

الجمهورية الجزائرية الديمقراطية الشعبية

وزارة التعليم العالي والبحث العلمي

UNIVERSITE BADJI MOKHTAR - ANNABA
BADJI MOKHTAR – ANNABA UNIVERSITY



جامعة باجي مختار – عنابة

Faculté : TECHNOLOGIE

Département : GENIE DES PROCEDES

Domaine : SCIENCES ET TECHNOLOGIES

Filière : GENIE DES PROCEDES

Spécialité : GENIE CHIMIQUE

Mémoire

Présenté en vue de l'obtention du Diplôme de Master

Thème:

**Élaboration des Relations Quantitatives Structure-Activité
(QSAR) pour l'estimation de l'adsorption des polluants
organiques sur le charbon actif**

Présenté par : *Oughlici Razan et Gasmi Islem*

Encadrant : *Dr Kherouf soumaya* Grade: MAB Université : UBMA

Jury de Soutenance :

Dr Lahmar Amina	MAB	UBMA	Président
Dr Kherouf Soumaya	MAB	UBMA	Encadrant
Pr BECHIRI Ouahiba	Pr	UBMA	Examineur

Année Universitaire : 2023/2024

Dédicace

Je dédie ce travail à :

- ❖ *A mon cher **PERE** : Je vous remercie pour tes sacrifices et pour votre présence et soutien tout au long de ma vie dans les meilleurs moments comme dans les mauvais pour devenir ce que je suis, vous étiez ma source d'inspiration.*
- ❖ *A ma chère **MERE** : Tous les mots ne sauraient exprimer la gratitude, l'amour, Le respect et la reconnaissance que J'ai pour vous. Votre amour et votre patience m'ont accompagné chacun de mes pas, merci d'être toujours à mes cotés pour m'aider et m'encourager, merci pour tout maman. ♥*
- ❖ *A mes chers frères : **ANIS&ACIL**, que dieu vous protège pour nous.*
- ❖ *A toute ma famille petite et grande.*
- ❖ *A tous mes amis, mes camarades, et toutes les personnes qui m'aime et me souhaite le bonheur de la vie.*

ISLAM

Dédicace

Je dédie ce modeste travail ...

A mes très chers parents Hannachi Amel et Oughlici Salim.

Aucune dédicace aussi parfaite et douce soit-elle, ne saurait exprimer toute ma reconnaissance et tout l'amour que je vous porte. Ce travail représente le fruit de votre soutien, vos sacrifices, et vos encouragements.

Que Dieu vous protège et vous accorde une longue vie pleine de santé et de bonheur!

A mes très chères sœurs Ritej et Rihem pour leur soutien et leurs encouragements tout au long de la réalisation de ce travail. Merci d'être toujours là pour moi, merci de m'avoir supportée.

A Mes chères amies : Ferial, Khadija et Ines pour notre amitié et tous les bons moments passés .Pour votre présence, et vos bons conseils.

*Last but not least, I want to thank me
I want to thank me for believing in me
I want to thank me for doing all this hard work
I want to thank me for having no days off
I want to thank me for, for never quitting
I want to thank me for always being a giver
And trying to give more than I receive
I want to thank me for trying to do more right than wrong
I want to thank me for just being me at all times*

Razan...s

Remerciement

Notre premier remerciement à " ALLAH " le tout puissant de nous avoir offert la force pour accomplir ce travail "Dieu merci".

Nous tenons à remercier vivement notre encadreur Professeur Kherouf Soumaya maitre pour sa patience, sa disponibilité, sa gentillesse

A toute l'équipe pédagogique de département génie des procédés ;

Aux membres du jury qui s'apprêtent à évaluer notre travail ;

A nos familles et amis qui nous ont tellement soutenus ;

A toutes les personnes qui ont contribué de près ou de loin à l'élaboration de ce mémoire ;

Nous exprimons nos forts remerciements

_ MERCI _

Résumé

Les relations quantitatives structure-activité / propriété (QSAR/QSPR) sont très utiles pour comprendre comment la structure chimique se corrèle avec la propriété des produits chimiques naturels et synthétiques. Dans le présent travail, des modèles QSPR ont été développés pour la prédiction d'adsorption d'une série des composés organiques.

Des techniques de régression linéaire multiple (MLR) ont été utilisées pour les études quantitatives de la relation structure propriété (QSPR) de 56 composés organiques basés sur des descripteurs moléculaires calculés.

Nous avons obtenu un coefficient de détermination (R^2) de 0.8649, la valeur de Q^2_{LOO} du modèle est 0.808, Le résultat de la validation externe du modèle construit a montré une grande capacité prédictive, car les valeurs de Q^2_{ext} est 82.30.

Abstract

Quantitative structure-activity/property relationships (QSAR/QSPR) are very useful for understanding how chemical structure correlates with the property of natural and synthetic chemicals. In the present work, QSPR models were developed for the prediction of adsorption of a series of organic compounds.

Multiple linear regression (MLR) techniques were used for quantitative structure property relationship (QSPR) studies of 56 organic compounds based on calculated molecular descriptors.

We obtained a coefficient of determination (R^2) of 0.8649, the value of Q^2_{LOO} of the model is 0.808, The result of external validation of the built model showed a high predictive ability, because the values of Q^2_{ext} is 82.30.

المخلص

تعتبر العلاقات الكمية بين البنية والنشاط/الخاصية (QSAR/QSPR) مفيدة جدًا لفهم كيفية ارتباط البنية الكيميائية بخواص المنتجات الكيميائية الطبيعية والصناعية. في العمل الحالي، تم تطوير نماذج QSPR للتنبؤ بالامتزاز لسلسلة من المركبات العضوية.

تم استخدام تقنيات الانحدار الخطي المتعدد (MLR) لدراسات العلاقة الكمية بين البنية والخاصية (QSPR) لـ 56 مركبًا عضوياً بناءً على الواصفات الجزيئية المحسوبة.

حصلنا على معامل التحديد (R^2) بقيمة 0.8649، وقيمة Q^2_{LOO} للنموذج هي 0.808. أظهر نتيجة التحقق الخارجي للنموذج المبني قدرة تنبؤية كبيرة، حيث كانت قيمة Q^2_{ext} هي 82.30.

Liste des tableaux

Tableau 1 Type d'isotherme.....	19
Tableau 2 Noms, valeurs expérimentale et valeurs des descripteurs	41
Tableau 3 Matrice de corrélation pour les quatre descripteurs sélectionnés dans le modèle. ..	43

Liste des figures

Figure 1 Les types d'adsorption	15
Figure 2 Domaines d'existence d'un soluté lors de l'adsorption sur un matériau microporeux.	17
Figure 3 Classification des isothermes d'adsorption selon Giles et al.....	18
Figure 4 Structure du charbon actif.....	20
Figure 5 CA en grain (CAG).....	20
Figure 6 CA en extrudé.	21
Figure 7 CA en poudre (CAP).....	21
Figure 8 Adsorption sur le charbon actif.....	22
Figure 9 Modèle de l'étude de relation structure activité/propriété.....	27
Figure 10 Partage des données expérimentales pour le développement d'un modèle	31
Figure 11 Les valeurs Prédites versus les valeurs expérimentales	43
Figure 12 Diagramme de Williams	44
Figure 13 Test de randomisation	45
Figure 14 Contribution (%) des descripteurs au modèle.	46

Symboles et Abréviations

OCDE : Organisation de coopération et de développement économiques

QSAR : Relations Quantitatives Structure-Activité

QSPR : Relations Quantitatives Structure-propriété

POA : Procédés d'oxydation avancés

CA : Charbon actif

CAG : Charbon actif en grain

OLS : moindres carrés ordinaires

CAP : Charbon actif en poudre

AG-SSV : La sélection de sous-ensembles de variables par algorithme génétique

QUIK: Q under Influence of K

DA : Domaine d'applicabilité

MLR : Régression linéaire multiple

ZM1V : Indice de Zagreb de premier ordre par degrés de sommets de valence

FDI : Comptage des Groupes Fonctionnels

VEv2 : Deuxième Valeur Propre de la Matrice d'Adjacence

RDF020v : Fonction de Distribution Radiale à 2,0 Å pondérée par les volumes de van der Waals

IUPAC: International Union for Pure and Applied Chemistry

Sommaire

Liste des tableaux	I
Liste des figures	II
Symboles et Abréviations	III
Introduction générale	11
Chapitre 1	13
Introduction	14
1 Méthodes de traitement des effluents :	14
1.1 Traitements biologiques.....	14
1.2 Traitements physiques	14
1.3 Traitement physico-chimique.....	14
1.4 Traitements chimiques.....	14
2 Généralités sur l'adsorption	15
2.1 Principe.....	15
2.2 Types d'adsorption	15
2.2.1 Adsorption physique.....	16
2.2.2 Adsorption chimique	16
2.3 Description du mécanisme d'adsorption	16
2.4 Facture influençant sur l'adsorption	17
2.5 Isothermes d'adsorption	18
2.6 Modélisations de l'adsorption.....	19
3 L'adsorption sur charbon actif.....	20
3.1 Généralités sur charbon actif	20
3.2 Origine	20
3.3 Structure du charbon actif.....	20
3.4 Classification des charbons actifs.....	21
3.5 Propriété physique du charbon actif.....	21
3.6 Le charbon actif et l'adsorption	22
3.7 Utilisation du charbon actif	22
Références	23

Chapitre 2.....	25
Introduction	26
1 Historique	26
2 Définition et principe.....	27
3 Méthodologie générale d'une étude QSPR	28
3.1 Base de données.....	28
3.2 Représentation des structures	29
3.3 Génération des descripteurs	29
3.4 Réduction des descripteurs	30
3.5 Choix de l'ensemble de calibrage et validation.....	30
3.6 Sélection des descripteurs.....	31
4 L'analyse des données par la méthode« Régression linéaire multiple (MLR) ».....	32
5 Validation du modèle	32
5.1 Validation interne du modèle.....	33
5.1.1 Validation croisée LOO (leave- one-out)	33
5.1.2 Validation par le test de randomisation	34
5.2 Validation externe	34
5.2.1 Domaine d'applicabilité.....	35
Conclusion	36
Références	37
Chapitre 3.....	39
Résultats et discussion	40
1 Base de données et calcul des descripteurs.....	40
2 Optimisation et calcul des descripteurs	40
3 Développement du modèle QSPR-MLR.....	40
3.1 Le modèle QSPR-MLR	41
3.2 Matrice de corrélation :.....	43
4 La qualité de l'ajustement :	43
5 Domaine d'application (Diagramme de Williams).....	44
6 Test de Randomisation :	45
7 Relatif importance des descripteurs	46
8 Interprétation Mécanistique du modèle développé.....	46
Conclusion	48
Référence	49
Conclusion générale	51

Introduction générale

Introduction générale

La pollution de l'eau est une dégradation physique, chimique, biologique ou bactériologique de ses qualités naturelles, provoquée par l'homme et ses activités. Elle perturbe les conditions de vie de la faune et la flore aquatiques, et l'équilibre du milieu naturel.

Les matières organiques ont longtemps été les principaux polluants des milieux aquatiques. Elles proviennent des déchets domestiques agricoles ou industriels, lorsque ceux-ci sont rejetés sans traitement préalable.

Pour ce faire, un vaste sujet d'études vise à réduire cette pollution à la source si possible ou dans les effluents si nécessaires avec des traitements curatifs appropriés

Le charbon actif grâce à son grand pouvoir adsorbant est un moyen très efficace pour éliminer les polluants organiques des eaux usées. Pour la sélection et la conception appropriées d'un système de traitement, la connaissance de l'adsorption de différents produits chimiques est souhaitable. Cependant, il est courant que les isothermes d'adsorption ne soient pas disponibles pour ces produits. Face à la diversité et au nombre considérable de polluants, Il serait bénéfique de disposer d'une méthode permettant de prédire l'adsorption des matières organiques sans expérimentation.

Des relations quantitatives structure-activité (QSAR) ont été développées pour la première fois pour prédire la toxicité et le devenir des produits chimiques dans l'environnement. Désormais, ils sont également utilisés pour corréler les propriétés d'adsorption et les caractéristiques moléculaires des produits chimiques (QSPR).

Des principes ont même été mis en place par l'OCDE en 2009 afin de valider scientifiquement et réglementairement les modèles QSAR/QSPR. Cela dans le but d'augmenter la confiance en ces modèles mathématiques prédictifs et de favoriser ainsi leur utilisation par les industriels

Nous nous proposons, dans ce travail de développer et d'évaluer le potentiel de tels modèles QSPR pour la prédiction : d'adsorption d'une série des composés organiques

Notre étude est constituée de deux parties : une théorique et l'autre expérimentale ou vous continuer avec l'idée des chapitres

Dans la première partie de ce mémoire, nous avons donné une synthèse bibliographique qui rassemble des données essentielles sur :

- Généralités sur la pollution des eaux usées ;
- Généralités sur l'adsorption ; Les études QSAR/QSPR

Introduction générale

La deuxième partie porte sur

- Application sur une série des composés organiques pour la prédiction de leur adsorption sur charbon actif
- Et finalement une Conclusion générale.

Chapitre 1

Synthèse bibliographique

**Sur la dépollution par adsorption sur
charbon actif**

Introduction

Les effluents des activités industrielles et agricoles ainsi que les effluents ménagers (eaux usées urbaines) doivent être traités avant leur injection dans le cycle naturel de l'eau. Actuellement, la moitié des eaux usées est introduite dans le milieu naturel sans aucune épuration.

D'où la nécessité de développer des procédés de traitement efficaces, économiques, portables et de taille adaptable aux diverses applications. Il faut noter que la dépollution n'est jamais totale et coûte très cher aux collectivités et aux consommateurs [1].

1 Méthodes de traitement des effluents :

Différents procédés sont utilisés dans le traitement des eaux usées : procédés biologiques, physiques, physico-chimiques, chimiques

1.1 Traitements biologiques

Les traitements biologiques des effluents aqueux utilisent des micro-organismes (microflore de bactéries, microfaune d'animaux, protozoaires, métazoaires ...) pour éliminer les polluants biodégradables dans l'eau.

1.2 Traitements physiques

Les traitements physiques sont appliqués pour une simple séparation mécanique solide-liquide, sans introduction de réactif chimique. Il s'agit de la décantation et de la filtration.

1.3 Traitement physico-chimique

Ces technologies d'oxydation sont regroupées sous le terme de procédés d'oxydation avancés (POA). L'activation catalytique et/ou photochimique du processus oxydatif est utilisée dans le but d'obtenir une oxydation totale et une minéralisation complète des polluants organiques.

1.4 Traitements chimiques

Dans le cas des polluants non –biodégradables, il est indispensable de trouver d'autres méthodes de traitement comme le traitement chimique qui consiste à extraire les polluants par exemple par **adsorption** sur des matériaux minéraux.

L'**adsorption**, est l'un des procédés les plus utilisés dans le monde en vue d'élimination des polluants dans les réseaux des eaux usées. L'adsorption est un phénomène d'interface, pouvant se manifester entre un solide et un gaz ou un liquide. Le phénomène est général pour toutes les surfaces mais on cherche à l'exploiter en employant des solides à porosité élevée. Cependant, les charbons actifs commerciaux, principaux adsorbants utilisés en traitement des eaux, sont relativement chers et par conséquent ces matériaux sont peu accessibles à grande échelle.

2 Généralités sur l'adsorption

2.1 Principe

L'adsorption est le processus où des molécules d'une espèce appelée adsorbat (gaz ou liquide) viennent se fixer sur la surface d'un solide, appelé adsorbant. Ce phénomène dépend à la fois de cette surface et des propriétés physico-chimiques de l'adsorbat [2]. L'adsorption peut être physique ou chimique selon la nature des interactions qui se produisent entre l'adsorbat et la surface de l'adsorbant [3].

2.2 Types d'adsorption

La nature des liens entre l'adsorbant et la substance adsorbée, et la grandeur de l'énergie mise en jeu ont permis de distinguer deux types d'adsorption [4-5].

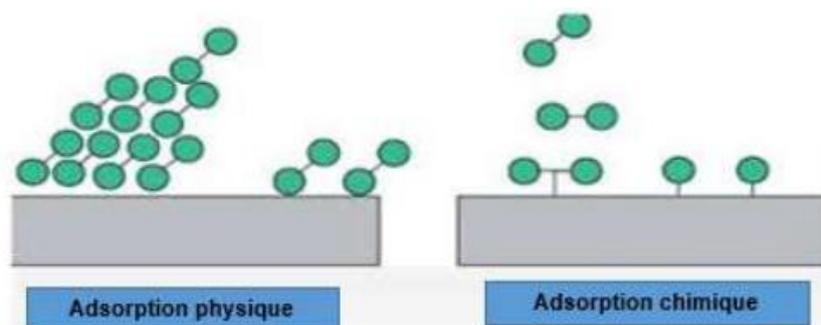


Figure 1 Les types d'adsorption

2.2.1 Adsorption physique

Dans la physio sorption, il n'y a pas de transfert d'électrons. Elle n'est pas spécifique et peut-être multicouche. L'adsorption physique est généralement réversible car l'énergie mise en cause est faible, (de l'ordre de quelques kilocalories par mole). Les forces impliquées sont des forces d'attraction de type VAN DER WALLS [6].

2.2.2 Adsorption chimique

Elle correspond à la création de véritables liaisons chimiques par le transfert électronique sont de type covalent, mais entre les molécules de l'interface des deux phases. Ces liaisons peuvent être légèrement polarisées

Les énergies mises en jeu sont plus grandes que dans le cas de physio sorption. Ce type d'adsorption spécifique n'apparaît le plus souvent qu'à des températures assez élevées et ne devient réversible qu'à des températures encore plus élevées [7].

2.3 Description du mécanisme d'adsorption

L'adsorption se produit principalement en quatre étapes. La figure 1 représente un matériau (adsorbant) avec les différents domaines dans lesquels peuvent se trouver les molécules organiques ou inorganiques qui sont susceptibles de rentrer en interaction avec le solide [8]. Avant son adsorption, le soluté va passer par plusieurs étapes :

- 1) Diffusion de l'adsorbat de la phase liquide externe vers celle située au voisinage de la surface de l'adsorbant.
- 2) Diffusion extra granulaire de la matière (transfert du soluté à travers le film liquide vers la surface des grains).
- 3) Transfert intra granulaire de la matière (transfert de la matière dans la structure poreuse de la surface extérieure des grains vers les sites actifs).
- 4) Réaction d'adsorption au contact des sites actifs, une fois adsorbée, la molécule est considérée comme immobile.

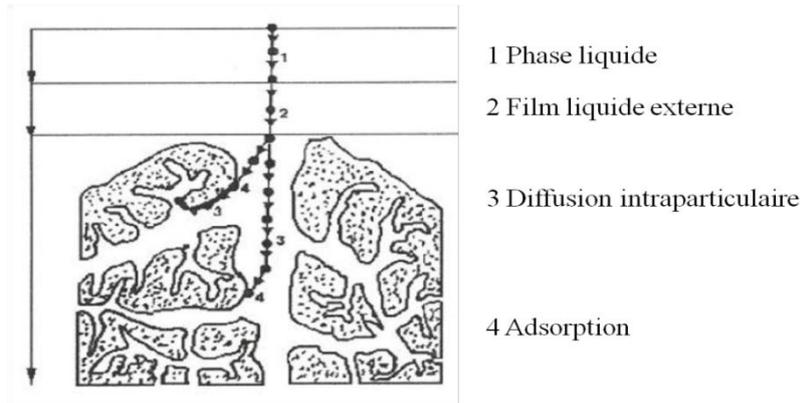


Figure 2 Domaines d'existence d'un soluté lors de l'adsorption sur un matériau microporeux [9].

2.4 Facture influençant sur l'adsorption

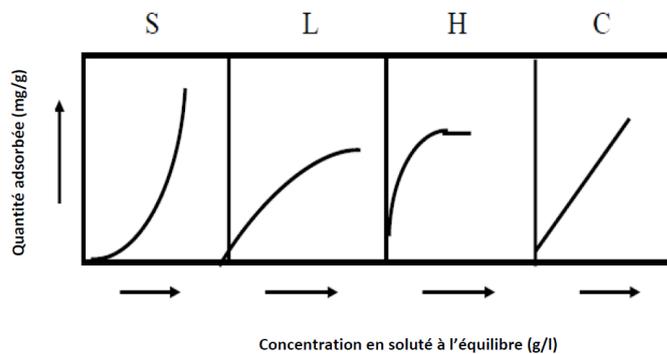
L'équilibre d'adsorption entre un adsorbant et un adsorbât dépend de nombreux facteurs dont voici les principaux et qui sont :

- 1. pH** : D'après la règle de Traube, l'optimisation des paramètres analytiques et particulièrement le pH augmente le rendement d'adsorption des polluants contenus dans les solutions aqueuses [10].
- 2. Température** : l'adsorption est un processus exothermique et par conséquent son déroulement doit être favorisé par un abaissement de température.
- 3. Concentration** : l'adsorption d'une substance croît avec l'augmentation de sa concentration dans la solution. Toutefois, cet accroissement n'est pas proportionnel car il se produit lentement.
- 4. Polarité** : Pour qu'il ait une bonne adsorption il faut qu'il ait une affinité entre le solide et le soluté. En règle générale, les solides polaires adsorbent préférentiellement d'autres corps polaires. Par contre les solides non polaires adsorbent préférentiellement des substances non polaires.
- 5. Dimensions des pores** : Si les dimensions des pores de l'adsorbant sont inférieures aux diamètres des molécules de l'adsorbât, l'adsorption de ce composé ne se fait pas même si la surface de l'adsorbant a une grande affinité pour ce composé [11].
- 6. Surface spécifique** : Plus la surface d'adsorption est grande, plus la quantité de matière adsorbable est importante [12].
- 7. La nature de l'adsorbât** : D'après la règle de Lundenius « Moins une substance est soluble dans le solvant, mieux elle est adsorbée » [10].

2.5 Isothermes d'adsorption

On peut décrire un processus d'adsorption à l'aide d'une isotherme d'adsorption. Une telle isotherme est une courbe qui représente la relation entre la quantité d'impuretés adsorbées par unité de masse de charbon activé et la concentration d'impuretés en solution. On obtient une telle courbe à partir des résultats d'essais de laboratoire effectués à une température constante. Expérimentalement, on distingue quatre classes principales nommées: S (Sigmoide), L (Langmuir), H (Haute affinité) et C (Partition Constante). La figure 3, illustre la forme de chaque type d'isotherme d'après la classification de Giles et al. [13].

Figure 3 Classification des isothermes d'adsorption selon Giles et al. [13]



- **Les courbes de type S** : s'obtiennent lorsque les molécules du soluté ne s'accrochent au solide, que par l'intermédiaire d'un seul groupement;
- **Les isothermes de type L** (dite Langmuir): sont les plus fréquentes. Ce comportement se rencontre dans le cas où l'adsorption est faible et lorsque les molécules de l'adsorbat sont orientées à plat;
- **Les isothermes de type H** (haute affinité) : s'obtiennent lorsqu'il y a chimie sorption du soluté;
- **Les isothermes de type C**: sont sous forme de ligne droite. Ce type de courbe est obtenu lorsqu'il y a compétition entre le solvant et le soluté pour occuper les sites de l'adsorbant [14].

2.6 Modélisations de l'adsorption

Plusieurs modèles ont été proposés pour l'étude de l'adsorption. Ces derniers expriment la relation entre la quantité adsorbée et la concentration en soluté dans un solvant à une température donnée. On exprime, ci-dessous, les principales lois, les plus utilisées, celle de **Langmuir** (1918), et celle de **Freundlich**(1962) [15].

Tableau 1 Type d'isotherme

Type d'isotherme	Equation	Description des termes
LANGMUIR	$\frac{C}{x/m} = \frac{1}{kb} + \frac{c}{b}$	C : concentration à l'équilibre x/m : quantité adsorbée/masse d'adsorbant b : maximum d'adsorption K : conste qui prend en compte l'énergie de liaison
FREUNDLICH	$\frac{X}{m} = kC^{1/n}$	C : concentration à l'équilibre x/m : quantité adsorbée/masse d'adsorbant K et n : conste qui prend en compte l'énergie de liaison

Les données d'adsorption de notre étude sont adaptées à un modèle de l'isotherme de Freundlich, qui utilise deux paramètres, K et n, pour modéliser l'adsorption

La valeur de 1/n donne une indication sur la validité de l'adsorption du système adsorbant-adsorbât.

Une valeur de 1/n comprise entre 0 et 1 indique une adsorption favorable [16].

Si n = 1 adsorption linéaire « la partition entre les deux phases est Indépendamment de la concentration », si n < 1 adsorption chimique et si n > 1 adsorption physique est favorable [17].

3 L'adsorption sur charbon actif

3.1 Généralités sur charbon actif

Le charbon actif ou encore charbon activé est un matériau constitué essentiellement de matière carbonée à structure poreuse, ayant subi une préparation particulière et qui, de ce fait, possède, à un haut degré la propriété de fixer et de retenir les fluides amenés à son contact. Il s'agit d'une structure amorphe composée principalement d'atomes de carbone et généralement obtenue après une étape de carbonisation à haute température, présentant une très grande surface spécifique qui lui confère un fort pouvoir adsorbant [18].

3.2 Origine

D'origine organique ou minérale, il provient de diverses matières animales et végétales (coque de noix de coco, tourbe, houille...), les charbons actifs issus de ces derniers produits sont les plus souvent employés pour le traitement de l'eau naturelle ou résiduaire. Pour des recherches spécifiques, on peut avoir recours à des produits carbonisés comme le sucre qui donne du charbon actif chimiquement pur [14].

3.3 Structure du charbon actif

Le **charbon actif** peut être défini comme une forme rudimentaire de graphite avec une structure aléatoire ou amorphe qui est extrêmement poreuse sur une large gamme de tailles de pores, de fissures et de crevasses visibles à l'œil nu à celles de dimensions moléculaires.

La purification au **charbon actif** se fonde principalement sur un phénomène appelé **adsorption**, dans lequel les molécules d'un liquide ou d'un gaz sont piégées par sa surface externe ou interne [19].

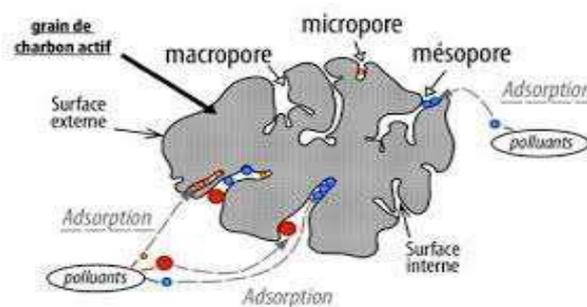


Figure 4 Structure du charbon actif.

3.4 Classification des charbons actifs

Il ya plusieurs type de charbon actif nous mentionnons trois types célèbres :

Charbon actif en grain (CAG)

Charbon actif en extrudé

Charbon actif en poudre (CAP)



Figure 5 CA en grain (CAG).



Figure 6 CA en extrudé.

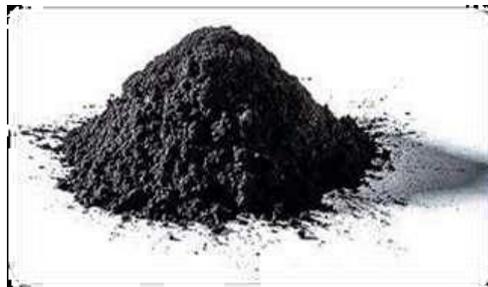


Figure 7 CA en poudre (CAP).

3.5 Propriété physique du charbon actif

La distribution poreuse ou porosité est liée à la répartition de la taille des pores. Ces pores sont caractérisés par leur volume poreux (cm^3/g), leur rayon ($^{\circ}\text{A}$) et leur surface (cm^2) [14].

3.6 Le charbon actif et l'adsorption

Le principe de l'adsorption repose sur l'accumulation à la surface ou à l'intérieur du charbon (dans les pores) des molécules contenues dans l'eau ou dans l'air à traiter. Ceci est possible grâce aux interactions physiques et chimiques entre la surface du charbon et les molécules à adsorber.

L'adsorption peut être décomposée en quatre étapes [20].

1. Transfert de la particule (phénomène très rapide);
2. Déplacement de l'eau jusqu'à être au contact du charbon (rapide);
3. Diffusion à l'intérieur du charbon sous l'influence du gradient de concentration (étape assez lente);
4. Adsorption dans un micropore (très rapide);

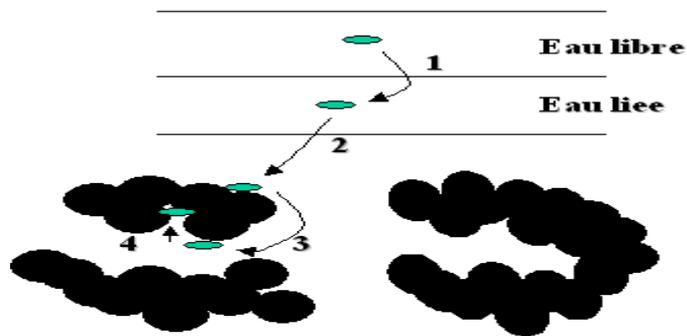


Figure 8 Adsorption sur le charbon actif [20]

3.7 Utilisation du charbon actif

L'utilisation du charbon actif est répartie en différents domaines comme

- Traitement des eaux potables et résiduaires.
- Purification de produits industriels.
- Le charbon actif est employé dans l'équipement d'air conditionné ainsi que pour l'élimination des odeurs d'égouts et d'émanations chimiques.
- Les filtres à charbon actif contenus dans certaines voitures, fixent les émanations d'hydrocarbures imbrûlés qui s'échappent des véhicules à l'arrêt.
- Ces dernières années, l'utilisation du charbon actif s'est élargie au traitement des eaux usées en raison non seulement de ses qualités d'adsorbant mais aussi du rôle épurateur joué par les bactéries qui y sont fixées. [21].

Références

- [1] <https://new.societechimiquedefrance.fr/wp-content/uploads/2019/12/2004-277-278-aout-sept-Oturan-p.57.pdf>
- [2] OUBAGHA, Noura. Décontamination des eaux contenant les colorants textiles et les adjuvants par des matériaux naturels et synthétique, 151p, Mémoire de Magister : Chimie de l'environnement : TiziOuzou, Université Mouloud Mammeri : 2011.
- [3] CARMEN, ManoleCreangă, Procédé Ad-Ox d'élimination de polluants organiques non biodégradables (par adsorption puis oxydation catalytique) [en ligne], Thèse de Doctorat : Génie des Procédés et de l'Environnement : Toulouse, Institut National Polytechnique : 2007, [Consultée le 15 mars 2017], Disponible à l'adresse :
- [4] Rachidi M. H. Etude de l'adsorption de l'Octaethylporphyrinede Nickel sur les apatites phosphocalciques. (1994).
- [5]Mechrafi E. Adsorption, desorption et mobilite des herbicides au contact des adsorbants organiques et minéraux (2002).
- [6] Desjardins R. Le traitement des eaux 2 ème édition revue (1990).
- [7]Mekaoui M. Etude de l'adsorption, de la persistance et de la photodegradation du (TBM) dans le sol et l'eau (2001).
- [8] M. Belmouden, Thèse de Doctorat, Faculté des Sciences d'Agadir, 2000.
- [9]F.Benamraoui, «Elimination des colorants cationiques par des charbons actifs synthétisés à partir des résidus de l'agriculture,». MAGISTER en Génie Chimique. Setif, 2014.
- [10] ABBAS, moussa. Valorisation du noyau d'abricot dans la dépollution des eaux, 205p, Thèse de Doctorat : Chimie des Matériaux : Boumerdes, Université M'hamedBougara : 2015. [11] KROU, N, Etude expérimentale et modélisation d'un procédé séquentiel AD-OX d'élimination de polluants organiques, Thèse de Doctorat : Génie des Procédés et de l'Environnement : Toulouse, Université de Toulouse : 2010.
- [11] MEKHALIF, Faiza, Réutilisation des eaux résiduaires industrielles épurées comme eau d'appoint dans un circuit de refroidissement, 158p, Mémoire de Magister : Chimie : Skikda, Université de 20 Aout 1955 : 2009

- [12] KROU, N, Etude expérimentale et modélisation d'un procédé séquentiel AD-OX d'élimination de polluants organiques, Thèse de Doctorat : Génie des Procédés et de l'Environnement : Toulouse, Université de Toulouse : 2010.
- [13] YAHIAOUI, Nouara, Etude de l'adsorption des composés phénoliques des margines d'olive sur carbonate de calcium, hydroxyapatite et charbon actif, 129p, Mémoire de Magister : Chimie de l'environnement : TiziOuzou, Université Mouloud Mammeri : 2012.
- [14]] H. Maiza, Modélisation des cinétiques d'adsorption dans le cas du phénol et du bleu de méthyle sur le charbon actif en grain, Ingénieur d'état, Ecole Nationale Supérieure Polytechnique, p 3,4 (2000)
- [15] Etude expérimentale des interactions point Argile naturelle / tensioactif cationique, master en Génie Chimique, Université EchahidHamma Lakhdar. El Oued
- [16] FAYOUD, N., ALAMI YOUNSSI, S., TAHIRI, S., ALBIZANE, A., Etude cinétique et thermodynamique de l'adsorption de bleu de méthylène sur les cendres de bois, J, Mater, Environ, Sci, Novembre, 2015, vol, 6, n° 11, P, 3295-3306.
- [17] DADA, A,O,, OLALEKAN, A,P,, OLATUNYA, A,M,, DADA, O, Langmuir, Freundlich, Temkin and Dubinin–Radushkevich Isotherms Studies of Equilibrium Sorption of Zn²⁺ Unto Phosphoric Acid Modified Rice Husk, IOSR Journal of Applied Chemistry, Dec, 2012, Vol, 3, Issue 1, P, 38-45.
- [18] SAKR, F , SENNAOUI, A : ELOUARDI, M., TAMIMI, M., ASSABBANE, M. Étude de l'adsorption du Bleu de Méthylène sur un biomatériau à base de Cactus, J, Mater, Environ, Sci, 2015, vol, 6, n° 2, p, 397-406.
- [19] «Chemical structure of activated carbon».
- [20] (Lu Jinyan, 2005) :
- [21] LOURGHI Lamia, *Adsorption du furfural issue des rejets liquides de la raffinerie d'Arzew par charbon actif en poudre*, Mémoire de Master, 2014.

Chapitre 2

QSAR /QSPR

Introduction

Malgré l'amélioration des équipements de laboratoire, l'analyse expérimentale de tous les produits récemment synthétisés est impossible. L'utilisation des méthodes alternatives à l'expérience est devenue plus qu'indispensable. Parmi ces méthodes, on trouve les méthodes de modélisation moléculaire qui permettent de justifier les données expérimentales disponibles et prédire les propriétés pour des composés nouveaux ou des composés pour lesquels les données expérimentales ne sont pas disponibles [1]. Ainsi, une approche QSPR fournit un compromis nécessaire à l'estimation des paramètres physicochimiques de grandes classes de composés qui sont importants du point de vue théorique ou pour les applications industrielles [2]. 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.

1 Historique

Il est communément admis que la QSAR moderne a vu le jour au début des années 1960. Cependant, dès 1816, les scientifiques faisaient des prédictions sur les propriétés physiques et chimiques. Les premières recherches sur la corrélation entre les activités biologiques et les propriétés physicochimiques telles que le poids moléculaire et la solubilité aqueuse ont débuté en 1841, près de 60 ans avant les travaux importants d'Overton et Meyer qui ont établi un lien entre la toxicité aquatique et le partage lipide-eau. Tout au long du XXe siècle, la QSAR a progressé, malgré de nombreuses années de vaches maigres. En 1962, les travaux fondamentaux de Corwin Hansch et de ses collaborateurs ont suscité un intérêt considérable pour la prédiction des activités biologiques. Au départ, cet intérêt concernait principalement la chimie médicinale et la conception de médicaments, mais dans les années 1970 et 1980, avec les préoccupations écotoxicologiques croissantes, la modélisation QSAR des toxicités environnementales a commencé à se développer, en particulier lorsque les autorités réglementaires ont commencé à s'en préoccuper. Depuis lors, les QSAR ont continué à se développer, avec plus de 1400 publications par an depuis 2011 [3].

2 Définition et principe

L'acronyme QS/XR [X= A (activité), P (propriété), R (indice de rétention), T (toxicité)...] est utilisé lorsqu'un effet spécifique est relié à la structure moléculaire.

En outre, lorsqu'on exprime des propriétés et des structures physicochimiques par des nombres, on peut proposer une relation mathématique ou une relation quantitative structure-propriété entre les deux. Pour notre cas il s'agit de relier l'adsorption sur du charbon actif à la structure moléculaire de quelques composés organiques dans l'eau.

Un modèle QSAR/QSPR a la forme d'une équation mathématique

$$\text{Activité/propriété} = f(x)$$

X = descripteur moléculaire

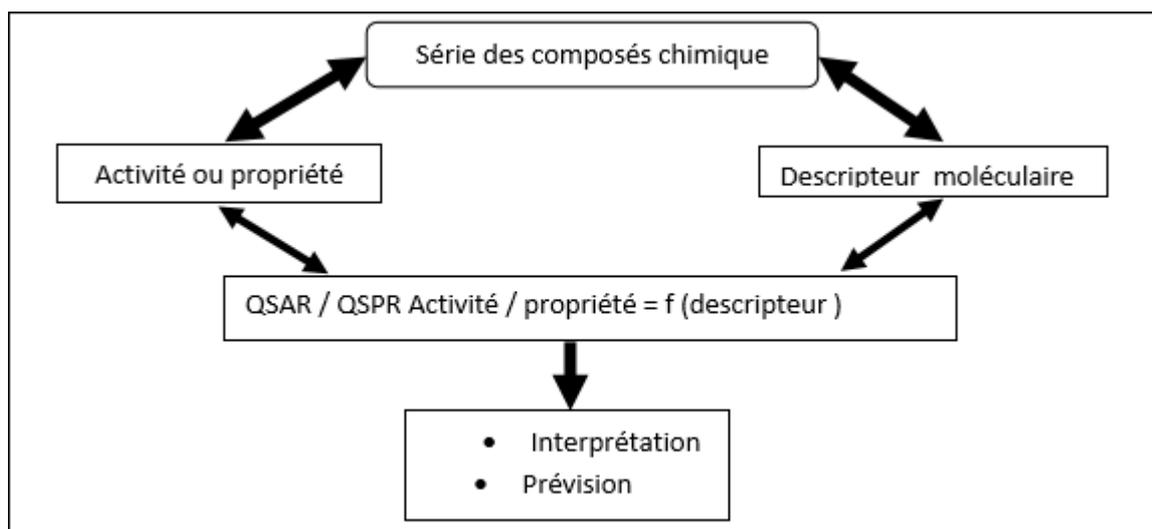


Figure 9 Modèle de l'étude de relation structure activité/propriété.

Le point de départ de telles méthodes se construit sur la définition des descripteurs.

Un descripteur moléculaire est toute propriété (une valeur numérique) propre à une structure chimique donnée. Ces valeurs peuvent être obtenues expérimentalement (descripteurs empiriques) ou calculées à partir de la structure de la molécule (théorique). Le lien entre les descripteurs et la propriété est déterminé grâce à des outils d'analyse comme les régressions multilinéaires (MLR), les arbres de décisions, les réseaux de neurones, les algorithmes génétiques....etc

Pour aboutir à un modèle QSAR explicatif et prédictif on doit :

1) Elaborer le modèle

- a- Collecter une base de données fiable
- b- Rechercher et calculer des descripteurs adéquats pour la propriété étudiée.
- c- Utiliser une méthode d'analyse des données pour relier la propriété aux descripteurs

2) Validation du modèle

- a- Validation interne : cette validation a pour but de vérifier la stabilité et le pouvoir explicatif du modèle QSPR construit.
- b- Validation externe : on utilise une série de test dans le but de vérifier le pouvoir prédictif du modèle élaboré.

3 Méthodologie générale d'une étude QSPR

3.1 Base de données

Une étape cruciale dans le processus de développement et de validation des modèles QSPR/QSAR est la sélection de la base de données. Il est essentiel d'opter pour des données de haute qualité, à savoir fiables et cohérentes, afin de réduire au maximum les erreurs expérimentales. De plus, il est primordial que ces données soient obtenues selon un protocole uniforme et dans des conditions expérimentales identiques, car ces facteurs exercent une influence significative sur les résultats et, par conséquent, sur la robustesse des modèles. [4] Il faut également que la distribution des données soit la plus homogène et normale, car la plupart des méthodes statistiques sont basées sur ce type de distribution [5].

Effectivement, la taille de la base de données revêt une importance capitale : une base de données trop restreinte rend la création de modèles prédictifs difficile, voire parfois impossible à valider. Il est également crucial de garantir l'exactitude des structures moléculaires, qui forment le fondement des modèles QSPR, leur représentation correcte étant essentielle. Ainsi, il est impératif de rester vigilant quant aux structures et de rectifier toute erreur potentielle, telles que les incohérences entre les numéros CAS et les noms de molécules, ou la présence de valeurs de propriétés différentes pour une même substance. Une seule erreur à ce niveau peut nécessiter une refonte complète des modèles et exercer un impact considérable sur leur capacité prédictive si elle n'est pas décelée. Il est également crucial de vérifier l'absence de redondance [6]. et de

données aberrantes (outliers), ces dernières étant des molécules dont la structure diffère trop des autres dans la base de données ou dont les valeurs de propriété sont complètement différentes ; elles doivent être éliminées afin de ne pas perturber le développement des modèles. Les données utilisées dans les évaluations QSAR proviennent soit de la littérature, soit sont spécifiquement générées pour les analyses de type QSAR.

3.2 Représentation des structures

On peut visualiser les structures en utilisant un graphique en 2D, en supprimant l'hydrogène. On peut utiliser des logiciels pour créer cette structure en deux dimensions. ChemDraw est un exemple de représentation des structures utilisant le nom IUPAC (International Union of Pure and Applied Chemistry) [7]. En utilisant le logiciel de modélisation moléculaire HyperChem, il est possible de créer une structure tridimensionnelle (3-D) en utilisant des structures 2-D nécessaires pour générer des descripteurs géométriques.

3.3 Génération des descripteurs

Les descripteurs moléculaires jouent un rôle fondamental dans les études de la relation quantitative structure propriété. Ils sont utilisés en tant que variables indépendantes pour prédire une variable dépendante (la propriété : l'adsorption pour notre cas) [8].

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 (1D, 2D, 3D) [9].

Les descripteurs 1D : sont remportés directement à partir de la formule brute de la molécule. Exemples : le nombre d'atomes et la masse moléculaire, ... etc.

Les descripteurs 2D : la formule développée d'une molécule peut nous offrir ce type de descripteurs. On distingue :

- Les indices 2D : constitutionnels : ce type d'indices caractérise les différents composants de la molécule. Il s'agit par exemple du nombre de liaisons simples ou multiples, du nombre de cycles...
- Les indices 2D : topologiques : Ils donnent des informations sur la taille, la forme globale et les ramifications d'une structure. Exemples: indice de Wiener [10], indice de Randić [11], indice de connectivité de valence de Kier-Hall [12], indice de Balaban [13], ... etc.

Les descripteurs 3D : sont estimés à partir des positions relatives des atomes dans l'espace, et décrivent des caractéristiques plus complexes. Leurs calculs nécessitent donc de connaître la géométrie 3D de la molécule.

- Les descripteurs 3D : géométriques : les plus importants sont le volume moléculaire, la surface accessible au solvant, le moment principal d'inertie.
- Les descripteurs 3D : électroniques : permettent de quantifier les différents types d'interactions inter- et intramoléculaires, de grande influence sur l'activité biologique des molécules. Le calcul de la plupart de ces descripteurs nécessite la recherche de la géométrie pour laquelle l'énergie est minimale, et fait souvent appel à la chimie quantique.

3.4 Réduction des descripteurs

Des méthodes de sélection de variables sont disponibles dans le but de réduire ce nombre, notamment pour éviter la surparamétrisation de l'équation. En général, la réduction des descripteurs commence par l'élimination des données redondantes, c'est-à-dire celles qui présentent une corrélation élevée entre elles. De plus, les descripteurs considérés comme pertinents sont ceux qui sont fortement corrélés avec la propriété étudiée et qui présentent une variance significative, indispensable pour permettre la distinction entre les différentes données.

L'ensemble des descripteurs doit donc être aussi concis que possible tout en étant le plus informatif possible. C'est pourquoi les méthodes utilisées doivent sélectionner les descripteurs les plus pertinents. En outre, il est essentiel de prendre en compte le sens chimique des descripteurs utilisés, car plus ces descripteurs sont liés chimiquement au phénomène, moins il est probable qu'ils soient sélectionnés par hasard [14].

3.5 Choix de l'ensemble de calibrage et validation

Avant que nous débutions le développement des modèles, nous cherchons d'abord à diviser les données en deux sous-ensembles : Un pour le calibrage (training set) qui contient généralement les 2/3 de la base de données, et un pour la validation (test set) constituée par le 1/3 restant. Le sous-ensemble de calibrage représente des données initiales pour l'élaboration du modèle. Le sous ensemble de la validation doit être choisis pour évaluer la qualité prédictive du modèle. Pour sélectionner ces deux sous-ensembles, il existe de nombreux algorithmes de sélection, qui se différencient par leurs techniques de base. Sélection aléatoire : cette sélection

est parmi les méthodes les plus utilisés [15], on divise l'ensemble de données par un simple processus de sélection aléatoire de l'ensemble de calibrage et de test (validation externe).

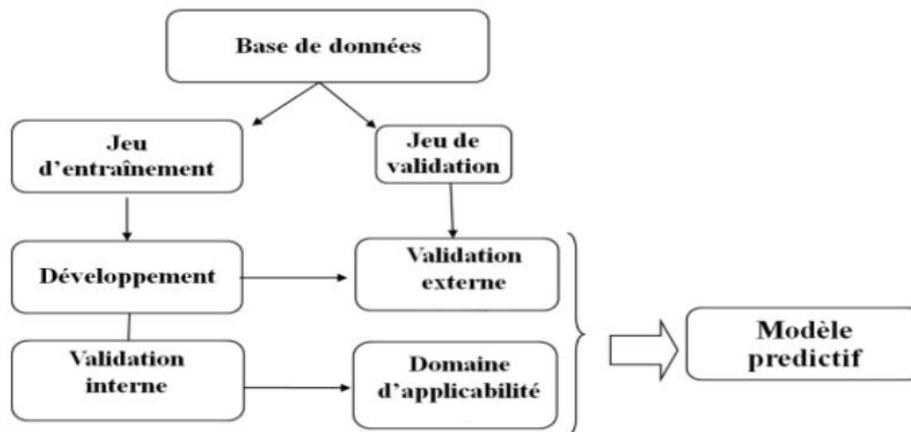


Figure 10 Partage des données expérimentales pour le développement d'un modèle

3.6 Sélection des descripteurs

Les meilleurs descripteurs sont sélectionnés en évaluant la qualité statistique de toutes les combinaisons possibles des descripteurs disponibles. Cette évaluation est réalisée à l'aide de techniques telles que la régression par les moindres carrés ordinaires (OLS) et la sélection de sous-ensembles de variables par algorithme génétique (AG-SSV) intégrés dans le logiciel MobyDigs. Ce programme est équipé d'une variété d'outils pour l'analyse exploratoire des données, incluant la division des ensembles de données en ensembles de calibrage et validation « d'entraînement et de prédiction », la détection des valeurs aberrantes avec le diagramme de Williams, ainsi que la validation interne et externe à l'aide de différents paramètres. Ils utilisent également des techniques de modélisation consensuelle et fournissent des visualisations graphiques pour faciliter la compréhension des résultats. Cette procédure de "sélection de variables" aboutit à une "population" de modèles, classés selon leur performance en termes de R^2 décroissant. Les meilleurs modèles sont identifiés en utilisant le Q^2_{LOO} comme critère d'optimisation, tout en respectant le principe de parcimonie, qui préconise des modèles aussi simples que possible en termes de dimension. De plus, la corrélation entre les descripteurs et la réponse modélisée est examinée à l'aide de la règle QUIK (Q Under Influence of K) afin d'écartier les modèles présentant une forte colinéarité entre les descripteurs [16].

4 L'analyse des données par la méthode « Régression linéaire multiple (MLR) »

Elle repose sur l'hypothèse qu'il existe une relation linéaire entre une variable dépendante (à expliquer) Y (ici, l'adsorption) et une série de p variables indépendantes (explicatives) Xi (ici, les descripteurs).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon \quad (1)$$

Cette équation est linéaire par rapport aux paramètres (coefficients de régression) $\beta_{0,1,\dots}$.

La détermination de l'équation (1) se fait alors à partir d'une base de données de n échantillons (observations) pour laquelle à la fois les variables indépendantes et la variable dépendante sont connues.

Il s'agit donc de considérer un système d'équations qui peut être donné sous la forme matricielle suivante [17] :

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \dots \\ Y_n \end{pmatrix} X = \begin{pmatrix} 1 & X_{11} & \dots & X_{1p} \\ 1 & & & \\ 1 & & & \\ \dots & & & \\ 1 & X_{n1} & \dots & X_{np} \end{pmatrix} \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \dots \\ \beta_p \end{pmatrix} \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \dots \\ \varepsilon_n \end{pmatrix} \quad (2)$$

5 Validation du modèle

Après avoir obtenu l'équation du modèle, il est primordial d'évaluer non seulement la stabilité et la qualité de l'ajustement du modèle, mais aussi d'estimer sa puissance et sa validité avant de l'utiliser pour prédire la propriété étudiée [18].

La validité est vérifiée par ce que l'on appelle validation, la validation des modèles QSAR a été reconnue ces dernières années, par des groupes d'experts spécifiques de l'OCDE [19] comme un point crucial et urgent, ce qui a conduit au développement, pour des raisons réglementaires, des « principes de l'OCDE pour la validation des modèles (QSAR) ».

5.1 Validation interne du modèle

La validation interne est souvent la technique la plus utilisée dans les études QSAR pour déterminer la stabilité et la robustesse du modèle et tester l'influence de chaque échantillon de l'ensemble d'entraînement sur le modèle final. Pour ce faire, nous utilisons des techniques de validation croisée leave-one-out (validation croisée CV)[20].

5.1.1 Validation croisée LOO (leave-one-out)

Pour déterminer la validation croisée LOO, l'ensemble de calibrage est principalement modifié en éliminant un composé de l'ensemble. Le modèle QSPR est ensuite reconstruit en fonction des molécules restantes de l'ensemble de calibrage en utilisant les descripteurs choisis, et la propriété du composé supprimé est calculée en fonction de l'équation QSPR.

Ce cycle est répété jusqu'à ce que toutes les molécules de l'ensemble de calibrage aient été supprimées une fois, et les données de la propriété prévue pour tous les composés de calibrage sont utilisées pour le calcul de divers paramètres de validation internes. Enfin, la prédiction du modèle est jugée à l'aide de la somme des carrés des erreurs de prédiction (PRESS) et du Q^2 pour le modèle tandis que la valeur de l'écart quadratique moyen de prédiction (SDEP) est calculée à partir du PRESS

$$\begin{aligned}
 PRESS &= \sum (Y_{obs} - Y_{pred})^2 \\
 SDEP &= \sqrt{\frac{PRESS}{n}} \\
 Q^2 &= 1 - \frac{\sum (Y_{obs(train)} - Y_{pred(train)})^2}{\sum (Y_{obs(train)} - \bar{Y}_{training})^2} = 1 - \frac{PRESS}{\sum (Y_{obs(train)} - \bar{Y}_{training})^2}
 \end{aligned}
 \tag{3}$$

Dans les équations (3), obs et pred correspondent aux valeurs de la propriété observée et prédite par LOO, n est le nombre d'observations, obs (train) est la propriété observée, pred (train) est la propriété prédite des molécules du jeu d'entraînement basée sur la technique LOO. Une valeur $Q_{LOO}^2 > 0.5$ est considérée comme satisfaisante, une valeur $Q_{LOO}^2 > 0.9$ est excellente.

5.1.2 Validation par le test de randomisation

Il n'est pas rare d'obtenir des corrélations fortuites (ou « chance corrélation »), c'est-à-dire un modèle affichant de bons résultats statistiques (R^2 , Q^2) pour l'apprentissage, mais impliquant des descripteurs qui ne sont pas reliés à la propriété modélisée. Ces modèles aléatoires peuvent être détectés par la procédure Y-randomisation [21]. Elle consiste à mélanger aléatoirement les propriétés expérimentales pour la série d'apprentissage en utilisant les mêmes descripteurs, de nouveaux modèles sont obtenus. Ces derniers doivent avoir des performances très faibles. La validation externe du modèle est indispensable pour vérifier le pouvoir prédictif d'un modèle QSPR. D'après certains chercheurs, la validation externe est une norme et une partie obligatoire dans la modélisation QSPR [22- 23].

5.2 Validation externe

Un modèle avec des valeurs élevées des indices internes Q^2 (ou R^2_{cv}) n'est pas encore dit valide, par conséquent, la validation interne est nécessaire mais insuffisante.

La puissance prédictive réelle d'un modèle QSPR est de tester leur capacité à prédire parfaitement l'activité/propriété des composés à partir d'un ensemble de test externe (composés non utilisés pour la construction du modèle). Le but d'un bon modèle QSPR est non seulement de prédire la propriété des composés d'ensemble d'apprentissage, mais aussi de prévoir les propriétés des molécules de test.

L'application du modèle, calculé sur l'ensemble de calibration, aux composés de l'ensemble de validation externe permet de vérifier de manière fiable la capacité prédictive du modèle obtenu.

L'équation(4) permet le calcul de Q^2_{ext} :

$$Q^2_{ext} = 1 - \frac{\sum_{i=1}^{n_{ext}} (y_i - \hat{y}_i)^2 / n_{ext}}{\sum_{i=1}^{n_{tr}} (y_i - \bar{y}_{tr})^2 / n_{tr}} \quad (4)$$

L'indice (ext) se rapportant aux objets de l'ensemble de validation externe, et l'indice (tr) à ceux de l'ensemble de calibrage. Seuil du Q^2

Avec Q^2_{ext} , le paramètre $EQMP_{ext}$ est également utile. On le calcule selon :

$$\text{EQMP}_{ext} = \sqrt{\frac{\sum_{i=1}^{n_{ext}} (y_i - \hat{y}_i)^2}{n_{ext}}} \quad (5)$$

La somme portant sur les objets de l'ensemble de validation (n_{ext})

5.2.1 Domaine d'applicabilité

L'établissement du domaine d'applicabilité pour chaque modèle représente l'un des défis majeurs dans l'analyse de QSPR. Le domaine d'applicabilité [24] du modèle (AD) est l'espace chimique dans lequel le modèle est fiable et peut être interpolé. Il est déterminé sur les données du jeu d'entraînement à partir des descripteurs du modèle (voir Fig. 6) permet de déterminer si le modèle peut être utilisé pour prédire la propriété pour une nouvelle molécule donnée. Le DA a été vérifié par la méthode du levier « leverage ». Cette méthode est basée sur la variation des résidus standardisés de la variable dépendante avec le levier (la distance entre les valeurs des descripteurs et leurs moyennes). L'effet de levier d'un composé dans l'espace est défini par [25]:

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (6)$$

L'analyse du domaine d'applicabilité est généralement réalisée à l'aide du graphe de Williams. Ce graphe représente graphiquement les valeurs résiduelles standardisées pour chaque composé de l'ensemble de données (TSET), en fonction de leurs valeurs de levier h_{ii} .

Où x_i est la valeur de descripteur du composé considéré, \bar{x} la valeur moyenne du descripteur dans l'ensemble d'apprentissage alors que n est le nombre de substances dans l'ensemble d'apprentissage (calibrage)

Si un composé a un résidu et un levier qui dépasse le seuil [$h^* = 3p/n$] (ou p est le nombre de descripteurs plus 1 et n le nombre d'observations), ce composé est considéré en dehors du domaine d'applicabilité du modèle élaboré.

En effet, un modèle QSPR n'est pas destiné à être employé en dehors de son domaine d'application, c'est-à-dire en dehors de l'espace chimique couvert par son jeu d'entraînement. Mis à part l'importance de la détermination des DA d'autres paramètres sont pris en considération pour le choix du modèle comme la possibilité d'interprétation des descripteurs.

Conclusion

La nouvelle politique des produits chimiques de la Commission Européenne (REACH : enregistrement, évaluation et autorisation des produits chimiques), énonce explicitement la nécessité d'utiliser des modèles QSPR pour réduire les tests expérimentaux (y compris les essais sur les animaux). De toute évidence, pour satisfaire aux exigences de la législation REACH, il est essentiel d'utiliser des modèles QSPR qui produisent des estimations fiables, c'est-à-dire des modèles prédictifs et validés

Une bonne pratique de modélisation QSPR, bien que l'utilisation des lignes directrices recommandées par l'OCDE, puisse développer de bons modèles prédictifs avec des applications pratiques démontrées dans divers domaines de la chimie biologique, ce qui pourrait renforcer son acceptabilité par la communauté scientifique.

Références

- [1] «**The Role of Molecular Modeling in Drug Discovery**'' by Andrew R. Leach (2006) (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6539951/>).
- [2] Alan R.Katritzky, Victor S.Lobanov and MatiKarelson, Normal Boiling points of Organic compounds: Correlation and prediction by a Quantitative Structure Property Relationship, *J.Chem.Inf.Comput.Sci.* 1998.38.28.
- [3]John DEARDEN | Cited by 8,329 | of Liverpool John Moores University, Liverpool (LJMU) | Read 230 publications | Contact John DEARDEN.
- [4]Saihi. S, Thèse de doctorat, Etude de la relation quantitative structure-activité inhibitrice des enzymes hydrolytiques : cas des alpha-glucosidases, Université Badjimokhtar – Annaba, 2015.
- [5] Tropsha, *Mol Inform.* 29, 2010, 476
- [6]Dearden, J. C.; Cronin, M. T. D.; Kaiser, K. L. E. How Not to Develop a Quantitative Structure–activity or Structure–property Relationship (QSAR/QSPR). *SAR and QSAR in Environmental Research* 2009, 20, 241–266.
- [7]Advanced Chemistry Development, Inc. IUPAC Nomenclature of Organics Chemistry <http://www.acdlabs.com/iupac/nomenclature>.
- [8] E. H. H. O. L.M. Arterburn, « Distribution, interconversion, and dose response of n– 3 fatty acids in humans,» *The American journal of clinical nutrition*, 83 , pp. 1467S-1476S, 2006.
- [9] M. C. T. S. J. M. S. Enoch, « An evaluation of global QSAR models for the prediction of the toxicity of phenols to Tetrahymenapyriformis,» *Chemosphere*, 71, pp. 1225-1232, 2008
- [10] *H. W. J. A. C. Soc.*, 69, 17, 1947.
- [11] *M. R. J. A. C. Soc.*, «,» 97, 6609, 1975.
- [12] *L. H. W. M. J. P. S. L.B. Kier*, 64,1971, 1975.
- [13] *A. B. C. P. Lett.*, 89, 399, 1982.
- [14] Witten I H, Frank E, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Publishers, San Francisco, 2005
- [15] N BOUARRA, *Etudes QSPR des propriétés contrôlant l'évolution dequelques HAP dans l'environnement* 2018.
- [16] Gramatica P, Chirico N, Papa E, Kovarich S, Cassani S, *QSARINS: A new software for the development, analysis, and validation of QSAR MLR models*, *J. Comput. Chem.* 2013, 34,2121.

- [17] O. FADIA, ETUDE QSAR DE LA TOXICITE D'UNE SERIE DE, TLEMCEN: UNIVERSITE ABOU-BEKR BELKAID , 2019.
- [18] J, Shao. Linear model selection by cross-validation. Journal of the American statistical Association, 88(422), (1993), 486-494.
- [19] OCDE, Principles for the Validation, for Regulatory Purposes, of (Quantitative) Structure-Activity Relationship Models, Organisation de Coopération et de Développement Economique, Paris, 2009.
- [20] Tropsha A. f. A. J. Mol. Graph. Model. 2002. 20. 269.
- [21] H. Linnan, P.C. Jurs, J. Mol. Graph. Model. 23, 2005, 503.
- [22] A. Tropsha, P. Gramatica, V.K. Gombar, QSAR Comb. Sci. 22, 2003, 69.
- [23] P.P. Roy, S. Paul, I. Mitra, K. Roy, Molecules, 14, 2009, 1660.
- [24] E. PAGA; J.C Dearden; P. Gramatica Chemospher 67, 2007, 351-358.
- [25] Karelson M, Molecular descriptors in QSAR/QSPR. New York: Wiley-InterScience, 2000.448.

Chapitre 3

**Application sur une série des composés
organiques utilisant la connectivité moléculaire
pour adsorption de charbon actif**

Résultats et discussion

1 Base de données et calcul des descripteurs

L'ensemble de données utilisé dans cette étude a été obtenu à partir du travail de *Blum et al.* [1] ; il se compose de 56 composés organiques. Le jeu de données a été divisé aléatoirement en un ensemble d'entraînement (40 composés) pour construire le modèle et un ensemble de validation externe (16 composés), également connu sous le nom de l'ensemble de prédiction. Les valeurs expérimentales des molécules de l'ensemble de données complet sont présentées dans le **Tableau 2** ainsi les valeurs des descripteurs du modèle développé.

2 Optimisation et calcul des descripteurs

Les descripteurs moléculaires théoriques ont été calculés par le processus suivant : Premièrement, les structures moléculaires ont été optimisées par la méthode semi-empirique PM3 avec un niveau Hartree-Fock restreint sans interaction de configuration, en appliquant un gradient limite standard de 0,001 Å kcal.mol⁻¹ comme critère d'arrêt par le logiciel de modélisation moléculaire HyperChem 6.03 (HyperChem, 1999) [2].

Les molécules optimisées par HyperChem ont été transférées dans le logiciel Dragon [3] pour le calcul des descripteurs. Ce programme permet de calculer plusieurs familles de descripteurs moléculaires tels que : indices constitutionnels, décomptes des groupes fonctionnels, auto corrélations 2D, indices topologiques et propriétés moléculaires, entre autres. Ensuite, les descripteurs moléculaires avec des valeurs constantes ou quasi constantes ont été éliminés, ils n'ont donc pas été utilisés dans les analyses ultérieures. Par conséquent, nous avons finalement utilisé 450 descripteurs pour effectuer l'algorithme génétique (implémenté dans le logiciel Mobydigs) pour la sélection des variables.

3 Développement du modèle QSPR-MLR

Pour le développement du modèle de régression linéaire multiple (MLR), le logiciel Mobydigs [4] a été utilisé. Il permet d'obtenir des modèles MLR par la méthode des moindres carrés ordinaires (OLS). Le modèle a été développé selon des critères tels que l'ajustement (R^2 le plus élevé), la robustesse (Q^2_{LOO} le plus grand), la stabilité ($R^2 - Q^2_{\text{LOO}}$ le plus bas) et la faible corrélation des descripteurs (K_{XX} le plus bas), de sorte qu'il présente une petite différence entre les paramètres d'ajustement, de validation croisée et de validation externe.

3.1 Le modèle QSPR-MLR

Plusieurs modèles ont été obtenus en explorant les meilleures combinaisons de descripteurs moléculaires montrant une forte corrélation avec la variable à expliquer ($\log(X/C)$). Par conséquent, en tenant compte du principe de parcimonie (décrire le maximum d'informations avec le moins de descripteurs possible), une analyse des paramètres du modèle a été réalisée. Sur cette base, un modèle QSPR-MLR a été développé pour évaluer les l'adsorption avec quatre variables. L'équation et les paramètres statistiques du meilleur modèle sont les suivants :

$$\text{Log}(X/C) = -13.0 - 0.00981 \text{ ZM1V} + 22.8 \text{ FDI} - 16.3 \text{ VEv2} + 0.802 \text{ RDF020v} \quad (7)$$

R²	RMSE tr	s	F	Q²_{LOO}	R²- Q²_{LOO}	RMSE_{cv}	RMSE_{ext}	Q²_{ext}
0.8649	0.4081	0.4363	56.0311	0.808	0.057	0.4867	0.4711	82.30

Les statistiques prouvent la stabilité, la robustesse et la capacité prédictive du modèle construit (Équation 7). Par conséquent, le modèle a été accepté avec des valeurs de R² supérieures à 0,85. De plus, ce modèle présente les plus petites valeurs de RMSE_{tr} indiquant que ce modèle a l'erreur la plus faible, c'est-à-dire les plus petites différences par rapport aux données prédites. En outre, la valeur de Q²_{LOO} du modèle est significative, supérieure à 0,6 et proche de R². Le résultat de la validation externe du modèle construit a montré une grande capacité prédictive, car les valeurs de Q²_{ext} sont supérieures à 0,7.

Tableau 2 Noms, valeurs expérimentale et valeurs des descripteurs

ID	Object	Log (X/C)	ZM1V	FDI	VEv2	RDF020v
1	Aldicarb	3.85	158.444	0.991	0.285	0.926
2	Bromodichloromethane	2.01	10.277	1	0.493	0.801
3	Bromoform	2.87	9.202	1	0.491	0.806
5	Carbontetrachloride	2.2	18.42	1	0.441	0.007
7	cis-1,2-Dichloroethene	2.01	21.21	1	0.492	0.144
8	Dalapon	2.23	95.21	1	0.372	0.203
9	Dibromochloromethane	2.51	9.739	1	0.492	1.101
10	Dibromochloropropane	4.19	17.739	1	0.402	1.422
11	Dibromomethane	1.45	4.134	1	0.568	1.267
13	Hexachlorocyclopentadiene	4.14	83.63	1	0.295	0.005
15	Lindane	3.99	99.63	1	0.283	0
16	Methylenechloride	0.55	5.21	1	0.572	0.007
18	ter t-Butylmethylether	1.29	56	0.911	0.4	0.479

CHAPITRE 3

19	trans-1,2-Dichloroethene	2.17	21.21	1	0.492	0.144
21	1,1-Dichloroethane	1.48	9.21	1	0.494	0.078
22	1,1-Dichloroethene	2.27	21.21	1	0.492	0.144
23	1,1-Dichloropropene	2.42	27.21	1	0.438	0.237
25	1,1,1,2-Tetrachloroethane	3	22.42	1	0.402	0.057
26	1,1,2-Trichloroethane	2.13	14.815	1	0.44	0.065
27	1,2-Dibromoethane	3.32	8.134	1	0.493	2.306
28	1,2-Dichloroethane	1.35	9.21	1	0.494	0.078
32	2,4,5-Trichlorophenoxyacetic acid	3.59	200.815	1	0.262	0.398
33	Alachlor	5.08	207.605	0.987	0.23	0.874
34	Atrazine	4.15	155.605	0.974	0.261	1.177
35	Benzene	2.86	54	0.999	0.408	0.295
36	Bromobenzene	3.37	61.067	1	0.375	0.472
37	Chlorobenzene	3.19	61.605	1	0.374	0.235
38	Cyanazine	3.37	219.605	0.964	0.246	0.945
41	Ethylbenzene	3.03	66	0.983	0.348	0.501
42	Metachlor	3.89	310.445	1	0.209	1.279
43	Metribuzin	3.39	172.444	0.983	0.263	0.652
44	ortho-Chlorotoluene	3.88	69.605	1	0.348	0.316
45	ortho-Dichlorobenzene	3.57	69.21	1	0.348	0.192
47	para-Dichlorobenzene	3.46	69.21	1	0.347	0.173
48	para-Xylene	3.98	70	0.993	0.347	0.424
50	Picloram	4.94	192.815	1	0.271	0.751
51	Simazine	5.81	165.605	0.995	0.271	1.061
53	Toluene	3.49	62	0.995	0.374	0.36
54	1,3,5-Trichlorobenzene	4.24	76.815	1	0.327	0.11
56	Pichloroam	3.34	34.42	1	0.4	0
4	Carbofuran	3.94	238	0.963	0.244	1.029
6	Chloroform	1.6	10.815	1	0.493	0.007
12	Glyphosate	3.69	187.309	1	0.312	0.857
14	Isophorone	3.42	114	0.982	0.311	0.246
17	Oxamy	3.16	226.444	0.939	0.263	0.871
20	Trichloroethene	2.84	26.815	1	0.438	0.054
24	1,1, 1-Trichloroethane	1.98	18.815	1	0.441	0.092
29	1,2-Dichloropropane	2.25	15.21	0.996	0.44	0.143
30	1,2,3-Trichloropropane	2.83	18.815	1	0.401	0.111
31	1,3-Dichloropropane	2.65	13.21	0.997	0.441	0.156
39	Dicamba	3.39	197.21	1	0.269	0.627
40	Dinoseb	4.45	276	0.979	0.235	0.907
46	Para	4.1	69.605	1	0.347	0.301
49	Pentachlorophenol	4.89	124.025	1	0.282	0.087
52	Styrene	3.53	74	0.995	0.35	0.483
55	2,4-Dinitrotoluane	4.19	230	0.985	0.268	0.931

3.2 Matrice de corrélation :

Le Tableau 3 présente la matrice de corrélation, où il est clair que les quatre descripteurs sélectionnés sont presque totalement non corrélés. Il s'agit d'une condition essentielle qui nous permet d'affirmer que les variables sélectionnées se comportent de manière indépendante.

Tableau 3 Matrice de corrélation pour les quatre descripteurs sélectionnés dans le modèle.

	log (X/C)	ZM1V	FDI	VEv2
ZM1V	0.631	1		
FDI	0.097	-0.212	1	
VEv2	-0.804	-0.887	0.22	1
RDF020v	0.307	0.275	-0.133	-0.116

4 La qualité de l'ajustement :

La Figure 11, reproduit les valeurs prédites de log (X /C) en fonction de celles mesurées. La dispersion des points autour de la première bissectrice montre que les valeurs prédites et calculées pour l'ensemble de validation et l'ensemble de calibrage respectivement sont en adéquation avec les valeurs expérimentales. Les valeurs proches de Q^2_{LOO} et R^2 indiquent que pour la majorité des composés, le modèle proposé a un bon ajustement.

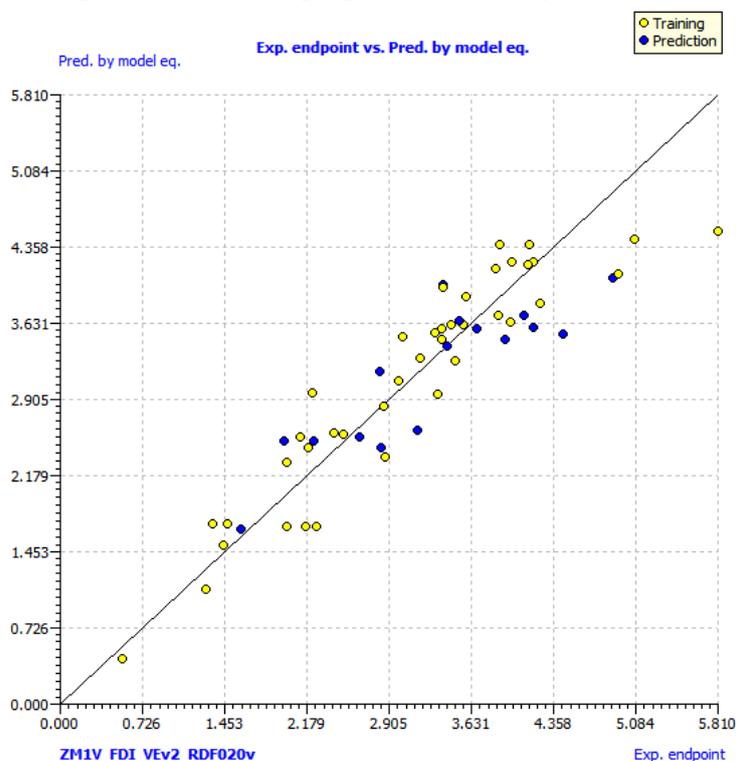


Figure 11 Les valeurs Prédites versus les valeurs expérimentales

5 Domaine d'application (Diagramme de Williams)

Un aspect crucial dans les études de QSPR, est la définition du domaine d'applicabilité (DA), car seules les prédictions pour les composés qui se trouvent dans ce domaine peuvent être considérées comme fiables. Dans ce travail, nous utilisons l'approche du levier (h) et des résidus standardisés. La Figure 12 montre le graphique de Williams pour les ensembles d'entraînement et de prédiction ; comme on peut le voir sur cette figure, la plupart des composés se situent dans le DA du modèle. Seuls quelques composés de l'ensemble d'entraînement et un composé de l'ensemble de prédiction (validation externe) présentent des valeurs de levier supérieures à la valeur critique du levier ($h^* = 3 \times (4+1)/40 = 0,375$), bien qu'ils affichent des valeurs d'écart type dans les limites, ce qui signifie qu'ils doivent être considérés comme des composés influents plutôt que comme des valeurs aberrantes. De plus, le composé 51 dans ce jeu est considéré comme une valeur aberrante en Y en raison de ses résidus standardisés dépassant 3 écarts types. Par conséquent, nous pouvons dire que ce modèle, pour son domaine d'applicabilité, peut être utilisé avec précision.

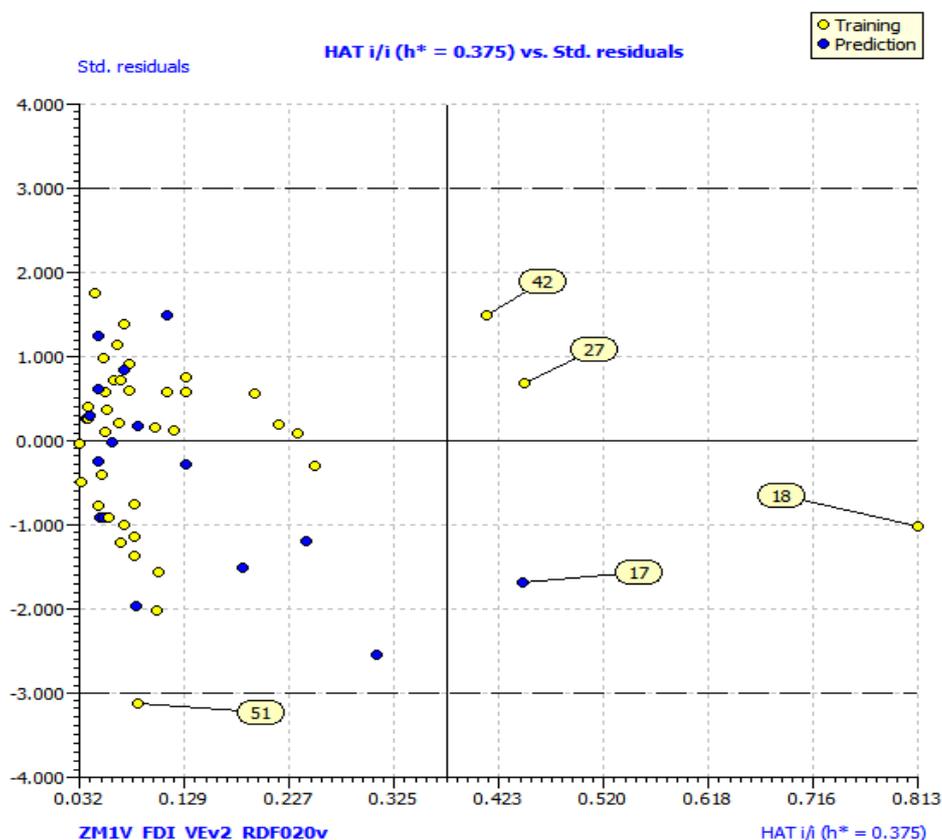


Figure 12 Diagramme de Williams

6 Test de Randomisation :

La dernière technique de validation interne a été la procédure de test de randomisation, comme nous l'avons mentionné précédemment ; elle a été réalisée pour démontrer que le modèle n'est pas le résultat d'une corrélation fortuite. Ici, les réponses ont été placées au hasard, de sorte qu'il n'y a aucune corrélation avec les descripteurs. En conséquence, les performances du modèle se dégradent considérablement. Les valeurs de R^2 et Q^2 de chaque itération, et leurs moyennes (R^2_{Y-scr} et Q^2_{Y-scr}), offrent les critères montrant que le modèle est bon, puisque ces paramètres sont toujours inférieurs par rapport aux valeurs du modèle ($R^2_{Y-scr} = 0.117$ et $Q^2_{Y-scr} = -0.246$). Les valeurs de R^2_{Y-scr} et Q^2_{Y-scr} par rapport à R^2 et Q^2 du modèle sont présentées dans la Figure 13. Il est à noter que les valeurs de R^2 et Q^2 du modèle sont éloignées des valeurs obtenues pour ces paramètres dans le test de randomisation, ce qui indique que le modèle n'est pas obtenu à la suite d'une corrélation aléatoire.

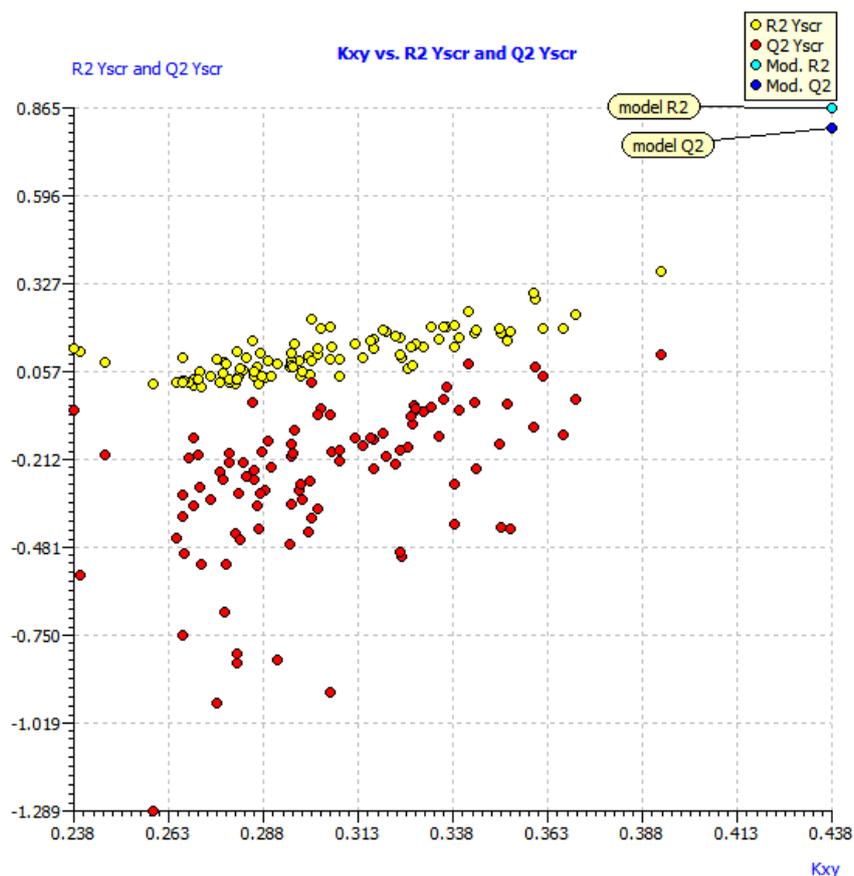


Figure 13 Test de randomisation

7 Relatif importance des descripteurs

L'influence des quatre variables d'entrée les plus importantes, identifiées à l'aide d'un algorithme génétique sur le $\log(X/C)$, a été étudiée. Selon la Figure 14, le paramètre VE2v était le plus influent avec une importance relative d'environ 35,42 %, tandis que l'influence de RDF020v et ZM1V était respectivement de 23,11 % et 21,39 %. Le FDI avait une influence à niveau de 20,08 %. La différence de contribution entre les quatre descripteurs n'est pas significative ce qui indique que tous ces descripteurs sont nécessaires pour la construction du modèle.

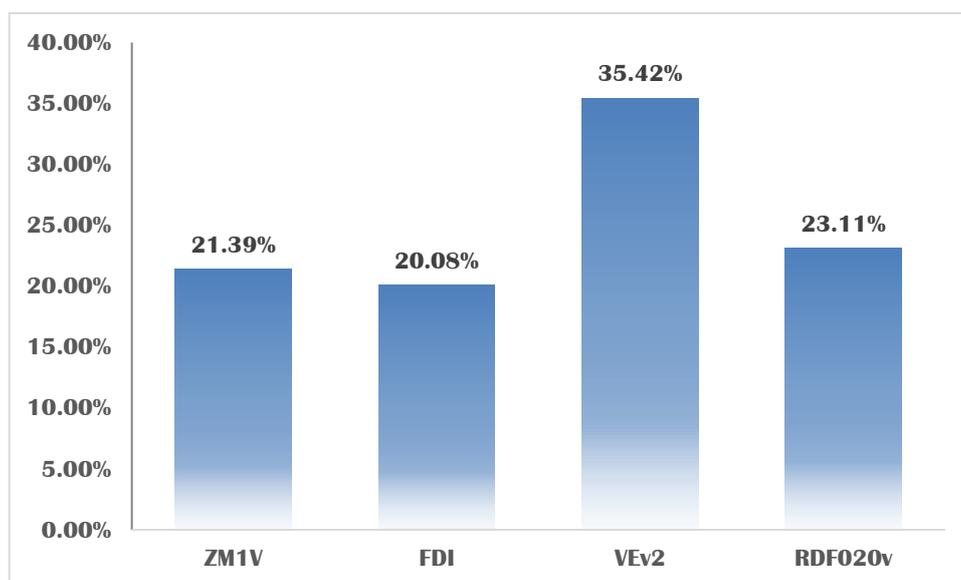


Figure 14 Contribution (%) des descripteurs au modèle.

8 Interprétation Mécanistique du modèle développé

ZM1V (Indice de Zagreb de premier ordre par degrés de sommets de valence). Ce descripteur capture la connectivité et les informations liées à la valence des atomes au sein d'une molécule. L'indice de Zagreb de premier ordre somme les carrés des degrés de valence (le nombre de liaisons qu'un atome forme, en considérant ses électrons de valence).

Les molécules avec des valeurs **ZM1V** plus élevées ont souvent des structures plus complexes ou plus densément connectées. Dans le contexte de l'adsorption, de telles molécules peuvent avoir plus de points d'interaction ou des affinités de liaison plus fortes avec la surface du charbon actif en raison de leurs sites de liaison multiples et de leurs régions de densité électronique plus élevée.

CHAPITRE 3

Le deuxième descripteur, **FDI** (Comptage des Groupes Fonctionnels). Ce descripteur compte la présence de groupes fonctionnels spécifiques au sein de la molécule, tels que les groupes hydroxyle, amino, carboxyle, etc.

Les groupes fonctionnels influencent significativement les caractéristiques d'adsorption. Par exemple, les groupes fonctionnels polaires (comme les hydroxyles ou les carboxyles) peuvent former des liaisons hydrogène ou d'autres interactions dipolaires avec le charbon actif, tandis que les groupes non polaires peuvent interagir par des forces de van der Waals. Le type et le nombre de groupes fonctionnels ont un impact direct sur l'efficacité et la spécificité de l'adsorption.

Le troisième descripteur, **VEv2** (Deuxième Valeur Propre de la Matrice d'Adjacence). Ce descripteur basé sur les valeurs propres reflète la topologie globale du graphe moléculaire. Il est influencé par la connectivité et la ramification de la molécule.

La deuxième valeur propre de la matrice d'adjacence peut indiquer la diffusion et la portée au sein de la structure moléculaire. Les molécules avec certaines caractéristiques topologiques peuvent mieux s'adapter à la structure poreuse du charbon active, améliorant l'adsorption. Par exemple, une molécule avec une diffusion optimale peut interagir de manière plus étendue avec la surface des pores.

Le quatrième descripteur, **RDF020v** (Fonction de Distribution Radiale à 2,0 Å pondérée par les volumes de van der Waals). Ce descripteur fournit des informations sur la distribution de probabilité de trouver des paires d'atomes à une distance de 2,0 Å, pondérée par leurs volumes de van der Waals.

Ce descripteur capture l'arrangement spatial des atomes et les contributions volumétriques aux interactions moléculaires. Les molécules avec des caractéristiques RDF spécifiques peuvent avoir des orientations et des distances favorables qui améliorent les interactions de van der Waals avec la surface du carbone. Ce descripteur spatial aide à comprendre comment la forme et la taille moléculaires contribuent à s'adapter aux micro et méso pores du charbon actif.

Ensemble, ces descripteurs fournissent une vision holistique de la manière dont un produit chimique pourrait interagir avec le charbon active :

Connectivité Structurale (ZM1V) : Indique les points d'interaction potentiels et la stabilité moléculaire globale, contribuant à des sites d'adsorption plus forts ou multiples.

CHAPITRE 3

Groupes Fonctionnels (FDI) : Met en évidence la nature chimique et la réactivité de la molécule, indiquant comment elle pourrait se lier chimiquement ou interagir avec la surface du charbon active.

Topologie Moléculaire (VEv2) : Reflète la forme et la distribution de la molécule, influençant sa capacité à pénétrer et à interagir dans la structure poreuse du charbon active.

Arrangement Spatial (RDF020v) : Fournit des informations sur l'arrangement tridimensionnel de la molécule, crucial pour s'adapter aux pores du carbone et maximiser les interactions de van der Waals.

Conclusion

Le Modèle développé peut prédire efficacement le comportement d'adsorption de différents produits chimiques sur le charbon active, tenant compte à la fois des interactions chimiques et physiques à un niveau détaillé. Cette compréhension mécanistique aide à concevoir et à sélectionner des adsorbants appropriés pour des contaminants spécifiques en fonction de leurs propriétés moléculaires.

Référence

- [1] Blum, D. J., Suffet, I. H., & Duguet, J. P. (1994). Quantitative structure-activity relationship using molecular connectivity for the activated carbon adsorption of organic chemicals in water. *Water research*, 28(3), 687-699.
- [2] HyperChem 6.03 Package. Hypercube, Inc., Gainesville, Florida, USA, 1999; software available at: <http://www.hyper.com>.
- [3] DRAGON Software for the calculation of molecular descriptors. Ver. 5.4 for Windows, 2006, Taletesrl, Milan, Italy.
- [4] MOBY DIGS Software for Multilinear Regression Analysis and Variable Subset Selection by Genetic Algorithm, Release 1,1 for windows, Milano.

Conclusion générale

Conclusion générale

La caractérisation expérimentale de l'ensemble des propriétés de composés chimiques est contraignante pour des raisons de temps, de coûts, d'éthique (essais sur animaux) et de faisabilité au niveau recherche et développement. Ainsi, le développement de méthodes prédictives, est recommandé

En particulier l'utilisation des modèles QSPR fournit un compromis nécessaire qui permet l'estimation des paramètres physicochimiques de grandes classes de composés qui sont importants du point de vue théorique ou pour les applications industrielles.

Les matières organiques ont longtemps été les principaux polluants des milieux aquatiques. Elles proviennent des déchets domestiques agricoles ou industriels, lorsque ceux-ci sont rejetés sans traitement préalable

L'objectif de cette étude est de déterminer une méthode permettant de prédire l'adsorption de la matière organique sans réaliser d'expériences.

Un ensemble de données composé de 56 composés a été sélectionné, puis différents types de descripteurs moléculaires ont été calculés par le programme Dragon. L'ensemble de données a été divisé en un ensemble d'entraînement (40 composés) pour construire le modèle et un ensemble de validation externe (16 composés).

Dans ce contexte, l'objectif de cette thèse était de développer des modèles QSAR Fiables pour la prédiction de l'adsorption des produits chimiques organiques sur le charbon actif.

Un grand nombre de descripteurs moléculaires a été calculé (Descripteurs constitutionnels, électroniques, topologiques, géométriques, physicochimiques).

La méthodologie basée sur la MLR, a été utilisée principalement dans la prédiction.

Les principales techniques de validation ont été utilisées (la validation interne, la validation externe, la randomisation, les domaines d'applications...).

Des validations internes et externes ont été utilisées pour juger la stabilité, la justesse et la capacité prédictive des modèles obtenus.

L'approche mécanistique présentée dans cette étude constitue une avancée significative dans la compréhension et la prédiction du comportement d'adsorption sur le charbon actif.

En combinant les interactions chimiques et physiques à un niveau détaillé, elle offre une méthode robuste pour la conception et la sélection d'adsorbants adaptés à différents contaminants. Cette approche ouvre la voie à des solutions de dépollution plus efficaces et mieux adaptées.