الجمهورية الجزائرية الديمقراطية الشعبية

وزارة التعليم العالي والبحث العلمي

UNIVERSITÉ BADJI MOKHTAR - ANNABA
BADJI MOKHTAR – ANNABA UNIVERSITY

جامعة باجي مختار – عنابـــة

Faculté : TECHNOLOGIE

Département : GENIE DES PROCEDES

Domaine : SCIENCES ET TECHNOLOGIES

Filière : GENIE DES PROCEDES

Spécialité : GENIE DES PROCEDES DE ENVIRONNEMENT

Mémoire

Présenté en vue de l'obtention du Diplôme de Master

Thème:

Conception d'un Modèle QSPR pour Prédire l'Indice de Réfraction des matrices polymères

Présenté par : *SEMICHI Rania et SENNAOUI Oumaima*

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

Jury de Soutenance :

Pr BECHIRI Ouahiba	Pr	*UBMA*	Président
Dr Kherouf Soumaya	MAB	*UBMA*	Encadrant
Dr LAHMAR Amina	MAB	*UBMA*	Examinateur

Année Universitaire : 2023/2024

« Sic mundus creatus est. Hincerunt adaptation es mirabiles, quarum modus est hic » -La Table d'émeraude d'Hermès Trismégiste, père des Philosophes.

Semichi Rania

Remerciements

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

Mme ***Kherouf Soumaya*** pour avoir d'abord proposé ce thème, pour suivi continuel 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 aidées, de près ou de loin, à accomplir ce travail.

Nos remerciements vont aussi à tous les enseignants du département de génie des procèdes 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….

Dédicace

Le devoir de reconnaissance m'oblige de dédier ce modeste mémoire à tous ceux qui me sont chers, ce sont ceux à qui je dois e mon succès.

A mon cher père Noureddine, chaque mot semble bien fade pour exprimer l'amour profond et la gratitude infinie que j'ai pour toi, pour les innombrables sacrifices que tu as consentis pour mon éducation. Tu as été bien plus qu'un guide, tu as été mon modèle d'honnêteté, de sérieux et de responsabilité. Ta présence incarne pour moi la quintessence de la persévérance, de la créativité et du dévouement sans bornes.

A ma merveilleuse mère **Karima**, aucun mot ne saurait capturer la profondeur de l'amour et de l'affection que je ressens pour toi. Tu es bien plus qu'une mère, tu es mon phare de générosité et mon exemple de dévouement. Ta tendresse infinie est une source inépuisable de réconfort, et je te suis infiniment reconnaissante pour chaque instant où tu as été là pour moi, sans jamais faillir. Merci pour ta présence rassurante et pour tous ces instants où ton amour inconditionnel a été ma plus grande force.

À mon grand frère **Seif**, tu es bien plus qu'un frère, tu es mon guide, mon confident et mon modèle. Ta présence dans ma vie est un cadeau précieux que je chéris chaque jour. Merci pour ta force, ton soutien inconditionnel et ton amour sans faille. Je suis infiniment reconnaissante de t'avoir comme frère., et mes chers frères aussi **Mohamed**, mon soutien et ma force.

À tous mes amis avec lesquelles j'ai partagé de merveilleux moments.

À certaines des personnes spéciales que j'ai connues au cours de cette période de trois ans et tous les **gpe** de la promo 2024.

Pour toute la famille **Boulbadaoui** et la famille **Semichi**

Merci pour tout votre amour et votre confiance. Je te dédie ce travail en témoignage de mon amour et ma profonde reconnaissance.

Semichi rania

À ma mère et à mon père, qui m'ont soutenu à chaque étape de ma réussite et qui m'ont motivé à rester ici, sans qui je ne serais pas là où je suis aujourd'hui.

À mon grand frère, mon ami, sur lequel j'ai compté les yeux fermés pour qu'il me soutienne quel que soit mon choix.

À mes jeunes frères et sœurs et à leur motivation.

À ma merveilleuse et unique petite amie, qui a été mon bras droit et m'a aidé dans tous mes choix.

À ma grande et belle famille, qui ne ressemble à aucune autre dans ce monde.

Senaoui oumaima

Résumé

Comprendre les bases moléculaires de la réfractivité des polymères est essentiel pour faire progresser les applications en science des matériaux et en ingénierie dans les domaines de l'optique et de la photonique. Dans cette étude, nous développons un modèle de relation structure-activité quantitative (QSAR) pour prédire la réfractivité des polymères, en utilisant des descripteurs moléculaires dérivés de la structure moléculaire.

Le logiciel de modélisation moléculaire HyperChem (version 6.03) a été utilisé pour représenter les structures moléculaires, tandis que la méthode semi-empirique PM3 a été appliquée pour optimiser les géométries moléculaires. Les descripteurs moléculaires ont ensuite été calculés à l'aide du logiciel Dragon (version 5.5).

Un modèle de relation quantitative structure-propriété (QSPR) ont été développé pour prédire l'indice de réfraction d'une série de polymères. La régression linéaire multiple (MLR) a été utilisée pour construire ces modèles, qui varient en qualité. Pour améliorer la sélection des descripteurs, des algorithmes génétiques (GA) ont été adoptés.

Les modèles obtenus ont été validés conformément aux cinq principes directeurs de l'Organisation de Coopération et de Développement Économiques (OCDE). Le domaine d'application des modèles a été évalué à l'aide du diagramme de Williams, permettant d'identifier les composés aberrants en termes de valeurs prédictives X et Y.

Des procédures de validation internes et externes ont été menées pour évaluer la robustesse et la capacité prédictive des modèles, visant à combler les lacunes des polymères.

Mots-clés : QSPR, Propriétés optique, Descripteurs moléculaires, Régression linéaire multiple, polymères.

Abstract

Understanding the molecular basis of polymer refractivity is essential for advancing applications in materials science and engineering, particularly in the fields of optics and photonics. In this study, we develop a quantitative structure-activity relationship (QSAR) model to predict polymer refractivity using molecular descriptors derived from molecular structure.

The molecular modeling software HyperChem (version 6.03) was used to represent molecular structures, while the semi-empirical PM3 method was applied to optimize molecular geometries. Molecular descriptors were then calculated using the Dragon software (version 5.5).

A quantitative structure-property relationship (QSPR) model was developed to predict the refractive index of a series of polymers. Multiple linear regression (MLR) was used to construct these models, which vary in quality. To improve descriptor selection, genetic algorithms (GA) were employed.

The obtained models were validated according to the five principles established by the Organisation for Economic Co-operation and Development (OECD). The applicability domain of the models was evaluated using the Williams plot, allowing the identification of outlier compounds in terms of predictive values X and Y.

Internal and external validation procedures were conducted to assess the robustness and predictive capability of these models, aiming to fill gaps in the physicochemical data of polymers.

Keywords: QSPR, Optical properties, Molecular descriptors, Multiple linear regression, Polymers.

الملخص

فهم الأسس الجزيئية لانكسار البوليمرات أمر ضروري لتعزيز التطبيقات في علم وهندسة المواد، خاصة في مجالات البصريات والفوتونيات. في هذه الدراسة، نقوم بتطوير نموذج العلاقة الكمية بين البنية والنشاط (QSAR) للتنبؤ بانكسار البوليمرات باستخدام واصفات جزيئية مشتقة من البنية الجزيئية.

تم استخدام برنامج النمذجة الجزيئية) HyperChem الإصدار 6.03) لتمثيل البنى الجزيئية، بينما تم تطبيق الطريقة شبه التجريبية PM3 لتحسين الأشكال الهندسية الجزيئية. تم حساب الواصفات الجزيئية باستخدام برنامج) Dragon الإصدار 5.5.).

تم تطوير نموذج العلاقة الكمية بين البنية والخواص (QSPR) للتنبؤ بمعامل الانكسار لمجموعة من البوليمرات. تم استخدام الانحدار الخطي المتعدد (MLR) لبناء هذه النماذج، والتي تختلف في الجودة. لتحسين اختيار الواصفات، تم اعتماد الخوارزميات الجينية.(GA)

تم التحقق من صحة النماذج التي تم الحصول عليها وفقًا للمبادئ الخمسة التي وضعتها منظمة التعاون والتنمية الاقتصادية .(OECD)تم تقييم مجال تطبيق النماذج باستخدام مخطط Williams ، مما يسمح بتحديد المركبات الشاذة من حيث القيم التنبؤية X و.Y

تم إجراء إجراءات تحقق داخلية وخارجية لتقييم استقرار وقدرة التنبؤ لهذه النماذج، بهدف سد الثغرات في بيانات البوليمرات الفيزيائية والكيميائية.

الكلمات المفتاحية : QSPR، الخواص البصرية، الواصفات الجزيئية، الانحدار الخطي المتعدد، البوليمرات.

Listes figures

Figure 1:polymére du polystyréne 5
Figure 2:Types des polymères : homopolymères et copolymères 7
Figure 3:Polymères linéaires.....7
Figure 4:Homopolymère ramifié (a) et copolymère ramifié (b) 8
Figure 5:Polymère réticulé avec ponts di-sulfure reliant deux chaînes 8
Figure 6:Représentation schématique d'un polymère semi-cristallisé.....9
Figure 7:Classification des architectures macromoléculaires selon Tamolia. 10
Figure 8:Plaque de thermoplastique et quelque produit. 11
Figure 10:Quelques objets sont des élastomères 12
Figure 9:Quelques objets thermodurcissables 12
Figure 11:présentation de la méthodologie QSPR 18
Figure 12:La stratégie générale d'une étude QSPR 19
Figure 13:Quelques exemples de descripteurs et leur classification en 1D, 2D et 3D 22
Figure 14:Résidus versus valeurs expérimentales de IR.....35
Figure 15:Diagramme de Williams.....36
Figure 16:. Les valeurs Prédites versus les valeurs expérimentales de IR.....37
Figure 17:Test de randomisation.....38

Listes tableaux

Tableau 1:Les domaines d'application des polymères.....15
Tableau 2:Définition des descripteurs intervenant dans le modèle.....29
Tableau 3:Noms des valeurs des descripteurs utilisés dans le modèle 30
Tableau 4:paramètres statistiques du modèle.....32
Tableau 5:paramètre statistique du modèle choisi. 33
Tableau 6:Valeurs de IR (expérimentales et prédites), résidus de prédiction (standardisés et ordinaires), valeurs des leviers (hii). 33

Sommaire

Introduction générale 5

Introduction 6

1. Définition des Polymères : 7
2. Historique 7
3. Degré de polymérisation : 8
4. Types de polymères: 8
   4.1. Homopolymères 8
   4.2 Copolymères : 9
5. Structure despolymères: 9
   5.1.Polymères linéaires : 9
   Formés par des monomères connectés en une seule chaîne longue. 9
   5.2 Polymères ramifiés : 9
   Possèdent des chaînes latérales courtes ou longues qui se ramifient àpartir de la chaîne principale 9
   5.3. Polymères réticulés : 10
   Présentent des liaisons covalentes entre les chaînes polymériques, formant une structure tridimensionnelle 10
   5.4. Polymères amorphes et polymères cristallisés : 11
6. Classification des Polymères 11
   6.1 Selon l'Origine : 11
   6.2 Selonleur nature chimique : 11
   6.3. Selon l'architecture macromoléculaire : 12
   6.4 Selon leurs usages technologiques : 12
   6.5 Selon le comportement thermique: 12
   6.7 Selon l'importance économique: 14
7. Quelques propriétés des polymères : 15
   7.1 Propriétés chimiques : 15
   7.2 Propriétés physiques : 15
8. Applications des polymères : 15
9. L'indice de réfraction des polymères : 17
10. Les facteurs influençant l'indice de réfraction des polymères incluent : 17

1.Principe : 19

2.Objectifs : 20

3.Méthodologie générale des études QSPR : 20

3.1 Élaboration du modèle : 22

3.1.1.Base de données : 22

3.1.3 Génération Les descripteurs moléculaires : 22

3.1.4 Réduction des descripteurs 25

3.1.5. Régression linéaire multiple (MLR) 25
3.1.6 Constitution des jeux d'apprentissage et de test 26
4.Qualité des modèles QSPR : 26
4.1. L'ajustement : 26
4.2. Le pouvoir de prévision interne : 27
4.3. Test de randomisation : 28
4.4. Le pouvoir de prévision externe : 29
4.5. Domaine d'Application (DA) des modèles : 29
1.Analyse MLR : 31
2.Modèle relatif 31
3.Équation et Analyse de régression 34
4.Analyse des résidus : 35
5.Domaine d'application : 38
6.Qualité de l'ajustement 39
7.Test de Randomisation : 40
8. Mécanistique interprétation du modèle: 41

Introduction générales

Les polymères sont des matériaux essentiels dans de nombreuses applications industrielles, allant des emballages aux dispositifs électroniques avancés, en passant par les biomatériaux. L'indice de réfraction des polymères joue un rôle crucial dans leurs performances optiques, telles que la transmission de la lumière, la réflexion et la réfraction. Par conséquent, la capacité à prédire avec précision l'indice de réfraction des matrices polymères à partir de leur structure moléculaire est d'un intérêt significatif pour les scientifiques et les ingénieurs des matériaux.

Les modèles QSPR (Quantitative Structure-Property Relationship) offrent une approche prometteuse pour établir des relations quantitatives entre la structure moléculaire des polymères et leurs propriétés physico-chimiques. En utilisant des descripteurs moléculaires appropriés et des techniques de modélisation avancées, il est possible de développer des outils prédictifs robustes qui peuvent accélérer le processus de conception de nouveaux polymères avec des propriétés optiques optimisées.

Dans cette étude, nous proposons la conception d'un modèle QSPR visant à prédire l'indice de réfraction des matrices polymères. En utilisant une combinaison judicieuse de descripteurs moléculaires et de méthodes de modélisation statistique, notre objectif est de développer un modèle précis et fiable capable de prédire l'indice de réfraction pour une large gamme de polymères. Ce modèle pourrait non seulement fournir des informations précieuses pour la conception de nouveaux polymères avec des propriétés optiques spécifiques, mais aussi contribuer à une meilleure compréhension des facteurs moléculaires influençant l'indice de réfraction dans les matériaux polymériques.

Le manuscrit de cette thèse est articulé sur trois grandes parties :

- Le premier chapitre : nous présentons des généralités sur les polymères en exposants les différentes classifications et leur indice de réfraction
- Deuxième chapitre : portera sur les modèles QSPR, en exposant les différentes étapes de leur développement.
- Dans le troisième chapitre : nous présentons nos résultats obtenus lors de cette étude, Nous exposants les modèles mathématiques prédictif et fiables que nous avons développés en utilisant la technique GA-MLR

Introduction

Depuis des siècles, l'Homme exploite les polymères naturels présents dans les végétaux et les animaux. Le bois, le caoutchouc, le coton, la laine, le cuir et la soie ce sont des exemples frappants. Ces matériaux, ont joué un rôle crucial dans le développement de notre civilisation. Utilisés dans diverses applications, allant de la construction à l'habillement, ces polymères naturels ont servi de fondement à de nombreuses avancées technologiques et culturelles.

Grâce aux progrès de la chimie et de la recherche scientifique, les scientifiques ont pu percer les secrets de la structure moléculaire des polymères naturels et les reproduire en laboratoire. Cette compréhension approfondie a permis la synthèse de polymères artificiels, ouvrant la voie à une multitude d'innovations dans divers domaines tels que la médecine, l'industrie automobile, l'électronique et bien d'autres.

De nos jours, les polymères, qu'ils soient naturels ou synthétiques, constituent une classe de matériaux incontournable. Ils concurrencent de plus en plus les matériaux traditionnels, comme les métaux et les matériaux minéraux, grâce à de nombreux avantages :

Leur faible coût de production, permettant une accessibilité économique pour diverses industries et consommateurs ;

Leurs propriétés très spécifiques, notamment dans le cas des polymères fonctionnels, qui peuvent être conçus pour répondre à des besoins très précis, comme une résistance accrue, une flexibilité ou une conductivité électrique ;

Leur possibilité de recyclage, particulièrement pour les matières thermoplastiques, qui répond à une exigence croissante de durabilité et de respect de l'environnement.

1. Définition des Polymères :

Les polymères sont des macromolécules constituées de nombreuses unités de répétition, appelées monomères, reliées entre elles par des liaisons covalentes. Leur structure moléculaire diversifiée leur confère une vaste gamme de propriétés physiques et chimiques. En conséquence, ils sont largement utilisés dans l'industrie pour la fabrication d'une multitude de produits, allant des plastiques et fibres synthétiques aux adhésifs, revêtements et dispositifs médicaux[1].

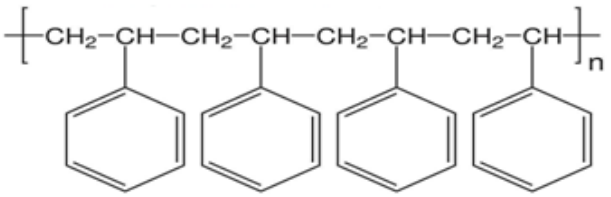


Figure 1 :polymére du polystyréne

2. Historique

Le terme polymère est utilisé pour la première fois en 1832 par Jöns Jacob Berzelius. En 1835, Henri Victor Regnault prépare du chlorure de vinyle à partir du dichlore et de l'éthylène, observant sa transformation en une poudre blanche non identifiée alors comme du PVC.
En 1863, John Wesley Hyatt et son frère, motivés par un concours offrant 10 000 dollars pour remplacer l'ivoire dans les boules de billard, inventent le celluloïd en 1870, une matière plastique issue de la nitrocellulose et du camphre [2].

En 1897, apparaît la galalithe, une corne artificielle faite de caséine et de formol. Le premier polymère entièrement synthétisé est créé en 1907 par Hendrick Baekeland : la bakélite, utilisée dans de nombreux objets domestiques.
Les années suivantes voient l'émergence de divers polymères :
1928 : Plexiglas (PMMA) par Otto Röhner.
1931 : PVC et polystyrène (PS).
1935 : Polyéthylène (PE).
1937 : Nylon.

Après la Seconde Guerre mondiale, les matériaux polymères connaissent un essor considérable grâce aux travaux de Hermann Staundinger, Carothers, et Ziegler et Natta. Vers les années 1960, le PVC commence à remplacer le verre dans les emballages alimentaires, et la pétrochimie prend le relais de la carbochimie pour la production de polymères.

En 1977, les professeurs Mac Diarmid, Heeger et Shirakawa découvrent la possibilité de moduler la conductivité électrique du polyacétylène, permettant de transformer le polymère de l'état d'isolant à celui de conducteur[3]. Aujourd'hui, les matériaux polymères sont omniprésents, remplaçant de nombreux matériaux traditionnels dans divers domaines.

3. Degré de polymérisation :

Le degré de polymérisation (DP) représente le nombre moyen de monomères dans les macromolécules d'un polymère. Il peut être calculé par le rapport de la masse moléculaire moyenneM d'une macromolécule du polymère considéré, à la masse moléculaire m du monomère correspondant

$$DP = \frac{M}{m} \quad (1)$$

Ce résultat ne peut être déterminé que statistiquement, car dans un polymère, il existe toute une distribution de tailles de chaînes macromoléculaires, fonction des conditions de polymérisation (pression, température, concentrations en monomères…). DP atteint couramment 102 à 104. Sa valeur a une grande influence sur les propriétés du polymère considéré : un DP élevé conduit à un matériau globalement plus rigide et plus résistant mécaniquement [4].

4. Types de polymères:

On distingue deux principaux types de polymères :

- Les homopolymères
- Les copolymères

4.1. Homopolymères

Les homopolymères sont des polymères constitués d'un seul type de monomère, ce qui signifie que chaque unité répétitive de la chaîne macromoléculaire est identique. Cette structure simple leur confère des propriétés bien définies et les rend faciles à synthétiser. Un exemple classique de

homopolymère est le polyéthylène[5].

4.2 Copolymères :

Les copolymères sont des polymères constitués de deux ou plusieurs types de monomères différents. Cette diversité confère aux copolymères une large gamme de propriétés et de fonctionnalités inaccessibles aux homopolymères [6].

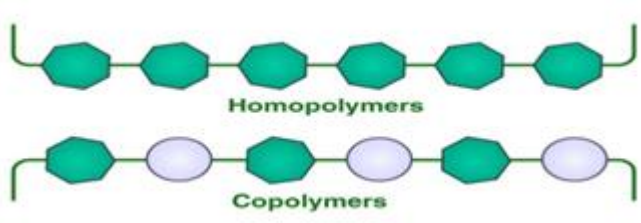


Figure 2 : Types des polymères : homopolymères et copolymères

5. Structure despolymères:

Les polymères peuvent présenter des architectures extrêmementvariables. Ils peuvent être linéaires, ramifiés ou réticulés.

5.1.Polymères linéaires : Formés par des monomères connectés en une seule chaîne longue.

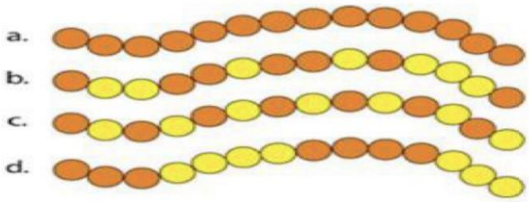


Figure 3: Polymères linéaires[7]

5.2 Polymères ramifiés : Possèdent des chaînes latérales courtes ou longues qui se ramifient àpartir de la chaîne principale

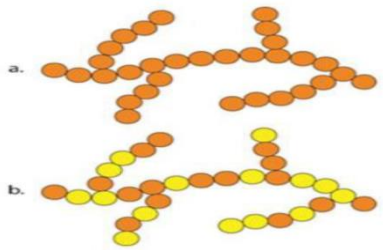


Figure 4 : Homopolymère ramifié (a) et copolymère ramifié (b)[7]

5.3. Polymères réticulés : Présentent des liaisons covalentes entre les chaînes polymériques, formant une structure tridimensionnelle.

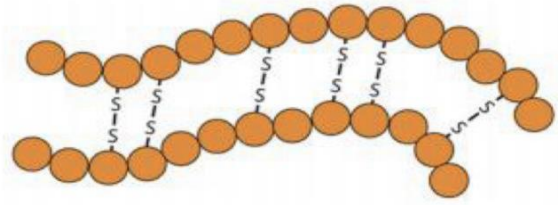


Figure 5: Polymère réticulé avec ponts di-sulfure reliant deux chaînes [7].

5.4. Polymères amorphes et polymères cristallisés :

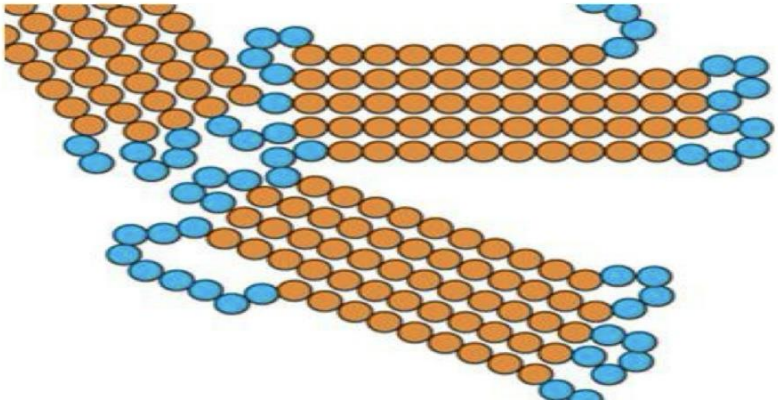


Figure 6: Représentation schématique d'un polymère semi-cristallisé[7].

Ces différentes structures influencent fortement sur les propriétés physiques et mécaniques des polymères, telles que la flexibilité, la résistance à la traction et la température de transition vitreuse.

6. Classification des Polymères

6.1 Selon l'Origine :

- **Polymères naturels :** Provenant de sources naturelles telles que les plantes et les animaux, comme la cellulose et le caoutchouc naturel.
- **Polymères artificiels :** Dérivés de polymères naturels par modification chimique, tels que le nitrate et l'acétate de cellulose, ainsi que l'ébonite (caoutchouc fortement réticulé par le soufre).
- **Polymères synthétiques :** Fabriqués par l'homme à partir de petites molécules (monomères) via des réactions de polymérisation. [8]

6.2 Selonleur nature chimique :

- **Polymères minéraux :** Composés de chaînes constituées d'un seul élément (comme le diamant et le graphite) ou de plusieurs éléments différents (comme les silicates).
- **Polymères organiques :** Représentant la classe la plus vaste, ils comprennent tous les polymères couramment utilisés. Les principaux polymères organiques de synthèse incluent les polyvinyles, les polyesters, les polyacryliques et les polydiènes.
- **Polymères mixtes :** Combinaison de polymères organiques et minéraux, ces matériaux possèdent des propriétés intéressantes, notamment une résistance thermique de 300-350 °C [9].

6.3. Selon l'architecture macromoléculaire :

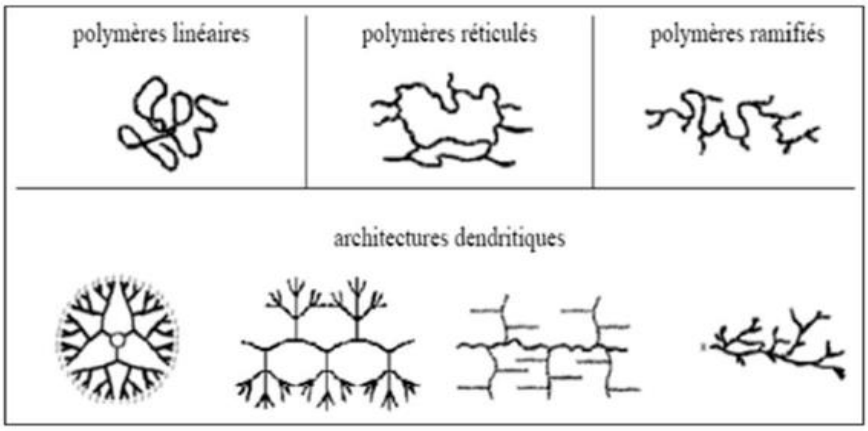


Figure 7: Classification des architectures macromoléculaires selon Tamolia.[10]

6.4 Selon leurs usages technologiques : On peut distinguer :

- **Fibres synthétiques:** on peut citer le nylon, le tergal, le coton et la soie.
- **Plastomères**: ce sont les plastiques au sens large regroupant les thermodurcissables et les thermoplastiques.
- **Elastomères**: sont par définition des polymères ayant des propriétés élastiques réversibles (allongement réversible important).

Exemple: caoutchouc naturel, polybutadiène et polyisoprène[11]

6.5 Selon le comportement thermique:

Les polymères peuvent être classés en trois types selon leurs propriétés thermodynamiques [12] :

- **Les thermoplastiques:** Les thermoplastiques ramollissent sous l'effet de la chaleur. Ils deviennent souples, malléables et durcissent à nouveau quand on les refroidit. Comme cette transformation est réversible, ces matériaux conservent leurs propriétés et ils sont facilement recyclables. Leurs polymères de base sont constitués par des macromolécules linéaires, reliées par des liaisons faibles qui peuvent être rompues sous l'effet de la chaleur ou de fortes contraintes.

Figure 8 : Plaque de thermoplastique et quelque produit. [13]

• Les thermodurcissables : Les thermodurcissables sont des plastiques qui prennent une forme définitive au premier refroidissement. La réversibilité de forme est impossible car ils ne se ramollissent plus une fois moulés. Sous de trop fortes températures, ils se dégradent et brûlent (carbonisation). Les molécules de ces polymères sont organisées en longues chaînes dans

lesquelles un grand nombre de liaisons chimiques solides ne peuvent pas être rompues et se renforcent quand le plastique est chauffé.

Figure 9 : Quelques objets thermodurcissables[13]

• **Les élastomères**. Ces polymères présentent les mêmes qualités élastiques que le caoutchouc. Un élastomère au repos est constitué de longues chaînes moléculaires repliées sur elles- mêmes. Sous l'action d'une contrainte, les molécules peuvent glisser les unes par rapport aux autres et se déformer. Pour que le matériau de base présente une bonne élasticité il subit une vulcanisation. C'est un procédé de cuisson et de durcissement qui permet de créer un réseau plus ou moins rigide sans supprimer la flexibilité des chaînes moléculaires. Les élastomères sont employés dans la fabrication des coussins, de certains isolants, dessemelles de chaussures ou des pneus .[13]

Figure 10: Quelques objets sont des élastomères[13]

6.7 Selon l'importance économique:

Pour les polymères, comme pour tout produit industriel, il existe une corrélation entre tonnage et prix. Ceci permet de distinguer trois grandes catégories de polymères commerciaux :

- Polymères largement utilisés : ventes et production élevées, prix bas, tels que : polyéthylène (PE), polypropylène (PP), polystyrène (PS), polychlorure de vinyle (PVC).
- Polymères hautes performances : volume le plus faible, prix le plus élevé.
- Polymères techniques : leurs volumes et prix se situent entre les deux premières catégories Exemples : polyamide, polyéthylène téréphtalate[14].

7. Quelques propriétés des polymères :

7.1 Propriétés chimiques :

Comme tous les composés, les propriétés chimiques des polymères sont déterminées par les fonctions qu'ils portent ou par les fonctions de leurs chaînes latérales. le polyéthylène et le polypropylène ont l'inertie chimique des hydrocarbures saturés et le polyester peut être hydrolysé par l'acide. Cependant, la grande taille des macromolécules leur confère un comportement réactionnel particulier. D'une manière générale, les réactions d'addition ou de substitution ne se produisent pas régulièrement sur les unités monomères, mais se produisent de manière aléatoire sur des modèles macromoléculaires.[15]

7.2 Propriétés physiques :

Selon la structure physique des polymères, ils peuvent être divisés en quatre états :

- État amorphe dans lequel les molécules sont disposées sans ordre prédéterminé.

Les grosses molécules s'emmêlent en boules. Ce désordre se traduit par l'absence de point de fusion, la présence d'une température de transition vitreuse qui marque le passage de l'état vitreux (où le matériau se comporte comme du verre, c'est à dire dur et cassant) à l'état caoutchouteux (où les chaînes peut glisser plus facilement), et Transparence visible, comme c'est le cas du polystyrène "cristallin" ou du poly(chlorure de vinyle)

8. Applications des polymères :

Les polymères constituent une des principales révolutions techniques du XXème siècle .ceux-ci sont utilisés pour un nombre extraordinaires d'applications à tous les échelons de la vie. (Voir Tableau.01) [16].

Tableau .1 : Les domaines d'application des polymères.

Domaines	Exemples
Le secteur de la construction	Les portes, conduites d'eau et l'assainissement, peinture des murs des plan chaires et des plafonds, revêtement du sol.
Industrie de l'emballage	Bouteilles, pots de yaourt, boites aux lettres, problème thermique, gainage films vidéo.
Médecine et santé	Poches de sang, gants, lentilles, verres de lunettes, les organes artificiels, Seringues, industrie dentaire, prothèses, outils de chirurgie
Articles ménagers	Tupperware, poubelles, seaux, vaisselle.
Matériel électrique et électronique et les communications	Tension d'isolement, isolation, laveuses, ordinateurs et les caméras, Radio et télévision, téléphone.
Industrie au Tommobile	Système vitre claire, Tapie, carrosserie, optiques, planches de bord, habillage intérieur, Façades de téléviseurs, gainage de câbles.
Industrie textile	Vêtement, fibres textiles naturelles et synthétiques, sacs, skaï, similicuir, non tissés.
Dans le domaine agricole	Colles, vernis, mousses
Sports-loisirs	Certaines parties des bâtiments, des avions et des bateaux, DVD, bandes magnétiques, Files de pêche et les cordons utilisés, piscines, coques de bateaux.
Industrie chimique	Tuyauterie, cuves, revêtements.
Le domaine alimentaire	Procédés de fabrication d'aliments, emballages (bouteilles, pots de yaourt, briques de lait, boites à oeufs) ou industriel (flacons de détergent, sachets et sacs, casiers)…
Le matériel de maison	Meuble, vaisselle, accessoires… .

9. L'indice de réfraction des polymères :

L'indice de réfraction d'un matériau, y compris les polymères, est une mesure de la vitesse à laquelle la lumière se déplace à travers ce matériau par rapport à la vitesse de la lumière dans le vide. Il est défini comme le rapport entre la vitesse de la lumière dans le vide et la vitesse de la lumière dans le matériau.

L'indice de réfraction est une propriété optique fondamentale qui dépend de la structure moléculaire du polymère et de la longueur d'onde de la lumière utilisée. Pour les polymères, cette propriété est cruciale dans diverses applications telles que les fibres optiques, les lentilles et les matériaux transparents.

10. Les facteurs influençant l'indice de réfraction des polymères incluent :

La densité du polymère : Plus la densité est élevée, plus l'indice de réfraction est généralement élevé.

La polarisation des liaisons chimiques : Les polymères avec des liaisons chimiques fortement polarisables auront un indice de réfraction plus élevé.

La longueur d'onde de la lumière : L'indice de réfraction varie avec la longueur d'onde, phénomène connu sous le nom de dispersion.

- **Équation de Calcul**

L'indice de réfraction (n) est calculé à l'aide de la relation suivante :

$\eta = \frac{c}{v}$ où :η est l'indice de réfraction,

C :est la vitesse de la lumière dans le vide ($\approx 3\times10^8$m/s),

v :est la vitesse de la lumière dans le matériau.

Les polymères comportant des groupements polaires ou polarisables (PVC, PS) ont des indices de réfraction plus élevés que les polymères à structures fortement symétriques ou peu polarisables (PE, PTFE). Il est possible de prévoir approximativement l'indice de réfraction sur la base des contributions des groupes à la réfraction molaire : chaque groupe (C-H, C-F, C=0) apporte sa propre contribution à la réfraction molaire et la réfraction molaire s'obtient par simple additivité.L'indice de réfraction de la plupart des polymères à l'état isotrope est situé entre 1.3 et 1.7. [17]

Pour les polymères, on utilise souvent des relations empiriques ou des modèles théoriques basés sur leur structure chimique pour estimer leur indice de réfraction. Ces modèles permettent une meilleure compréhension des interactions microscopiques entre les composants du polymère et la lumière, ce qui peut être difficile à saisir uniquement par des méthodes expérimentales

Références Bibliographiques

[1] CNR « les emballages plastiques de la fabrication à la valorisation », Rapport de l'association Cercle national de recyclage, avril 1999.

[2] U. Schulte, in: Proceedings of Plastics Pipes XIII, Washington, DC, USA, 2006.

[3]https://www.memoireonline.com/01/14/8519/m_Les-polymeres-dans-l-industrie3.html

[4] Aide-mémoire science des matériaux livre écrit par Michel Dupeux.

[5] P. Weiss, La chimie des polymères, «Support de Cours (Version PDF)», Université Medicale Virtuelle Francophone, Société Francophone de Biomatériaux Dentaires (2010).

[6] M.Daci, « Analyses et dégradation de polystyrene contenant l'azote et le soufre », mémoire magister, Université de Biskra, Algérie (2007).

[7]http://dspace.univbiskra.dz:8080/jspui/bitstrream/123456789/4326/3/chapitre%20%2001.pdf

[8] http://www.cndp.fr/portailsdisciplinaires/fileadmin/userupload/Phsiquechimie/Ressources-RNSTL/Polymeres.pdf, Date de consultation : 24/01/2020(13.42).

[9] http://mms2.ensmp.fr/mat-paris/elaboration/polycop/ch-4- Elab-polymeres.pdf, Date de consultation : 19/01/2020 (9h.45min).

[10] F. Annab, K. Zdira, « Etude de la polymérisation cationique et modification du 4-vinyle chlorure de benzyle », mémoire de master, université Larbi Ben M′Hidi Oum El Bouaghi Algérie (2015).

[11] D.Tabli, « Etude de Nanocomposite de Semiconducteur TiO2 Dispersé dans la Matrice de Polyépoxyde» , , mémoire de master, Université de Bouira, Algérie (2015).

[12] N. Guemaz, « Préparation et étude de la stabilité thermique des polyurethanes », mémoire de magister, l'université Ferhat Abbas Setif, Algérie (2012).

[13] l.Djoudi, « Etude de la stabilisation du polychlorure de vinyle (PVC) et les phénomènes de dégradation », mémoire magister, Université de Biskra, Algérie (2007).

[14] M.Bensaid, « Etude des polymères par dynamiques moléculaire», thèse doctorat, université d'Oran des sciences et de la technologie Mohamed Boudiaf, Algérie (2015).

[15] :A. Ait-zoura Ali, S. Chikbouni, « Modélisation de l'influence d'une charge minérale et d'un cycle de recyclage sur la diffusion de l'irganox 1076 dans un emballage alimentaire à base de polyéthylène haute densité», mémoire de master, Université Abderrahmane Mira – Bejaia, Algérie (2017).

[16] S.Ithri, L.Teklal, « caractérisation d'une géomembrane soudée utiliser pour l'etanchéité d'une station d'epuration des eaux usées», mémoire de master, Université Mouloud Mammeri De TiziOuzou, Algérie (2016)

[17] Y. Bendahma, « Etude de la diffusion des molécules de bas poids moléculaire dans des réseaux de polymères acryliques», mémoire de magister, Université Abou BekrBelkaid Tlemcen, Algérie (2011).

Les Relation Quantitative Structure –Activité /Propriété (QSAR /QSPR)

L'idée que l'activité biologique d'une molécule est intimement liée à sa structure chimique n'est pas récente. Dès le milieu du 19ème siècle, des observations précoces ont établi des corrélations entre des structures chimiques simples et certaines propriétés physiques. Cependant, c'est à partir de la fin du 20ème siècle que les Relations Quantitatives Structure-Activité (QSAR) ont connu un développement formel, ouvrant la voie à une compréhension plus profonde des relations entre la structure et l'activité biologique suivant un modèle mathématique :

[Activité biologique] = f [Descripteurs]

L'acronyme **QS/XR** (où X peut désigner une activité biologique (A), une propriété physico-chimique (P), un indice de rétention (R), une toxicité (T), etc.) est couramment utilisé lorsqu'un effet spécifique est relié à la structure moléculaire. Notamment on utilise les QSPR pour désigner la relation entre la propriété (l'indice de réfraction) et la structure moléculaire.

1. Principe :

L'étude QSPR (Quantitative Structure PropertyRelationships) consiste à établir une relation mathématique quantitative entre une propriété mesurée, et des descripteurs moléculaires qui représentent la structure chimique d'une série de composés similaires. Cette relation est obtenue en utilisant des méthodes statistiques et permet d'analyser les données structurales pour identifier les facteurs moléculaires qui influencent la propriété étudiée [18].

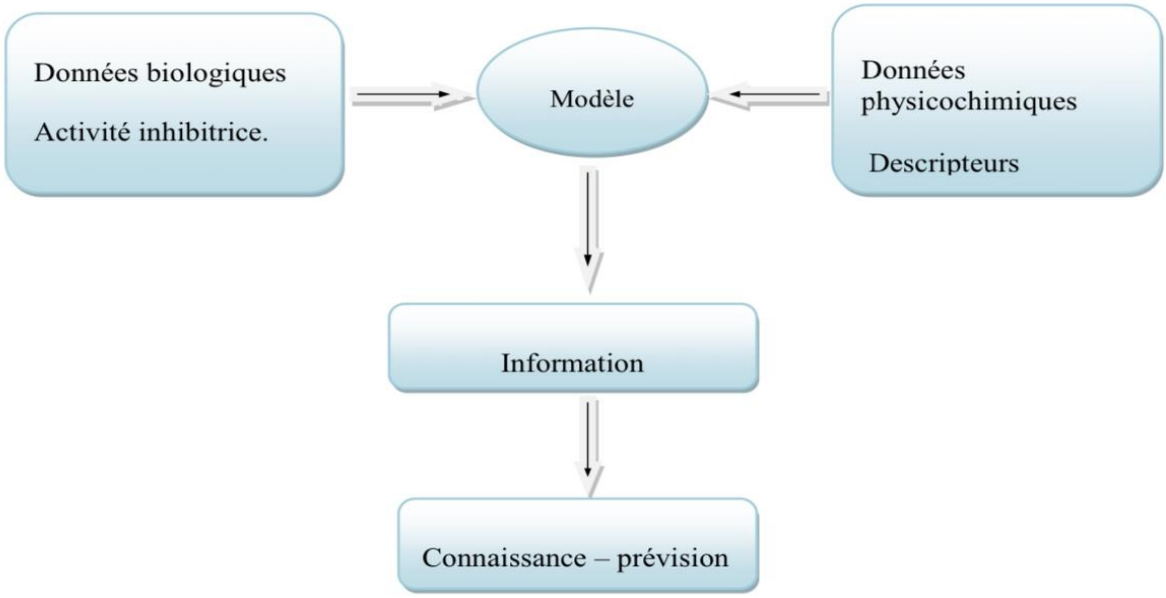


Figure 11: présentation de la méthodologie QSPR [19].

2. Objectifs :

Les objectifs pratiques d'une étude QSPR sont nombreux et ces techniques sont largement utilisées dans de nombreuses situations. Le but des études QSPR comprend donc ce qui suit [20] :

• Prédire les propriétés physico-chimiques par des moyens rationnels.

• Comprendre les mécanismes d'action au sein d'une série de produits chimiques.

• Économies sur les coûts de développement de produits.

• Les prévisions pourraient réduire la nécessité de tests sur les animaux longs et coûteux.

3. Méthodologie générale des études QSPR :

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

a- Constituer une base de données à partir des mesures expérimentales fiables de la propriété de chaque composé.

b- Sélectionner les descripteurs en relation avec la propriété ou l'activité étudiée.

c- Diviser cette base de données, en une série d'apprentissage (training set) qui contient généralement les 2/3 de la base de données et une série de test (test set) constituée par le 1/3 restant.

d- Etablir des modèles mathématiques en utilisant la série d'apprentissage.

e- Caractériser les modèles élaborés par leurs indices de validation internes et vérifier leur robustesse par un test de randomisation de la variable dépendante Y (réponse).

f- Valider les modèles élaborés en utilisant la série de test et calculer leurs paramètres statistiques de validation externe.

g- Elaborer le domaine d'applicabilité du modèle retenu.

h- Explorer et exploiter les modèles validés pour comprendre les mécanismes et les modes d'action [21].

Le schéma fonctionnel de la figure 1 représente la stratégie générale d'une étude QSPR.

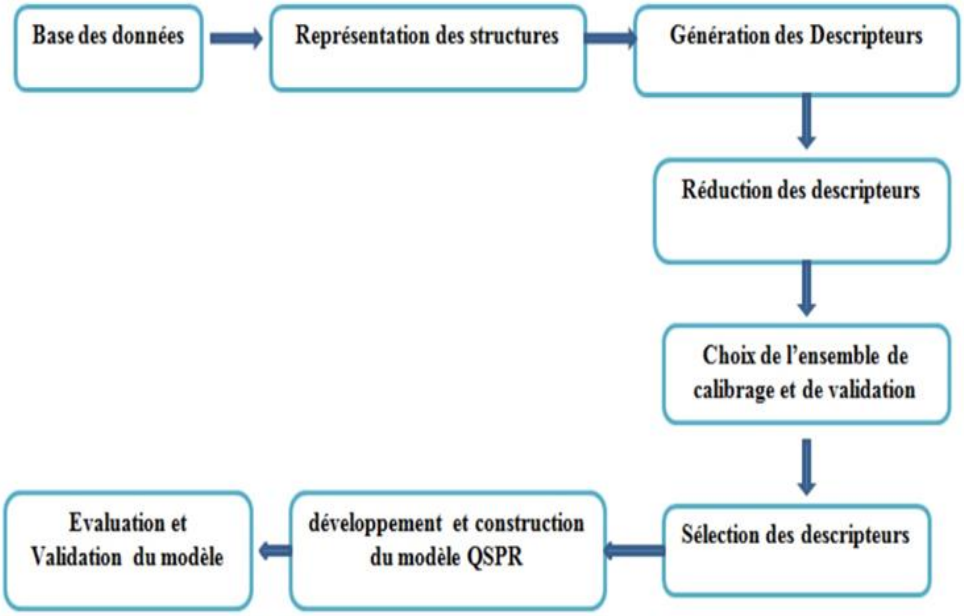


Figure 12 : La stratégie générale d'une étude QSPR

3.1 Élaboration du modèle :

3.1.1.Base de données :

La sélection de la base de données expérimentale initiale constitue une étape cruciale dans le développement de modèles QSPR performants et fiables. En effet, la qualité et la cohérence des données utilisées pour entraîner et évaluer les modèles influencent directement leur précision et leur généralisabilité[22].

3.1.2. Représentation des structures.

Les structures peuvent être représentées sous forme de graphique 2D avec suppression de l'hydrogène. Des logiciels sont disponibles pour dessiner cette structure 2D. On peut citer ChemDraw qui utilise le nom IUPAC (International Union of Pure and Applied Chemistry) [23] pour la représentation des structures. Et le logiciel de modélisation moléculaire HyperChem qui peut générer une structure tridimensionnelle (3-D) obtenue à partir de structures 2-D qui sont nécessaires à la génération de descripteurs géométriques.

3.1.3GénérationLes descripteurs moléculaires :

Au cours des dernières décennies, de nombreuses recherches ont été menées pour trouver la meilleure manière de représenter l'information contenue dans la structure moléculaire, ainsi que dans les structures elles-mêmes, sous forme de nombres réels appelés descripteurs.

Donc les descripteurs réalisent un codage numérique de l'information chimique en vecteurs réels.Actuellement, on en dénombre plus de 5000, quantifiant diverses caractéristiques physico-chimiques ou structurelles de molécules. Ces descripteurs, qu'ils soient empiriques ou

calculés, jouent un rôle crucial dans la prédiction des propriétés moléculaires sans nécessiter la synthèse préalable des molécules, ce qui est un objectif fondamental de la modélisation. [24].

De nombreux logiciels, tels que DRAGON, sont utilisés pour calculer ces descripteurs, chacun offrant une gamme variée de méthodes et de types de descripteurs, allant des autocorrélations 2D aux descripteurs de 3D-MoRSE en passant par les WeightedHolistic Invariant Molecular (WHIM) et GETAWAY (GEometry, Topology, and Atom-WeightsAssembly) [25].

Les descripteurs moléculaires sont fréquemment classés par rapport à la dimensionnalité de la représentation moléculaire sur laquelle ils sont calculés : on parlera alors de descripteurs 0D, 1D, 2D, ou 3D [26.].

- **Les descripteurs 1D :** sont accessibles à partir de la formule brute de la molécule (par exemple C_6H_6O pour le phénol), et décrivent des propriétés globales du composé, telles que sa composition ou sa masse molaire. Cependant, ils ne permettent pas de distinguer les isomères de constitution.
- **Les descripteurs 2D** sont calculés à partir de la formule développée de la molécule et incluent les indices constitutionnels et topologiques.Ces derniers fournissent des informations sur la taille, la forme globale et les ramifications de la molécule.On trouvera dans cette catégorie les descripteurs suivants :
  - **Les indices constitutionnels** caractérisent les différents composants de la molécule. Il s'agit par exemple du nombre de liaisons simples ou multiples, du nombre de cycles.
  - **Les indices topologiques** dérivés de la structure 2D des molécules, offrent des perspectives sur leur taille, leur forme globale et leurs ramifications. Parmi les plus couramment utilisés figurent l'indice de Wiener [27], l'indice de Randić [28], l'indice de connectivité de valence de Kier-Hall [29] et l'indice de Balaban [30]. Par exemple, l'indice de Wiener caractérise à la fois le volume moléculaire et la complexité de la ramification : il est défini comme la somme de toutes les distances topologiques entre les différentes paires d'atomes de la molécule. Quant à l'indice de Randić, il est largement employé et peut être interprété comme une mesure de la surface moléculaire accessible au solvant.

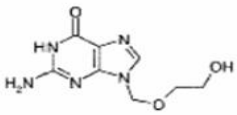
Bien que ces descripteurs 2D fournissent des informations précieuses sur les propriétés physiques des molécules, ils se révèlent souvent insuffisants pour expliquer certaines propriétés ou activités complexes. Par conséquent, pour aborder des questions plus complexes, notamment celles liées à la biologie, il est nécessaire de recourir à des

descripteurs dérivés de la structure 3D des molécules, rendus accessibles grâce au développement des techniques instrumentales et des nouvelles méthodes théoriques**.**

- **Les descripteurs 3D** d'une molécule sont évalués en fonction des positions relatives de ses atomes dans l'espace, et ils fournissent des informations détaillées sur sa structure. Leur calcul implique généralement de déterminer la géométrie 3D de la molécule, souvent par le biais de modèles moléculaires empiriques ou ab initio. Bien que cette démarche soit chronophage, les informations ainsi obtenues sont cruciales pour la modélisation des propriétés ou des activités qui dépendent de la structure tridimensionnelle.

Il existe plusieurs catégories importantes de descripteurs 3D :

- **Les descripteurs géométriques**, tels que le volume moléculaire, la surface accessible au solvant et le moment principal d'inertie, sont parmi les plus significatifs [31].
- **Les descripteurs électroniques** permettent de quantifier différentes interactions inter et intramoléculaires, influençant fortement l'activité biologique des molécules. Leur calcul nécessite souvent la recherche de la géométrie minimisant l'énergie stérique, impliquant souvent des méthodes de chimie quantique. Des exemples incluent les énergies de la plus haute orbitale moléculaire occupée (HOMO) et de la plus basse vacante (LUMO), ainsi que le moment dipolaire, le potentiel d'ionisation et d'autres paramètres énergétiques relatifs à la molécule [32].
- **Les descripteurs spectroscopiques** reposent sur des mesures spectrales, telles que les fonctions d'onde vibrationnelles, qui fournissent des informations sur la structure et la conformation de la molécule. Les spectres infrarouges peuvent être obtenus expérimentalement ou par simulation théorique après avoir déterminé la géométrie optimale de la molécule. Ces spectres sont ensuite convertis en vecteurs de descripteurs de taille fixe. Par exemple, le descripteur EVA est obtenu à partir des fréquences de vibration de chaque molécule, tandis que les descripteurs de type MoRSE sont calculés à partir d'une simulation du spectre infrarouge, en se basant sur les intensités théoriques de diffraction des électrons [33].

	Représentation Type	Descripteurs Type
1D	C8H10N5O3	Masse moléculaire Nombre d'atomes
2D		Fragments Indices Topologiques Connectivité
3D		Surface Moléculaire Volume Moléculaire Energie d'interaction

➢

Figure 13: Quelques exemples de descripteurs et leur classification en 1D, 2D et 3D

3.1.4 Réduction des descripteurs

Les descripteurs chimiques présentent souvent des corrélations entre eux, fournissant ainsi des informations similaires. L'objectif central de l'analyse statistique consiste à démêler ces descripteurs, à identifier ceux qui sont significativement corrélés à la variable cible - ceux qui révèlent un signal - par opposition à ceux qui ne le sont pas, générant du bruit. En outre, cette analyse permet de détecter les descripteursqui présentent des corrélations entre eux, afin de ne conserver que les plus pertinents et ainsi réduire la redondance d'informations.

Dans le langage statistique, la caractéristique étudiée est désignée par différents termes : variable cible, variable dépendante ou variable Y. C'est cette caractéristique que l'on s'efforce d'expliquer et d'optimiser. Quant aux descripteurs, souvent d'ordre chimique, ils sont appelés variables explicatives, variables indépendantes ou variables X.

L'analyse statistique permet de déterminer et de quantifier les corrélations entre les descripteurs et la variable cible, ainsi que de mettre en évidence la contribution relative de chaque descripteur à l'explication globale de la propriété. Le modèle statistique établit alors une équation décrivant la valeur de la variable cible en fonction de la somme pondérée des valeurs des descripteurs.

Lesprincipaux outils statistiques pour obtenir un modèle sont :

- La régression linéaire multivariée (MultivariateLinearRegression - MLR) ;
- La régression en composantes principales (Principal Component Regression - PCR) ;
- La régression des moindre carrés partiels (Partial Least Squares - PLS) [34] ;
- Les réseaux de neurones artificiels (Artificial Neural Network - ANN) [35, 36].

Le choix de la méthode dépend principalement de la question qui est posée et de lanature des données à traiter.

3.1.5. Régression linéaire multiple (MLR)

Dans notre travail, nous avons opté pour la régression linéaire multiple (MLR). Il est indéniable que la régression linéaire représente l'une des méthodes statistiques les plus couramment employées. Traditionnellement, on différencie la régression simple (avec une seule variable explicative) de la régression multiple (avec plusieurs variables explicatives), bien que leur cadre conceptuel et leurs méthodes de calcul soient identiques [37]. La régression linéaire multiple est ainsi considérée comme l'approche la plus élémentaire en modélisation.

Elle consiste à rechercher une équation linéaire par rapport à ses paramètres, reliant la variable à modéliser y au vecteur d'entrées x = {Xk, k = 1...n}. L'équation linéaire recherchée est de la forme :

$$Y_i = A_0 + A_1 X_{i,1} + A_2 X_{i,2} + \ldots + A_p X_{i,p} + \varepsilon_i \; ; \; i = 1,..,n$$

Où (A0, A1, A2) Sont les paramètres/coefficients du modèle que l'on veut estimer à l'aide desdonnées. ε : quantifie les écarts entre les valeurs réellement observées et les valeurs prédites par lemodèle.

3.1.6 Constitution des jeux d'apprentissage et de test

La constitution des jeux d'apprentissage et de test est une étape cruciale dans le processus de modélisation. Le modélisateur dispose généralement d'un seul ensemble de structures avec les propriétés associées, nécessitant ainsi une division en un jeu d'apprentissage et un jeu de test pour évaluer le pouvoir de prévision du modèle.

Plusieurs méthodes de sélection d'échantillons sont disponibles, parmi lesquelles :

- **La Sélection aléatoire des échantillons :**

La méthode la plus simple consiste à prélever un nombre déterminé de composés de façon aléatoire dans l'ensemble de départ. Les composés restants constituent le jeu d'apprentissage, et cette opération est répétée pour obtenir différents couples de jeux de tests et d'apprentissage [38].

4 .Qualité des modèles QSPR :

La validation des modèles constitue une étape critique dans le processus de QSAR. Étant donné que le modèle QSAR est le résultat d'une analyse statistique, son interprétation et son application doivent se faire dans un cadre précis correspondant au domaine couvert par l'analyse. Toute extrapolation en dehors de ce cadre nécessite une grande prudence, car elle devient de plus en plus risquée à mesure que l'on s'éloigne du domaine d'application initial. Afin de minimiser les erreurs, tant lors de la validation que lors de l'application, il est essentiel

d'établir clairement les limites du modèle : sa robustesse, son pouvoir de prédiction et sa pertinence dans l'espace chimique concerné.

4.1. L'ajustement :

Une méthode statistique, telle que la Régression Linéaire Multiple (RLM), fournit un premier indicateur de la qualité du modèle, à savoir le coefficient de détermination multiple R^2. Ce coefficient mesure l'écart entre les valeurs observées et les valeurs prédites par le modèle pour les structures présentes dans le jeu d'apprentissage. Un R^2 proche de 1 indique que le modèle s'ajuste bien aux données expérimentales, étant capable de les reproduire. En revanche, un R^2 proche de 0 indique un faible niveau de corrélation entre le modèle et les observations.

R^2= Le coefficient de détermination multiple :

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

L'analyse de régression (multi) linéaire a été réalisée avec le logiciel MOBYDIGS [39] en utilisant la méthode des moindres carrés ordinaires.

La qualité de l'ajustement a été évaluée par le coefficient de détermination, R^2, et l'écart quadratique moyen calculé sur l'ensemble de calibration :

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

yi et yˆ i étant les valeurs mesurées et prédites de la variable dépendante.

4.2. Le pouvoir de prévision interne :

Le pouvoir de prévision interne est évalué à travers la validation croisée, qui nous fournit des informations essentielles sur la capacité d'un modèle à prédire avec précision. Ce pouvoir de prévision est qualifié de "interne" car il est estimé à partir des mêmes structures utilisées pour construire le modèle. Lorsque plusieurs modèles présentent des capacités prévisionnelles équivalentes, celui comportant le moins de composantes est souvent préféré. En effet, un modèle moins complexe est généralement plus robuste.

La validation croisée implique de réitérer le processus de modélisation en excluant un ou plusieurs éléments du jeu d'apprentissage, puis en utilisant le modèle ainsi obtenu pour prédire l'activité de ces éléments omis. Cette opération est répétée jusqu'à ce que chaque élément du jeu ait été exclu au moins une fois. La procédure la plus courante est la LOO

(LeaveOne Out), où un seul composé est omis avant chaque itération et son activité est prédite par le modèle correspondant.

Le coefficient de corrélation Q^2 entre les activités prédites et les activités observées reflète le pouvoir de prévision interne du modèle. Une valeur de Q^2 proche de 1 indique une excellente capacité de prévision, tandis qu'une valeur proche de zéro ou négative invalide le modèle. Pour être considéré comme acceptable, le pouvoir de prévision interne doit généralement dépasser 0,6.Une variante plus robuste de cette méthode consiste à exclure plusieurs composés simultanément (LeaveMany Out - LMO).

$$Q^2_{LOO} = \frac{SCT - PRESS}{SCT}$$

"La somme des carrés des erreurs de prédiction", désignée par l'acronyme PRESS (pour : PredictiveResidualSum of Squares) :

$$\text{PRESS} = \sum_{i=1}^{n} \left(y_i - \hat{y}_{(i)}\right)^2$$

SCT : la somme des carrés totaux

$$\text{SCT} = \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2$$

La racine de l'erreur quadratique moyenne de prédiction

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

Ce paramètre revêt une importance cruciale, mais il n'est pas en soi suffisant. En plus de cela, il est nécessaire de déterminer le nombre optimal de composantes, également appelées tendances, à intégrer au modèle. Il est également souhaitable de connaître l'intervalle de confiance des coefficients associés à chaque descripteur. Ces informations sont obtenues respectivement par les méthodes de rééchantillonnage (bootstrapping) et de validation croisée (cross-validation).

4.3. Test de randomisation :

Le pouvoir de prévision interne Q^2 a tendance à être surestimé par la procédure LOO. Une valeur élevée de cet indice peut résulter d'une corrélation due au hasard ou de la redondance des structures lorsque les différences entre les composés du jeu d'apprentissage sont minimes, phénomène connu sous le nom d'auto-corrélation. Pour contrôler la robustesse d'un modèle, l'approche la plus couramment utilisée est le test de randomisation des réponses (Y-

randomization test). Dans cette méthode, les valeurs de la variable cible sont redistribuées de manière aléatoire sur l'ensemble du jeu d'apprentissage, et un nouveau modèle est dérivé. Cette opération est répétée plusieurs fois (généralement 100 fois), et si la moyenne des indices R^2 et Q^2 reste élevée, cela indique qu'aucun modèle QSAR acceptable ne peut être obtenu par cette méthode statistique sur ce jeu de données.

4.4. Le pouvoir de prévision externe :

Des valeurs élevées pour les indices internes R^2 et Q^2 sont nécessaires mais pas suffisantes pour valider la qualité d'un modèle. Une validation véritable ne peut être obtenue que par l'évaluation d'un jeu de données externe qui n'a pas été utilisé pour construire le modèle. On peut alors calculer le pouvoir de prévision externe, noté Q^2_{EXT}.

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

avec yi et (ŷi) les valeurs observées et calculées par validation externe sur le jeu de test et ȳtrlavaleur moyenne sur l'ensemble des observations disponibles d'apprentissage.

Il a été démontré sur plusieurs jeux de données qu'il n'existe aucune corrélation entre Q^2 etQ^2_{EXT} [40-41]. La validation par un jeu de test externe est donc une étape essentielle de lavalidation d'un QSAR.

4.5. Domaine d'Application (DA) des modèles :

Le domaine d'application (DA) représente une zone théorique dans l'espace délimitée par les descripteurs du modèle et la réponse modélisée, où l'on attend ce qu'un modèle QSRR donné puisse générer des prédictions fiables. Cette zone, influencée par la composition des composés dans l'ensemble de calibration, peut être caractérisée de différentes manières. Dans cette étude, l'évaluation de la structure du DA s'est appuyée sur l'approche des leviers, définie par les éléments diagonaux de la matrice H. Cette matrice, en permettant une simple multiplication, facilite la transition du vecteur y au vecteur ŷ. L'élément diagonal hii est ainsi défini.

$$h_{ii} = x_i \left(X^T X \right)^{-1} x_i^T$$

Où : xi est le vecteur-ligne des descripteurs du composé i, et X la matrice du modèle déduitedes valeurs des descripteurs de l'ensemble de calibration ; l'exposant T désignant le vecteur(ou la matrice) transposé(e).

L'élément hii évalue l'impact de l'observation i sur les estimateurs obtenus par la méthode des moindres carrés. Un point levier correspond à une observation ayant une influence significative sur ces estimateurs. En pratique, une observation i est identifiée comme un point levier si :

$$h_{ii} > h^* \quad \text{Où} \quad h^* = 3(p+1)/n$$

Où n est le nombre de composés dans l'ensemble de calibration, p est le nombre de descripteurs dans le modèle.

Le diagramme de Williams, qui illustre les résidus de prédiction standardisés par rapport aux valeurs des leviers hii, a été employé pour une détection graphique. Il permet à la fois d'identifier les valeurs aberrantes en X (les points leviers) et les valeurs aberrantes en Y (les résidus standardisés supérieurs à 3 unités d'écart-type, 3s en valeur absolue).

Références

[18]. S Chtita, Modélisation de molécules organiques hétérocycliques biologiquement actives par des méthodes QSAR/QSPR. Recherche de nouveaux médicaments, 2017

[19]. Y.Saihi , Etude de la relation quantitative structure-activité inhibitrice des enzymes hydrolytiques : cas des alpha-glucosidases, thèse doctorant , université badjimokhtar – Annaba , 2015.

[20] T. P. J. L. M. T. C. Mark T. D. Cronin, Recent Advances in QSAR Studies: Methods and Applications, Page: 4.

[21] A. Fortuné, «Techniques de Modélisation Moléculaire appliquées à l'Etude et à l'Optimisation de Molécules Immunogènes et de Modulateurs de la Chimiorésistance,,» Université Joseph-FourierGrenoble I, , 2006.

[22]. Crum-Brown A.; & Frazer T., 1868. On the connection between chemical constitution and physiological action. Transactions of the Royal Society of Edinburgh.Vol. 25(1).151-203.

[23]. Cramer C.J. Essentials of Computational Chemistry: Theories and Models. John Wiley andSons, New York, 2002.

[24]. Gonzalez-Diaz H.; Olazabal E.; Santana L.; Uriarte E.; Gonzalez-Diaz Y. & Castanedo N.,2007. QSAR study of anticoccidial activity for diverse chemical compounds: Prediction and experimental assay of trans-2-(2-nitrovinyl)furan. Bioorganic & Medicinal Chemistry, Vol. 15 (2) 962-968.

[25]. Wiener H., 1947. Structural determination of paraffin boiling points. Journal of Chemical Information and Computer Sciences, Vol. 69 (1) 17-20.

[26]. Du-Cuny L, Huwyler J, Wiese M, Eur. Kansy M, Computational aqueous solubility prediction for drug-like compounds in congeneric series. J. Med. Chem. 2008, 43, 501..

[27]. Kier L.B. & Hall L.H., 1976. Molecular connectivity in chemistry and drug research. New-York : Academic Press.

[28]. Balaban A.T., 1982. Highly discriminating distance-based topological index. Chemical Physics Letters. Vol. 89 (5) 399-404.

[29]. Heritage T.W. & EVA., 1998 . A novel theoretical descriptor for QSAR studies. Perspectives in Drug Discovery and Design, 9-11 (0), p. 381-398.

[30]. Schuur J.H.; Selzer P. & Gasteiger J., 1996. The coding of the three-dimensional structure of molecules by molecular transforms and its application to structure-

spectracorrelations and studies of biological activity. Journal of Chemical Information and Computer
Sciences, 36 (2) 334-344.
[31]. Scotti M.T; Fernandes M. B.; Ferreira M. J. P.&Emerenciano V,. P. 2007.Quantitative structure-activity relationship of sesquiterpene lactones with cytotoxic activity.Bioorganic& Medicinal Chemistry. Vol 15 (8): 2927-2934.
[32].Vapnik V.N., 1979. Estimation of dependences based on empirical data. Nauka ed. Moscow, traduction anglaise : Springer Verlag, B., 1982.
[33]. Vapnik V. &Chervonenkis A., 1971. On the uniform convergence of relativerequencies of events to their probabilities. Theory of Probability and its Applications. Vol. 16 (2). 264-280.
[34]. Monari G. & Dreyfus G., 2002. Local overfitting control via leverages. Neural Computation. Vol. 14 (6), 1481-1506.
[35] M. Lejeune, Statistiques : la théorie et ses applications, Springer-Verlag, Paris, 2004
[36]. Boser, B.E., Guyon, I.M., &Vapnik, V.N. A. 1992. Training algorithm for optimal margin classifiers. 5th Annual ACM Workshop on COLT, Pittsburgh, PA: ACM Press,p. 144-152.
[37].Vapnik V.N., 1979. Estimation of dependences based on empirical data. Nauka ed. Moscow, traduction anglaise : Springer Verlag, B.
[38]. Vapnik V. &Chervonenkis A., 1971. On the uniform convergence of relative frequencies of events to their probabilities. Theory of Probability and its Applications.Vol.16 (2). 264-280.
[39].Kennard R. W. & Stone L. A., Computer aided design of experiments, Technometrics, 1969. Vol. 11 (1) 137-148.
[40]. Dantas Filho, H. A.; HarropGalvao, R. K.; Ugulino Araujo, M. C.; Da Silva, E. C.; Bezerra Saldanha, T. C.; Jose, G. E.; Pasquini, C.; Raimundo, I. M.; Rodrigues Rohwedder, J. J. 2004. A strategy for selecting calibration samples for multivariate modeling. Chemometrics and Intelligent Laboratory Systems. Vol.72 (1) 83-91.
[41].de Groot, P. J.; Postma, G. J.; Melssen, W. J. &Buydens, L. M. C., 1999. Selecting a representative training set for the classification of demolition waste using remote NIR sensing. Analytica Chimica Acta. Vol. 392 (1), 67-75.

1. Analyse MLR :

La régression linéaire multiple (MLR) est l'une des méthodes de modélisation les plus utilisées en QSPR. Dans l'étude actuelle, la MLR a été utilisée pour la prédiction de l'indice de réfraction des polymères. Dans l'analyse MLR, les descripteurs dans l'équation de régression doivent être indépendants, et le nombre de composés dans les échantillons doit être au moins cinq fois plus grand que le nombre de descripteurs. Les descripteurs codant des informations structurales significatives sont utilisés pour représenter les caractéristiques de l`indice de réfraction des composés. La meilleure équation est sélectionnée sur la base du plus haut coefficient de corrélation multiple (R^2). Différents types de descripteurs ont été générés pour chaque molécule par le logiciel DRAGON. La valeur des descripteurs pour les composés a été calculée et plus de 3000 descripteurs moléculaires ont été calculés pour chaque molécule. Mais quatre descripteurs ont été sélectionnés pour chaque composé dans l'ensemble de données pour refléter les propriétés des composés. 41 (70 %) polymères ont été utilisés dans l'ensemble d'entraînement et 16 polymères (30 %) ont été utilisés dans l'ensemble de prédiction.

2. Modèle relatif

Le tableau(2) ci-dessous présente les descripteurs moléculaires intervenant dans le modèle choisis avec une brève définition. Les valeurs des descripteurs sont réunies dans le **tableau 3**

Tableau 2.Définition des descripteurs intervenant dans le modèle

Descripteur	Définition
Mv	Volume moléculaire moyen des atomes de van der Waals
Me	Électronégativité moyenne des atomes selon Sanderson
Mor15p	3D-MoRSE - signal 15 / pondéré par les polarisabilités atomiques
BAC	Indice centré de Balaban

Tableau 3. Noms des valeurs des descripteurs utilisés dans le modèle.

N°	NOM	Mv	Me	S3K	Mor15p	BAC
1	Poly(trifluoroethylmethacrylate	0.57	1.09	5.617	0.18	45
2	Poly(vinylethylether	0.53	0.99	3.701	-0.023	9
3	Poly(vinyl n-butyl ether	0.53	0.98	5.701	-0.041	13
4	Poly(3-butoxypropylene oxide	0.53	0.99	4.427	0.13	7
5	Poly(vinyl n-pentyl ether	0.53	0.98	6.914	-0.039	16
6	Poly(3-hexoxypropylene oxide	0.53	0.97	2.263	-0.059	16
7	Poly(vinyl n-hexyl ether	0.53	0.98	7.701	-0.031	17
8	Poly(vinyl n-octyl ether	0.53	0.97	9.701	-0.017	21
9	Poly(vinyl 2-ethylhexyl ether	0.53	0.97	6.114	-0.195	27
10	Poly (vinyl n-decyl ether	0.53	0.97	11.701	-0.026	25
11	Poly(t-butylmethacrylate	0.55	0.99	4.945	0.155	44
12	Poly(4-methyl-1-pentene	0.53	0.96	5.105	-0.098	14
13	Poly(n-butyl acrylate	0.55	1	5.244	0.062	21
14	Poly(vinylacetate	0.57	1.03	4.797	0.059	14
15	Poly(vinyl formate	0.58	1.05	3.377	-0.032	9
16	Polyethylene	0.53	0.96	1.45	-0.025	4
17	Poly(2-fluoroethyl methacrylate)	0.56	1.03	3.859	0.151	25
18	Polyn-hexylmethacrylate)	0.55	0.99	6.84	0.133	32
19	Poly(n-butylmethacrylate)	0.55	0.99	4.945	0.129	28
20	Poly(ethylmethacrylate	0.56	1	3.112	0.127	24
21	Poly(3,3,5-trimethylcyclohexyl methacrylate)	0.56	0.98	3.789	-0.032	40
22	Poly (methylmethacrylate	0.56	1.01	2.121	0.181	21
23	Poly(4-methylcyclohexyl methacrylate	0.56	0.99	3.313	-0.354	20
24	Poly(ethyl a-chloroacrylate	0.61	1.03	3.368	-0.019	24
25	Poly (cyclohexyl a-bromoacrylate	0.61	1	3.423	-0.096	13
26	Poly(2-bromoethyl methacrylate	0.62	1.02	4.366	0.23	25
27	Poly(1-phenylethyl methacrylate	0.61	0.99	3.116	0.396	21
28	Poly(p-cyclohexylphenylmethacrylate	0.61	0.99	3.543	0.283	13

29	Poly(1-(o-chlorophenyl)ethyl methacrylate	0.64	1.01	3.186	0.321	30
30	Poly(benzylmethacrylate	0.62	1	3.22	0.354	14
31	Poly (phenylmethacrylate	0.64	1	2.635	0.34	13
32	Poly(vinyl benzoate)	0.65	1.01	2.024	0.346	8
33	Poly(o-methoxystyrene	0.63	0.99	1.464	0.484	9
34	Poly(N-benzylmethacrylamide	0.62	0.99	3.22	0.391	14
35	Poly(p-methoxystyrene)	0.63	0.99	1.699	0.494	9
36	Poly(pentachlorophenylmethacrylate	0.8	1.08	3.074	-0.282	68
37	Poly (N-vinylphthalimide)	0.69	1.02	0.945	0.423	11
38	Poly(2,6-dichlorostyrene	0.74	1.01	1.536	0.03	11
39	Polya-naphthylmethacrylate	0.66	1	2.204	0.897	13
40	Poly(styrenesulfide	0.68	0.98	1.276	0.305	0
41	Poly(a-naphthalene	0.69	0.97	0.918	0.604	0
42	**Poly(vinylidenefluoride**	0.57	1.13	0	0.005	10
43	Poly(vinylisobutylether	0.53	0.98	5.701	-0.012	17
44	Poly(propyleneoxide	0.53	1	0.961	0.07	2
45	Poly(vinylmethylether	0.53	1	4.127	0.027	8
46	Poly(ethyl acrylate	0.56	1.01	3.2	0.031	17
47	Polypropylene (atactic	0.53	0.96	2	0.027	5
48	Poly(vinyl sec-butyl ether)	0.53	0.98	3.534	0.001	17
59	Poly(isobutylmethacrylate)	0.55	0.99	4.945	0.132	34
50	Poly (methyl acrylate	0.57	1.03	2.377	0.108	14
51	Polyn-propylmethylacrylate	0.55	1	3.919	0.118	25
52	Poly(vinylbutyral	0.56	0.99	2.154	0.048	19
53	Poly (3-methylcyclohexyl methacrylate	0.56	0.99	3.313	-0.266	20
54	Poly(vinylchloroacetate)	0.63	1.05	3.494	0.089	17
55	Poly(methyl a-chloroacrylate	0.63	1.05	2.369	0.051	21
56	Poly(2-chloroethyl methacrylate	0.6	1.02	4.184	0.139	25
57	Poly(vinylchloride	0.65	1.02	2	-0.069	5

3. Équation et Analyse de régression

La régression linéaire multiple (RLM) a été utilisée pour calculer le modèle sur l'ensembles de calibration de 41 observations pour le choix Aléatoire, a pour équation **:**

IR = 1.58 + 0.951 **Mv** - 0.627 **Me** - 0.000536 **BAC** + 0.0678 **Mor15p**

Les paramètres statistiques dans le **tableau 4** du modèle élaboré montrent une corrélation substantielle entre les valeurs IR et les descripteurs choisis. La valeur du coefficient de régression de la validation croisée (Q^2_{LOO} = 0,961) suggère que la précision de prédiction de ce modèle QSPR est d'environ 96 %. En revanche, la valeur du coefficient de régression (R^2 = 0,972) indique une corrélation de 97 % entre les valeurs IR et les descripteurs sélectionnés dans l'ensemble d'entraînement. Le rapport de variance (F) est calculé pour évaluer la signification des coefficients de régression. La valeur de F doit être élevée pour que les coefficients de régression soient statistiquement significatifs. Le potentiel prédictif supérieur du modèle MLR est confirmé par les valeurs élevées de Q^2ext. Les ensembles d'entraînement et de validation présentent des valeurs de RMSE comparables. Le modèle MLR offre un bon ajustement, une prédictivité interne, une capacité prédictive externe, ainsi qu'un haut niveau de stabilité en validation interne.

Tableau 4 : paramètres statistiques du modèle

R^2	R^2_{adj}	$RMSE_{tr}$	MAE_{tr}	S	F	Q^2_{LOO}	$RMSE_{ext}$	Q^2_{ext}
0.972	0.969	0.011	0.008	0.011	321.173	0.961	0.010	0.973

Le tableau 4 montre que tous les paramètres statistiques évaluant le modèle sont très satisfaisants les degrés de signification inférieur à 5 % donc nous remarquons que tous les descripteurs sont significatifs. La valeur du VIF de chaque descripteur (***Mv, Me, Mor15p, BAC***) utilisé dans le modèle MLR est 1 < VIF < 5, ce qui indique une légère corrélation d'un descripteur donné avec les trois autres, alors le modèle développé est excellent.

Tableau 5 : paramètre statistique du modèle choisi.

Prédicteur	Coef	SE Coef	T	P	VIF
Constant	1.58441	0.07664	20.67	0	
Mv	0.95079	0.03575	26.59	0	1.534
Me	-0.62682	0.08513	-7.36	0	1.65
BAC	-0.0005357	0.0001775	-3.02	0.005	1.54
Mor15p	0.067778	0.008988	7.54	0	1.428

4. Analyse des résidus :

Les valeurs de hii, ainsi que les valeurs des résidus, ordinaires et standardisés sont présentées dans le **tableau 6**, dont les colonnes 3 et 4 reproduisent les valeurs expérimentales et prédites des indices de réfractivité des composés considérés. Tous les résidus standardisées (ei_std) (colonne 7), sont inférieurs, en valeur absolue, à **3** fois l'erreur standard sauf pour l'observation N°23.

Tableau 6 : Valeurs de IR (expérimentales et prédites), résidus de prédiction (standardisés et ordinaires), valeurs des leviers (hii).

ID	Name	IR. Exp.	IR. Pred.	Pred.Mod. Eq.Res.	H i/i	Ei.Std. Pred.
1	Poly(trifluoroethyl methacrylate	1.437	1.4312	-0.0058	0.4234	-0.6446
2	Poly(vinylethylether	1.454	1.4614	0.0074	0.0799	0.6548
3	Poly(vinyl n-butyl ether	1.4563	1.4643	0.008	0.0619	0.7015
4	Poly(3-butoxypropylene oxide	1.458	1.4728	0.0148	0.0782	1.3117
5	Poly(vinyl n-pentyl ether	1.459	1.4628	0.0038	0.0562	0.3352
6	Poly(3-hexoxypropylene oxide	1.459	1.4678	0.0088	0.0691	0.7694
7	Poly(vinyl n-hexyl ether	1.4591	1.4628	0.0037	0.0545	0.3267
8	Poly(vinyl n-octyl ether	1.4613	1.4679	0.0066	0.0701	0.5823
9	Poly(vinyl 2-ethylhexyl ether	1.4626	1.4526	-0.01	0.1122	-0.8965

10	Poly (vinyl n-decyl ether	1.4628	1.4652	0.0024	0.0833	0.2097
11	Poly(t-butylmethacrylate	1.4638	1.4737	0.0099	0.2015	0.9433
12	Poly(4-methyl-1-pentene	1.465	1.4724	0.0074	0.0964	0.6646
13	Poly(n-butyl acrylate	1.466	1.4735	0.0075	0.0386	0.6476
14	Poly(vinylacetate	1.467	1.4772	0.0102	0.1195	0.926
15	Poly(vinyl formate	1.4757	1.4707	-0.005	0.278	-0.4966
16	Polyethylene	1.476	1.4828	0.0068	0.1056	0.6056
17	Poly(2-fluoroethyl methacrylate)	1.4768	1.4681	-0.0087	0.0941	-0.7773
18	Polyn-hexylmethacrylate)	1.4813	1.4787	-0.0026	0.0841	-0.2328
19	Poly(n-butylmethacrylate)	1.483	1.4805	-0.0025	0.061	-0.2149
20	Poly(ethylmethacrylate	1.485	1.4858	0.0008	0.038	0.0685
21	Poly(3,3,5-trimethylcyclohexyl methacrylate)	1.485	1.479	-0.006	0.1558	-0.5557
22	Poly (methylmethacrylate	1.4893	1.4848	-0.0045	0.0502	-0.3925
23	Poly(4-methylcyclohexyl methacrylate	1.4975	1.4616	-0.0359	0.1436	-3.2905
24	Poly(ethyl a-chloroacrylate	1.502	1.5046	0.0026	0.0713	0.2316
25	Poly (cyclohexyl a-bromoacrylate	1.542	1.5241	-0.0179	0.089	-1.59
26	Poly(2-bromoethyl methacrylate	1.5426	1.5367	-0.0059	0.0436	-0.5075
27	Poly(1-phenylethyl methacrylate	1.5487	1.5594	0.0107	0.0686	0.9439
28	Poly(p-cyclohexylphenylmethacrylate	1.5575	1.5561	-0.0014	0.0366	-0.124
29	Poly(1-(o-chlorophenyl)ethyl methacrylate	1.5624	1.5655	0.0031	0.0749	0.2753
30	Poly(benzylmethacrylate	1.5679	1.5636	-0.0043	0.0434	-0.3745
31	Poly (phenylmethacrylate	1.5706	1.5822	0.0116	0.048	1.0071
32	Poly(vinyl benzoate)	1.5775	1.5885	0.011	0.0733	0.9701
33	Poly(o-methoxystyrene	1.5932	1.5908	-0.0024	0.076	-0.2076
34	Poly(N-benzylmethacrylamide	1.5965	1.5724	-0.0241	0.0531	-2.1045
35	Poly(p-methoxystyrene)	1.5967	1.5915	-0.0052	0.0785	-0.4573

36	Poly(pentachlorophenylmethacrylate	1.608	1.6125	0.0045	0.6998	0.7039
37	Poly (N-vinylphthalimide)	1.62	1.6239	0.0039	0.1044	0.3482
38	Poly(2,6-dichlorostyrene	1.6248	1.6511	0.0263	0.2687	2.6044
39	Polya-naphthylmethacrylate	1.641	1.639	-0.002	0.2698	-0.2033
40	Poly(styrenesulfide	1.6568	1.6373	-0.0195	0.1504	-1.7905
41	Poly(a-naphthalene	1.6818	1.6734	-0.0084	0.1954	-0.7957
42	**Poly(vinylidenefluoride**	1.42*	1.413	-0.007	1.137	1.1708
43	**Poly(vinylisobutylether**	1.450*	1.4641	0.0134	0.0527	0.748
44	**Poly(propyleneoxide**	1.457*	1.4652	0.0082	0.1375	-0.7088
45	**Poly(vinylmethylether**	1.467*	1.4591	-0.0079	0.0974	0.7208
46	**Poly(ethyl acrylate**	1.468*	1.4768	0.0083	0.0534	1.0901
47	**Polypropylene (atactic**	1.473*	1.4857	0.0122	0.0928	-0.7828
48	**Poly(vinyl sec-butyl ether)**	1.474*	1.465	-0.009	0.0518	0.0477
59	**Poly(isobutylmethacrylate)**	1.477*	1.4775	0.0005	0.0979	0.1412
50	**Poly (methyl acrylate**	1.479*	1.4806	0.0016	0.1165	-0.7695
51	**Polyn-propylmethylacrylate**	1.484*	1.4751	-0.0089	0.0459	0.378
52	**Poly(vinylbutyral**	1.485*	1.4894	0.0044	0.0323	-2.4305
53	**Poly (3-methylcyclohexyl methacrylate**	1.494*	1.4676	-0.0271	0.1032	0.8481
54	**Poly(vinylchloroacetate)**	1.513*	1.5222	0.0092	0.1567	0.0423
55	**Poly(methyl a-chloroacrylate**	1.517*	1.5175	0.0005	0.139	-0.4706
56	**Poly(2-chloroethyl methacrylate**	1.517*	1.5116	-0.0054	0.0405	1.5889
57	**Poly(vinylchloride**	1.539*	1.5557	0.0167	0.2032	-0.6446

La figure 14 présente un diagramme de dispersion des résidus de prédiction par rapport aux valeurs expérimentales d'IR pour les ensembles de données de prédiction et d'entraînement.

D'après la figure 14, le modèle ne présente aucune erreur proportionnelle ou systématique dans le développement du modèle MLR. Cela est dû à la propagation aléatoire des résidus de part et d'autre du zéro.

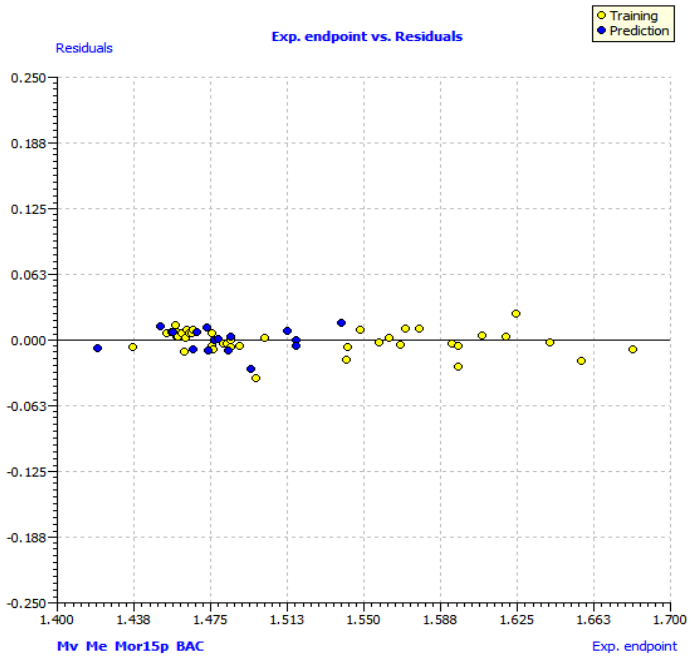


Figure 14. Résidus versus valeurs expérimentales de IR

5. Domaine d'application :

Selon les principes de l'OCDE (Organisation de coopération et de développement économiques) pour la régulation des modèles QSAR, le domaine d'application doit être défini et seules les prédictions pour les polymères qui se situent dans ce domaine peuvent être considérées comme fiables. Le graphique des résidus standardisés par rapport aux valeurs de levier (le graphique de Williams), qui peut être utilisé pour une détection graphique immédiate et simple à la fois des valeurs aberrantes IR (Y outliers) et des composés structurellement influents dans un modèle (X outliers), est présenté dans la **figure 15**.

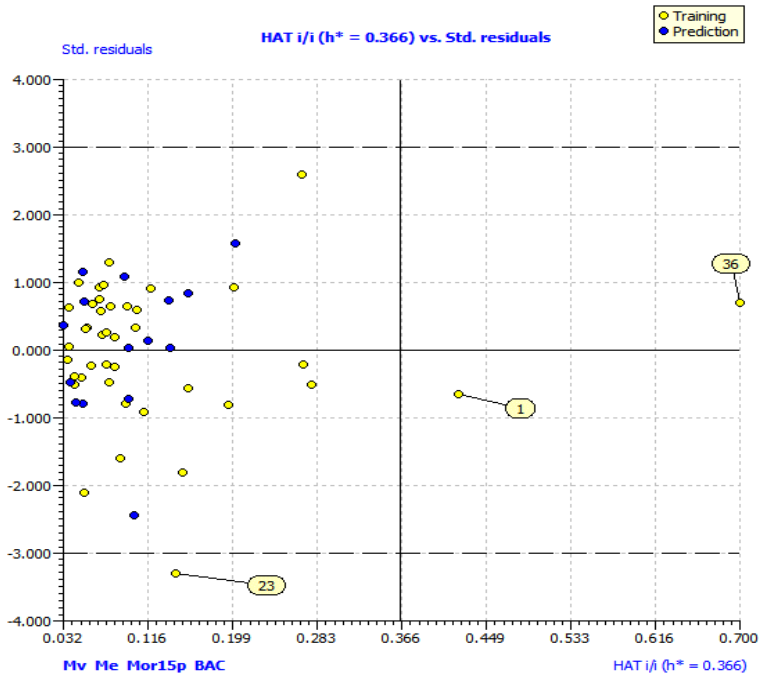


Figure 15. Diagramme de Williams

Les graphiques de Williams (Figure 3) révèlent qu'il y a seulement deux valeurs aberrantes X (la valeur de levier $h*=3\times (4+1)41=0,366$), à savoir le composé 1 et le composé 36 dans l'ensemble d'entraînement. De plus, le composé 23 dans l'ensemble d'entraînement est une valeur aberrante Y, car ce composé présente des résidus standardisés supérieurs à 3 unités d'écart type. L'erreur de prédiction de ce composé est significativement plus grande que celle des autres composés.

6. Qualité de l'ajustement

Une autre preuve de la qualité du modèle est la forte corrélation entre les valeurs observées et prédites des points d'ébullition pour les ensembles d'entraînement et de prédiction. La figure 16 illustre les valeurs prédites l'indice de réfractivité par rapport aux valeurs expérimentales. Le coefficient de corrélation (R^2 = 0,972) de ce graphique indique une bonne concordance entre ces valeurs, ce qui prouve la fiabilité du modèle.

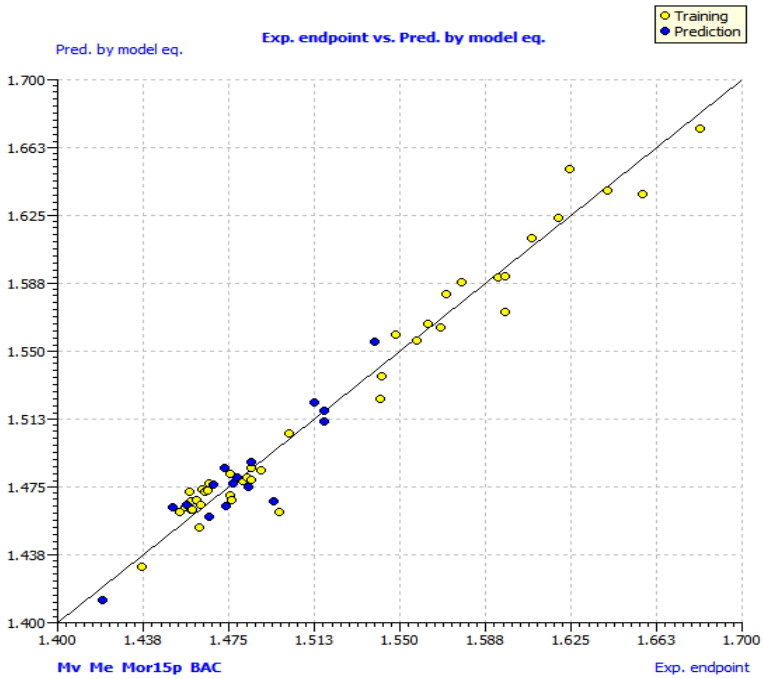


Figure 16. Les valeurs Prédites versus les valeurs expérimentales de IR

7. Test de Randomisation :

Le test derandomisation est appliqué pour prouver que le modèle n'est pas simplement dû au hasard, ce qui placerait aléatoirement les valeurs IR. Cela assure qu'il n'y a pas de corrélation avec les descripteurs. En conséquence, la performance du modèle diminue significativement. Comparé aux valeurs du modèle ($R^2_{Y\text{-}SCR}$ et $Q^2_{Y\text{-}SCR}$), les valeurs de R^2 et Q^2 de chaque itération et leurs moyennes ($R^2_{Y\text{-}SCR}$ et $Q^2_{Y\text{-}SCR}$) nous fournissent le critère que le modèle est bon. La figure 17 trace les valeurs de $R^2_{Y\text{-}SCR}$ et $Q^2_{Y\text{-}SCR}$ du modèle par rapport à leurs valeurs respectives de R^2 et Q^2. Il est important de noter que les valeurs de R^2 et Q^2 du modèle sont considérablement différentes des valeurs obtenues pour ces paramètres dans technique de y-scrambling, indiquant que le modèle n'a pas été obtenu en raison d'une corrélation fortuite.

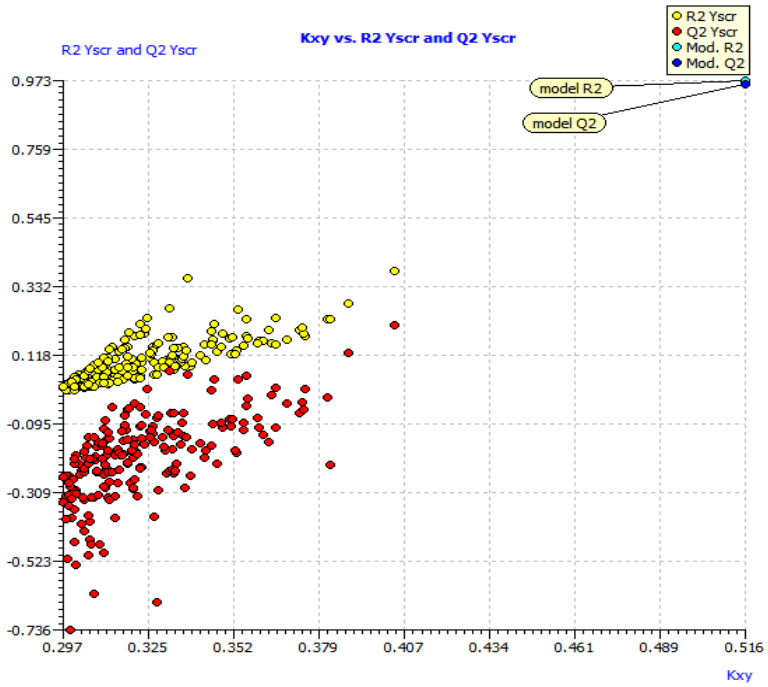


Figure 17. Test de randomisation

8. Mécanistique interprétation du modèle:

Le premier descripteur Mv (Volume moléculaire moyen des atomes de van der Waals). Ce descripteur mesure la taille moyenne de la molécule en prenant en compte les volumes des atomes individuels selon les interactions de van der Waals.

Les molécules plus grandes ont tendance à avoir une plus grande surface et à occuper plus d'espace, ce qui les rend plus susceptibles d'interagir avec la lumière. Cette interaction accrue conduit à une polarisabilité plus élevée, ce qui signifie que la molécule peut se déformer plus facilement en réponse à un champ électrique externe. Par conséquent, le descripteur Mv est une mesure de la capacité de la molécule à se polariser, ce qui influe sur son indice de réfraction.

Le deuxième descripteur Me (Électronégativité moyenne des atomes selon Sanderson).L'électronégativité est une mesure de la capacité d'un atome à attirer les électrons vers lui dans une liaison chimique. Ce descripteur représente la moyenne de l'électronégativité des atomes constitutifs de la molécule selon la méthode de Sanderson. Une électronégativité élevée indique une forte capacité de l'atome à attirer les électrons vers lui, ce qui crée des régions de charge partielle positive et négative dans la molécule. Ces régions de charge déséquilibrée

peuvent influencer la distribution de la densité électronique dans la molécule, affectant ainsi sa polarisabilité et son indice de réfraction.

Le troisième descripteur Mor15p (3D-MoRSE - signal 15 / pondéré par les polarisabilités atomiques). Ce descripteur capture la distribution tridimensionnelle de la densité électronique dans la molécule, en tenant compte des polarisabilités atomiques. Le MoRSE (MolecularRepresentation of Structure based on Electron diffraction) est une méthode de représentation de la structure moléculaire basée sur la distribution de la densité électronique. Le signal 15 se réfère à un aspect spécifique de cette représentation. En pondérant ce signal par les polarisabilités atomiques, le descripteur Mor15p tient compte des variations locales de la densité électronique dues à la polarisabilité différentielle des atomes. Ces variations de densité électronique affectent la capacité de la molécule à se polariser et donc son indice de réfraction.

Le dernier descripteur BAC (Indice centré de Balaban). Cet indice quantifie la centralité des atomes dans la structure moléculaire. Il mesure dans quelle mesure les atomes d'une molécule sont centralisés ou dispersés. Les molécules avec une centralisation plus élevée ont tendance à avoir une structure plus compacte et symétrique, ce qui peut influencer la manière dont la densité électronique est répartie dans la molécule. Une distribution électronique plus uniforme peut conduire à une polarisabilité plus élevée et donc à un indice de réfraction plus élevé. Le BAC fournit donc des informations sur la manière dont la topologie moléculaire peut influencer les propriétés optiques du polymère.

Conclusion générale

Un modèle QSPR simple a été présenté pour la prédiction de l'indice de réfraction, l'une des caractéristiques importantes des polymères. Ce modèle est un modèle linéaire multivarié simple et interprétable, qui ne comporte que quatre variables ou descripteurs moléculaires. Ces quatre descripteurs moléculaires ont été sélectionnés à l'aide de la technique GA-MLR. Le calcul des variables est principalement basé sur la structure chimique des molécules concernées. Les procédures de validation, grâce à la validation croisée, la séparation des données en ensembles d'entraînement et de validation indépendants ainsi que la Y-randomisation, ont démontré l'exactitude et la robustesse du modèle QSPR produit, non seulement en calculant son aptitude sur des ensembles de données d'entraînement, mais aussi en testant la capacité prédictive du modèle. Le modèle QSPR avec des descripteurs moléculaires simplement calculés pourrait être utilisé pour estimer les indices de réfraction (IR) de polymères similaires ou nouveaux.