

République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la recherche
scientifique

Université Echahid Hamma Lakdhar- EL Oued

Faculté de la Technologie
Département de Génie des Procédés et Pétrochimie

MÉMOIRE

Présenté en vue de l'obtention du diplôme de

MASTER

Spécialité : Génie chimique

Présenté par :

- Mostafa LASSOUED
- Hocine HAMMOUDA

**Prédiction QSPR de propriétés physico-chimiques pour
l'enregistrement et l'évaluation des produits chimiques :
Cas des composés organiques phase solide**

Soutenue le :/..... /2020

Devant le jury composé de :

Dr :

Président

Dr :

Examineur

Dr : SEROUTI Abdelghani

Encadreur

Année universitaire 2019/2020

Prédiction QSPR de propriétés physico-chimiques pour l'enregistrement et
l'évaluation des produits chimiques : Cas des composés organiques phase solide

Présentée par

Mostafa LASSOUED

Hocine HAMMOUDA

Mémoire présenté au Département Génie des Procédés
en vue de l'obtention du grade de master académique en sciences et Technologie
Spécialité : Génie Chimique

FACULTÉ DES SCIENCES ET TECHNOLOGIE
UNIVERSITÉ D'EL OUED

El Oued, Alger, Septembre 2020

Résumé

Pour l'enregistrement et l'évaluation des produits chimiques, nécessite des informations sur les propriétés physico-chimiques pertinentes du produit chimique. Les valeurs de propriétés prédites peuvent être utilisées lorsqu'il est possible de démontrer que les prédictions sont valides et adéquates. Les propriétés physico-chimiques pertinentes pouvant faire l'objet de prévisions sont les suivantes : point de fusion, point d'ébullition, densité, pression de vapeur, solubilité dans l'eau, point d'éclair, inflammabilité, propriétés explosives...etc.

Le but de cette thèse est de développer une méthode des descripteurs moléculaires permettant d'évaluer les enthalpies de combustion de composés solides contenant des atomes de carbone, d'hydrogène, d'oxygène et surtout d'azote.

Pour obtenir les résultats on utilise les méthodes de relation structure-propriété (QSPR), ainsi que les logiciels de prédiction de propriété pertinente comme Dragon (descripteurs moléculaires) et XLSTAT (régression multilinéaire).

MOTS CLES : QSPR, Enthalpie de combustion, Composés organiques, Propriétés physico-chimiques, Descripteurs moléculaires, XLSTAT.

ملخص:

يتطلب تسجيل وتقييم المواد الكيميائية معلومات عن الخصائص الفيزيائية الكيميائية ذات الصلة للمادة الكيميائية. يمكن استخدام قيم الممتلكات المتوقعة عندما يكون من الممكن إثبات أن التنبؤات صحيحة وكافية. الخصائص الفيزيائية والكيميائية ذات الصلة التي يمكن التنبؤ بها هي التالية: نقطة الانصهار، نقطة الغليان، الكثافة، ضغط البخار، الذوبان في الماء، نقطة الوميض، القابلية للاشتعال، الخواص التفجيرية، إلخ.

الهدف من هذه الأطروحة هو تطوير طريقة للوصفات الجزيئية التي تجعل من الممكن تقييم المحتوى الحراري لاحتراق المركبات الصلبة التي تحتوي على ذرات الكربون والهيدروجين والأكسجين وخاصة النيتروجين.

للحصول على النتائج، نستخدم طرق علاقة خصائص البنية (QSPR)، بالإضافة إلى برامج التنبؤ بالخصائص ذات الصلة مثل Dragon (الواصفات الجزيئية) و XLSTAT (الانحدار متعدد الخطوط).

الكلمات المفتاحية : QSPR، المحتوى الحراري لاحتراق، المركبات العضوية، الخواص الفيزيائية الكيميائية، الوصفات الجزيئية، XLSTAT.

ABSTRACT:

For the registration and evaluation of chemicals, requires information on the relevant physicochemical properties of the chemical. Predicted property values can be used when it is possible to demonstrate that the predictions are valid and adequate. The relevant physicochemical properties that can be predicted are the following: melting point, boiling point, density, vapor pressure, water solubility, flash point, flammability, explosive properties.... etc.

The aim of this thesis is to develop a method of molecular descriptors making it possible to evaluate the enthalpies of combustion of solid compounds containing atoms of carbon, hydrogen, oxygen and especially nitrogen.

To obtain the results we use the methods of structure-property relationship (QSPR), as well as the relevant property prediction software such as Dragon (molecular descriptors) and XLSTAT (multilinear regression).

KEYWORDS: *QSPR, Enthalpy of combustion, Organic compounds, Physico-chemical properties, Molecular descriptors, XLSTAT.*

Remerciements

Tout d'abord :

Merci à Dieu pour tous ce qu'il nous a donné et tout puissant pour la volonté, la santé, et la patience, qu'il nous a données durant toutes ces années d'étude.

Nous exprimons nos profondes gratitudes à nos parents pour leurs encouragements, leur soutien et pour les sacrifices qu'ils ont enduré.

Ainsi, nous tenons également à exprimer nos vifs remerciements à notre encadreur

Mr. SEROUTI Abdelghani, Maître assistant à l'Université d'El- Oued,

pour avoir d'abord proposé ce thème, pour suivi continuuel tout le long de la réalisation de ce mémoire, et qui n'a pas cessé de nous donner ses conseils et remarques.

Nous tenons à remercier vivement toutes les personnes qui nous ont aidés à élaborer et réaliser ce mémoire, ainsi à tous ceux qui nous ont aidés, de près ou de loin, à accomplir ce travail.

Nos remerciements vont aussi à tous les enseignants du département de génie des procédés qui a contribué à notre formation. Enfin nous tenons à exprimer notre reconnaissance à tous nos amis et collègues pour le soutien moral et matériel...

Table des matières

Résumé.....	iii
Remerciements.....	v
Table des matières	vi
Liste des abréviations.....	ix
Liste des tableaux.....	x
Liste des figures	xi
Introduction.....	1
Chapitre 1 Revue bibliographie	3
Introduction.....	3
1.1 Composés organiques.....	3
1.2 Propriétés physiques et thermodynamiques	4
1.2.1 Importance	4
1.3 Enthalpie de réaction.....	5
1.3.1 Définition de l'enthalpie.....	5
1.3.2 Calcule de l'enthalpie standard de combustion	7
1.3.3 Calcule de l'enthalpie de formation standard	9
1.3.4 Détermination expérimentale.....	9
1.4 Méthodes de calculs	15
1.4.1 Méthodes de contribution de groupes	15
1.4.2 Méthode de Benson.....	19
1.4.3 Extension de la méthode de Benson à la phase liquide et à la phase solide	21
1.4.4 Méthodes de descripteurs moléculaires	22
Chapitre 2 Outils et techniques utilisées.....	25

Introduction.....	25
2.1 Analyse quantitative de relation structure-activité/propriété	26
2.1.1 Construction du modèle	26
2.2 Descripteur moléculaire	29
2.2.1 Les types de descripteurs moléculaires.....	29
2.3 Techniques des régressions	34
2.3.1 Régression linéaire multiple (MLR)	35
2.3.2 Régression non linéaire.....	37
2.3.3 Différences entre Régression : non linéaire /linéaire multiple.....	37
2.3.4 Logiciel XLSTAT.....	39
2.4 Sélection du modèle	40
2.4.1 Méthode générale de sélection.....	41
2.4.2 Sélection du modèle par estimation de l'erreur de généralisation.....	42
2.4.3 La validation croisée	43
Chapitre 3 Modélisation de la propriété $\Delta_c H^\circ$	44
3.1 Etablissement de la base de donne et détermination de descripteurs moléculaires ..	44
3.2 Méthodologie de préparation de base des données	46
3.2.1 Conception de la base des données.....	46
3.2.2 Préparation de base des données.....	48
3.3 Méthodes d'analyse de données	53
3.3.1 Régression multi-linéaire	53
3.3.2 Régression non linéaire.....	55
3.3.3 Le résidu moyen et la moyenne quadratique	56
Chapitre 4 Résultats et discussion	58
Introduction.....	58
4.1 Traitement le base de donne.....	58
4.2 Détail les résultats à base de donne.....	59
4.2.1 Résultat de régression linéaire multiple.....	60
4.3 Analyse des résultats	64

4.3.1	Mise en évidence de problèmes générale.....	65
4.4	Paramètres des modèles de régression multilinéaire.....	66
4.4.1	Coefficients d'ajustement.....	66
4.4.2	Paramètres du modèle	66
Conclusion		68
Bibliographie.....		70
Annexe A		73
A.1	HperChem	73
A.1.1	Définition.....	73
A.1.2	Construire et Afficher des Molécules.....	73
A.1.3	Construire et Afficher des Molécules.....	73
A.1.4	Utilisant les Méthodes des Mécaniques Moléculaires.....	74
A.1.5	Disponibilité de Paramètres.....	74
A.1.6	Traits du Champ de la force	74
A.1.7	MM+	74
A.1.8	Interactions électrostatiques	75
A.2	Autres paramètres du modèle.....	76

Liste des abréviations

QSAR	Relation structure-activité quantitative
QSPR	Relation structure-propriété quantitative
EQMA	Erreur quadratique moyenne commise sur l'ensemble d'apprentissage
EQMT	Erreur quadratique moyenne commise sur l'ensemble du test
MCE	Moyenne carrés des erreurs
RMCE	Racine carrée de la MCE.
MAPE	Pourcentage d'erreur absolue moyenne
DW	Coefficient de Durbin-Watson
Cp	Coefficient de Mallows
AIC	Critère d'information d'Akaike (Akaike's Information Criterion).
SBC	Critère Bayésien de Schwarz (Schwarz's Bayesian Criterion)
PC	Critère de prédiction d'Amemiya (Amemiya's Prediction Criterion)
F	Coefficient de Fisher
STD	La déviation Standard
AAD	La déviation Absolu moyenne
RLM	Régression linéaire multiple

Liste des tableaux

Tableau 1 - Chaleur de combustion (quelques substances) calculée par l'équation I-12	8
Tableau 2 - Découpage de deux molécules en groupes d'ordre 0, 1 et 2.	17
Tableau 3 - Appellations de l'atome de carbone.	20
Tableau 4 - Les descripteurs moléculaires de l'équation (I.24) et leur signification.	23
Tableau 5 - Les valeurs des constantes de l'Eq (I.24) et leurs interprétations statistiques.	23
Tableau 6 - Quelque descripteurs constitutionnels générés par le logiciel Dragon.....	33
Tableau 7 - Quelques descripteurs moléculaires (bloc de groupements fonctionnels).....	33
Tableau 8 - Les différents modèles prédictif dans la fonctionnalité Easy Fit	40
Tableau 9 - Montre les résultats de calcul des descripteurs.....	51
Tableau 10 - Les descripteurs : Descripteurs d'anneau et Descripteurs de type P_VSA.....	52
Tableau 11 - Les coefficients d'ajustement pour chaque bloc de descripteur	54
Tableau 12 - Coefficients d'ajustement pour RL du descripteur de type P_VSA.....	60
Tableau 13 - Résidus moyens et moyennes quadratiques obtenus (Desc de type P_VSA)....	60
Tableau 14 - Coefficients d'ajustement pour RN du descripteur Descripteurs d'anneau.....	61
Tableau 15 - Résidus moyens et moyennes quadratiques obtenus (Descripteurs d'anneau) ..	61
Tableau 16 - Coefficients d'ajustement (Descripteurs de type P_VSA).....	66
Tableau 17 - Coefficients d'ajustement (Descripteurs d'anneau).....	66
Tableau 18 - Paramètres du modèle (Descripteurs de type P_VSA).....	66
Tableau 19 - Paramètres du modèle (Descripteurs d'anneau).....	67
Tableau 20 - Analyse de la variance (Descripteurs d'anneau).....	76
Tableau 21 - Analyse de la variance (Descripteurs de type P_VSA)	76
Tableau 22 - Matrice de corrélation (Descripteurs de type P_VSA).....	76
Tableau 23 - Les étapes de recherche des données à partir de nbre d'enregistrement CAS ...	77
Tableau 24 - Nbre d'enregistrement CAS et la masse moléculaire pour les données étudier.	79
Tableau 25 – Liste de Descripteurs d'anneau (en anglais).....	86
Tableau 26 – Liste de Descripteurs de type P_VSA (en anglais).....	87

Liste des figures

Figure 1 - Schéma simplifié du calorimètre et de la bombe calorimétrique	11
Figure 2 - Thermogramme idéal et réel	13
Figure 3 -Exemple de deux molécules, isomères de constitution.....	17
Figure 4 -Principes de la méthode QSAR/QSPR.....	27
Figure 5 - Procédure de validation des modèles	28
Figure 6 - Procédure de la validation croisée.....	43
Figure 7 - Modèle de l'étude de relation structure activité/propriété	45
Figure 8 - Schéma du cadre hiérarchique pour Modèle QSPR.	46
Figure 9 - Représentation schématique de l'architecture la base de données relationnelles ...	46
Figure 10 - Schéma d'une relation selon le modèle de bases de données relationnelles	47
Figure 11 - Application de l'option "Field Force" comme méthode de calcul MM.....	50
Figure 12 - Correction valeur de $\Delta_c H^\circ$ (Composant N° CAS : 33719-66-3-2d)	59
Figure 13 - Le résultats des $\Delta_c H^\circ$ aux descripteurs de type P_VSA.....	62
Figure 14 - Corrélation entre les $\Delta_c H^\circ$ Exp et Préd pour les descripteurs de type P_VSA	62
Figure 15 - Le résultats des $\Delta_c H^\circ$ aux descripteurs d'anneau.....	63
Figure 16 - Corrélation entre les $\Delta_c H^\circ$ Exp et Préd pour les descripteurs d'anneau	63

Introduction

L'utilisation des méthodes théoriques pour l'obtention de modèles qui puissent prédire et comprendre les structures, les propriétés et les interactions moléculaires est connue sous le nom de « Modélisation Moléculaire ». Celle-ci permet de fournir des informations qui ne sont pas disponibles par l'expérience et joue donc un rôle complémentaire à celui de la chimie expérimentale. Ainsi, la modélisation moléculaire peut par exemple permettre de faire une idée précise de la structure de l'état de transition pour une réaction donnée, ce qui est difficile, voire impossible, par la chimie expérimentale.

Identifier et déterminer aussi tôt et de manière aussi fiable que possible les propriétés dangereuses des substances chimiques, quelle que soit leur utilisation finale, représente une étape initiale importante dans la maîtrise des risques industriels. Ceci est d'autant plus vrai pour les propriétés explosives intrinsèques des substances chimiques qui se classent au premier rang des dangers physico-chimiques pouvant être craints au cours du stockage, transport ou manipulation. Aux vues de la complexité des phénomènes liés à ces propriétés d'explosibilité dans les cas réels, l'approche expérimentale reste, jusqu'à présent, la voie privilégiée pour l'évaluation de cette catégorie de dangers (FAYET, 2010).

Le coût d'une telle démarche, le temps nécessaire et même la disponibilité des laboratoires d'essais agréés pour la réalisation des essais rendent le processus particulièrement difficile du fait du nombre de substances concernées (plus de 143 000)

Aussi, l'utilisation de méthodes alternatives à l'expérimentation, parmi lesquelles les Relations Quantitatives_Structure_Activité/Propriété (QSAR/QSPR) sont devenues d'un grand intérêt et sont même recommandées dans la nouvelle réglementation, afin d'obtenir les données nécessaires à l'enregistrement des substances.

Ces méthodes sont à l'heure actuelle principalement utilisées pour le criblage des propriétés toxiques, des nitroaromatiques par exemple.

La procédure de criblage consiste à utiliser ces méthodes pour prédire a priori les performances (ou la dangerosité) des substances pour des applications données. Ainsi, il est possible d'identifier les molécules présentant le plus fort potentiel pour la propriété attendue (recherche d'explosifs performants par exemple) ou celles présentant d'éventuels problèmes de sécurité.

Les premières applications de ces méthodes, basées sur des analyses statistiques, ont concerné principalement des applications biologiques, toxicologiques ou pharmaceutiques.

Dans le domaine des propriétés explosives, des méthodes reposant sur des groupes de contribution ont déjà été développées pour l'évaluation de l'énergie libérée pendant la décomposition (CHETAH) et le calcul des réactions adiabatiques (CART) (FAYET, 2010).

Le but ultime de ce travail est de développer un modèle capable de prédire l'enthalpie standard de combustion en phase solide de molécules organiques (matériaux énergétiques), en se basant sur les descripteurs des molécules et en utilisant une méthodologie de type QSPR reliant les propriétés expérimentales aux structures moléculaires, calculées à l'aide d'outils de chimie quantique.

Il s'agit de mettre en place une équation mathématique (ou un algorithme) entre la propriété macroscopique à prédire et une série de descripteurs à l'échelle moléculaire de plusieurs types : constitutionnels, topologiques, géométriques ou quantiques. Pour se faire, différents outils de traitement de données peuvent être employés : algorithmes génétiques, réseaux de neurones ou régressions multi-linéaires.

Dans le cadre de ce travail, nous commençons le présent manuscrit, par décrire brièvement la méthodologie du plan de ce mémoire. Ce dernier est constitué de quatre chapitres, qui sont abordés successivement dans ce qui suit :

On présente dans le premier chapitre l'importance des propriétés physiques et thermodynamiques (enthalpie standard de combustion), et une revue bibliographique des méthodes prédictives.

Le deuxième chapitre est, quant à lui, consacré à la description du dispositif des descripteurs moléculaires et les techniques utilisées.

Dans le troisième chapitre, nous expliquons la méthodologie de préparation de la base de données, les étapes de méthode de calcul par les logiciels utilisés.

La méthode de calcul retenue est finalement décrite en détail dans le quatrième chapitre est consacré à la présentation des résultats obtenus, la discussion de ces résultats ainsi leurs comparaisons avec celles obtenues par d'autres méthodes. Enfin, la technique développée est appliquée à de nombreux composés énergétiques.

Chapitre 1

Revue bibliographie

Introduction

Le terme de chimie organique apparaît pour la première fois en 1807 grâce au suédois Jacob Berzélius. Il définit comme composé organique toute espèce chimique provenant d'un matériau d'origine vivante. D'autres chimistes, tels que Friedrich Wöhler ou August Kékulé, ont beaucoup œuvré pour le développement de la chimie organique.

Plus tard, de nombreuses synthèses organiques permettent l'obtention de composés synthétiques, une nouvelle définition est alors nécessaire. La chimie organique est aujourd'hui la chimie des composés carbonés, naturels ou synthétiques.

Outre le carbone, les espèces organiques ne contiennent qu'une gamme réduite d'éléments : l'hydrogène H, l'oxygène O, l'azote N ou plus rarement le soufre S, le phosphore P et le chlore Cl (DEUDON, 2008).

Nous définissons dans ce chapitre les Composés organiques, puis nous montrons les propriétés physiques et thermodynamiques (enthalpie de combustion), et enfin une revue bibliographique des méthodes prédictives.

1.1 Composés organiques

Les composés organiques sont, au niveau le plus fondamental, des composés qui contiennent du carbone et de l'hydrogène. Ces composés sont nommés organiques puisque jadis on croyait qu'ils venaient des organismes vivants, mais ce n'est pas toujours forcément le cas. Il y a des milliers de composés organiques qui se trouvent naturellement ou qui sont produits de façon synthétique. Parmi les exemples de composés organiques sont les hydrates de carbone, gras (lipides), protéines, et acides nucléiques, qui sont la base même des molécules de la vie. Les composés organiques incluent aussi le pétrole et le gaz naturel, qui sont les composantes principales des combustibles fossiles.

Les composés organiques sont divisés en deux groupes principaux. Les hydrates de carbone sont les composés qui ne contiennent que des atomes de carbone et d'hydrogène. Ils peuvent être subdivisés en alcanes (une seule liaison carbone), alcènes (double liaison carbone), alcynes (triple

liaison carbone), et arènes (les carbones forment un anneau cyclique appelé benzène) (Feunang, 2016).

Le deuxième groupe de composés organiques est défini par un groupement fonctionnel qui remplace l'un des atomes d'hydrogène dans la molécule. Les groupements fonctionnels peuvent être un seul atome (comme F, Cl, Br, ou I) ou une combinaison d'atomes (NH₂, OH, C=O, ou SH). Les groupements fonctionnels influencent la réactivité chimique des molécules organiques et sont importants dans la synthèse organique. Les groupements fonctionnels communs incluent les groupements contenant de l'oxygène (alcools, éthers, aldéhydes, cétones, acides carboxyliques, et esters d'acide carboxylique), les groupements contenant de l'azote (amines et amides), et les groupements contenant le soufre (thiols, thioéthers, et thioesters). Afin de trouver ces groupements fonctionnels particuliers, utilisez la recherche Fisher Scientific par structure chimique.

1.2 Propriétés physiques et thermodynamiques

1.2.1 Importance

La connaissance des propriétés physiques et thermodynamiques des fluides ou solides est indispensable pour le fonctionnement des procédés qui utilisant par exemple les températures et pressions élevées. Pour cette raison, plusieurs littératures faites des recherches pour obtenir des méthodes d'estimation qui facilite les calculs physiques et thermodynamiques des composés organiques purs ou mélange.

En outre, les propriétés physiques et thermodynamiques sont très importantes dans le domaine de stockage et transport des produits chimiques.

Les transformations chimiques des produits sont basées sur les propriétés physiques des composés. Cependant ce n'est pas toujours possible de trouver des valeurs expérimentales des composés dans la littérature (SAIDAT, 2004).

L'enthalpie de formation est élevée et plus l'énergie libérée au cours de la combustion du matériau énergétique n'est plus grande. Pour cela, il est essentiel de favoriser l'introduction de groupements azotés dans les structures moléculaires à développer, ceux-ci présentant une enthalpie de formation importante (Ex : N-NO₂, C-N₃) (SALMON, 2006).

1.3 Enthalpie de réaction

Les réactions chimiques cèdent la place à des transferts vigoureux entre l'énergie potentielle associée à la position des atomes et l'énergie cinétique associée aux vitesses microscopiques des molécules (translation, rotation et vibration), produisant ainsi un échauffement ou un refroidissement macroscopique (changement de température; le plus le cas courant étant que les réactions soient exothermiques, c'est-à-dire dégagent de la chaleur lorsqu'elles sont maintenues à température constante) Plusieurs nouvelles variables d'équilibre sont définies pour analyser les systèmes réactifs (par exemple l'enthalpie de réaction, H_r) (ISIDORO, 1992).

1.3.1 Définition de l'enthalpie

Le terme « Enthalpie » désigne une grandeur thermodynamique correspondant à une fonction d'état extensive, c'est-à-dire une fonction d'état proportionnelle à la quantité de matière en présence. En revanche, l'enthalpie est indépendante de la façon dont un système a atteint son état. Elle s'appuie sur la situation initiale et sur la situation finale sans considérer les états intermédiaires.

En d'autres termes, L'enthalpie (se note H), est la mesure de l'énergie du système qui peut être dégagée sous forme de chaleur. (Enthalpie = quantité de chaleur libérable par un système). Elle est définie par la relation suivante :

$$H = E + P \quad (I.1)$$

L'enthalpie représente l'énergie totale d'un système thermodynamique : son **énergie interne (E)** (c'est-à-dire l'énergie nécessaire à la création du système), plus le travail de détente (**PV**), qui correspond au travail que le système doit fournir contre **la pression** pour occuper son **volume** (GUILLAUME, 2016).

Comme l'énergie interne, pression et volume sont toutes les fonctions de l'Etat, d'enthalpie est aussi une fonction d'état. Mais quel est exactement l'enthalpie ? Pour aider à répondre à cette question, considérons un processus réalisé à la pression constante et où le seul travail autorisé est de pression-volume de travail ($W = - P\Delta V$). Dans ces conditions, l'expression :

$$\Delta E = QP + W \quad (I.2)$$

Devient :

$$\Delta E = QP - P\Delta V \quad (I.3)$$

Où

$$QP = \Delta E + P\Delta V \quad (I.4)$$

Où QP est la chaleur à pression constante.

Nous allons maintenant se rapportent QP à une variation d'enthalpie. La définition de l'enthalpie est $H = E + PV$.

Par conséquent, nous pouvons dire : *Variation de H = Variation de E + Variation de PV*

Où

$$\Delta H = \Delta E + \Delta(PV) \quad (I.5)$$

Puisque P est constant, la variation de PV est due uniquement à une variation de volume. Ainsi :

$$\Delta(PV) = P\Delta V \quad (I.6)$$

$$\Delta H = \Delta E + P\Delta V \quad (I.7)$$

Cette expression est identique à celui que nous avons obtenu pour Q_p :

$$Q_p = \Delta E + P\Delta V \quad (I.8)$$

Ainsi, pour un processus réalisé à la pression constante et où le seul travail autorisé est celui d'une variation de volume, nous avons :

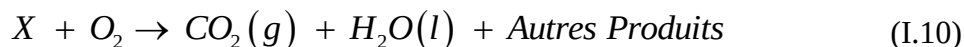
$$\Delta H = QP \quad (I.9)$$

À pression constante (où seul le travail PV est autorisé), la variation de ΔH enthalpie du système est égale à la circulation de l'énergie sous forme de chaleur. Cela signifie que pour une réaction étudiée à pression constante, le flux de chaleur est une mesure de la variation d'enthalpie pour le système. Pour cette raison, la chaleur termes de réaction et la variation d'enthalpie sont utilisés entre la variation pour les réactions étudiées à pression constante. (SEROUTI, 2012).

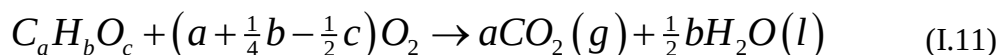
1.3.2 Calcul de l'enthalpie standard de combustion

La chaleur de combustion d'une substance à 25°C peut être calculée à partir des données d'enthalpie de formation ($\Delta_f H^\circ$) dans le livre de "Thermochemical Data of Organic Compounds"

(J. B. Pedley, 1986). Nous pouvons écrire la réaction de combustion générale comme suit :



Pour un composé contenant uniquement du carbone, de l'hydrogène et de l'oxygène, la réaction est simplement :



Et la chaleur standard de combustion $\Delta_c H^\circ$, qui est définie comme la valeur négative du changement d'enthalpie pour la réaction (c'est-à-dire la chaleur libérée lors du processus de combustion), est donnée par :

$$\Delta_c H^\circ = -a\Delta_f H^\circ(CO_2, g) - \frac{1}{2}b\Delta_f H^\circ(H_2O, l) + \Delta_f H^\circ(C_a H_b O_c) \quad (I.12)$$
$$393.51a + 142.915b + \Delta_f H^\circ(C_a H_b O_c)$$

Cette équation s'applique si les réactifs commencent dans leur état standard (25°C et une pression atmosphérique) et que les produits reviennent aux mêmes conditions.

La même équation s'applique à un composé contenant un autre élément si cet élément se termine dans son état de référence standard (par exemple, l'azote, si le produit est N_2) ; en général, cependant, les produits exacts contenant les autres éléments doivent être connus pour calculer la chaleur de combustion (Chemistry-Archive, 2007).

Le tableau 1 donne la chaleur de combustion standard calculée de cette manière pour quelques substances représentatives.

Formule M	Nom	$\Delta_c H^\circ$ kJ mo ⁻¹ -1	Formule M	Nom	$\Delta_c H^\circ$ kJ mo ⁻¹ -1
Substances inorganiques					
C	Carbon (graphite)	393,5	C ₃ H ₈ O	1-Propanol (1)	2021.3
CO	Carbon monoxyde (g)	283	C ₃ H ₈ O ₃	Glycérol (1)	1655.4
H ₂	Hydrogène (g)	285.8	C ₄ H ₁₀ O	D'éthyle éther (1)	2723.9
H ₃ N	Ammoniac (g)	382.8	C ₅ H ₁₂ O	1-Pentanol (1)	3330.9
H ₄ N ₂	Hydrazine (g)	667.1	C ₆ H ₆ O	Phénol (s)	3053.5
N ₂ O	Nitrous oxide (g)	82.1	Composés carbonylés		
Hydrocarbonés			CH ₂ O	Formaldéhyde (g)	570,7
CH ₄	Méthane (g)	890.8	C ₂ H ₂ O	Cétène (g)	1025.4
C ₂ H ₂	Acétylène (g)	1301.1	C ₂ H ₄ O	Acétaldéhyde (1)	1 166.90
C ₂	Ethylène (g)	1411.2	C ₃ H ₆ O	Acétone (1)	1789.9
C ₂ H ₆	Ethane (g)	1560.7	C ₃ H ₆ O	Propenal (1)	1822.7
C ₃ H ₆	Propylène (g)	2058	C ₄ H ₈ O	2-Butanone (1)	2444.1
C ₃ H ₆	Cyclopropane (g)	2091.3	Acides et esters		
C ₃ H ₈	Propane (g)	2219.2	CH ₂ O ₂	Formique Acid (1)	254.6
C ₄ H ₆	1,3-Butadiene (g)	2541.5	CiH ₄ O ₂	Acétique Acid (1)	874.2
C ₄ H ₁₀	Butane (g)	2877.6	C ₂ H ₄ O ₂	Methyl format (1)	972.6
C ₅ H ₁₂	Pentane (1)	3509	C ₃ H ₆ O ₂	Méthyl acétate (1)	1592.2
C ₆ H ₆	Benzène (1)	3267.6	C ₄ H ₈ O ₂	Ethyle acétate (1)	2238.1
C ₆ H ₁₂	Cyclohexane (1)	3919.6	C ₇ H ₆ O ₂	Benzoïque Acid (s)	3226.9
C ₆ H ₁₄	Hexane (1)	4163.2	Composés azotés		
C ₇ H ₈	Toluène (1)	3910.3	CHN	Hydrogène cyanide (g)	671.5
C ₇ H ₁₆	Heptane (1)	4817	CH ₃ NO ₂	Nitrométhane (1)	709.2
C ₁₀ H ₈	Naphtalène (s)	5156.3	CH ₅ N	Méthylamine (g)	1085.6
Alcools et éthers			C ₂ H ₃ N	Acétonitrile (1)	1247.2
CH ₄ O	Méthanol (1)	726.1	C ₂ H ₅ NO	Acétamide (s)	1184.6
C ₂ H ₆ O	Ethanol (l)	1 366.80	C ₃ H ₉ N	Triméthylamine (g)	2443.1
C ₂ H ₆ O	Diméthyle éther (g)	1460.4	C ₅ H ₅ N	Pyridine (1)	2782.3
C ₂ H ₆ O ₂	Éthylène glycol (1)	1189.2	C ₆ H ₇ N	Aniline (1)	3392.8

Tableau 1 - Chaleur de combustion (quelques substances) calculée par l'équation I-12

1.3.3 Calcul de l'enthalpie de formation standard

L'énergie libérée au cours de la combustion d'une substance dépend de l'enthalpie de formation de cette même substance, selon :

$$\Delta_c H^0 = \sum_j \nu_j \Delta_f H_j^0 - \sum_i \nu_i \Delta_f H_i^0 \quad (I.13)$$

Où $\Delta_c H^0$: Enthalpie standard de combustion ;

$\Delta_f H^0$: Enthalpie standard de formation ;

i : réactif (s) ; j : produit (s) ;

ν_j et ν_i : coefficients stœchiométriques.

Il est ainsi primordial de favoriser la synthèse de composés à enthalpie de formation élevée (SALMON, 2006).

1.3.4 Détermination expérimentale

1.3.4.1 Histoire du calorimètre

L'étude énergétique de changements de phase à l'état solide nécessite l'utilisation de techniques qui relèvent principalement de la mesure calorimétrique, de l'analyse thermique ou des deux à la fois car l'une et l'autre se complètent : la première fournit le bilan thermique du phénomène étudié sans en préciser le mécanisme, les informations transmises par la seconde sont qualitatives et ne permettent pas en général la mesure des énergies mises en jeu.

Les premiers calorimètres à glace ont été construits sur la base du concept de chaleur latente de Joseph Black, introduit en 1761. Antoine Lavoisier a inventé le terme de calorimètre en 1780 pour décrire l'appareil qu'il utilisait pour mesurer la chaleur provenant de la respiration des cobayes utilisés pour faire fondre la neige. En 1782, Lavoisier et Pierre-Simon Laplace ont expérimenté les calorimètres à glace, dans lesquels la chaleur nécessaire pour faire fondre la glace pouvait être utilisée pour mesurer la chaleur provenant de réactions chimiques (Dailypress, 2019).

Un calorimètre de bombe est un calorimètre à volume constant, construit pour résister à la pression produite par la réaction lorsqu'il chauffe l'air à l'intérieur du conteneur. Le changement

de température de l'eau est utilisé pour calculer la chaleur de combustion (M.E. Minas da Piedade, 1999).

Bombes calorimétriques à oxygène est la méthode la plus largement utilisée pour obtenir des enthalpies de formation de composés organiques solides et liquides.

Un très bon compte rendu de l'histoire de la calorimétrie par combustion, il y a jusqu'à vingt ans, a été donné par Domalski. La naissance de cette technique remonte aux travaux de Berthelot et de ses collègues de la période 1881 à 1905. Ces chercheurs ont développé le précurseur des calorimètres à bombe statique modernes, qui peuvent être appliqués à des composés organiques contenant du carbone, de l'hydrogène, oxygène et azote.

Afin de dériver des enthalpies de combustion et de formation à partir de mesures calorimétriques de combustion, les réactifs et produits, ainsi que leurs états, doivent être rigoureusement définis.

Les états des produits de la combustion des composés organo-soufrés et organo-halogénés dans une bombe statique sont cependant pratiquement impossibles à caractériser. Des solutions acides de différentes concentrations se forment dans différentes parties de l'intérieur de la bombe et, de plus, lorsque des composés de chlore et de brome sont brûlés, les produits de combustion comprennent à la fois l'halogène libre et l'hydracide halogène dans des proportions qui ne peuvent pas être déterminées avec précision. Le problème de l'atteinte d'un état final bien défini dans des expériences impliquant des composés organiques contenant du soufre et des halogènes a été abordé par plusieurs chercheurs et a finalement conduit au développement des calorimètres à bombe mobile.

1.3.4.2 Importance

Bombes calorimétriques à oxygène est l'un des dispositifs de base nécessaires à certaines recherches thermochimiques. Bien qu'il s'agisse d'une méthode classique pour déterminer les énergies de combustion de certains éléments, composés organiques, métallorganiques, etc., la calorimétrie de combustion est largement appliquée pour obtenir les enthalpies molaires standard de formation de nombreuses substances organiques importantes dans l'industrie et la recherche scientifique basée sur les données d'enthalpies de combustion et de certaines grandeurs thermodynamiques auxiliaires. Ces données thermodynamiques jouent un rôle très important

dans la production et l'utilisation des carburants et des agents propulsifs, le calcul des différentes propriétés difficile à obtenir à travers certaines expériences sur certaines substances organiques importantes au moyen de la théorie de la chimie quantique, la molécule assistée par ordinateur. La conception de médicaments et de nouveaux matériaux, la destruction des déchets dans les incinérateurs, etc. Les enthalpies molaires standard de formation, ainsi que l'entropie standard ou spécifiée, sont des données tout aussi importantes pour déterminer tout équilibre chimique.

(J. Chen, Yy Di, Zc Tan, Ht Zhang, L. Sun , 2005)

1.3.4.3 La bombe calorimétrique

Une bombe calorimétrique est un « porte-échantillon en acier, habituellement cylindrique, muni d'un fil de mise à feu intérieure par effet Joule et fermé par un bouchon fileté étanche, capable de supporter une combustion explosive en présence de 30 bar d'oxygène ». Placée dans un calorimètre, elle permet de mesurer le dégagement de chaleur au cours d'une réaction chimique mettant en jeu des gaz et effectuée à volume constant. Dans ces conditions, la chaleur mise en jeu est égale à la variation de la fonction d'état énergie interne U du système réactionnel : $\Delta U = Q_v$. La chaleur ne dépendant plus de la manière de procéder, on peut alors la déterminer.

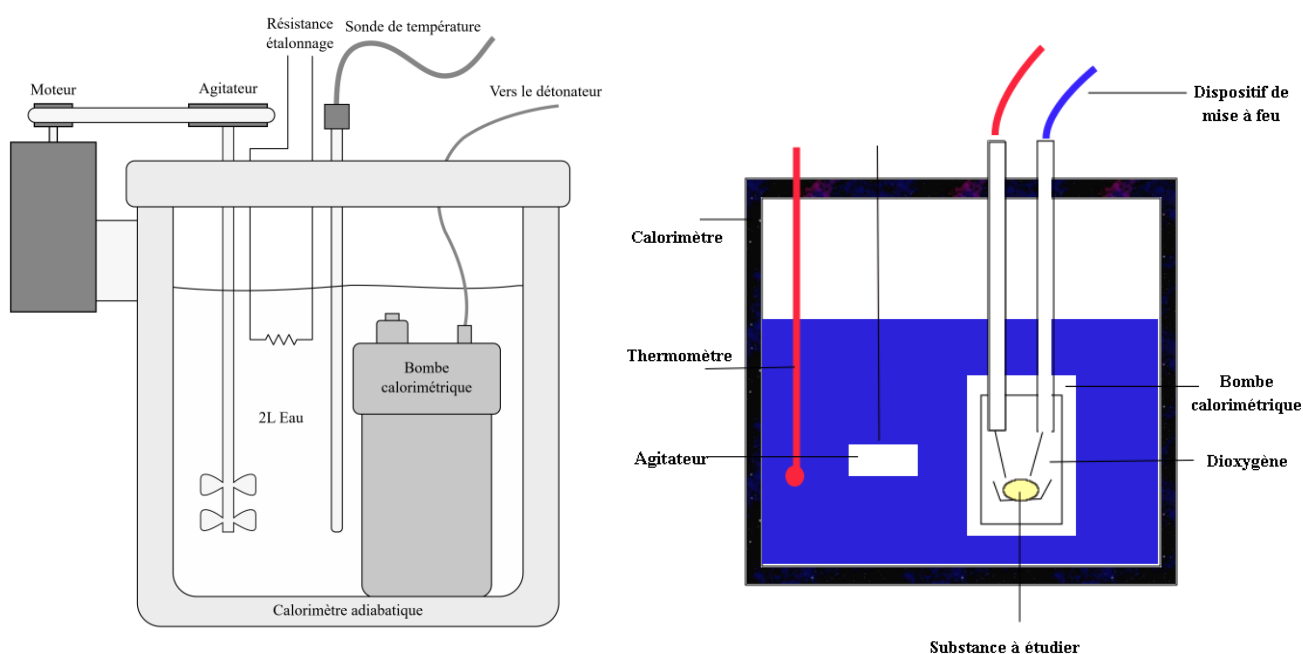


Figure 1 - Schéma simplifié du calorimètre et de la bombe calorimétrique

Le principe de fonctionnement de la bombe calorimétrique remplit ces trois conditions. Il s'agit d'un cylindre en acier, bon conducteur de la chaleur et très résistant à la pression, qui peut atteindre instantanément après la réaction plusieurs centaines de bars. On dispose au fond de ce cylindre un creuset dans lequel sera déposée la substance à étudier. Le cylindre est alors rempli de comburant (dioxygène à une pression de 20 à 30 bars, pour être sûr qu'il sera en excès afin d'assurer une combustion complète). Le dispositif de mise à feu est constitué d'un fil métallique au contact de la substance, qui sera traversé par un courant de forte intensité pour déclencher la combustion. La bombe est placée dans un calorimètre comprenant une grande quantité d'eau distillée, muni d'un dispositif d'agitation et d'un thermomètre de précision. La quantité d'eau est importante pour que l'élévation de température ne dépasse pas quelques degrés. Après avoir mesuré la température initiale, on déclenche la mise à feu.

Lorsque la combustion est effectuée dans la bombe, la chaleur de combustion sera transférée à l'eau et à toutes les composantes à l'intérieur du calorimètre pour élever la température légèrement. Afin d'assurer une température homogène dans le système, l'eau sera agitée à l'aide d'un agitateur mécanique relié à un moteur par une courroie d'entraînement. La mesure de température dans le calorimètre (on mesure la température de la fin de la combustion lorsque l'équilibre thermique du calorimètre est atteint) s'effectue grâce à une thermistance de grande précision qui est reliée à un pont de Wheatstone permettant de convertir la résistance en différence de potentiel. Cette différence de potentiel sera ensuite mesurée par un voltmètre interfacé à un ordinateur qui recueillera les données (TURGEON, 2013).

1.3.4.4 Mesures d'enthalpie de combustion en phase solide

Les chaleurs de réactions sont souvent déterminées expérimentalement par des mesures calorimétriques à P constante ou à V constant. Pour étudier les changements d'énergie dans les réactions dans des conditions de volume constant, une "bombe calorimétrique" est utilisée.

(*) La mesure de chaleur de réaction par calorimétrie s'applique aux réactions rapides, totales, uniques et exothermiques (TURGEON, 2013).

Afin de connaître l'énergie dégagée par la réaction, il est primordial de connaître la capacité calorifique du système (eau + bombe + produits de réaction dans la bombe). Pour y arriver, il est nécessaire de faire un étalonnage, c'est-à-dire introduire une quantité connue d'énergie dans le système et mesurer l'élévation de la température causée par cet ajout. En assumant

que la dépendance en température de la capacité calorifique est faible, il est possible de la relier à la chaleur et l'élévation de température de la façon suivante :

$$q = C_{\text{sys}} \Delta T \quad (\text{I.14})$$

Une méthode efficace pour introduire une quantité connue d'énergie connue dans le système est d'y insérer une résistance électrique et d'y faire circuler un courant pendant un temps t . L'énergie produite par la résistance pourra alors être déterminée de la façon suivante :

$$q_{\text{elec}} = P \times t = V \times I \times t \quad (\text{I.15})$$

Où P est la puissance électrique (en Watt), V la tension (en V) et I le courant (en ampères).

Une fois la capacité calorifique du système connue, il devient possible déterminer précisément l'énergie dégagée par la combustion d'une substance. Il est important de faire un étalonnage après chaque détonation puisque la capacité calorifique du calorimètre peut changer d'un essai à l'autre. En effet, elle dépend fortement de la quantité d'eau à l'intérieur.

1.3.4.4.1 Interprétation des thermo-grammes

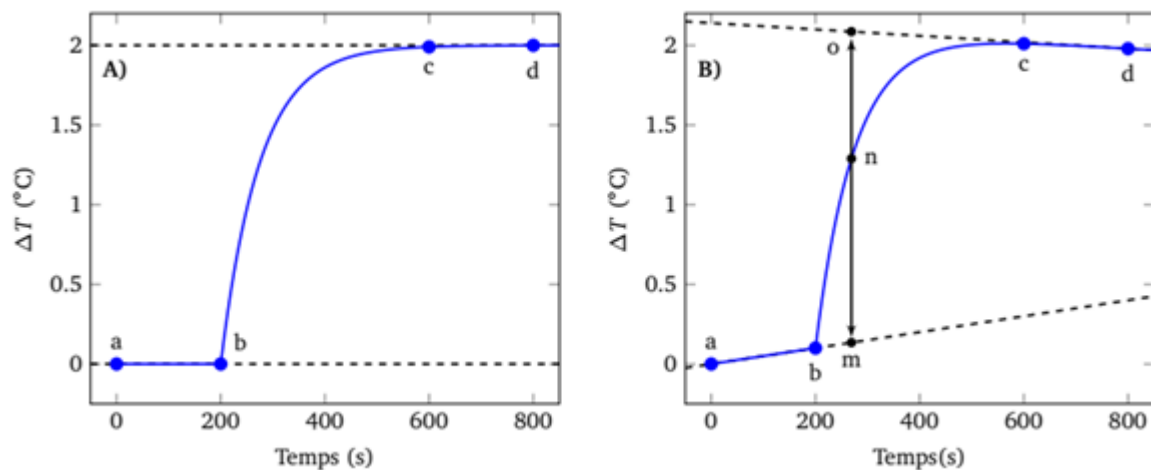


Figure 2 - Thermogramme idéal et réel

(*) **Figure 2** :

- **A** → Thermogramme idéal qui comprend la période précombustion a-b, la combustion b-c et la période postcombustion c-d.

- **B** → Thermogramme réel où les périodes a-b et c-d ont une pente non nulle qui provient des échanges de chaleur avec l'environnement

Les données recueillies lors de la combustion d'une substance à l'intérieur de la bombe calorimétrique seront représentées sous la forme d'un thermogramme tel qu'illustré à la figure 6A). La période entre les points a et b est appelée période précombustion. Pendant cette période, la température reste stable dans la mesure où le calorimètre est adiabatique ou encore s'il est à l'équilibre thermique avec l'environnement. La combustion se produit au point b et la température s'élève en suivant une décroissance exponentielle inverse jusqu'au point c. Au-delà de ce point (segment c-d), la température se stabilise à nouveau, c'est la période postcombustion. Pour le cas idéal, l'élévation de température due à la combustion est évaluée simplement par la différence entre T_c et T_b .

Pour un thermogramme comme celui de la figure 6B, l'élévation de la température due à la combustion est plus difficile à évaluer et il faudra procéder de la façon suivante :

- Évaluer le point où l'élévation de température atteint 60% entre les températures T_b et T_c . Ce point sera dénoté n .
- Extrapoler la droite a-b jusqu'au temps du point n pour obtenir la température m
- Extrapoler la droite c-d jusqu'au temps du point n pour obtenir la température o
- Évaluer ΔT à partir des points o et m : $\Delta T = T_o - T_m$ (I.16)

1.3.4.5 Calcul de l'enthalpie de combustion à partir des données expérimentales

Après avoir réalisé la détonation pour déterminer l'élévation de température due à la combustion et avoir fait l'étalonnage électrique du calorimètre pour en déterminer la capacité calorifique, il est possible de déterminer l'énergie dégagée par la combustion :

$$q_c = C_{\text{sys}} \Delta T_c \quad (\text{I.17})$$

Cependant, il faut tenir compte de la masse de fil d'ignition qui a été brûlé lors de la détonation puisque la chaleur de combustion qui en résulte est non-négligeable. Il sera nécessaire d'apporter une correction de 5,8576 kJ/g pour le fil brûlé

$$q_c = C_{sys} \Delta T_c - q_{fil} \quad (I.18)$$

- **Pour un processus à volume constant** (p.ex. bombe calorimétrique), le dégagement de chaleur q_c correspondra directement à la variation d'énergie interne ΔU .
- **Pour un processus à pression constante** (p.ex. réaction chimique acide-base), le dégagement de chaleur q_c correspondra directement à la variation d'enthalpie ΔH .
- **Si l'équation chimique d'une réaction de combustion à volume constant est connue**, il est possible de calculer la variation d'enthalpie ΔH à partir de l'équation :

$$\Delta H = \Delta U + \Delta_{ng} RT \quad (I.19)$$

Pour un échantillon pur dont la formule chimique est connue, il est intéressant de ramener les quantités mesurées en **quantités molaires**. Pour les échantillons qui sont des mélanges (p. ex. carburant de voiture) ou qui sont des matériaux complexes (p. ex. bois ou charbon), il convient de ramener les quantités mesurées en **quantités massiques**.

1.4 Méthodes de calculs

1.4.1 Méthodes de contribution de groupes

Le concept de contributions de groupes repose sur le fait qu'il existe une infinité de molécules qui peuvent être « construites », d'un point de vue structure moléculaire, par assemblage d'un nombre restreint de groupes élémentaires. S'il est illusoire de penser pouvoir recueillir des données expérimentales sur les propriétés de toutes les molécules, on peut envisager le développement de méthodes permettant l'estimation de ces propriétés en décomposant les valeurs de ces propriétés en contributions que l'on attribue aux groupes constitutifs de la molécule.

La correspondance entre un groupe et sa contribution à une propriété permet ainsi la prédiction de la propriété considérée pour une structure moléculaire originale, constituée de plusieurs groupes dont la contribution est connue (VITU, 2007).

Une méthode de contribution de groupe utilise le principe selon lequel certains aspects simples des structures de composants chimiques sont toujours les mêmes dans de nombreuses molécules différentes. Les plus petits constituants courants sont les atomes et les liaisons. La grande majorité des composants organiques, par exemple, sont construits de carbone, l'hydrogène, l'oxygène, l'azote, les halogènes, et peut - être le soufre ou le phosphore. Ensemble avec un seul, un double, et une liaison triple il n'y a que dix types d'atomes (à l'exclusion

astatine) et trois types de liaisons pour construire des milliers de composants. Les prochains blocs de construction légèrement plus complexes de composants sont des groupes fonctionnels qui sont eux - mêmes construits de quelques atomes et liaisons.

Un procédé de contribution de groupe est utilisé pour prédire les propriétés des composants purs et des mélanges en utilisant des propriétés de groupe ou atome. Cela réduit le nombre de données nécessaires de façon spectaculaire. Au lieu d'avoir besoin de connaître les propriétés des milliers ou des millions de composés, seules les données pour quelques dizaines ou des centaines de groupes doivent être connus (Article, 2020).

Certaines méthodes plus récentes introduisent des groupes de second ordre. Ceux-ci peuvent être super-groupes contenant plusieurs groupes de premier ordre (standard). Cela permet à l'introduction de nouveaux paramètres pour la position des groupes. Une autre possibilité consiste à modifier les contributions des groupes de premier ordre si d'autres groupes spécifiques sont également présents.

Plusieurs types de groupes fonctionnels peuvent être définis. Ils sont généralement organisés en un système hiérarchique :

- Les groupes d'ordre 0 sont des atomes, et le calcul d'une propriété est effectué en sommant les contributions de chacun des atomes de la molécule considérée ;
- La décomposition en groupes d'ordre 1 consiste à découper la molécule en groupes d'atomes (tels que -CH₂-, -CH₃ ou -OH). Leurs contributions à une propriété donnée sont sommées sans que l'environnement de chacun des groupes dans la molécule ne soit pris en considération. Ainsi, le groupe -CH₂- a une contribution fixe, qu'il soit relié à un carbone ou à un groupe oxygéné. Ces groupes sont assez souvent employés, car ils permettent d'estimer rapidement la valeur d'une propriété, avec une précision parfois suffisante (par exemple pour l'enthalpie de formation). Cependant, les résultats obtenus, pour la température d'ébullition par exemple, ne sont pas toujours satisfaisants. De plus, certains isomères peuvent conduire à la même décomposition : il est alors impossible de les distinguer par cette méthode ;
- Les groupes d'ordre 2 sont constitués des atomes centraux de la molécule (autres que H), accompagnés de leurs plus proches voisins, c'est-à-dire de tous les atomes auxquels ils sont reliés. Contrairement aux groupes d'ordre 1, ceux d'ordre 2 tiennent compte de l'environnement des atomes (SEROUTI , 2012).

Exemple :

Le Tableau 2 Présente une comparaison des décompositions des molécules de :

Butan-2-ol et 2-méthylpropan-1-ol, représentées sur la Figure 3, On observe que : Seule la décomposition en groupes d'ordre 2 permet de distinguer ces deux molécules.

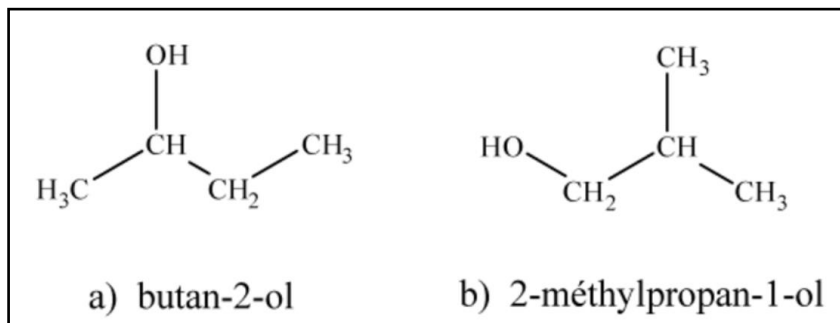


Figure 3 -Exemple de deux molécules, isomères de constitution

Le tableau ci-dessous décrit le découpage à effectuer dans le cas de l'emploi d'une méthode de contribution de groupes d'ordre 0, 1 ou 2.

Méthode	Groupe	Nbre de groupe (a)	Nbre de groupe (b)
Groupe 0	C	4	4
	H	10	10
	O	1	1
Groupe 1	CH	1	1
	-CH ₂ -	1	1
	-CH ₃ -	2	2
	-OH	1	1
Groupe 2	C - (C)(H) ₃	2	2
	C - (C) ₂ (H) ₂	1	0
	C - (C)(O)(H) ₂	0	1
	C - (C) ₃ (H)	0	1
	C - (C) ₂ (O)(H)	1	0

Tableau 2 - Découpage de deux molécules en groupes d'ordre 0, 1 et 2.

On observe que seule la décomposition en groupes d'ordre 2 permet de distinguer ces deux molécules.

De nombreuses méthodes s'appuient donc sur des groupes des trois ordres pour améliorer la précision des prédictions et différencier les isomères. Elles sont principalement utilisées pour prédire des propriétés thermodynamiques, par exemples des propriétés critiques (température

ou pression critique) et de nombreuses grandeurs énergétiques. Elles présentent également l'avantage de permettre l'estimation de propriétés de mélanges, par addition des contributions des composants du mélange.

1.4.1.1 Classification des méthodes contributions des groupes

Les méthodes des contributions des groupes peuvent être classées comme suit : méthodes basées sur les contributions des groupes simples et des méthodes basées sur les contributions des interactions des groupes.

1.4.1.1.1 Méthode basée sur le principe de contributions des groupes simples

Cette classification est basée sur la décomposition de la structure moléculaire des composées, en plusieurs groupes fonctionnelles simple. Sont définit comme étant de méthodes qui basées sur l'addition de groupes principaux qui constituent les molécules tous les méthodes additionnelles utilisent de corrélations correspondant à :

$$P = f(P) = \sum_i N_i C_i \quad (I.20)$$

Où

C_i : Est la contribution de groupe.

N_i : Est le nombre de fois de contribution de groupe.

1.4.1.1.2 Méthodes basées sur contribution des interactions de groupe

C'est la somme des contributions des interactions formées de deux ou plusieurs groupes simples et voisines. Elles sont basées sur la relation suivant (MAZOUZI Aziza, 2013) :

$$P = f(P) = \sum_i M_i L_i \quad (I.21)$$

Où

M_i : Nombre des groupes d'interactions.

L_i : Groupes d'interactions.

1.4.2 Méthode de Benson

1.4.2.1 Définition

La **méthode de Benson** utilise des contributions de groupes de deuxième ordre. Les relations suivantes permettent d'estimer des enthalpies standard de formation en phase gazeuse à 298,15K, des capacités calorifiques, des entropies standard à 298,15 K et des enthalpies libres standard de formation en phase gazeuse à 298,15 K :

$$\begin{aligned}\Delta H_f^0 &= \sum_k N_k (d_f h_k) \\ c_p^0(T) &= \sum_k N_k (c_{pk}) \\ S^0 &= \sum_k N_k (s_k) + S_s^0, \text{ avec} \\ S_s^0 &= \sum_k R \ln(N_{oi}) - R \ln(N_{ts}) \\ \Delta_f G_g^0 &= \Delta_f H_g^0 - 298,15 \times (S^0 - S_{el}^0), \text{ avec} \\ S_{el}^0 &= \sum_e N_e (S_e^0)\end{aligned}\tag{I.22}$$

Où

N_k : nombre de groupes de type k dans la molécule ;

$d_f h_k$, c_{pk} , s_k : contributions de groupes ;

S_s^0 : entropie standard de symétrie de la molécule ;

N_{oi} : nombre d'isomères optiques de la molécule (*i.e.* énantiomères) ;

N_{ts} : nombre de symétrie total ;

S_{el}^0 : entropie absolue des éléments formant la molécule à T et 101 325 Pa ;

S_e^0 : entropie absolue d'un élément contenu dans la molécule à T et 101 325 Pa.

La relation qui nous intéresse ici est l'équation (40), permettant d'évaluer des enthalpies standard de formation de composés à l'état de gaz parfait.

A noter que la méthode de Benson présente un domaine d'application assez étendu qui comprend, outre les éléments C, H, O, N, S et P, les halogènes (X), des métaux (Al, Ge, Pb, ...), des métaux de transition (Ti, V, Cr, Fe, ...) et des alcalins et alcalino-terreux (Na, K, Mg, ...) (SALMON, 2006).

1.4.2.1.1 Définition des pseudo-éléments

Chaque élément constitutif d'un groupe est appelé pseudo-élément. Ainsi, dans le groupe C-(N)(CO)(C)(H), les atomes ou groupes d'atomes C, N, CO et H sont des pseudo-éléments du groupe de deuxième ordre précité.

La méthode de Benson différencie plusieurs types d'atomes de carbone, notamment suivant l'hybridation des orbitales atomiques :

Hybridation	Type de liaison	Appellation
sp^3	Liaison simple	C
sp^2	Liaison double	C _d
sp	Hybridation	Type de liaison

Tableau 3 - Appellations de l'atome de carbone.

Par ailleurs, CB représente un atome de carbone appartenant à un noyau benzénique et CBF un atome de carbone appartenant à un noyau benzénique, lui-même accolé à un autre noyau benzénique. CA correspond à un atome de carbone allénique et CI représente un atome de carbone dans un groupement imino.

De même, il existe différents atomes d'azote. A l'élément N viennent s'ajouter les atomes nommés NI (N d'un groupement imino) et NA (N d'un groupement azo).

Des pseudo-éléments polyatomiques se distinguent : CO (groupement carbonyle), CN (groupement cyano), NO (groupement nitroso) et NO₂ (groupement nitro). De plus, N₃ correspond à la fonction azidure.

La technique de Benson est connue pour être la plus précise des méthodes de contribution de groupes utilisées en thermochimie. Elle permettait à l'origine l'estimation d'enthalpie standard de formation en phase gazeuse, d'entropie standard et de capacités calorifiques à différentes températures. Mais, telle que Benson l'a développée, elle ne s'appliquait qu'aux composés à l'état de gaz parfait. Or, c'est la phase condensée qui intervient dans les projets industriels (SALMON, 2006).

1.4.3 Extension de la méthode de Benson à la phase liquide et à la phase solide

1.4.3.1 Méthode de Domalski et Hearing

Domalski et Hearing ont publié, en 1993, une étude portant sur 1 512 molécules contenant les éléments carbone, hydrogène, oxygène, azote, soufre, halogènes, dans les trois états. La liste des composés étudiés ne prétend pas être exhaustive. L'objectif essentiel de cette étude est de démontrer l'applicabilité de la méthode de Benson à la phase condensée. Concernant la phase solide, 57 composés C_xH_y , 140 composés $C_xH_yO_z$ et 147 composés $C_xH_yN_wO_z$ ont permis d'établir l'ensemble des contributions de groupes de Benson C-H, C-HO et C-H-O-N utiles à l'estimation d'enthalpies standard de formation, soit 344 molécules au total.

Une procédure systématique a été utilisée afin de déterminer les différentes contributions de groupes. Les auteurs se sont tout d'abord intéressés aux n-alcanes. Des contributions ne contenant que des atomes de carbone et d'hydrogène ont été calculées. Ces valeurs ont été fixées et ont ensuite permis de déterminer des contributions de groupes intervenant dans des alcanes substitués, puis, des alcènes, des alcynes, des composés aromatiques, etc...

La différence essentielle de cette technique par rapport à la méthode introduite par Benson est le traitement des interactions répulsives engendrées par les atomes d'hydrogène des groupes méthyles liés à des carbones tertiaires ou quaternaires. Une contribution de groupe est attribuée aux molécules présentant ces atomes de carbone. Par exemple, le 2-méthylbutane requiert 2 corrections de ce type tandis que le 3-éthylpentane n'en nécessite pas. Cette approche remplace celle de Benson vis-à-vis des interactions gauches.

L'étude comparative, dans les trois phases, des valeurs mesurées et calculées d'enthalpies de formation, de capacités calorifiques et d'entropies est présentée. Les 26 estimations effectuées sur les trois phases ont montré que 67% des écarts entre les enthalpies de formation issues de la littérature et celles qui ont été calculées étaient inférieurs à 4 kJ.mol⁻¹. 16% d'entre eux étaient compris entre 4 et 8 kJ.mol⁻¹, et les 17% restants étaient supérieurs à 8 kJ.mol⁻¹.

Il convient de souligner que les auteurs ne différencient pas les trois phases, mais donnent un résultat global. Les bilans pour chacune des trois phases, prises séparément, sont absents de l'étude publiée.

1.4.4 Méthodes de descripteurs moléculaires

L'application de modèles de relation quantitative structure-propriété (QSPR) à la prédiction et à l'estimation des propriétés physiques des matériaux se développe largement. Dans QSPR, des méthodes mathématiques avancées (algorithme génétique, réseaux de neurones, etc.) sont utilisées pour trouver une relation entre la propriété d'intérêt et les propriétés moléculaires de base qui sont obtenues uniquement à partir de la structure chimique des composés et appelées "descripteurs moléculaires" (Ali Vatani, 2007).

Il existe de nombreuses méthodes de calcul de ΔH . L'une des méthodes est classée en méthodes de contribution de groupe : méthode de "Ali Vatani, Mehdi Mehrpooya et Farhad Gharagheizi".

1.4.4.1 Méthode Vatani, Mehrpooya et Gharagheizi

Vatani, Mehrpooya et Gharagheizi ont développé une méthode pour estimer l'enthalpie de formation standard. Les modèles développés sont applicables pour les composés organiques. Le principe de la technique utilisée est basé sur les descripteurs moléculaires. Dans cette méthode, étudié un nouveau modèle de QSPR pour la prédiction de ΔH_f^0 de 1115 composés organiques est présentée. Ces composés appartiennent à toutes les familles de matériaux, à la suite le modèle obtenu peut être appliqué pour la prédiction de ΔH_f^0 pour un composé. Le meilleur modèle obtenu a six paramètres et est présenté ci-dessous (VATANI, et al., 2007) :

$$\Delta H_f^0 = 50.1688 - 80.52012nSK + 5364546CBO - 169.21889SCBO - 174.75477nF - 266.57659nHM \quad (I.23)$$

Les équations caractéristiques des modèles développés peuvent s'écrire sous la forme suivante :

$$\Delta H_f^0 = (w_0 - w_1) * nSK + w_2 * CBO + w_3 * SCBO + w_4 * nF + w_5 * nHM \quad (I.24)$$

Où les descripteurs moléculaires de l'équation (I.24) et leur signification sont présentés dans le tableau ci-dessus.

Variable	Signification du descripteur moléculaire
w_i	Coefficient de régression i= 1, 2, 3, 4, 5 ;
nSK	Nombre d'atomes non-H
SCBO	Somme des ordres de liaison classiques (H-appauvri)
nO	Nombre d'atomes d'oxygène
nF	Nombre d'atomes de fluor
nHM	Nombre d'atomes lourds

Tableau 4 - Les descripteurs moléculaires de l'équation (I.24) et leur signification.

Les paramètres statistiques d'ajustement pour l'équation :

$$\Delta H_{\text{réaction}} = \sum_p \Delta H_f - \sum_r \Delta H_f \quad (\text{I.25})$$

Sont les suivants :

$$R^2 = 0.983, F=10239.02, s=58.541, Q^2=0.9826,$$

Où R^2 est le coefficient de corrélation, F est le facteur de Fisher, S c'est le deviation standard, et Q^2 est le coefficient de corrélation croisée validée.

Les paramètres statistiques des coefficients de l'Eq (I.24) sont présentés dans le tableau 5.

N°	Variable	Coefficient de régression (w _i)	Erreurs coefficient de régression	Intervalles de confiance (0,95)	Standard coefficient de régression
0	intercept	50.16088267	4.075696	0	0
1	nSK	-80.52012489	1.175295	-1.081325919	3.60196
2	SCBO	53.64545627	0.9837223	0.858696667	3.593514
3	nO	-169.2188895	1.424115	-0.552855974	1.061814
4	nF	-174.7547718	1.178939	-0.680669749	1.047946
5	nHM	-266.5765885	6.857063	-0.171389892	1.006101

Tableau 5 - Les valeurs des constantes de l'Eq (I.24) et leurs interprétations statistiques.

Pour obtenir le meilleur modèle linéaire multi-variée, d'abord, ils ont commencé avec un modèle de descripteur moléculaire et trouvent le meilleur modèle linéaire multivariée, 40 les deux modèles moléculaires descripteurs ont été testés, et le meilleur linéaire multi-variée modèle à deux descripteurs a été trouvé. Ce travail a été repris et le nombre de descripteurs a été augmenté, ils aient constaté que l'augmentation du nombre de descripteurs moléculaires ne pas affecter la régularité du meilleur modèle.

Chapitre 2

Outils et techniques utilisées

Introduction

L'expérience est un moyen direct pour obtenir des données de l'activité des composés organiques, mais elle coûte chère et nécessite beaucoup de temps. Il est donc crucial d'utiliser des méthodes théoriques avant de passer à l'expérience pour prédire les activités des composés. Le développement important de l'informatique, ainsi que des études théoriques de la chimie quantique permettent aux chercheurs d'obtenir des paramètres physico-chimiques et quantiques des composés en un temps plus court.

Il existe de nombreuses méthodes de chimie théorique ayant pour but de déterminer des propriétés physiques ou chimiques de molécules isolées, que ce soient des propriétés thermodynamiques comme des enthalpies de liaison, des énergies relatives de différents conformères...

Les paramètres structuraux ainsi que l'introduction des études de relation quantitative structure-propriété QSPR peuvent interpréter et prédire les propriétés pour de nouveaux composés organiques. Ceci est utilisé pour examiner la relation entre des descripteurs moléculaires d'un ensemble de composés et leurs propriétés biologiques ou propriétés physico-chimiques (EN-NAHLI, 2019).

Toutes les méthodes en rédigé dans premier chapitre est utilisé les contributions des groupes, mais nous on utilise descripteurs moléculaires, pour la prévision d'un modèle mathématique pour calcule l'enthalpie standard de formation en phase solide de molécules organiques. Pour concept la prévision a ce modèle on utilise deux méthodes prédictives : Régression Multi-Linéaire (RML) et Régressions Non-Linéaires (RNL). Toutes ces les méthodes en bases aux descripteurs moléculaires comme variables.

2.1 Analyse quantitative de relation structure-activité/propriété

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

Le principe des méthodes QSPR est, comme leur nom l'indique, de mettre en place une relation mathématique reliant de manière quantitative des propriétés moléculaires aussi bien électroniques que géométriques, appelées descripteurs, avec une observable macroscopique (activité biologique, toxicité, propriété physico-chimique, etc.), pour une série de composés chimiques similaires à l'aide de méthodes d'analyses de données (FAYET, 2010).

L'objectif d'une telle méthode est donc d'analyser les données structurales afin de détecter les facteurs déterminants pour la propriété mesurée. Pour ce faire, différents types d'outils peuvent être employés :

- Régressions multi-linéaires (MLR) (Ghasemi et al., 2007),
- Régressions aux moindres carrés partiels (PLS) (Geladi et Kowalski, 1986),
- Arbres de décision (Myles et al., 2004),
- Réseaux de neurones (Dupart et al., 1998; Tetko et al., 1996; Gasteiger et Zupan, 1993),
- Algorithmes génétiques (Leardi, 2001),
- Machines à vecteur support (Gasteiger et Zupan, 1993).

Une fois cette relation 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, voire pour des molécules non encore synthétisées. De tels modèles peuvent être également utilisés pour mieux comprendre les phénomènes moléculaires mis en jeu dans la propriété /activité d'intérêt et les modes d'action (CHTITA, 2017).

2.1.1 Construction du modèle

La construction d'un modèle QSAR/QSPR est très dépendante des données expérimentales de référence. Le choix de la base de données est donc un point critique de son développement. Dans la plupart des cas, les données expérimentales sont issues de la littérature. Habituellement, les données l'activité à modéliser peut être considéré comme étant un effet

biologique ou d'une propriété physico-chimique. La relation quantitative structure-activité/propriété (QSAR/QSPR) modélisation se rapporte à la construction des modèles prédictifs d'activités biologiques différentes, en fonction de l'information de structure moléculaire et d'une banque de composés. Consiste à trouver une relation mathématique reliant de manière quantitative une activité biologique, ou une propriété, mesurée pour une série de composés similaires dans les mêmes conditions expérimentales, avec des descripteurs moléculaires à l'aide des méthodes statistiques.

L'expression mathématique obtenue peut alors être utilisée comme moyen prédictif de l'activité/propriété étudiée pour de nouvelles molécules ou des molécules pour lesquelles les données expérimentales ne sont pas disponibles (BOUARRA, 2018). Ceci peut être traduit par l'équation suivante : $Propriété = f(Descripteurs)$

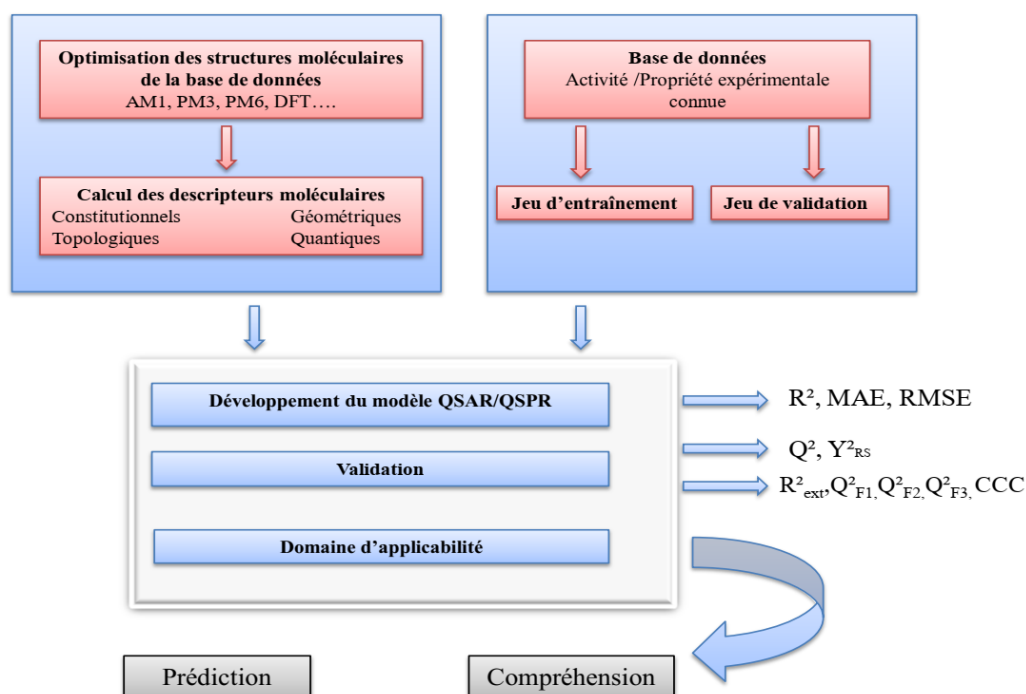


Figure 4 -Principes de la méthode QSAR/QSPR

La stratégie de développement d'un QSAR et plus généralement d'un QSPR va donc s'articuler autour de 5 points (FORTUNÉ, 2006) :

- Constituer une base de données Structure - Propriété à partir de mesures quantitatives, fiables et normalisées de la propriété cible, pour chaque composé. Sélectionner des descripteurs chimiques en relation avec la propriété cible.

- Diviser ce jeu de données en un jeu d'apprentissage et un jeu de test ;
- Construire des modèles à partir de jeu d'apprentissage avec les outils statistiques souhaités. Caractériser ces modèles par leurs indices de validation internes et vérifier leur robustesse par un test de hasardisation ;
- Valider les modèles avec le jeu de test et calculer leur indice de corrélation externe. Répéter l'opération de division pour obtenir d'autres jeux d'apprentissage et de test. La division optimale donne le plus petit jeu d'apprentissage capable de bonnes prévisions pour le plus grand jeu de test;
- Explorer et exploiter les modèles validés pour comprendre les mécanismes possibles et faire des prévisions.

L'ensemble de la procédure de validation est représenté sur la Figure 5

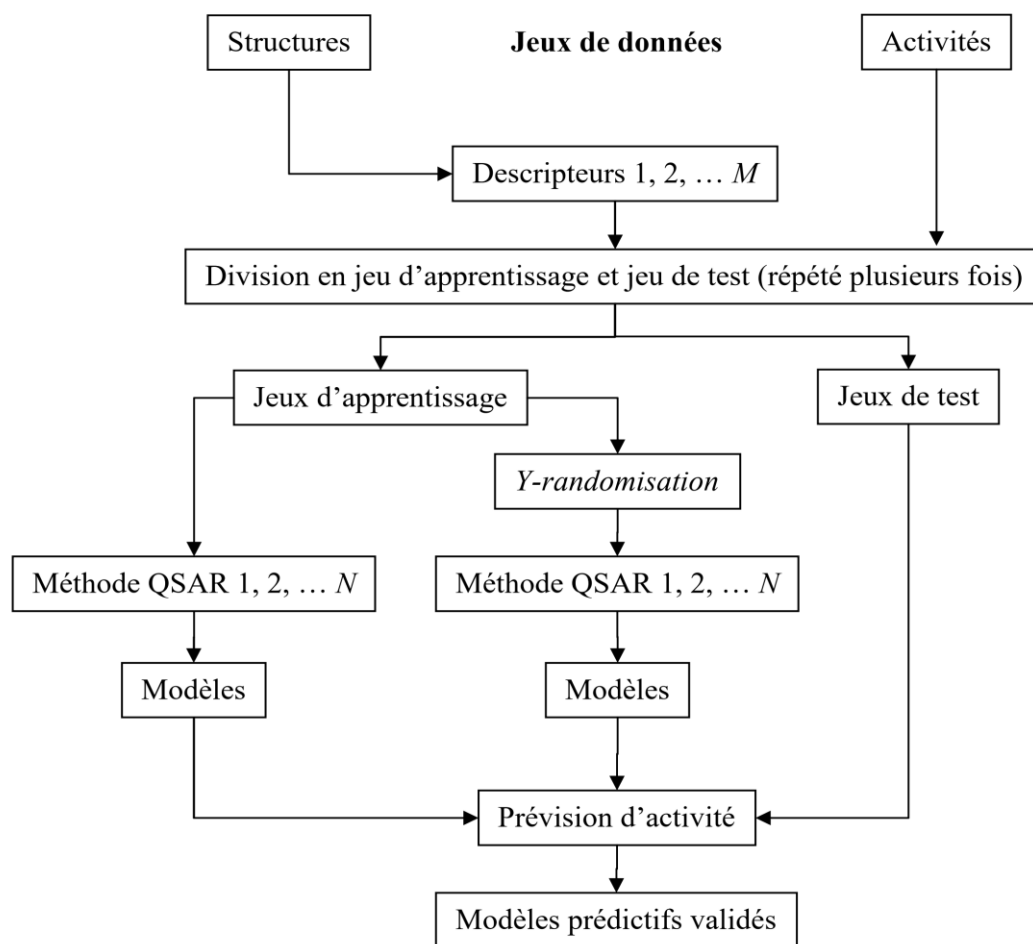


Figure 5 - Procédure de validation des modèles

2.2 Descripteur moléculaire

Un descripteur moléculaire est un paramètre (une valeur numérique) propre à une structure chimique donnée. Ces valeurs peuvent être obtenues expérimentalement ou calculées à partir de la structure de la molécule. Les descripteurs calculés, permettent d'effectuer des prédictions sans avoir à synthétiser les molécules, ce qui est l'un des objectifs de la modélisation moléculaire.

Les descripteurs moléculaires jouent un rôle fondamental dans les études de la relation quantitative structure activité/propriété. Ils sont utilisés en tant que variables indépendantes pour prédire une variable dépendante (activité ou propriété).

L'utilisation des descripteurs moléculaires dans le développement de modèles RQSA/RQSP n'est pas une tâche aisée. Tout d'abord, un très grand nombre de descripteurs moléculaires, de différentes complexités et de conceptions diverses a été introduit au cours des dernières années. Ensuite, pendant ce temps, aucune règle stricte n'a été établie pour la sélection de descripteurs adaptés parmi le grand nombre de descripteurs disponibles. Ce choix a souvent été basé sur l'intuition chimique des chercheurs, ou en se pliant à la tradition. (CHTITA, 2017)

2.2.1 Les types de descripteurs moléculaires

Dans ce qui suit, nous allons présenter que les descripteurs moléculaires les plus utilisés et ceux qui ont été utilisés dans l'ensemble de nos travaux, en commençant par les descripteurs les plus simples, qui nécessitent peu de connaissances sur la structure moléculaire, Nous verrons ensuite comment les progrès de la modélisation moléculaire ont permis d'accéder à la structure 3D de la molécule, et de calculer des descripteurs à partir de cette structure.

Historiquement, deux grands schémas pour la classification des descripteurs moléculaires ont été établis : l'un en fonction de leur **origine** (constitutionnel, topologique, géométrique, quantique, thermodynamique...), et un autre sur leur **dimensionnalité** (0D, 1D, 2D ou 3D). (CHTITA, 2017)

2.2.1.1 Classification des descripteurs en fonction de leur dimensionnalité

2.2.1.1.1 Les descripteurs 0D

Tous les descripteurs moléculaires pour lesquels aucune information sur la structure moléculaire et les connectivités atomiques n'est nécessaire appartiennent à la classe des

descripteurs 0D. Les nombres d'atomes et de liaisons, ainsi que la somme ou la moyenne des propriétés atomiques sont typiques de cette classe de descripteurs.

Ces descripteurs peuvent toujours être facilement calculés, interprétés naturellement, ne nécessitent pas d'optimisation de la structure moléculaire et sont indépendants de tout problème conformationnelle, et leurs valeurs sont précises, essentielles et interviennent régulièrement dans les modèles RQSA/RQSP. Leur contenu informationnel est faible, mais ils peuvent néanmoins jouer un rôle important dans la modélisation de plusieurs propriétés physico-chimiques ou prendre part à des modèles plus complexes

(*) On aura des problèmes au niveau de l'interprétation des mécanismes d'interaction mis en jeu pour l'activité ou la propriété étudiée

2.2.1.1.2 Les descripteurs 1D

Tous les descripteurs moléculaires qui peuvent être calculés à partir des informations de la sous-structure de la molécule appartiennent aux descripteurs 1D. Le comptage des groupes fonctionnels et des fragments de sous-structure, ainsi que des descripteurs centrés sur les atomes, sont les descripteurs 1D les plus connus. Ces descripteurs sont souvent présentés comme des empreintes digitales, c'est-à-dire un vecteur binaire où 1 indique la présence d'une sous-structure et 0 son absence. Un avantage pertinent dans la description des molécules par les empreintes digitales est la possibilité d'effectuer des calculs rapides pour la similarité des molécules/problèmes de diversité. Comme les descripteurs 0D, ces descripteurs peuvent généralement être facilement calculés, sont naturellement interprétés, ne nécessitent pas une optimisation de la structure moléculaire et sont indépendants de tout problème conformationnel. Ils montrent généralement une dégénérescence moyenne élevée et sont souvent très utiles dans la modélisation à la fois des propriétés physico-chimiques et biologiques. (BOUARRA, 2018)

2.2.1.1.3 Les descripteurs 2D

Les descripteurs moléculaires utilisant la représentation des molécules sous forme de graphes sont dits « descripteurs 2D » et contiennent des informations relatives à la connectivité ou à certains fragments moléculaires, mais aussi des estimations des propriétés physicochimiques. C'est à partir de ce niveau que l'on peut espérer la capture d'informations chimiques pertinentes pour la prédiction de la majorité des propriétés moléculaires. On

trouvera dans cette catégorie les descripteurs suivants : **Les indices constitutionnels** et **Les indices topologiques**

De nombreux indices quantifiant la connectivité moléculaire ont été développés en se basant sur cette approche, comme par exemple l'indice de Wiener, Schultz 1989, Balaban 1982, Randić 1975, ...etc.

2.2.1.1.4 Les descripteurs 3D

Ce type de descripteurs nécessitent une conformation 3D de la molécule ; Ils sont évalués à partir des positions relatives de leurs atomes dans l'espace et décrivent des caractéristiques plus complexes ; leurs calculs nécessitent donc de connaître, le plus souvent par « modélisation moléculaire empirique » ou « ab-initio », la géométrie 3D de la molécule. La plupart de ces descripteurs s'avèrent relativement coûteux en temps de calcul, mais apportent davantage d'informations et sont nécessaires à la modélisation de propriétés ou d'activités qui dépendent de la structure 3D (CHTITA, 2017). On distingue plusieurs familles de descripteurs

3D :

- **Les descripteurs géométriques** : Les plus importants sont le volume moléculaire, la surface accessible au solvant, le moment principal d'inertie.
- **Les descripteurs physico-chimiques** : Les descripteurs physicochimiques, (ou indices physicochimiques) certains d'entre eux reflètent la composition moléculaire du composé (le nombre et le type d'atomes et de liaisons présents dans la molécule, le nombre de cycle, les propriétés donneur/accepteur de liaison H, cation, anion, etc....). D'autres représentent le caractère hydrophile ou lipophile de la molécule généralement évalué à partir du coefficient de partage Octanol/eau représenté par le log P.
- **Les descripteurs quantiques/électroniques** : Ces descripteurs caractérisent la distribution de charge des molécules (polarité des molécules) mais aussi les paramètres de la chimie quantique qui, pour être calculés de manière fiable, nécessitent des calculs plus sophistiqués. Les approches de la chimie quantique nous donnent accès à des informations supplémentaires telles que des données structurales, énergétiques, électroniques et spectroscopiques des systèmes étudiés.

- **Les descripteurs thermodynamiques :** Ce sont des descripteurs peu utilisés dans les études RQSA/RQSP. Ils peuvent être exprimés par la fonction de partition Q de la molécule utilisée en thermodynamique statistique ainsi que de ses dérivées. Cette fonction décrit la façon avec laquelle l'énergie d'un système de molécules est répartie parmi les individus moléculaires. Sa valeur dépend du poids moléculaire, de la température, du volume moléculaire, des distances inter nucléaires, des mouvements moléculaires et des forces intermoléculaires. La fonction de partition est le point le plus commode entre les propriétés microscopiques des molécules indépendantes (niveaux d'énergie, moments d'inertie) avec les propriétés macroscopiques (point de fusion, point d'ébullition, entropie). Nous avons utilisé dans nos travaux : La chaleur de combustion (appelé aussi l'enthalpie de combustion), notée H_c , mesurée en KJ/mol, est l'énergie résultant de la combustion d'une mole d'une substance à partir de ses éléments constitutifs à l'état standard (T à 298.15 °K et P à 1 atm)

2.2.1.2 Classification des descripteurs en fonction de leur origine

2.2.1.2.1 Les descripteurs constitutionnels

Ces descripteurs sont très couramment utilisés du fait de leur extrême simplicité non seulement d'un point de vue conceptuel mais surtout calculatoire. Ils sont notamment à la base des analyses par groupes de contribution, comme la méthode de Benson, dans laquelle l'effet des substituants est introduit sous la forme d'une valeur tabulée.

La manière la plus simple de représenter un système moléculaire est de considérer sa composition chimique, sans prise en compte de quelconques considérations géométriques ou électroniques. Des exemples caractéristiques de descripteurs constitutionnels sont : les nombres absolus et relatifs d'atomes (C, H, N...), les nombres absolus et relatifs de groupes fonctionnels (NO_2 , COOH , OH ...), les nombres absolus et relatifs de liaisons (simples, doubles, aromatiques...), les nombres absolus et relatifs de cycles (aromatiques ou non), la masse moléculaire. (CHTITA, 2017)

N°	Nom	Description
01	MW	molecular weight
02	AMW	average molecular weight
03	Sv	sum of atomic van der Waals volumes (scaled on Carbon atom)
04	Se	sum of atomic Sanderson electronegativities (scaled on Carbon atom)
05	Sp	sum of atomic polarizabilities (scaled on Carbon atom)
06	Si	sum of first ionization potentials (scaled on Carbon atom)
07	Mv	mean atomic van der Waals volume (scaled on Carbon atom)
08	Me	mean atomic Sanderson electronegativity (scaled on Carbon atom)
09	Mp	mean atomic polarizability (scaled on Carbon atom)
10	Mi	mean first ionization potential (scaled on Carbon atom)
11	nAT	number of atoms
12	nSK	number of non-H atoms
13	nBT	number of bonds
14	nBO	number of non-H bonds
15	nBM	number of multiple bonds

Tableau 6 - Quelques descripteurs constitutionnels générés par le logiciel Dragon

2.2.1.2.2 Descripteurs de groupements fonctionnels

Ce sont des descripteurs moléculaires simples, définis comme étant le nombre de groupes fonctionnels spécifiques dans une molécule. Ils sont calculés sur la base de la composition moléculaire et des connectivités atomiques. (SAIHI, 2015)

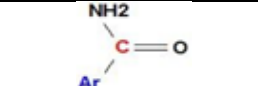
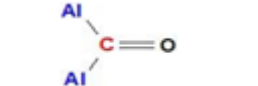
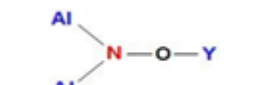
N°	Symbole	Structure	Définition
01	nArOCN		Nombre de cyanates (aromatique)
02	nArNCO		Nombre d'isocyanates (aromatique)
03	nArCONH ₂		Nombre d'amides primaires (aromatique)
04	nRCSOH		Nombre de thioacides (aliphatique)
05	nRCO		Nombre de cétones (aliphatique)
06	nRNHO		Nombre d'hydroxylamines (aliphatique) Y=H, Al ou Ar

Tableau 7 - Quelques descripteurs moléculaires (bloc de groupements fonctionnels)

2.2.1.2.3 Descripteurs topologiques

Peuvent être obtenus à partir de la structure 2D de la molécule, et donnent des informations sur sa taille, sa forme globale et ses ramifications. Ces descripteurs, ou indices topologiques, s'inspirent de la théorie des graphes appliquée à la table de connectivité qui n'est autre qu'une représentation compacte de la connectivité interatomique au sein de la molécule.

De nombreux indices quantifiant la connectivité moléculaire ont été développés en se basant sur cette approche, comme par exemple l'indice de Wiener, Schultz 1989, Balaban 1982, Randić 1975, ...etc.

2.2.1.2.4 Descripteurs topologiques géométriques

Il s'agit de descripteurs évalués à partir des positions relatives des atomes d'une molécule dans l'espace, ainsi que des rayons et masses atomiques. Ils nécessitent donc de connaître la structure 3D de la molécule. Celle-ci peut être obtenue expérimentalement bien entendu mais le plus souvent par modélisation moléculaire, empirique ou *ab initio*. Par conséquent, contrairement aux descripteurs précédents, ils nécessitent un certain temps de calcul. (FAYET, 2010)

2.2.1.2.5 Descripteurs quantiques

Ces descripteurs qui, pour être calculés de manière fiable, nécessitent des calculs plus sophistiqués sont regroupés ici sous l'étiquette de « descripteurs quantiques ». A partir de tels calculs, sont également évaluées des propriétés électroniques telles que le moment dipolaire ou la polarisabilité. Des descripteurs permettent également de quantifier les différentes interactions inter- et intramoléculaires, ces dernières ayant une grande influence sur la plupart des propriétés des systèmes moléculaires.

2.3 Techniques des régressions

Les techniques les plus courantes pour établir des modèles QSRR utilisent l'analyse de régression (régression linéaire multiple : MLR ; projection des structures latentes par les moindres carrés partiels : PLS), les réseaux neuronaux, et les méthodes de classification.

L'analyse statistique détermine et quantifie les corrélations entre les descripteurs et la variable cible. Elle indique également la contribution relative de chaque descripteur dans

l'explication globale de l'activité. Le modèle statistique est une équation donnant la valeur de la variable cible en fonction de la somme des valeurs pondérées des descripteurs.

Les principaux outils statistiques pour obtenir un modèle sont :

- La régression linéaire multivariée (Multivariate Linear Regression - MLR) ;
- La régression en composante principale (Principale Component Regression - PCR) ;
- La régression des moindres carrés partiels (Partial Least Squares - PLS) ;
- Les réseaux de neurones artificiels (Artificial Neural Network - ANN).

2.3.1 Régression linéaire multiple (MLR)

La régression linéaire multiple (MLR) est l'une des méthodes de modélisation les plus fondamentales et les plus courantes pour la régression QSAR.

MLR est favorisée par sa simplicité et la facilité d'interprétation en tant que modèle suppose une relation linéaire entre la propriété du composé, Y, et de son vecteur de caractéristiques, noté X, qui est habituellement les descripteurs moléculaires. Ainsi, avec la notion de X, la propriété d'un composé inconnu peut être prédite par le modèle ajusté. L'équation suivante représente une expression générale d'un modèle MLR :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (\text{II.1})$$

Où β_0 est la constante du modèle, $X_1; \dots; X_k$ sont des descripteurs moléculaires avec leurs coefficients correspondants $\beta_1; \dots; \beta_k$ (pour les descripteurs moléculaires 1 à k). Ces coefficients peuvent être obtenus en utilisant des estimateurs comme la méthode des moindres carrés qui minimise la somme des carrés des résidus.

La taille des coefficients peut révéler le degré d'influence des descripteurs moléculaires correspondants sur la propriété cible. De plus, un coefficient positif suggère que le descripteur moléculaire correspondant contribue positivement à la propriété cible, tandis qu'un coefficient négatif suggère une contribution négative. Cependant, ces interprétations peuvent ne pas être exactes car les descripteurs colinéaires ont le potentiel d'influencer les coefficients de sorte que des valeurs erronées peuvent être attribuées. Ainsi, les descripteurs moléculaires du modèle doivent être indépendants les uns des autres et le nombre d'instances pour la construction du modèle doit être au moins cinq fois supérieur au nombre de descripteurs utilisés. De plus, l'hypothèse d'une relation linéaire rend la MLR moins appropriée pour modéliser des problèmes

complexes comme la toxicité, où de multiples mécanismes peuvent interagir pour provoquer une réponse toxique. Néanmoins, le MLR a été utilisé pour modéliser les systèmes de toxicité.

À ce jour, MLR reste utilisé avec des améliorations ou en combinaison avec la sélection de fonctionnalités pour améliorer ses performances. Des exemples d'améliorations sont: l'utilisation de l'analyse des composants indépendants - MLR dans QPSR de la solubilité aqueuse, la régression paresseuse locale, la rétro-régression appliquée sur les points d'ébullition des nonanes, la sélection des caractéristiques d'ensemble, et d'autres méthodes de sélection de caractéristiques telles que l'algorithme génétique, la régression des crêtes, la méthode des moindres carrés partiels, la méthode de corrélation de paires, la sélection directe et la sélection du meilleur sous-ensemble dans l'application de MLR. (Liew Chin Yee, 2012)

2.3.1.1 Coefficients d'ajustement

2.3.1.1.1 Le coefficient de détermination (R^2)

Ce coefficient, dont la valeur est comprise entre 0 et 1, n'est affiché que si la constante du modèle n'est pas fixée par l'utilisateur. Sa valeur est définie par :

$$R^2 = \frac{\sum_{i=1}^n w_i (y_i - \hat{y}_i)^2}{\sum_{i=1}^n w_i (y_i - \bar{y})^2} \quad (\text{II.2})$$

Avec

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n w_i y_i \quad (\text{II.3})$$

Le R^2 s'interprète comme la proportion de la variabilité de la variable dépendante expliquée par le modèle. Plus le R^2 est proche de 1, meilleur est le modèle.

L'inconvénient du R^2 est qu'il ne prend pas en compte le nombre de variables utilisées pour ajuster le modèle.

2.3.1.1.2 Le coefficient de détermination ajusté (R^2 ajusté)

Le R^2 ajusté peut-être négatif si le R^2 est voisin de zéro. Ce coefficient n'est affiché que si la constante du modèle n'est pas fixée par l'utilisateur. Sa valeur est définie par :

$$\hat{R}^2 = 1 - (1 - R^2) \frac{W - 1}{W - p - 1} \quad (\text{II.4})$$

Le R^2 ajusté est une correction du R^2 qui permet de prendre en compte le nombre de variables utilisées dans le modèle.

- **MCE** : la moyenne des carrés des erreurs

$$MCE = \frac{1}{w - p^*} \sum_{i=1}^n w_i (y_i - \hat{y}_i)^2 \quad (\text{II.5})$$

- **La racine de la moyenne des carrés des erreurs (RMCE)** : la racine carrée de la MCE.

$$RMCE = \sqrt{MCE} = \sqrt{\frac{1}{w - p^*} \sum_{i=1}^n w_i (y_i - \hat{y}_i)^2} \quad (\text{II.6})$$

- **L'erreur moyenne de pourcentage absolu (MAPE)** : est calculée comme suit :

$$MAPE = \frac{100}{W} \sum_{i=1}^n w_i \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (\text{II.7})$$

(*) W : Somme des poids

2.3.2 Régression non linéaire

La régression non linéaire génère une équation permettant de décrire la relation non linéaire entre une variable de réponse continue et une ou plusieurs variables de prédiction, et prévoit de nouvelles observations. Utilisez la régression non linéaire plutôt que la régression sur les moindres carrés lorsque vous ne pouvez pas modéliser de manière adéquate la relation avec des paramètres linéaires. Les paramètres sont linéaires lorsque chaque terme du modèle est additif et contient seulement un paramètre qui multiplie le terme.

2.3.3 Différences entre Régression : non linéaire /linéaire multiple

La différence fondamentale entre la régression linéaire et la régression non linéaire (à laquelle ces analyses doivent leur nom) tient aux formes fonctionnelles acceptables du modèle. De façon plus spécifique, la régression linéaire requiert des paramètres linéaires, ce qui n'est pas le cas de la régression non linéaire. Utilisez la régression non linéaire plutôt que la régression

linéaire lorsque vous ne pouvez pas modéliser de manière adéquate la relation avec des paramètres linéaires.

Une fonction de régression linéaire doit être linéaire dans les paramètres, ce qui limite l'équation à une seule forme basique. Les paramètres sont linéaires lorsque chaque terme du modèle est additif et contient seulement un paramètre qui multiplie le terme :

$$\text{Réponse} = \text{constante} + \text{paramètre} * \text{prédicteur} + \dots + \text{paramètre} * \text{prédicteur}$$

Ou
$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (\text{II.8})$$

En revanche, une équation non linéaire peut prendre différentes formes. En fait, comme il existe un nombre infini de possibilités, vous devez spécifier la fonction de prévision utilisée par Minitab pour effectuer une régression non linéaire. Ces exemples illustrent la variété des possibilités (les θ représentent les paramètres) :

- $y = \theta X$ (Convexe 2, 1 paramètre, 1 prédicteur)
- $y = \theta_1 * X_1 / (\theta_2 + X_1)$ (Equation de Michaelis-Menten, 2 paramètres, 1 prédicteur)
- $y = \theta_1 - \theta_2 * (\ln(X_1 + \theta_3) - \ln(X_2))$ (Equation de Nernst, 3 paramètres, 2 prédicteurs)

Le choix de la fonction de prévision dépend souvent de vos connaissances préalables sur la forme de la courbe de la réponse ou du comportement des propriétés physiques et chimiques du système. Les formes de courbes non linéaires possibles sont notamment : concave, convexe, à croissance ou décroissance exponentielle, à courbe sigmoïde (S) et asymptotique. Vous devez spécifier la fonction qui satisfait à la fois aux exigences liées à vos connaissances préalables et aux hypothèses de régression non linéaire.

Bien que le nombre de fonctions de prévision pouvant être spécifiées offre une grande flexibilité, déterminer la fonction fournissant l'ajustement optimal pour vos données peut requérir des efforts considérables. Cela demande souvent une connaissance du domaine concerné, des analyses des erreurs et des essais ainsi que des recherches supplémentaires. Pour les équations non linéaires, déterminer l'effet de chaque prédicteur sur la réponse peut s'avérer moins intuitif que pour les équations linéaires (Minitab, 2019).

→ La régression non linéaire utilise une procédure différente de celle utilisée dans la régression linéaire pour réduire la somme des carrés de l'erreur résiduelle (SCE).

2.3.4 Logiciel XLSTAT

Nul besoin de dépenser une fortune pour réaliser des régressions linéaires multiples (RLM), il existe des logiciels libres pour cela. Nul besoin non plus de se lancer dans des programmations compliquées, l'add-in d'Excel XLSTAT (produit d'Addinsoft) est d'une utilisation particulièrement simple. Certes, le logiciel complet est payant mais vous pouvez télécharger une version d'essai et, même si vous ne l'achetez pas, vous bénéficiez de la version gratuite en fin de période d'essai.

2.3.4.1 Principe

XLSTAT permet de corriger les matrices de covariance pour les effets d'hétéroscédasticité et d'autocorrélation qui peuvent survenir notamment dans des cas où le temps intervient (séries chronologiques, données longitudinales). Différentes méthodes sont proposées dans XLSTAT pour ces corrections, comme l'estimateur suggéré par Newey et West (1987) (Addinsoft-Team, 2020).

2.3.4.2 Sélection du modèle par Logiciel XLSTAT

On utilise la fonction **Easy Fit** afin de tester et comparer différents modèles prédictifs sur un même jeu de données. En fonction du type de variable à prédire (quantitative ou qualitative) et du type de variables explicatives, différents modèles prédictifs sont proposés. La fonction Easy Predict vous permet ensuite de prédire les valeurs de nouvelles observations sur les modèles précédemment générés.

Ce tutoriel montre comment générer et sélectionner plusieurs modèles de classification (cas où la variable à prédire Y est qualitative). Cependant la fonction **Easy Fit** permet, de la même manière que dans ce tutoriel, de traiter des modèles de régression (cas où la variable à prédire Y est quantitative).

Le tableau ci-dessous montre les différents modèles prédictifs actuellement disponibles dans la fonctionnalité **Easy Fit** en fonction de la nature de la variable Y à expliquer et de la nature des variables explicatives X (Addinsoft-Team, 2020) :

Variable explicatives X	Régression : Y quantitative	Classification : Y qualitative
X quantitatives	• Régression linéaire • Random Forests	• Régression logistique • Random Forests
X qualitatives	• ANCOVA • Random Forests	
X quantitatives et X qualitatives	• ANCOVA • Random Forests	

Tableau 8 - Les différents modèles prédictif dans la fonctionnalité **Easy Fit**

2.4 Sélection du modèle

La modélisation vise à fournir un modèle qui soit non seulement ajusté aux données d'apprentissage, mais aussi capable de prédire la valeur de la sortie sur de nouveaux exemples, c'est-à-dire de généraliser. Soit R_i l'erreur commise par le modèle considéré sur l'exemple i (également appelé résidu) :

$$R_i = y(x^i) - g(x^i, \theta) \quad (\text{II.9})$$

Où $y(x^i)$ est la valeur mesurée de la grandeur à modéliser pour l'exemple i , et $g(x^i, \theta)$ est l'estimation du modèle pour ce même exemple. L'erreur du modèle sur les données d'apprentissage, E_A , peut être évaluée par l'erreur quadratique moyenne en apprentissage, appelée également EQMA :

$$E_A = \sqrt{\frac{1}{N_A} \sum_{i=1}^{N_A} (R_i)^2} \quad (\text{II.10})$$

Où N_A est le nombre d'exemples servant à établir le modèle. La qualité du modèle en apprentissage est souvent visualisée sur un diagramme de dispersion, sur lequel sont portées les valeurs de la grandeur d'intérêt estimées par le modèle, en fonctions des valeurs mesurées de cette grandeur. La qualité de la modélisation est d'autant meilleure que les points de ce graphique sont proches de la première bissectrice. L'ajustement des points à cette droite peut être évalué par **le coefficient de détermination** :

$$R^2 = \frac{\sum_{i=1}^{N_A} (y^i - \bar{y})^2}{\sum_{i=1}^{N_A} (g(x^i, \theta) - \bar{y})^2} \quad (\text{II.11})$$

Où $\bar{y}$ est la moyenne des valeurs mesurées. Ce coefficient est égal au rapport de la variance expliquée à la variance totale de la sortie. Plus il est proche de 1, plus la corrélation entre les valeurs mesurées et prédites est forte. Les valeurs intermédiaires renseignent sur le degré de dépendance linéaire entre les deux variables.

L'erreur sur la base d'apprentissage seule n'est pas un bon indicateur de la qualité du modèle ; un bon modèle doit être capable de rendre compte de la relation déterministe entre les variables et la grandeur modélisée, sans s'ajuster au bruit des données d'apprentissage. Il convient donc de trouver comment estimer cette capacité. (GOULON, 2008)

2.4.1 Méthode générale de sélection

Lors de la modélisation, nous recherchons à la fois la complexité la mieux adaptée au problème étudié, et les paramètres du modèle correspondant. Le modèle doit en effet être de complexité suffisante pour rendre compte de la relation entre les variables explicatives et la grandeur modélisée, mais il ne doit pas être trop complexe, afin de ne pas être trop sensible au bruit présent dans les données. En effet, lorsque la complexité du modèle envisagé augmente, on observe généralement une diminution du coût d'apprentissage, car le modèle s'ajuste de plus en plus précisément aux données d'apprentissage. En revanche, l'erreur de généralisation diminue pour atteindre un minimum, puis augmente avec la complexité du modèle, car celui-ci est alors sûr ajusté aux données d'apprentissage, et au bruit présent dans ces données. Ce compromis est également connu sous le nom de dilemme biais-variance : le biais caractérise l'écart entre les estimations et les mesures, et la variance reflète l'influence du choix de la base d'apprentissage sur le modèle. De plus, lorsque les modèles envisagés ne sont pas linéaires par rapport à leurs paramètres, la sélection du modèle optimal ne consiste pas uniquement à choisir la meilleure famille de fonctions, comme c'est le cas pour des modèles linéaires (modèles polynomiaux ou combinaisons linéaires de fonctions non paramétrées). La sélection s'effectue donc souvent en plusieurs étapes :

- Pour les modèles non linéaires en leurs paramètres, la première sélection s'effectue au sein d'une famille de fonctions donnée. En effet, la fonction de coût n'étant pas quadratique en les paramètres du modèle, elle ne possède pas un minimum unique. L'optimisation de cette fonction peut donc conduire à des minima différents, donc à différents modèles de même complexité. Il est donc nécessaire de réaliser plusieurs apprentissages, pour une complexité donnée, et de choisir celui qui est susceptible de posséder les meilleures propriétés de généralisation.
- Il faut ensuite sélectionner, parmi les meilleurs modèles de complexités différentes, celui qui présente les meilleures capacités de généralisation (SEROUTI , 2012).

2.4.2 Sélection du modèle par estimation de l'erreur de généralisation

L'erreur de généralisation ne peut pas être évaluée exactement ; la base de données disponible est en effet de taille limitée, et la distribution de probabilité des données, dont dépend cette erreur, est généralement inconnue. Il est donc nécessaire de trouver une approximation de cette erreur de généralisation. Nous allons ainsi présenter les méthodes les plus utilisées pour effectuer ces sélections.

La méthode la plus commune, dite méthode de validation simple ou hold-out, consiste à construire, à partir de l'ensemble des données, deux ensembles : les paramètres du modèle sont ajustés sur **la base d'apprentissage**, et l'erreur de généralisation est évaluée sur **la base de validation**.

L'erreur en validation est définie de façon analogue au coût d'apprentissage :

$$E_V = \sqrt{\frac{1}{N_V} \sum_{i=1}^{N_V} (R_i)^2} \quad (\text{II.12})$$

Où N_V est le nombre d'exemples dans la base de validation.

Pour des modèles de complexité donnée, la méthode de sélection consiste à effectuer plusieurs apprentissages à partir de différentes initialisations des paramètres, lorsque l'on a affaire à un modèle non linéaire en ses paramètres. Il faut alors choisir, parmi les modèles obtenus, celui qui fournit la plus faible erreur en validation. Il s'agit d'une méthode très rapide,

mais d'efficacité limitée, surtout lorsque le nombre d'exemples disponible est faible. En effet, si le nombre d'exemples est petit, la variance de l'estimation de l'erreur de généralisation est grande ; le coût de validation dépend alors fortement du choix de la base de validation. (GOULON, 2008)

2.4.3 La validation croisée

Afin de déterminer la stabilité prédictive d'un modèle et de tester l'influence de chaque échantillon (composé) sur le modèle final, des procédures de validation croisée (en anglais : cross-validation) sont souvent utilisées.

Généralement, Ces techniques de validation permettent l'évaluation de la robustesse du modèle, autrement dit la stabilité des paramètres du modèle RQSA/RQSP vis-à-vis des molécules du jeu d'entraînement. Cela dit, qu'elles ne permettent en aucun cas de démontrer le pouvoir prédictif des modèles

Le principe de ces méthodes consiste à extraire un certain nombre de molécules du jeu d'apprentissage et à construire un nouveau modèle avec les molécules restantes à l'aide des descripteurs choisis (seules les constantes de la régression changent). Ce nouveau modèle est alors utilisé pour la phase de prédiction sur les molécules retirées. Ce processus est ensuite répété pour retirer et prédire les valeurs de toutes les molécules du jeu d'entraînement (Figure 8) (CHTITA, 2017).

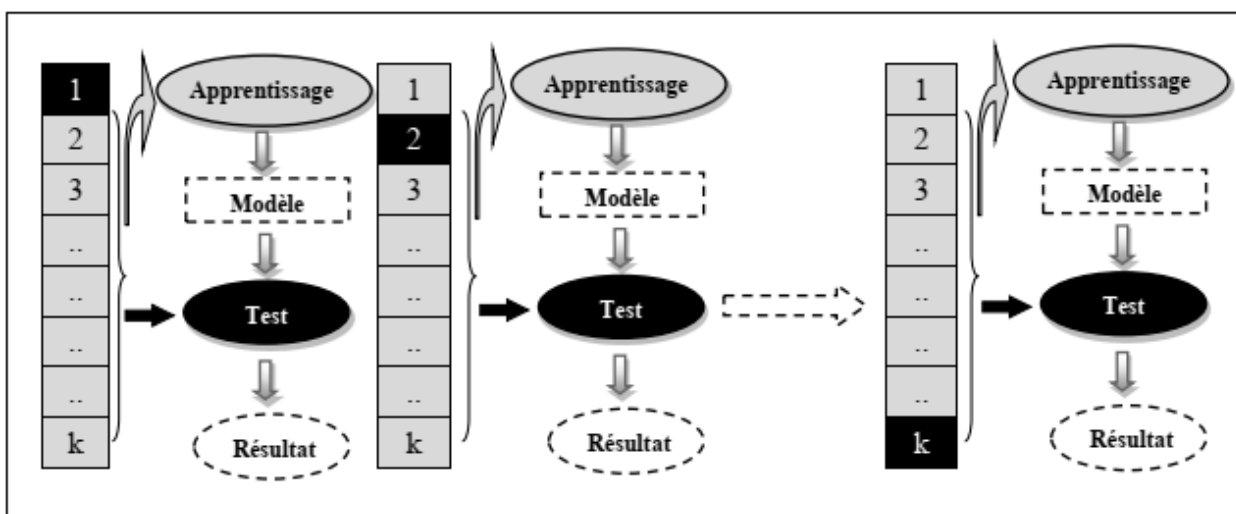


Figure 6 - Procédure de la validation croisée

Chapitre 3

Modélisation de la propriété $\Delta_c H^\circ$

Les relations entre les structures moléculaires et leurs propriétés sont généralement établies à l'aide de méthodes de modélisation par apprentissage statistique. Les techniques usuelles reposent sur la caractérisation des molécules par un ensemble de nombres réels mesurés ou calculés à partir des structures moléculaires "descripteurs".

L'enthalpie standard de combustion des composés de tous les groupes chimiques a été prédite en utilisant une régression linéaire multivariée basée sur un modèle mathématique. Le modèle linéaire multivarié à quelque descripteur obtenu par MLR a un coefficient de corrélation. Tous les descripteurs moléculaires qui sont entrés dans ce modèle sont calculés à partir de la structure chimique de n'importe quelle molécule. En conséquence, l'application de ce modèle pour n'importe quel composé est facile et précise (GUENDOUZI, 2015).

Nous allons présenter les descripteurs moléculaires les plus courants, en commençant par les descripteurs les plus simples, qui nécessitent peu de connaissances sur la structure moléculaire, mais véhiculent peu d'informations. Nous verrons ensuite comment les progrès de la modélisation moléculaire ont permis d'accéder à la structure 3D de la molécule, et de calculer des descripteurs à partir de cette structure (SEROUTI, 2012).

3.1 Etablissement de la base de données et détermination de descripteurs moléculaires

Le principe des méthodes QSPR est, comme leur nom l'indique, de mettre en place une relation mathématique reliant de manière quantitative des propriétés moléculaires aussi bien électroniques que géométriques, appelées descripteurs, avec une observable macroscopique (activité biologique, toxicité, propriété physico-chimique, etc.), pour une série de composés chimiques similaires à l'aide de méthodes d'analyses de données. Aussi la forme générale d'un tel modèle est la suivante :

$$\text{Propriété} = f(\text{Descripteurs}) \quad (\text{III.1})$$

En pratique, le développement proprement dit d'un modèle débute par la collecte de données expérimentales fiables et en nombre le plus important possible. Il s'agit ensuite de développer une série de descripteurs qui caractérisent les structures moléculaires électroniques et géométriques des composés de la base de données en vue de les relier à la propriété expérimentale étudiée. Des outils d'analyse de données sont alors employés pour aider à choisir les descripteurs adéquats et mettre en place le modèle proprement dit. Une fois développée, le modèle doit alors être validé en termes de corrélation (sur le jeu de données d'entraînement). Sa robustesse, c'est-à-dire l'influence des composés du jeu d'entraînement sur le modèle, est estimée par des méthodes de validation interne. Pour estimer son pouvoir prédictif, il est ensuite nécessaire de disposer de données expérimentales supplémentaires afin de déterminer la capacité du modèle à prédire ces valeurs. Enfin, pour tout modèle, il est important de savoir pour quel type de molécules il est utilisable ou non, c'est-à-dire connaître son domaine d'applicabilité



Figure 7 - Modèle de l'étude de relation structure activité/propriété

3.2 Méthodologie de préparation de base des données

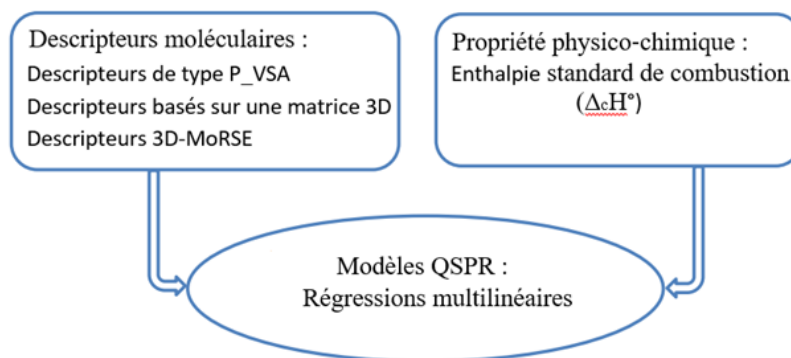


Figure 8 - Schéma du cadre hiérarchique pour Modèle QSPR.

→ Modèle QSPR établit des relations mathématiques entre des caractéristiques de la structure moléculaire et les propriétés observées

3.2.1 Conception de la base des données

Fondamentaux pour la conception de bases de données de pointe est le concept de la « base de données relationnelle » où les objets et les différents types d'informations sont reliés et rendus accessibles à l'utilisateur dans le contexte. Par exemple, en chimie, une base de données relationnelle simple peut stocker des structures composées, de disponibilité et de données analytiques, et une version plus avancée de cette base de données pourrait également inclure dosage biologique ou des données d'activité de ces composés. L'utilisation de bases de données chimiques relationnelles est un des aspects les plus importants impliqués dans de nombreuses applications chimio-informatique. Par ailleurs, bases de données relationnelles offrent une plateforme pratique pour les interactions entre les scientifiques informatiques et expérimentales (SEROUTI, 2012).

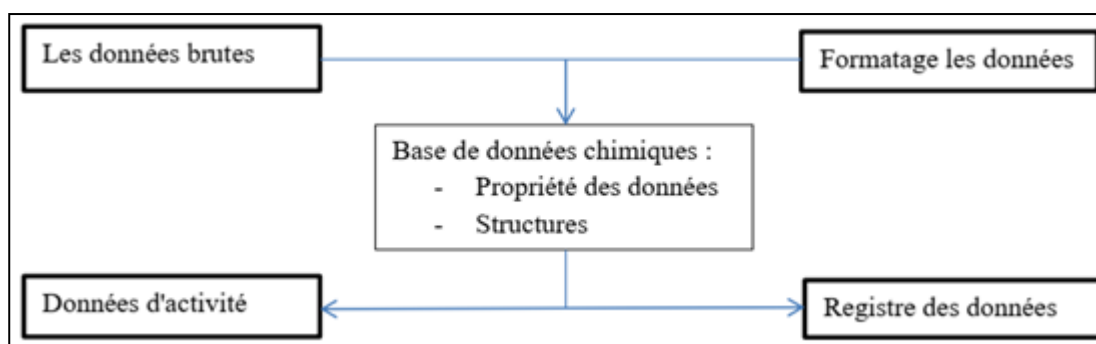


Figure 9 - Représentation schématique de l'architecture de la base de données relationnelles

Le concept central du schéma relationnel de la base de données est **la relation**. C'est le mathématicien et théoricien britannique, Edgar F. Codd qui inventa les modèles relationnels. Selon lui, une relation représente un ensemble d'entités ayant les mêmes propriétés. Chaque relation est constituée d'une série d'enregistrements de données (appelés tuples) dont les valeurs sont affectées à certains attributs.

Les attributs qu'une relation contient et le type de données auquel correspondent les valeurs affectées aux attributs sont définis à l'aide d'un **schéma de relations** selon la syntaxe suivante :

$$R = (A1 : Type1, A2 : Type2, \dots, An : Typen)$$

Le schéma de relation (R) comprend les attributs A1 à An. Un type de données (Type1, Type2,etc.) est affecté à chaque attribut. Ceci peut être illustré par un exemple concret.

Afin de pouvoir représenter clairement l'affectation des valeurs individuelles d'un tuple aux attributs définis dans le schéma relationnel, un concept classique d'organisation de l'information est utilisé dans le modèle de base de données relationnelle : **la table**. Une base de données relationnelle n'est donc rien de plus qu'une collection de tables liées entre elles.

Les tables sont des schémas d'ordre composés de lignes horizontales et de colonnes verticales qui permettent de collecter et d'afficher les informations sous une forme ordonnée. Chaque ligne d'une table de base de données correspond à un tuple. Les valeurs des tuples listés sont affectées aux attributs définis dans le schéma de relation via les colonnes de la table (IONOS, 2019).

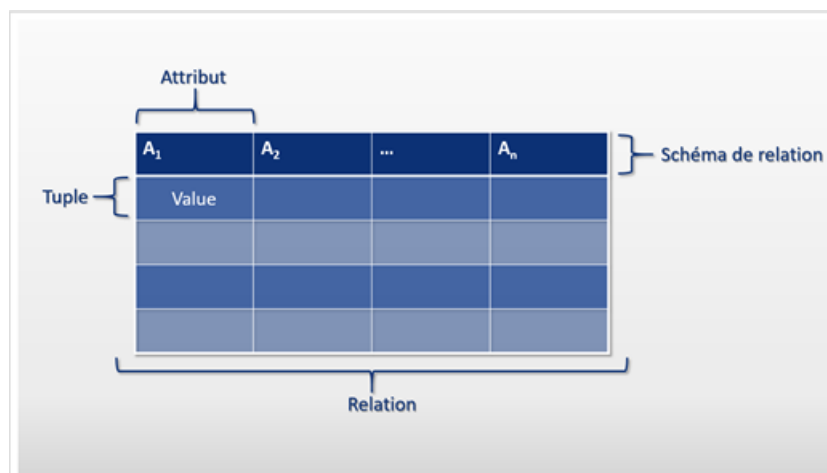


Figure 10 - Schéma d'une relation selon le modèle de bases de données relationnelles

(*) Dans le modèle de bases de données relationnelles, un ensemble de tuples ayant les mêmes attributs est appelé « relation ».

3.2.2 Préparation de base des données

Toutes les valeurs des propriétés expérimentales ont été obtenues de la Banque de données du labo de Génie de procédés. Le nombre des composés est atteint 861 constituants, présent de nombreux familles chimiques à savoir (acides, esters, composés oxygénés polyfonctionnels, alcools).

On recherche des valeurs d'enthalpie standard de Combustion à partir des CAS et les structures moléculaires on registre sous forme de fichier "2D Mol " (voir Annexe A).

Toutes les valeurs rapportées ont été dérivées de mesures expérimentales publiées, et la plupart des données ont été sélectionnées à partir de la compilation suivante :

- D. D. Wagman, W. H. Evans, I. Halow, V. B. Parker, S. M. Bailey, and R. H. Schumm, "Selected values of chemical thermodynamic properties: Tables for the alkaline earth elements “. (1965)
- J. B. Pedley, R. D. Naylor, S. P. Kirby (auth.), "Thermochemical Data of Organic Compounds-Springer Netherlands". (1986)
- **NIST** (National Institute of Standards and Technology) Chemistry WebBook [<http://webbook.nist.gov/chemistry/>].

(*) L'ensemble de ces valeurs est présenté dans le tableau (Annexe). Celui-ci contient des données sur 861 molécules. Les valeurs expérimentales d'enthalpies de combustion en phase solide dans le tableau, sont des valeurs moyennes.

3.2.2.1 Récupération et stabilisation les molécules de fichier Hin

3.2.2.1.1 Stabilisation structure des molécules (minimisation de l'énergie)

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

recherche une structure dans laquelle l'énergie ne diminue pas avec les changements dans la géométrie infinitésimale.

À ce stade, la conformation est relativement stable, et est considéré comme un minimum. Si l'énergie diminue par de petits changements dans une ou plusieurs dimensions, mais pas dans toutes les dimensions, il est un point selle (Point B). Un système moléculaire peut avoir plusieurs minima. L'un avec l'énergie la plus basse est appelé le minimum global et le reste sont visées à minima que local.

3.2.2.1.2 Mécanique moléculaire

L'expression « Mécanique Moléculaire » désigne actuellement une méthode de calcul qui permet, à priori, d'obtenir des résultats de géométries et d'énergies moléculaires en se basant sur la mécanique classique. La MM est apparue en 1930, mais s'est développée à partir des années soixante, quand les ordinateurs furent plus accessibles et plus performants. La MM est basée sur l'approximation de Born-Oppenheimer selon laquelle les électrons sont beaucoup plus rapides que les noyaux (MOSTEFAOUI, 2011).

La mécanique moléculaire calcule l'état le plus stable au zéro absolu. Ce calcul de minimisation de l'énergie potentielle ne fait pas intervenir l'énergie cinétique, donc la température. Des méthodes de gradient sont utilisées pour minimiser cette énergie.

La mécanique moléculaire ressemble aux modèles de type « tiges et boules ». Un calcul MM aboutit à une disposition des noyaux telle que la somme de toutes les contributions énergétiques est minimisée ; ses résultats concernant surtout la géométrie et l'énergie du système. L'idée directrice de cette méthode est d'établir, par le choix des fonctions énergétiques et des paramètres qu'elles contiennent, un modèle mathématique, le « CHAMP DE FORCE », qui représente aussi bien que possible les variations de l'énergie potentielle avec la géométrie moléculaire.

La mécanique moléculaire a pour but de calculer l'énergie potentielle d'une molécule (ou d'un système de molécules) en fonction des coordonnées des atomes.

Cette discussion se concentre sur les composants individuels d'un champ de force mécanique moléculaire typique. Il illustre les fonctions mathématiques utilisées, pourquoi ces fonctions sont choisies, et les circonstances dans lesquelles les fonctions deviennent des approximations de pauvres. Théories et méthodes, comprend des détails sur la mise en œuvre de

la MM+, AMBER, BIO+, et des champs de force modèles OPLS (Optimized Potentials for Liquid Simulations) dans HyperChem (SEROUTI , 2012).

À partir d'une structure dessinée en deux dimensions (HyperChem). La transformation en structure en trois dimensions est ensuite obtenue par ajustement global des paramètres. La stéréochimie aléatoire peut alors être modifiée.

HyperChem comprend quatre haut-mécanique moléculaire des champs de force : de nouvelles implémentations de techniques élaborées et publiées par des groupes de recherche respectés, et présente dans figure 4 suivante :

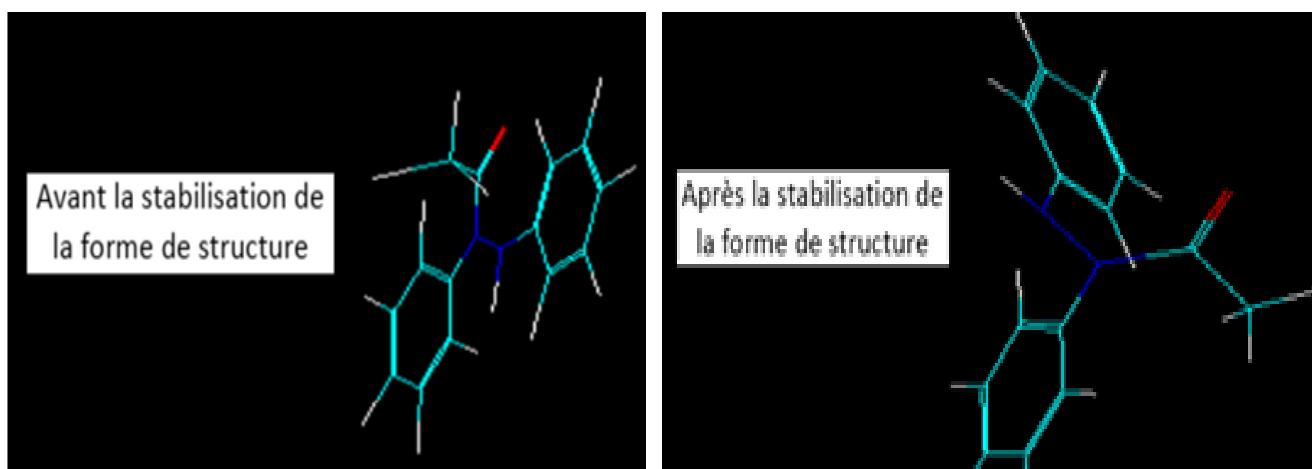


Figure 11 - Application de l'option "Field Force" comme méthode de calcul MM

3.2.2.2 Calcul des descripteurs moléculaires

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

- Indiquer la description de la configuration de la molécule à étudier.
- Décrire tous les paramètres descriptifs de la molécule

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

La représentation numérique de la structure chimique (descripteurs moléculaires) est une étape importante de l'investigation QSAR/QSPR. Les performances du modèle élaboré et la précision des résultats sont étroitement liées au mode de détermination de ces descripteurs. Les descripteurs moléculaires ont été calculés par le logiciel Dragon. Après le calcul de descripteurs moléculaires pour toutes les 861 structures chimiques ;

- Nous devons rejeter les descripteurs non informatifs de la sortie de Dragon.
- Les descripteurs avec un écart type inférieur à 0,0001, ont été rejetés parce que ces descripteurs ont été à peu près constants.
- Les descripteurs avec une seule valeur différente de celles qui restent sont rejetés.
- La corrélation de chaque paire de deux descripteurs a été vérifiée et l'un des deux descripteurs avec une corrélation d'un coefficient égal (comme valeur de seuil) a été exclu.
- Pour chaque paire de descripteurs de corrélation, l'une montrant la corrélation la plus élevée avec les autres descripteurs rejetés de la piscine de descripteurs.
- Enfin, la piscine de descripteurs moléculaires a été réduite par la suppression de descripteurs qui n'ont pas pu être calculé pour chaque structure dans notre jeu de données.

En conséquence, à partir du 1664 calculée descripteurs moléculaires par le DRAGON, dans la première étape, seulement 657 descripteurs moléculaires est resté dans la piscine de descripteurs moléculaires.

Les blocs des descripteurs moléculaires choisis et les résultats de calcul des descripteurs sont résumés dans le tableau suivant :

Les blocs des descripteurs	Nombre de descripteurs dans le bloc	Nombre de descripteurs après la réduction
Descripteurs d'anneau	39	39
Descripteurs de type P_VSA	07	07

Tableau 9 - Montre les résultats de calcul des descripteurs

(*) Nous notons qu'il n'y a pas de changement sur le nombre de descripteurs sélectionnés pour notre étude (pour nous ces descripteurs sont idéaux et adaptés à la modélisation).

3.2.2.2.1 Le Logiciel DRAGON

Le logiciel DRAGON a été conçu pour fournir à l'utilisateur une variété de descripteurs moléculaires dérivés de différentes représentations moléculaires, permettant à l'utilisateur de choisir les descripteurs moléculaires qui conviennent le mieux à sa recherche spécifique.

DRAGON a été conçu comme un logiciel convivial qui effectue des calculs de descripteurs selon une séquence logique simple : 1. Chargement des fichiers moléculaires ; 2. Sélection des descripteurs ; 3. Calcul des descripteurs ; 4. Sauvegarde des descripteurs calculé. Les calculs du logiciel de DRAGON les descripteurs moléculaires est donne 20 blocs, et nous on présente les descripteurs des deux blocs (les descripteurs en utilisent dans cette étude) : Descripteurs d'anneau, Descripteurs de type P_VSA et dans le Tableau 8 suivant :

Descripteurs d'anneau				Descripteurs de type P_VSA	
N°	Desc	N°	Desc	N°	Desc
1	Qindex	21	PW2	40	LP1
2	HNar	22	PW3	41	Eig1v
3	Xt	23	PW4	42	SEigZ
4	Ram	24	PW5	43	SEigv
5	VDA	25	PJ12	44	VRA2
6	MSD	26	DECC	45	VEZ2
7	SPI	27	Lop	46	VRe1
8	WA	28	D/Dr03	/	/
9	TI1	29	D/Dr04	/	/
10	TI2	30	D/Dr05	/	/
11	STN	31	D/Dr06	/	/
12	Rww	32	D/Dr07	/	/
13	Wap	33	D/Dr08	/	/
14	J	34	D/Dr09	/	/
15	Jhetv	35	D/Dr10	/	/
16	MAXDN	36	D/Dr11	/	/
17	MAXDP	37	D/Dr12	/	/
18	DELS	38	T(N..N)	/	/
19	S3K	39	T(O..O)	/	/
20	BLI	/	/	/	/

Tableau 10 - Les descripteurs : Descripteurs d'anneau et Descripteurs de type P_VSA

3.3 Méthodes d'analyse de données

En fait, de nombreux outils existent et il s'agit de trouver le moyen le plus adapté pour obtenir un modèle fiable à partir des données disponibles. Selon les cas, plusieurs approches sont envisageables, il faut alors choisir celle permettant de caractériser au mieux le système.

Les différentes méthodes présentées dans la suite sont celles employées au cours de l'étude, pour développer des modèles (plus ou moins de paramétrés, linéaires ou non linéaires, interprétables ou non), choisir les paramètres les plus pertinents, valider ces modèles (en interne ou en externe) et déterminer leurs domaines d'applicabilité.

Dans cette étape, notre problème est de trouver le meilleur modèle linéaire multivarié qui a le plus de précision ainsi que le nombre minimum de descripteurs moléculaires possibles.

Pour obtenir le meilleur modèle, il faut considérer l'effet de l'augmentation du nombre de descripteurs moléculaires sur l'augmentation de la valeur du coefficient de validation croisée.

Lorsque le coefficient de validation croisée était assez constant avec l'augmentation du nombre de descripteurs moléculaires, nous devons arrêter notre recherche, et le meilleur résultat a été obtenu.

3.3.1 Régression multi-linéaire

La régression multi-linéaire (MLR, pour Multiple Linear Regression) est la méthode la plus simple et la plus communément employée pour le développement de modèles prédictifs. Elle repose sur l'hypothèse qu'il existe une relation linéaire entre une variable dépendante y (ici, la propriété) et une série de n variables indépendantes x_i (ici, les descripteurs). L'objectif est d'obtenir une équation de la forme suivante :

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n \quad (\text{III.2})$$

Où a_i sont les coefficients de la régression.

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 l'enthalpie de combustion standard. Dont les valeurs expérimentales ont été prélevées dans la littérature.

Cette analyse est dévisée à l'analyse de régression simple (training set 80%) et la validation du modèle par 165 molécules qui présentent 20% du nombre globale. Nous avons fait les calculs avec XLSTAT software et les résultats pour ces descripteurs sont rapportés avec

référence aux valeurs expérimentales. Avec la probabilité d'entré est : $\alpha = 0.05$.

La forme de l'équation est donnée qui les calculs de XLSTAT s'écrivent comme suit :

$$P(x) = A + \sum_{i=0}^n w_i x_i \quad (\text{III.3})$$

Ou P: quelque soit propriété ($\Delta_c H^\circ$: Enthalpie de combustion) ;

A : constante ;

w : les poids (coefficients) de les variables x, $i=0, 1, 2, \dots, n$;

x : les variables de la propriété (descripteurs moléculaires), $i=0, 1, 2, \dots, n$.

3.3.1.1 La régression pour choix les meilleurs blocs

En premier régression dans ce travail on calcule par RML à chaque bloc de descripteur moléculaire sont seul et présente les résultats dans le tableau suivant :

N°	Descripteurs moléculaire	Nbr Desc	R ²	R ² ajusté	RMCE	MAPE
01	Descripteurs constitutionnels	17	0.989	0.988	725.895	152.898
02	Descripteurs d'anneau	39	0.997	0.997	384,64	7,386
03	Indices topologiques	12	0.895	0.893	2174.544	174.092
04	Marche et Chemin comptent	13	0.975	0.975	1050.925	160.375
05	Indices de connectivité	19	0.972	0.971	1121.047	164.862
06	Indices d'information	21	0.952	0.951	1487.052	149.077
07	Descripteurs basés sur une matrice 2D	30	0.957	0.955	1393.392	150.089
08	Auto corrélations 2D	13	0.963	0.962	1293.439	153.969
09	Valeurs propres de charge	15	0.745	0.738	3481.265	181.521
10	Descripteurs de type P_VSA	07	0.991	0.991	620.197	10.166
11	Indices ETA	04	0.848	0.847	2614.085	145.886
12	Indices de contiguïté des bords	21	0.968	0.967	1210.283	179.227
13	Descripteurs géométriques	42	0.985	0.984	855.745	13.479
14	Descripteurs basés sur une matrice 3D	85	0.998	0.998	312.016	5.277
15	Autocorrélations 3D	24	0.956	0.954	1419.913	170.157
16	Descripteurs RDF	100	0.988	0.985	808.135	146.565
17	Descripteurs 3D-MoRSE	78	0.997	0.996	408.918	6.706
18	Descripteurs WHIM	37	0.987	0.986	781.642	153.662
19	Descripteurs GETAWAY	07	0.898	0.891	2778.832	31.003
20	Profils moléculaires aléatoires	20	0.986	0.986	775.288	155.528

Tableau 11 - Les coefficients d'ajustement pour chaque bloc de descripteur

(*) Deux Blocs N° : 02 et 10, qui a donné de bons résultats, les deux blocs sera adapté pour la modélisation par RLM.

On observe dans le tableau présenter, le dix-septième modèle est bien précis par rapport les autres modèles (R^2 élevé et moindre Nbre des descripteur), et on a fait la même chose pour les autres modèles, et chaque fois en prend le modèle qui nous donne une constante de corrélation le plus élevé, et une valeur d'erreur plus faible.

Cette procédure (sélection de variable) constitue une partie cruciale lors du développement du modèle de QSAR. Elle sert à réduire le nombre de descripteurs en deux étapes successives : Sélection objective et subjective.

La sélection objective utilise seulement les variables indépendantes (les descripteurs) pour filtrer tous les descripteurs non utiles. Ce procédé implique :

- Tous les descripteurs présentant les mêmes valeurs pour toutes les molécules ont été éliminés.
- Ensuite, nous ordonnons par voie décroissante les descripteurs en fonction e leurs variances.

Entre temps, on calcule la matrice de corrélation de tout l'ensemble de descripteurs. A partir de cette matrice, nous éliminons un des deux descripteurs qui ont prouvé un coefficient de corrélation audessus de 0,9 ($R > 0,9$) et présentant la petite variance.

Après élimination par la sélection objective, le nombre de descripteurs est, encore une fois, réduits par **la sélection subjective** en cherchant un sous ensemble optimal riche en informations.

3.3.2 Régression non linéaire

La régression non linéaire permet de modéliser des phénomènes complexes n'entrant pas dans le cadre du modèle linéaire. XLSTAT propose des fonctions préprogrammées parmi lesquelles l'utilisateur pourra éventuellement trouver le modèle décrivant le phénomène à modéliser.

3.3.2.1 Méthode d'analyse

L'analyse des données par XLSTAT doit être réalisée à partir de 2 ensembles distincts servant au cours de 2 phases, une phase d'apprentissage (3/4 des données) et une phase de validation (1/4 des données). Les prélèvements des données sont réalisés par tirage aléatoire, sans

remise, la répartition statistique devant être la plus proche possible entre les 2 sous bases constituées lors des 2 phases.

Le logiciel XLSTAT analyse les données d'entrées suivant ces deux phases :

3.3.2.1.1 La phase d'apprentissage

C'est la phase pendant laquelle les paramètres du modèle (poids w_i du réseau) sont ajustés de façon à ce que Le logiciel XLSTAT matérialise au mieux la relation Entrées/Sortie telle qu'elle est représentée par les données. Etant donné que 2 apprentissages successifs ne conduisent pas au même résultat (poids initiaux différents), il est habituel de lancer successivement plusieurs apprentissages. Dans le cadre de notre étude, nous lancerons 10 apprentissages successifs.

3.3.2.1.2 La phase de validation

Il convient de prendre des dispositions qui permettront d'estimer la capacité à valider convenablement, en phase d'utilisation, sur des données nouvelles qui n'étaient pas disponibles lors de l'apprentissage.

Afin de déterminer la qualité d'un modèle, différents indicateurs statistiques sont employés (MCE, RMCE, coefficient de Durbin-Watson, Critère Bayésien de Schwarz...). Le plus répandu d'entre eux est le coefficient de corrélation R^2 qui évalue la part de la variance expliquée par le modèle.

3.3.3 Le résidu moyen et la moyenne quadratique

Une fois familiarisé avec le fonctionnement et l'utilisation du Logiciel XLSTAT, nous allons exercer celui-ci sur les données expérimentales de l'enthalpie de formation standard en phase solide de molécules organiques (matériaux énergétiques).

Pour comparer entre les résultats Les résidus moyens et les moyennes quadratiques obtenus sont présentés dans le résultat ultérieur. Rappelons que :

Le résidu moyen s'agit donc de la moyenne de la valeur absolue des résidus valent :

$$R_{moyen} = \frac{\sum_i |x_i(litt) - x_i(calc)|}{n_i} \quad (III.4)$$

La moyenne quadratique est quant à elle obtenue de la manière suivante :

$$M_{quadratique} = \sqrt{\frac{\sum_i (x_i(litt) - x_i(calc))^2}{n_i}} \quad (III.5)$$

Elle permet d'amplifier l'influence des écarts absolus les plus élevés par rapport à l'écart moyen calculé

Chapitre 4

Résultats et discussion

Introduction

Dans ce chapitre, nous avons construits une base de données qui a été arrange à partir des variétés de familles chimiques, suivi d'étude de 20 blocs de descripteurs et du choix des blocs des descripteurs moléculaires, et l'étude de chaque bloc comprend deux parties, la première concerne la présentation des résultats de calcul et graphiques, et la deuxième pour l'analyse de ces résultats. Cette analyse se devise en analyse statistique simple, analyse des résidus et validation du modèle par 165 molécules qui présente 20% du nombre globale. Nous avons réalisé les calculs à l'aide du logiciel de modélisation moléculaires XLSTAT 2020. Les modèles ont été calculés à l'aide de la méthode de sélection pas à pas des variables de la régression linéaire multiple.

Aussi, dans cette partie, et afin d'illustrer la fiabilité de la méthode développée, nous allons représenter quelques sommaires statistiques des résultats obtenues à partir la régression pour la propriété étudiée.

4.1 Traitement le base de donne

Après la première étape de calcul, dans laquelle le meilleur groupes a été choisi, la banque de données a également été traitée, et nous avons remarqué à travers le graphe composite 8,13-Diethyl-3,7,12,17,20-pentamethyl-21H,23H-porphine-2-propanoic acide méthyl avec une valeur de l'enthalpie standard de combustion -18,035 kJ/mol loin de l'axe $y = x$ (Figure 14), et après s'être assuré de l'exactitude de la valeur -18,035 kJ/mol référencée du NIST, que nous avons trouvée complètement différente de Il est indiqué dans la référence Pedley et al. 1986, qui équivaut à -18034.7 kJ/mol.

Lors de la modification de la valeur du composé mentionné ci-dessus, nous avons remarqué une diminution significative des valeurs d'erreur.

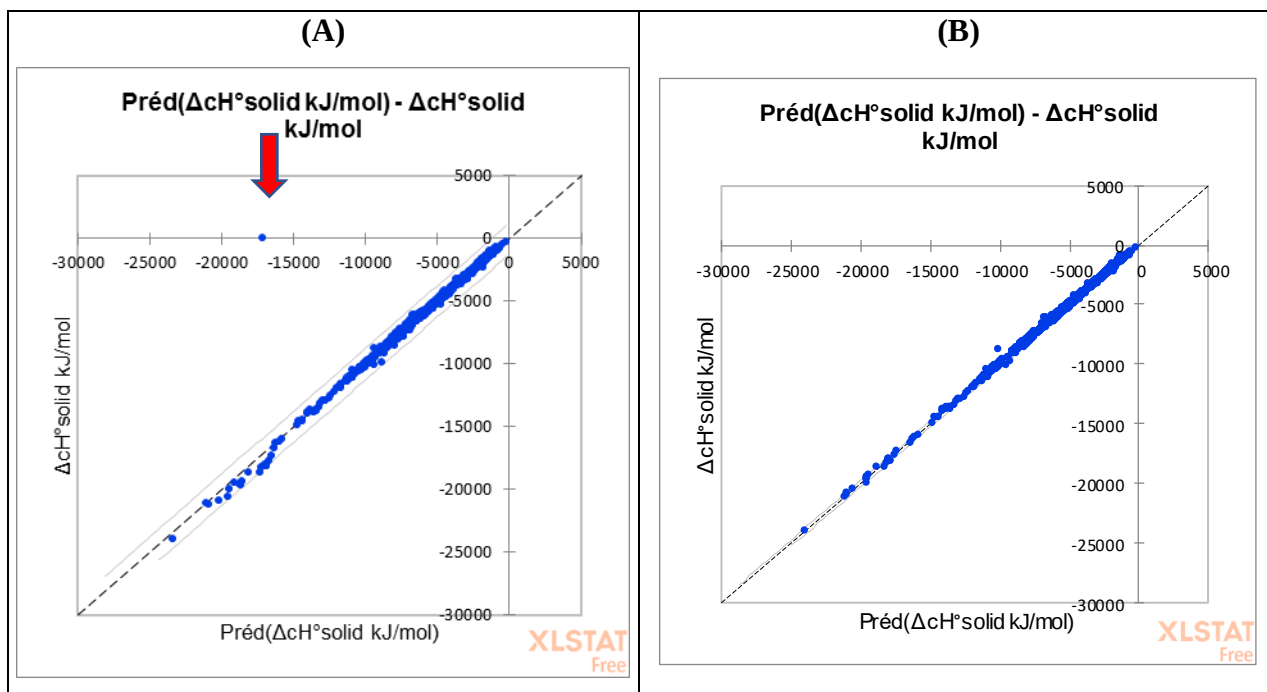


Figure 12 - Correction valeur de $\Delta_c H^\circ$ (Composant N° CAS : 33719-66-3-2d)

(*) **Figure 14 :**

- A)** Le graphe présente la composition 8,13-Diethyl-3,7,12,17,20-pentamethyl-21H,23H-porphine-2-propanoic acid methyl avec la valeur de $\Delta_c H^\circ$ est égal -18.035 kJ/mol.
- B)** Le graphe présente la composition 8,13-Diethyl-3,7,12,17,20-pentamethyl-21H,23H-porphine-2-propanoic acid methyl avec la valeur de $\Delta_c H^\circ$ est égal -18034.7 kJ/mol.

4.2 Détail les résultats à base de donne

Dans cette partie, et afin d'illustrer la fiabilité de la méthode développée, nous allons représenter quelques sommaires statistiques des résultats obtenues à partir la régression pour la propriété étudiée.

4.2.1 Résultat de régression linéaire multiple

4.2.1.1 Coefficients d'ajustement

4.2.1.1.1 Descripteurs de type P_VSA

L'analyse de la régression de la propriété étudiée nous a donné l'unique modèle statistique à sept descripteurs. Le modèle QSPR sélectionné (Bloc N°10) est représenté par l'équation :

$$\Delta_c H_{solid}^{\circ} \text{ (kJ / mol)} = -2907.00314083498 * LP1 - 28.5817841381056 * Eig1v + 5345.49305902646 * SEigZ + 607.015724749816 * SEigv + 4205.81821895717 * VRA2 + 1128.76577811646 * VEZ2 - 300.112741034339 * VRe1 \quad (IV.1)$$

$$\Delta_c H_{Solid}^{\circ} = b - a_1 \times LP1 - a_2 \times Eig1v + a_3 \times SEigZ + a_4 \times SEigv + a_5 \times VRA2 + a_6 \times VEZ2 - a_7 \times VRe1 \quad (IV.2)$$

Ce tableau affichés les statistiques relatives à l'ajustement du modèle de régression :

Tableau 12 - Coefficients d'ajustement pour RL du descripteur de type P_VSA

Descripteurs Moléculaire	N° Desc	R ²		R ² Ajusté		RMCE		MAPE	
		App	Val	App	Val	App	Val	App	Val
Descripteurs d'anneau	39	0.997	0.982	0.997	/	384.64	501.348	7.386	8.043
Descripteurs de type P_VSA	07	0.991	0.916	0.991	/	620.197	1393.537	10.166	483.367

Après les calculs on a obtenu plusieurs modèles, le choix de meilleur modèle se base sur les constantes de corrélation les plus élevés et l'erreur (On constate que tous les **R²** sont supérieures à la valeur 0.99)

Les résidus moyens et la moyennes quadratiques pour les descripteurs de type P_VSA obtenus sur la représentation de l'enthalpie standard de combustion par la régression linéaire dans le tableau suivant :

Tableau 13 - Résidus moyens et moyennes quadratiques obtenus (Desc de type P_VSA)

Descripteurs moléculaire	Résidu moyen	Moyenne quadratique
Descripteurs d'anneau	10.719	308.071
Descripteurs de type P_VSA	14.055	349,966

4.2.1.1.2 Descripteurs d'anneau

Le modèle QSPR sélectionné (Bloc N°02) est représenté par l'équation :

$$\begin{aligned}
\Delta C H^{\circ} \text{solid } kJ / mol = & -68,8698971623475 * Q_{index} + 664,313894976995 * H_{Nar} - 230,625251672444 * X_t \\
& -341,446469757628 * Ram - 17,2615326134865 * VDA + 2697,28048121994 * MSD + 60,5843207475749 * SPI \\
& -1436,17466095941 * WA - 4,49165664949838 * TI1 + 208,579292265838 * TI2 - 320,906422974245 * STN \\
& -66,7660565166725 * Rww - 5,06265504952271E - 03 * Wap + 786,156590101644 * J - 710,066013494212 * Jhetv \\
& +94,1353081368827 * MAXDN - 22,4740640405812 * MAXDP + 85,5511955215963 * DELS - 32,3814525185314 \\
& * S3K - 4901,76648378949 * BLI + 7359,0624430195 * PW2 - 3458,79396725955 * PW3 + 902,990009346316 * \\
& PW4 + 2809,75337299349 * PW5 - 493,897513508314 * PJI2 + 1002,83086106684 * DECC - 116,189519316619 * \\
& Lop + 45,5945891563099 * D / Dr03 + 36,7310183311231 * D / Dr04 + 12,241219611963 * D / Dr05 + 2,28839203012489 * \\
& D / Dr06 + 4,52973508971946 * D / Dr07 + 5,85326471423665 * D / Dr08 + 10,3980649910717 * D / Dr09 + 4,74504175272597 \\
& * D / Dr10 - 6,79786115251199 * D / Dr11 - 0,246431604593439 * D / Dr12 + 5,19095092917012 * T (N..N) + 1,29763602238378E \\
& -03 * T (O..O)
\end{aligned}
\tag{IV.3}$$

Ce tableau affichés les résultats statistiques relatives à l'ajustement du modèle de régression :

Tableau 14 - Coefficients d'ajustement pour RN du descripteur Descripteurs d'anneau

Descripteurs Moléculaire	N° Desc	R ²		R ² Ajusté		RMCE		MAPE	
		App	Val	App	Val	App	Val	App	Val
Descripteurs d'anneau	39	0.997	0.982	0.997	/	384.64	501.348	7.386	8.043
Descripteurs de type P_VSA	07	0.991	0.916	0.991	/	620.197	1393.537	10.166	483.367

Les résidus moyens et la moyennes quadratiques pour les descripteurs d'anneau obtenus sur la représentation de l'enthalpie standard de Combustion par la régression linéaire dans le tableau suivant :

Tableau 15 - Résidus moyens et moyennes quadratiques obtenus (Descripteurs d'anneau)

Descripteurs moléculaire	Résidu moyen	Moyenne quadratique
Descripteurs d'anneau	10.719	308.071
Descripteurs de type P_VSA	14.055	349,966

4.2.1.2 Les graphiques

Une autre manière de représentation des données, est la représentation graphique. Dans cette partie nous avons basés sur trois types des graphes. Ces trois graphiques permettent de visualiser respectivement la distance entre les prédictions et les observations (valeurs expérimentales ; pour un modèle idéal, les points seraient tous sur la bissectrice), les résidus normalisés sous la forme d'un diagramme en bâtons de la collection active, et les résidus normalisés sous la forme d'un diagramme en bâtons de la collection de validation (une partie des données sélectionnées aléatoirement ou non pour valider le modèle).

Ces deux derniers graphiques permettent de rapidement voire si un nombre anormal de données sort de l'intervalle $]-2,2[$ sachant que ce dernier, sous hypothèse de normalité, doit

contenir environ 95% des données. Il est à noter que la collection des données de validation est choisie aléatoirement.

4.2.1.2.1 Descripteurs de type P_VSA

Les graphiques qui suivent permettent de visualiser les résultats mentionnés ci-dessus. Les graphes présentent une évaluation globale des résultats de représentation des enthalpies de combustion en phase solide d'état étudiées par XLSTAT (régression linéaire).

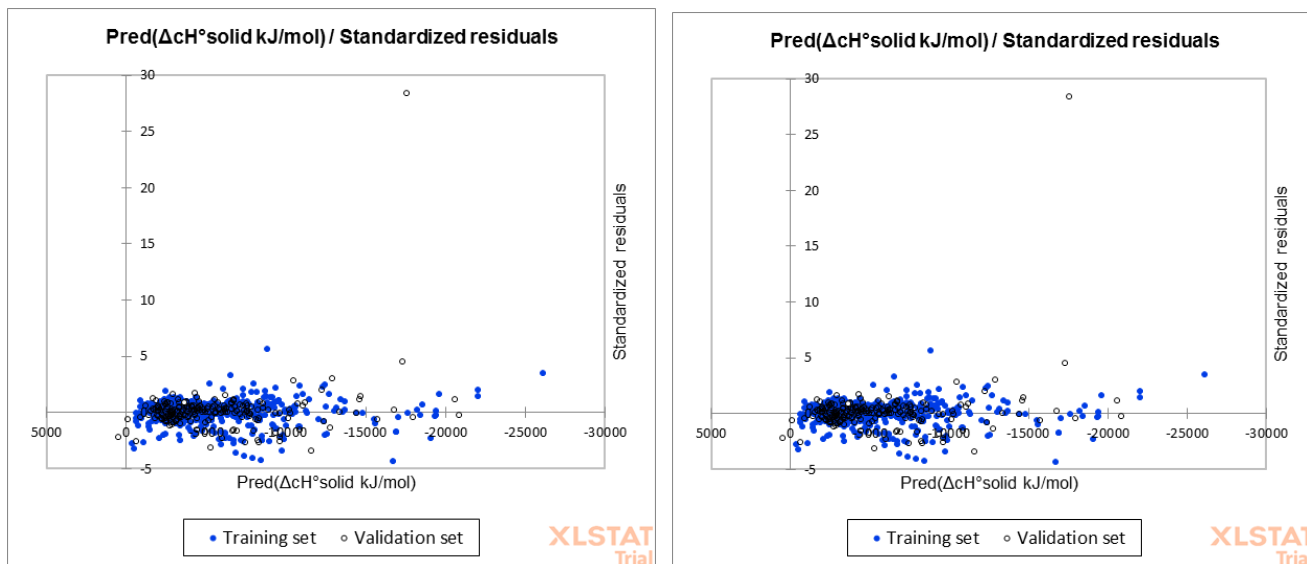


Figure 13 - Le résultats des $\Delta_c H^\circ$ aux descripteurs de type P_VSA

La Figure 15 montre clairement la bonne prédictivité pour les ensembles d'apprentissage et de validation. Tous les composés de calibrage et de validation sont dans le domaine chimique, ce qui suggère qu'il n'y a pas de valeurs aberrantes (en X et/ou en Y) et que la prédiction du modèle est fiable.

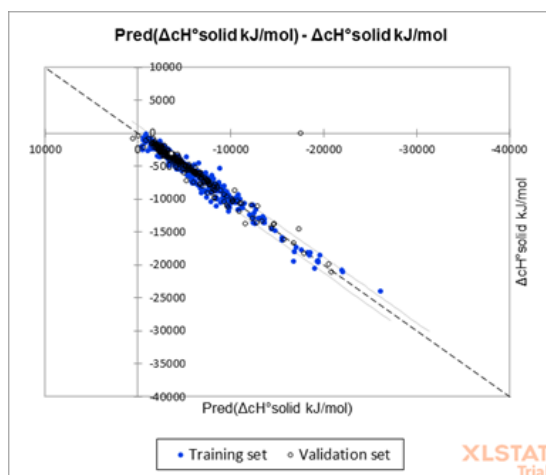


Figure 14 - Corrélation entre les $\Delta_c H^\circ$ Exp et Préd pour les descripteurs de type P_VSA

4.2.1.2.2 Descripteurs d'anneau

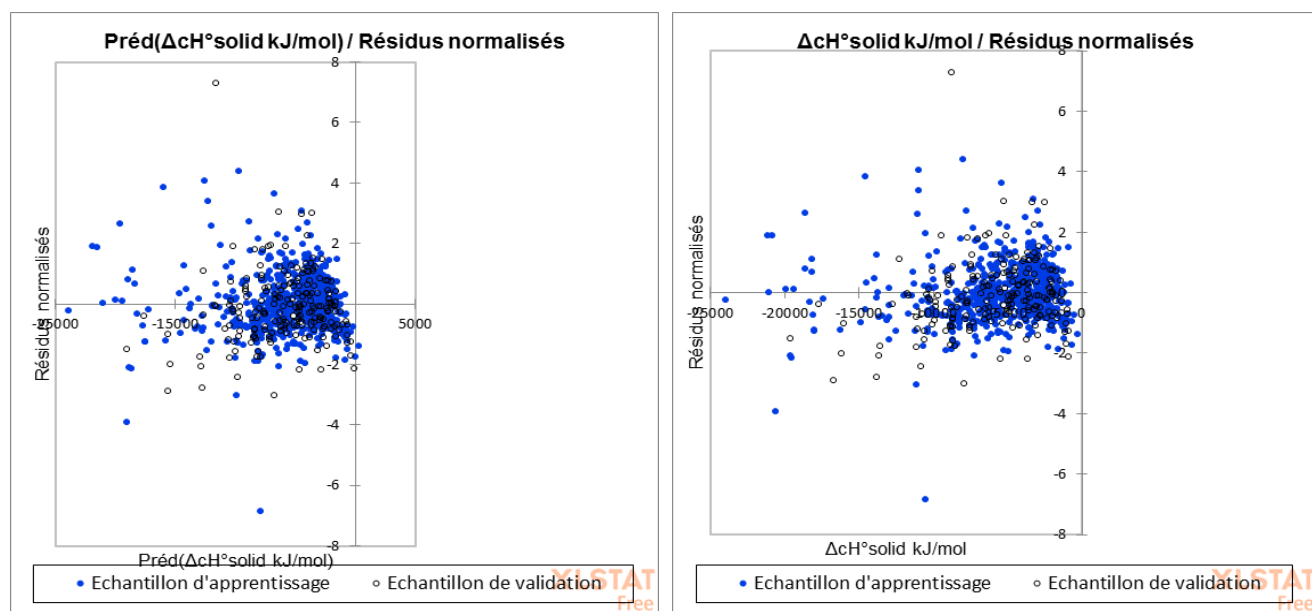


Figure 15 - Le résultats des $\Delta_c H^\circ$ aux descripteurs d'anneau

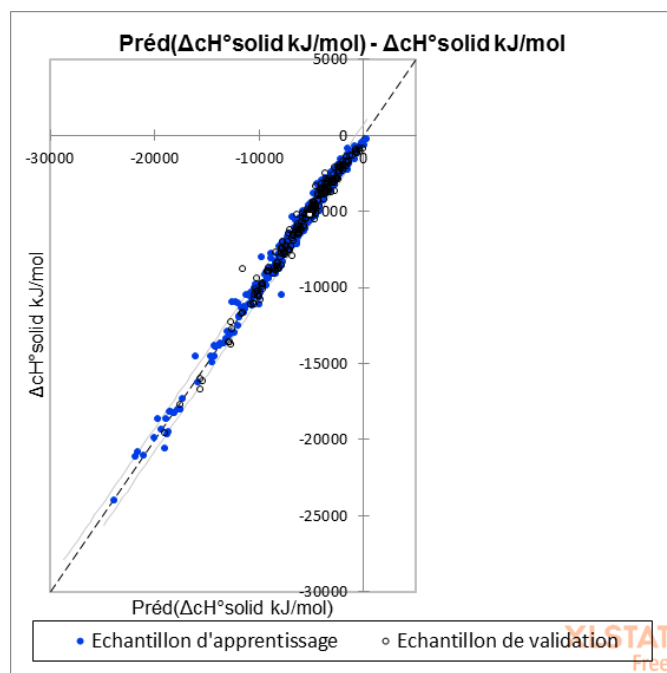


Figure 16 - Corrélation entre les $\Delta_c H^\circ$ Exp et Préd pour les descripteurs d'anneau

4.3 Analyse des résultats

Le **coefficient de détermination (R^2)** 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.

- $R^2 = 0.991$ et $R^2 \text{ ajusté} = 0.991$
- $R^2 = 0.997$ et $R^2 \text{ ajusté} = 0.997$

Donc les pourcentages de la variation des variables à modéliser : **99.7 % et 99.7 %**.

Cela nous indique que les : **99.7 % et 99.7 %** de la variabilité du $\Delta_c H^\circ_{\text{solid}}$ sont expliqués 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

Les résidus :

- $(1 - R^2) \times 100 = (1 - 0.991) \times 100 = 0.9\%$
- $(1 - R^2) \times 100 = (1 - 0.997) \times 100 = 0.3\%$

Cela signifie, entre autres, que 95% des résidus doivent se trouver dans l'intervalle (-2, +2), toute valeur en dehors de cet intervalle est révélatrice d'une donnée suspecte. Afin de mettre en évidence les valeurs se trouvant hors de l'intervalle (-2, +2).

Les paramètres permettent d'expliquer **99.7 % et 99.7 % de** la variabilité de $\Delta_c H^\circ_{\text{solid}}$. Pour expliquer la variabilité restante, d'autres sources de variabilité doivent donc être prises en compte

L'étude de corrélation entre les descripteurs deux à deux ont conduit à classer les descripteurs entre deux extrêmes : Les descripteurs moins corrélés (voir Annexe A), signifie qu'ils ne donnent pas la même information sur la propriété à estimer $\Delta_c H^\circ_{\text{solid}}$ (chaque descripteur à une information différente à celle de l'autre). Ces descripteurs peuvent être importants pour le modèle. Les descripteurs possédants une grande corrélation entre eux (voir Annexe A) signifie qu'ils vont nous donner des informations plus proches sur la propriété à estimer $\Delta_c H^\circ_{\text{solid}}$. Ces descripteurs peuvent être éliminés du modèle.

L'interprétation du graphe **Préd ($\Delta_c H^\circ_{\text{solid}}$) / ($\Delta_c H^\circ_{\text{solid}}$)**, nous montre qu'il y a **785** observations qui pressentent 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.

4.3.1 Mise en évidence de problèmes générale

Tout d'abord, il est important de souligner le fait qu'il existe parfois plusieurs valeurs d'enthalpies standard de formation en phase solide pour un même composé. Nous avons donc analysé à nouveau ce type de molécule pour vérifier sa validité, En effet, il est parfois très compliqué de faire « le bon choix » de valeur d'enthalpie standard de formation permettant la détermination de descripteur moléculaires. Lorsque le type de cas évoqué plus haut s'est présenté, il a donc fallu sélectionner la valeur la plus plausible, i.e. souvent, la valeur d'enthalpie standard de formation qui était retenue dans plusieurs banques de données. Par ailleurs, seule une valeur d'enthalpie standard de formation en phase solide a été publiée pour un nombre non négligeable de substances organiques étudiées. Il est donc, dans ces cas précis, difficile d'être absolument certain de disposer d'une valeur juste.

Enfin, certains groupes sont rarement présents au sein des molécules organiques. C'est pour cette raison que plusieurs descripteurs moléculaires n'ont pu être estimés qu'à l'aide d'un nombre de composés peu élevé. Or, plus ce chiffre est faible et plus une incertitude élevée pèse sur la valeur de descripteur. En réalité, le faible nombre d'occurrences d'un groupe n'est pas directement responsable d'une incertitude élevée (car des mesures expérimentales tout à fait fiables engendrent la détermination de descripteur moléculaires exactes). Cela empêche simplement de certifier la validité de la valeur de descripteur trouvée. Dans le cadre de la méthode établie, il faut considérer que les descripteurs moléculaires qui ont été déterminées grâce à un nombre de molécules inférieur à 3 sont les moins fiables.

4.4 Paramètres des modèles de régression multilinéaire

4.4.1 Coefficients d'ajustement

Statistique	Apprent	Validation	Statistique	Apprent	Validation
Observations	620	206	Observations	626	200
Somme des poids	620	206	Somme des poids	626	200
DDL/ DF	613	199	DDL/ DF	587	161
R²	0.991	0.916	R²	0.997	0.982
R² ajusté	0.991	/	R² ajusté	0.997	/
MCE	384643.778	1941944.438	MCE	147953.881	251349.790
RMCE	620.197	1393.537	RMCE	384.648	501.348
MAPE	10.166	483.367	MAPE	7.386	8.043
DW	1.702	/	DW	1.975	/
Cp	7.000	/	Cp	39.000	/
AIC	7980.205	/	AIC	7490.047	/
SBC	8011.213	/	SBC	7663.181	/
PC	0.009	/	PC	0.004	/
Tableau 16 - Coefficients d'ajustement (Descripteurs de type P_VSA)			Tableau 17 - Coefficients d'ajustement (Descripteurs d'anneau)		

4.4.2 Paramètres du modèle

Source	Valeur	Erreur Standard
LP1	0,029	0,019
Eig1v	0,416	0,032
SEigZ	-1,517	0,231
SEigv	-0,673	0,235
VRA2	-0,127	0,020
VEZ2	-0,188	0,021
VRe1	0,690	0,033

Tableau 18 - Paramètres du modèle (Descripteurs de type P_VSA)

Source	Valeur	Erreur standard	Source	Valeur	Erreur standard	Source	Valeur	Erreur standard
Qindex	-0,119	0,075	J	0,302	0,037	Lop	-0,023	0,014
HNar	0,177	0,090	Jhetv	-0,258	0,018	D/Dr03	0,010	0,003
Xt	-0,013	0,044	MAXDN	0,032	0,017	D/Dr04	0,012	0,003
Ram	-0,311	0,062	MAXDP	-0,011	0,010	D/Dr05	0,093	0,016
VDA	-0,222	0,037	DELS	0,426	0,021	D/Dr06	0,038	0,022
MSD	0,119	0,073	S3K	-0,023	0,015	D/Dr07	0,005	0,004
SPI	0,094	0,030	BLI	-0,688	0,024	D/Dr08	0,017	0,004
WA	-0,798	0,064	PW2	0,628	0,081	D/Dr09	0,021	0,003
TI1	-0,050	0,025	PW3	-0,167	0,042	D/Dr10	0,029	0,005
TI2	0,089	0,023	PW4	0,022	0,025	D/Dr11	-0,008	0,004
STN	-0,167	0,060	PW5	0,035	0,018	D/Dr12	-0,001	0,003
Rww	-0,196	0,036	PJI2	-0,060	0,012	T(N..N)	0,046	0,006
Wap	-0,034	0,012	DECC	0,153	0,018	T(O..O)	0,000	0,009

Tableau 19 - Paramètres du modèle (Descripteurs d'anneau)

Conclusion

Ce travail a porté sur l'étude de composé organique et concerne en particulier l'estimation de L'enthalpies de combustion $\Delta_c H^\circ$ en phase solide. Nous avons cherché à corréler L'enthalpies de combustion expérimentale avec des descripteurs moléculaire.

Après le calcul de descripteurs 2D et les descripteurs 3D par le logiciel Dragon. Ce travail a été effectué sur 1664 paramètres de 20 blocs différents et une base de données contenant 657 composés chimiques de différentes familles. Pendant la prédiction nous avons trouvés une relation mathématique : $\Delta_c H^\circ = f(\text{paramètres structuraux})$, $\Delta_c H^\circ = f(b + \sum_i (a_i \times x_i))$. Ce modèle donne des résultats de corrélation acceptables avec des descripteurs des blocs moléculaire des propriétés, Descripteurs d'anneau et Descripteurs de type P_VSA ($R^2 = 0.991$), ($R^2 = 0.997$), respectivement.

La validité des modèles obtenus a été établie par la détermination des paramètres statistiques appropriés. Les valeurs prédites sont plus proches que les valeurs expérimentales. Les faibles valeurs résiduelles de l'activité biologique et les grandes valeurs du coefficient de validation croisé obtenus suggèrent une bonne capacité prédictive des modèles QSPR développés.

Les paramètres statistiques relatifs à la validation du modèle démontrent que le modèle est simple, fiable avec une bonne capacité prédictive. Puisque ce modèle a été développé sur la base de descripteurs moléculaires théoriques calculés exclusivement à partir de la structure moléculaire.

On peut conclure que la capacité de prédiction de la méthode dépend du type et du nombre de composés dans la base de données. Pour améliorer les résultats, nous devons exclure les particules non corrélées (rejetées).

Il serait intéressant de continuer à développer la méthode établie au cours de cette étude en déterminant des valeurs de descripteur moléculaire contenant d'autres atomes que C, H, O et N. Cela précisé, plus le nombre d'atomes traités augmente et plus il faut déterminer de valeurs de descripteur moléculaire. De très nombreuses données expérimentales seraient donc nécessaires au développement de la méthode présentée dans ces travaux afin d'être le plus complet possible.

Mais il n'est pas toujours aisé de disposer de ces valeurs d'enthalpies mesurées expérimentalement.

Il serait en outre envisageable de créer un logiciel permettant une automatisation du calcul des enthalpies de combustion. Ceci serait fort pratique et permettrait aux personnes non initiées aux méthodes de descripteur moléculaire d'utiliser notamment les résultats présentés dans ce mémoire. Par ailleurs, l'élaboration d'un logiciel permettrait de valoriser de façon notable les valeurs de descripteur moléculaire déterminées au cours de la présente étude. Ceci serait appliqué d'autres méthodes, par exemple méthode de signature de descripteur moléculaire, graphe machine, etc...

A la fin, et pour les travaux du futur, on propose les recommandations suivantes :

- Elargir la base de données pour couvrir toutes les familles chimiques, et la plupart des composés organiques.
- Chercher une méthode pour réduire le nombre des interactions binaires, si on décide d'élargir, en plus, la base des données.
- Appliquer cette technique sur d'autres propriétés physiques et thermodynamiques.
- Procéder une autre étude avec l'utilisation des autres simulateurs

Bibliographie

1. **VITU, Stéphane.** Développement d'une méthode de contributions de groupes pour le calcul du coefficient d'interaction binaire de l'équation d'état de Peng-Robinson et mesures d'équilibres liquide-vapeur de systèmes contenant du CO₂. Lorraine : Institut National Polytechnique, Laboratoire de thermodynamique Polyphasés, 2007.
2. **VATANI , Ali , MEHRPOOYA, Mehdi et GHARAGHEIZI, Farhad.** Prediction of Standard Enthalpy of Formation by a QSPR Model. International Journal of Molecular Sciences. 22 May 2007, pp. 402-432.
3. **SEROUTI , Abdelghani.** Développement d'une méthode prédictive de calcul la variation des enthalpies de formation en phase solide de molécules organiques. Application à la conception des matériaux énergétiques. Ouargla : Université Kasdi Merbah, Faculté des Sciences et de la Technologie, Département de Génie des Procédés, 2012.
4. **SALMON, Anna .** Développement d'une méthode prédictive de calcul des enthalpies de formation en phase solide de molécules organiques - Application aux matériaux énergétiques. Paris : Ecole des Mines, 2006.
5. **SAIHI , Youcef.** Etude de la relation quantitative structure-activité. Annaba : Université Badji Mokhtar, Faculté des Sciences, Département de Chimie, 2015.
6. **MOSTEFAOUI, Larbi.** Contribution à la description et à la compréhension de la solvation des biomolécules. Tlemcen : Université Abou Bekr Belkaid UABT, Département de Chimie , 2011.
7. **MAZOUZI Aziza, NOUI Chahira.** Méthode de prédiction des propriétés physiques et. Ouargla : Faculté des Sciences et de la Technologie et Science de la matière, Département de Génie des Procédés, 2013.
8. **Liew Chin Yee, Yap Chun Wei.** Current Modeling Methods Used in QSAR/QSPR. Germany : Wiley-VCH Verlag GmbH & Co. KGaA, 2012 .
9. **J. B. Pedley, R. D. Naylor, S. P. Kirby (auth.** Thermochemical Data of Organic Compounds- Springer Netherlands . 2nd. London, New York : Chapman and Hall Ltd, 1986.
10. **ISIDORO, Martínez.** Termodinámica básica y aplicada. 1st. Madrid : Dossat, 1992.
11. **HMAMOUCI, Rachid, BOUACHRINE, Mohammed et LAKHLIFI, Tahar.** Tentative Pratique du Relation Quantitatives Structure-Activité/Propriété. Meknes, Morocco : University Moulay Ismail, Faculté des sciences, Laboratoire de chimie moléculaire et substances naturelles., 2016.
12. **GUENDOUI, Abdelkrim .** Élaboration des modèles QSPR prédictifs des propriétés physico-chimiques à l'aide des descripteurs moléculaires. Tlemcen : Université Abou-bekr belkaïd, Faculté des sciences, Département de chimie, 2015.

- 13. GOULON, SIGWALT, ABRAM.** Une nouvelle méthode d'apprentissage de données structurées : applications à l'aide à la découverte de médicaments. Paris : Université Pierre et Marie Curie, 2008.
- 14. FORTUNÉ, Antoine.** Techniques de Modélisation Moléculaire Appliquées à l'Etude et à l'Optimisation de Molécules Immunogènes et de Modulateurs de la Chimiorésistance. Grenoble : Université Joseph Fourier, 2006.
- 15. FAYET, Guillaume.** Développement De modèles QSPR Pour La Prédiction Des Propriétés D'explosibilité Des Composes NITROAROMATIQUES. Paris : Université Pierre Et Marie Curie, 2010.
- 16. EN-NAHLI, Fatima.** Etude de la Relation Quantitative Structure Activité de l'acide ursolique et ses dérivés. Fès, Morocco : Université Sidi Mohamed Ben Abdellah, Faculté des Sciences et Techniques, 2019.
- 17. DEBBA , Fatima Zohra.** Un modèle chimio-informatique pour une synthèse virtuelle. Annaba : Université Badji Mokhtar, Faculté des sciences de l'ingénieur, Département d'informatique, 2010.
- 18. CHTITA, Samir.** Modélisation de molécules organiques hétérocycliques biologiquement actives par des méthodes QSAR/QSPR. Recherche de nouveaux médicaments. Meknès : Université Moulay Ismail, Faculté des sciences, 2017.
- 19. BOUARRA, Nabil.** Etudes QSPR des propriétés contrôlant l'évolution de quelques HAP dans l'environnement. Annaba : Université Badji Mokhtar, Faculté des sciences, Département de chimie, 2018.
- 20. SAIDAT, Mustapha.** Etude et développement d'une bibliothèque orientée objet de calcul de propriétés physiques et thermodynamique basée sur le concept de contribution des groupes organiques fonctionnels. Ouargla : Université Kasdi Merbah , 2004.
- 21. M.E. Minas da Piedade.** Energetics of Stable Molecules and Reactive Intermediates. Lisboa, Portugal : kluwer academic, 1999.
- 22. J. Chen, Yy Di, Zc Tan, Ht Zhang, L. Sun.** Determination of Standard Molar Enthalpy of Formation. Dalian, China : Thermochemistry Laboratory, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, 2005.
- 23. Article, Wikipedia.** Méthode de contribution du groupe. Wikipedia. 2020, https://fr.qwe.wiki/wiki/Group_contribution_method.
- 24. Addinsoft-Team.** XLSTAT 2020. Paris, France : Addinsoft, 2020.
- 25. DEUDON, Valentin.** Le Cnrs, Themes, composes organique. 2008. <https://www.cnrs.fr/cnrs-images/chimieaulyce/THEMES/organique/composes.htm>.
- 26. Dailypresse.** Définition Calorimètre. 2019. <http://www.dailypresse.com/definition/calorimetre/>.

27. **GUILLAUME, Josse.** Futura-sciences, Sciences, Physique-Enthalpie. 2016. <https://www.futura-sciences.com/sciences/definitions/physique-enthalpie-2079/>.
28. **Feunang, D. et al.** Fisher Scientific, Produits, Composés organiques. 2016. <https://www.Fisher.sci.ca/ca/en/products/80003050/organic-compounds.html>.
29. **TURGEON, Pierre Alexandre.** Détermination de l'enthalpie de réaction. 2013. http://glace.recherche.usherbrooke.ca/wp-content/uploads/2013/11/CPH316_calorimetrie20180118.pdf.
30. **Minitab, Assistance.** Présentation de la fonction Régression non linéaire. 2019. <https://support.minitab.com/fr-fr/minitab/18/help-and-how-to/modeling-statistics/regression/supporting-topics/nonlinear-regression/understanding-nonlinear-regression/>.
31. **IONOS, Digital guide.** Le modèle de bases de données relationnelles. 2019. <https://www.ionos.fr/digitalguide/hebergement/aspects-techniques/base-de-donnees-relationnelle/>.
32. **Chemistry-Archive.** Index of /Ionium/Rhodium/pdf/chemical-data: Heat of Combution. 26 Février 2007. <https://www.thevespiary.org/rhodium/Rhodium/pdf/chemical-data/heat.pdf>.

Annexe A

A.1 HperChem

A.1.1 Définition

HyperChem est un modeler moléculaire flexible et éditeur et un paquet computationnel puissant. Il offre beaucoup de types de calculs des mécaniques moléculaires et quantiques.

A.1.2 Construire et Afficher des Molécules

Soumission d'une structure à un calcul peut être coûteuse en termes de temps et l'effort humains. HyperChem vous permet de créer et d'afficher facilement des molécules. Depuis HyperChem contient une interface graphique, vous pouvez suivre la construction de molécules.

Utilisation de l'outil de dessin, vous pouvez dessiner une représentation en deux dimensions (2D) d'une molécule, puis utiliser le générateur de modèle pour générer une structure tridimensionnelle (3D). Le constructeur du modèle ajoute de l'hydrogène est implicite à la molécule à votre demande. Vous pouvez également manipuler des obligations individuelles, des géométries des obligations, des angles, des torsions et charges atomiques au cours de la construction de modèles (en utilisant les commandes limitent dans le menu Générer) ou après la construction de modèles (en utilisant les commandes Set dans le menu Edition).

A.1.3 Construire et Afficher des Molécules

Pour calculer les propriétés d'une molécule, vous avez besoin de produire une structure précise. Un calcul exige souvent une structure qui représente un minimum sur une surface d'énergie potentielle. HyperChem contient plusieurs optimiseurs de la géométrie pour faire

ceci. Vous pouvez calculer alors les point seules propriétés d'une molécule ou utiliser la structure optimisée comme un point de départ pour calculs subséquents, tel que simulations de la dynamique moléculaire.

A.1.4 Utilisant les Méthodes des Mécaniques Moléculaires

HyperChem offre quatre champs de force de mécanique moléculaires : MM+, AMBER, BIO+, et OPLS. Pour exécuter un calcul de mécanique moléculaire, vous devez d'abord choisir un champ de force. Les sections suivantes discutent les considérations dans le choix d'un champ de force.

A.1.5 Disponibilité de Paramètres

Les champs de la force donnent les meilleurs résultats pour les molécules semblable à ce développait ses paramètres. Choisissez un champ de la force développé pour une gamme de molécules semblables à votre système moléculaire.

A.1.6 Traits du Champ de la force

Les équations du champ de la force pour MM+, l'AMBRE, BIO+, et OPLS est semblable dans les types de termes qu'ils contiennent : liaison, angle, van der Waals, et électrostatique. Il y a des différences dans les formes des équations qui peuvent affecter votre choix de champ de la force.

A.1.7 MM+

MM+ est unique parmi les champs de la force dans le chemin il traite des liaisons et des angles. La liaison et termes de l'angle peuvent contenir de plus hauts termes de l'ordre que le niveau du second degré. Ceux-ci lient et les possibilités de l'angle expriment mieux qu'une possibilité harmonique un mouvement harmonique.

MM+ contient aussi un étirement coude terme en colère, a appelé un terme Urey-Bradley. Les autres champs de la force dans HyperChem évaluent normalement des interactions de la nonliaison pour atomes séparés par trois ou plus de liaison (1–4 et plus grandes interactions). UN

terme Urey-Bradley inclut 1–3 interactions qui est critique pour simuler des molécules correctement.

Par exemple, la liaison orientée pour les atomes lourds dans cyclo-butane est comprimée, comparée à l'angle de la liaison naturel pour les orbitales C sp³, et les liaisons du carbone-carbone ont le plus haut caractère p. Cela résulte en une liaison plus faible et une liaison carbone-carbone durée augmentée. Un terme Urey-Bradley tient compte de ces changements structuraux et reconnaît les 1–3 interactions uniques qui se produisent dans les molécules filtrées.

HyperChem augmente les MM2 standard en fournissant des paramètres supplémentaires (constantes de la force) pour utiliser deux plans alternatifs (voyez la deuxième partie de ce livre, Théorie et Méthodes). Cela étend la gamme de composés chimiques que MM+ peut accommoder. MM+ fournit aussi des raccourcis pour les interactions nonliaisons calculatrices et les conditions de la limite périodiques.

A.1.8 Interactions électrostatiques

Une autre différence entre les champs de la force est le calcul d'interactions électrostatiques. L'AMBRE, BIO+, et OPLS utilisent les charges ponctuelles pour modéliser des interactions électrostatiques. MM+ calcule des interactions électrostatiques qui utilisent des dipôles de la liaison. La méthode du dipôle de la liaison ne peut pas simuler suffisamment les systèmes très polaires ou chargés.

A.2 Autres paramètres du modèle

Source	DDL	Somme des carrés	Moyenne des carrés	F	Pr > F
Modèle	39	27726594564,244	710938322,160	4805,135	<0,0001
Erreur	587	86848928,178	147953,881		
Total corrigé	626	27813443492,421			

Tableau 20 - Analyse de la variance (Descripteurs d'anneau)

Source	DDL	Somme des carrés	Moyenne des carrés	F	Pr > F
Modèle	7	27344980939,708	3906425848,530	10155,957	<0.0001
Erreur	613	235786635,705	384643,778		
Total corrigé	620	27580767575,413			

Tableau 21 - Analyse de la variance (Descripteurs de type P_VSA)

	LP1	Eig1v	SEigZ	SEigv	VRA2	VEZ2	VRe1	$\Delta cH^\circ_{\text{solid}}$ kJ/mol
LP1	1	0,310	0,226	-0,229	0,740	-0,634	0,564	0,375
Eig1v	0,310	1	0,680	-0,682	0,224	-0,724	0,878	0,545
SEigZ	0,226	0,680	1	-0,999	-0,151	-0,362	0,454	-0,165
SEigv	-0,229	-0,682	-0,999	1	0,155	0,369	-0,454	0,160
VRA2	0,740	0,224	-0,151	0,155	1	-0,595	0,535	0,583
VEZ2	-0,634	-0,724	-0,362	0,369	-0,595	1	-0,865	-0,709
VRe1	0,564	0,878	0,454	-0,454	0,535	-0,865	1	0,761
$\Delta cH^\circ_{\text{solid}}$ kJ/mol	0,375	0,545	-0,165	0,160	0,583	-0,709	0,761	1

Tableau 22 - Matrice de corrélation (Descripteurs de type P_VSA)

Tableau 23 - Les étapes de recherche des données à partir de nbre d'enregistrement CAS

An official website of the United States government [Info](#)

NIST National Institute of Standards and Technology
U.S. Department of Commerce

Careers | Shop | Contact

Search NIST

Topics ▾ Publications Labs & Major Programs ▾ Services & Resources ▾ News & Events ▾ About NIST ▾

Standards and Measurements
Calibration Services
Laboratory Accreditation (NVLAP)
Quality System
Standard Reference Materials (SRMs)
Standards.gov
Time Services
Office of Weights and Measures

Software
Data
[Chemistry WebBook](#)
National Vulnerability Database
Physical Reference Data
Standard Reference Data (SRD)
Storefront
License & Patents

Computer Security Resource Center (CSRC)
NIST Research Library

NIST National Institute of Standards and Technology
U.S. Department of Commerce

NIST Chemistry WebBook, SRD 69

Home Search ▾ NIST Data ▾ About ▾

NIST Chemistry WebBook

Reference Database Number 69

Last updated: 10/18/2017 DOI: <https://doi.org/10.18434/T4D303>

Search for species by CAS number

View: [Search Options](#), [Models and Tools](#), [Special Data Collections](#), [Documentation](#), [Changes](#), [Notes](#)

► Credits

NIST reserves the right to charge for access to this database in the future.

NIST recently released a new version of the [NIST Inorganic Crystal Structure Database \(ICSD\)](#). For more information visit the [ICSD web site](#).

NIST National Institute of Standards and Technology
U.S. Department of Commerce

NIST Chemistry WebBook, SRD 69

Home Search ▾ NIST Data ▾ About ▾

Search for Species Data by CAS Registry Number

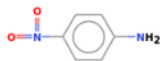
Please follow the steps below to conduct your search:

- Enter a registry number (e.g., 74-82-8):
- Select the desired units for thermodynamic data:
☒ SI ☐ calorie-based
- Select the desired type(s) of data:

Thermodynamic Data	Other Data
☒ Gas phase	☒ IR spectrum
☒ Condensed phase	☒ THz IR spectrum
☒ Phase change	☒ Mass spectrum
☒ Reaction	☒ UV/Vis spectrum
☒ Ion energetics	☒ Gas Chromatography
☒ Ion cluster	☒ Vibrational & electronic energy levels
	☒ Constants of diatomic molecules
	☒ Henry's Law
- Press here to search:

p-Nitroaniline

- Formula: $C_6H_6N_2O_2$
- Molecular weight: 138.1240
- IUPAC Standard InChI:
 - InChI=1S/C6H6N2O2/c7-5-1-3-6(4-2-5)8(9)10/h1-4H,7H2
 - Download the identifier in a file.
- IUPAC Standard InChIKey: TYMLOMAKGOJONV-UHFFFAOYSA-N
- CAS Registry Number: 100-01-6
- Chemical structure:



This structure is also available as a 2d Mol file or as a computed 3d SD file

The 3d structure may be viewed using Java or Javascript.

- Other names: Benzenamine, 4-nitro-; Aniline, p-nitro-; p-Aminonitrobenzene; p-Nitrophenylamine; Azoamine Red ZH; C.I. Azoic Diazo Component 37; C.I. Developer 17; C.I. 37035; Developer P; Devol Red GG; Fast Red Base GG; Fast Red Base 2J; Fast Red GG Base; Fast Red MP Base; Fast Red P Base; Fast Red 2G Base; Naphtolan Red GG Base; Nitrazol CF extra; PNA; Red 2G Base; Shinnippon Fast Red GG Base; 1-Amino-4-nitrobenzene; 4-Nitraniline; 4-Nitroaniline; 4-Nitrobenzenamine; p-Nitraniline; Aniline, 4-nitro-; Azofix Red GG Salt; Azoic Diazo Component 37; Diazo Fast Red GG; Fast Red GG Salt; Fast Red P Salt; Fast Red Salt GG; Fast Red Salt 2J; Fast Red 2G Salt; NCI-C60786; p-Nitroanilina; Paranitroaniline; Rcra waste number P077; 4-Aminonitrobenzene; Azoamine Red ZH; NSC 9797

Condensed phase thermochemistry data

Go To: Top, Gas phase thermochemistry data, Phase change data, Reaction thermochemistry data, Gas phase ion energetics data, IR Spectrum, Mass spectrum (electron ionization), UV/Visible spectrum, Gas Chromatography, References, Notes

Data compilation copyright by the U.S. Secretary of Commerce on behalf of the U.S.A. All rights reserved.

Data compiled as indicated in comments:

ALS - Hussein Y. Afeefy, Joel F. Liebman, and Stephen E. Stein

DH - Eugene S. Domalski and Elizabeth D. Hearing

Quantity	Value	Units	Method	Reference	Comment
$\Delta_f H^\circ_{\text{solid}}$	-45.6 $\pm$ 1.2	kJ/mol	Ccr	Nishiyama, Sakiyama, et al., 1983	ALS
$\Delta_f H^\circ_{\text{solid}}$	-43.1 $\pm$ 0.8	kJ/mol	Ccb	Lebedeva, Gutner, et al., 1971	Hfusion=5.1 $\pm$ 0.2, see Lebedeva, Rjadnenko, et al., 1969; ALS
$\Delta_f H^\circ_{\text{solid}}$	-38.5 $\pm$ 3.2	kJ/mol	Ccb	Medard and Thomas, 1954	Heat of combustion corrected for pressure; ALS
$\Delta_f H^\circ_{\text{solid}}$	-41.5	kJ/mol	Cm	Cole and Gilbert, 1951	ALS
Quantity	Value	Units	Method	Reference	Comment
$\Delta_c H^\circ_{\text{solid}}$	-3172.98 $\pm$ 0.90	kJ/mol	Ccr	Nishiyama, Sakiyama, et al., 1983	ALS
$\Delta_c H^\circ_{\text{solid}}$	-3176. $\pm$ 0.8	kJ/mol	Ccb	Lebedeva, Gutner, et al., 1971	Hfusion=5.1 $\pm$ 0.2, see Lebedeva, Rjadnenko, et al., 1969; ALS
$\Delta_c H^\circ_{\text{solid}}$	-3180.0 $\pm$ 3.2	kJ/mol	Ccb	Medard and Thomas, 1954	Heat of combustion corrected for pressure; ALS
$\Delta_c H^\circ_{\text{solid}}$	-3177.1	kJ/mol	Cm	Cole and Gilbert, 1951	ALS

Tableau 24 - Nbre d'enregistrement CAS et la masse moléculaire pour les données étudier

N°	CAS N°	MW	N°	CAS N°	MW	N°	CAS N°	MW
01	100-01-6-2d	138.140	50	108-80-5-2d	129.090	99	126-58-9-2d	254.320
02	100-09-4-2d	152.160	51	109-00-2-2d	95.110	100	129-00-0-2d	202.260
03	100-10-7-2d	149.210	52	109-77-3-2d	66.070	101	13114-72-2-2d	226.300
04	100-21-0-2d	166.140	53	110-14-5-2d	116.140	102	13184-80-0-2d	524.310
05	100-25-4-2d	168.120	54	110-15-6-2d	118.100	103	131-91-9-2d	173.180
06	1002-84-2-2d	242.450	55	110-16-7-2d	116.080	104	13222-85-0-2d	254.300
07	1003-29-8-2d	95.110	56	110-17-8-2d	116.080	105	13225-10-0-2d	374.210
08	100-54-9-2d	104.120	57	110-21-4-2d	118.120	106	13228-22-3-2d	272.430
09	1006-67-3-2d	145.170	58	110-61-2-2d	80.100	107	132-53-6-2d	173.180
10	100-76-5-2d	111.210	59	110-85-0-2d	86.160	108	13358-02-6-2d	196.230
11	1008-74-8-2d	159.200	60	110-94-1-2d	132.130	109	134-32-7-2d	143.200
12	1008-75-9-2d	159.200	61	111-16-0-2d	160.190	110	13576-20-0-2d	114.130
13	100-97-0-2d	140.220	62	111-20-6-2d	202.280	111	135-88-6-2d	219.300
14	101-25-7-2d	186.210	63	1120-12-3-2d	173.290	112	136-35-6-2d	197.260
15	101-60-0-2d	310.380	64	1121-25-1-2d	109.140	113	13980-04-6-2d	174.150
16	101-68-8-2d	250.270	65	1121-78-4-2d	109.140	114	13985-60-9-2d	241.180
17	10170-03-3-2d	943.060	66	112-37-8-2d	186.330	115	140-79-4-2d	144.160
18	1017-93-2-2d	208.240	67	113-00-8-2d	59.090	116	141-82-2-2d	104.070
19	101-81-5-2d	168.250	68	1145-32-0-2d	236.250	117	141-86-6-2d	109.150
20	1022-07-7-2d	242.170	69	115-77-5-2d	136.170	118	14213-16-2-2d	160.200
21	10230-68-9-2d	103.090	70	118-55-8-2d	214.230	119	142-18-7-2d	274.450
22	102-96-5-2d	149.160	71	1188-01-8-2d	146.170	120	142-45-0-2d	114.060
23	10311-74-7-2d	208.320	72	118-92-3-2d	137.150	121	143-07-7-2d	200.360
24	10323-20-3-2d	150.150	73	118-96-7-2d	227.150	122	144-62-7-2d	90.040
25	103-29-7-2d	182.280	74	119-26-6-2d	198.160	123	14474-71-6-2d	246.340
26	103-30-0-2d	180.260	75	119-27-7-2d	198.150	124	14474-78-3-2d	218.280
27	10349-06-1-2d	146.160	76	119-32-4-2d	152.170	125	1450-63-1-2d	358.500
28	10349-09-4-2d	160.190	77	119-53-9-2d	212.260	126	1456-22-0-2d	162.160
29	10349-12-9-2d	278.280	78	119-75-5-2d	214.240	127	1459-11-6-2d	246.480
30	10349-14-1-2d	161.180	79	1198-98-7-2d	160.190	128	1459-93-4-2d	194.200
31	10359-20-3-2d	105.110	80	1199-00-4-2d	160.190	129	1460-02-2-2d	246.480
32	103-84-4-2d	135.180	81	120-12-7-2d	178.240	130	14760-99-7-2d	206.140
33	10444-59-4-2d	188.220	82	120-61-6-2d	194.200	131	14771-77-8-2d	156.210
34	104-71-2-2d	236.340	83	120-72-9-2d	117.160	132	147-85-3-2d	115.150
35	10507-71-8-2d	188.300	84	121-14-2-2d	182.150	133	14794-69-5-2d	288.370
36	1055-23-8-2d	354.460	85	121-82-4-2d	222.150	134	14794-71-9-2d	288.370
37	105-60-2-2d	113.180	86	121-91-5-2d	166.140	135	148-24-3-2d	145.170
38	10565-52-3-2d	222.140	87	121-92-6-2d	167.130	136	1483-64-3-2d	362.540
39	10639-26-6-2d	342.340	88	122-39-4-2d	169.240	137	1484-12-4-2d	181.250
40	106-50-3-2d	108.160	89	122-42-9-2d	179.240	138	149-32-6-2d	122.140
41	106-57-0-2d	114.120	90	1225-13-4-2d	270.300	139	1499-10-1-2d	330.440
42	1072-67-9-2d	98.120	91	122-66-7-2d	184.260	140	150-13-0-2d	137.150
43	107-97-1-2d	89.110	92	123-56-8-2d	99.100	141	1502-47-2-2d	218.220
44	1079-71-6-2d	186.320	93	123-77-3-2d	116.100	142	150-30-1-2d	165.210

45	1080-16-6-2d	182.240	94	123-94-4-2d	358.630	143	150-69-6-2d	180.230
46	108-13-4-2d	102.110	95	123-99-9-2d	188.250	144	1518-16-7-2d	204.200
47	1083-56-3-2d	210.340	96	124-04-9-2d	146.160	145	1520-42-9-2d	258.380
48	108-45-2-2d	108.160	97	126-11-4-2d	151.140	146	1523-11-1-2d	186.220
49	108-78-1-2d	126.150	98	1263-63-4-2d	594.820	147	1538-74-5-2d	193.270

148	1540-80-3-2d	188.340	198	1943-08-4-2d	284.550	248	22293-38-5-2d -	error -999
149	156-10-5-2d	198.240	199	1943-09-5-2d	340.670	249	22361-85-9-2d	566.760
150	1561-49-5-2d	286.360	200	1943-87-9-2d	196.180	250	22401-53-2-2d	231.180
151	1563-87-7-2d	177.220	201	195-19-7-2d	228.300	251	2243-32-5-2d	192.280
152	157-06-2-2d	174.240	202	19740-34-2-2d	164.320	252	2243-95-0-2d	263.180
153	1576-69-8-2d	206.300	203	19768-02-6-2d	222.260	253	22706-20-3-2d	160.200
154	15770-21-5-2d	160.190	204	19816-73-0-2d	244.420	254	22719-28-4-2d	274.260
155	1585-40-6-2d	298.170	205	19836-28-3-2d	388.150	255	2277-23-8-2d	246.390
156	15897-81-1-2d	180.270	206	1985-48-4-2d	127.210	256	2292-79-7-2d	188.340
157	1610-39-5-2d	240.420	207	198-55-0-2d	252.320	257	2294-94-2-2d	334.480
158	16132-20-0-2d	239.300	208	19947-75-2-2d	112.150	258	230-17-1-2d	180.220
159	16225-26-6-2d	234.370	209	20062-22-0-2d	450.250	259	230-27-3-2d	179.230
160	1625-84-9-2d	306.340	210	2034-53-9-2d	124.250	260	2305-32-0-2d	172.200
161	1630-08-6-2d	243.160	211	2039-49-8-2d	221.270	261	2308-61-4-2d	294.380
162	1633-22-3-2d	208.320	212	2040-73-5-2d	234.360	262	2319-97-3-2d	208.320
163	16421-52-6-2d	86.070	213	2044-37-3-2d	198.340	263	23242-92-4-2d	846.410
164	1670-17-3-2d	211.200	214	2044-88-4-2d	197.170	264	2325-17-9-2d	174.230
165	1678-45-1-2d	274.450	215	20532-03-0-2d	354.460	265	233-59-0-2d	169.200
166	16820-12-5-2d	142.180	216	206-44-0-2d	202.260	266	23470-00-0-2d	330.570
167	1687-36-1-2d	192.380	217	20666-86-8-2d	264.300	267	23559-36-6-2d	162.160
168	1694-20-8-2d	225.260	218	20686-66-2-2d	183.220	268	23579-46-6-2d	161.190
169	17082-12-1-2d	182.240	219	20743-50-4-2d	160.200	269	23627-80-7-2d	342.160
170	17088-73-2-2d	94.080	220	208-96-8-2d	152.200	270	2364-46-7-2d	263.180
171	17096-47-8-2d	165.130	221	20919-97-5-2d	238.140	271	2386-26-7-2d	209.270
172	17199-29-0-2d	152.160	222	20919-99-7-2d	387.160	272	2388-12-7-2d	216.360
173	17215-44-0-2d	454.250	223	2103-16-4-2d	160.190	273	23987-32-8-2d	874.430
174	17267-51-5-2d	113.150	224	2103-18-6-2d	202.280	274	2400-02-4-2d	358.720
175	1740-57-4-2d	164.180	225	21084-84-4-2d	162.160	275	2407-87-6-2d	344.450
176	17557-81-2-2d	136.080	226	2113-58-8-2d	199.220	276	2424-32-0-2d	142.130
177	17760-68-8-2d	320.450	227	21211-65-4-2d	146.210	277	24331-09-7-2d	116.090
178	17854-07-8-2d	486.680	228	21565-41-3-2d	112.110	278	2436-79-5-2d	239.300
179	1798-20-5-2d	256.490	229	2161-85-5-2d	194.200	279	2466-76-4-2d	110.130
180	18039-42-4-2d	146.170	230	2163-12-4-2d	194.200	280	24973-49-7-2d	179.230
181	18039-45-7-2d	222.270	231	217-59-4-2d	228.300	281	2510-76-1-2d	208.320
182	1804-15-5-2d	102.110	232	218-01-9-2d	228.300	282	25155-29-7-2d	120.130
183	18168-01-9-2d	240.330	233	21898-53-3-2d	267.310	283	25321-14-6-2d	182.150
184	18212-12-9-2d	295.200	234	2199-44-2-2d	167.230	284	2540-31-0-2d	253.320
185	1852-04-6-2d	216.310	235	2199-45-3-2d	167.230	285	25767-20-8-2d	710.900
186	18610-59-8-2d	131.150	236	2199-47-5-2d	195.290	286	25894-11-5-2d	193.270
187	18615-86-6-2d	109.140	237	2199-51-1-2d	167.230	287	259-79-0-2d	152.200
188	1874-14-2-2d	269.230	238	2199-52-2-2d	167.230	288	2603-10-3-2d	151.180
189	1874-22-2-2d	167.130	239	2199-53-3-2d	167.230	289	260-36-6-2d	179.230
190	1874-23-3-2d	171.120	240	2199-54-4-2d	181.260	290	260-94-6-2d	179.230
191	1885-34-3-2d	224.290	241	2199-59-9-2d	195.240	291	26447-28-9-2d	112.090

192	18869-29-9-2d	208.320	242	2199-64-6-2d	195.240	292	26460-76-4-2d	202.370
193	1892-35-9-2d	191.200	243	22092-57-5-2d	202.380	293	26655-34-5-2d	180.180
194	19092-03-6-2d	242.170	244	22092-58-6-2d	216.410	294	26670-16-6-2d	166.180
195	19112-42-6-2d	410.580	245	22092-59-7-2d	230.440	295	2672-58-4-2d	252.240
196	19159-68-3-2d	452.230	246	2210-07-3-2d	127.210	296	2683-82-1-2d	534.860
197	1924-47-6-2d	281.220	247	2226-96-2-2d	173.290	297	269-12-5-2d	169.200

298	27032-01-5-2d	143.160	348	30236-29-4-2d	702.330	398	38677-56-4-2d	241.180
299	2717-39-7-2d	184.300	349	302-72-7-2d	89.110	399	3891-59-6-2d	390.380
300	2721-31-5-2d	126.230	350	3042-22-6-2d	143.200	400	389-36-6-2d	192.140
301	2721-32-6-2d	96.150	351	30545-28-9-2d	202.370	401	3964-18-9-2d	176.200
302	27256-18-4-2d	119.110	352	3058-38-6-2d	258.180	402	39900-53-3-2d	282.220
303	273-59-6-2d	120.120	353	3061-90-3-2d	236.300	403	4023-65-8-2d	174.120
304	27497-47-8-2d	300.470	354	31178-70-8-2d	150.150	404	40484-82-0-2d	223.300
305	275-51-4-2d	128.180	355	313-74-6-2d	256.360	405	4055-94-1-2d	225.140
306	2763-67-9-2d	157.290	356	3155-16-6-2d	242.240	406	40673-57-2-2d	390.600
307	277-10-1-2d	104.160	357	316-51-8-2d	256.360	407	4076-43-1-2d	256.360
308	2782-04-9-2d	192.140	358	3211-87-8-2d	121.200	408	4122-45-6-2d	132.090
309	2783-98-4-2d	138.120	359	32255-35-9-2d	407.230	409	41463-68-7-2d	194.200
310	2789-88-0-2d	206.300	360	32597-34-5-2d	110.130	410	41463-69-8-2d	194.200
311	279-23-2-2d	96.190	361	3279-76-3-2d	109.140	411	4146-43-4-2d	146.180
312	280-33-1-2d	110.220	362	328-38-1-2d	131.200	412	4164-28-7-2d	90.100
313	280-57-9-2d	112.200	363	328-39-2-2d	131.200	413	4164-37-8-2d	176.160
314	280-65-9-2d	124.250	364	328-50-7-2d	146.110	414	4185-47-1-2d	240.160
315	28071-99-0-2d	192.380	365	3327-06-8-2d	368.320	415	420-04-2-2d	42.050
316	281-23-2-2d	136.260	366	3332-38-5-2d	208.320	416	42216-29-5-2d	161.150
317	28211-96-3-2d	638.830	367	3333-52-6-2d	136.220	417	42216-41-1-2d	143.130
318	2831-60-9-2d	228.180	368	33491-88-2-2d	590.310	418	4286-49-1-2d	148.140
319	283-24-9-2d	125.240	369	33504-43-7-2d	121.150	419	4384-23-0-2d	320.560
320	283-50-1-2d	138.280	370	3366-61-8-2d	226.300	420	4418-61-5-2d	85.090
321	283-68-1-2d	220.440	371	33687-74-0-2d	195.240	421	4427-97-8-2d	226.350
322	28375-86-2-2d	202.370	372	33719-66-3-2d	522.750	422	4433-16-3-2d	424.210
323	28453-24-9-2d	347.220	373	33719-95-8-2d	548.740	423	443-79-8-2d	131.200
324	28464-24-6-2d	326.220	374	3376-48-5-2d	246.390	424	448-70-4-2d	478.740
325	28464-26-8-2d	310.220	375	3376-59-8-2d	196.220	425	448-71-5-2d	478.740
326	28787-21-5-2d	147.120	376	338-69-2-2d	89.110	426	4488-57-7-2d	134.240
327	288-13-1-2d	68.090	377	33902-83-9-2d	534.760	427	4589-97-3-2d	164.180
328	288-32-4-2d	68.090	378	3443-83-2-2d	302.510	428	461-58-5-2d	84.100
329	288-94-8-2d	error-999	379	3458-28-4-2d	180.180	429	471-46-5-2d	88.080
330	2892-51-5-2d	114.060	380	3470-17-5-2d	252.120	430	4732-14-3-2d	257.180
331	2896-70-0-2d	171.270	381	34815-01-5-2d	129.100	431	4744-47-2-2d	222.270
332	290-37-9-2d	80.100	382	34880-53-0-2d	error-999	432	4758-64-9-2d	209.320
333	2913-24-8-2d	236.380	383	34882-73-0-2d	372.150	433	476-73-3-2d	254.160
334	2919-19-9-2d	208.320	384	35364-31-9-2d	212.280	434	477-75-8-2d	254.340
335	2919-20-2-2d	208.320	385	35392-77-9-2d	174.220	435	4789-34-8-2d	196.420
336	2921-92-8-2d	269.200	386	35392-80-4-2d	174.220	436	4789-35-9-2d	196.420
337	29415-95-0-2d	152.310	387	36360-34-6-2d	376.440	437	4793-98-0-2d	308.180
338	294-62-2-2d	168.360	388	3637-10-3-2d	173.290	438	479-45-8-2d	287.170
339	295-17-0-2d	196.420	389	3637-11-4-2d	171.270	439	479-47-0-2d	254.160
340	295-37-4-2d	200.380	390	3646-73-9-2d	180.180	440	4815-44-5-2d	254.320

341	295-48-7-2d	210.450	391	3674-69-9-2d	206.300	441	489-84-9-2d	198.330
342	295-65-8-2d	224.480	392	36888-18-3-2d	208.320	442	49558-02-3-2d	145.090
343	29585-45-3-2d	326.160	393	3697-27-6-2d	256.360	443	495-69-2-2d	179.190
344	295-97-6-2d	238.510	394	3741-24-0-2d	196.180	444	49570-30-1-2d	138.240
345	29720-92-1-2d	139.170	395	37789-64-3-2d	239.300	445	4957-16-8-2d	252.430
346	297-35-8-2d	420.900	396	38082-89-2-2d	621.340	446	496-46-8-2d	142.140
347	3010-82-0-2d	164.180	397	3815-86-9-2d	132.150	447	49753-54-0-2d	714.380

448	498-21-5-2d	132.130	498	538-81-8-2d	206.300	548	572-45-2-2d	240.280
449	498-24-8-2d	130.110	499	539-47-9-2d	138.130	549	573-56-8-2d	184.120
450	498-66-8-2d	94.170	500	539-71-9-2d	204.260	550	57-48-7-2d	180.180
451	501-65-5-2d	178.240	501	541-35-5-2d	87.140	551	57-50-1-2d	342.340
452	50296-60-1-2d	197.210	502	5415-99-6-2d	131.200	552	5769-11-9-2d	217.340
453	5037-75-2-2d	204.250	503	54166-22-2-2d	112.200	553	57694-62-9-2d	146.160
454	505-48-6-2d	174.220	504	5422-44-6-2d	99.120	554	5794-13-8-2d	132.140
455	505-52-2-2d	244.370	505	5422-45-7-2d	113.150	555	5796-86-1-2d	300.540
456	505-71-5-2d	150.120	506	5422-89-9-2d	255.300	556	579-75-9-2d	152.160
457	506-12-7-2d	270.510	507	542-44-9-2d	330.570	557	57-97-6-2d	256.360
458	506-30-9-2d	312.600	508	5442-92-2-2d	346.470	558	5807-76-1-2d	206.300
459	506-33-2-2d	338.640	509	544-63-8-2d	228.420	559	581-40-8-2d	156.240
460	50-69-1-2d	150.150	510	5448-17-9-2d	211.240	560	581-42-0-2d	156.240
461	50-78-2-2d	180.170	511	544-85-4-2d	450.980	561	581-96-4-2d	186.220
462	50-81-7-2d	176.140	512	5467-78-7-2d	161.190	562	582-16-1-2d	156.240
463	50853-29-7-2d	310.280	513	54798-92-4-2d	190.180	563	5831-43-6-2d	388.580
464	50892-90-5-2d	420.230	514	54986-63-9-2d	256.360	564	5840-76-6-2d	115.100
465	51-28-5-2d	184.120	515	55106-96-2-2d	647.340	565	5845-66-9-2d	220.250
466	5144-11-6-2d	98.130	516	55-21-0-2d	121.150	566	58505-78-5-2d	711.410
467	51449-77-5-2d	162.170	517	552-16-9-2d	167.130	567	58505-80-9-2d	902.450
468	51460-84-5-2d	635.310	518	553-12-8-2d	562.720	568	585-84-2-2d	174.120
469	517-25-9-2d	151.050	519	553-90-2-2d	118.100	569	586-38-9-2d	152.160
470	5174-83-4-2d	508.720	520	554-95-0-2d	210.150	570	58-72-0-2d	256.360
471	517-51-1-2d	532.700	521	55644-07-0-2d	120.080	571	588-68-1-2d	208.280
472	517-60-2-2d	342.180	522	556-50-3-2d	132.140	572	589-68-4-2d	302.510
473	519-44-8-2d	200.120	523	556-88-7-2d	104.090	573	5910-19-0-2d	197.170
474	519-73-3-2d	244.350	524	557-30-2-2d	88.080	574	593-45-3-2d	254.560
475	519-87-9-2d	211.280	525	55850-41-4-2d	188.200	575	594-70-7-2d	103.140
476	520-69-4-2d	137.250	526	56092-91-2-2d	204.120	576	594-82-1-2d	114.260
477	52073-04-8-2d	178.260	527	56-12-2-2d	103.140	577	595-49-3-2d	134.110
478	522-34-9-2d	240.280	528	5625-41-2-2d	174.180	578	595-86-8-2d	165.080
479	52459-90-2-2d	153.200	529	5625-50-3-2d	212.330	579	5965-66-2-2d	342.340
480	526-87-4-2d	146.160	530	5638-92-6-2d	163.200	580	59-67-6-2d	123.120
481	5271-39-6-2d	258.380	531	56-40-6-2d	75.080	581	597-09-1-2d	149.170
482	528-29-0-2d	168.120	532	56-41-7-2d	89.110	582	597-43-3-2d	146.160
483	528-44-9-2d	210.150	533	56-45-1-2d	105.110	583	59-87-0-2d	198.160
484	53010-03-0-2d	99.120	534	56-55-3-2d	228.300	584	60-00-4-2d	292.280
485	53130-19-1-2d	136.260	535	56818-06-5-2d	272.410	585	6004-38-2-2d	136.260
486	531-52-2-2d	300.390	536	56-84-8-2d	133.120	586	60-18-4-2d	181.210
487	5324-72-1-2d	207.170	537	56-85-9-2d	146.170	587	602-03-9-2d	183.140
488	5336-95-8-2d	144.100	538	56-86-0-2d	147.150	588	602-38-0-2d	218.180
489	5343-02-2-2d	232.220	539	5692-97-7-2d	174.220	589	60-27-5-2d	113.140

490	5343-98-6-2d	159.200	540	569-41-5-2d	156.240	590	602-99-3-2d	243.150
491	5376-15-8-2d	100.100	541	56941-72-1-2d	338.340	591	60302-02-5-2d	338.430
492	5378-52-9-2d	146.170	542	569-51-7-2d	210.150	592	603-02-1-2d	196.180
493	53816-93-6-2d	188.200	543	57-00-1-2d	131.160	593	60-32-2-2d	131.200
494	538-39-6-2d	210.340	544	57-10-3-2d	256.480	594	603-52-1-2d	241.310
495	538-51-2-2d	181.250	545	57-11-4-2d	284.540	595	603-54-3-2d	212.270
496	5385-36-4-2d	208.320	546	57-13-6-2d	60.070	596	60-35-5-2d	59.080
497	538-56-7-2d	278.320	547	572-43-0-2d	240.280	597	604-83-1-2d	206.300

598	605-60-7-2d	173.180	648	62155-31-1-2d	163.200	698	66591-60-4-2d	127.140
599	605-69-6-2d	234.180	649	621-59-0-2d	152.160	699	6682-67-3-2d	188.340
600	605-71-0-2d	218.180	650	621-61-4-2d	358.630	700	66907-70-8-2d	165.230
601	606-20-2-2d	182.150	651	62209-60-3-2d	356.190	701	670-54-2-2d	128.100
602	606-22-4-2d	183.140	652	62-23-7-2d	167.130	702	67242-61-9-2d	214.290
603	606-35-9-2d	243.150	653	623-26-7-2d	128.140	703	673-66-5-2d	127.210
604	606-87-1-2d	259.370	654	6237-59-8-2d	426.360	704	67-43-6-2d	393.400
605	607-86-3-2d	254.300	655	624-40-8-2d	116.100	705	675-20-7-2d	99.150
606	608-39-9-2d	146.160	656	6250-80-2-2d	151.230	706	676-22-2-2d	162.300
607	608-40-2-2d	146.160	657	626-17-5-2d	128.140	707	6869-07-4-2d	166.260
608	608-66-2-2d	182.200	658	6263-74-7-2d	374.160	708	6893-26-1-2d	147.150
609	609-36-9-2d	115.150	659	626-64-2-2d	95.110	709	68-94-0-2d	136.130
610	609-39-2-2d	113.080	660	626-97-1-2d	101.170	710	69155-29-9-2d	210.260
611	610-09-3-2d	172.200	661	627-12-3-2d	103.140	711	693-23-2-2d	230.340
612	610-41-3-2d	183.140	662	62786-89-4-2d	624.800	712	6957-81-9-2d	376.440
613	610-54-8-2d	212.180	663	628-02-4-2d	115.200	713	69-65-8-2d	182.200
614	611-48-3-2d	268.370	664	6280-33-7-2d	165.230	714	69-72-7-2d	138.130
615	611-72-3-2d	152.160	665	628-47-7-2d	131.200	715	6973-58-6-2d	263.310
616	611-92-7-2d	240.330	666	6285-54-7-2d	346.470	716	698-63-5-2d	141.090
617	6123-87-1-2d	256.330	667	629-01-6-2d	143.260	717	69-89-6-2d	152.130
618	6123-88-2-2d	450.580	668	630-51-3-2d	174.220	718	69-93-2-2d	168.130
619	612-71-5-2d	306.420	669	630-72-8-2d	176.060	719	700-12-9-2d	148.270
620	6130-98-9-2d	162.300	670	630-76-2-2d	320.450	720	700-56-1-2d	150.290
621	613-31-0-2d	180.260	671	632-14-4-2d	102.160	721	701-51-9-2d	156.110
622	613-33-2-2d	182.280	672	632-50-8-2d	334.480	722	702-79-4-2d	164.320
623	613-35-4-2d	268.340	673	632-51-9-2d	332.460	723	70288-75-4-2d	237.280
624	613-75-2-2d	198.190	674	632-52-0-2d	336.460	724	704-02-9-2d	158.260
625	613-90-1-2d	131.140	675	632-92-8-2d	241.180	725	70-47-3-2d	132.140
626	614-16-4-2d	145.170	676	6343-21-1-2d	194.210	726	70-54-2-2d	146.220
627	6154-04-7-2d	99.120	677	634-63-9-2d	148.140	727	705-60-2-2d	163.190
628	615-82-7-2d	188.260	678	635-10-9-2d	310.280	728	707-35-7-2d	178.350
629	6158-77-6-2d	127.130	679	635-51-8-2d	194.200	729	7074-00-2-2d	256.320
630	616-03-5-2d	114.120	680	636-48-6-2d	146.160	730	709-50-2-2d	194.210
631	616-06-8-2d	131.200	681	636-61-3-2d	134.100	731	71-00-1-2d	155.180
632	616-07-9-2d	132.190	682	638-53-9-2d	214.390	732	71217-51-1-2d	638.830
633	616-72-8-2d	196.180	683	63-91-2-2d	165.210	733	71-30-7-2d	111.120
634	616-74-0-2d	200.120	684	644-08-6-2d	168.250	734	7144-15-2-2d	376.440
635	6169-94-4-2d	278.470	685	645-12-5-2d	157.090	735	71742-94-4-2d	402.180
636	61-82-5-2d	84.100	686	6453-67-4-2d	548.740	736	72-18-4-2d	117.170
637	618-87-1-2d	183.140	687	645-92-1-2d	127.130	737	72-19-5-2d	119.140
638	61-90-5-2d	131.200	688	645-93-2-2d	128.110	738	725-12-2-2d	222.260

639	619-18-1-2d	183.140	689	646-30-0-2d	298.570	739	72762-00-6-2d	95.110
640	619-93-2-2d	270.260	690	653-49-6-2d	243.160	740	7311-29-7-2d	272.480
641	62030-33-5-2d	318.180	691	65-71-4-2d	126.130	741	73-22-3-2d	204.250
642	62030-34-6-2d	273.180	692	660-88-8-2d	117.170	742	73-24-5-2d	135.150
643	62030-36-8-2d	195.240	693	6620-95-7-2d	192.200	743	73-32-5-2d	131.200
644	62030-37-9-2d	435.260	694	66-22-8-2d	112.100	744	73-40-5-2d	151.150
645	620-40-6-2d	287.430	695	6624-53-9-2d	225.260	745	7343-06-8-2d	234.360
646	620-84-8-2d	183.270	696	6626-64-8-2d	231.310	746	736-31-2-2d	270.260
647	62154-79-4-2d	130.090	697	6634-01-1-2d	366.400	747	7396-38-5-2d	234.360

748	7409-48-5-2d	130.220	798	86-57-7-2d	173.180	848	97-65-4-2d	130.110
749	74418-40-9-2d	95.080	799	86-74-8-2d	167.220	849	97-67-6-2d	134.100
750	7477-73-8-2d	222.270	800	86-87-3-2d	186.220	850	98-73-7-2d	178.250
751	74999-19-2-2d	168.160	801	87123-46-4-2d	201.260	851	99-03-6-2d	135.180
752	74999-20-5-2d	196.200	802	87-79-6-2d	180.180	852	99-05-8-2d	137.150
753	74999-22-7-2d	125.160	803	87-85-4-2d	162.300	853	99-09-2-2d	138.140
754	74999-25-0-2d	222.260	804	87-99-0-2d	152.170	854	99-34-3-2d	212.130
755	74999-26-1-2d	125.160	805	886-66-8-2d	202.260	855	99-35-4-2d	213.120
756	74999-34-1-2d	269.330	806	88-74-4-2d	138.140	856	99-55-8-2d	152.170
757	74999-36-3-2d	153.200	807	888-71-1-2d	222.260	857	99-65-0-2d	168.120
758	75125-45-0-2d	282.370	808	88-89-1-2d	229.120	858	996-98-5-2d	118.120
759	75534-73-5-2d	280.400	809	88-99-3-2d	166.140	859	997-76-2-2d	103.090
760	7570-26-5-2d	120.080	810	89-05-4-2d	254.160	860	99-92-3-2d	135.180
761	7584-72-7-2d	270.300	811	90-41-5-2d	169.240	861	99-99-0-2d	137.150
762	75946-40-6-2d	264.360	812	90-45-9-2d	194.250	/	/	/
763	762-98-1-2d	162.170	813	91-15-6-2d	128.140	/	/	/
764	76-39-1-2d	119.140	814	91-20-3-2d	128.180	/	/	/
765	764-42-1-2d	78.080	815	91-55-4-2d	145.220	/	/	/
766	768-91-2-2d	150.290	816	91-56-5-2d	147.140	/	/	/
767	77-49-6-2d	135.140	817	91-57-6-2d	142.210	/	/	/
768	77-71-4-2d	128.150	818	918-37-6-2d	300.070	/	/	/
769	77-85-0-2d	120.170	819	918-54-7-2d	181.080	/	/	/
770	77-86-1-2d	121.160	820	92-24-0-2d	228.300	/	/	/
771	77-92-9-2d	192.140	821	92-52-4-2d	154.220	/	/	/
772	77-99-6-2d	134.200	822	92-55-7-2d	243.190	/	/	/
773	78-11-5-2d	316.170	823	92-62-6-2d	209.270	/	/	/
774	78-67-1-2d	164.240	824	92-67-1-2d	169.240	/	/	/
775	787-84-8-2d	240.280	825	92-70-6-2d	188.190	/	/	/
776	78956-69-1-2d	223.300	826	92-82-0-2d	180.220	/	/	/
777	79-05-0-2d	73.110	827	92-87-5-2d	184.260	/	/	/
778	793-25-9-2d	268.340	828	929-17-9-2d	145.230	/	/	/
779	79-33-4-2d	90.090	829	92-93-3-2d	199.220	/	/	/
780	796-97-4-2d	302.600	830	931-64-6-2d	108.200	/	/	/
781	79-92-5-2d	136.260	831	93-44-7-2d	248.290	/	/	/
782	80-05-7-2d	228.310	832	93-59-4-2d	138.130	/	/	/
783	80308-88-9-2d	448.210	833	936-12-9-2d	153.200	/	/	/
784	80308-90-3-2d	596.310	834	938-91-0-2d	164.230	/	/	/
785	806-71-3-2d	358.500	835	93-99-2-2d	198.230	/	/	/
786	80-68-2-2d	119.140	836	94-09-7-2d	165.210	/	/	/
787	826-36-8-2d	155.270	837	94-67-7-2d	137.150	/	/	/

788	828-51-3-2d	180.270	838	947-84-2-2d	198.230	/	/	/
789	830-81-9-2d	186.220	839	95-14-7-2d	119.140	/	/	/
790	83-32-9-2d	154.220	840	952-80-7-2d	210.340	/	/	/
791	84-61-7-2d	330.460	841	95-45-4-2d	116.140	/	/	/
792	84-62-8-2d	318.340	842	95-93-2-2d	134.240	/	/	/
793	85-01-8-2d	178.240	843	961-68-2-2d	259.240	/	/	/
794	85-02-9-2d	179.230	844	962-16-3-2d	258.240	/	/	/
795	85-15-4-2d	183.220	845	97-02-9-2d	183.140	/	/	/
796	85-98-3-2d	268.390	846	97-30-3-2d	194.210	/	/	/
797	86-30-6-2d	198.240	847	97-59-6-2d	158.140	/	/	/

Tableau 25 – Liste de Descripteurs d'anneau (en anglais)

N°	Name	Description
49	nCIC	number of rings (cyclomatic number)
50	nCIR	number of circuits
51	TRS	total ring size
52	Rperim	ring perimeter
53	Rbrid	ring bridge count
54	MCD	molecular cyclized degree
55	RFD	ring fusion density
56	RCI	ring complexity index
57	NRS	number of ring systems
58	NNRS	normalized number of ring systems
59	nR03	number of 3-membered rings
60	nR04	number of 4-membered rings
61	nR05	number of 5-membered rings
62	nR06	number of 6-membered rings
63	nR07	number of 7-membered rings
64	nR08	number of 8-membered rings
65	nR09	number of 9-membered rings
66	nR10	number of 10-membered rings
67	nR11	number of 11-membered rings
68	nR12	number of 12-membered rings
69	nBnz	number of benzene-like rings
70	ARR	aromatic ratio
71	D/Dtr03	distance/detour ring index of order 3
72	D/Dtr04	distance/detour ring index of order 4
73	D/Dtr05	distance/detour ring index of order 5
74	D/Dtr06	distance/detour ring index of order 6

N°	Name	Description
75	D/Dtr07	distance/detour ring index of order 7
76	D/Dtr08	distance/detour ring index of order 8
77	D/Dtr09	distance/detour ring index of order 9
78	D/Dtr10	distance/detour ring index of order 10
79	D/Dtr11	distance/detour ring index of order 11
80	D/Dtr12	distance/detour ring index of order 12

Tableau 26 – Liste de Descripteurs de type P_VSA (en anglais)

N°	Name	Description
1209	P_VSA_LogP_1	P_VSA-like on LogP, bin 1
1210	P_VSA_LogP_2	P_VSA-like on LogP, bin 2
1211	P_VSA_LogP_3	P_VSA-like on LogP, bin 3
1212	P_VSA_LogP_4	P_VSA-like on LogP, bin 4
1213	P_VSA_LogP_5	P_VSA-like on LogP, bin 5
1214	P_VSA_LogP_6	P_VSA-like on LogP, bin 6
1215	P_VSA_LogP_7	P_VSA-like on LogP, bin 7
1216	P_VSA_LogP_8	P_VSA-like on LogP, bin 8
1217	P_VSA_MR_1	P_VSA-like on Molar Refractivity, bin 1
1218	P_VSA_MR_2	P_VSA-like on Molar Refractivity, bin 2
1219	P_VSA_MR_3	P_VSA-like on Molar Refractivity, bin 3
1220	P_VSA_MR_4	P_VSA-like on Molar Refractivity, bin 4
1221	P_VSA_MR_5	P_VSA-like on Molar Refractivity, bin 5
1222	P_VSA_MR_6	P_VSA-like on Molar Refractivity, bin 6
1223	P_VSA_MR_7	P_VSA-like on Molar Refractivity, bin 7
1224	P_VSA_MR_8	P_VSA-like on Molar Refractivity, bin 8

N°	Name	Description
1225	P_VSA_m_1	P_VSA-like on mass, bin 1
1226	P_VSA_m_2	P_VSA-like on mass, bin 2
1227	P_VSA_m_3	P_VSA-like on mass, bin 3
1228	P_VSA_m_4	P_VSA-like on mass, bin 4
1229	P_VSA_m_5	P_VSA-like on mass, bin 5
1230	P_VSA_v_1	P_VSA-like on van der Waals volume, bin 1
1231	P_VSA_v_2	P_VSA-like on van der Waals volume, bin 2
1232	P_VSA_v_3	P_VSA-like on van der Waals volume, bin 3
1233	P_VSA_v_4	P_VSA-like on van der Waals volume, bin 4
1234	P_VSA_e_1	P_VSA-like on Sanderson electronegativity, bin 1
1235	P_VSA_e_2	P_VSA-like on Sanderson electronegativity, bin 2
1236	P_VSA_e_3	P_VSA-like on Sanderson electronegativity, bin 3
1237	P_VSA_e_4	P_VSA-like on Sanderson electronegativity, bin 4
1238	P_VSA_e_5	P_VSA-like on Sanderson electronegativity, bin 5
1239	P_VSA_e_6	P_VSA-like on Sanderson electronegativity, bin 6
1240	P_VSA_p_1	P_VSA-like on polarizability, bin 1
1241	P_VSA_p_2	P_VSA-like on polarizability, bin 2
1242	P_VSA_p_3	P_VSA-like on polarizability, bin 3
1243	P_VSA_p_4	P_VSA-like on polarizability, bin 4
1244	P_VSA_i_1	P_VSA-like on ionization potential, bin 1
1245	P_VSA_i_2	P_VSA-like on ionization potential, bin 2
1246	P_VSA_i_3	P_VSA-like on ionization potential, bin 3
1247	P_VSA_i_4	P_VSA-like on ionization potential, bin 4
1248	P_VSA_s_1	P_VSA-like on I-state, bin 1
1249	P_VSA_s_2	P_VSA-like on I-state, bin 2
1250	P_VSA_s_3	P_VSA-like on I-state, bin 3

N°	Name	Description
1251	P_VSA_s_4	P_VSA-like on I-state, bin 4
1252	P_VSA_s_5	P_VSA-like on I-state, bin 5
1253	P_VSA_s_6	P_VSA-like on I-state, bin 6
1254	P_VSA_ppp_L	P_VSA-like on potential pharmacophore points, L – lipophilic
1255	P_VSA_ppp_P	P_VSA-like on potential pharmacophore points, P – positive
1256	P_VSA_ppp_N	P_VSA-like on potential pharmacophore points, N – negative
1257	P_VSA_ppp_D	P_VSA-like on potential pharmacophore points, D – hydrogen-bond donor
1258	P_VSA_ppp_A	P_VSA-like on potential pharmacophore points, A – hydrogen-bond acceptor
1259	P_VSA_ppp_ar	P_VSA-like on potential pharmacophore points, ar – aromatic atoms
1260	P_VSA_ppp_con	P_VSA-like on potential pharmacophore points, con – conjugated atoms
1261	P_VSA_ppp_hal	P_VSA-like on potential pharmacophore points, hal – halogen atoms
1262	P_VSA_ppp_cyc	P_VSA-like on potential pharmacophore points, cyc – atoms belonging to cycles
1263	P_VSA_ppp_ter	P_VSA-like on potential pharmacophore points, ter – terminal atoms