République Algérienne Démocratique et Populaire.

Ministère de l'Enseignement Supérieur et la Recherche Scientifique.

Université de Médéa



Faculté de technologie

Département de: Génie des Procédés.

Option: Génie des procédés Pharmaceutiques.

En vu d'obtention du diplôme de Master en génie des procédés.

Thème:

MODELE QSPR POUR LA PREDICTION DE LA PERMEABILITE DES POLYMERES

Réalisé par :

- Melle : YAHI Karima

- Melle : KHALFI Meriem

Proposé et dirigés par :

-Pr.KREA Mohamed.

-Mme : HASNAOUI Hanaa

ANNEE UNIVERSITAIRE

2016/2017

REMERCIEMENT

On remercie Dieu le tout puissant de nous avoir donné la santé et la volonté d'entamer et de terminer ce mémoire, pour être notre meilleur confident et pour nous permettre de réaliser nos rêves. Merci pour nous guider et être toujours avec nous.

La réalisation de ce mémoire a été possible grâce au concours de plusieurs personnes à qui on voudrait témoigner toute notre reconnaissance.

Tout d'abord, ce travail ne serait pas aussi riche et n'aurait pas pu avoir le jour sans l'aide et l'encadreur : Monsieur ***KREA Mohamed.****, on la remercie pour la qualité de son encadrement exceptionnel, pour sa patience, sa rigueur et surtout pour sa confiance, ses remarques et ses conseils, sa disponibilité et sa bienveillance durant notre préparation de ce mémoire.*

Qu'il trouve ici le témoignage de notre profonde gratitude. Notre remerciement s'adresse à madame ***HASNAOUI Hanaa*** *pour son aide pratique et son soutien moral et ses encouragements, et pour sa patience, sa rigueur et surtout pour sa confiance, ses remarques et ses conseils, sa disponibilité et sa bienveillance durant notre préparation de ce mémoire.*

À tout les membres du jury pour avoir accepté d'évaluer ce travail et pour toutes leurs remarques et critiques, ainsi que nos professeurs pour leurs générosités et la grande patience dont ils ont su faire preuve malgré leurs charges académiques et professionnelles

A tous nos enseignants qui nous ont initié aux valeurs authentiques, en signe d'un profond respect et d'un profond amour !!!

Nos profonds remerciements vont également à toutes les personnes qui nous ont aidés et soutenue de prés ou de loi.

DEDICACE

Du profond de mon cœur, je dédie ce modeste travail à tous ceux qui me sont chères :

A ma raison d'être, ma raison de vivre, la lanterne qui éclaire mon chemin et m'illumine de douceur et d'amour.

**** Ma chère mère ****

A mon signe d'amour, de reconnaissance et de gratitude pour tous les soutiens et les sacrifices dont il a fait preuve à mon égard.

**** Mon cher père ****

A la témoignage de la fraternité, de l'encouragement, de l'amour et du sacrifice avec mes souhaite de bonheur de santé et de sucées.

Ma chère sœur (Djamila Nassima) ,Ma chère frère (Hamza Ahmed Hakim)

A tous les membres de ma famille avec mes souhaite de bonheur et de réussite.

A mon signe de l'amitié, de la sincérité, de la fidélité et de

l'amour qui Nous a Liées et des bons moments

passées ensembles, je vous Souhaitant un avenir

Radieux et plein de bonnes promesses.

Mes chers amis

A tous toute personne Ayant Consacré un jour, un moment de sa vie pour m'aider, ou simplement me

faire Un sourire…….

Karima

DEDICACE

Du profond de mon cœur, je dédie ce modeste travail à tous ceux qui me sont chères :

A ma raison d'être, ma raison de vivre, la lanterne qui éclaire mon chemin et m'illumine de douceur et d'amour.

**** Ma chère mère ****

A mon signe d'amour, de reconnaissance et de gratitude pour tous les soutiens et les sacrifices dont il a fait preuve à mon égard.

**** Mon cher père ****

A la témoignage de la fraternité, de l'encouragement, de l'amour et du sacrifice avec mes souhaite de bonheur de santé et de sucées.

Ma chère sœur (Massouda Asma Saida),Ma chère frère Abd raouf ,Salah adin et Bellal)

A tous les membres de ma famille avec mes souhaite de bonheur et de réussite.

A mon signe de l'amitié, de la sincérité, de la fidélité et de

l'amour qui Nous a Liées et des bons moments

passées ensembles, je vous Souhaitant un avenir

Radieux et plein de bonnes promesses.

Mes chers amis

A tous toute personne Ayant Consacré un jour, un moment de sa vie pour m'aider, ou simplement me

faire Un sourire…….

Meriem

ملخص:

في هذه الدراسة تم تطوير نموذج العلاقة الكمية البنية –النشاط للتنبؤ بنفاذية 73 متماثر لغازالاكسجين و ثاني أكسيد الكربون في درجات حرارة مختلفة.تم اختيار أهم الواصفات التي تتعلق ببنية الوحدة في سلسلة من ثلاث وحدات متكررة من المتماثرات عن طريق أسلوب الخوارزمية الجينية من مجموعة تتكون أكثر من 1586 واصفة .استخدمت الواصفات المختارة بالإضافة إلى درجات الحرارة كمداخل لتطوير نماذج من الشبكة العصبية الاصطناعية. النماذج غير خطية و تم التأكد منها ببعض أساليب التحقق الداخلي و الخارجي.النتائج المتحصل عليها أظهرت قدرة ممتازة في التنبؤ بنفاذية هذه المجموعة من المتماثرات للغازات مع المعلمات الإحصائيةبالنسبة ل O_2 ,$R=0.988$, $R=0.83$, $REQM=0.459$, $Q^2=0.82$ CO_2بالنسبة لERQM=0.459,$Q^2=0.83$

الكلمات المفتاحية

المتماثرات،الشبكة النور ونية،الغاز، النفاذية،المواصفات،التنبؤ.

Résumé:

Dans ce travail, un modèle quantitatif de relation structure-propriété (QSPR) est établi pour la prédiction de la perméabilité de 73 polymères au dioxyde de carbone et l'oxygène à différentes températures collectionné dans une base de données. 1586 descripteurs pour chaque polymère on tétés déterminés à partir de trois unités répétitives (trimères) et les plus importants ont été sélectionnés au moyen de la méthode d'algorithme génétique (AG).

Les descripteurs sélectionnés en plus de la température ont été utilisés comme des entrées pour développer des modèles de réseaux neuronal artificiel (RNA) . Les modèles sont non linéaires et ont été évalués par des méthodes de validation internes et externes. Les résultats obtenus indiquent l'excellente aptitude de RNA obtenu à prédire la perméabilité des gaz dans cet ensemble de polymères avec des paramètres statistiques satisfaisants R = 0.988, REQM = 0.126, Q^2 =0.82 pour CO_2 et R = 0.83, REQM = 0.459, Q^2 =0.8 pour O_2.

Les mots clés : Polymère, réseaux de neurone, gaz, perméabilité, descripteur, prédiction

Abstract:

In this work, a quantitative structure–property relationship (QSPR) model was established to predict permeability of73 polymers at different temperatures for carbon dioxide and oxygen gases collected in a database. For each polymer1586 descriptors was determined from three repeating units (trimer) and the most important were selected by means of a genetic algorithm function (GAF).The selected descriptors in addition to the temperature were used as the inputs to develop an artificial neural network (ANN) model. These nonlinear models were assessed by some internal and external validation methods. The obtained results indicated the excellent ability of ANN to predicti the permeability of this set polymers for O_2

and CO_2gases with satisfying statistical arameters,R=0.988, REQM=0.126,Q^2=0.82 for CO_2 and R=0.83, REQM=0.459,Q^2=0.8 for O_2.

.Key words: Polymer, artificial neural network, gas, permeability, descriptors, prediction

TABLE DES MATIERE

REMERCIMENT
DEDICACE
RESUMUE
TABLE DES MATIERES
LISTE DES FIGURES
LISTE DES TABLEAUX
ABREVIATION
NOMENCLATURE
INTRODUCTION GENERALE…… 1

CHPITRE I. PERMEABILITE DES POLYMERES

I.1.Généralités sur la structure et les propriétés des polymères…… 3
I.1.1 Définition…… 3
I.1.2. Classification des polymères …… 4
I.1.2.1. Selon leur structure…… 4
I.1.2.2. Selon leur origine…… 5
I.1.2.3. Selon leurs unités…… 5
I.1.2.4. Selon leur domaine d'application…… 6
I.1.2.5.Selon le comportement thermique…… 7
I.1.2.6. En fonction du cycle de vie …… 7
I.2.Méthodes de synthèse des polymères…… 8
I.2.1. Réaction de polycondensation:(StepGrowthPolymerization) …… 8
I.2.2.La polymérisation en chaîne (Chain polymérisation)…… 8
I.2.2.1. Définition …… 8
I.2.2.2.les caractéristiques des polymérisations en chaine…… 8
I.3.Généralité sur les polymères étudiés …… 9
I.3.1.Les polyacétylènes …… 10
I.3.2.Les fluoropolymères…… 10
I.3.3.les Norbornènes…… 11
I.3.4.Les silicones …… 11
I.4.Généralité sur les membranes …… 12
I.4.1. La Classification des membranes…… 12
I.4.1.1. Selon la structure des matériaux et la morphologie …… 12
I.4.1.2.La nature des matériaux constitutifs…… 13
I.4.1.3.Le mécanisme de séparation et la taille des pores…… 13
I.5.Notion sur le transport de masse dans les polymères …… 14
I.5.1.La diffusion…… 15
I.5.1.1.Les lois de diffusion de Fick …… 15
I.5.1.2.Type de diffusion …… 17
I.5.2.La sorption…… 17
I.5.2.1.La sorption d'après la loi de Henry …… 19
I.5.2.2.Mécanisme de sorption selon Langmuir …… 19
I.5.2.3.Mécanisme Brunauer-Emmet-Teller (BET)…… 20
I.5.2.4.Le mécanisme de Flory-Huggins…… 20
I.5.2.5.Le mécanisme de sorption double « dual » …… 20
I.5.3.Perméabilité…… 20

CHAPITRE II. THEORIE DE LA PREDICTION

II.1.Introduction…… 23
II.2. Historique de la prédiction de la perméabilité des polymères …… 23

II.3.Les principales méthodes de prédiction …… 26
II.3.1.Méthode de contribution d groupes…… 26
a. Méthode du Permachor…… 26
b. Méthode du Yampolskii et al…… 27
c. Méthode proposée par Paul et Park…… 28
II.3.2.Théorie des graphes…… 30
II.3.3.Réseaux de neurones artificiels …… 31
II.3.4.Conclusion…… 33

CHAPITRE III :
RELATION QUANTITATIVES STRUCTURE-PROPRIETES (QSPR)

III.1.Introduction…… 34
III.2.Les relations structure-propriétés quantitatifs "QSPR"…… 34
III.2.1.Définition d QSPR …… 34
III.2.2.Principe de QSPR…… 35
III.2.3.Les méthodes mathématiques utilises par le model QSPR…… 36
III.2.4.Les types d'apprentissages d'un modèle QSPR…… 36
III.3.Les descripteurs moléculaires…… 37
III.3.1.Définition d'un descripteur…… 37
III.3.2.L'importance des descripteurs…… 37
III.3.3.Le choix des descripteurs…… 37
III.3.4.Choix du nombre de descripteurs…… 38
III.3.5.Réduction le nombre de descripteurs…… 38
III.3.6. Les méthodes utilisées…… 39
A. Analyse en Composantes Principales (ACP)…… 39
B.Algorithmes génétiques…… 39
C.Régression linéaire multiple (MLR)…… 39
D.La Régression Linéaire Simple(RLS) …… 39
III.3.7.Le classement des descripteurs…… 40
III.3.8.Calcul des descripteurs moléculaires…… 40
III.3.9.Les types des descripteurs moléculaires…… 41
III.3.10.Bloc des descripteurs…… 41
III.4. Les réseaux de neurones…… 42
III.4.1.Le neurone formel (artificiel)…… 42
III.4.2.Analogie avec le modèle biologique…… 44
III.4.3.Topologie de connexion des neurones…… 44
III.4.3.1.Réseaux statiques non bouclés …… 44
III.4.3.2.Réseaux à une couche cachée…… 45
III.4.4.Apprentissage des réseaux de neurones…… 46
III.4.4.1.Algorithme de rétropropagation du gradient…… 47
III.4.4.2.Algorithme de Levenberg – Marquardt…… 47
III.4.5.Conception d'un réseau de neurones…… 47
III.4.5.1. Choix des échantillons…… 48
III.4.5.2.Elaboration de la structure du réseau …… 48
III.4.5.3. Apprentissage…… 48
III.4.5.4.Validation…… 48
III.5.Conclusion…… 48

CHAPITRE IV :
MODÉLISATION NEURONALE POUR LA PRÉDICTION DE LA PERMÉABILITÉ

IV.1 Introduction…… 50

Partie 1 : Élaboration de la base de données

IV.1.1.Les logiciels et Matériels utilisés …… 51
IV.1.2.Collecte des données…… 51
IV.1.3 Polymères collectés utilisés dans ce travail…… 51
IV.1.4 Dessin des molécules par ChemDraw…… 56
IV.1.5Modélisation moléculaire des trimères par HyperChem…… 57
IV.1.6.Calcul des descripteurs par DRAGON…… 59
IV.1.7.Collection des donnés dans un fichier EXCEL…… 63
Partie 2 : Réduction du nombre des descripteurs
IV.2.1.Prétraitement de la base de données…… 64
IV.2.2.Application de la méthode des Algorithme génétiques…… 65
Partie3 : Modélisation neuronal pour la prédiction de la perméabilité
IV.3.1.Répartition des données…… 70
IV.3.2.Optimisation et conception du réseau de neurone…… 71
IV.3.3.Optimisation de l'architecture de réseau de prédiction d'O2…… 74
IV.3.3.1.Étape 1 …….. 74
A .Effet de l'algorithme d'apprentissage…… 74
IV.3.3.2.Etape2 ……. 75
A. Effede normalisation…… 75
IV.3.3.3.Étape 3 …… 76
A. Effet de fonction d'activation…… 76
IV.3.3.4.Etape4 …… 77
A. Effet de nombre de neurone…… 77
B. Effet de nombre des couches cachées…… 78
IV.3.4.Optimisation de l'architecture de réseau de prédiction de CO_2 :…… 78
IV.3.4.1.Étape 1 :…… 78
A. Effet de l'algorithme d'apprentissage…… 78
IV.3.4.2.Étape 2……. 79
A. Effet de la normalisation…… 79
IV.3.4.3.Étape 3…….. 81
A.Effet de fonction d'activation…… 81
IV.3.4.4.Étape 4 …….. 81
A. Effet du nombre de neurones …… 82
B. Effet du nombre des couches cachées …… 82
IV.3.5.Résultat de l'optimisation…… 83
IV.3.6.Paramètres statistiques…… 87
IV.3.6.1.Coefficient de corrélation R…… 87
IV.3.6.2.Le coefficient de prédict*ion (*$\mathbf{Q^2_{train}}$*)*…… 88
IV.3.6.3.Le coefficient de prédiction *(*$\mathbf{Q^2_{test}}$*)*…… 88
IV.3.6.4.Facteur de performance (C)…….. 88
IV.3.6.5.Erreur relative absolue moyenne (ERAM)…… 88
IV.3.6.6.Test de Fisher : Significativité globale du modèle…… 88
IV.3.7.Analyse des résultats…… 90
CONCLUSION GENERALE…… ….. 91
REFERENCE BIBLIOGRAPHIQUE
ANNEXE

LISTE DES FIGURES

Figure I.1 : Représentation de la structure linéaire d'un polymère. 4
Figure I.2 : Représentation de structure ramifiée d'un polymère 4
Figure I.3 : Polymère réticulé avec ponts disulfure reliant deux chaînes. 5
Figure I.4: Segment de tans-polyacétylène. 10
Figure I.5 : Structure de norbornène. 11
Figure I.6 : Polydiméthylsiloxanes linéaires 12
Figure I.7 : Pénétration à travers la membrane. 12
Figure I.8 : Représentation schématique de la différente structure membranaire. 13
Figure I.9 : Schéma de fonctionnement d'une membrane permsélective 13
Figure I.10 : Schéma de principe du contrôle du transfert de matière en pervaporation (modèle de sorption-diffusion). 15
Figure I.11: Les familles de représentation de l'équilibre. 18
Figure I.12: Les cinq types d'isothermes de sorption. 18
Figure II.1: Représentation schématique de (a.) une topologie de réseau neural (b) un élé élément de traitement (neurones**)** 32
Figure III.1: Principe de la méthode QSPR. 36
Figure III.2: Représentation des blocs des descripteurs moléculaires. 42
Figure III.3 : Le neurone formel. 43
Figure III.4 : Interaction de signaux provenant de n neurones et analogie sommation de signaux dans un neurone artificiel. 44
Figure III.5 : Un perceptron multicouche. 45
FigureIII.6 : Réseau à une couche cachée. 46
Figure III.7: Réseau associatif avec apprentissage supervisé. 46
Figure IV.1 : Organigramme des déférentes étapes de réalisation de la base de données 56
Figure IV.2: Etape de représentation des structures par ChemDraw. 57
Figure IV .3 :HyperChem version8.0.7 pour la modélisation moléculaire des trimères 58
Figure IV.4 : Etape des pré-optimisations des molécules respectivement 58
Figure IV .5 :L'interface du logiciel DRAGON version 5.5 59
Figure IV.6: Première étapes pour le calcul des descripteurs par DRAGON 60
Figure IV.7: Deuxième étapes pour le calcul des descripteurs par DRAGON. 61
Figure IV.8 : Troisième étape pour le calcul des descripteurs par DRAGON. 62
Figure IV.9 : Quatrième étape pour le calcul des descripteurs par DRAGON. 61
Figure IV.10 : Cinquième étape pour le calcul des descripteurs par DRAGON. 62
Figure IV.11 : La base de données des descripteurs finale. 63

Figure IV12.: Architecture générale d'un algorithme génétique. 65
Figure IV.13 : Schématisation générale et répartition du contenu de la base de données 69
Figure IV.14 : Organigramme de la méthodologie de calcul de réseau de neurones statique. 72
Figure IV.15 : Effet de l'algorithme d'apprentissage sur la performance du réseau de neurones. 73
Figure IV.16 : Effet de la normalisation sur la performance d'apprentissage du réseau de neurones train SCG 75
Figure IV.17: Effet de la fonction d'activation sur les performances du réseau de neurones. 76
Figure IV.18 : Effet de nombre de neurones de la couche cachée sur la performance d'apprentissage du réseau de neurones 76
Figure IV.19 : Effet de nombre de couches couche cachées sur la performance d'apprentissage 77
Figure IV.20: Effet de l'algorithme d'apprentissage sur la performance du réseau de neurones 78
Figure IV.21: Effet de la normalisation sur la performance d'apprentissage du réseau de neurones train RP 79
Figure IV.22: Effet de la normalisation sur la performance d'apprentissage du réseau de neurones train BR 80
Figure IV.23: Effet de la fonction d'activation sur les performances du réseau de Neurone 80
Figure IV.24 : Effet de nombre de neurones de la couche cachée sur la performance d'apprentissage du réseau de neurones 81
Figure IV.25 : Effet du nombre des couches cachés sur la performance d'apprentissage 81
Figure IV.26 : Architecture de réseau de neurone optimisé d'O2. 84
Figure IV.27 : Architecture de réseau de neurone optimisé de CO_2 85
Figure IV.28 : Comparaison entre la perméabilité expérimentale et la perméabilité estimée pour l'O_2 et CO_2 86
Figure IV.29 : Résidus des valeurs de log(p résidu) obtenus en fonction de log(p expérimentale) pourO_2 86
Figure IV.30 : Résidus des valeurs de log(P résidu) obtenus en fonction de log(p expérimentale) pour CO_2 86

LISTE DES TABLEAUX

Tableau IV.1 : Les perméabilités expérimentales des Norbornenes 51

Tableau IV.2 : Perméabilité expérimentales des fluoropolymères 53

Tableau IV.3 : Perméabilités expérimentales des polyacétylènes 53

Tableau IV.4: Perméabilités expérimentales des silicones 55

Tableau IV.5 : Nombre de descripteurs calculés par DRAGON 62

Tableau IV.6 : Résultats de calcul des descripteurs moléculaires pour CO_2 66

Tableau IV.7 :Résultats de calcul des descripteurs moléculaires pour O_2 67

Tableau IV.8 : Signification des principaux descripteurs moléculaires utilisés. 68

Tableau VI.9 : Les significations de quelque instruction de Matlab. 71

Tableau VI.10 : Les codes des différentes fonctions d'apprentissage utilisés dans le test 71

Tableau IV.11 : les algorithmes de normalisation. 74

Tableau IV.12 : Notation des réseaux normalisés 74

Tableau IV.13 : Différentes fonctions d'activations utilisées dans la phase de conception du modèle neuronal 75

Tableau IV.14 : Structure des réseaux de neurones optimisé. 83

Tableau IV.15 : Les valeurs des paramètres statistique 88

Abréviations

ACP	Analyse Composantes Principales.
ANN	Artificiel neurone network
AFC	Analyse correspondance facile.
AG	Algorithmes Génétiques.
DP	degré de polymérisation.
FFV	fraction de volume libre.
IUPAC	International Union of Pure and AppliedChemistry.
MF	Microfiltration.
MLR	Régression linéaire multiple.
MORSE	molécule représentation of structures based on electron diffraction
NF	Nano filtration.
PCR	Régression en composantes principale.
PLS	Régression moindre carre.
QSAR	Relations Quantitatives Structure Activité.
QSPR	Relations Quantitatives Structure Propriété.
REACH	Enregistrement, Evaluation et Autorisation des produits Chimiques.
RN	Réseaux de Neurone.
RNA	Réseaux de Neurone Artificiel.
RNS	Réseaux de Neurone Supervisé
SFV	volume libre spécifique.
UF	Ultrafiltration.

Nomenclature

C : Concentration du gaz dans le polymère (mol/l)
A : La surface soumise à un fluide (m^2)
J : Le flux diffusif de la molécule pénétrant(Kg/ m^2.s)
Q : La quantité de pénétrant qui passe à travers la membrane
t : Temps de pénétration d'un fluide(s)
D : Le coefficient diffusion ($m^2.s^{-2}$)
$c(x,t)$: La concentration locale de pénétrant à une position x et à un temps t.
T_g : Température de transition vitreuse (°C)
K_D: La constante de Henry
m_∞:La masse qui serait adsorbée
S : coefficient de sorption
m : La masse réellement adsorbée.
P : La pression partielle de soluté (atm,cmHg).
d : l'épaisseur du film polymère (cm)
p_0: La tension de vapeur du soluté
P : Le coefficient de perméabilité(Barrer ou cm^3(vap).cm/cm^2.s.cmHg)
α : Coefficient de sélectivité.
T : La température absolue(°C ou°k)
T_c : La température critique (°C)
n : le facteur espace.
$\left(\frac{\varepsilon}{k}\right)$: Le paramètre d'énergie de Lenard Jones du gaz pénétrant (°k)
V : Volume de polymère (cm^3)
V_0 : Volume occupé par les chaines polymériques
V_w: Volumes de van der Waals (cm^3/mol)
Φ_i : Fraction volumique des composantes i
d_{eff} : Diamètre effectif du pénétrant (cm)
π: Valeur numérique segmentaire « Permachor »(Cal/cm^3)
k : Le coefficient de normalisation.
FFV : Le volume libre fractionnel.
SFV : Volume libre spécifique (cm^3)
Υ: Paramètre représente les interactions entre le gaz et le polymère
G : Paramètre dépend de la nature du gaz pénétrant
Y : Fonction d'activation d'un neurone
U_i: Le potentiel
W_{ij}: Poids de connexions



Introduction générale



INTRODUCTION GENERALE

Les polymères naturels ont été parmi les premiers matériaux utilisés par l'homme : bois et fibres végétales, cuir, tendons d'animaux, laine, etc. Les polymères sont devenus l'élément essentiel d'un nombre très important d'objets de la vie courante, dans lesquels ils ont souvent remplacé les substances naturelles. Ils sont présents dans de nombreux domaines industriels, les recherches sont entreprises pour collecté les matériaux adéquats pour la séparation des mélange gazeux, en utilisant des procédés de séparation membranaire **[1]**.

Les caractéristiques des transports en masse des réseaux macromoléculaires ont fait l'objet de recherches intensives depuis plusieurs années. Une meilleure compréhension des phénomènes de transport à travers les polymères est très importante pour obtenir une amélioration significative dans les domaines des revêtements protecteurs, des procédés de séparation des membranes et de l'industrie de l'emballage. Récemment, une combinaison d'économie améliorée et de meilleure technologie a abouti à des produits à membrane qui ont signé une nouvelle ère dans l'utilisation commerciale des membranes pour la séparation des liquides et des gaz. Les techniques de séparation des membranes ont été largement utilisées pour de nombreuses applications industrielles comme le dessalement de la saumure et de l'air enrichi en oxygène.

La clé de réussite de ces procédés réside dans le bon choix des matériaux filmogènes. Ainsi, au cours de la dernière décennie, un nombre considérable de travaux sur la perméation des gaz à travers des films polymériques a été réalisé. Souvent, ces travaux sont orientés en vue de trouver des relations étroites entre les paramètres de perméabilité et la microstructure du matériau filmogène candidat à une séparation/récupération des gaz ou de vapeurs condensables. Cela a engendré un cumule très important des données expérimentales des paramètres de perméation (coefficients de perméabilité, coefficient de diffusion et de solubilité, l'énergie d'activation, etc.) des gaz dans les polymères

Pas mal de théories élaborées pour prédire ces propriétés sont utilisées des méthodes alternatives plus fiables, moins coûteuses et aux résultats plus rapides ne cessent de se développer.

L'élaboration d'un modèle qui permettra de prédire la perméabilité des polymères aux gaz au moyen de la chémoinformatique.Cette dernière est une discipline scientifique qui

consiste à étudier et résoudre des problèmes relatifs à la chimie en appliquant des méthodes informatiques. Nous nous intéresserons aux modèles "in silico", appelés modèles Quantitative Structure-proprieteRelationships (*QSPRs*), mettant en relation la structure d'une molécule avec la perméabilité). Ces modèles quantifieront cette relation et mettront en exergue l'impact des propriétés de la molécule sur sa perméabilité. En se basant sur le modèle obtenu, on pourra faire la prédiction de la perméabilité de tous nouveaux polymères en un temps record et sans recourir à l'expérimentation.

Ce travail s'inscrit dans ce contexte et constitue une contribution dans l'exploitation des données expérimentales pour la collection et la validation des modèles publiés de prédiction de perméabilité des polymères.

Le présent travail s'organise en quatre chapitres selon le plan suivant :

Le premier chapitre de ce mémoire présente les polymères, leurs différentes classifications ainsi que le transfert des gaz dans les films polymériques et les mécanismes de transfert de masse.

Le chapitre suivant introduit les méthodes de prédiction de la perméabilité décrites dans la littérature et les travaux antérieurs.

Des connaissances concernant, les relations quantitatives structure-propriétés "QSPR"**,** les réseaux de neurones, les algorithmes d'apprentissage sont représentés dans un troisième chapitre.

Le dernier chapitre est divisé en trois parties : la première détaille la préparation de la base de données ainsi que la réduction du nombre des descripteurs et les résultats obtenus sont analysés et expliqués dans la seconde partie, la troisième partie détaille la modélisation neuronal pour la prédiction de la perméabilité

La conclusion tirée à partir de ce travail est présentés à la fin du mémoire.



Chapitre -I-

Perméabilité des polymères



CHAPITRE I

PERMEABILITE DES POLYMERES

Plusieurs types d'interactions existent entre un emballage (polymère) et le produit emballé. L'inertie d'un emballage est rarement totale ce qui peut engendrer des modifications indésirables au niveau de l'efficacité et des effets secondaires du principe actif ou même des excipients et par conséquent menace la stabilité du médicament/aliment.

Et pour améliorer les performances des emballages, il faut bien connaître la morphologie et bien comprendre le comportement mécanique des polymères: c'est-à-dire de bien comprendre les relations étroites structure du matériau polymère- ses propriétés de permettions vis-à-vis des gaz et de l'humidité.

I.1 Généralités sur la structure et les propriétés des polymères

I.1.1 Définition:

Monomères : Ce sont les composés de base des polymères. Relevant de la chimie organique, ils associent par des liaisons covalentes des atomes de carbone et des atomes d'hydrogène, d'oxygène, d'azote, plus éventuellement des atomes de chlore, de fluor, de soufre, etc. Chaque atome de carbone, tétravalent, est relié aux atomes voisins par quatre liaisons covalentes, orientés dans l'espace vers les quatre sommets d'un tétraèdre régulier **[1]**.

Polymères : Un polymère est une substance composée de longues chaînes moléculaires appelées macromolécules. Une macromolécule résulte de l'enchaînement par liaison covalente, d'unités constitutives (ou groupes d'atomes) appelées « mères » (provenant du grec «meros» qui signifie «partie») **[2]**.

– A – A – A – A – A – A – A –

Où -A- représente l'unité constitutive

Un polymère peut être caractérisé par son degré de polymérisation ou sa masse moléculaire. Le degré de polymérisation est le nombre total de monomères contenus dans une macromolécule. Lorsque ce degré de polymérisation (DP) est inférieur à 30, on parle d'oligomère et lorsqu'il est supérieur à 30, c'est un polymère **[2]**.

I.1.2 Classification des polymères :

I.1.2.1 Selon leur structure :

- polymères linéaire : les polymères linéaires sont constitués de grandes chaînes de monomères reliés entre eux par des liaisons covalentes. Ces macromolécules sont liées entre elles par des liaisons secondaires qui assurent la stabilité du polymère. Ces liaisons secondaires sont des liaisons sou ponts hydrogène ou des liaisons de Van der Waals.

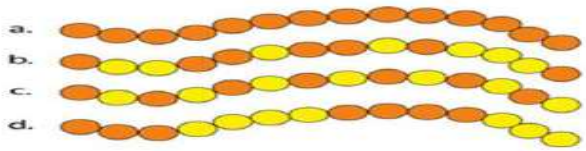
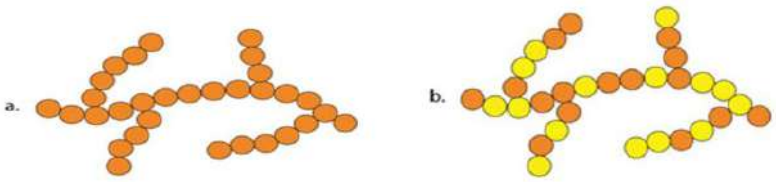


Figure I.1 représentation de la structure linéaire d'un polymère.

a: homopolymère. b : copolymère statistique.

c: copolymère alterné. d : copolymère séquencé.

- Polymères ramifiés : Des chaînes homopolymériques ou copolymériques peuvent se greffer sur d'autres chaînes au cours de la polymérisation. Au-dessus de la température de transition vitreuse, ces matériaux présenteront comportement visqueux plus marqué que les polymères linéaire.



a. Homopolymère ramifié b.copolymère ramifié

Figure I.2 : représentation de structure ramifiée d'un polymère.

- Polymères réticulés : La réticulation correspond à la formation de liaisons chimiques suivant les différentes directions de l'espace au cours d'une polymérisation, d'une polycondensation ou d'une polyaddition, et qui conduit à la formation d'un réseau **[3]**.

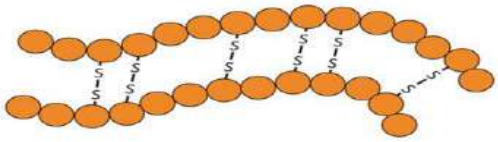


Figure I.3 : Polymère réticulé avec ponts disulfure reliant deux chaînes.

I.1.2.2 Selon leur origine: Les polymères classés selon l'origine peuvent être:

- des polymères naturels : ce sont des composés organiques formant la matière vivante, comme les protéines, les acides nucléiques, la cellulose, la chitine…
- les polymères artificiels : sont obtenus par modification chimique d'un polymère naturel, exemple: méthylcellulose.
- des polymères synthétiques: ce sont les matières plastiques, les élastomères, les fibres, les adhésifs **[4]**.

I.1.2.3 Selon leurs unités :

- Homopolymères **:** Le nom d'un polymère est généralement constitué du préfixe poly suivi du nom chimique :

- soit du monomère dans le cas où le polymère résulte d'une simple polymérisation

-soit du motif structural unitaire du polymère lorsque ce dernier résulte de la

polyaddition ou de la polycondensation de plusieurs monomères.

- Copolymères **:** les noms des copolymères sont constitués du préfixe poly, suivi, entre parenthèses, des noms chimiques des monomères donnés dans l'ordre décroissant des fractions massiques ou molaires de ces composants dans le copolymère, chaque monomère étant séparé des autres par une barre oblique **[5]**.

On distingue les familles suivantes **[2]:**

- Les copolymères statistiques **:** les unités constitutives A et B se répartissent le long de la chaîne en suivant une statistique. Les copolymères complètement désordonnés font partie de cette famille. Dans ce cas particulier, les unités constitutives sont reparties de façon aléatoire

—A—B—A—A—A—B—B—A—B—B—B—B—A—A—

- Les copolymères alternés **:** les unités constitutives A et B sont disposées en alternance le long de la chaîne.

─A─B─A─B─A─B─A─B─A─B─A─B─A─B─

- Les copolymères séquencés linéaires ou copolymères à bloc **:** ce sont des molécules où chaque unité constitutive est répétée plusieurs fois de suite (block) pour former de longues séquences. Par exemple :

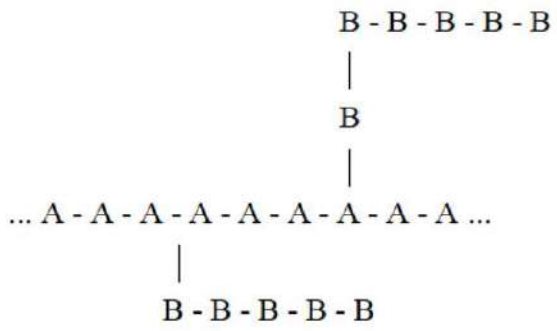
- Copolymère «di bloc» [AB] :

─A─A─A─A─A─A─A─B─B─B─B─B─B─B─

- Copolymère «tri bloc» [ABA] :

─A─A─A─A─A─B─B─B─B─B─A─A─A─A─

- Les copolymères séquencés et greffés **:** l'exemple le plus simple est celui d'une chaîne d'un homopolymère principal sur lequel des chaînes secondaires d'un autre polymère sont greffées. Par exemple :



I.1.2.4 Selon leur domaine d'application et économique :

- les polymères de grande diffusion :(encore appelés polymères de commodité), dont la production annuelle s'évalue en millions de tonnes, sont devenus d'un emploi quotidien. Le polyéthylène, le polystyrène, le poly (chlorure de vinyle) et quelques

autres sont à classer dans cette catégorie ; ils présentent une importance économique considérable.

- les polymères techniques : ont des caractéristiques mécaniques qui leur permettent de se substituer, de plus en plus, aux matériaux traditionnels (métaux, céramiques…) dans de nombreuses applications ; les polyamides, les polyacétals… font partie de cette famille.
- Les polymères spéciaux (ou polymères de fonction) : présentent généralement une propriété spécifique qui induit leur utilisation pour une application particulière. C'est dans cette catégorie que se trouvent les polymères conducteurs, photo-actifs, thermostables, adhésifs, etc. **[6]**.

I.1.2.5 Selon le comportement thermique:

Les polymères sont souvent classés d'après leurs propriétés thermodynamiques en trois types **:**

- Les thermoplastiques : ramollissent sous l'effet de la chaleur, ils deviennent souples, malléables et durcissent à nouveau quand on les refroidit, ces matériaux conservent leurs propriétés et ils sont facilement recyclables **[7]**.

- Les thermodurcissables : les matières thermodurcissables sont les produits dont la transformation conduit, par une réaction chimique, à des composés macromoléculaires tridimensionnels qui sont des matières thermo durcies ou thermo rigides. Dans ce cas les chaînes de départs sont beaucoup plus courte et plus réactives, ces chaînes vont se lier ensemble chimiquement, cette réaction conduit à des liens chimiques rigides et met en jeu toutes les molécules présentes pour former un réseau tridimensionnel **[8]**.

- Les élastomères: ces polymères présentent les même qualités élastiques que le caoutchouc, un élastomère au repos est constitué de longues chaînes moléculaire repliées sur elles même, sous l'action d'une contrainte, les molécules peuvent glisser les unes par rapport aux autres et se déformer **[9]**.

I.1.2.6 En fonction du cycle de vie [10] :

- Les polymères jetables (<1an).
- Les polymères durables (1- 20ans).

- Les polymères pour infrastructure (>20ans).

I.2 Méthodes de synthèse des polymères:

Les polymères sont synthétisés à partir des molécules de monomères par des réactions de polymérisation. Selon le comportement cinétique et le mécanisme de la réaction qui conduit à la formation des composés macromoléculaires, on distingue deux types principaux de réaction de polymérisation; la polycondensation et la polymérisation en chaîne.

La polymérisation est par définition la réaction chimique ou le procédé permettant la synthèse d'un polymère à partir de monomères, et le produit obtenu est un polymère synthétique **[11]**.Les réactions de polymérisation permettent notamment de transformer un mélange liquide, plus ou moins visqueux en un matériau solide dont le taux de réticulation dépend notamment de la fonctionnalité des monomères employés **[12-13]**.

I.2.1 Réaction de polycondensation:(StepGrowthPolymerization) :

- La polycondensation est une réaction chimique entre molécules de base possédant des groupements fonctionnels différents, cette réaction mène à l'élaboration des macromolécules nommés polycondensats **[14]**.
- Au cours de la polycondensation, les monomères s'associent avec élimination simultanée d'atomes ou de groupes d'atomes. Les polyamides, les polyesters et certains polyuréthannes sont des exemples typiques de polymères obtenus par polycondensation.

I.2.2 La polymérisation en chaîne (Chain polymérisation): [11]

I.2.2.1 Définition :

Dans la polymérisation en chaîne, les monomères s'associent sans réaction d'élimination simultanée. C'est le procédé le plus utilisé dans l'industrie: le polyéthylène, le polypropylène, le polystyrène, l'alcool polyvinylique et le polytétrafluoroéthylène (Téflon) sont des exemples de polymères obtenus par polymérisation en chaîne.

I.2.2.2 Les caractéristiques des polymérisations en chaine :

- La polymérisation en chaine se déroule en 3 phases :
    - La première phase est la phase d'Amorçage: elle correspond à l'activation d'une molécule de monomère qui se fait grâce à un amorceur A, le début de l'équation de la polymérisation est de la forme:

$$A+M \longrightarrow M^*$$

Où M*: est le monomère activé ou centre actif.

- La deuxième phase est la phase de propagation : elle correspond à la propagation du centre actif à d'autres monomères. L'activité de ce monomère et activé à d'autres monomères.

$$M+M \longrightarrow M\,M^*$$

Ainsi, lapropagation va être la répétition de cette forme d'équation:

$$M+M^* \longrightarrow MM^* \longrightarrow MM^* \longrightarrow MMM^* \longrightarrow MMMM^* \longrightarrow \ldots.$$

- La troisième phase est la phase de terminaison:Ilcorrespond à la rencontre d'un polymère ayant un monomère activé au bout de la chaîne et d'une espèce qui désactive le monomère.



La polymérisation peut être radicalaire, cationique ou anionique selon la nature de l'espèce active ou la polymérisation par transfert de groupe.

- Polymérisation radicalaire: la Polymérisation radicalaire est une réaction en chaîne et fait intervenir comme espèce active des radicaux ; elle se déroule en trois étapes: amorçage, propagation et terminaison.
- Polymérisation anionique**:** dans ce type de polymérisation, le centre actif est un carbocation, le groupement R a tendance à être donneur d'électron, exemple:

  CH2, -CH2, –O…

- Polymérisation cationique **:** le centre actif est un carbanion, le groupement R à tendance à être accepteur d'électron, exemple :

  CN, COOR,…

I.3 Généralité sur les polymères étudiés :

on utilisé quatre types des polymères sont :

I.3.1 Les polyacétylènes

Le polyacétylène est un polymère organique dans lequel l'unité de répétition est l'éthyne : $(C_2H_2)_n$a été synthétisé pour la première fois en 1958 par NATTA **[15]**. Le polymère est formé d'une longue chaîne d'atomes de carbone avec une alternance de liaisons simples et liaisons doubles, portant chacun un atome hydrogène. Cette alternance confère au polyacétylène deux structures planes cis et trans. La configuration trans est thermodynamiquement la plus stable **[16]**. La structure schématique du polyacétylène est montrée ci-dessous :

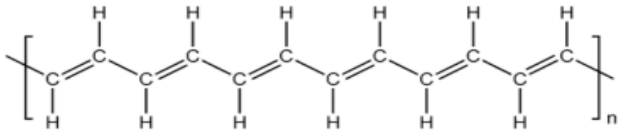


Figure I.4: Segment de tans-polyacétylène.

Plusieurs méthodes de polymérisation de l'acétylène conduisent à la formation du polyacétylène (PA), soit sous forme de film d'aspect métallique **[17]** sois se forme de poudre noire, ou encore de gel **[18,19].**

I.3.2 Les fluoropolymères :

Les fluoropolymères ont été découverts en 1938 par le Dr Roy J. Punkettes un polymère fluorocarboné à base de plusieurs liaisons carbone-fluor solides. Il se caractérise par une grande résistance aux solvants, aux acides et aux bases **[20]**. Ce sont des matériaux plastiques de haute performance utilisés dans des environnements chimiques et des températures élevées, principalement là où une spécification de performance critique doit être respectée. Ils sont utilisés par les industries liées à la défense et dans l'automobile, l'aérospatiale, l'électronique et les télécommunications. Ils sont également utilisés dans de nombreux produits de consommation **[21]**.

Les fluoropolymères partagent les propriétés des fluorocarbures en ce qu'ils ne sont pas aussi sensibles à la force de van der Waals que les hydrocarbures . Ceci contribue à leurs propriétés antiadhésives et de réduction du frottement. De plus, ils sont stables en raison de la stabilité de plusieurs liaisons carbone-fluor ajoutées à un composé. Les fluoropolymères sont des thermoplastiques ou thermodurcissables , peuvent être des homopolymères ou des copolymères.

I.3.3 Les Norbornènes :

Le norbornène, norbornylène ou norcamphène est un hydrocarbure ponté de formuleC_7H_{10}. C'est un solide blanc à l'odeur aigre piquante. La molécule consiste en un cycle de cyclohexène avec un pont méthylène entre les carbone C-3 et C-6. Du fait de sa liaison double, elle subit une certaine contrainte cyclique et possède une réactivité significative.

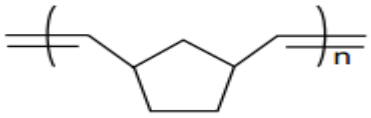


Figure I.5 : Structure de norbornène

Le norbornène à une structure semblable à d'autres composés bicycliques, le norbornadiène qui possède la même structure avec une liaison double supplémentaire, et le norbornane sans aucune liaison double, elle synthétisé par la réaction entre le cyclopentadiène et l'éthylène **[22,23]**. Le norbornène n'a pas beaucoup d'applications pratiques. Il est utilisé pour synthétiser des intermédiaires de composés pharmaceutiques, de pesticides, de certains parfums spéciaux et en synthèse organique en général. Les polynorbornènes sont des polymères avec de hots températures de transition vitreuse et une grande transparence optique.

I.3.4 Les silicones

Les silicones, ou polysiloxanes, sont des polymères macromoléculaires caractérisés par des liaisons silicium-oxygène (Si-O) et des liaisons silicium-carbone (Si-C) **(**Figure I.6). Les silicones se déclinent dans une gamme de matériaux présentant des propriétés mécaniques et physiques différentes: élastomères, pâtes, résines, gels ou huiles. Les silicones sont caractérisée par :

- Des forces intermoléculaires, énergie de liaison, énergie de surface faible.
- Une viscosité faible, peu dépendante de la température.
- Pas de variation importante des propriétés physique entre 40°C et 250°C.
- Ils sont neutres sur le plan physiologique **[24]**.

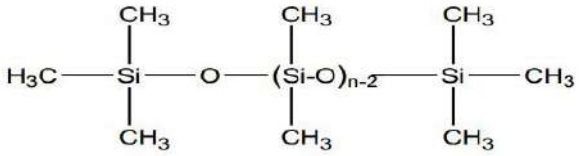


Figure I.6 : Polydiméthylsiloxanes linéaires

I.4 Généralité sur les membranes :

- Une membrane est une barrière de quelques centaines de nanomètres à quelques millimètres d'épaisseur, elle permet le transfert ou la rétention de certaines substances entre les deux milieux liquides ou gazeux qu'elle sépare. Ces substances pouvant être des ions, des molécules ou des particules. Mulder donne à la séparation membranaire la représentation schématique fournie par la figure I.7 **[25]**.

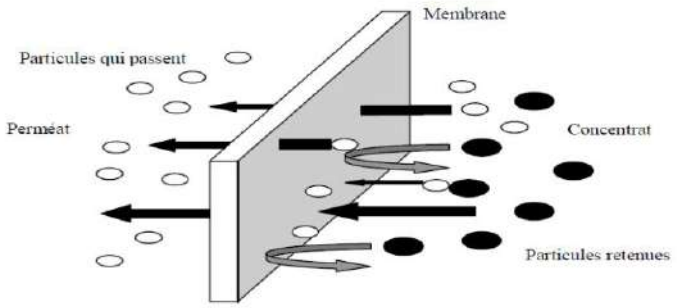


Figure I.7 : pénétration à travers la membrane.

- La séparation des espaces chimique est fondée sur leur affinité physico chimique à la membrane, sur la taille et/ou la charge électrique des molécules ou des espèces ionique **[26]**.

I.4.1 La Classification des membranes

la classification des membranes industrielles est basé sur :

I.4.1.1 La structure des matériaux et la morphologie : on distingue deux types

❖ Les membranes symétriques (ou isotropes):dont la structure est homogène sur toute l'épaisseur (figure I.8.a).

❖ Les membranes asymétriques (ou anisotropes): le concept de membraranes asymétriques a été développé dans les années soixante par Loeb et Sourirajan **[27]**. Elles présentent une structure hétérogène, dont le transport varie d'une couche à l'autre (figure I.8.b).

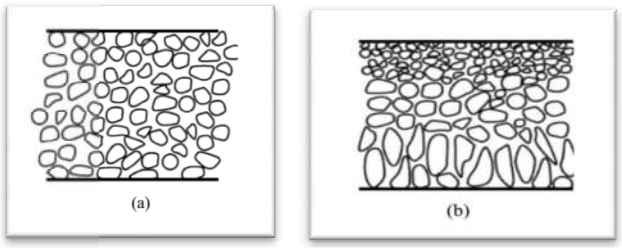


Figure I.8 : Représentation schématique de la différente structure membranaire **[28]**.

(a) Membrane symétrique, (b) Membrane asymétrique.

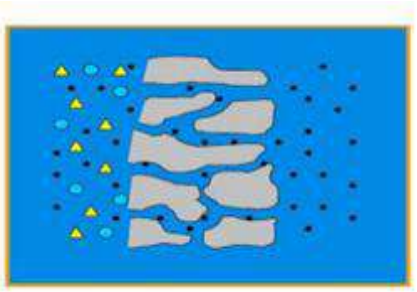
I.4.1.2 La nature des matériaux constitutifs:

Les membranes organiques : sont fabriquées à base de polymères tels que le polyamide, le polysulfone, le polyéthersulfone…etc. Sont les membranes utilisées dans les procédés de filtration

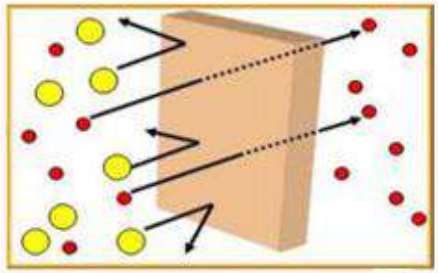
Les membranes inorganiques : sont constituées de matériaux céramiques tels que l'alumine, le dioxyde de titane ou la zircone.

I.4.1.3 Le mécanisme de séparation et la taille des pores :

- Membrane poreuses macropores : d>5μm (Microfiltration(MF), Ultrafiltration (UF))
- Membrane poreuses mésopores : 2à50 nm (Ultrafiltration (UF))
- Membrane poreuses micropores : d<2nm (Nanofiltration(NF))
- membranes denses : **(**perméation gazeuse, pervaporation, osmose inverse) **[29]**. (Voire la figure I.9)



Membrane poreuses (MF, NF, UF)



membrane non poreuses

Figure I.9 : Schéma de fonctionnement d'une membrane permsélective

I.5 Notion sur le transport de masse dans les polymères :

Il est généralement admis que les gaz traversent les films polymères denses par un processus de "dissolution-diffusion". L'application d'un gradient de pression gazeuse (force motrice) entre les faces amont (en surpression) et aval (en dépression) de la membrane provoque les trois phénomènes successifs suivants :

- Absorption du gaz à la surface amont.
- Dissolution et diffusion moléculaire du gaz dans le matériau.
- Désorption du gaz de la surface aval **[30]**.

La première étape et la dernière étape sont considérées comme instantané par rapport à la seconde, qui sort de la diffusion devient l'étape cinétiquement militante pour le transport. Dans les processus de diffusion pure, le gradient de concentration en molécules diffusantes et la seule force motrice **[31]**.

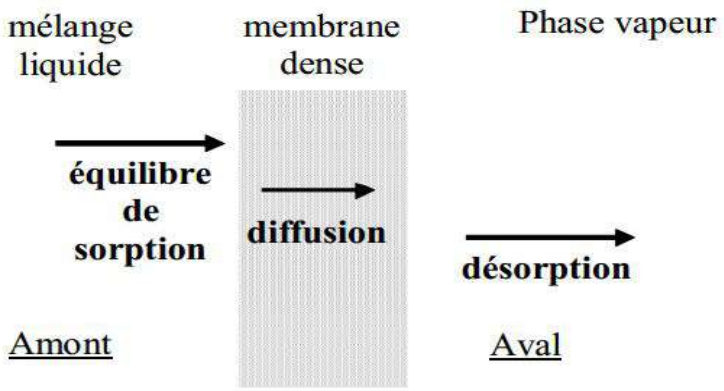


Figure I.10 : Schéma de principe du contrôle du transfert de matière en pervaporation (modèle de sorption-diffusion) **[32]**.

Deux revues récentes présentent les propriétés de transport de gaz **[33]** et d'arômes **[34]** des déférents films et montrent les nombreux facteurs qui influencent ces propriétés. Dans la partie qui suit, nous allons donc détailler les déférents phénomènes de transport et les facteurs qui les influencent,

I.5.1 La diffusion :

La diffusion est le processus responsable du déplacement de la matière d'un point d'un système vers un autre point. Pour les gaz, cette diffusion est la plus rapide (environ 10^{-3}m.s^{-1}), elle diminue en passant aux liquides (environ10^{-5}m.s^{-1}) et aux petits solutés dans les solides (environ10^{-9}m.s^{-1}) **[35]**. Dans l'étude de la diffusion de solvants dans les polymères, différents comportements peuvent être observés. La diffusion du solvant est liée aux propriétés physiques du polymère et aux interactions entre le polymère et le solvant.

I.5.1.1 Les lois de diffusion de Fick :

Considérons une membrane de polymère d'épaisseur *l*, de surface *A* en contact avec le fluide. Soit *Q* la quantité de pénétrant passant au travers de la membrane pendant un temps *t*. On peut définir le flux *J* de fluide passant au travers de la membrane par unité de temps et de surface par :

$$J = \frac{Q}{At} \qquad (\mathrm{I}, 1)$$

La première loi de Fick établit une relation proportionnelle entre le flux de substance diffusante à travers la membrane et le gradient de concentration entre les deux côtés de la membrane :

$$J = -D\nabla C \qquad \text{(I.2)}$$

Ou D est le coefficient de diffusion ($m^2.s^{-2}$). Cette première loi est applicable à l'état d'équilibre qui est atteint quand la concentration ne varie plus et que le flux est constant. Quand la diffusion se passe seulement dans une direction x (unidirectionnelle), la relation se réduit à

$$J = -D\frac{\partial \mathrm{c}}{\partial \mathrm{x}} \qquad \text{(I.3)}$$

La simplification est valide quand l'épaisseur de la membrane considérée est beaucoup plus petite que les autres dimensions (par exemple, au diamètre dans le cas d'une membrane circulaire). Si cette condition n'est pas vérifiée, les phénomènes de diffusion dans les autres directions ne peuvent plus être négligés.

En dynamique, le transfert de matière par diffusion est lié à la concentration du pénétrant dans la matrice polymérique et du temps. La seconde loi de diffusion de Fick décrit cet état de non-équilibre et donné (dans le cas unidirectionnel) par :

$$\frac{\partial \mathrm{c(x,t)}}{\partial \mathrm{t}} = \frac{\partial J_x}{\partial x} = \frac{\partial (D\nabla C)}{\partial x} = \frac{\partial D}{\partial x}\frac{\partial C}{\partial x} + D(C)\frac{\partial^2 C}{\partial x^2} \qquad \text{(I.4)}$$

C(x, t)Est la concentration locale de pénétrant à une position x et à un temps t.cette équation différentielle peut être intégrée en tenant compte de la condition initiale(t=0).

Et des conditions limites (x=0, l).Il est constitué la référence théorique pour de nombreux problèmes communs de diffusion.

La résolution de cette équation donne le profil de concentration dans la zone de diffusion à déférents intervalles de temps. Pour de nombreux systèmes polymères/pénétrants, et dans des conditions définies, D peut-être considérées comme constant dans toute l'épaisseur de la membrane. L'équation précédente se réduit alors à :

$$\frac{dC}{\partial t} = \frac{d^2 C}{dx^2} \qquad \text{(I.5)}$$

I.5.1.2 Type de diffusion :

- Diffusion Fickienne ou Cas I : est généralement observé quand on travaille au-dessus de la température de transition vitreuse (Tg). Quand le polymère est à l'état caoutchoutique, les chaînes macromoléculaires ont une mobilité élevée qui permet une diffusion du solvant facilitée. Donc la vitesse de relaxation du polymère est supérieure à la vitesse de diffusion du solvant.
- Diffusion non Fickienne ou cas II : ce cas est principalement observé pour les polymères vitreux, quand la température de l'étude est en dessous de Tg **[36]**.

En général, l'étape de diffusion est le facteur limitant dans le transfert de matière. Cette diffusion, pour une température constante, est un phénomène complexe qui dépend fortement de la concentration locale en solvant dans le film polymère et donc du degré de gonflement de ce dernier. Ce qui veut dire que le coefficient de diffusion varie à son tour entre l'amont et l'aval du film polymère **[37]**.

I.5.2 La sorption

L'adsorption peut être définie comme l'opération fondamentale de Génie Chimique qui exploite l'aptitude de certains solides à concentrer spécifiquement à leur surface les constituants d'une solution permettant ainsi leur séparation.

Il existe deux types d'adsorption qui se diffèrent complètement par les énergies mises en jeu et par leur nature :

- L'adsorption physique ou adsorption de Van der Waals.
- L'adsorption chimique ou chimie sorption ou encore adsorption activée.

La performance d'une adsorption dépend en grande partie de l'équilibre entre les deux phases. Cet équilibre fixe en effet la capacité maximale qui peut être atteinte dans les conditions opératoires. L'équilibre est généralement représenté sous forme graphique. Il rend compte de la relation entre la concentration en soluté adsorbé et la concentration en soluté dans la phase fluide. Il ya trois grandes familles de représentation de l'équilibre :

- les isothermes où l'on porte la masse de soluté (m) adsorbé par l'unité de masse de l'adsorbant en fonction de la pression partielle du gaz dans la phase vapeur à température constante.
- les isobares qui traduisent les variations de (m) en fonction de la température à pression partielle constante de l'adsorbat dans la phase gazeuse.

- les isostères qui donnent la pression partielle du soluté dans la phase gazeuse en fonction de la température à masse adsorbée constante.

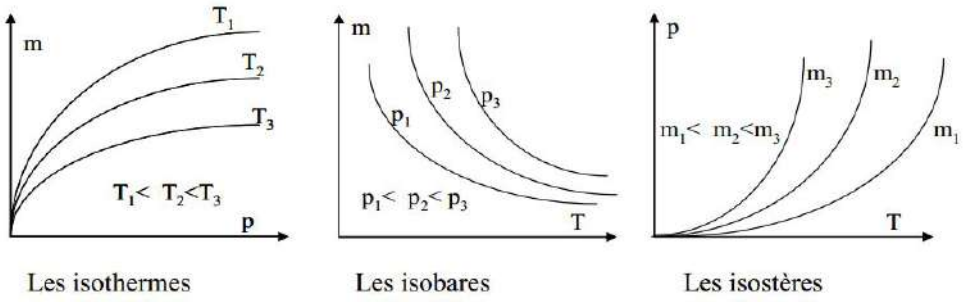


Figure I.11: les familles de représentation de l'équilibre **[37]**.

Le phénomène d'adsorption est un processus non-stationnaire qui s'arrête au moment ou s'établit un équilibre thermodynamique entre l'adsorbat et l'adsorbant. Cet équilibre est conditionné par la nature des deux phases, la pression partielle (ou concentration) de l'adsorbat dans les deux phases et la température.

En fonction du couple adsorbat - adsorbant, il possible de classer les isothermes en 5 catégories (figure I.12) **[38]**.

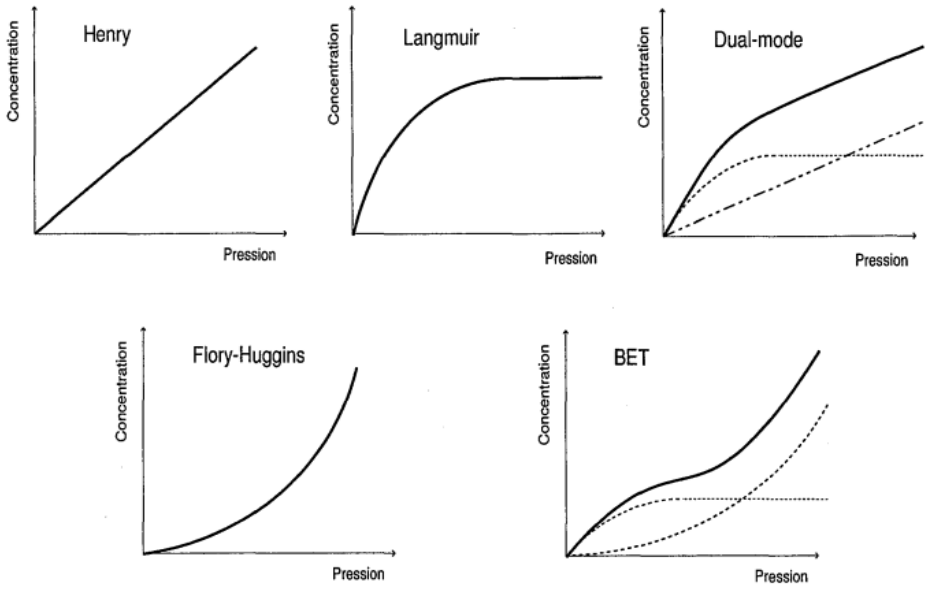


Figure I.12: Les cinq types d'isothermes de sorption **[38]**.

I.5.2.1 La sorption selon la loi de Henry :

Dans les polymères vitreux, ce modèle décrit correctement la dissolution de petites molécules, en particulier de certains gaz simples présentant peu d'interactions avec la matrice polymère. Cette dissolution est définie par l'équation de l'isotherme linéaire **[31]** :

$$C=K_D.p \qquad (I.6)$$

OùK_D, pente de l'isotherme, est la constante de Henry.

Le coefficient de solubilité S est alors défini comme une constante :

$$S=K_D \qquad (I.7)$$

I.5.2.2 Mécanisme de sorption selon Langmuir :

Moyennant les hypothèses suivantes :

- les molécules sont adsorbées sur des sites bien définis à la surface de l'adsorbant
- tous les sites sont identiques.
- chaque site ne peut fixer qu'une seule molécule, donc l'adsorption s'effectue suivant une couche mono moléculaire.
- l'énergie de chaque molécule adsorbée est indépendante de sa position sur la surface.

Ce qui explique l'allure de l'isotherme de sorption illustré par la figure I.11.La vitesse d'adsorption,r_{ads}, par unité de surface est décrite par l'équation

$$r_{ads}=k_{des}.(1-Q) \qquad (I.8)$$

Q c'est la fraction des sites non occupés

$$r_{des} = k_{des}.Q \qquad (I.9)$$

Par égalisation des deux vitesses à l'équilibre, on obtient l'équation de l'isotherme de Langmuir :

$$Q = k_{des}.\frac{p}{k_{des}.p+k_{des}} = \frac{k.p}{k.p+1} \qquad (I.10)$$

I.5.2.3 Mécanisme Brunauer-Emmet-Teller (BET) :

Brunauer-EmmetetTelleront proposé une généralisation de la théorie de Langmuir à l'adsorption en couche multimoléculaires à la surface du solide. Les auteurs ont adopté des hypothèses semblables à celles émises par Langmuir. La principale différence résulte du fait que les molécules de soluté peuvent s'adsorber sur les sites déjà occupés. La chaleur libérée au cours de l'adsorption sur de tels sites est alors égale à la chaleur normale de liquéfaction. L'équation d'équilibre obtenue à partir de cette théorie et pour un nombre infini découche est :

$$\frac{m}{m_\infty} = \frac{c\left(\frac{p}{p_0}\right)}{\left(1-\frac{p}{p_0}\right)\left[1+(c-1)\frac{p}{p_0}\right]} \qquad \text{(I.11)}$$

Où : m est la masse réellement adsorbée par unité de masse d'adsorbant. m_∞ Lamasse qui serait adsorbée si l'adsorbant était entièrement recouvert d'une monocouche la pression partielle de soluté.p_0La tension de vapeur du soluté.

Le paramètre C est une fonction de la température, de la chaleur d'adsorption des molécules sur la première couche et de la chaleur de liquéfaction **[39]**.

I.5.2.4 Le mécanisme de Flory-Huggins :

Ce type de sorption se produit lorsque les interactions entre les molécules du perméeant sont fortes relativement aux interactions perméeant/polymère. Le coefficient de solubilité augmente continuellement avec la pression.

I.5.2.5 Le mécanisme de sorption double « dual » :

Cette modalité était proposée pour décrire les courbes observées dans les cas de sorption des composés volatils de faible activité dans les polymères. Elle postule l'existence de deux populations de molécules diffusantes. Cette situation correspond à la combinaison des deux modes précédents. Néanmoins, ce modèle est valide pour des pressions modérées, en l'absence de fortes interactions. Il ne peut pas décrire le phénomène de sorption qui a lieu lors du gonflement ou de la plastification de la matrice par les molécules sorbées.

I.5.3 Perméabilité

La perméabilité est une caractéristique physique qui représente la facilité qu'a un matériau à permettre le transfert de fluide au travers d'un réseau connecté **[40]**.

La première loi de Fick est l'équation la plus couramment utilisée pour décrire le transport d'un gaz à travers un film polymère dans le cas idéal (gaz simples et non-plastifiants).Si l'on considère une relation linéaire entre la pression de gaz et sa concentration à l'interface (Loi de Henry), l'équation suivante s'applique :

$$C = S.P \qquad \text{(I.12)}$$

Avec :

C : concentration du gaz dans le film polymère.

S : solubilité du gaz dans le film polymère.

P : pression du gaz dans le film polymère.

En supposant que S et D sont indépendants de la concentration du gaz, l'intégration de la loi de Fick sur toute l'épaisseur d du film polymère donne l'expression du flux spécifique de gaz J à travers le film.

$$J = D.S\left(\frac{\Delta P}{d}\right) \qquad \text{(I.13)}$$

La capacité d'un film à transporter un gaz est décrit par le coefficient de perméabilité $P_e$défini par

$$P_e = D.S \qquad \text{(I.14)}$$

$$P_e = \frac{(\text{quntité du perméante})(\text{épaisseur du film})}{(\text{la sur face})(\text{temp})(\text{gradient de pression})}$$

Plusieurs autres grandeurs sont aussi utilisées pour caractériser le passage du gaz à travers un film :

- Taux de Perméation (Permeation rate) : C'est le coefficient de perméabilité sans tenir compte de la différence de pression.
- Perméance : C'est le coefficient de perméabilité sans tenir compte de l'épaisseur du film **[41].**
- Le flux de gaz varie, lui, avec l'épaisseur et la pression. Il est inversement proportionnel à l'épaisseur.

- Sélectivité : L'aptitude d'un film à séparer les constituants d'un mélange gazeux est décrite par la sélectivité α. Dans le cas d'un mélange binaire idéal (gaz A et B), la sélectivité $\alpha_{\frac{A}{B}}$est définie par le rapport des perméabilités$P_e(A)$et $P_e(B)$**[30]**.



Chapitre -II-

Théorie de la prédiction



CHAPITRE II

THEORIE DE LA PREDICTION

II.1 Introduction

Aujourd'hui, un grand nombre de paramètres de perméation des gaz dans les polymères (perméabilité des gaz, coefficients de diffusion, les facteurs de séparation, l'énergie d'activation de la diffusion et permeation) ont été accumulés. En particulier de nombreuses données ont été recueillies au cours de la dernière décennie. En vue d'apprendre les principes qui régissent la relation entre la perméabilité des gaz et l'unité structurale répétitive du polymère.

Un autre objectif, qui est souvent présent, mais pas toujours manifesté, c'est un développement de règles de prédiction des paramètres de perméation des polymères qui n'ont pas étés étudiés ou même en cours de préparation. Ces règles, si elles avaient été formulées, simplifier ou faciliter significativement la tâche de synthèse de nouveaux matériaux membranaires aux propriétés améliorées. Aujourd'hui,vu le vaste matériel expérimental déjà accumulé il est ambitieux et meme réaliste objectif ,pour les polymères ayant des structures chimiques et les propriétés l' opportunité pour l'interprétation et l'examen de diverses hypothèses.

II.2 Historique de la prédiction de la perméabilité des polymères :

La 1ère tentative de prédiction des perméabilités était la corrélation perméabilité / théorie du volume libre des polymères en 1980 .Lee a proposé d'utiliser la notion du volume libre spécifique définie comme su

$$SFV = (V - V0) \qquad \text{(II.1)}$$

Ou V est le volume spécifique du polymère calculé à partir de mesure de la densité et V_0 est le volume spécifique occupé par les molécules de polymère **[42]**.

Une autre corrélation proposée par Lee **[43]**, basée sur la notion du volume libre de Fujita **[44 ,45]** est définie comme suit :

$$P = A\,exp\,(-B\,/\,SFV) \qquad \text{(II.2)}$$

Où P est le coefficient de perméabilité pour un système gaz /polymère, A et B sont des constantes caractéristiques qui sont indépendantes de la concentration du gaz pénétrant et de la température, et SFV est le volume libre spécifique.

Une des corrélations les plus anciennes est celle de Stannett et Szwrarc **[46],**qui ont exprimé le coefficient de perméabilité d'un système gaz/polymère comme le produit de trois facteurs, l'un étant déterminé par la nature du film polymère, le second par la nature du gaz, tandis que le troisième, qui est de l'ordre de l'unité, tient compte de la spécificité interaction entre le gaz et le film. Ce dernier facteur représente, donc, seulement un petit terme de correction. Ainsi, la constante de perméabilité $P_{i,k}$, est donée par l'équation suivante:

$$P_{i,k} = F\,(polymer\ i).G\,(gaz\ k).\Upsilon(i,k) \qquad \text{(II.3)}$$

Ou i et k désignent respectivement le polymère et le gaz.

$$\frac{P_{ik}}{P_{il}} = \frac{G(gas\ k)\ \Upsilon(i,k)}{G(gas\ l)\ \Upsilon(i,l)} \qquad \text{(II.4)}$$

$$\frac{P_{ik}}{P_{jk}} = \frac{F(polymer\ i)\ \Upsilon(i,k)}{F(polymer\ j)\ \Upsilon(j,k)} \qquad \text{(II.5)}$$

Le premier ratio serait approximativement indépendant de la nature du polymère alors que le second serait approximativement indépendant de la nature du gaz.

La méthode de contribution des groupes existant dans la littérature est utilisée pour la 1ère fois par Salame**[47]**.Il a publié de nombreux ouvrages sur ce qu'il a appelé "l'approche Permachor " pour prédire la perméabilité des gaz tels que l'O_2dans les polymères de type barriére.Le paramètre pour un polymère donné est calculé à partir des facteurs empiriques de chaque groupement chimique du motif de répétition.

Alentiev et al **[48]**.ont proposé une autre corrélation pour la prédiction de la perméabilité des polymères complètement miscibles ou des copolymères aléatoires donnée par l'équation

$$Log\ P_m = a_1.logP_1 + a_2.logP_2 \qquad \text{(II.6)}$$

Ou : $a_1$et $a_2$sont les concentrations des comonomères

Teplyakov et Meares **[49]** ont proposé de corréler séparément le coefficient de diffusion et de sorption, au lieu de la perméabilité des différents polymères. Selon leur corrélation, les valeurs de Permachor ne sont pas les mêmes pour tous les gaz, ni pour tous les homopolymères, et ne sont pas un outil fiable de prédiction de nouveaux copolymères, ou un mélange de polymère. Dans ce modèle, le coefficient de la perméabilité calculé par les équations suivantes :

$$Ln\ p\ =\ \alpha\ +\ b\ (T_c\ /\ T)\ +\ c\ (d\ /\ T)^n \qquad \text{(II.7)}$$

Où: a , b et c sont des constantes en fonction du polymère, T la température absolue ,T_c la température critique du perméant; d le diamètre moléculaire du permanant qui est fonction de la température; et n est le facteur espace.

$$Log\ D\ =\ K_1 - K_2 d_{eff}^2 \qquad \text{(II.8)}$$

$$Log\ S\ =\ K_3\ -\ K_4\ (\frac{\varepsilon}{K}) \qquad \text{(II.9)}$$

Où K_i sont des constantes qui dépendent de la nature du polymère, d_{eff} est le diamètre effectif du pénétrant et $\left(\frac{\varepsilon}{k}\right)$est le paramètre d'énergie de Lenard Jones du gaz pénétrant.

Paul **[50]** a examiné la littérature sur le transport dans des mélanges et trouve que la règle de mélange simple, mais jusqu'à maintenant empirique, pour le coefficient de perméabilité, P, est souvent utilisée dans la discussion des mélanges et des copolymères:

$$Ln\ P\ =\ \Phi_1\ ln\ P_1\ +\ \Phi_2\ ln\ P_2 \qquad \text{(II.10)}$$

OùΦ_i fraction volumique des composantes i, P_1 etP_2 le coefficient de perméabilité. La base fondamentale des règles de mélange pour la perméabilité des petites molécules dans les systèmes polymères multicomposants sont considérées ensuite en utilisant deux différents points de vue.

Robeson et al **[51]** ont proposé une autre corrélation pour la prédiction de la perméabilité des polymères aromatiques donnée par l'équation :

$$Ln\ P\ =\ \sum_{i=1}^{n} \Phi_i\ ln\ P_i \qquad \text{(II.11)}$$

Ou Φ_i est la fraction volumique d'une unité structurale i,P_i la perméabilité de cette unité.

Une autre approche **[52]** pour l'estimation de la perméabilité des polymères a été note par Jia et Xu. Ils ont notés que la perméabilité de gaz pourrait être prévue à partir du rapport : volume libre molaire énergie molaire de cohésion (V_f/E_{coh}).Bien que cette approche peut donner une estimation de la perméabilité, mais sa précision est limitée et sans valeur pour la prédiction de la permselectivité.

Paul et Park **[53]** ont utilisé, dans une nouvelle publication, le concept de la fraction du volume libre FFV pour la prédiction de la perméabilité ainsi que la permselectivité.

II.3 Les principales méthodes de prédiction :

II.3.1 Méthode de contribution des groupes

a. Méthode du Permachor [54].

Salame étudié la relation entre la structure et de la morphologie des polymères avec la barrière des gaz et la prédiction essentiellement de sa perméabilité envers trois gaz proposés : O_2, N_2etCO_2. Le travail peut être étendu à d'autres gaz et vapeurs, pourvu qu'il n'y ait pas des interactions ou gonflement entre le perméant et le polymère. L'extension de ce travail aux polymères non couverts est évidente. Bien que seulement 27 polymères soient listés, l'approche a réussi avec plus de 60 polymères.

Salame décompose la structure d'un polymère en petits segments, et assigne à chaques segement une valeur numérique pour prédire une propriété physique ou chimique particulière.

Le total des valeurs numériques attribuées, pour toute la chaîne est le Permachor (π) du polymère calculé selon l'équation :

$$\pi = (\sum \pi_i)/\,\mathrm{n} \qquad \text{(II.12)}$$

Où π_i sont les valeurs des segments individuels des groupes latéraux et de la chaîne centrale et n est le nombre de segments (par unité de répétition de polymère).

Le coefficient de perméabilité P pour un système gaz/polymère est ensuite prévu à partir de la relation (II.13)

$$P = A\, exp\, (-s\pi) \qquad \text{(II.13)}$$

Où *A* et s sont des constantes d'un gaz particulier.

Les unités structurelles compilées par Salame n'incluent pas un grand nombre de polymères d'intérêt pour les séparations membranaires (exemple: polysulfones, polycarbonates, polyarylates, polyimides).

Cette méthode nous permet aussi d'estimer la diffusion et la solubilité .Cependant, l'emploi excessif de paramètres entièrement empiriques a rendu l'utilisation de cette méthode pour les nouveaux polymères tout à fait difficile. Ceci poussa à chercher une méthode encore plus simple.

b.Méthode du Yampolskii et al:

La méthode des contributions de groupe est basée sur la validité de plusieurs hypothèses:

- la structure des molécules complexes et, en particulier- les polymères, peut être représenté comme la somme de plusieurs groupes ou fragments, qui sont communs pour l'ensemble des structures (atomes, mono- ou multivalents radicaux), dont chacun est le caractérisé par un incrément selon la propriété considérée (coefficient de diffusion ou de perméabilité) ;
- Les incréments sont constants pour l'ensemble des structures complexes, c'est-à-dire ne dépendent pas d'une mutuelle entre les groupes ou leurs interactions possibles
- Les propriétés des molécules complexes peuvent être représentées la somme des incréments correspondants en tenant compte de facteurs du poids

Dans cette formulation, aucun élément des régularités fondamental ou semi-empirique n'est utilisés concernant les propriétés et structures chimiques des polymères. Cependant, il a utilisé une propriété A d'un polymère comme étant la somme:

$$Log\ A = \frac{1}{k}\sum_i q_i x_i \qquad \text{(II.14)}$$

k est un coefficient de normalisation, x_i est l'incrément caractéristique d'un groupe du i^{ème} type dans l'unité répétitive, $q_i$et est le nombre de ces groupes dans cette unité.

Il ya une différence fondamentale dans l'application de cette méthode pour la prédiction des propriétés des composés de faible masse moléculaire, comme cela a été fait, par exemple dans le livre de Benson **[55]** et de polymères. Dans le cas des composés de haut poids moléculaire, l'unité répétitive n'est pas une structure réelle d'une macromolécule. Par

conséquent, l'introduction de certains facteurs de normalisation est nécessaire. Il est évident que sans ces facteurs, on peut obtenir des résultats absurdes, lorsque l'on essaie de prédire les paramètres de permutation pour les polymères avec une structure d'unité répétitive plus complexe. Afin de surmonter cette difficulté, Van Krevelen **[56]** a fait une normalisation en divisant chaque contribution par la masse moléculaire du groupe, alors que Askadskii **[57]** et Robeson et al **[58]** ont utilisé le volume de Van der Waals ou le volume moléculaire, respectivement, dans le même but.

En principe, le nombre d'atomes dans une unité répétitive (k = Q) peut être utilisé comme un facteur de normalisation en plus de la masse moléculaire de cette unité (k = M), ou le volume de Van der Waals (k =V_W). Yampolskii a choisi dans son travail trois approches pour l'analyse: sans normalisation (k ≡ 1), et la normalisation par Q et M.

Yampolskii a choisi l'ensemble de groupe de base contenant: H, C (SP^3), C (sp^2), C (ar-sp^2), O=, ─O─, Si, Cl, Br, F, N≡ (nitriles), ─N<, =S=, ─S─. D'autre part, les paramètres de permutation ont été rapportés pour l'ensemble des gaz les plus connus : He, H_2, N_2, O_2, CO_2, CH4, ceci pour des centaines de structures polymériques

c. Méthode proposée par Paul et Park

Paul et Park ont essayé de développer une méthode qui permet de prédire la permselectivité des nouvelles structures polymériques. Cette méthode prédit effectivement la densité et utilise une estimation affinée du volume spécifique occupé par chaque gaz.

Le coefficient de perméabilité dans les cas simples est le produit d'une solubilité et d'un coefficient de diffusion, bien que pour les polymères vitreux, le processus de perméation soit plus complexe **[59]**.

$$P = D.S \qquad \text{(II.15)}$$

Toutefois, pour un gaz donné le coefficient de diffusion, D, varie considérablement d'un polymère à un autre par rapport au coefficient de solubilité. Le volume libre du polymère influence d'une manière remarquable sur le coefficient de diffusion. Le coefficient de solubilité dépend de ce paramètre aussi. Ainsi, la perméabilité d'un gaz dans une série de polymère peut être corrélée en termes de volume

$$P = A\,exp\,(-B\,/\,FFV) \qquad \text{(II.16)}$$

Où A et B sont des constantes pour un gaz particulier.

Le volume libre fractionnaire (*FFV*) est une mesure du volume théorique du polymère divisé par le volume réel de polymère. Le volume libre fractionnaire est généralement défini comme suit :

$$FFV = (V - V_0)/V \qquad \text{(II.17)}$$

V est le volume du polymère (défini par le volume par unité de masse ou mole de l'unité répétitive), V_0 est le volume occupé par les chaînes polymériques.

La manière la plus simple et la plus populaire utilisée est la définition proposée par Bondi **[60]**. Selon cette méthode, le volume libre spécifique dans un polymère, V_f est la différence entre le volume spécifique d'un polymère V_{sp} et son volume occupé V_{oc}, Bondi a proposé de calculer V_{oc} en $1.3V_w$, où V_w est la somme des volumes de Van der Waals par conséquent:

$$V_f = V_{sp} - 1.3V_w \qquad \text{(II.18)}$$

Paul et Park ont proposé une définition de FFV, qui suppose que Le FFV n'est pas constant pour tous les types des gaz

$$(FFV)_n = [(V - (V_0)_n]/V \qquad \text{(II.19)}$$

$$(V) = \sum_{k=1}^{k} Y_{nk}(V_w)_k \qquad \text{(II.20)}$$

Où : Y_{nk} est un ensemble de facteurs empiriques qui dépendent du gaz n et de groupe k. $(V_w)_k$ est le volume de Van der Waals pour le groupe k **[61].**

Il ya plusieurs autres méthodes de prédiction des propriétés permsélectives dont les une découlent du principe de la méthode de contribution des groupes et d'autre proposent leurs propres bases de départ, parmi ces méthodes ,nous avons la méthode de contribution des groupes proposés par Robeson, elle est basée sur le volume fractionnel de l'unité structurale i, la perméabilité est calculée selon l'équation :

$$\text{Ln } P = \sum_{i=1}^{n} \Phi_i \ln P_i \qquad \text{(II.21)}$$

Où P est la perméabilité du polymère, Φ la fraction volumique du groupe i, et P_i les perméabilités du groupe i. Les valeurs de Φ_i sont calculées en utilisant des logiciels de prédiction du volume molaire.

Les groupes (i) se composent de liaisons chimiques, ou de parties de liaisons chimiques,et les unités structurelles associées. Ils sont définis par le volume moléculaire contenu entre les lignes pointillées et pas nécessairement définies le long des liaisons chimiques ou complètes unités fonctionnelles **[62]**.

II.3.2 Théorie des graphes

La théorie des graphes a été utilisée avec succès pour prédire les propriétés importantes de composés organiques. En conséquence, elle a été appliquée pour prédire la perméabilité des gaz dans les polymères **[63]**.En utilisant le formalisme développé par Kier et Hall. La théorie des graphes est une approche topologique qui attribue un ensemble d'indices à différentes écritures, la connectivité ou les indices de Randic sont particulièrement appropriés pour cette tâche **[64]**.

Les indices de connectivité moléculaire sont des descripteurs quantitatifs des caractéristiques de la structure moléculaire qui sont essentiellement géométriques et topologique dans le caractère .La connectivité moléculaire peut être définie par la manière dont laquelle les atomes sont liés dans une molécule **[65]**

L'hypothèse de base, pour la théorie de la connectivité moléculaire, est que les structures moléculaires des unités répétitives peuvent être représentées par certains graphes, et elles sont décrites par les indices ${}^{m}\chi_{t}$. Ces derniers sont calculés pour les différents sous-graphes qui forment cette structure. Ici, m désigne le nombre de tranchants (liaisons reliant deux sommets ou atomes) dans le graphique et t représente le type de sous-graphe.

Afin de calculer ${}^{m}\chi_{t}$, un paramètre spécial δ est affecté à chaque atome inclus dans l'unité répétitive. Pour la première approximation, ce paramètre peut être considéré comme étant la valence de l'$i^{ème}$ atome moins le nombre d'atomes d'hydrogène qui sont liés à cet atome (par exemple 1 pour méthyle, 2 pour méthylène, etc.). Certaines règles semi-empiriques pour le calcul de δ ont été proposées par Kier et Hall.

$$ {}^{m}\chi_{t} = \sum_{j=1}^{n_m} {}^{m}S_{j} \qquad \text{(II. 22)} $$

$$^{m}S_{j} = \prod_{i=1}^{d}(\delta_i)_j^{-0.5} \qquad (II.23)$$

Dans ces équations, i est le numéro de l'atome, j est le nombre de sous-graphe, n_m est le numéro du sous-graphe de tranchant m et t sont le descripteur du x , alors que le calcul du nombre m d'arêtes (obligations), simples et doubles liaisons sont tous considérés comme un seul bord, d est le nombre total représentant les unités répétitives dans cette théorie.

Les propriétés d'intérêt, c'est la forme logarithmique du coefficient de perméabilité *log p(x)*, peut être exprimé en fonction de $^{m}\chi_t$:

$$\text{Log p}(\chi) = \Sigma C_i \left(^{m}\chi_t\right) z_i + b_0 \qquad (II.24)$$

Le paramètre b_0 peut être trouvé par la régression de certaines valeurs fixe arbitraires de z_i (par exemple, z = 0,5, 1, 2).

Une autre version d'application de la théorie du graphe pour la prédiction da la perméabilité des gaz a été décrite par Bicerano. Dans ce cas, un modèle semi-empirique, qui prend en considération la densité d'énergie de cohésion(CED) et le terme du volume libre, a été utilisé. Les δ paramètre ont été calculés selon le système de Kier et Hall.

Pour un ensemble qui comprenait 60 polymères (élastomères, vitreux, amorphes et semi-cristallins), une équation pour le calcul de la perméabilité à l'oxygène a été proposée comme suit :

$$P = A^* \, exp\,(-B^* v) \qquad (II.25)$$

Dans cette formule, $A^* > 0$ et $B^* > 0$ sont des constante *($A^* = 0.047\ cm^3(STP)cm/cm^3 s(cmHg)$; $B^* = 0.017622$)* **[64]**.

II.3.3 Réseaux de neurones artificiels

Une autre démarche récente pour la prédiction des perméabilités peut être fournie par la méthodologie des réseaux de neurones artificiels (ANN). ANN est un modèle théorique de traitement de l'information qui est inspiré à partir des systèmes nerveux biologiques. ANN représente une technique de modélisation prometteuse, en particulier pour les ensembles des équations non-linéaires **[66].**

Un réseau neuronal se compose d'un réseau de traitement des éléments, soi-disant «Neurones».Ces neurones sont organisés en couches dans le réseau comme indiqué sur la figure II.1. Chaque neurone calcule une sortie O_i de toutes les entrées X_j. Généralement, pour un neurone i, toutes les entrées $X_j$sont pondérées par un facteur W_{ij},ce qui résumé la dernière entréeU_i.

$$U_i = \sum W_{ij} X_j \qquad \text{(II.26)}$$

Cette entrée peut être transférée dans la sortie par un transfert mathématique. Différents types de neurones peuvent avoir différentes fonctions de transfert. Tout d'abord, un nombre suffisant de neurones dans la couche d'entrée est choisi. Deuxièmement, un certain nombre de numérisations de spectres infrarouges IR, avec perméabilités, est présenté au réseau. La troisième étape consiste à entraîner le réseau en changeant le poids de telle sorte qu'une erreur minimale soit obtenue entre les perméabilités prédites et mesurées.

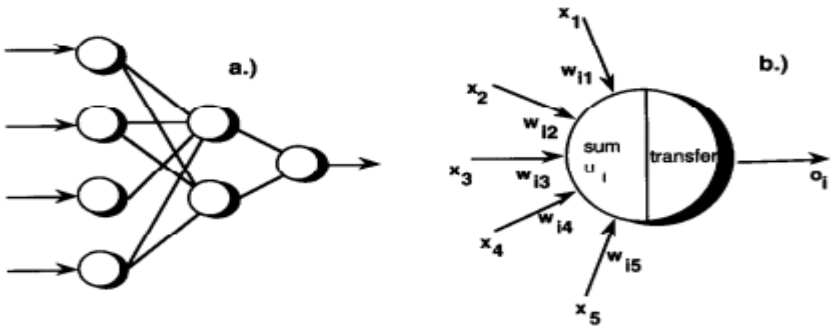


Figure II.1: Représentation schématique de (a.) une topologie de réseau neural (b) un élément de traitement (neurones**) [67]**.

La perméabilité de ce polymère est prévue sur la base de son spectre IR. La procédure est répétée pour tous les polymères dans l'ensemble d'entraînement. La qualité de l'ajustement a été caractérisée par ce qu'on appelle le facteur de performance, C:

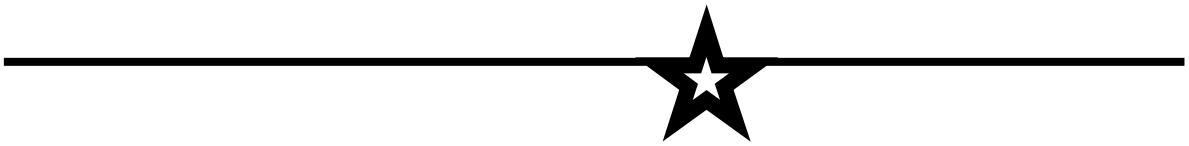
$$C = \frac{\sum |P_{pred} - P_{exp}|}{\sum |P_{av} - P_{exp}|} \qquad \text{(II.27)}$$

Où P_{av}est la moyenne des coefficients de perméabilité de l'ensemble. Le réseau neuronal fonctionne de manière prédictive si C est inférieur à 1. Pour C =l, la prédiction de la perméabilité pour un polymère inconnu serait égale à la perméabilité moyenne de tous les polymères présentés dans l'ensemble **[68]**.

II.3.4 Conclusion :

La prédiction des perméabilités constitue une étape préliminaire très importante dans l'étude de la séparation des gaz en utilisant des nouveaux polymères. Si on arrive à trouver des modèles prédictifs qui assurent des bonnes prédictions de perméabilité, on pourra par la suite les utilisé pour estimer ces propriétés pour les nouveaux polymères qui n'ont pas fait encore l'objet de synthèse.

L'objectif de l'étude de ces modèles, est de trouver un modèle applicable tout les types des polymères.



Chapitre -III-

Relation quantitatives structure- propriétés (QSPR)



CHAPITRE III

RELATION QUANTITATIVES STRUCTURE-PROPRIETES (QSPR)

III.1 Introduction :

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 **[69]**, qui recommande leur utilisation pour limiter le recours à l'expérience, donne un nouvel essor au développement de tels modèles prédictifs.

Dans les dernières années, l'utilisation des méthodes QSAR n'a pas cessé de progresser. Elle est même devenue indispensable en chimie pharmaceutique pour la conception de médicaments. Leur développement dans une gamme plus large d'applications leur ouvre d'ailleurs de grandes perspectives (ex : points d'ébullition, températures critiques, densité). Il s'agit de présenter ici le principe des modèles QSPR ainsi que ceux des différents outils employés pour leur mise en place et leur évaluation : bases de données expérimentales, descripteurs, outils d'analyse de données **[69]**.

III.2 Les relations quantitatives structure-propriétés "QSPR"

III.2.1 Définition de QSPR :

Le QSPR (Quantitative Structure-PropertyRelationship) est le procédé par lequel des liens quantitatifs sont établis entre la structure moléculaire d'un ensemble de composés avec une propriété physico-chimique. Les grandes phases de développement d'un modèle QSPR peuvent être décrites comme suit :

- Choisir des descripteurs adaptés au problème structure-propriété.
- Exploiter les valeurs des descripteurs comme variables, afin de définir une relation qui les corrèle à la propriété en question, à l'aide de machines d'apprentissage. C'est la fouille de données.
- Établir des critères de performance et de validation qui aideront au choix des meilleurs modèles pour le problème posé et estimer des incertitudes de prédiction **[70]**.

III.2.2 Principe de QSPR:

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

$$\text{Propriété} = f(\text{Descripteurs})$$

L'objectif d'une telle méthode est donc d'analyser les données structurales afin de détecter les facteurs déterminants pour la propriété mesurée. Pour ce faire, différents types d'outils peuvent être employés : régressions multilinéaires (MLR), régressions aux moindres carrés partiels (PLS), arbres de décision, réseaux de neurones, algorithmes génétiques. Une fois cette relation est mise en place et validée sur un jeu de validation, elle peut alors être employée pour la prédiction de la propriété de nouvelles molécules, pour lesquelles la valeur expérimentale n'est pas disponible, voire pour des molécules encore non synthétisées. De tels modèles peuvent également, dans certains cas, être utilisés pour mieux appréhender les phénomènes moléculaires mis en jeu dans la propriété d'intérêt, la figure III.1represnte le principe de la méthode QSPR **[71]**.

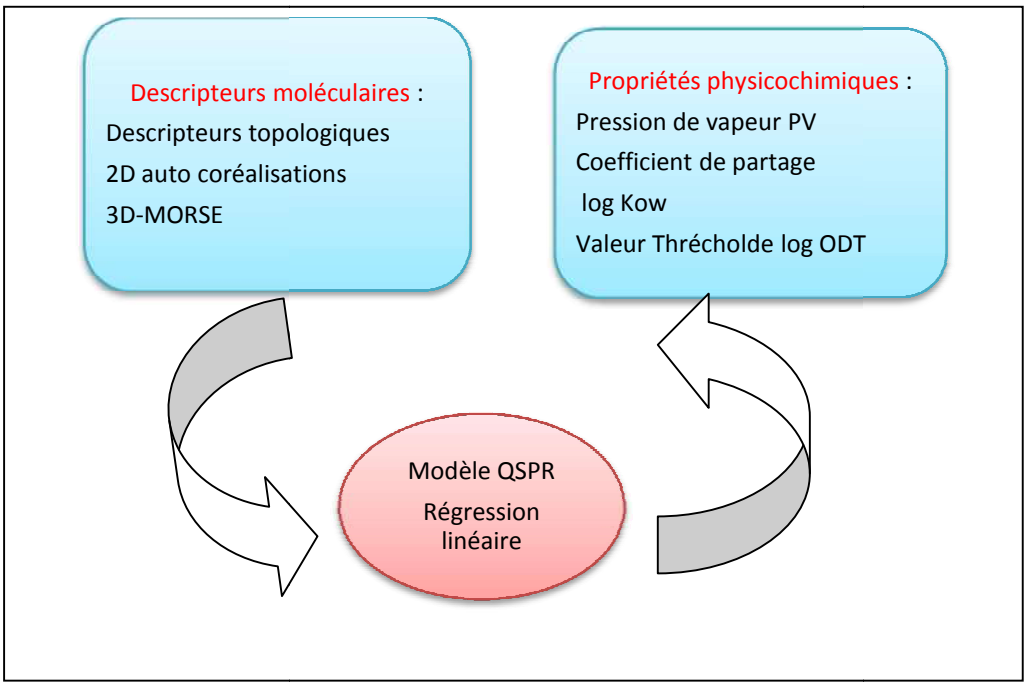


Figure III-1: Principe de la méthode QSPR **[71]**.

III.2.3 Les méthodes mathématiques utilisé par le model QSPR :

- Régression linéaire.
- Régression linéaire multiple (MRL).
- Régression en composantes principales(PCR).
- Régression moindre carre (PLS).
- Les réseaux de neurone (RN).

III.2.4 Les types d'apprentissages d'un modèle QSPR [72] :

Les techniques d'apprentissage ont étés développées en informatiques ces 15 dernières années. Il 'y a deux types d'apprentissages :

Non-supervisé: les activités et propriétés ne sont pas utilisées **(**seuls les descripteurs**)**. Ce type utilisé pour grouper les composés chimiques.

Supervisé: établi une relation entre les données entrées / input data (composés chimiques)et les données produites output data (propriétés).

III.3. Les descripteurs moléculaires:

III.3.1 Définition d'un descripteur

Un descripteur est un paramètre numérique propre à une structure chimique donnée, également, c'est un résultat d'un procédé mathématique ou d'une certaine expérience normalisée. Ces descripteurs représentent un moyen de corréler une structure chimique et une valeur de propriété physico-chimique ou biologique. Les descripteurs sont finalement des intermédiaires numériques se substituant à la molécule elle-même **[73]**. Ces descripteurs peuvent être utilisés pour évaluer l'influence de la structure moléculaire ou les relations propriétés-structure, aussi pour l'analyse de symétrie et la projection des bases de données des molécules **[74]**.

III.3.2 L'importance des descripteurs :

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

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

Les descripteurs moléculaires sont utilisés pour, une connaissance de statistiques, chimio métriques, et les principes des approches QSAR/QSPR sont nécessaires en plus de la connaissance spécifique du problème **[75]**.

III.3.3 Le choix des descripteurs :

Pour faire le bon choix de descripteurs moléculaires les points suivants doivent respectés :

1- Les descripteurs doivent avoir une relation directe, et une description suffisante à la structure moléculaire des composés chimiques à étudier.

2- Les descripteurs doivent avoir des liens avec les propriétés chimiques et physiques

et l'activité des composés chimiques.

3- La représentation descriptive de la totalité de la structure moléculaire des composées.

4- La diversité (variétés) des descripteurs.

5-La sélection exacte et la distinction entre les descripteurs décisifs (principaux), et d'écarts (secondaires).

6- La simplicité et l'efficacité des descripteurs, puisque la complexité provoque une perturbation sur les résultats d'estimation **[76]**.

7-Il n'est pas nécessaire d'avoir des relations parfaites, entre les variables (descripteurs) prédictives, parce qu'elles causent des difficultés sur la régression linéaire (confusion parfaite d'effets et perturbation dans l'interprétation des résultats) **[77].**

III.3.4 Choix du nombre de descripteurs :

- L'estimation des propriétés nécessite certain nombre de descripteurs moléculaires (certains blocs).

- Le choix du nombre de descripteurs convenables est très important.

III.3.5 Réduction le nombre de descripteurs :

Dans un calcul quanto-chimique un grand nombre de descripteurs sont collectés pour la modélisation d'une grandeur donnée, car les facteurs déterminants du processus étudié ne sont a priori pas connus. Cependant, les descripteurs envisagés n'ont pas tous une influence significative sur la grandeur modélisée, et les variables ne sont pas toujours mutuellement indépendantes.

De plus, le nombre de descripteurs, c'est-à-dire la dimension du vecteur d'entrée, détermine la dimension du vecteur des paramètres à ajuster. Si cette dimension est trop importante par rapport au nombre d'exemples de la base d'apprentissage, le modèle risque d'être sur ajusté à ces exemples, et incapable de prédire la grandeur modélisée sur de nouvelles observations. Il est donc nécessaire de réduire la dimensionnalité des variables d'entrée, plusieurs approches sont possibles pour résoudre ce problème **[78].**

- ✓ Réduire la dimension de l'espace des entrées.
- ✓ Remplacer les variables corrélées par de nouvelles variables synthétiques, obtenues à partir de leurs combinaisons.
- ✓ Sélectionner les variables les plus pertinentes.

III.3.6 Les méthodes utilisées :

A. Analyse en composantes principales(ACP)

L'Analyse en Composantes Principales (ACP) fait partie du groupe des méthodes descriptives multidimensionnelles appelées méthodes factorielles. L'ACP est une technique qui permet de trouver des espaces de dimensions plus petites dans lesquels il est possible d'observer au mieux les individus. Sa démarche essentielle consiste à transformer les variables quantitatives initiales, plus ou moins corrélées entre elles, en des variables quantitatives, non corrélées, combinaisons linéaires des variables initiales et appelées composantes principales. Les composantes principales sont donc de nouvelles variables indépendantes, combinaisons linéaires des variables initiales, possédant une variance maximale **[79].**

B. Algorithmes génétiques :

Les algorithmes génétiques (AG) sont des méthodes d'optimisation stochastiques qui font partie des algorithmes ´évolutionnaires. Les AGs sont des algorithmes itératifs basées sur la reproduction et l'évolution naturelle des individus en utilisant les principes de la survie des individus considérés comme les plus forts ou les mieux adaptés à l'environnement. Avec les AGs on cherche à optimiser une fonction (objectif) donnée dans un espace de recherche, des individus. Pour l'optimiser, on définit une fonction dévaluation reliée à cette fonction objective et appliquée sur chaque individu. En général, le fonctionnement d'un AG est basé sur les phases suivantes : Initialisation, évaluation, reproduction, retour **[80]**.

C. Régression linéaire multiple (MLR):

La régression multilinéaire (MLR) est la méthode la plus simple et la plus communément employée pour le développement de modèles prédictifs. Elle repose sur l'hypothèse qu'il existe une relation linéaire entre une variable dépendante **y** (ici, la propriété physico-chimique) et une série de n variables indépendantes xi (ici, les descripteurs moléculaires).

D. Régression linéaire simple(RLS) :

La régression linéaire simple traite la question suivante :

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

L'objectif central de la RLS est de trouver de bons estimateurs de ces quantités ainsi que d'identifier les liens de dépendance (covariance) entre ces estimateurs **[81]**.

III.3.7 Le classement des descripteurs:

Le classement de ces descripteurs moléculaires doit être suivant l'importance et l'efficacité (le pouvoir descriptif), la qualité d'informations sur la propriété à estimer. (Les descripteurs moléculaires décisifs et d'autres écarts).

III.3.8 Calcul des descripteurs moléculaires :

Afin de calculer les descripteurs moléculaires d'une molécule donnée, on doit passer par plusieurs étapes avant d'établir la base de données contenant les valeurs de ces descripteurs, la première de ces étapes consiste à représenter la structure moléculaire dans forme bien déterminée après avoir dessiné sa structure moléculaire.

Différents formats de fichiers ont été décrits pour représenter la structure des molécules. Les plus connus sont hin, SMILES, SDF, mol2 (Tripos) et PDB (Protein Data-Bank). Le format hin permet de représenter les molécules sur le logiciel hyperchem qui fait une optimisation moléculaire qui se fait en deux étapes (la mécanique moléculaire de la force (MM + et AM1 (Austin Model 1)).

Afin d'exploiter au maximum les informations contenues dans les structures moléculaires, celles-ci sont traduites en une série de grandeurs (en général scalaires) qui quantifient leurs caractéristiques physico-chimiques et structurelles. Ces descripteurs peuvent être utilisés pour évaluer des relations structure moléculaire-activité ou structure-propriété, ainsi que pour l'analyse de similitude et de dépistage. Pour calculer ces descripteurs il existe plusieurs logiciels, parmi les plus utilisé on peut citer :

Academic package

- ChemAxon (Marvin, jchem): calcul de descripteurs topologiques
- OpenEye Sci. Soft. (FILTER)
- Treweren Consultant (THINK)
- DRAGON
- CODESSA

Calcul en ligne

- Molgen

- PreADME
- CORINA (autorise la transformation de structure 2D→3D de 1000 molécules gratuitement)

III.3.9 Les types des descripteurs moléculaires :

Il existe quatre types de descripteurs moléculaires 0D, 1D, 2D, 3D.

Les descripteurs 0-D **:** sont calculés à partir de la formule brute de la molécule. Les descripteurs 1-D **:** contient le nombre d'atomes, les groupements fonctionnels, les propriétés des molécules, et les charges. Ils sont calculés à partir de la formule brute de la molécule et ils représentent des propriétés très générales, comme le poids moléculaire, nombre d'atomes, etc.

Les descripteurs 2-D: contient des descripteurs simples et autres dérivés des algorithmes appliqués à une représentation topologique. Ils sont obtenus à partir de la structure 2D constitutionnels, fragmentées ou des propriétés physico-chimiques (donneur de liaison H, accepteur de liaison H, cation, anion, etc.) de la molécule, dans cette catégorie rentrent les indices topologiques.

Les descripteurs 3-D: contient des descripteurs moléculaires dérivés d'une représentation géométrique. Ils sont calculés à partir de la structure 3D de la molécule. Ils peuvent être quantiques, de surfaces moléculaires ou de volume moléculaire **[82]**.

III.3.10 Bloc des descripteurs [83]:

Les descripteurs moléculaires calculés par DRAGON sont divisés en 22 blocs logiques. Ces descripteurs moléculaires sont représentés dans la figure III.2.

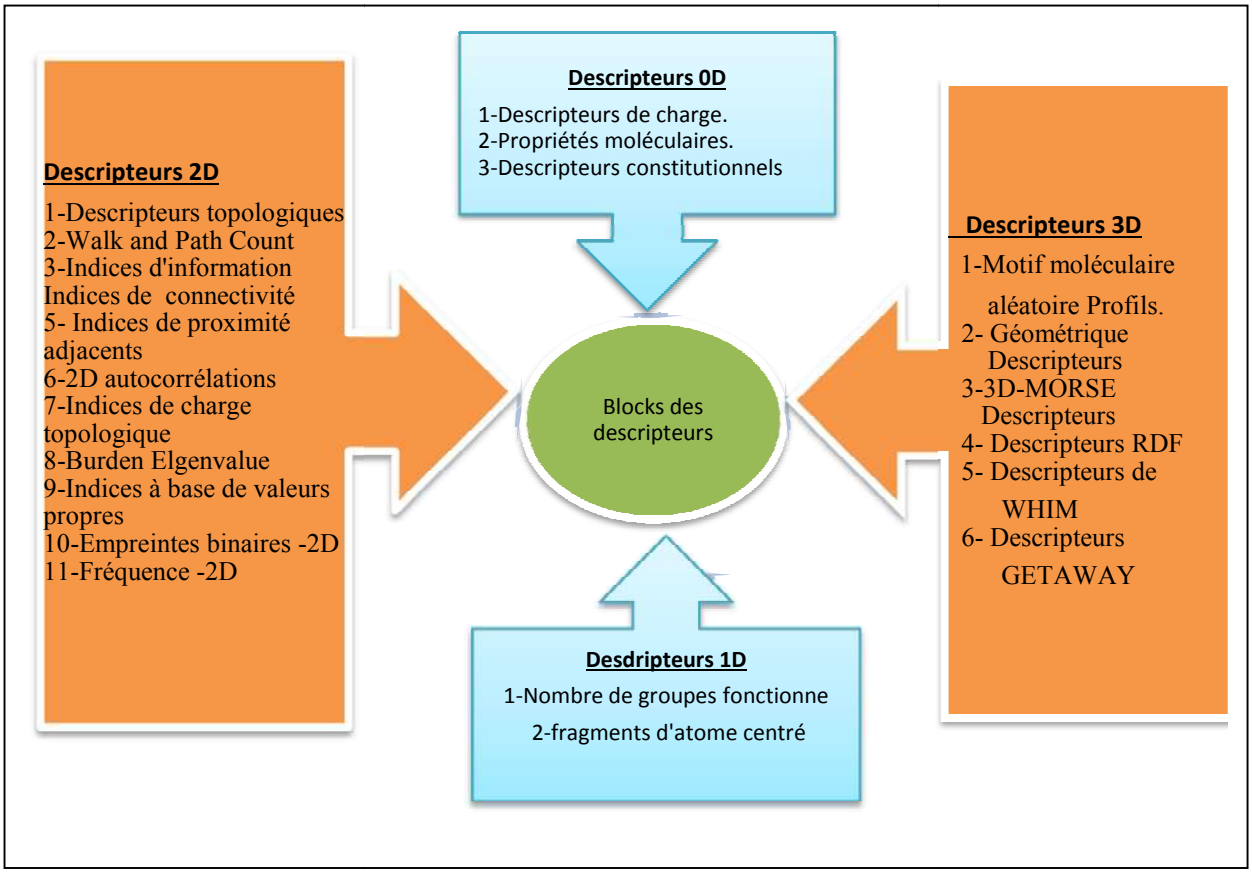


Figure III.2: Représentation des blocs des descripteurs moléculaires **[83]**.

III.4 Les réseaux de neurones :

Les succès récents des réseaux de neurones formels ne doivent rien à la biologie, mais sont dus exclusivement à une meilleure compréhension des propriétés mathématiques fondamentales de ces réseaux. On est bien éloigné de la réalisation de machines susceptibles de reproduire une partie, même infime, des capacités de calcul des systèmes nerveux les plus simples. Àl'heure actuelle, les réseaux de neurones formels doivent être considérés comme des extensions puissantes de techniques statistiques classiques.

III.4.1 Le neurone formel (artificiel)

Un neurone formel est un processeur très simple (simulé sur ordinateur ou réalisé sur circuit intégré) imitant grossièrement la structure et le fonctionnement d'un neurone biologique. La première version du neurone formel est celle de W. S. Mc Culloch et W. Pitts, et date de 1943. C'est un automate binaire qui réalise une somme pondérée de ses entrées, le

potentiel, et compare ce potentiel à un seuil (nul) : s'il est supérieur, la sortie vaut +1 et le neurone est actif, s'il est inférieur, la sortie vaut -1 et le neurone est inactif (figure III.3) **[84]**.

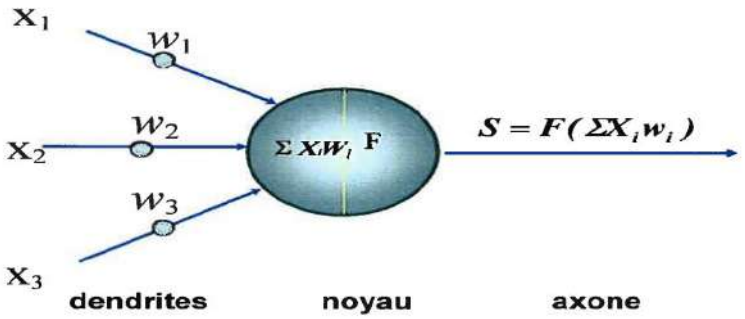


Figure III.3 : le neurone formel [83].

Les signaux (x_1, x_2, x_3) représentent les entrées (inputs), et le noyau du neurone est représenté par un sommateur et par une fonction d'activation. Lorsque le neurone reçoit les signaux d'entrées, ceux-ci sont d'abord multipliés par des coefficients de pondération appelés encore 'poids de connexion (W_1, W_2, W_3). Ensuite la somme pondérée $\sum X_i Y_i$ est comparée à une valeur de seuil du neurone. Si le seuil est franchi, une fonction d'activation est appliquée sur la somme pour donner une valeur de sortie (S), sinon, le neurone reste inhibé. La fonction d'activation semble être un élément constitutif important. Souvent, des fonctions non linéaires sont nécessaires **[85]**.

À titre d'exemple, les fonctions les plus couramment utilisées sont:

a. La fonction sigmoïde standard (encore appelée fonction logistique)

$$Y = \frac{1}{1+e^{-x}} \qquad \text{(III.1)}$$

b. La tangente hyperbolique

$$Y = \frac{2}{(1+e^{-2x})^{-1}} \qquad \text{(III.2)}$$

c. La fonction Gaussienne

$$Y = e^{-x^2/2} \qquad \text{(III.3)}$$

III.4.2 Analogie avec le modèle biologique [86] :

La figure III.4 illustre n neurones biologiques avec plusieurs signaux d'intensité x et force synaptique w pénétrant un neurone avec un seuil b, ainsi que le système de neurone artificiel correspondant.

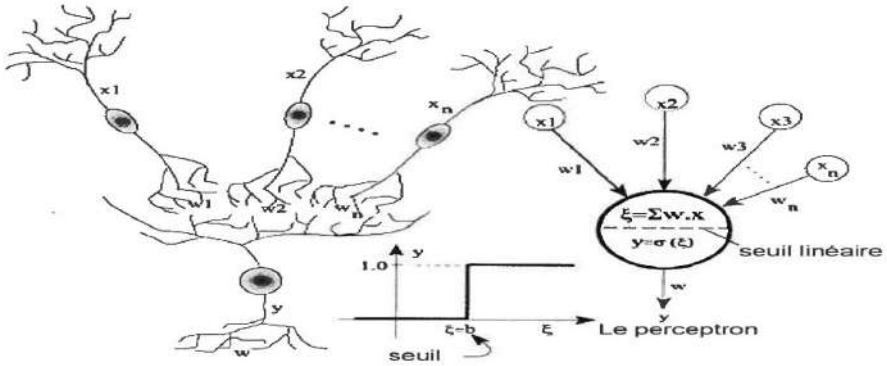


Figure III.4 : Interaction de signaux provenant de n neurones et analogie, la sommation de signaux dans un neurone artificiel **[86]**.

Les points d'analogie entre le neurone artificiel et le neurone biologique sont résumés comme suit :

- Les connexions entre les neurones représentent les axones et les dendrites.
- Les poids des connexions représentent les synapses.
- Le seuil représente l'activité dans le soma.

III.4.3 Topologie de connexion des neurones :

La topologie du réseau est la façon dont les neurones sont reliés, elle se décline de deux manières : dans la première, chaque neurone communique avec tous les autres, et même avec lui-même (réseau totalement connecté), dans la seconde les neurones sont organisés en couches, chaque couche reçoit alors des signaux de la couche précédente et en envoie à la couche suivante, mais les neurones appartenant à la même couche ne communiquent pas entre eux. On distingue deux grands types d'architectures de réseaux de neurones : les réseaux de neurones non bouclés et les réseaux de neurones bouclés **[87]**.

III.4.3.1 Réseaux statiques non bouclés :

Un réseau de neurones non bouclé réalise une (ou plusieurs) fonction algébrique de ses entrées, par composition des fonctions réalisées par chacun

de ses neurones. Il est représenté graphiquement par un ensemble de neurones connectés entre eux, l'information circulant des entrées vers les sorties sans retour en arrière.La figure III.5 représente un réseau de neurones non bouclé qui a une structure particulière, très fréquemment utilisée : il comprend des entrées, une couche de neurones cachés et des neurones de sortie. Les neurones de la couche cachée ne sont pas connectés entre eux. Cette structure est appelée perceptron multicouche.

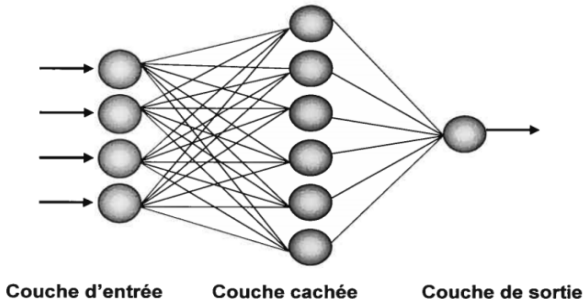


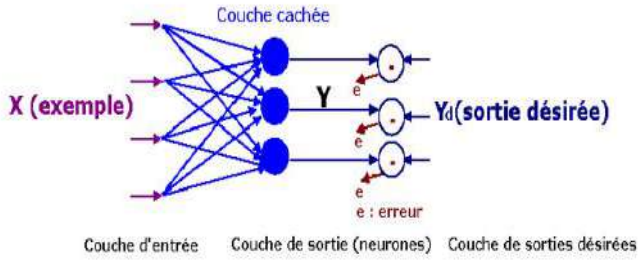
Figure III.5 : un perceptron multicouche **[88]**.

Les réseaux de neurones non bouclés sont des objets statiques, si les entrées sont indépendantes du temps, les sorties le sont également. Ils sont utilisés principalement pour effectuer des tâches d'approximation de fonction non linéaire, de classification ou de modélisation de processus statiques non linéaires **[89]**.

III.4.3.2 Réseaux à une couche cachée :

Le perceptron à une couche cachée (figure III.6) est une forme réduite du perceptron multicouches, il est caractérisé par :

- Entrée et sortie vectorielles.
- Une fois l'entrée X (exemple d'apprentissage) est présentée aux réseaux, les activations de neurones sont calculées une fois les réseaux sont stabilisés.
- Les erreurs ($e = Y - Y_{\square}$) sont calculées.
- Les erreurs (e) sont minimisées par un algorithme d'apprentissage **[90]**.



FigureIII.6 : Réseau à une couche cachée **[90]**.

III.4.4 Apprentissage des réseaux de neurones:

L'apprentissage d'un réseau de neurones signifie qu'il change son comportement de façon à lui permettre de se rapprocher d'un objectif défini. Ce but est normalement l'approximation d'un ensemble d'exemples ou l'optimisation de l'état du réseau en fonction de ses poids pour atteindre l'optimum d'une fonction objective fixée a priori.

Il existe deux types d'apprentissages principaux. Ce sont l'apprentissage supervisé et l'apprentissage non-supervisé **[91]**.

L'apprentissage supervisé : consiste à présenter au réseau les exemples ainsi que les réponses désirées et à modifier les poids en fonction des réponses fournies par le réseau. Ce type d'apprentissage est généralement utilisé par les réseaux multicouches **[92,93]**. L'ajustement des poids se fait souvent selon les règles de l'algorithme de rétropropagation (figure III.7)

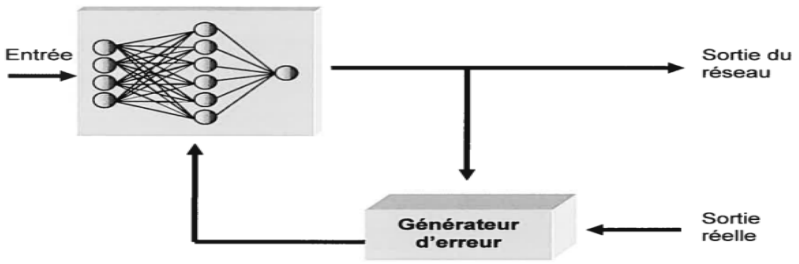


Figure III.7. Réseau associatif avec apprentissage supervisé **[92]**.

L'apprentissage non-supervisé : le réseau décide lui-même quelles sont les bonnes sorties. Cette décision est guidée par un but interne au réseau qui exprime une configuration idéale à atteindre par rapport aux exemples introduits **[94]**.

III.4.4.1 Algorithme de rétropropagation du gradient :

L'algorithme de rétro-propagation du gradient est utilisé dans le perceptron multicouche. Un réseau de type rétropropagation se compose d'une couche d'entrée, une couche de sortie et zéro ou plusieurs couches cachées **[95]**. Les connections sont permises seulement d'une couche inférieure (plus proche de la couche d'entrée) vers une couche supérieure (plus proche de la couche de sortie). Il est aussi interdit d'avoir des connections entre des neurones de la même couche.

La couche de sortie se compose normalement des neurones linéaires qui calculent seulement une somme pondérée de toutes ses entrées. Les couches cachées contiennent des neurones avec des fonctions d'activation non-linéaires, normalement la fonction sigmoïde.

La procédure d'apprentissage se décompose en deux étapes. Pour commencer, les valeurs d'entrées sont fournies à la couche d'entrée. Le réseau propage ensuite les valeurs jusqu'à la couche de sortie et donne ainsi la réponse du réseau. A la deuxième étape les bonnes sorties correspondantes sont présentées aux neurones de la couche de sortie, qui calculent l'écart, modifient leurs poids et rétropropagent l'erreur jusqu'à la couche d'entrée pour permettre aux neurones cachés de modifier leurs poids de la même façon.

III.4.4.2 Algorithme de Levenberg – Marquardt :

L'algorithme de Levenberg-Marquardt n'est pas calculé de manière explicite mais approché en utilisant une fonction dépendant du gradient **[96]**. Est une méthode du second ordre utilisé comme algorithme de minimisation itérative des poids du modèle pour résoudre le problème du temps. Cette technique permet en effet de converger vers le minimum de la fonction d'erreur plus rapidement du fait qu'elle considère la dérivée seconde de la fonction d'erreur. Il importe également d'éviter que cette optimisation ne s'arrête comme nous l'avons vu, non pas sur le minimum vrai, mais sur un des minima locaux de la fonction d'erreur.

III.4.5 Conception d'un réseau de neurones :

Pour construire un réseau de neurones, la première chose à faire n'est pas de choisir le type de réseau mais de bien choisir les échantillons de données d'apprentissage, test et validation. Ce n'est qu'ensuite que le choix du type du réseau interviendra. Voici chronologiquement les quatre grandes étapes qui doivent guider la création d'un RNA.

III.4.5.1 Choix des échantillons :

Cette étape est cruciale et va aider le concepteur à déterminer le type du réseau le plus approprié pour résoudre son problème. La façon dont se présente l'échantillon conditionne :

- Le type du RNA.
- Le nombre de cellules d'entrée et de sortie.
- La façon dont il faudra mener l'apprentissage, les tests et la validation.

III.4.5.2 Elaboration de la structure du réseau :

La structure du réseau dépend étroitement du type des échantillons. Il faut tout d'abord choisir le type de réseau : Un perceptron standard, un réseau de HOPFIELD, un réseau de KOHONEN etc. Dans le cas du perceptron par exemple, il faudra aussi choisir le nombre de neurones dans la couche cachée.

III.4.5.3 Apprentissage :

L'apprentissage est l'avant dernière phase d'élaboration d'un réseau de neurones, elle consiste à calculer les pondérations optimales des différentes liaisons, en utilisant un échantillon. La méthode la plus utilisée est la rétropropagation : On entre des valeurs dans la cellule d'entrée, et en fonction de l'erreur obtenue en sortie, on corrige les poids accordés aux pondérations. C'est un cycle qui est répété jusqu'à ce que la courbe d'erreur du réseau ne soit pas croissante

III.4.5.4 Validation :

Une fois le réseau calculé, il faut procéder à des tests pour vérifier qu'il réagit bien comme on le souhaite : c'est la validation. La méthode la plus simple étant de garder une partie de l'échantillon réservée à l'apprentissage, pour la validation et faire ainsi une validation hors échantillon.

III.5 Conclusion :

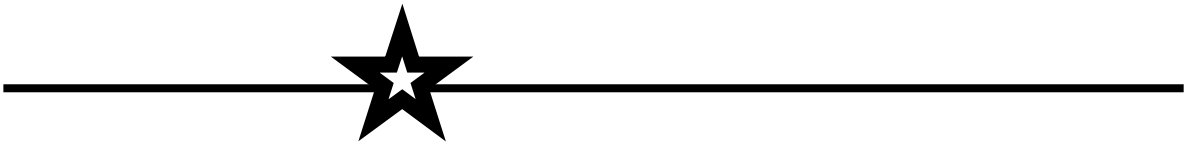
Dans la pratique, l'objectif d'une modélisation statistique n'est pas d'ajuster finement un modèle sur un ensemble d'apprentissage, mais d'obtenir un bon compromis entre performances d'apprentissage et performances de généralisation. Lors de la construction d'un réseau de neurones, l'objectif donc est de déterminer la meilleure architecture possible et le

meilleur jeu de poids au sens d'un critère donné. Sous cet angle, la construction du réseau de neurones peut être vue comme un problème de recherche d'une solution dans l'espace des architectures et des poids.



Chapitre -IV-

Modélisation neuronale pour la prédiction de la perméabilité



CHAPITRE IV

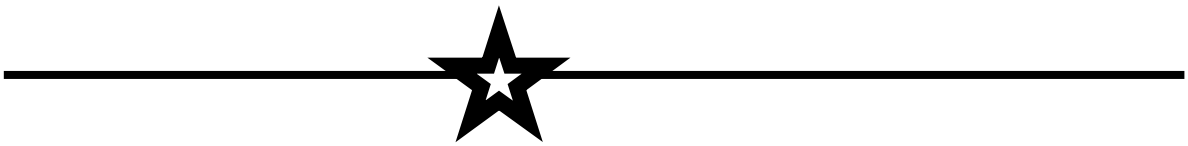
MODÉLISATION NEURONALE POUR LA PRÉDICTION DE LA PERMÉABILITÉ

Ce chapitre traite la prédiction de perméabilité des polymères. Pour ce faire, on commencera par une première partie intitulée « élaboration de la base de données ».Cette dernière comportera la récolte et la représentation des structures moléculaires des polymères appartenant à différentes classes jusqu'au calcul des descripteurs par les logiciels appropriés ; la deuxième partie présente l'algorithme génétique comme méthode pour la réduction de la taille de la base de données. La troisième partie est consacrée à la modélisation neuronale des données et l'optimisation des paramètres des réseaux de neurones.



Partie-1-

Elaboration de la base de données



IV.1.1 Les logiciels et Matériels utilisés :

Nous avons démarré les calculs avec une station de travail Toshiba intel (R)core ™ i3cpu ; 2.93 GHz de 4 Go RAM avec système d'exploitation 64 bits.

IV.1.2 Collecte des données

Dans cette étape nous avons collecté et sélectionné 73 molécules, ces molécules appartiennent à 4 principales classes de polymères :norbornènes , fluoropolymères , polyacétylènes et silicones à partir des différents articles leurs formules monomérique se trouve dans l'annexe.les différentes étapes de réalisation de la base de données sont présentées sur La figure IV.1.

IV.1.3 Polymères collectés utilisés dans ce travail :

Les polymères dans ce travail sont les plus utilisés dans les travaux publiés concernant la perméabilité des gaz à cause de leurs valeurs dans plusieurs industries.

- **Les norbornènes :** sont des polymères vitreux ayant une température de transition vitreuse supérieure à la température ambiante,est un hydrocarbure ponté de formule $\mathbf{C_7H_{10}}$ les structures des norbornènes sur lesquelles on a travaillé sont citées au tableau .1 de l'annexe, Les perméabilités expérimentales de ces polymères aux O_2et CO_2sont illustrées dans le tableau IV.1.

Tableau IV.1: Les perméabilités expérimentales des Norbornenes**[97,98,99 ,100,101]**.

Polymère	T_{exp}(°C)	P_{exp}(barrer)	
		$P(CO_2)$	$P(O_2)$
PDSNB	35	445	95
PTFmPhONDI	35	164.6	39.4
DFHFPNB	35	102	27.7
PCYNDI	35	9.2	2.3
POLY-2A	35	77.5	15.5
PhNDI	30	11.4	1.44
P5FPNDI	30	25.17	6.08
PB3FMPNDI	30	67.26	13.53
P5FPONDI	30	16.51	2.66
P5FPNDI/PPNDI	30	16.30	3.55

PTMSNB	22	89	30
PhONDINB	30	6.14	0.99
PNB	24	15.4	2.8
PNBSi	24	79	21
PNBDSi	24	64	20
PNBSiCSi	24	67	16
PTCNSi	24	112	28
P(6.5-NBSi$_2$)	24	455	95
P(5.5-NBSi$_2$)	24	277	54
P(2,3- NBDSi$_2$)	24	358	71
P(3,4-TCNSi$_2$)	24	396	89
P(3,3-TCNSi$_2$)	24	258	50
APNB	24	33.6	6.9
APNBSi	24	4350	780
APTCNSi	24	2230	450
POFPNB	22	200	55
PTCNSi2V	20	11280	2380
PTCNSi2g	20	19900	4750

- **Les fluoropolymères :** sont des matériaux plastiques de haute performance utilisés dans desenvironnements chimiques contribue à leurs propriétés antiadhésives et de réduction du frottement, ces polymères appartient a la famille des polymères vitraux, le tableau IV.2 résume les coefficients de la perméabilité des CO_2et O_2de cette famille de polymères

Tableau IV.2: Perméabilité expérimentales des fluoropolymères **[102]**.

Polymère	T_{exp}(°C)	P_{exp}(barrer)	
		$P(CO_2)$	$P(O_2)$
PTFE	25	420	1200
PCTFE	25	30	160
AF1600	25	600	530
AF2400	25	1200	2800
HyflonAD80	25	194	473
HyflonAD60	25	51	124
Cytop	25	16	17

- **Les polyacétylènes** : les données expérimentales de la perméabilité de certains gaz à travers les polyacétylènes sont illustrées dans le tableau suivant :

Tableau IV.3: Perméabilités expérimentales des polyacétylènes**[103 ,104]**.

Polymère	T_{exp}(°C)	P_{exp}(barrer)	
		$P(CO_2)$	$P(O_2)$
PMP	25	10700	2700
PTMSP	25	180	40
MeSiMe3	25	19000	3000
H tBu	25	560	130
$MeSiMe2CH2SiMe_3$	25	310	75
SiMe2CH2CH2SiMe3	25	150	50
Me n-C7H15	25	130	50
Cl C8H17	25	170	47
C6H13	25	130	32
C4H9	25	180	35

SiMe2C6H13	25	71	18
HCHn-C5H11SiMe3	25	120	27
CH C3H7Si	25	70	19
PhMe	25	25	6.3
Ph Et	25	40	12
Ph n-C6H13	25	48	14
Ph Cl	25	23	5.1
HCH(C3H7)SiMe2	25	15	8.1
HCH(C3H7)SiMe2Ph	25	54	9.5
CH=C(CF3)	25	130	25
CH=C(Sime3)	25	290	78
CCl=CPh	25	23	5.1
CMe=C(GEMe3)	25	/	1100
CMe=C=SiMeEt	25	/	970
Me=C(SiEt3)	25	/	860
Me=C(SiMe2Pr)	25	/	100

- **Les silicones** :les silicones sont caractérisée par des forces intermoléculaires et énergie de liaison faible appartient a la famille des élastomères ,cette famille ayant la propriété de viscoélasticité. Les élastomères sont amorphes polymères existants dessus de leur température de transition vitreuse , de sorte que beaucoup de mouvement segmentaire sont possible, le tableau IV.4 présente les données expérimentales de la perméabilité de CO_2 à travers les silicones.

Tableau IV.4:Perméabilités expérimentales des silicones **[104]**.

Polymer	T_{exp}(C°)	$P(CO_2)$
CH3Si	35	4570
CH3C3H7SiO	35	1520
CH3C8H17SiO	35	917
CF3C2H4SiO	35	1210
C3H5CH3SiO	35	226
CH3SiCH2SiO	35	1310
CH3SiCH2	35	542
MeSiMe3	35	52.3

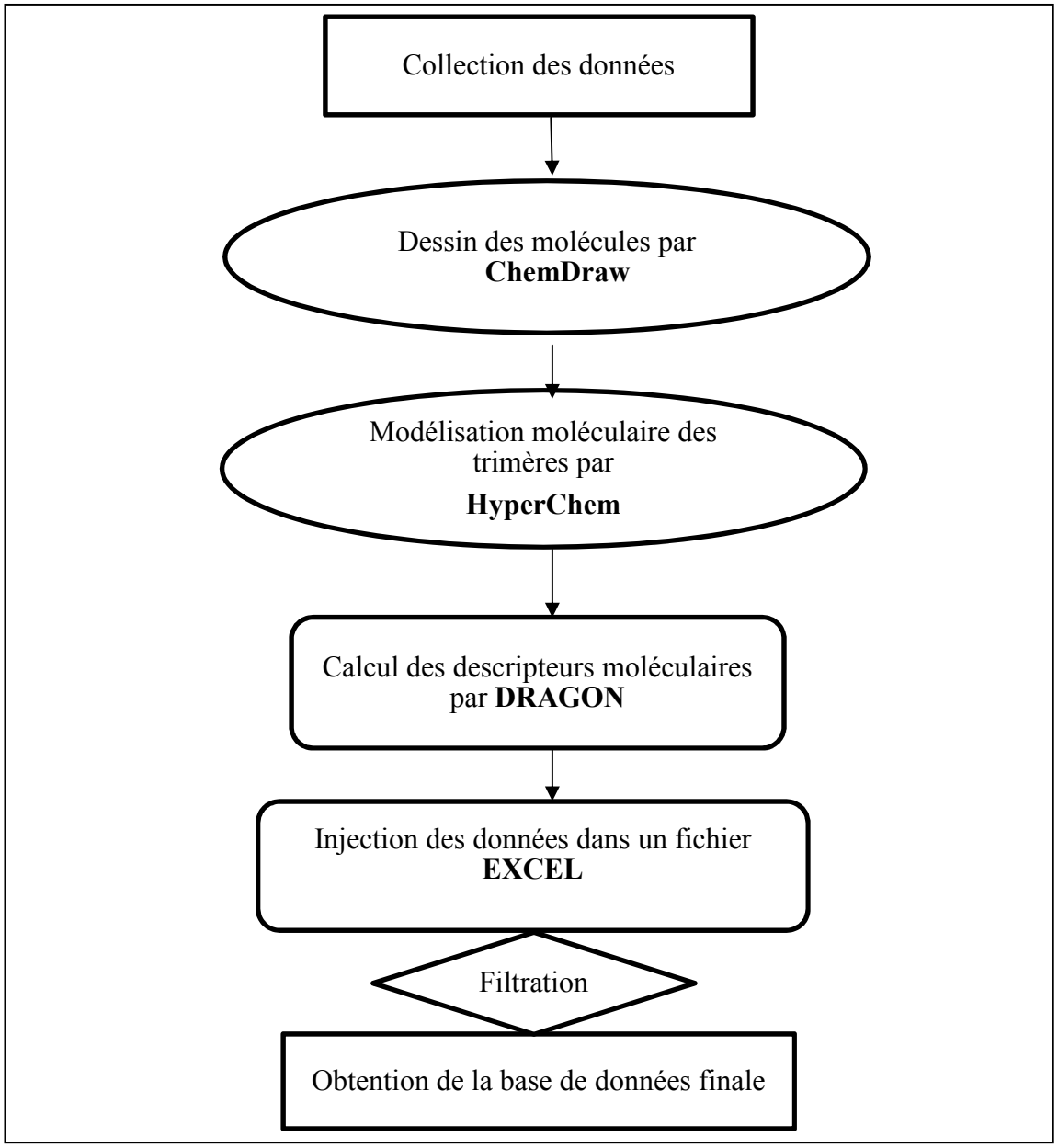


Figure IV.1 : Organigramme des déférentes étapes de réalisation de la base de données

IV.1.4 Dessin des molécules par ChemDraw

Pour réaliser cette tâche, nous avons choisi le logiciel ChemBioDraw comme un outil de dessin, nous avons dessiné les molécules (figure IV.2) sous forme de trimères pour les considérations suivantes :

- un trimer est une représentation très simple d'un polymère et est valable pour son optimisation géométrique à fin d'évaluer un maximum d'interaction intermoléculaire.
- L'avantage de travailler avec un trimère réside dans la rapidité de l'optimisation de la structure et la facilité du calcul des descripteurs.
- Le segment trimère représente le mieux la structure originale du polymère est celui du milieu (unité répétitive), tel qu'il est par les caractéristiques physico-chimiques, électronique et stérique des unités adjacentes et aussi préserve les caractéristiques structurales du polymère …

Une foi la molécule dessinée, elle est sauvegardée sous forme d'un fichier MDL Mol file (*.mol) pour l'étape suivante qui est la pré-optimisation par HyperChem

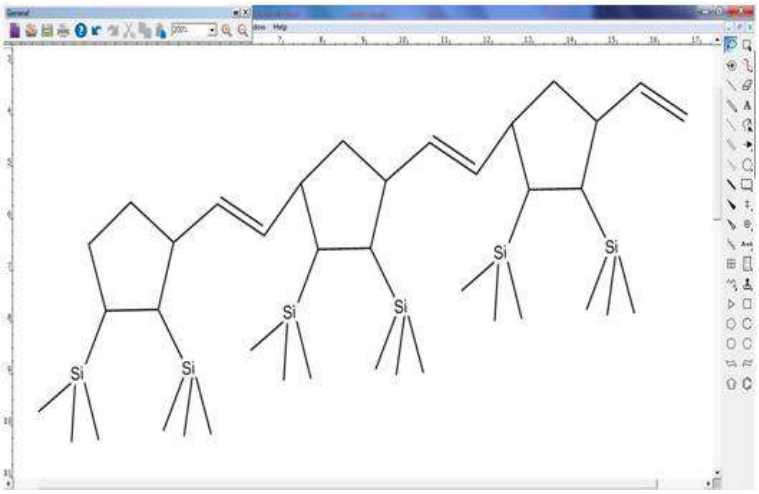


Figure IV.2: Etape de représentation des structures par ChemDraw.

IV.1.5 Modélisation moléculaire des trimères par HyperChem

L'HperChem (Figure IV .3) est un logiciel très apprécié par les scientifiques à cause de la diversité des modules qu'il présente ainsi que sa puissance présentée à travers les diverses méthodes de modélisation et d'optimisation géométriques qu'il procure **[105]** en suivant les étapes :

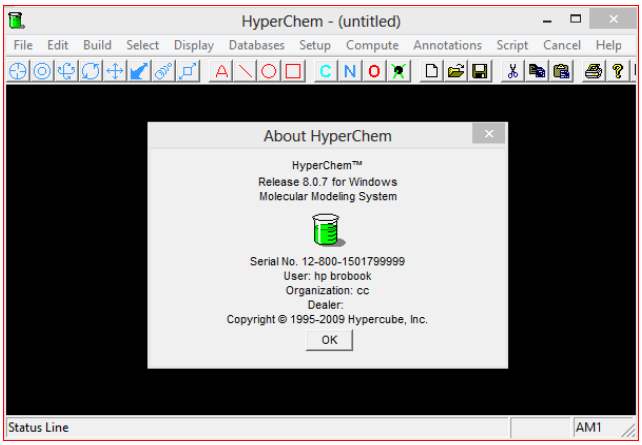


Figure IV .3 :HyperChem version8.0.7 pour la modélisation moléculaire des trimères

En ouvrant l'interface d'Hyperchem on extrait le dossier qui contient le fichier d'extension (*.mol) déjà préparé par ChemDraw puis on réalise une pré-optimisation des structures que consiste à une procédure basé sur la mécanique moléculaire de champs de force (MM^{+} :Force Field Molecular). Avec les instruction suivants :

(Setup/MolecularMechanics/MM^{+}/Compute/ GeometryOptimization/Gradient conjugate polak Ribière).Une fois le calculs converge la structure 3D préoptimisée est sauvgardé.

L'optimisation finale de notre structure par la méthode des orbitales moléculaires semi empirique AM1 avec l'algorithme de polak-Ribière d'une valeur limite du RMS du gradient =de 0.01 $KcalA^{\circ -1}$ mol^{-1}.cette est très longue, elle prend des temps considérables parfois plusieurs heures, mais indispensable au calcul des descripteurs.

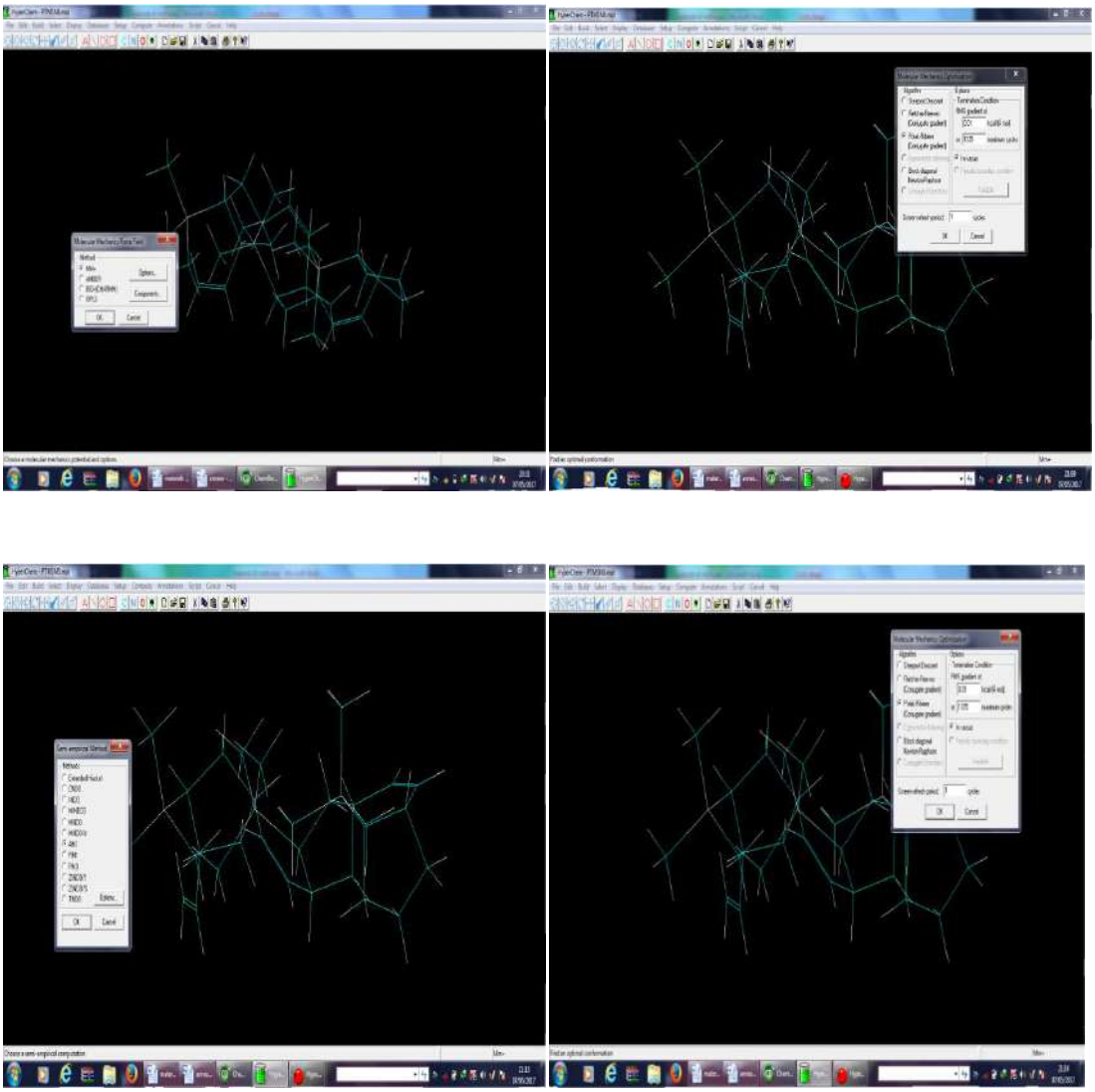


Figure IV.4 : Etape des pré-optimisations des molécules respectivement.

Dans notre travail, la modélisation moléculaire a été employée à fin d'obtenir la conformation énergétique la plus stable –celle dont l'énergie est plus petite possible.

Une fois la convergence atteinte, on enregistre le structure optimisée du polymère sous forme d'un fichier d'extension (*HIN) à fin de la rendre compatible au calcul des descripteurs par le logiciel Dragon.

IV.1.6 Calcul des descripteurs par DRAGON

Une fois que nos structures ont été optimisées, on procède au calcul des descripteurs par le logiciel de calcul des descripteurs moléculaires le plus employé dans la majorité des recherches de type modélisation structurale qui est DRAGON version 5.5- Année 2007(Figure IV .5).



Figure IV .5 :L'interface du logiciel DRAGON version 5.5.

Les blocs des descripteurs moléculaires choisis et la méthodologie de calcul des différents descripteurs par le logiciel DRAGON sont résumés dans les figures IV.6, 10.

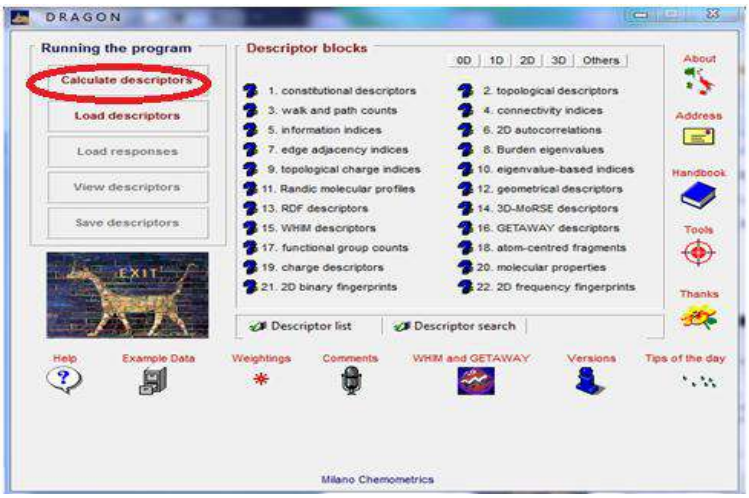


Figure IV.6: Première étapes pour le calcul des descripteurs par DRAGON

Pour calcule les descripteurs, on ouvre les fichiers *.HIN, car il ne lie pas les fichiers d'extension*.Mole et on choisit tous les22 blocs de descripteurs disponibles dont les deux derniers font la nouveauté de cette version pour toutes les structures (0D, 1D ,2D ,3D) puis une fois le calcules terminé enregistrer les fichiers obtenus sous forme de fichiers texte (.*txt)

.

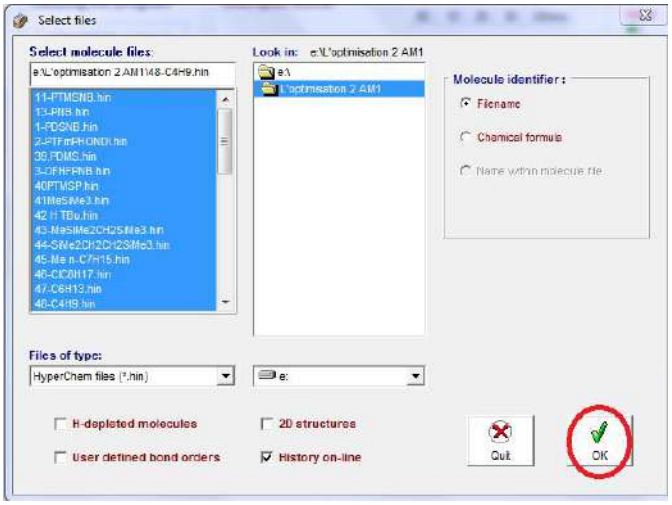


Figure IV.7: Deuxième étapes pour le calcul des descripteurs par DRAGON.

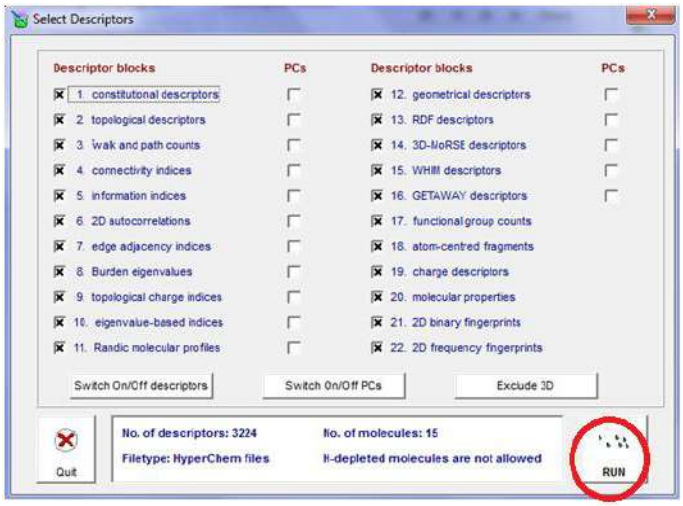


Figure IV.8 : Troisième étape pour le calcul des descripteurs par DRAGON.

Nous avons calculé tous les descripteurs (tableau 1) parce que dans notre étude nous avons besoin du maximum d'information structural pour pouvoir prédire la relation entre la structure des polymères et leur perméabilité aux gaz O_2, CO_2(QSPR).

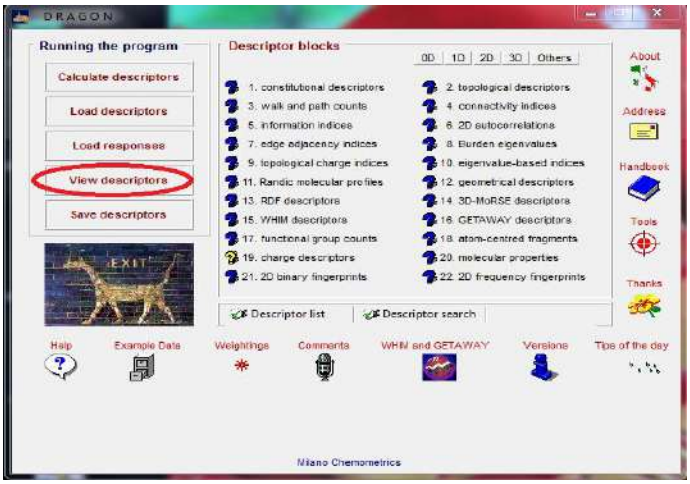
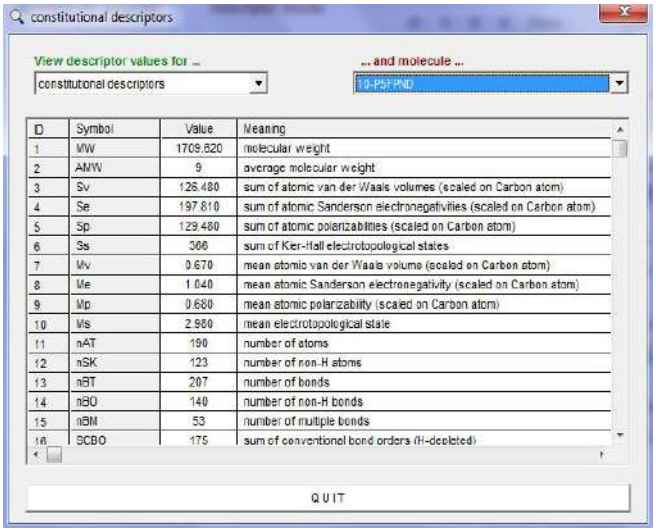


Figure IV.9 : Quatrième étape pour le calcul des descripteurs par DRAGON



ID	Symbol	Value	Meaning
1	MW	1709.820	molecular weight
2	AMW	9	average molecular weight
3	Sv	126.480	sum of atomic van der Waals volumes (scaled on Carbon atom)
4	Se	197.810	sum of atomic Sanderson electronegativities (scaled on Carbon atom)
5	Sp	129.480	sum of atomic polarizabilities (scaled on Carbon atom)
6	Ss	366	sum of Kier-Hall electrotopological states
7	Mv	0.670	mean atomic van der Waals volume (scaled on Carbon atom)
8	Me	1.040	mean atomic Sanderson electronegativity (scaled on Carbon atom)
9	Mp	0.680	mean atomic polarizability (scaled on Carbon atom)
10	Ms	2.960	mean electrotopological state
11	nAT	190	number of atoms
12	nSK	123	number of non-H atoms
13	nBT	207	number of bonds
14	nBO	140	number of non-H bonds
15	nBM	53	number of multiple bonds
16	SCBO	175	sum of conventional bond orders (H-depleted)

Figure IV.10 : Cinquième étape pour le calcul des descripteurs par DRAGON.

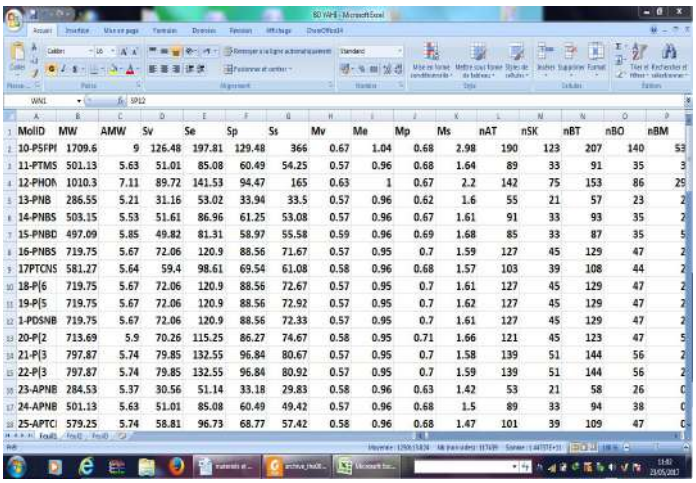
Tableau IV.5 : Nombre de descripteurs calculés par DRAGON

Nom de descripteurs	Nombre de descripteurs
Descripteurs constitutionnels	48
Descripteurs topologique	119
Walk et Path counts descripteurs	47
Indices de connectivity	33
Indices d'information	47
Auto-corrélations 2D	96
Indices edge Adjacency	107
Burden Eigen valeurs	64

Indices de charge topologique	21
Indices de valeurs propres	44
Profils random moléculaires	41
Descripteur géométrique	74
Descripteurs RDF	150
3D MoRSE descripteurs	160
WHIM Descripteurs	99
GetawayDescripteurs	197
Nombre de groupes fonctionnels	154
Fragments -atome centré	120
Descripteurs de charge	14
Propriétés moléculaires	29
Empreintes binaires 2D	780
Empreintes de fréquence 2D	780
Total	3224

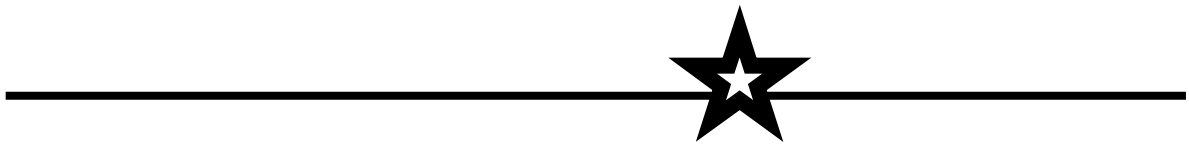
IV.1.7. Collection des données dans un fichier EXCEL

Nos descripteurs calculés par DRAGON sauvegardés dans un fichier de format texte contenant les noms de chaque descripteur, la valeur attribuée le nom du polymère et le nombre des descripteurs calculés (dans notre étude 1586 descripteurs). On importe le contenu des fichiers texte dans un fichier EXCEL la 1ère ligne contient les noms des descripteurs et la1ère colonne les noms des polymères (Figure IV.11).



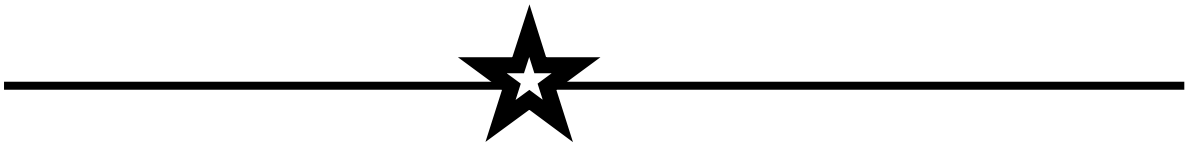
MolID	MW	AMW	Sv	Se	Sp	Ss	Mv	Me	Mp	Ms	nAT	nSK	nBT	nBO	nBM
10-PSFPI	1709.6	9	126.48	197.81	129.48	366	0.67	1.04	0.68	2.98	190	123	207	140	53
11-PTMS	501.13	5.63	51.01	85.08	60.49	54.25	0.57	0.96	0.68	1.64	89	33	91	35	3
12-PHOM	1010.3	7.11	89.72	141.53	94.47	165	0.63	1	0.67	2.2	142	75	153	86	29
13-PNB	286.55	5.21	31.16	53.02	33.94	33.5	0.57	0.96	0.62	1.6	55	21	57	23	2
14-PNBS	503.15	5.53	51.61	86.96	61.25	53.08	0.57	0.96	0.67	1.61	91	33	93	35	2
15-PNBD	497.09	5.85	49.82	81.31	58.97	55.58	0.59	0.96	0.69	1.68	85	33	87	35	5
16-PNBS	719.75	5.67	72.06	120.9	88.56	71.67	0.57	0.95	0.7	1.59	127	45	129	47	2
17PTCNS	581.27	5.64	59.4	98.61	69.54	61.08	0.58	0.96	0.68	1.57	103	39	108	44	2
18-P(6	719.75	5.67	72.06	120.9	88.56	72.67	0.57	0.95	0.7	1.61	127	45	129	47	2
19-P(5	719.75	5.67	72.06	120.9	88.56	72.92	0.57	0.95	0.7	1.62	127	45	129	47	2
1-PDSNB	719.75	5.67	72.06	120.9	88.56	72.33	0.57	0.95	0.7	1.61	127	45	129	47	2
20-P(2	713.69	5.9	70.26	115.25	86.27	74.67	0.58	0.95	0.71	1.66	121	45	123	47	5
21-P(3	797.87	5.74	79.85	132.55	96.84	80.67	0.57	0.95	0.7	1.58	139	51	144	56	2
22-P(3	797.87	5.74	79.85	132.55	96.84	80.92	0.57	0.95	0.7	1.59	139	51	144	56	2
23-APNB	284.53	5.37	30.56	51.14	33.18	29.83	0.58	0.96	0.63	1.42	53	21	58	26	0
24-APNB	501.13	5.63	51.01	85.08	60.49	49.42	0.57	0.96	0.68	1.5	89	33	94	38	0
25-APTCI	579.25	5.74	58.81	96.73	68.77	57.42	0.58	0.96	0.68	1.47	101	39	109	47	0

Figure IV.11 : La base de données des descripteurs finaux.



Partie-2-

Réduction du nombre des descripteurs



Après optimisation des structures tridimensionnelles, les descripteurs sont déterminés dont certains peuvent être corrélés à la propriété qui nous intéresse qui est la perméabilité (P). La base de données obtenue a besoin d'un prétraitement pour réduire sa taille. La sélection des variables pertinentes par rapport à la grandeur à modéliser ; elle vise deux points essentiels :

- Réduction de la dimension de l'espace de la représentation des variables du modèle.
- Élimination des entrées qui n'ont pas ou possèdent peu d'influence sur la sortie.

IV.2.1 Prétraitement de la base de données

Le prétraitement de la base de données a pour finalité d'éliminer les descripteurs qui ne sont pas pertinents afin d'éviter le phénomène du sur-apprentissage.

Notre base de données qu'on appellera « base de données initiale » est un tableau de 73*1586 descripteurs, nous avons procédé à une élimination préliminaire des valeurs insignifiantes comme suit :

-les descripteurs invariants ont été supprimés manuellement.

- tout descripteur ayant plus de 75% de valeurs identiques et supprimé.

D'où une pré-réduction de plus de 15% des descripteurs peu ou non signifiants. Cette base de données appelée « nouvelle base de données » est devenue 73*1351 qui va subir une autre réduction par la méthode d'algorithmes génétique(AG)

Avec les AGs on cherche à optimiser une fonction (objectif) donnée dans un espace de recherche, celui des individus. Pour l'optimiser, on définit une fonction d'évaluation (fitness) reliée à cette fonction objective et appliquée sur chaque individu ou chromosome.

En général, le fonctionnement d'un AG est basé sur les phases suivantes (voir Figure IV.12) :

1. Initialisation : Générer aléatoirement une population initiale de taille N chromosomes.
2. Evaluation : Evaluer chaque individu de la population par la fonction d'évaluation appropriée au problème (fonction de fitness).
3. Reproduction : Créer une nouvelle population de N chromosomes par l'utilisation d'une méthode de sélection appropriée et l'application d'opérateurs génétiques (croisement et mutation) sur certains chromosomes au sein de la population courante.
4. Retour à la phase 2 tant que la condition d'arrêt du problème n'est pas satisfaite (erreur minimale de la validation croisé).

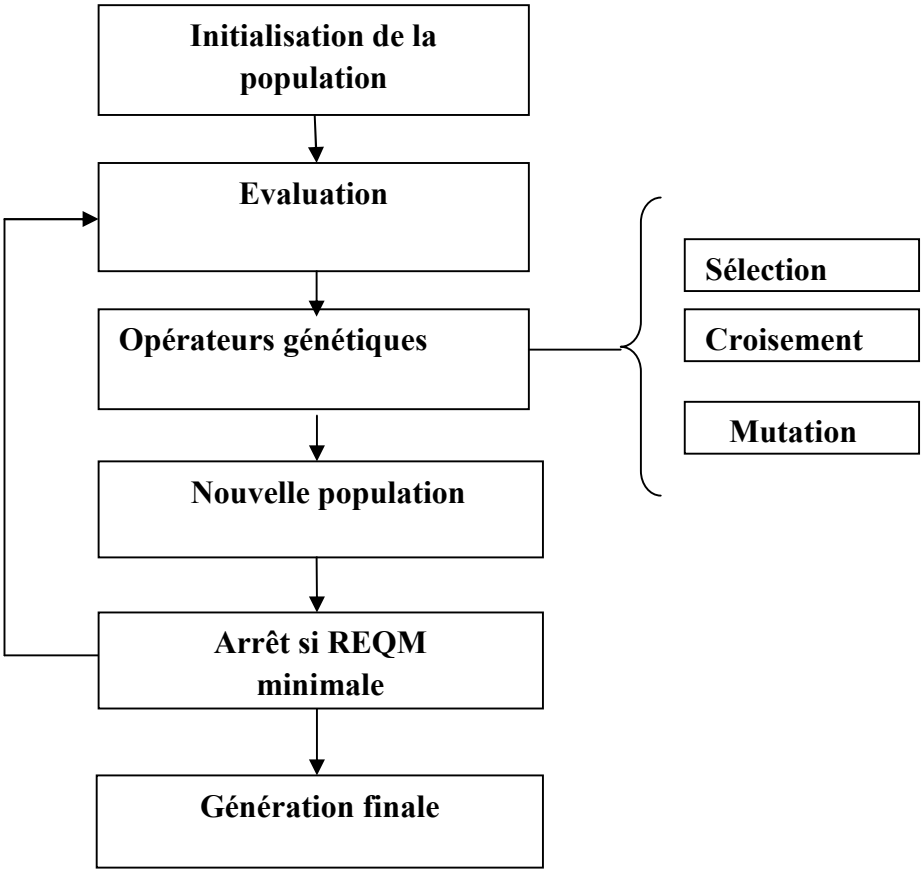


Figure IV12.: Architecture générale d'un algorithme génétique.

IV.2.2Application de la méthode des Algorithmes génétiques

Nous avons entré la base complète de 59*1351 pour le gaz (O_2) et 67*1351pour le gaz(CO_2) dans un algorithme génétique à base de PLS en prenant la valeur du coefficient de validation croisé LOO (leave one out) comme fonction d'objectif (fitness function) avec les caractéristiques suivants :

- Taille de la population: 30 chromosomes
- En moyenne, cinq variables par chromosome dans la population d'origine
- Réponse: variance validée en% a expliqué la variance (cinq groupes de suppression, le nombre des composants sont déterminés par validation croisée)
- Nombre maximal de variables sélectionnées dans le même chromosome: 30
- Probabilité de mutation: 1%
- Nombre maximal de composants: le nombre optimal de composants déterminés par validation crosées sur le modèle contenant toutes les variables (pas plus de 15)

- Nombre de courses: 100
- Élimination vers l'arrière après chaque 100ème évaluation et à la fin (si le nombre d'évaluations n'est pas un multiple de 100,taille de la fenêtre pour le lissage: 3)

Nous avons appliqué la méthode à travers l'interface du logiciel de calcul mathématique Matlab R2012b version 8.0.0.783, 64 bit. La GA-PLS est implémentée dans la bibliothèque du logiciel, les résultats sont représentés dans le tableau IV.6 et tableau IV.7:

Tableau IV.6 : Résultats de calcul des descripteurs moléculaires pour CO_2.

Les blocs des descripteurs	Le nombre des descripteurs dans chaque bloc	Le nombre des descripteurs après la réduction
Descripteurs constitutionnels	48	11
Descripteurs topologiques	119	2
Walk and pathcounts	47	1
Indices de connectivité	33	1
Indices de proximité adjacents	107	1
Indices de charge topologique	21	1
3D-morse descripteurs	160	1
WHIM descripteurs	99	1
Total	634	19

-Descripteurs 0D: AMW/Se/Sp/Ss/Me/Mp/Ms/nsk /nBT /ARR/nH /

-Descripteurs 2D :SRW02/DELS/ICR/GGI3/X2Sol/EEig06d/

-Descripteurs 3D : Mor12u/Ap/

Tableau IV.7 : Résultats de calcul des descripteurs moléculaires pour O_2.

Les blocs des descripteurs	Le nombre des descripteurs dans chaque bloc	Le nombre des descripteurs après la réduction
Descripteurs constitutionnels	48	4
Descripteurs topologiques	119	3
3D MoRSE descripteurs	160	1
GetawayDescripteurs	197	1
Indices de proximité adjacents	107	3
2D autocorrelations	96	1
Descripteurs géométriques	74	1
Indices d'information	47	4
Total	848	18

-Descripteurs 0 D: Sp/nsk /SCBO/nF /

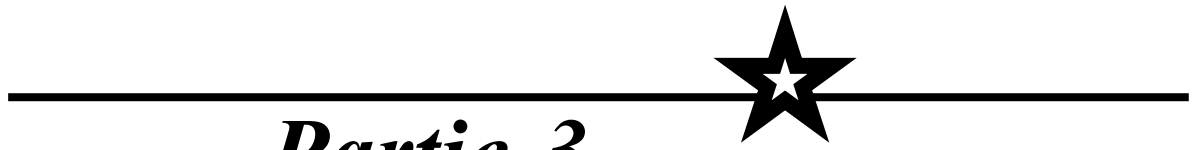
-Descripteurs2D:ZM2/Pol/BAC/GATS1m/HDCPx/TIC1/BIC4/Yindex/EEig01x/ESPm04x/ESPm12x/

-Descripteurs 3D : RTV/G1/Mor18u/

Tableau IV. 8 : Signification des principaux descripteurs moléculaires utilisés.

	Indice	Description de l'indice
Descripteurs 0D	AMW	Poids moléculaire
	Se	Somme atomiques d'électrégativités de Sanderson (mesurés sur l'atome de carbonne)
	Sp	Somme de polarisabilités atomiques
	Ss	Somme d'état électro-topologique de kier-Hall
	Me	Moyen d'électronégativité atomique de Sanderson
	Ms	Etat électro-topologique moyen
	Mp	Moyen de polarisabilités atomique
	nSK	Nombre d'atomes de non-H
	nBT	Numbre de liens
	ARR	Ration aromatique
	nH	Nombre d'atome d'hydrogène
Descripteurs 2D	SRW02	Compteur de commande d'ordre 02
	ICR	Indice d'information centralisé radial
	GGI3	Indices de charge topologique d'ordre 2
	X2Sol	Indices de solvation d'ordre 2
	EEig01x	Valeur propre 01 de la matrice edge adj, pondérée par des degrés de bord
	EEig06d	Valeur propre 06 de la matrice edge adj, pondérée par des degrés de bord
	ZM2	Deuxiéme indices M2 de Zagreb
	Pol	Nombre de polarité
	BAC	Indice central de Balaban
	GATS1m	Geary autocorrelation-retard 1/ pondéré par les polarisations atomiques
	TIC1	Indices de contenu d'information totale d'ordre 1
	BIC4	Bond contenu d'information (symétrie de quartier d'ordres 4)
	Yindex	Indices Y de Balaban
	ESPm04x	Moment spectral 04 à partir de la matrice de bordage pondérée par degrés de bord
	ESPm12x	Moment spectral 12 à partir de la matrice de bordage pondérée par degrés de bord
Descripteurs3D	Mor12u	3D-MoRSE-signale12/non pondéré
	Mor18u	3D-MoRSE-signale18/non pondéré
	AP	Indice de taille Atotal / pondéré par les polarisations atomiques
	RTv	Indices totale/podérés par les volumes atomique de Van Der Walls
	G1	Indices de gravité G1

Si en comparant les résultats entre les deux gaz, on a remarqué qu'ils ne réduisent pas toujours les mêmes descripteurs. Cependant les descripteurs 3D, 0Det 2D sont toujours présents dans les deux cas.



Partie-3-

Modélisation neuronal pour la prédiction de la perméabilité



Les première parties précédente de notre étude nous permet d'obtenir une base de données de dimension .Quand peut les utiliser pour établir une relation de structure propriété (la perméabilité) QSPR, pour réaliser cette relation, on utilise le réseau de neurone artificiel.

Cette partie de notre étude est divisée en trois étapes: la première s'intéresse à l'architecture du réseau et l'optimisation des paramètres qui l'affecte et la deuxième étape présente les résultats de prédiction du modèle obtenu et ses performances par rapport aux modèles existants et la dernière s'intéresse à la validation des modèles obtenus par des paramètres statistique.

IV.3.1 Répartition des données

Dans cette étape, les données sont subdivisées en deux sous-ensembles :

- Un ensemble pour l'apprentissage.
- Un ensemble pour le teste.

Certes il n'y a pas de règles pour la détermination de ce partage de manière quantitative. La base de données expérimentale comprend 37 descripteurs choisis pour les différentes structures polymériques, la structure de la base de données finale est schématisée sur la figure IV.13.(80%pour la phase d'apprentissage et 20% pour la hase de test)

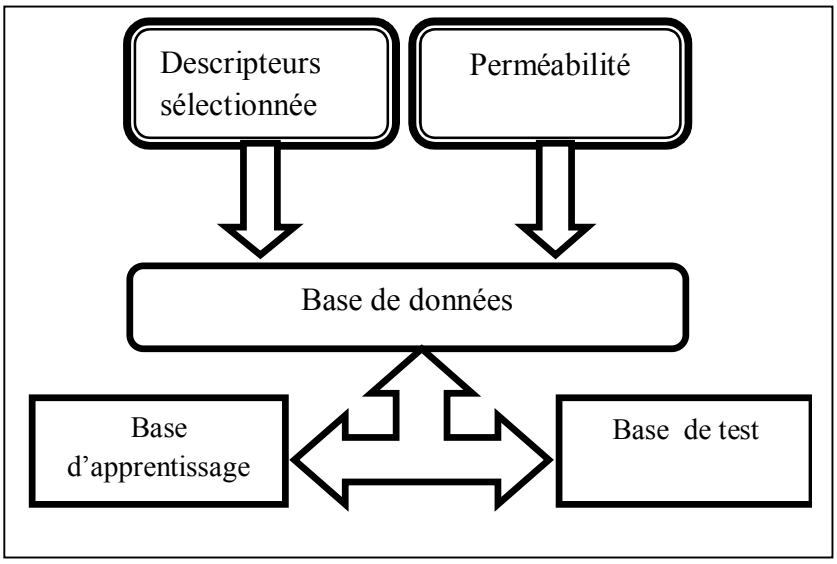


Figure IV.13 : Schématisation générale et répartition du contenu de la base de données.

IV.3.2 Optimisation et conception du réseau de neurones :

Il existe un grand nombre de types de réseaux de neurones, chacun a ses avantages et ses inconvénients, le choix dépend de :

- ✓ La tâche à effectuer (classification, association, contrôle…..)
- ✓ La nature de données.

Différents types de réseaux de neurones disponibles dans le logiciel de simulation (Matlab) que l'on compte utiliser ou le développent de son propre programme.

L'objectif majeur de cette partie est de savoir quelle est la meilleure architecture qui peut être capable de traiter la base de données disponible pour la conception et l'optimisation d'un modèle neuronal.

Dans ce travail, trois étapes ont été suivies: la première consiste à tester les paramètres fondamentaux telle que la fonction d'activation et l'algorithme d'apprentissage, la deuxième partie consiste à tirer les meilleurs réseaux et voir l'effet de quelques fonctions de normalisation sur leur performance et la troisième partie est consacrée à l'étude de l'effet du nombre de neurones et du nombre des couches cachées.

Il existe plusieurs langages ou logiciels où on peut réaliser un modèle de prédiction basé sur ANN comme : Lazy, e1071, Fortran 77 et Matlab.

Dans ce présent travail, ANN est réalisé par Matlab à cause de son utilisation fréquente dans les publications tel que wessling**[106]**, Mazdak**[107]** et notre maitrise de ce logiciel dans notre cursus pédagogique.

Dans Matlab, les étapes de création d'un modèle neuronal se traduisent par les instructions mentionnées dans le tableau IV.9.

Tableau IV.9 : Les significations de quelque instruction de Matlab.

Instruction	Signification
Newff	Créer un nouveau réseau MLP
Init	Initialise le réseau MLP
Train	La phase d'apprentissage
Sim	Simuler le réseau de neurones

La propriété émergente d'un ANN est sa capacité d'apprentissage, à partir d'un état quelconque. Il va en s'appuyant sur un algorithme d'apprentissage, adapter la valeur des poids synaptiques pour fournir une réponse vraie semblable vis-à-vis du problème traité .Il existe différentes techniques d'apprentissage pour déterminer les valeurs des poids synaptiques optimales .

L'apprentissage supervisé est retenu dans cette étude. Il consiste à fournir au réseau des jeux de données d'entrées (les descripteurs chimiques de structure polymérique) pour lesquels on connait la réponse souhaitée (le coefficient de perméabilité).L'architecture d'un réseau est optimisée pour minimiser la racine de l'erreur quadratique moyenne.

Le réseau de type" feedforward " est le réseau le plus utilisé pour l'approximation des fonctions des données statiques.

Les codes des différentes fonctions d'apprentissage utilisés dans les tests sont illustrés dans le tableau suivant :

Tableau VI.10 : Les codes des différentes fonctions d'apprentissage utilisés dans le test

L'algorithme	Code Matlab	Acronyme
Levenberg-Marguardt	Trainlm	LM
Scaledconjugate Gradient	Trainscg	SCG
Régularisation bayesienne	Trainbr	BR
Resilient Back propagation	Train rp	RP
BFGC quassi-Newton	Trainbfg	BFG

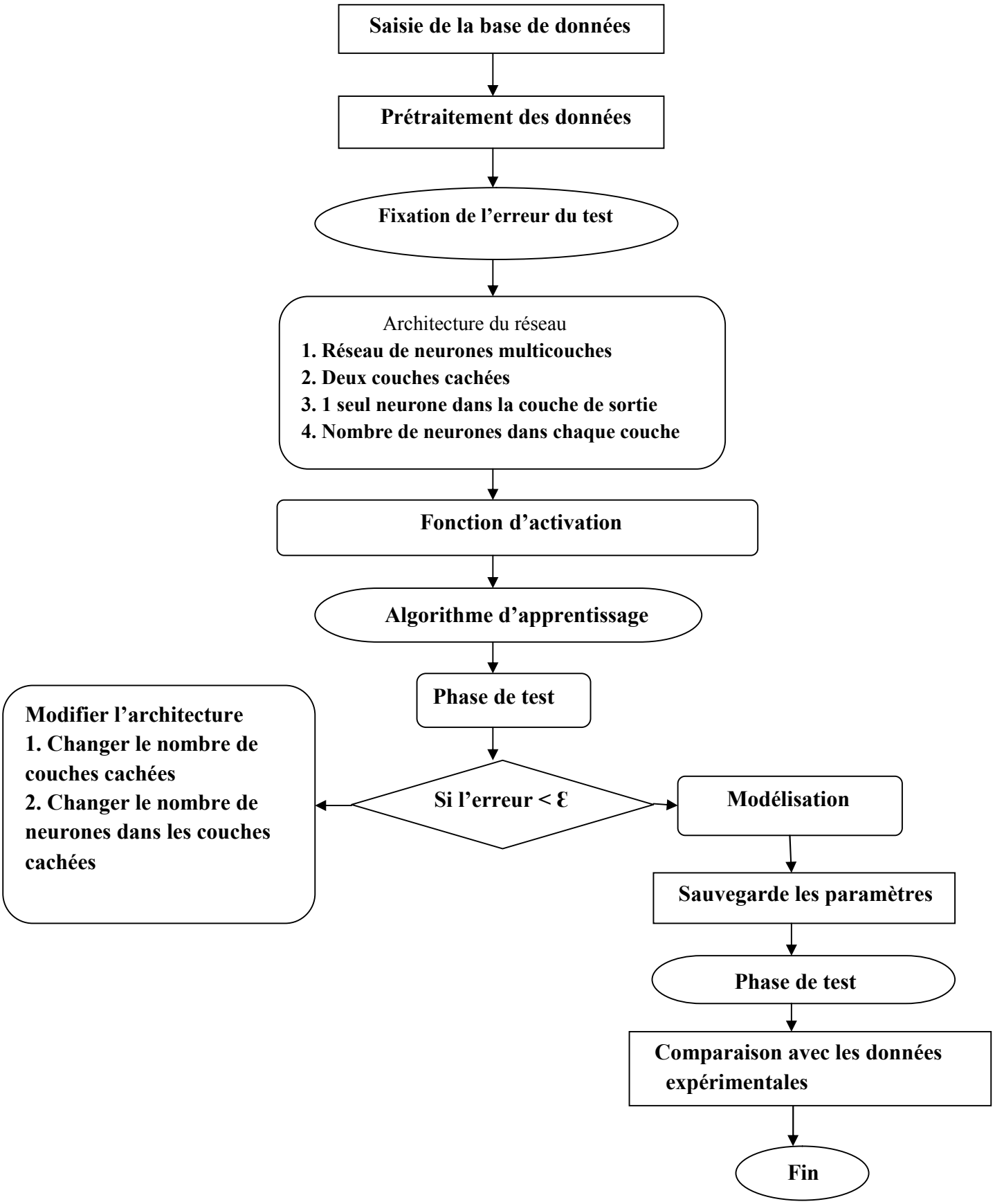


Figure IV.14: Organigramme de la méthodologie de calcul réseau de neurones statique optimale(RNSO).

IV.3.3 Optimisation de l'architecture de réseau de prédiction d'O2

IV.3.3.1 Étape 1 :

A. Effet de l'algorithme d'apprentissage

Pour suivre l'effet de la variation de l'algorithme d'apprentissage, on a utilisé plusieurs algorithmes cités dans le tableau VI.9.

Le but est d'avoir la racine de l'erreur quadratique moyenne (REQM) minimale quel que soit l'algorithme.

$$\text{REQM} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{\text{prédite}} - y_{\text{observée}})^2} \qquad \text{(IV.2)}$$

Avec $y_{\text{prédite}}$ et $y_{\text{observée}}$ les valeurs respectivement prédite et expérimentale.

Après la phase d'apprentissage, on a abouti l'algorithme, train SCG et les résultats d'apprentissage sont représentés sur la figure IV.15.

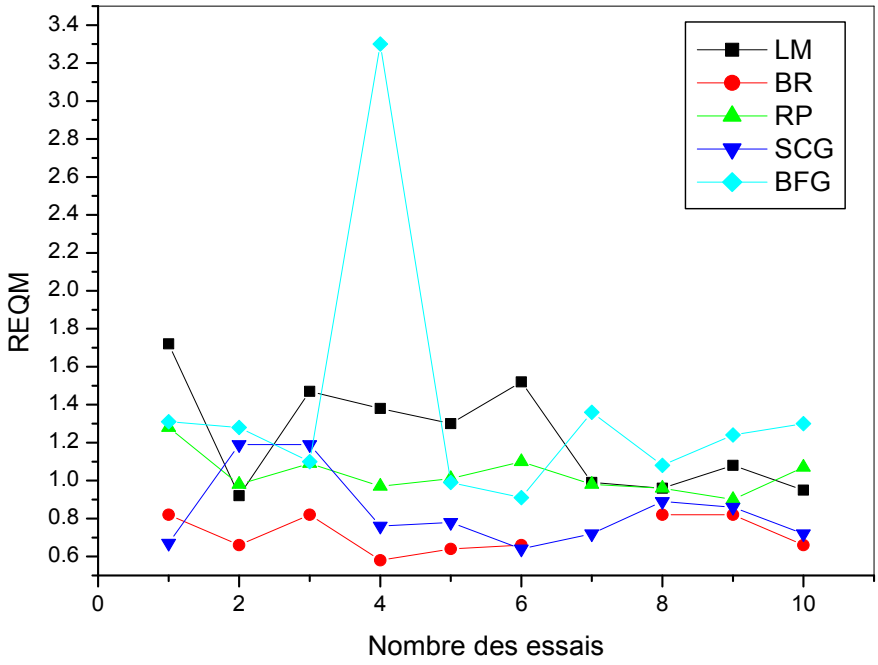


Figure IV.15 : Effet de l'algorithme d'apprentissage sur la performance du réseau de neurones.

IV.3.3.2 Etape2 :

A. Effet de normalisation

Les données numériques ont été normalisées afin d'améliorer l'optimisation. La normalisation est faite par l'une des fonctions suivantes : mapminmx, mapstd**[108]**.Ces fonctions suivent les formules illustrées dans le tableau IV.11.

Tableau IV.11 : les algorithmes de normalisation.

	L'algorithme
Mapminmx	$y = (y_{max} - y_{min}) * \frac{(x - x_{min})}{(x_{max} - x_{min})} + y_{min}$
Mapstd	$y = (x - x_{moy}) * \left(\frac{y_{std}}{x_{std}}\right) + y_{moy}$

Ou y : est la valeur normalisé.

X : est la valeur originale.

A partir de l'étape 1, on a abouti l'algorithmes train SCG. Ce résultat permet de tester l'effet de normalisation sur SCG.

L'effet de normalisation sur la convergence du modèle neuronale SCG est représenté sur les figures IV.14.

Tableau IV.12 : Notation des réseaux normalisés.

Méthodes de normalisation	Code de matlab
Sans normalisation	-
Méthode MMM	Mapminmax
Méthode de MSDM	Mapstd

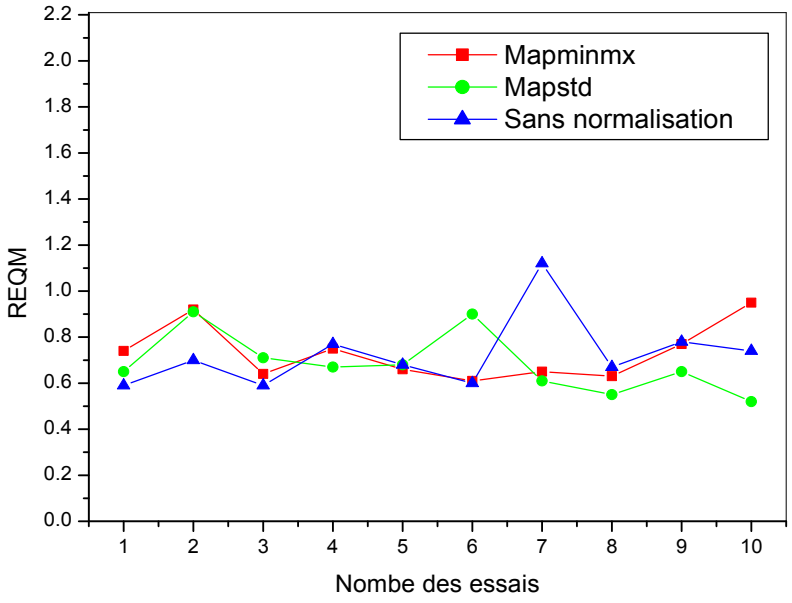


Figure IV.16 : Effet de la normalisation sur la performance d'apprentissage du réseau de neurones train SCG.

En termes d'REQM La figure IV.16 présente la méthode MSDM de normalisation qui donne la meilleure performance.

IV.3.3.3 Étape 3 :

A. Effet de fonction d'activation

Quatre fonctions d'activation ont été appliquées (tableau IV.13) pour avoir une seule fonction d'activation qui donne une erreur quadratique minimale, la figure IV.17 présente la variation de la racine de l'erreur quadratique moyenne avec différentes fonctions d'activation.

Tableau IV.13 : Différentes fonctions d'activations utilisées dans la phase de conception du modèle neuronal.

Fonction de transfert	Code MATLAB
Tan- Sigmoïde transferfunction	Tansig
Log- Sigmoïde transferfunction	Logsig
Triangular bais transferfunction	Tribas
Radial basis function	Radbas

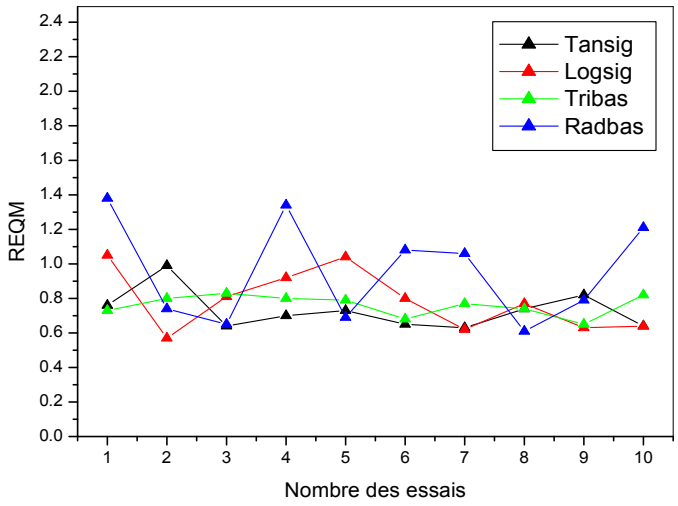


Figure IV.17: Effet de la fonction d'activation sur les performances du réseau de neurones.

La fonction d'activation qui a été choisie pour les couches cachées est log Sigmoïde et la fonction identité purelin pour la couche de sortie.

IV.3.3.4 Etape4 :

A. Effet de nombre de neurones :

Cette étape d'optimisation se base sur le test de différentes architectures du réseau de neurones par la modification du nombre de neurones dans chaque couche cachée.

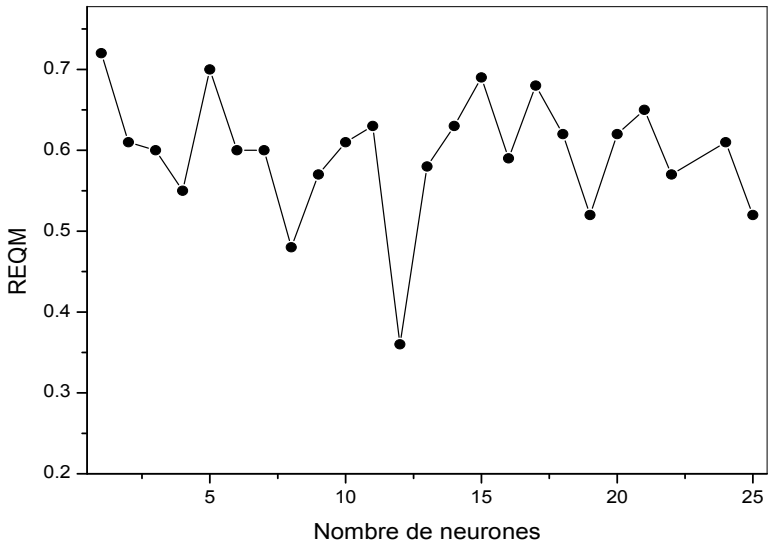


Figure IV.18 : Effet de nombre de neurones de la couche cachée sur la performance

d'apprentissage du réseau de neurones.

D'après la figure IV.18 on a choisi 12 neurones dans les couches cachées où le réseau donne une meilleure performance.

A. Effet de nombre des couches cachées :

L'effet du nombre de couches cachées sur la performance d'apprentissage est illustré dans la figure IV.19. On remarque qu'il n y a pas une différence remarquable au niveau de la performance par rapport au réseau a une seule couche cachée.

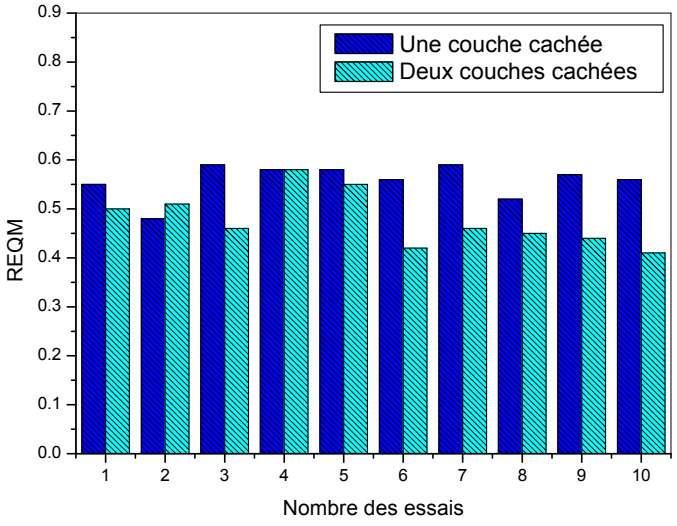


Figure IV.19 : Effet de nombre de couches cachées sur la performance d'apprentissage.

IV.3.4 Optimisation de l'architecture de réseau de prédiction de CO_2 :

IV.3.4.1 Étape 1 :

A. Effet de l'algorithme d'apprentissage

Pour suivre l'effet de la variation de l'algorithme d'apprentissage, on utilise plusieurs algorithmes qui sont cités dans le tableau IV.10.

Après la phase d'apprentissage, on a abouti l'algorithme, train BR et train RP, les résultats d'apprentissage sont représentés sur la figure IV.20.

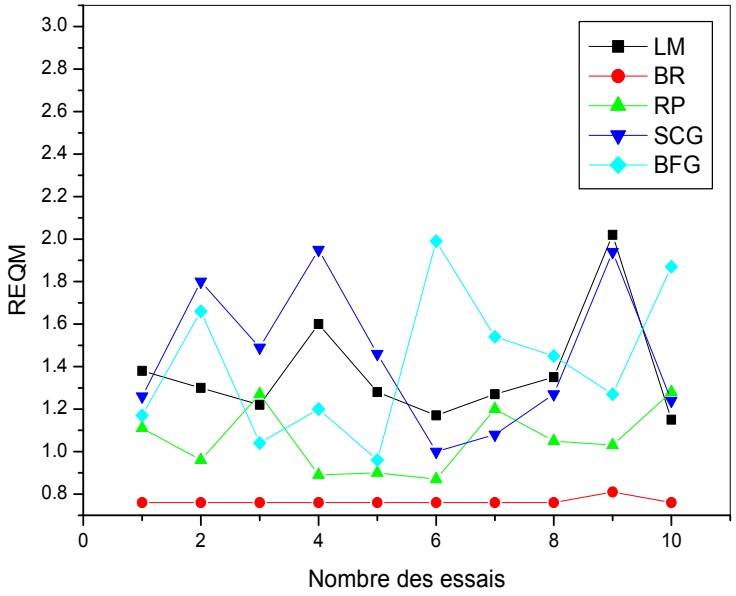


Figure IV.20: Effet de l'algorithme d'apprentissage sur la performance du réseau de neurones

IV.3.4.2 Étape 2 :

A. Effet de la normalisation

Les données numériques ont été normalisées afin d'améliorer l'optimisation ;la normalisation se fait par l'une des fonctions suivantes : mapminmx, mapstd.

A partir de l'étape 1, on a abouti à deux algorithmes train BR, train RP. Ces résultats permettent de tester l'effet de normalisation sur les deux réseaux obtenus. L'algorithme qui donne le meilleur résultat est train RP sans normalisation.

L'effet de normalisation sur la convergence du modèle neuronal est représenté sur la figure IV.21et la figure IV.22.

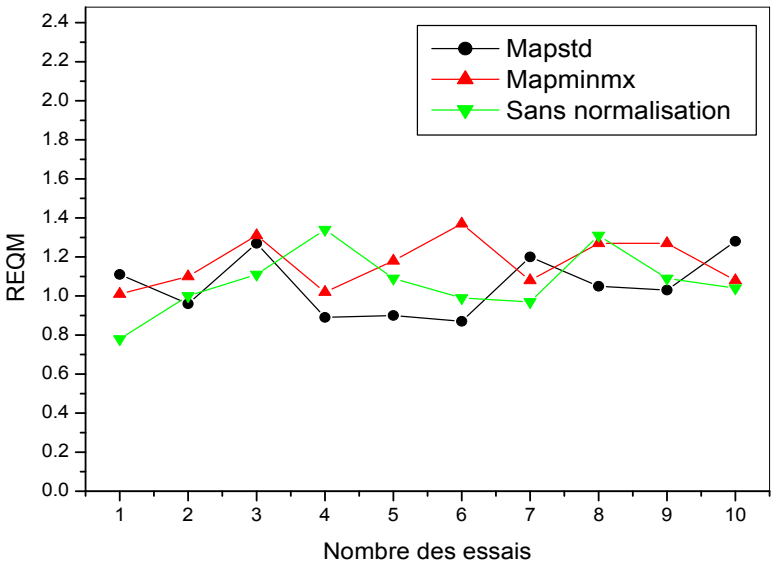


Figure IV.21: Effet de la normalisation sur la performance d'apprentissage du réseau de neurones, train RP.

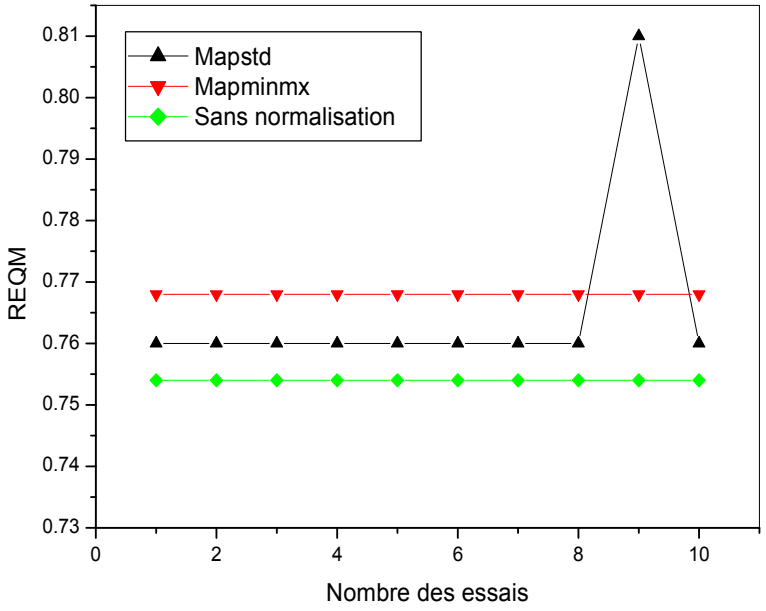


Figure IV.22: Effet de la normalisation sur la performance d'apprentissage du réseau de neurones, train BR.

IV.3.4.3 Étape 3 :

A. Effet de fonction d'activation

Quatre fonctions d'activation ont été appliquées (tableau IV.12) pour avoir une seule fonction d'activation qui donne la racine de erreur quadratique minimale, la figure IV.23 présente la variation de la racine de l'erreur quadratique moyenne (REQM) avec différentes fonctions d'activation.

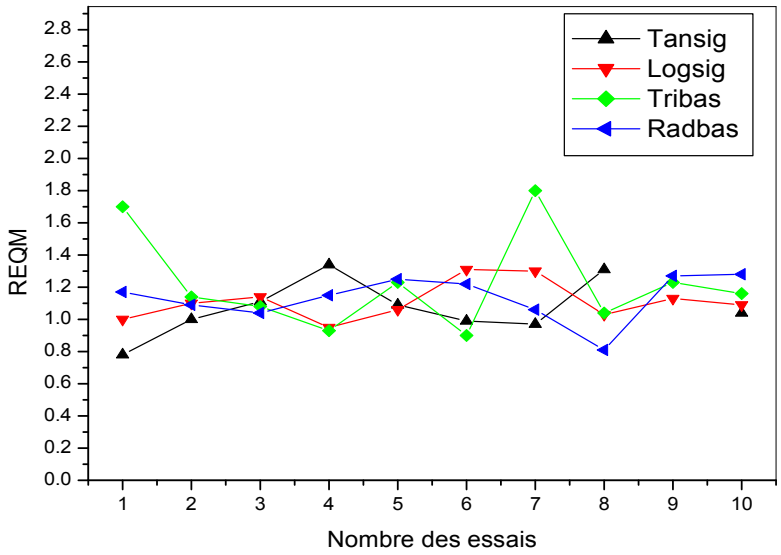


Figure IV.23: Effet de la fonction d'activation sur les performances du réseau de neurones.

La fonction d'activation qui a été choisie pour les couches cachées est Tan- Sigmoïde et la fonction identité purelin pour la couche de sortie.

IV.3.4.4 Étape 4 :

Cette étape a été consacrée au test des différentes architectures en modifiant le nombre de neurones dans chaque couche cachée et le nombre des couches cachées.

A. Effet du nombre de neurones

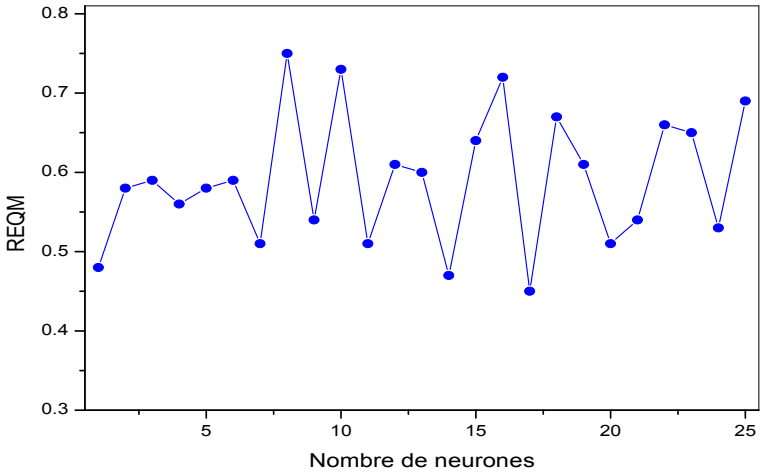


Figure IV.24 : Effet de nombre de neurones de la couche cachée sur la performanc

d'apprentissage du réseau de neurones.

D'après la figure IV.24, le réseau donne une meilleure performance pour 16 neurones dans la couche cachée.

A. Effet du nombre des couches cachées :

L'effet du nombre de couches cachées sur la performance d'apprentissage est illustré dans la figure IV.25. Celle-ci montre que la meilleure performance est celle de deux couches cachées.

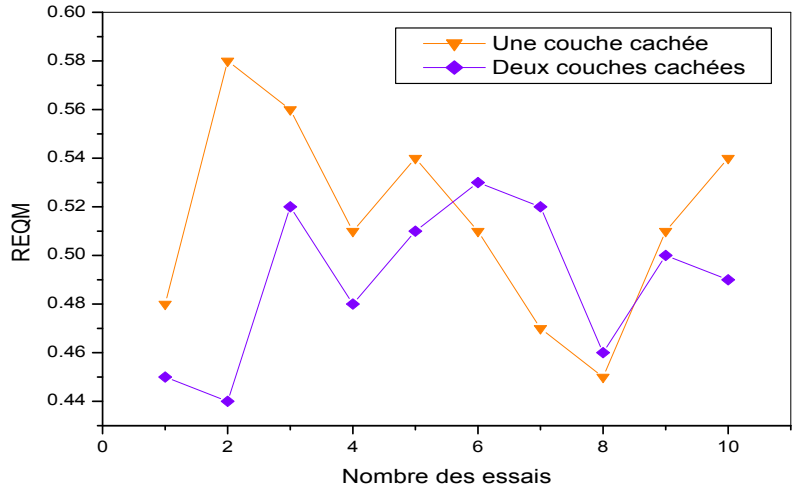


Figure IV.25 : Effet du nombre des couches cachés sur la performance d'apprentissage.

IV.3.5 Résultat de l'optimisation

La constitution des réseaux obtenus est présentée dans le tableau IV.13.Une fois que l'apprentissage est réalisé, les réseaux sont testés sur de nouveaux jeux d'entrée/sortie qui forment la base de généralisation .Cette étape permet de vérifier le comportement des réseaux sur des cas non appris.

Tableau IV.14 : Structure des réseaux de neurones optimisés.

	Type de réseau	Couche d'entrée	Première couche cachée		Deuxième couche cachée		Couche de sortie		Algorithme d'appren-tissage
		Nombre de neurones	Nombre de neurones	Fonction d'activation	Nombre de neurones	Fonction d'activation	Nombre de neurones	Fonction d'activation	
O_2	RNS MPL	18	12	Tansig	-		1	Purlin	Train SCG
CO_2		19	16	Tansig	6	Logsig	1	Purlin	Train RP

Ces réseaux sont représentés schématiquement par la figure IV.26 pour le réseau d'O_2et figure IV .27 pour le réseau CO_2.

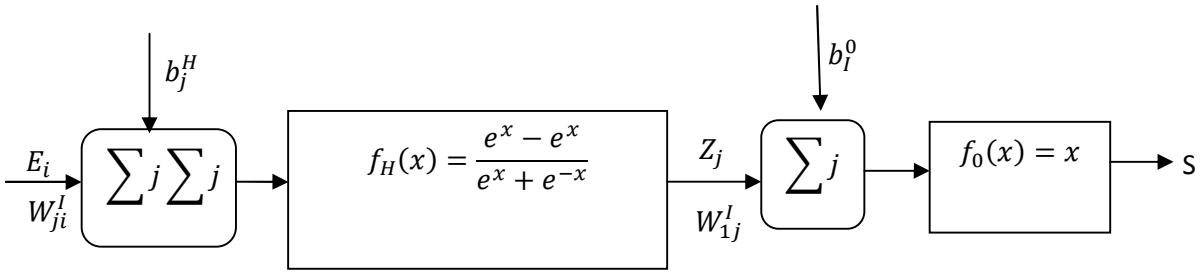


Figure IV.26 : Architecture du réseau de neurones optimisé d'O2.

E_i représente les entrées (i allant de 1 a 18)

W_{ij}^I Représentent les poids qui relient les entrées aux « j » neurones de la couche cachée (j allant de 1a 12).

b_j^H Sont les bais des poids reliant les entrées aux « j » neurones de la couche cachée.

Z_j Sont les sorties des « j » neurones de la couche cachée.

W_{1j}^HRepresentent les poids des « j » neurones de la couche allant vers la sortie.

b_I^0est le bais des poids de la couche cachée allant vers la sortie.

S est la sortie désirée (la perméabilité dans le cas de notre étude).

Les expressions mathématiques des sorties Z_j de la couche cachée et S de la couche de sortie sont données ci-après :

$$Z_j = f\sum_{i=1}^{18} w_{ij}^I E_i + b_j^H] = \frac{\exp(\sum_{i=1}^{18} w_{ij}^I E_i + b_j^H) - \exp(-\sum_{i=1}^{18} w_{ij}^I E_i + b_j^H)}{\exp(\sum_{i=1}^{18} w_{ij}^I E_i + b_j^H) + \exp(-\sum_{i=1}^{18} w_{ij}^I E_i + b_j^H)} \quad \text{(IV.3)}$$

$$S = f_0 \sum_{j=1}^{12} w_{1j}^H E_i + b_1^0 = \sum_{j=1}^{12} w_{1j}^H Z_j + b_1^0 \quad \text{(IV.4)}$$

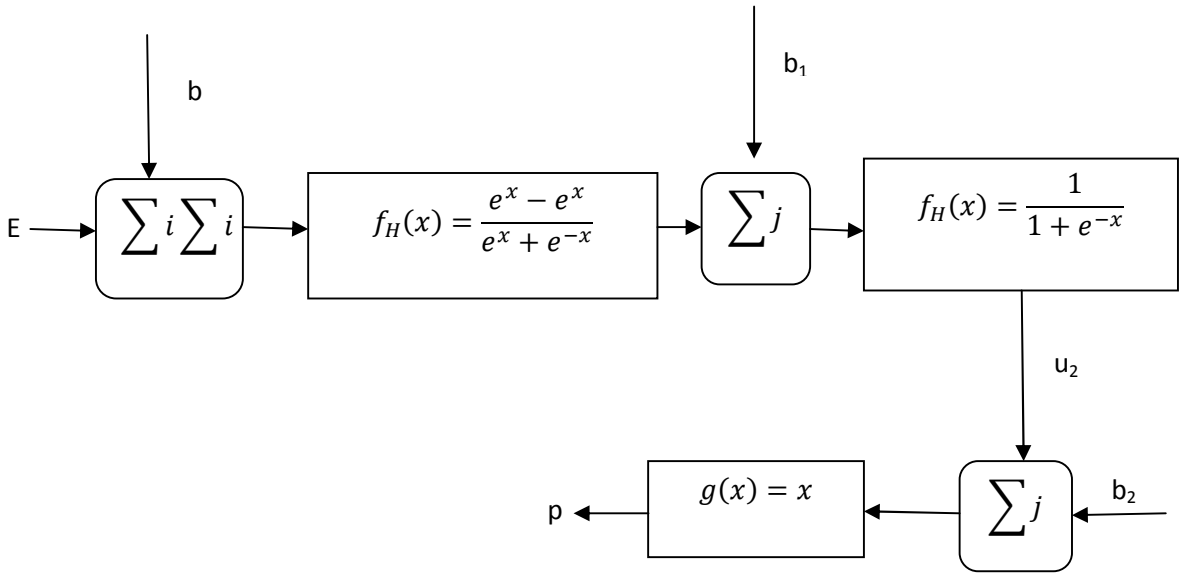


Figure IV.27 : Architecture de réseau de neurones optimisé de CO_2.

$$P_{CO_2} = g\{(u_2 f_2(\sum_{j=1}^{19} u_{j1}\, f[\sum_{i=1}^{16} w_{ij}\, x_N + b_j]) + b_1) + b_2\} \qquad \text{(IV .5)}$$

La présentation graphique entre la sortie désirée et la réponse des réseaux est réalisée par la fonction (Postreg, Matleb), Les résultats obtenus sont représentés sur la figure IV.28 et la figure IV.29.

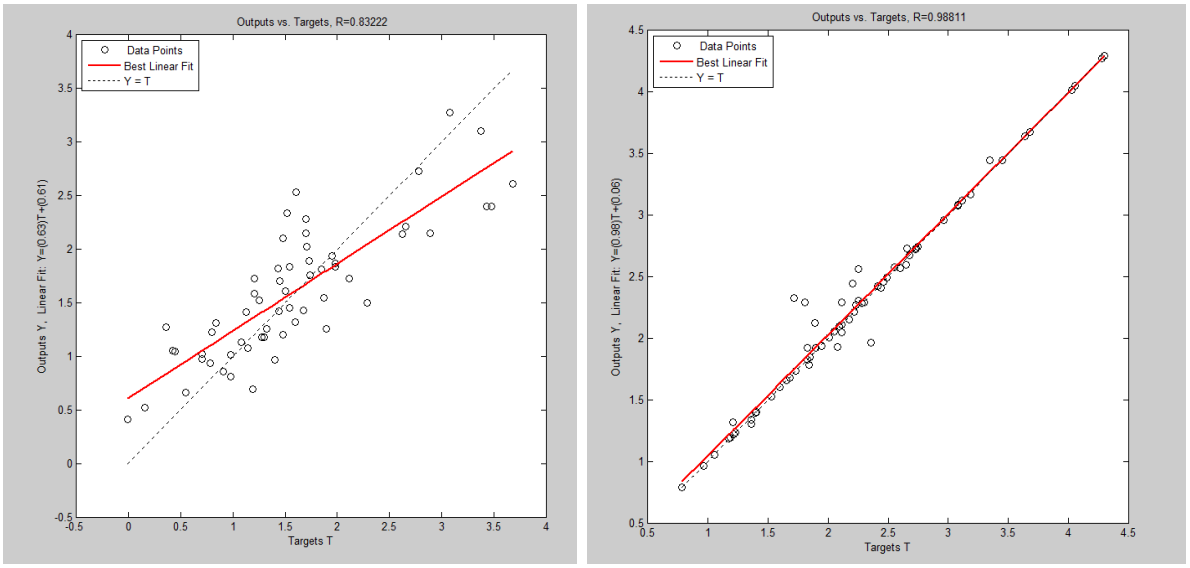


Figure IV.28 : la perméabilité estimée en fonction de la perméabilité expérimentale pour l'O_2 et CO_2et leur équation linéaire par la fonction postreg de MATLAB.

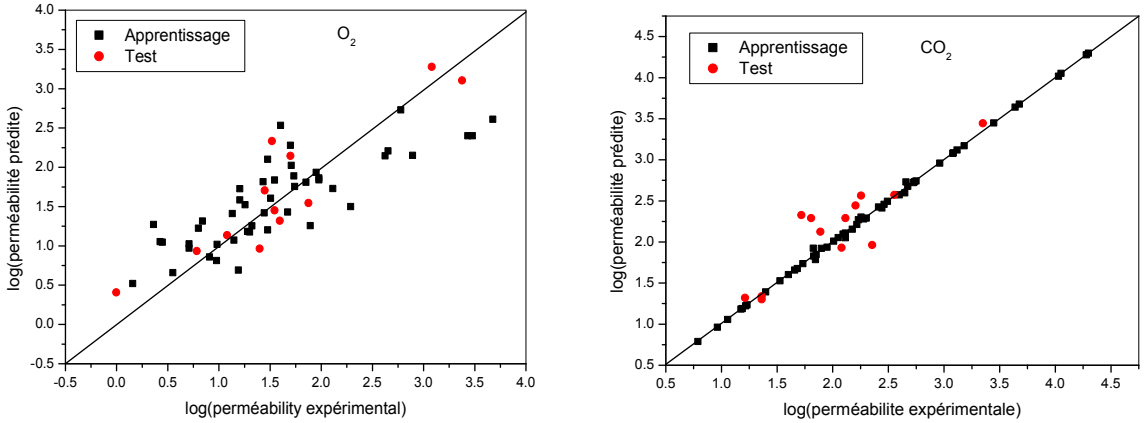


Figure IV.29 : Comparaison entre la perméabilité expérimentale et la perméabilité estimée pour l'O_2 et CO_2.

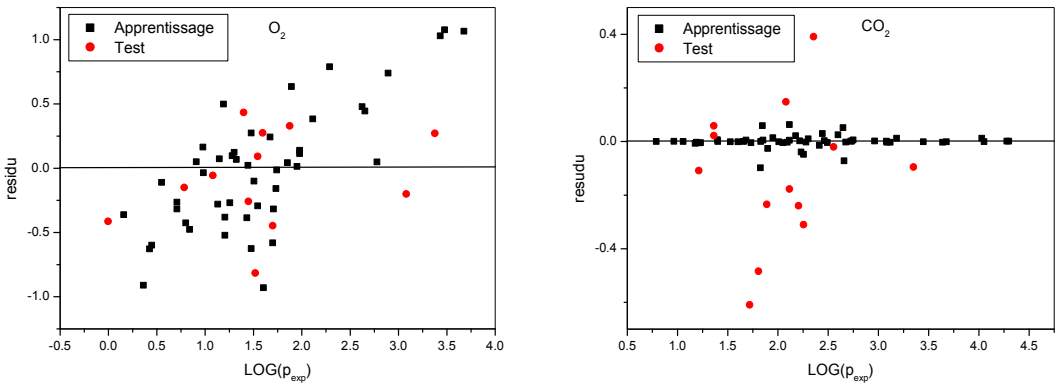


Figure IV.30 : Résidus des valeurs de log (p) obtenus en fonction de log(p expérimentale) pourO2 et CO2.

IV.3.6 Paramètres statistiques

Les paramètres statistiques sont requis pour évaluer la performance du modèle élaboré**.**

IV.3.6.1 Coefficient de corrélation R

$$R = 1 - \frac{\sum_1^n (y_i - y_i^t)}{\sum_1^n (y_i - y_0)^2} \qquad \text{(IV .6)}$$

Avec $$y_0 = \left[\frac{1}{n}\left(\sum y_i - y_i^t\right)^2\right]^{1/2} \qquad \text{(IV.7)}$$

Ou$y_i$est la valeur expérimentale, y_i^t la valeur prédite.

Le R est un indicateur simple, on comprend aisément que plus il s'approche de la valeur 1, plus le modèle est intéressant.

IV.3.6.2 Le coefficient de prédict*ion* (Q^2_{train})

$$Q^2_{train} = 1 - \frac{\sum(y_{exp(train)} - y_{pred(train)})^2}{\sum(y_{exp(train)} - \bar{y}_{exp(train)})} \quad \text{(IV.8)}$$

Ce paramètre nous renseigne sur la robustesse du modèle élaboré (en générale, il est préférable qu'il soit supérieure de 0,6.

IV.3.6.3 Le coefficient de prédiction (Q^2_{test})

$$Q^2_{test} = 1 - \frac{\sum(y_{exp(test)} - y_{pred(test)})^2}{\sum(y_{exp(test)} - \bar{y}_{exp(train)})^2} \quad \text{(IV.9)}$$

Ce paramètre nous renseigne sur le pouvoir prédictif du modèle une valeur supérieure de 0,8 stipule que le modèle est très performant

IV.3.6.4 Facteur de performance (C)

$$C = \frac{\sum_1^n |(y_{pred} - y_{exp})|}{\sum_1^n |(y_{exp} - \bar{y}_{exp})|} \quad \text{(IV.10)}$$

Le réseau neuronal fonctionne de manière prédictive si C est inférieur à 1.

IV.3.6.5 Erreur relative absolue moyenne (ERAM)

$$ERAM = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_{exp} - y_{cal}}{y_{exp}}\right) \quad \text{(IV.11)}$$

Ou y_{exp} est la valeur expérimentale, y_{cal} la valeur calculée.

IV.3.6.6 **Test de Fisher** : Significativité globale du modèle

La formulation du test d'hypothèse qui permet d'évaluer globalement le modèle est la suivante :

H0 : b1 = b2 = … = bk = 0 ;

H1 : un des coefficients au moins est non nul.

La statistique dédiée à ce test s'appuie (parmi les différentes formulations possibles) sur le R^2, il s'écrit :

$$F_c = \frac{R^2/K}{(1-R^2)/(n-K-1)} \quad \text{(IV.12)}$$

Avec

N le nombre des polymères

K Le nombre de variables explicatives

R^2 Coefficient de détermination (corrélation)

Les valeurs de ces paramètres sont représentées dans le tableau IV.14.

Tableau IV.15: Les valeurs des paramètres statistique

Paramètres statistiques	CO_2	O_2
n	67	59
$REQM_T$	0.283	0.367
$REQM_G$	0.126	0.459
$ERAM_T$	0.1153	8.06
$ERAM_g$	0.027	1.92
R_T	0.90	0.910
R_G	0.98	0.83
Q^2_{test}	0.8	0.82
Q^2_{train}	0.64	0.99
C	0.085	0.59
F_c	47.00	10.84
P_{min}	9.2	0.99
P_{max}	19900	3000
K	19	18
n/K	3.52	3.27

IV.3.7 Analyse des résultats

Pour évaluer les performances des modèles neuronaux obtenus, la représentation graphique des valeurs prédites de la perméabilité en fonction des valeurs expérimentales pour les ensembles d'apprentissage et de test obtenus est donnée sur les figures IV.27 et IV.28.

Une corrélation étroite entre les valeurs calculées par le modèle RNA et les valeurs expérimentales de la perméabilité a été observée.

La véritable utilité des modèles QSPR n'est pas seulement leur capacité à reproduire les données connues, mais aussi avoir un potentiel prédictif. Pour cette raison, on a établi les paramètres caractéristiques du modèle qui figurent sur le tableau IV.15.

La valeur de 0.99 et 0.64 pour Q^2_{train}suggère une cohérence de l'ensemble d'apprentissage et une bonne validation interne. Par conséquent la robustesse du modèle a été confirmée. D'autre part, la grande valeur deQ^2_{test}(0.8 et 0.82) pour l'O2 et CO2 respectivement préconise un bon pouvoir prédictif du modèle.

Les résidus pour les valeurs observées dans les ensembles d'apprentissage et de validation ont été déterminés et présentés sur la figure IV.29. le modèle n'a pas montré d'erreur proportionnelle et systématique, parce que la distribution des résidus des deux côtés de la ligne médiane est aléatoire.

Tous les autres paramètres de tableau IV.15 qui sont satisfaisants indiquent que les modèles QSPR que nous avons construits sont fiables et prédictifs.

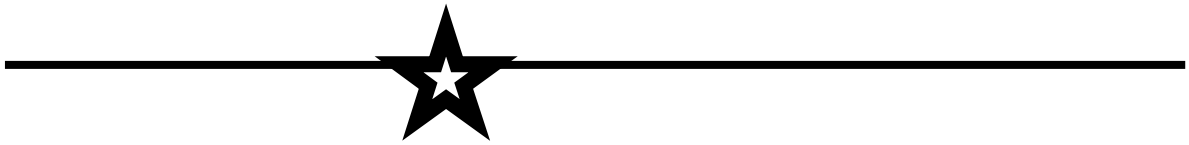
IV.3.8 Conclusion

Dans cette partie de notre étude, on est arrivé à établir une architecture neuronale convenable permettant de prédire la perméabilité des gaz ($CO_{2,\ O2}$), cette capacité de prédiction peut être évalué en réalisant un traitement préliminaire sur la base de données initiale à l'aide d'une méthode « l'algorithme génétique », cette dernière nous permet de déterminer le nombre optimale des variables d'entrée pour la modélisation neuronale qui assure une erreur (REQM) minimale. Les différentes études d'optimisations réalisées sur les deux gaz nous permettent d'évaluer des intervalles de confiance pour la capacité de prédiction des modèles neuronaux conçus pour chaque gaz.

On note ici que les meilleures performances de prédiction ont été trouvées dans le cas de CO_2 on signale des excellentes capacités de prédiction pour cette gaz.



Conclusion générale



CONCLUSION GENERALE

La modélisation de propriétés de molécules constitue un champ de recherche important, les méthodes traditionnelles de modélisation établissent des relations entre les propriétés étudiées et les caractéristiques structurelles des molécules, appelées descripteurs. Leurs principaux inconvénients résident dans la difficulté du choix des descripteurs. Nous avons mis au point une approche de modélisation qui s'affranchit de ces problèmes, en établissant une relation directe entre les descripteurs moléculaires et la propriété modélisée par l'utilisation d'une méthode de prédiction « Algorithme génétique » permettant de déterminer les descripteurs les plus pertinents et nous avons appliqué ensuite la modélisation neuronale.

Cette méthode de modélisation par réseaux de neurones artificiels fait partie de nouvelles techniques sous l'égide de l'intelligence computationnelle ou des sciences cognitives et elle permet de développer des modèles assurant la corrélation et la prédiction de permeabilité des polymères aux gaz.

Pour atteindre cet objectif**,** nous avons élaboré une base de données de dimension « 67-19 pour CO_2 » et « 59-18 pour O_2 » contenant les structures moléculaires des polymères, caractérisées par leurs descripteurs moléculaires qui représentent les différentes propriétés chimiques,et physiques . Le calcul des descripteurs est assuré par différents logiciels (Chembio office, Dragon), un grand nombre de descripteurs différents ont été calculéspour la modélisation de perméabilité.

Pour la recherche de la topologie du réseau de neurones optimal nous avons adopté la stratégie d'erreur, deux architectures neuronales ont été optimisées pour avoir des topologies finales de: {18 :12 :1}; {19 :16 :6 :1} permettant de prédire respectivement la perméabilité des polymères aux gaz (O_2) et perméabilité des polymères aux gaz CO_2).

Pour une meilleure performance de prédiction, nous avons utilisé :

- Des algorithmes d'apprentissage «Resilient Back propagation, Scaled conjugate Gradient
- Des fonctions d'activation Tansig ,logsig et Purlin, les deux première pour la couche cachée et l'autre pour la couche de sortie.

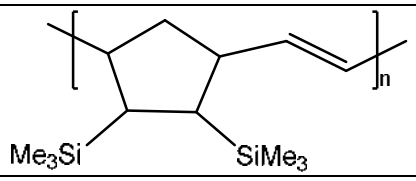
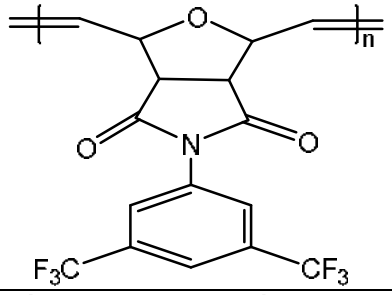
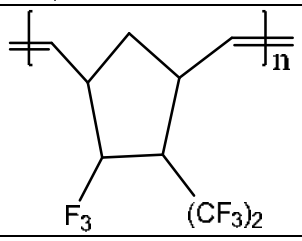
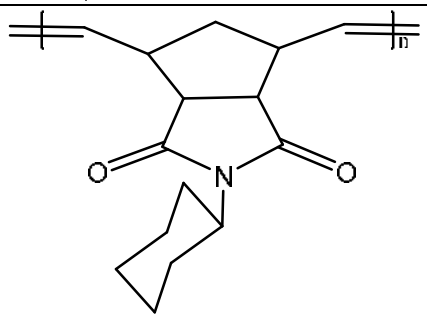
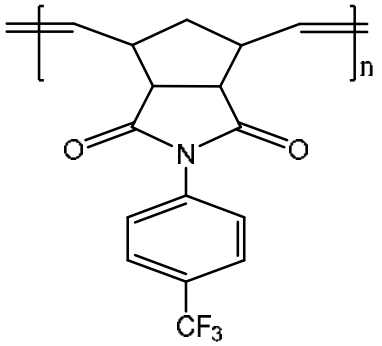
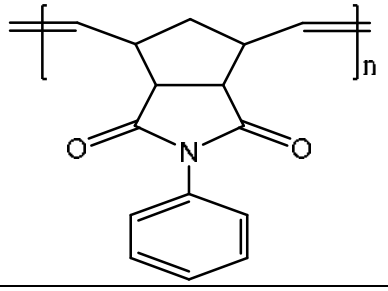
- ☞ La racine de erreur quadratique moyenne (REQM) de 0,12, 0,45, pour CO_2 ,O_2 applicable sur un ensemble des polymères.
- ☞ Un coefficient de corrélation(R) égale à 0,98 pour CO_2, et 0,83 pour O_2 .

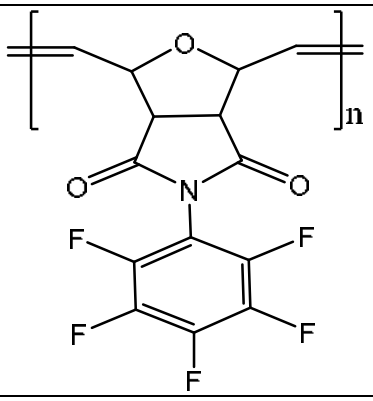
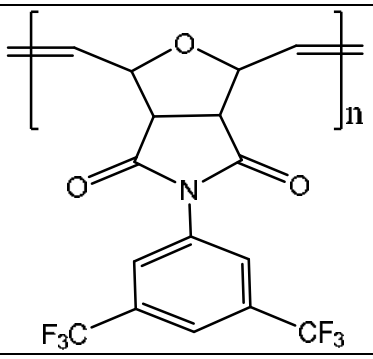
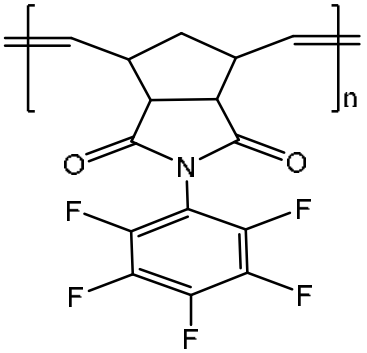
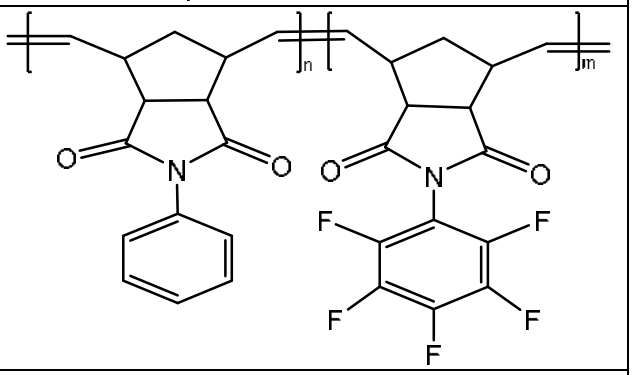
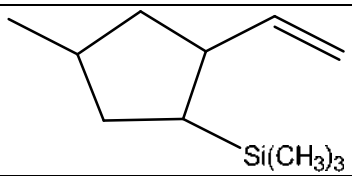
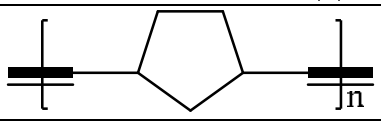
Les résultats obtenus montrent que les meilleures performances de prédiction ont été trouvées dans le cas des permeabilité aux gaz CO_2 ,il reste à étudier les facteurs vont intervenir sur la capacité prédictive notamment

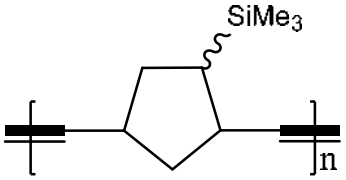
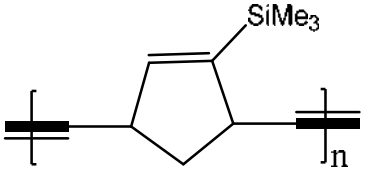
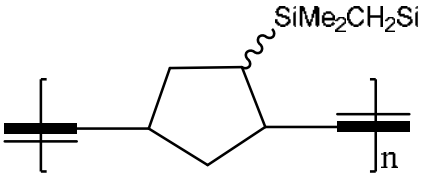
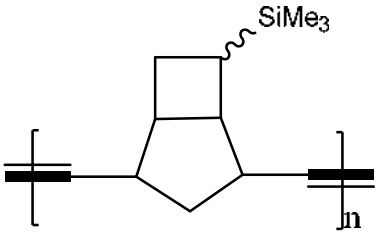
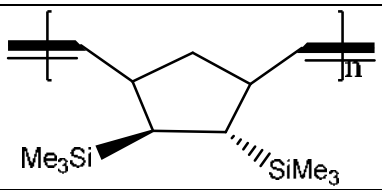
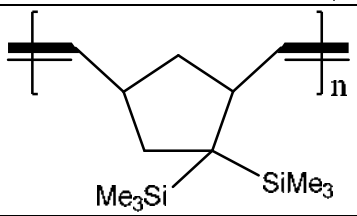
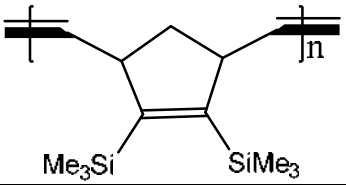
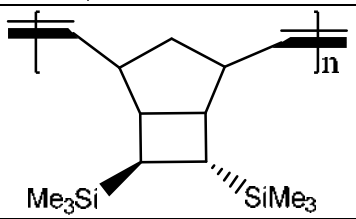
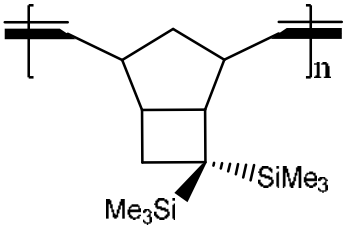
- ➢ La nature des polymèressélectionnées dans la base de données.
- ➢ Augmenter a taille de la base de données.
- ➢ choisir d'autre méthode pour déterminer le nombre et le type de descripteurs.

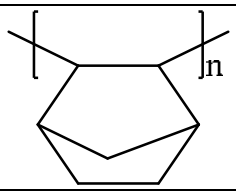
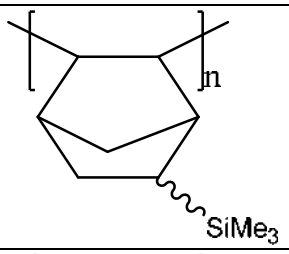
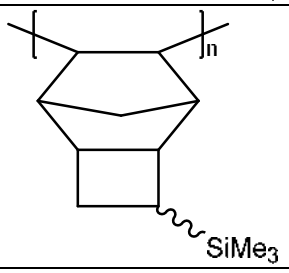
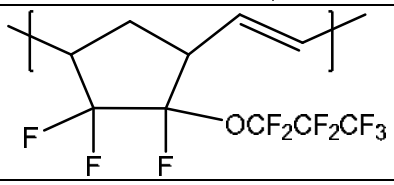
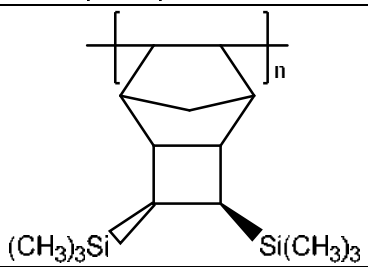
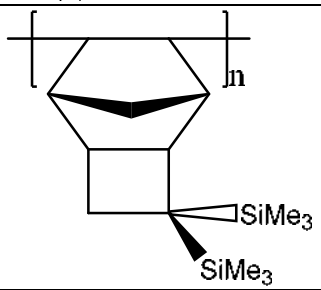
Annexe

Annexe IV.1 : Structure des norbornène

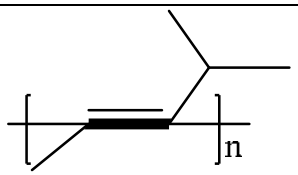
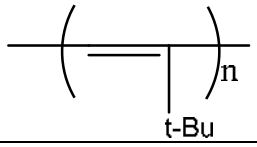
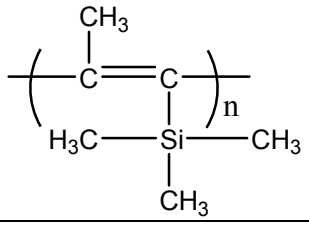
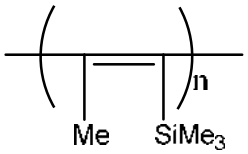
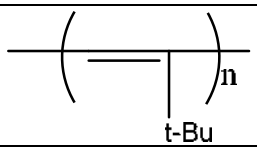
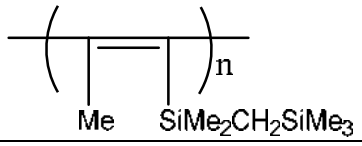
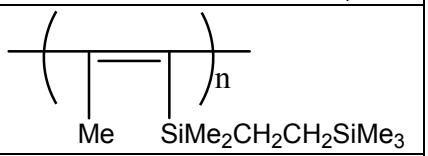
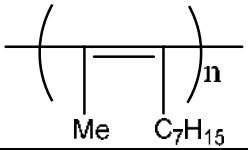
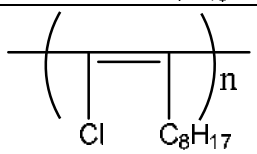
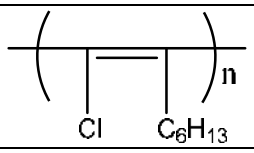
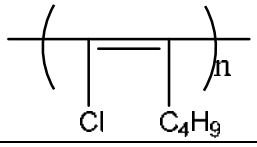
Polymère	Structure
PDSNB	
PTFmPhONDI	
DFHPNB	
PCYNDI	
POLY-2A	
PhNDI	

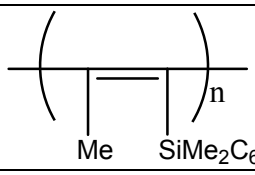
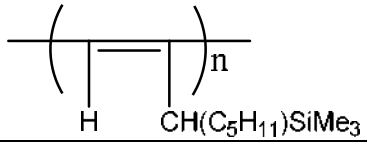
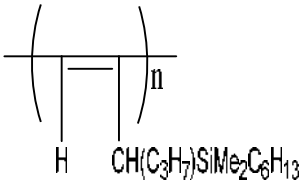
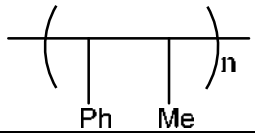
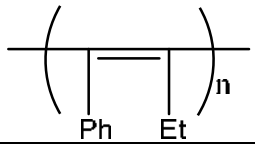
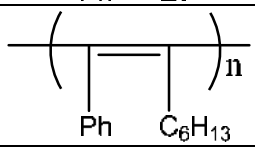
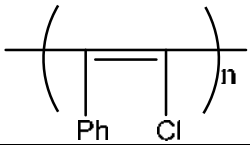
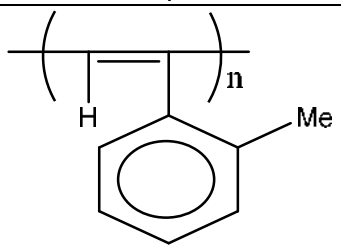
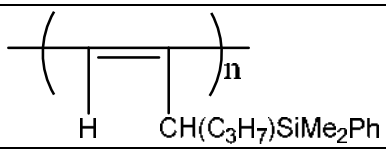
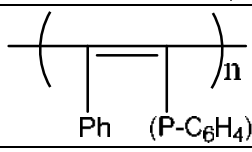
P5FPNDI	
PB3FMPNDI	
P5FPONDI	
P5FNDI	
PTMSNB	
PNB	

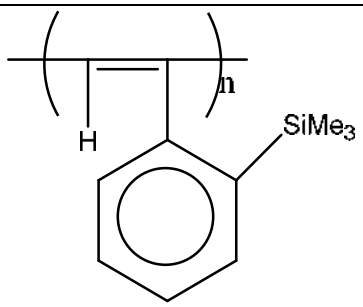
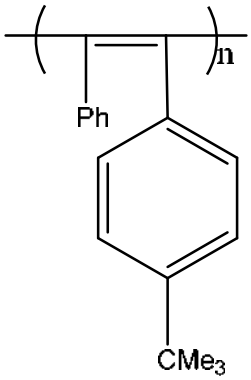
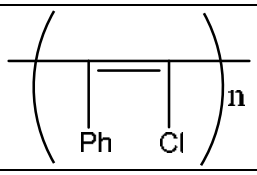
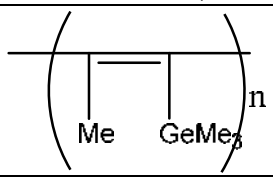
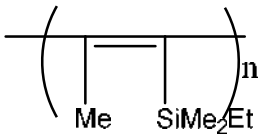
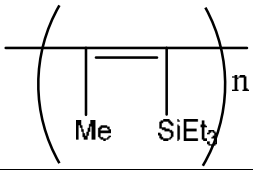
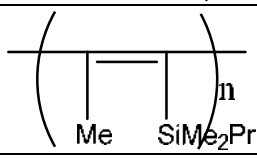
PNBSi	
PNBDSi	
PNBSiCSi	
PTCNSi	
P (6,5-NBSi$_2$)	
P (5,5-NBSi$_2$)	
P(2,3NBDSi$_2$)	
P(3,4-TCNSi$_2$)	
P(3,3-TCNSi$_2$)	

APNB	
APNBSi	
APTCNSi	
POFPNB	
PTCNSi2V	
PTCNSi2g	

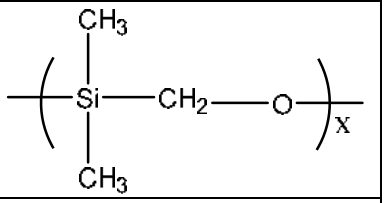
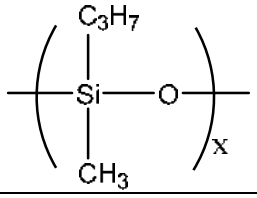
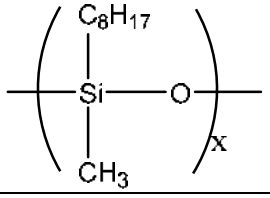
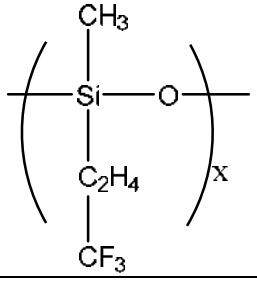
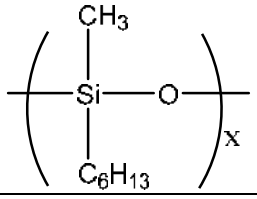
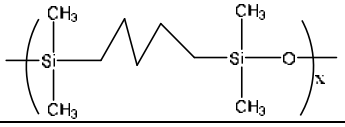
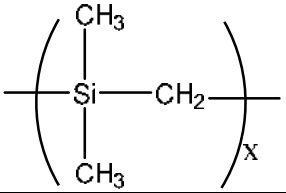
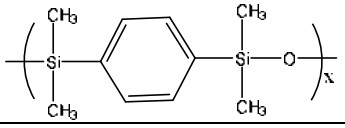
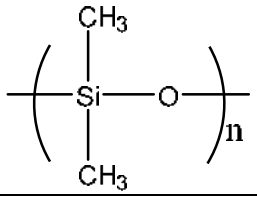
Annexe IV.2 : Structure des polyacétylènes

Polymères	**Structures**
PMP	
PVTMS	
PTMSP	
[C(Me)=C($SiMe_3$)]	
[C(H)=C(t-Bu)]	
[C(Me)=C($SiMe_2CH_2SiMe_3$)]	
[C(Me)=C($SiMe_2CH_2CH_2SiMe_3$)]	
[C(Me) =C(C_7H_{15})]	
[C(Cl)=C(C_8H_{17})]	
[C(Cl)=C(C_6H_{13})]	
[C(Cl)=C(C_4H_9)]	

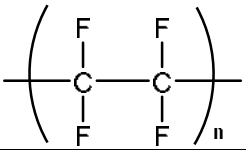
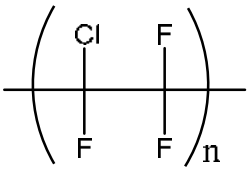
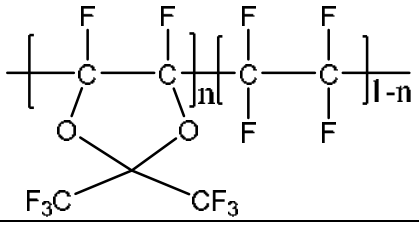
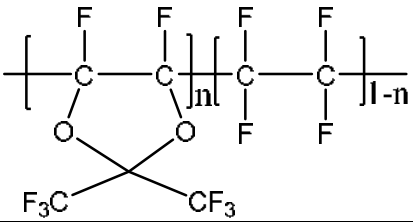
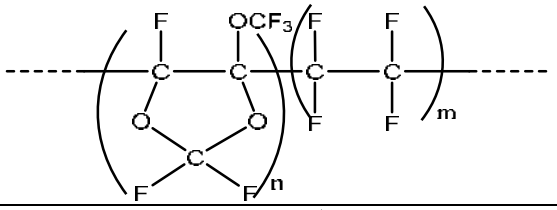
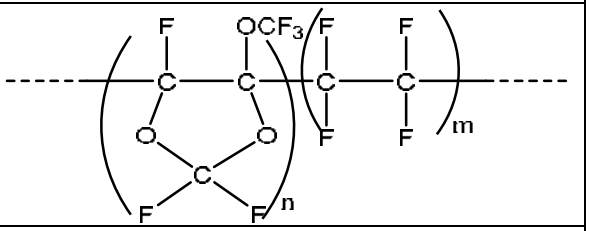
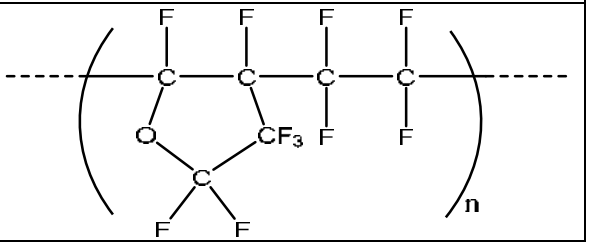
[C(Me)=C($SiMe_2C_6H_{13}$)]	
[C(H)=C(CH(C_5H_{11})$SiMe_3$)]	
C(H)=C(CH(C_3H_7)$SiMe_2C_6H_{13}$	
[C(Ph)=C(Me)]	
[C(Ph)=C(Et)]	
[C(Ph)=C(C6H13)]	
[C(Ph)=C(Cl)]	
[C(H)=C(o-C_6H_4Me)]	
[C(H)=C(CH(C_3H_7)$SiMe_2$Ph)]	
[C(Ph)=C((p-C_6H_4)$SiMe_3$)]	

[C(H)=C($SiMe_3$)]	
[C(Ph)=C(P-$C_6CM_{e_3}$)]	
[C(Cl)=C(Ph)]	
[C(M_e)=C($G_eM_{e_3}$)]	
(CM_e=C ($S_iM_{e_2}Et$)	
[C(M_e)=C($S_iE_{t_3}$)]	
[C(M_e)=C ($S_iM_{e_2}$Pr)]	

AnnexeIV.3: Structure des silicones

Polymere	**Structure**
$[(CH_3)_2SiCH_2O)]$	
$CH_3C_3H_7SiO)$	
$CH_3C_8H_{17}SiO)$	
$[(CF_3C_2H_4)CH_3SiO]$	
$(C_6H_{13}CH_3\text{SiO})$	
$[(CH_3)_2Si(CH_2)_6Si(CH_3)_2O]$	
$[(CH_3)_2SiCH_2\ ]$	
$[(CH_3)_2Si\text{-P-}C_6H_4Si(CH_3)_2O]$	
PDMS	

AnnexeIV.4: des fluoropolymeres

Polymeres	Structure
PTFE	
PCTFE	
AF1600	
AF2400	
HUFLON AD80	
HYFLON AD60	
CYTOP	

Références bibliographique

Bibliographie

[1]: Michel Dupeux. « livre Aide-mémoire science des matériaux » .

[2] : P. Combette et I. Ernoult , « physique des polymères tome I : structure fabricationet emplois » Hermann éditeurs (2005).

[3] : P. WEISS, « Société Francophone de Biomatériaux Dentaires », chimie polymères2009.

[4]: "Polymère" Microsoft ® Encarta ® [CD].Microsoft Corporation, (2007).

[5]: A.NAUDINI , « Nomenclature, classification et formules chimiques des polymères Cécile- génieur de l'École Supérieure de Chimie Industrielle de Lyon Responsable de traités auxTechniques de l'Ingénieur.

[6] : Michel Fontanille –Yves Gnanou , « chimie et physico –chimidespolymère3emédition », (2002).

[7]: http// : cap- science.net/upload/differents Plastique "Différents types de matières Plastique ".

[8] : " Aucun titre de diapositive " http//:www.civil.usherbrooke.Ca/cours/gci116/notes cours/5.2polyméres.pp6.

[9]: M. Fontanille, P. Vairon , « Polymérisation; Techniques de l'ingénieur, Traité plastiques et composites »,(A3 040),pp 1-13.

[10]: Jean-Marks , « elaboration des polymeres »,pp14 , (2005).

[11]: M. Medkour, « Polymères: synthèse macromoléculaire », offices de publications Universitaires (1991).

[12] : D.J. Lougnot, « Techniques d'utilisation des photons, chap.5 », collection Electra Dopee, 85, Paris, 245, (1992).

[13] : C. Roffey, John Wileng, « Photo degradation of reactive species for UV-curing», Chichester , (1997).

[14] : C. Decker ," Progr. Polym. Sci.", 21, 593, (1996).

[15]: G,Natta.,G, Mazzanti, Corradini.Patti.Acad.Nazi.Lincei, and Rend.ClasseSci.ris.Mat. Nat.8.3 .(1958).

[16] : T. ITD, H.Shirakawa and S.Ikeda,Kdbunshi R.

[17]: T. Ito, H. Shirakaua and S.Ikedaj,"polymer,Scipolym",Chem.Ed12.11.(1974).

[18] :G.Natta,G.Mazzanhi,P.Cdrradini,Atti.Acad.Naz.Lincci,Class.Sci.Fis.Mat,Nature , 25,3,(195.

[19]: W. Reppe , D. Schlichting, K. Klager, T. Tdepel Ann. (1948).

[20] : J,Scheirs . Modern Fluoropolymers; Wiley: Chichester, UK, (1997).

[21] : http://www.productknowledge.com/fluropolymers-what-are-fluropolymers-used-for.html

[22]: Paul Binger, Petra Wedemann et Udo H. Brinker. "Cyclopropene : A New Simple Synthesis and its DielsAlder Reaction with Cyclopentadiene". Org. Synth. ; Col Vol. 10 : 231.

[23] : Masaji Oda, Takeshi Kawase, Tomoaki Okada et Tetsuya Enomoto, "2- Cyclohexene-1,4-dione",Org. Synth. ; Coll. Vol. 9 : 186(1982).

[24] :Jean.Marie Michel, « Contribution à l'histoire industrielle des polymèree en France »,pp3-4, (1988).

[25] : A. kahina , « extraction et transport des ions de cadmium(II) par membranes d'affinite dans des dispositif microfluidique »,these de doctorat,pp21,(2015).

[26] : http:// vietsciences.free.fr/giaokhoa/chemistry/polimere/membrance.Htm (dernièreConsultation: Mai (2009).

[27] : S.Loeb ,S. Sourirajan, "UCLA Water Resources Centre Report WR CC-34" , Los Angeles California (1960).

[28] : K. Liamine, « Elaboration de membrane de nanofiltration sur un support composite carbone/zéolite », magistère , pp8, (2014).

[29] : http://www.ibmem.fr/france/membrane separation f .htm(Dernière consultation: Mai(2009).

[30]: www.biophyresearch.com, permeabilité des polymeres.

[31] : Fabienne Piroux, « transport de petites molecules (eau , gaz) dans les poliimides naphtaleniques sulfones- application a la pile a combustible H_2/O_2, deplome de doctorat »,pp57,8 mars (1992).

[32]: J.G. Wijmans, « The role of permeant molar volume in the solution-diffusion mod transport equations », Journal of Membrane Science, 237,pp39-50, (2004).

[33] : B.Flaconnèche et al,"Permeability, Diffusion and Solubility of Gases in Polyethylene, Polyamide 11 and Poly(vinylidene fluoride)". 261-278. , s.l. : Oil & Gas Science and Technology, Vol. 56, (2001).

[34] : C.Dury-Brun et al,"Multiple Mass Transfers of small volatill Molecules Through flexible Food Packaging". Food Rev, Int., 23, pp199-255, (2007).

[35]: E.L. Cussler, "Diffusion, mass transfer in fluids systems", 2nd ed. Cambridge:

Cambridge University Press, (1997).

[36] : M. Ercken, P. Adriaensens, G. Reggers, R. Carleer, D. Vanderzande, and J. Gelan, " Use of magnetic resonance imaging to study transport of methanol in poly (methylmethacrylate) at variable temperature", Macromolecules, 29 pp 5671-5677, (1996).

[37]: J. Crank and G.S. Park," Diffusion in polymers", Academic Press, London, (1968).

[38] : M. Hemati « l'adsorption industrielle », chapitre 2, pp8.

[39] :E.Koller, » Aide-mémoire Génie chimique : chapitre18, adsorption »,Dunod, 2^{eme}édition,pp363-385,(2004).

[40] : B. Urgelli, « perméabilité des roches et loi de darcy » , art , (2000).

[41]:L.M.Sun,J.Y.Thonnelier, « permeation gazzeuse »,technique de l'ingénieur,j2810,pp1-10

[42]: J.Y.Paark,D.R.Paul," Correlation and prediction of gas permeability in glassy polymer membrane materials via modified free based group contribution method ", journal of membrane science, vol.125,pp 23-39, (1997).

[43]: W.M.Lee ,"Selection of barrier materials from molecular structure ", polymer ingeneering science, V 20, 5669, (1980).

[44] : H .Fujita, " Diffusion in polymer-dilluent systems ", Fortschr. Hochpolym.Forsch, V 1-74, (1964).

[45] : H.Fujita ," Organic vapors above the glass transition temperature ", in J.crank and G.S.Park (Eds), Diffusion in polymers, Academic, New York, 75-105, (1968).

[46] : V.T Stannett and M .Szwarc," The permeability of polymer films to gases.A simple relation", Journal of Membrane Science, V 16, 89-91, (1955).

[47]: M .Salame, "Prediction of gas barrier properties of hight polymers ", polym. Eng Science, Vol. N° 22,PP 1543-1546, (1986).

[48] : A. Ya. Alentiev et al, "High transport parameters and free volume perfluorodioxole copolymers", Journal of membrane science 126 pp 123, (1997). 132.

[49] : V. Teplaykov and P.Meares, "Correlation aspects of the selective ga permeability of polymeric materials and membranes ", Gas Sep. Purif, V 4, p66-74, (1990).

[50]: D.R. PAUL," gas transport in homogeneous multicomponent olymers" Journal of Membrane Science, 18 ,pp 77,(1984).

[51]: Y.Yumpolskii and al,"Group contribution method for transport property prediction of glassy polymers: focus on polimides and polynorbornenes",journal of membran science,vol.149,pp 203-220, (1998).

[52]: L.Jia and J.Xu, " A simple method for prediction of gas permeability of polymers from their molecular structure",polymer jornal,vol.23,pp 417, (1991).

[53] : L.M. Robeson and al, "A group contribution approach to predict permeability apermselectivity of aaromatic polymers",journal of membrane Science , vol.132 ,pp33-54, (1997).

[54]: M .Salame, "Prediction of gas barrier properties of hight polymers ", polym. Eng. Sci,V 26,pp1543-1546, (1986).

[55] : S.W. Benson, "Thermochemical Kinetics", Wiley, New York, (1968).

[56] : D.W. Van Krevelen," Properties of Polymers", Elsevier, Amsterdam, 1st ed., 1972, 3rd ed, (1990).

[57]: A.A. Askadskii, Yu.I. Matveev, "Chemical Structure and Physical Properties of Polymers", Khimiya, Moscow, (1983).

[58]: L.M. Robeson, C.D. Smith, M. Langsam, A group contribution approach to Predict permeability and permselectivity of aromatic polymers, J. Membr. Sci. 132 33, (1997).

[59]: D.R. Paul, Ber. Bunsenges. Phys. Chem., 83 ,294,(1979).

[60] : Yu. Yampolskii, I. Pinnau and B. D. Freeman," Transport of Gases and Vapors in Glassy and Rubbery Polymers",pp16,(2006).

[61]: Clem E. Powell, Greg G. Qiao," Polymeric CO2/N2 gas separation membranes for the capture of carbon dioxide from power plant flue gases", Journal of Membrane Science 279 pp 3(2006).

[62] : D. V. Laciak, L . M . Robeson, and C. D. Smith" Group Contribution Modeling of Gas transport in Polymeric Membranes", pp152-153, (1999).

[63]: M.R.Surgi,A.J.Polak and R.C Sundhal,"Description of oxygen permeability in various high polymers using a graph theoretical approach", journal of membrane science,Part 1:. Polym. Chem., vol.27 ,pp 2761-2776,(1989).

[64] : Yu. Yampolskii, I. Pinnau and B. . Freeman,"material science of membrane for gaz andvaour permeation",John Wiley and sons,Ltd,England,(2006).

[65]: L.H. Hall, L.B. Kier," Issues in representation of molecular structure The development of molecular connectivity", Journal of Molecular Graphics and Modelling 20 ,pp 4–18(2001).

[66] : F. Yamashita,M.Hashida,"Mechanistic and empirical modeling of skin permeation of drug",advanced drug delivery reviews,vol.55,pp1185- 1199, (2003).

[67]: G.S.Patit and al,"Empirical correlations for prediction of permeability of gases/liquids through polymers",journal of membrane science,vol.101,pp14 5-152, (1995).

[68] : M.Wessling, M.H.V.Mulder. "Modelling the permeability of Polymers: a neural network approach", Journal of Membrane Science, V 86, 193-198, (1994).

[69]: A.R. Katritzky., D.C. Fara., R.O. Petrukhin, Top. Med. Chemm,pp1333,(2002).

[70]: C,Hansch , a quantitative approach to biochemical structure activity relationships. Accounts of chemical research, pp 232-239,(1969).

[71] : R. Leardi ,Chemometr. 15,pp559-569, (2001).

[72]:Taletesrl, Dragon Profesional Software, Via V.Pisani, 13-20124 Milano (ITALY), (2005).

[73]: IUPAC., Pure and Applied Chemistry, 74,pp1823–1830, (2002).

[74]: Neely., Partition Coefficients to Measure Bioconcentration Potential of Organic Chemicals in Fish. Environ. Sci. Technol, 8,pp 1113-1115, (1974).

[75]: AI ACCESS, 91940, Les Ulis, Franonce.

[76]: B ,SouyeiI ., contribution à la prédiction de coefficient de partage Octanol/eau par la technique QSPR, pp 64-68,(2007).

[77] : D,Chessel .,J Thioulouse ., S Dray ., Co-inertia Analysis and the Linking of Ecological Tables. Ecology, 84,pp3078-3089, (2003).

[78] : G.Abdelkrim. " Élaboration des modèles QSPR prédictifs despropriétés physico-chimiques à l'aidedes descripteurs moléculaires", thèse de doctorat , pp24,(2015).

[79] : CH. Hassan. , "Sélection de caractéristiques: méthodes et applications", thèse de Doctorat,pp23-24, (2011).

[80] : J.H,Holland,. Adaptation in Natural and Artificial Systems. The University o Michigan Press, Ann Arbor, Michigan, (1975).

[81] : Altervino– « Cours de dégustation»,(2006).

[82] : D ,Regadi ., Développement de modèle QSPR pour la prédiction des propriétés physique des quelques composés odorants, pp17-18,(2013).

[83] : K ,Saadi ., Contribution á l'étude de la Relation structure chimique- odeur Utilisation de la technique Random Forest (Application à la famille des pyrazines), p36,(2009).

[84]: C.M,Bishop , Neural Networks for Pattern Recognition, Clarendon Press Oxford(1995).

[85] : D.O, Hebb, . The organization of behaviour Wiley: New York, (1949).

[86] : I.A,Basheer, . and M .Hajmeer. ," artificial neural networks: fundamentals Computing, design, and application". J. Microbiol. Meth, 43,pp3-31,(2000)..

[87]: G.Dreyfus , « Les réseaux de neurones , Mécanique Industrielle et Matériaux », 51(1998).

[88]: Hecht-Nielsen," Kolmogorov's mapping neural network existence theorem. proc.First IEEE Int. Joint Conf Neural Networks", San Diego, June 21-24(1987).

[89] : C Bishop ," Neural Networks for Pattern Recognition", Oxford University Press, (1995).

[90]: G.Cybenko ,” Approximation by superpositions of a sigmoidal function”,Mathematics of Control, Signals, and Systems, 2 4,pp. 303-314,(1989).

[91]:Hecht-Nielsen R,”Neurocomputing”, Addison-Wesley Publishing Company p433 ,(1990).

[92] : J. Gasteiger, J. Zupan, Angew,“ Chem. Int. Ed. Engl”, 32 ,pp503-527,(1993).

[93] :Agatonovic-Kustrin,” S. and Beresford, R, Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research”, J. Pharm. Biomed. Anal, 22,pp 717-727, (2000).

[94] :Kohonen T,”Self-organization and associative memory”, Berlin: Springer-Verlag,(1984).

[95]: D.Rumelhart .E. G. Hinton et R. Williams Learning Representations by Back propagating Errors,Nature, 323,pp 533-536, (1986).

[96] : M.Hagan .T., Menhaj M,” IEEE Transactions on Neural Networks”, 5(1994).

[97]: Y.P. Yampolskii et al,” Thermodynamics of sorption in and free volume of poly(5,6-bis(trimethylsilyl)norbornene)”, Russian Academy of Sciences, 29, Leninsky Prospect, 119991, Moscow, Russian Federation,pp 6945–6952,2004.

[98] : J. Vargas et al,Synthesis and gas permeability of new polynorborne ne dicarboximide with fluorine pendant groups, Polymer 48,pp6546-6553, (2007).

[99]: J. Vargas et al, “The effect of fluorine atoms on gas trans port properties of new polynorbornenedicarboximides”,Journal of Fluorine Chemistry 130 ,162– 168, (2009).

[100] : M. L. Gringol’ts, M. V. Bermeshev,” Highly Permeable Polymer Materials Based on Silicon Substituted Norbornenes”,published in Neftekhimiya, Vol. 50, No. 5, pp. 363–371, (2010).

[101] : P.Pavel . Chapala, V Maxim . Bermeshev,” A Novel, Highly Gas-Permeable Polymer Representing a New Class of Silicon-Containing Polynorbornens As Efficient Membrane Materials», Macromolecules , 48, 8055 −8061,(2015).

[102]: ARCELLA et al, » High Performance Perfluoropolymer Films and Membranes », Ann. N.Y. Acad. Sci. 984: 226–244 (2003).

[103] : C.E. Powell, G.G. Qiao,” Polymeric CO2/N2 gas separation membranes for the capture of carbon dioxide from power plant flue gases», Journal of Membrane Science 279 ,1– 49, (2006).

[104] : S.A. Stern, » Polymers for gas separations: the next decade”, Membrane Sci. 94 ppl-65, (1994).

[105]: R. Leardi,”Application of genetic algorithm–PLS for feature selection in spectral data sets”, Journal of chemometrics J. Chemometrics, 14: 643– 655, (2000).

[106]: M.Wessling, M.H.V.Mulder,” Modelling the permeability of polymers: a neural network approach”, Journal of membrane Science, v86, 193-198, (1994).

[107] : M. Shokrian, M. Sadrzadeh, ”C3H8 separation from CH4 and H2 using a synthesized PDMS membrane: Experimental and neural network modeling”, Journal of Membrane Science, V346, 59-70, (2010).

[108]:H.Demuth, Mark Beale« Neural Network Toolbox for Use with MATLAB», (2002).