation de Clapeyron ainsi:

$$T_B = \left(\frac{R \ln(p_0)}{\Delta H_{vap}} + \frac{1}{T_0} \right)^{-1}$$

où:

T = Le point d'ébullition normal, (K)

R = La constant de gaz parfait, $8,314 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$

P = Est la pression de vapeur à une température donnée, (atm)

ΔH = La chaleur de vaporisation du liquide, (J / mole)

T = La temperature donnée, (K)

ln = Le logarithme naturel de la base **e**

1.1.5. Le Point d'ébullition comme une propriété de référence d'un composé pur:

Comme on peut le voit depuis le terrain au-dessus du logarithme de la pression de vapeur par rapport à la température pour tout compose chimique donné pur de son point d'ébullition normal peut servir comme une indication de l'ensemble que le composé de la volatilité. Un composé donné pur a un seul point d'ébullition normal, le cas échéant, et un composé de la normale à point d'ébullition et le point de fusion peut servir caractéristiques des propriétés physique de ce composé, énumérés dans les livres de référence. Le composé supérieur de point d'ébullition normal, l'moins volatil que le composé est dans l'ensemble, et inversement, l'un composé inférieure de point d'ébullition normal, plus volatil que le composé est dans l'ensemble. Certains composés se décomposer à des températures plus élevées avant d'atteindre leur point d'ébullition normal, ou parfois même leur point de fusion. Pour un composé stable, la température d'ébullition comprise entre son point triple de son point critique , en fonction de la pression externe. Delà de son point triple, un composé de la normale à point d'ébullition, le cas échéant, est supérieure à son point de fusion. Au-delà du point critique, les phases liquide et vapeur d'un composé fusionner en une seule phase, qui peut être appelé un gaz surchauffé. A une température donnée, s'il est normal d'un composé à point d'ébullition est plus faible, alors que le composé sera généralement exister en tant que gaz à la pression atmosphérique extérieure. Si normale du composé à point d'ébullition est plus élevé, puis ce composé peut exister sous forme liquide ou solide à la température que donnée à la pression atmosphérique extérieure, et de sorte sont en équilibre avec sa vapeur (si volatile) si les vapeurs sont contenus. Si les vapeurs d'un composé ne sont pas contenues, puis certains composés volatils peuvent finissent par s'évaporer l'écart en dépit de leurs points faibles [9].

I-2-Influence du point d'ébullition sur d'autres propriétés physico-chimiques:

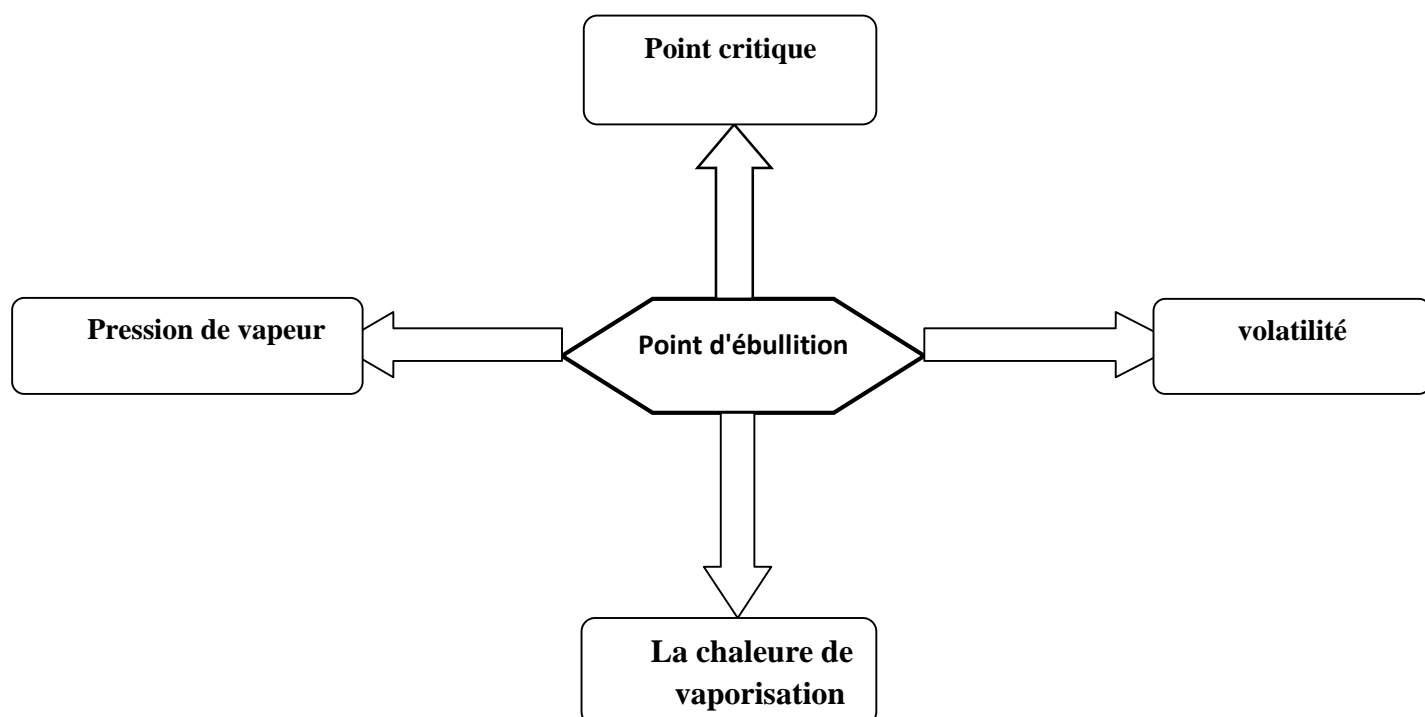


Figure I.2 : Influence du point d'ébullition sur d'autres propriétés physico-chimiques .

I.3.Les méthodes (QSAR/QSPR)

I.3.1. Définition de QSAR / QSPR :

Quantitatives structure-activité et propriété de la structure les études de relations (QSAR / QSPR) sont sans aucun doute d'une grande importance dans la chimie moderne et de la biochimie. Le concept de QSAR / QSPR est de transformer les recherches pour les composés avec souhaitée en utilisant les propriétés chimiques intuition et l'expérience en mathématique et informatique quantifiés formulaire. Une fois une corrélation entre la structure et activité / établissement se trouve, un certain nombre de composés, y compris ceux non encore synthétisé, peut être facilement projeté sur l'ordinateur afin de sélectionner les structures avec les propriétés souhaitées. Il est alors possible de sélectionner les composés les plus prometteurs pour synthétiser et de tester en laboratoire. Ainsi, les QSAR / QSPR approche qui préserve

les ressources et accélère la processus de développement de nouvelles molécules pour les utiliser comme médicaments, les matériaux, les additifs, ou pour toute autre fin.

1.3.2. Méthode quantitative structure-propriété relation (QSPR):

Les données expérimentales de la propriété ne sont disponibles que pour une petite fraction des composés purs ayant trait à la chimie, génie chimique, génie de l'environnement et l'évaluation de l'impact de l'environnement ,analyse de risque opérationnelle, etc. Par conséquent, des méthodes de prédiction fiable de données de propriété sont nécessaires.

L'introduction des descripteurs topologique, courbe théorique, topochemique, chimie quantique, structure, physico-chimiques, electrotopologique , et d'autres descripteurs moléculaires ont permis considérables amélioration de la présentation des structures chimiques et par conséquent, de l'exactitude des prédictions de propriétés.

QSPRs ont été développés afin de profiter de nouveaux descripteurs moléculaires sophistiqués pour la prédiction des propriétés physico-chimiques. Une étude récente a été publié dans ce domaine par Dearden.[12]

Dans cette étude, nous allons concentrés sur la plupart des caractéristiques importantes communes des méthodes **QSPR**, telles que définies par Wakeham et d'autres [13].

La méthode **QSPR** peut être schématisé par l'équation suivante:

$$y_p = f(x_{s1}, x_{s2}, \dots, x_{sk}, x_{p1}, x_{p2}, \dots, x_{pm}, \beta_0, \beta_1, \dots, \beta_n) \quad (I - 5)$$

où $x_{s1}, x_{s2}, \dots, x_{sk}$ sont les descripteurs de la structure moléculaire des composés particuliers pures, $x_{p1}, x_{p2}, \dots, x_{pm}$ sont mesurables propriétés d'un même composé (tels que la température d'ébullition, la température de fusion, la toxicité, etc.), $\beta_0, \beta_1, \dots, \beta_n$ sont les **QSPR** paramètres, et y_p est la propriété cible (à prévoir) de le même composé.

Pour calculer la **QSPR**, les données disponibles sont divisées en un "ensemble d'apprentissage" et un "ensemble de validation". En utilisant l'ensemble d'apprentissage pour la régression linéaire multiple ou non linéaire et des techniques de moindres carrés partiels sont utilisées pour sélectionner les descripteurs moléculaires et / ou les propriétés à être inclure dans la partie droite de l'équation (I – 5) et de calculer les valeurs des paramètres du modèle. La validation du modèle

est réalisée en utilisant l'ensemble de validation. Applications avec **QSPRs** utilisés pour la prédiction des propriétés physiques et thermodynamiques sont rapportés par de nombreux auteurs.[13.14]

Ne faut pas s'attendre prédire de façon fiable la propriété modélisée pour l'univers entier de produits chimiques. Sur la base de cette observation, ils ont recommandé la définition d'un "domaine d'applicabilité du modèle" afin de obtenir une prévision fiable de propriétés le domaine d'applicabilité du modèle est déterminé par la variation des structures chimiques dans la base de données et testé par l'ensemble d'apprentissage. Par conséquent, la sélection des composés dans le jeu de la formation ne doit pas être faite au hasard, mais il devrait contenir des composés structurellement liée à le composés ciblé, dont les propriétés sont à prévoir. En outre, de nouvelles méthodes pour l'identification de structure apparentée composés sont nécessaires non seulement pour les ensembles d'apprentissage, mais aussi pour la conception assistée par ordinateur moléculaire des composés structurellement semblable à un composé dont on connaît des propriétés intéressantes. Ces problèmes et les approches pour traiter avec eux ont été récemment passés en revue par Basak et d'autres.[15] Tropsha et d'autres.[16]

Une nouvelle **QSPRs**, est basé sur l'hypothèse que la structure (le vecteur de descripteurs moléculaires) d'un composé cible peut être représenté par une combinaison linéaire des vecteurs de descripteurs moléculaires des composés structurellement liés prédictive (à structure-structure " corrélation ")Shacham et d'autres.[17] et Brauner et d'autres.[18] .

Une régression linéaire pas à pas algorithmes a été utilisée pour identifier les paramètres de la structure-structure corrélation. Les paramètres de cette corrélation sont ensuite utilisés dans les corrélations propriété des biens, capable de prédire toutes les propriétés nécessaires d'un composé cible et de fournir une évaluation réaliste estimation de la prédiction.

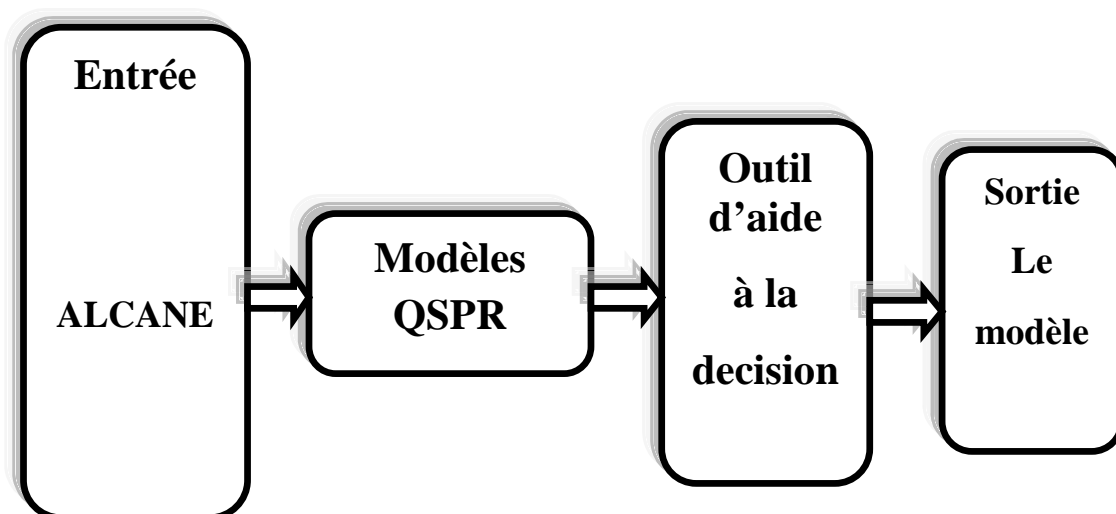
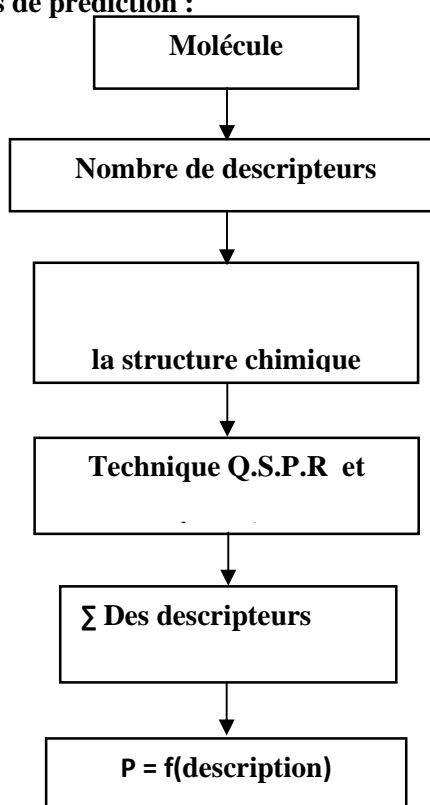


Figure I-3: Outil de prédiction des propriétés physico-chimique

I.4. Diagramme montre les étapes de prédiction :



I.5.L'objectif de la prédiction :

L'objectif principal est d'établir un modèle de prédiction de(**Bp**)en se basant sur la structure moléculaire et plus précisément les descripteurs moléculaires. Ce modèle permet, en suite la classification des composées selon la nature de pollution sans avoir recours à l'expérimental.

Le cycle de prédiction est :



I.6. Les descripteurs moléculaires :

Un descripteur moléculaire est le résultat final d'une logique et d'un procédé mathématique qui transforme l'information chimique codée dans une représentation symbolique d'une molécule en nombre utile ou résultat d'une certaine expérience normalisée [19]. Les descripteurs moléculaires jouent un rôle fondamental dans la chimie, les sciences pharmaceutiques, la politique de protection de l'environnement, la recherche de santé et le contrôle de qualité, étant obtenu quand des molécules sont transformées en représentation moléculaire permettant un certain traitement mathématique. On a proposé beaucoup de descripteurs moléculaires qui dérivent de différentes théories et approches dans le but de prévision des propriétés biologiques et physico-chimiques des molécules. [20] Les descripteurs moléculaires sont calculés par des logiciels conçus à cet effet, telle que le logiciel **DRAGON version 5.5** qui possède 20 blocs de Descripteurs qui sont [21]

- ❖ **Descripteurs constitutionnels (Constitutional descriptors (block 1))**
- ❖ **Descripteurs topologiques (Topological descriptors (block 2))**
- ❖ **Comptes de promenade et de chemin (Walk and path counts (block 3))**
- ❖ **Indices de connectivité (Connectivity indices (block 4))**

- ❖ Indices de l'information (Information indices (block 5))
- ❖ autocorrélations 2D (List of 2D autocorrelation indices (block 6)-
- ❖ Indices de contiguïté de bord (Edge adjacency indices (block 7))
- ❖ Descripteurs de valeur propre de fardeau (Burden eigenvalue descriptors (block8))
- ❖ Indices de charge topologiques (Topological charge indices (block 9))
- ❖ Indices de Valeur-basés (Eigenvalue-based indices (block 10))
- ❖ Profils moléculaires de Randic (Randic molecular profiles (block 11))
- ❖ Descripteurs géométriques (Geometrical descriptors (block 12))
- ❖ Descripteurs de RDF (RDF descriptors (block 13))
- ❖ Descripteurs 3D-MoRSE (3D-MoRSE descriptors (block 14)-
- ❖ Descripteurs Etranges (Bisard)(WHIM descriptors (block 15))
- ❖ Descripteurs de manière d'obtention (GETAWAY descriptors (block 16))
- ❖ comptes de groupe fonctionnels (Functional group counts)
- ❖ fragments de l' atome centrale (Atom-centred fragments (block 18))
- ❖ Descripteurs du charge (Charge descriptors (block 19))
- ❖ Propriétés moléculaires (Molecular properties (block 20))

I.6.1. La sélection des descripteurs moléculaire :

La sélection des descripteurs: implique de trouver le sous ensemble le plus informatif de descripteurs des modèles . L'algorithme génétique (GA) étés présenté pour être plus effectif dans la sélection des descripteurs performants dans le cas d'une large gamme de descripteurs .

I.6.2. Mapping (plan):

implique l'application des descripteurs utilisant les analyses de régression multi linéaire tel que heuristique et meilleur régression multilinéaire par méthodes (BMLR) .

méthodes multi variables statistique tel que les analyses de composant principal (P.C.A) et la moindre carré partiel (P.L.S) .

les méthodes non linéaires tel que réseaux de computational neural (nerveux) (NNS) pour construire les méthodes mathématiques et d'enchaîner directement les descripteurs au propriété a investiguer. [22]

Conclusion:

Dans ce chapitre nous avons donné une définition générale et spéciale du point d'ébullition. Et nous avons vu la méthode QSPR, les étapes de prédiction, l'objectif de prédiction et les descripteurs moléculaires.

II- La Régression Simple et Multiple

Introduction:

Dans ce chapitre nous avons étudiés l'analyse statistique qui est basé sur la régression simple et multiple pour avoir obtenir un modèle mathématique suivit d'une validation de cet modèle.

II.1. La régression Linéaire Simple (RLS):

La Régression Linéaire Simple (**RLS**) est un modèle particulier de régression dans lequel :

- ❖ Il n'y a qu'une seule variable explicative (numérique),
- ❖ Le modèle est linéaire dans la variable dans les paramètres.

Comme toutes les techniques prédictives, elle a deux objectifs :

Construire un modèle dont les paramètres soient interprétables par le praticien en termes de propriétés de la population dont est extrait l'échantillon. On espère bien entendu que les paramètres du modèle seront de bons estimateurs des paramètres de la population sous-jacente.

- ❖ Utiliser le modèle pour faire des prédictions.

La Régression Linéaire Simple traite la question suivante :

- ❖ Une grandeur y est mesurée
- ❖ Pour un certain nombre de valeurs d'une autre grandeur x .

Ces mesures se traduisent par le diagramme de dispersion suivant:

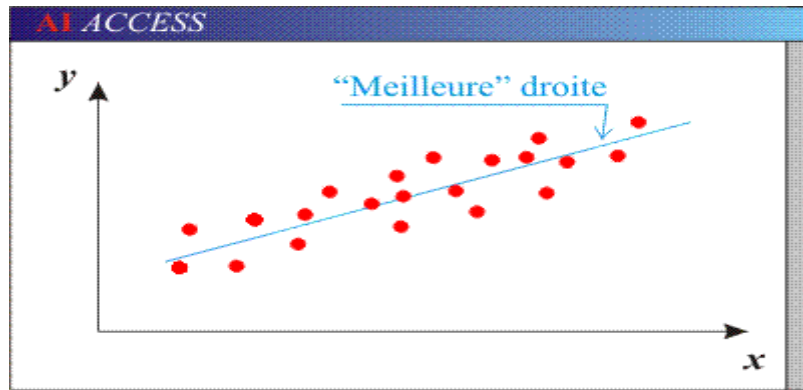


Figure II.1: Montre la meilleure droite de régression linéaire Simple.

Le Modèle Linéaire Simple (MLS) précise les conditions suivantes :

1) Il existe une droite d'équation :

$$y = \alpha + \beta x \quad (\text{II-1})$$

Telle que, pour chaque valeur x_i de x , la valeur y_i de Y_i est obtenue de la façon suivante :

- ❖ La hauteur de cette droite (donc y) est la valeur de la mesure si celle-ci n'était pas entachée d'erreur.
- ❖ Mais à cette valeur "parfaite" vient s'ajouter une **erreur de mesure**, réalisation d'une valeur. ε_i de Moyenne nulle et de variance σ^2 .

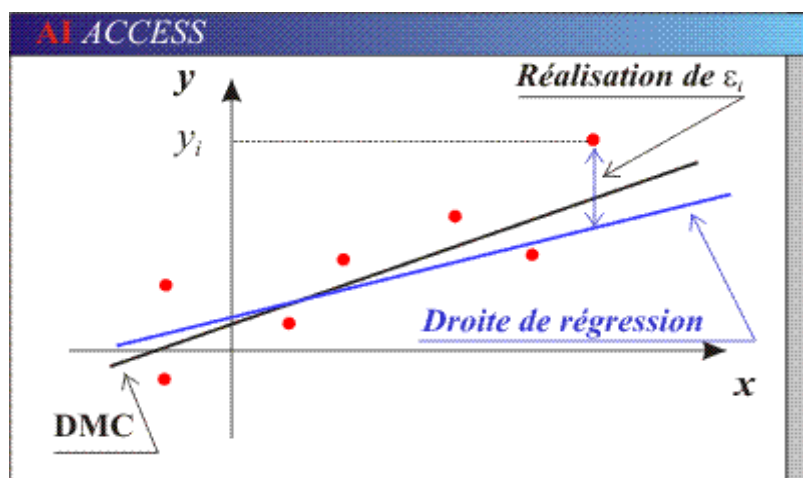


Figure II.2 :Montre la réalisation de valeur ε_i et la droite de régression

Cette droite (inconnue) s'appelle la **droite de régression** du mécanisme qui a engendré les données. Dans la littérature, la Droite des Moindres Carrés est souvent, appelée Droite de Régression. Nous réservons cette appellation à l'authentique droite de régression. Il est par contre justifié de baptiser la DMC "droite de régression empirique".

Les valeurs. Y_i sont donc définies par :

$$Y_i = \alpha + \beta x_i + \varepsilon_i \quad (\text{II-2})$$

Les ε_i étant de moyenne nulle, la moyenne de Y_i (pour un i donné) est égale à $\alpha + \beta x_i$. Les valeurs, y_i ont toutes leur moyenne sur la droite de régression (ce qui est d'ailleurs la définition de la régression). Le **MLS** est donc défini par trois paramètres :

- β , la pente de la droite de régression,
- α , l'ordonnée à l'origine de la droite de régression,
- σ^2 , la variance des erreurs.

L'objectif central de la RLS est de trouver de bons estimateurs de ces quantités ainsi que d'identifier les liens de dépendance (covariance) entre ces estimateurs.

II.2 Régression Linéaire Multiple (RLM) :

La **RLM** est une généralisation de la Régression Linéaire Simple (RLS) au cas où il y a plusieurs variables explicatives : on cherche alors dans un espace à plus de deux dimensions la "meilleure droite" passant dans un nuage de points de forme si possible effilée et rectiligne.

Dans la régression Linéaire Multiple. Le caractère linéaire dans les variables et les paramètres du modèle peut être maintenu lorsque y dépend non pas d'une seule variable x , mais de plusieurs variables $\{x_1, x_2, \dots, x_n\}$ [23].

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_px_p \quad (\text{II-3})$$

II.2.1 Principes de la Régression Linéaire Multiple:

Le principe des moindres carrés appliqué à l'hyperplan formé par l'ensemble des variables explicatives (chaque variable constitue un axe de l'hyperplan) est le même que pour le test de régression linéaire simple. Seules les méthodes de calcul changent car on fait ici appel aux calculs matriciels pour lesquels nous ne rentrerons pas dans les détails.

Toutefois, le calcul du vecteur colonne des paramètres du modèle se calcule de la façon suivante :

$$\beta = (X'X)^{-1} X'y \quad (\text{II-4})$$

où " X " est la matrice ($n \times p$) des données relatives aux variables explicatives avec une colonne de 1 pour le calcul de la constante, " y " est le vecteur colonne ($p \times 1$) des données relatives à la variable expliquée, n est le nombre d'observations et p le nombre de paramètres du modèle.

La conditions d'utilisation du test est :

La distribution normale des mesures sur chaque variable étudiée [24]

II.3 Valider un modèle :

Rappelons que la qualité réelle d'un modèle ne peut pas être estimée au vu de ses résultats sur les données qui ont servi à le construire, sauf cas exceptionnel (p. ex. Régression Linéaire dans les conditions idéales). En fait, beaucoup de types de modèles peuvent, lorsqu'ils sont

mal utilisés, conduire à des résultats excellents (et totalement illusoire) sur les données "de construction", et des résultats catastrophiques sur les données "d'utilisation".

Valider un modèle est difficile, et la qualité de la validation dépend grandement du temps que l'on est prêt à y consacrer. Les approches classiques sont :

- ❖ La théorie, qui, dans les cas les plus simples, permet d'estimer les performances en généralisation.
- ❖ Diverses techniques de simulation des conditions réelles d'utilisation à partir des données disponibles. Citons en particulier .
 - L'utilisation d'un ou de plusieurs ensemble de validation aléatoirement choisis.
 - L'utilisation de plusieurs ensembles de validation choisis de façon particulière par les techniques dites de "Validation Croisée", de "Leave One Out" et plus récemment, de "Bootstrap". Ces techniques (dites de "ré-échantillonnage") permettent en fait d'estimer le pouvoir de généralisation d'un certain type d'architecture de modèle, plutôt que d'un modèle particulier.
- ❖ Des approximations mathématiques de "linéarisation de modèle dans l'espace des paramètres", que nous ne pouvons malheureusement pas aborder dans ce glossaire, et qui donnent toute leur puissance dans le contexte des Réseaux de neurones supervisés.

Note : La phase de validation est parfois appelée phase de "test".[24]

Conclusion:

A prés les activités de la préparation de la base des données, nous avons arrivés premièrement a un ensemble de valeurs expérimentales du point d'ébullition de 170 composes alcanes et un ensemble de 54 descripteurs moléculaires a l'aide du programme **DRAGON version 5.5** .

III – la méthode de travail et la base des données

Introduction:

Pour la réalisation de cette étude il faut arriver premièrement la préparation de la base des données, qui est représenté par les valeurs expérimentales de point d'ébullition de ses alcanes suivit d'une modélisation **MM+** et **AM1** par **Hyperchem** et les descripteurs moléculaires qui sont sélectionnés par logiciel les logiciel **DRAGON** version **5.5**.

Construction d'un plan de travail illustrer les étapes nécessaires pour atteindre la base des données présentent.

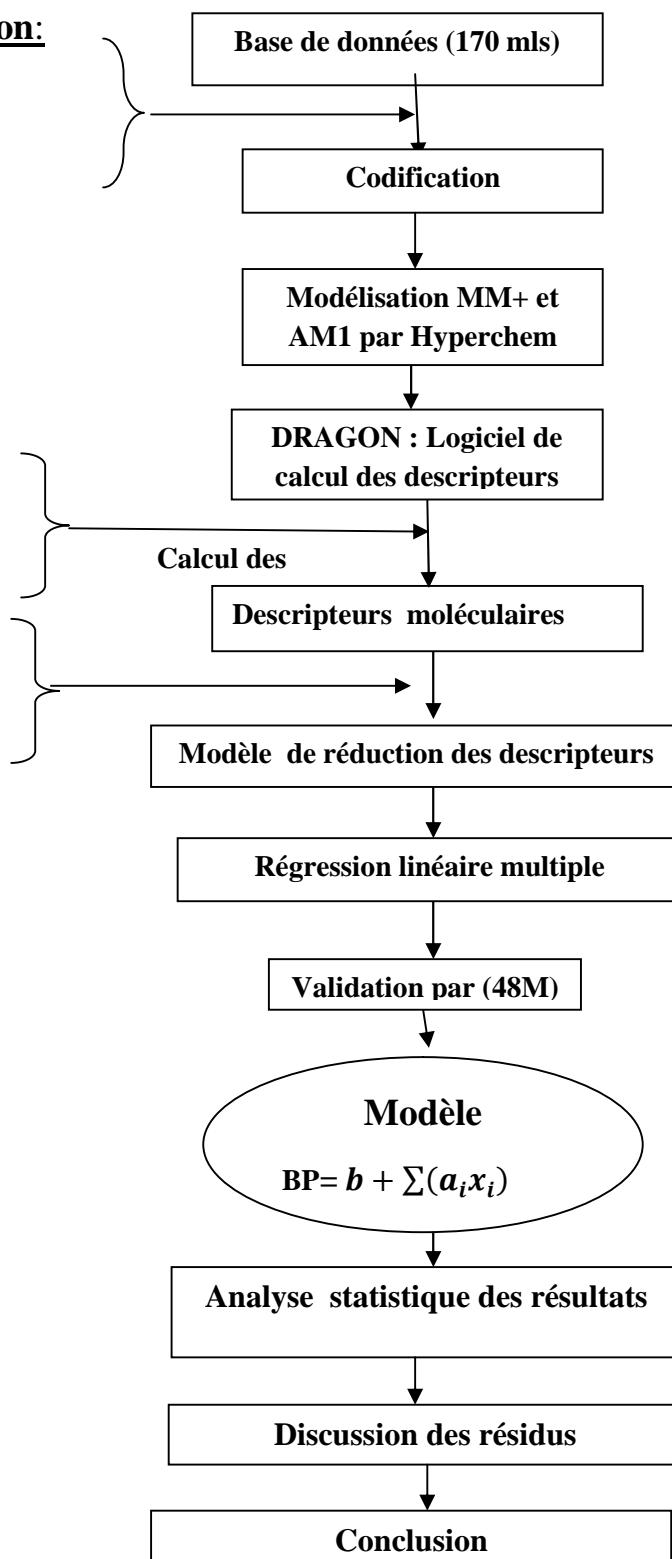
III.1 .Les étapes de travail (Diagramme de travail):

III.1. 1. Modélisation:

- Source
- Nombre
- Nature

- Définition
- Calcul des descripteurs

Réduction



III.2. Base des données :

III.2.1 source : la base de données utilisées dans ce travail a été compilée à partir de l'article: Gabriela Espinosa,[†] Denise Yaffe,[‡] Yoram Cohen,^{*,‡} Alex Arenas,[§] and Francesc Giralt[†]

III.2.2. nature: des composées : ces composées sont des alcanes.

- la raison du choix de cette base des données est pour l'étude des aspects chimiques des différents types des composées .

Enfinement nous avons trouvé cette base de données qui possède les caractéristiques et les propriétés précédentes.

- l'importance de la base de données dans l'estimation, est majeur puisqu'elle nous permet de prédire des propriété physico- chimiques des différents types des composées chimiques .

III.3. Génération des descripteurs moléculaires:

Afin d'exploiter au maximum les informations contenues dans les structures moléculaires, celles-ci sont traduites en une série de grandeurs (en général scalaires) qui quantifient leurs caractéristiques physico-chimiques et structurelles. Ces grandeurs sont appelées descripteurs. Depuis des décennies, de nombreux travaux ont été menés afin de développer des descripteurs capables de décrire de la manière la plus exhaustive possible les structures moléculaires. On en dénombre aujourd'hui des milliers [25,26]. Ils peuvent être calculés ou empiriques. Ces derniers nécessitant d'être mesurés, les descripteurs calculés seront privilégiés. Les détails les concernant sont disponibles dans l'ouvrage de Karelson [27].

Dans ce travail nous avons utilisés 5 bloc de descripteurs moléculaires qui sont :

- ❖ Descripteurs topological charge (block 1))
- ❖ Descripteurs conductivite indices (block 2))

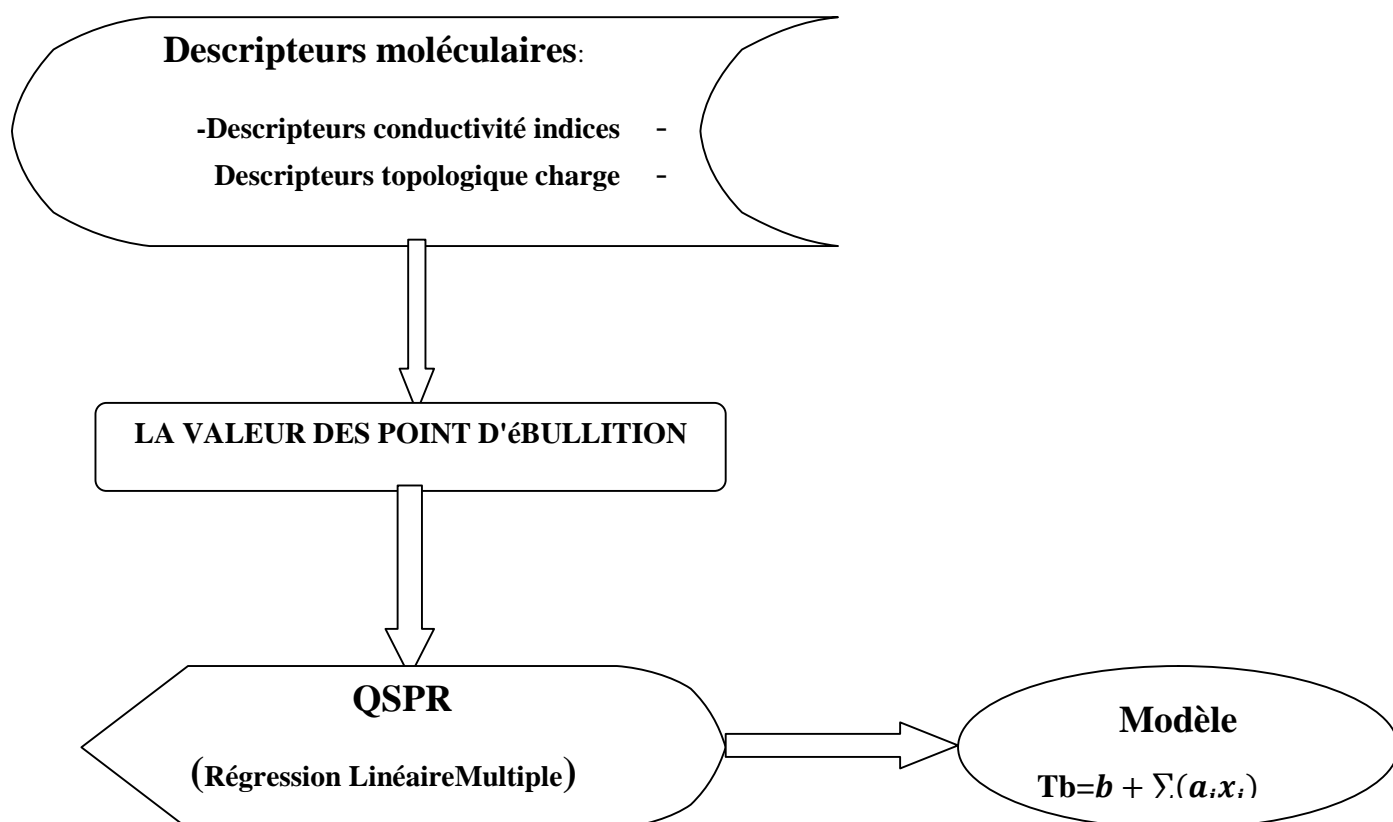


Figure III .2:relation descripteurs moléculaires Enter QSPR et point d'ébullition

III .3.1. Propriétés moléculaires :

Les blocs des descripteurs	Le nombre de descripteurs dans le bloc .	Le nombre de descripteurs après la réduction.
Descripteurs topologiques	33	33
Descripteurs constitutionnels	21	21

Tableau III.1: Montre les résultats de calcul des descripteurs[32]

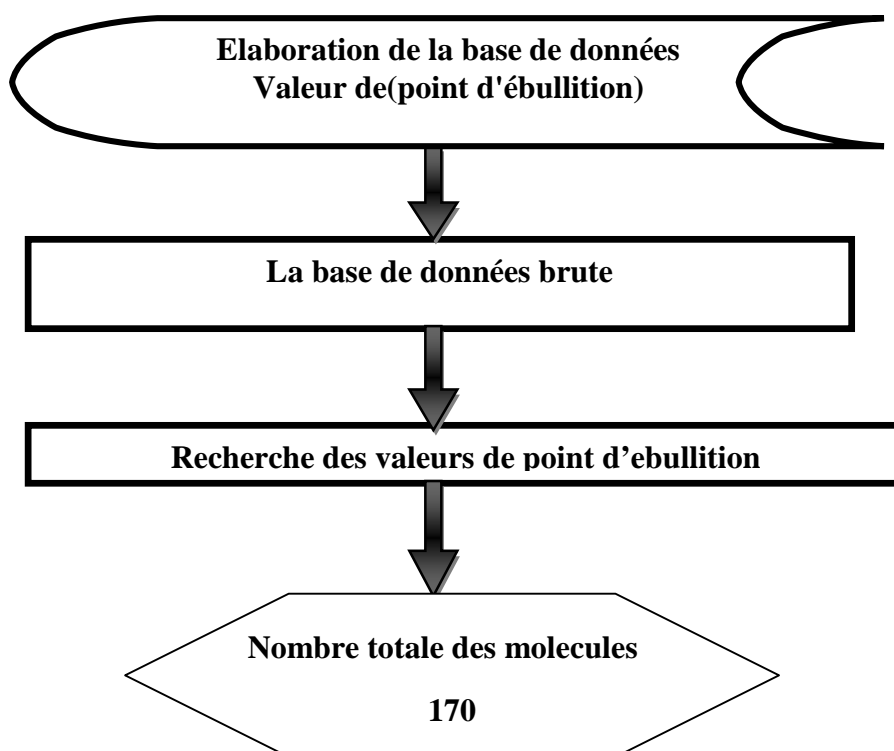
III .4. Organigramme:

Pour la contribution à la prédiction de **point d'ébullition** du composés chimiques et d'identifier les propriétés physiques et chimiques de ces composés, nous avons divisé cette étude en deux parties expérimental et par l'ordinateur.

Le travail expérimental est-ce la préparation de la base de données à un groupe spécifique de composés issus différentes familles chimiques, où chaque composé de la base de données sera défini par sa valeur expérimental de **point d'ébullition** . Le travail que nous faisons par l'ordinateur est-ce les deux programmes **DRAGON** et **XLSTAT** . Lorsque le programme **DRAGON** calcule les descripteurs moléculaires pour les composé inclus dans la base de données. Et le programme **XLSTAT** doit tenir compte statistiquement les résultats de **DRAGON** .

III .4.1. La préparation la base des données :

Nous apportons 170 composés chimiques différents (Composés de différentes familles chimiques) , Où chaque compose es



III .3: décrit les étapes de la préparation de la base de données

III.4.2. définition Logiciel DRAGON :

DRAGON est une application pour le calcul de descripteurs moléculaires. Ces descripteurs peuvent être utilisés pour évaluer la structure moléculaire d'activité ou des relations structure-

propriété, ainsi que pour l'analyse de similitude et de dépistage des bases de données des molécules.

DRAGON fournit plus de 1600 descripteurs moléculaires qui sont divisés en 20 blocs logiques. L'utilisateur peut calculer non seulement le type d'atome le plus simple, groupe fonctionnel et le nombre de fragments, mais aussi plusieurs descripteurs topologiques et géométriques. Certaines propriétés moléculaires telles que **logP**, réfractivité molaire, nombre de titres en rotation, H-bailleurs de fonds, H-accepteurs et de la surface topologique (**TPSA**) sont également calculés en utilisant des modèles communs tirés de la littérature.

III.4.3. Travail par Logiciel DRAGON :

Le cadre L'exécution du programme dans la fenêtre principale de **DRAGON** contient le menu de base du programme. Lorsque le programme démarre les éléments de menu est activé sont comme suit:

- -"descripteurs Calculer" pour obtenir de nouveaux descripteurs.
- -"descripteurs Load" pour afficher et modifier les descripteurs calculés précédemment.

Une fois les descripteurs moléculaires sont calculées ou chargées, trois autres éléments de menu sont activés:

- 1) Les réponses Load pour charger des variables définies par l'utilisateur.
- 2) descripteurs View pour afficher les valeurs des descripteurs, des statistiques et des graphiques.
- 3) Enregistrer les descripteurs pour enregistrer les descripteurs calculés.

Les descripteurs moléculaires **DRAGON** ont été divisés en 20 blocs logiques.

Symboles de courtes définitions et des références bibliographiques de descripteurs de chaque bloc peuvent être obtenus en cliquant sur le bouton correspondant dans "blocs descripteur" le cadre dans la fenêtre principale de l'application.

Ces blocs descripteurs ont été regroupées en fonction de la dimension descripteur. proposé par **DRAGON** sont **0D**, **1D**, **2D** et **3D**, où les chiffres se réfère à **0D** type d'atome et d'obligations, **1D** chiffres fragment, la **2D** à la descripteurs topologiques et connexes, et **3D** pour tous les descripteurs qui dépendent de la géométrie des coordonnées les atomes molécule.

Pour mettre en évidence les blocs contenant des descripteurs **0D**, **1D**, **2D** ou **3D** moléculaire cliquez sur les symboles correspondants sur le haut de "blocs Descriptor» le cadre. Pour désélectionner cliquez partout dans le châssis. Le bouton «Autres» permet à l'utilisateur de mettre en lumière mixte descripteurs moléculaires comme les descripteurs de charge et les propriétés moléculaires. [32]

III .5. Logiciel XLSTAT :

III.5.1. définition Logiciel XLSTAT :

XLSTAT est un Logiciel propose un nombre important de statistiques descriptives et de graphiques qui vous permettront d'avoir un premier aperçu pertinent de vos données. Bien que vous puissiez sélectionner plusieurs variables (ou échantillons) à la fois, **XLSTAT** calcul l'ensemble des statistiques descriptives pour chacun des échantillons indépendamment[31].

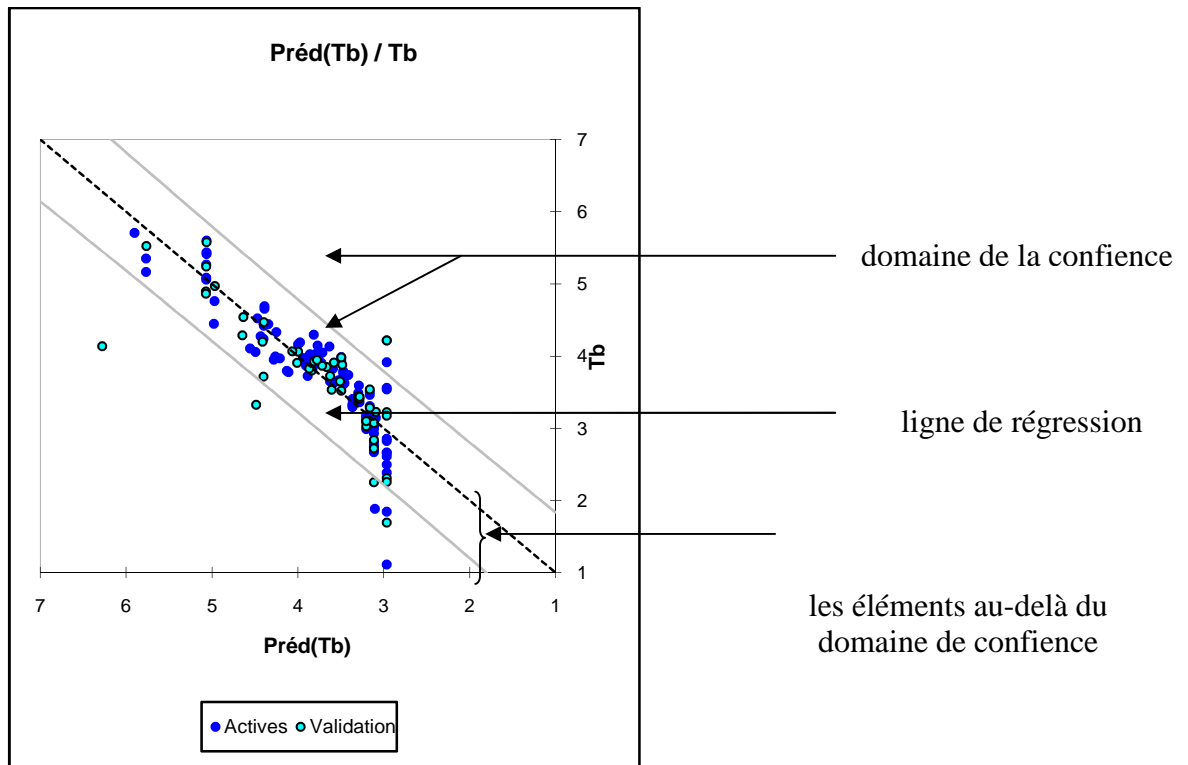
III .5. 2. Travail par Logiciel XLSTAT :

Nous a permis de faire le calcul et l'analyse statistique par la technique **QSPR** qui est couramment utilisé dans les études linéaire et aussi comme un excellent outil pour la sélection des descripteur qui précède la construction d'un modèle linéaire ou non linéaire [23]. et la method de régression multiple « stepwise » avec les Cinq blocs des descripteurs moléculaires. Cette analyse se devise en analyse statistique simple et analyse des résidus et validation du modèle par 180 molécules qui présentent 30% du nombre globale. Pour l'évaluation de la corrélation entre les descripteurs **2D** du bloc et les valeurs expérimentales de logk_{ow} qui ont été enregistrés durant la préparation de la base de données. Nous avons réalisé les calculs à l'aide du logiciel de modélisation moléculaires **XLSTAT** et rapporté les résultats obtenus pour ces descripteurs en nous référant à la prévision de la propriété physico-chimique logk_{ow}. dont les valeurs expérimentales ont été prélevées dans la littérature.

Les modèles ont été calculés à l'aide de la méthode de sélection pas à pas des variables de la régression linéaire multiple «**STEPWISE**». Avec la probabilité pour l'entrée est $\alpha = 0.05$

III .6.Déterminer les éléments au-delà du domaine de confiance

Après les calculs statistiques par le programme xlstat nous avons trouvés une courbe qui est constitué d'une droite (ligne de régression) et le domaine de la confiance .Tous les molécules de la base de données sont représentés par des points situer dans le domaine de confiance et autres qui sont hors (au-delà) de ce domaine



Conclusion

Après les activités de la préparation de la base des données, nous avons arrivés premièrement a un ensemble de valeurs expérimentales de point d'ébullition de 170 composé alcane et un ensemble de 235 descripteurs moléculaires a l'aide du programme **DRAGON**.

IV- Résultats et discussion

Introduction :

Dans ce travail nous étudions 2 blocs de descripteurs Chaque bloc séparément où les résultats sont obtenus à l'aide du programme XLSTAT (tableaux et graphique) puis analyse de ces résultats, la méthode utilisée dans cette étude est la régression linéaire multiple pas à pas «STEPWISE»

Cette analyse est une analyse statistique, Pour cela nous faisons un ensemble de validation contient 48 composés de la base de données. Il s'agit de déterminer et d'évaluer la corrélation entre les descripteurs pour chaque bloc et les valeurs expérimentales de BP.

Après avoir obtenu les résultats statistiques et l'analyse des résultats on a fait une discussion pour chaque bloc, nous avons mis la discussion générale.

IV-1-les résultats du bloc topologique charge indices:

IV-1-1-Equation du modèle :

Tb =

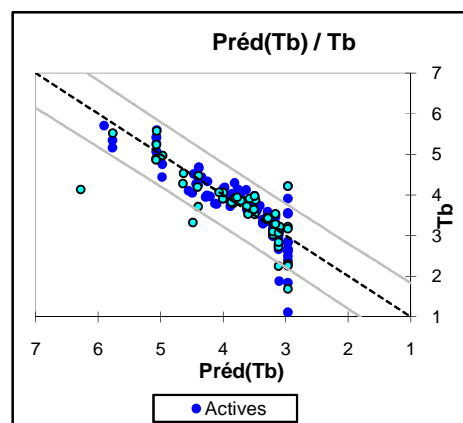
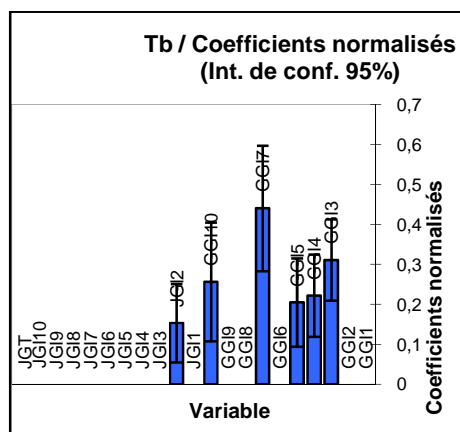
$$2,96895480470481 + 110,948058665004 * \text{GGI3} + 148,335526039738 * \text{GGI4} + 399,29742692296 * \text{GGI5} + 2937,23061909625 * \text{GGI7} + 4029,71296757575 * \text{GGI10} + 134,249042105795 * \text{JGI2}$$

IV-1-2-Les paramètres du modèle:

Variables	Valeurs	Coef	Ecart-type
Constante	2.969	b	0.068
GGI3	110.948	a ₁	18.292
GGI4	148.336	a ₂	34.716
GGI5	399.297	a ₃	108.879
GGI7	2937.231	a ₄	528.417
GGI10	4029.713	a ₅	1176.812
JGI2	134.249	a ₆	43.377

IV-1-3-Coefficients d'ajustement :

Observations	112.000
Somme des poids	112.000
DDL	105.000
R ²	0.761
Ajusté R ²	0.747
RMSE	0.392
MSE	0.154
MAPE	8.644
DW	0.964
Cp	22.160
AIC	-202.878
SBC	-183.849
PC	0.271



IV-1-4-Analyse des résultats:

R^2 = le coefficient de détermination donne une idée du pourcentage de la variation de la variable à modéliser, expliqué par les variables explicatives. Plus ce coefficient est proche de 1, plus le modèle est meilleur.

Dans ce cas $R^2 = 0.761$ et R^2 ajusté = 0.747

Donc le pourcentage de la variation de la variable à modéliser = **76.1%**.

Cela nous indique que **76.1%** de la variabilité du **point d'ébullition (BP)** est expliquée par les descripteurs. Le reste de la variabilité est dû à des effets (autres variables explicatives) qui ne sont pas pris en compte ou bien les descripteurs sélectionnés ne sont pas totalement descriptifs.

IV-1-5- discussion :

L'étude de corrélation entre les descripteurs deux à deux ont conduit à classer les descripteurs en deux types de descripteurs moins corrélés et descripteurs de grande corrélation. Cela signifie qu'ils vont nous donner des informations différentes sur la propriété à estimer **point d'ébullition**. Le calcul de la matrice de corrélation dans ce bloc nous a permis d'éliminer **6** autres descripteurs puisqu'ils nous ramènent à la même information que les **9** restants.

L'interprétation du graphe ($BP_{\text{préd}}/BP_{\text{exp}}$), nous montre qu'il y a **158** observations qui présentent **95%** de molécules de la base des données, et sont à l'intérieur de l'intervalle de confiance, ce qui nous permet de dire que les structures bidimensionnelles de ces molécules sont régulières. Par contre les **48** observations qui ne sont pas corrélées présentent les **5%** des observations (molécules) restants de la base des données, ces molécules possèdent des structures chimiques bidimensionnelles irrégulières.

IV-2-les résultats du bloc connectivity index:

IV-2-1-Equation du modèle:

Tb =

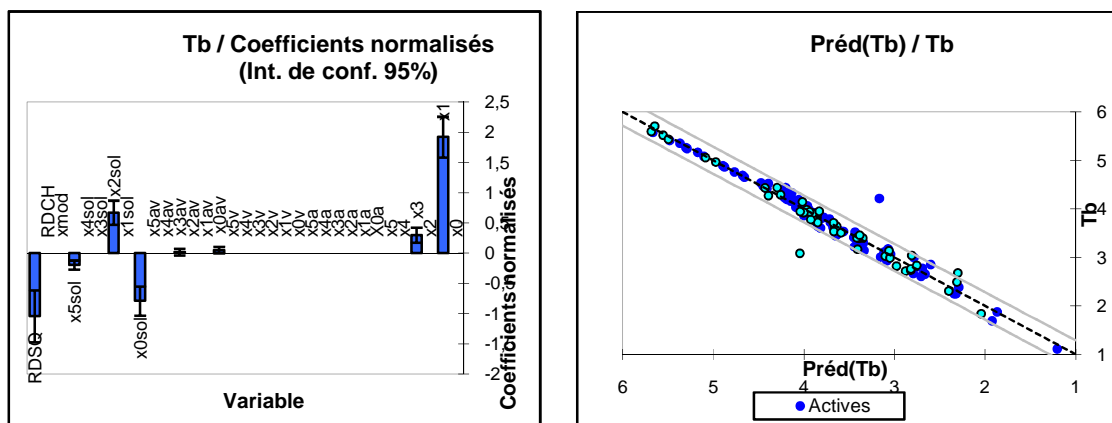
$$1,20788132747827+105,711384102632*x1+25,3909201391063*x3+42,2184701881824*x0av+19,6504493296501*x3av+29,4487803672978*x0sol+41,6014526105664*x2sol-6,42090157699217*x5sol-5,10061909040701*R$$

IV-2-2- Les paramètres du modèle:

Variables	Valeurs	Coef	Ecart-type
Constante	1.208	b	0.120
x1	105.711	a ₁	9.358
x3	25.391	a ₂	5.292
x0av	42.218	a ₃	21.658
x3av	19.650	a ₄	25.756
x0sol	-29.449	a ₅	4.530
x2sol	41.601	a ₆	6.236
x5sol	-6.421	a ₇	1.261
RDSQ	-5.101	a ₈	1.058

IV-2-3- Coefficients d'ajustement:

Observations	113.000
Somme des poids	113.000
DDL	104.000
R ²	0.969
Ajusté R ²	0.967
MSE	0.019
RMSE	0.139
MAPE	2.153
DW	1.907
Cp	47.438
AIC	-437.857
SBC	-413.311
PC	0.036



IV-2-4-Analyse des résultats:

R^2 = le coefficient de détermination donne une idée du pourcentage de la variation de la variable à modéliser, expliqué par les variables explicatives. Plus ce coefficient est proche de 1, plus le modèle est meilleur.

Dans ce cas $R^2 = 0.969$ et R^2 ajusté = 0.967

Donc le pourcentage de la variation de la variable à modéliser = **9.69%**.

Ce la nous indique que **9.69%** de la variabilité du **point d'ébullition** est expliquée par les descripteurs. Le reste de la variabilité est dû à des effets (autres variables explicatives) qui ne sont pas pris en compte ou bien les descripteurs sélectionnés ne sont pas totalement descriptifs.

IV-2-5- discussion:

L'étude de corrélation entre les descripteurs deux à deux ont conduit à classer les descripteurs en deux types de descripteurs moins corrélés et descripteurs de grande corrélation. Cela signifie qu'ils vont nous donner des informations différentes sur la propriété à estimer **point d'ébullition**. Le calcul de la matrice de corrélation dans ce bloc nous a permis d'éliminer **8** autres descripteurs puisqu'ils nous ramènent à la même information que les **3** restants.

L'interprétation du graphe ($BP_{\text{préd}}/BP_{\text{exp}}$), nous montre qu'il y a **158** observations qui présentent **95%** de molécules de la base de données, et sont à l'intérieur de l'intervalle de confiance, ce qui nous permet de dire que les structures bidimensionnelles de ces molécules sont régulières. Par contre les **3** observations qui ne sont pas corrélées présentent les **5%** des observations (molécules) restants de la base de données, ces molécules possèdent des stru

ctures chimiques bidimensionnelles irrégulières

IV-3-Analyse globale:

Blocs des descripteurs	topologique et charge indices	conectivité indices
R2	0.761	0.969
Nombre de variables	6	8
Nombre de molécules non corrélées (hors 95%)	9	3

IV-4-Discussion générale :

D'après les analyses statistiques des valeurs de ($BP_{\text{périd}}$ et BP_{exp}), résidus normalisés et l'interprétation des graphes ($BP_{\text{périd}}/BP_{\text{exp}}$) on voit que les valeurs de résidus varient de (-2+2) et les valeurs qui sont hors de cet intervalle présentent les points aberrants qui indiquent **les molécules non corrélées**. Ces molécules possèdent des structures chimiques bidimensionnelles irrégulières. Cette irrégularité est due à l'existence des liaisons d'hydrogène, van der Waals.

La qualité du modèle dépend du nombre de points aberrants, et du nombre des variables (paramètres).

La valeur de R^2 nous indique le degré de corrélation de la propriété (**BP**) avec les paramètres structuraux des différents blocs.

Si on compare les résultats des blocs entre-eux on peut dire que :

*Les paramètres **2D** du bloc (**conectivité indices**) donne une représentation suffisante pour la structure moléculaire des composées de la base des données utilisée.

D'après ces résultats, on observe que le degré d'estimation de la propriété (**BP**) change suivant la qualité de descripteurs de chaque bloc.

D'après cette étude et ces résultats on peut identifier et classer les composées (les alcanes) suivants la valeur des résidus ($BP_{\text{exp}} - BP_{\text{pred}}$) ou bien la valeur de (**BP**)

Conclusion:

Dans ce chapitre nous avons étudié la corrélation des descripteurs moléculaires qui sont calculés par le logiciel **Dragon version 5.5** (Descripteurs Topologique charge indices et Descripteurs Connéctivité indices) avec la propriété physico-chimiques qui est le point d'ébullition (**BP**) par la technique **QSPR**.

Conclusion générale :

Dans ce travail nous avons abordé le problème de prédiction du point d'ébullition (**BP**), pour cela nous avons fait la corrélation de différents types de paramètres (descripteurs) et la contribution pour le calcul de (**BP**) par la méthode **QSPR**. Après le calcul des descripteurs 2D (Descripteurs Topologique charge indices et Descripteurs Connectivité indices) par le logiciel **Dragon version 5.5**. Ce travail a été effectué sur **54** paramètres de **2** différents blocs et une base de données contienne **170** composés (alcanes) .

Pendant la prédiction nous avons trouvés une relation mathématique

$$\mathbf{BP} = f(\text{paramètres structuraux}) = \mathbf{BP} = \mathbf{b} + \sum (\mathbf{a}_i \times \mathbf{x}_i)$$

Ce modèle donne des résultats de corrélation remarquables avec les descripteurs du bloc Connectivité indices , ($\mathbf{R}^2 = 0.969$) ($\sigma=0.782$).

REFERENCE

- [1] C.E. Rechsteiner, in: W.J. Lyman, W.F. Reehl, D.H. Rosenblatt (Eds.), Handbook of Chemical Property Estimation Methods, McGraw-Hill: New York, 1982,
- [2] E. Panteli, E. Voutsas, K. Magoulas, D. Tassios, Fluid Phase Equilibr. 248 (2006) 70.
- [3] A. Komasa, Polish J. Chem. 77 (2003) 1491.
- [4] M. Pérez González, A.A. Toropov, P.R. Duchowicz D.A. Castro, Molecules 9 (2004) 1019.
- [5] Q. Li, X. Chen, Z. Hu, Chemom. Intel. Labl. Syst. 72 (2004) 93.
- [6] M. Randic, New. J. Chem. 24 (2000) 165.
- [7] M. Randic, Chemom. Intel. Labl. Syst. 10 (1991) 213.
- [8] M. Randic, J. Chem. Inf. Comput. Sci. 31 (1991) 970.
- [9] [http://fr.wiktionary.org/w/index.php?title=point d'Ébullition&oldid=10224209](http://fr.wiktionary.org/w/index.php?title=point_d%27%C3%A9bullition&oldid=10224209) »
Catégories :
- [10] <http://www.encyclopedie-gratuite.fr/redacteur.php>
- [11] <http://creativecommons.org/licenses/by-sa/3.0/deed.fr>
- [12] Niemelä1993 QTOXIN-program (ver 2.0). Danish Environmental Protection Agency.
- [13] Dearden, J. C. Quantitative Structure-Property Relationships for Prediction of Boiling Point, Vapor Pressure, and Melting Point. EnViron. Toxicol. Chem. 2003, 22, 1696.
- [14] Dearden, J. C. Quantitative Structure-Property Relationships for Prediction of Boiling Point, Vapor Pressure, and Melting Point. EnViron. Toxicol. Chem. 2003, 22, 1696.
- [15] Li, Q.; Chen, X.; Hu, Z. Quantitative structure-property relationship studies for estimating boiling points of alcohols using calculated molecular descriptors with radial basis-function neural networks. Chemom. Intell. Lab. Syst. 2004, 72, 93.
- [16] J Basak, S. C.; Gute, B. D.; Mills, D.; Hawkins, D. M. Quantitative Molecular Similarity Methods in the Property/Toxicity Estimation of Chemicals: A Comparison of Arbitrary versus Tailored Similarity Spaces. J. Mol. Struct. 2003, 622 (1-2), 127.
- [17] Tropsha, A.; Gramatica, P.; Gombar, V. K. The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. QSAR Combin. Sci. 2003, 22, 69.
- [18] J Shacham, M.; Brauner, N.; Cholakov, G. St.; Stateva R. P. Property Prediction by Correlations Based on Similarity of Molecular Structures. AIChE J. 2004, 50, 2481.

- [19] Brauner, N.; Shacham, M.; Cholakov, G. St.; Stateva, R. P. Property Prediction by Similarity of Molecular Structures - Practical Application and Consistency Analysis. *Chem. Eng. Sci.* 2005, 60, 5458
- [20] Manuel de descripteurs moléculaires, R.Todeschini et V.Consonni, Wiley-VCH, 2000)
- [21] Manuel des descripteurs moléculaires R.Todeschini, V.Consonni, , Wiley-VCH, Weinheim (GER), 2000)
- [22]] Dragon_ Aide blocs des descripteurs
- [23]] ELULOOKIRJELDUS Kaido Tamm QSPR Modeling Of Some Properties Of Organic Compounds
- [24] M. Karelson, *Molecular Descriptors in QSAR/QSPR*, Wiley, New York, **2000**.
- [25] R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley, Weinheim, **2000**.
- [26] M. Karelson, V.S. Lobanov, A.R. Katritzky, *Chem. Rev.*, **1996**, 96, 1027-1044.
- [27] S.W. Benson, J.H. Buss, *J. Chem. Phys.*, **1958**, 29, 546-572.
- [28] M. Randic, *J. Am. Chem. Soc.*, **2002**, 97, 6609-6615.
- [29] L.B. Kier, L.H. Hall, *J. Pharm. Sci.*, **1981**, 70, 583-589.
- [30] A.T. Balaban, *Chem. Phys. Lett.*, **1982**, 89, 399-404
- [31] S.C. Basak, D.K. Harriss, V.R. Magnuson, *J. Pharm. Sci.*, **1984**, 73, 429-437.
- [32] : Gabriela Espinosa,† Denise Yaffe,‡ Yoram Cohen,*,‡ Alex Arenas,§ and Francesc Giral

Résumé :

L'étude de la relation quantitative structure-propriété (QSPR) est réalisée pour établir des corrélations qui relient les structures moléculaires des composés organiques à leur les valeurs **expérimentales du point d'ébullition (BP)**, en utilisant les descripteurs moléculaires .

Les corrélations sont simples en application avec une précision ,qui fournissent une manière simple, directe et relativement précise pour calculer **BP**. Ces calculs nous donnent un modèle qui nous amène aux résultats remarquables en corrélation avec les descripteurs des bloc Descripteurs conndictivite indices ($R^2 = 0,969$). ($\sigma = 0.782$).

المخلص :

دراسة العلاقة الكمية التي تربط بنية المركب الكيميائي بالخاصية الفيزيوكيميائية (QSPR), استعملت هذه العلاقة لتحديد درجة الارتباط بين البنية الكيميائية للمركبات العضوية والقيم التجريبية للخاصية الفيزيوكيميائية: درجة الغليان, (**BP**) وهذا باستعمال الواصفات الجزئية

درجات الارتباط كانت سهلة التطبيق ودقيقة وهذا ما سهل حساب درجة الغليان النظرية (**BP_{pred}**). هذه لحسابات الاحصائية اعطت نموذج (المعادلة الرياضية) الذي بواسطته حصلنا على درجة ارتباط جيدة في مجموعة الواصفات الخاصة (**Connéctivité indices**)

, ($R^2 = 0.969$) ($\sigma = 0.782$)..
