REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE YAHIA FARES DE MEDEA



Faculté de **T**echnologie

Département de **G**énie des **P**rocédés et d'**E**nvironnement

En vue de l'obtention du diplôme de MASTER en Génie des Procédés

Filière : Génie des Procédés
Spécialité : Génie Chimique

Modélisation de l'effet de l'adsorption sur la rétention des produits pharmaceutiques lors de nanofiltration et d'osmose inverse

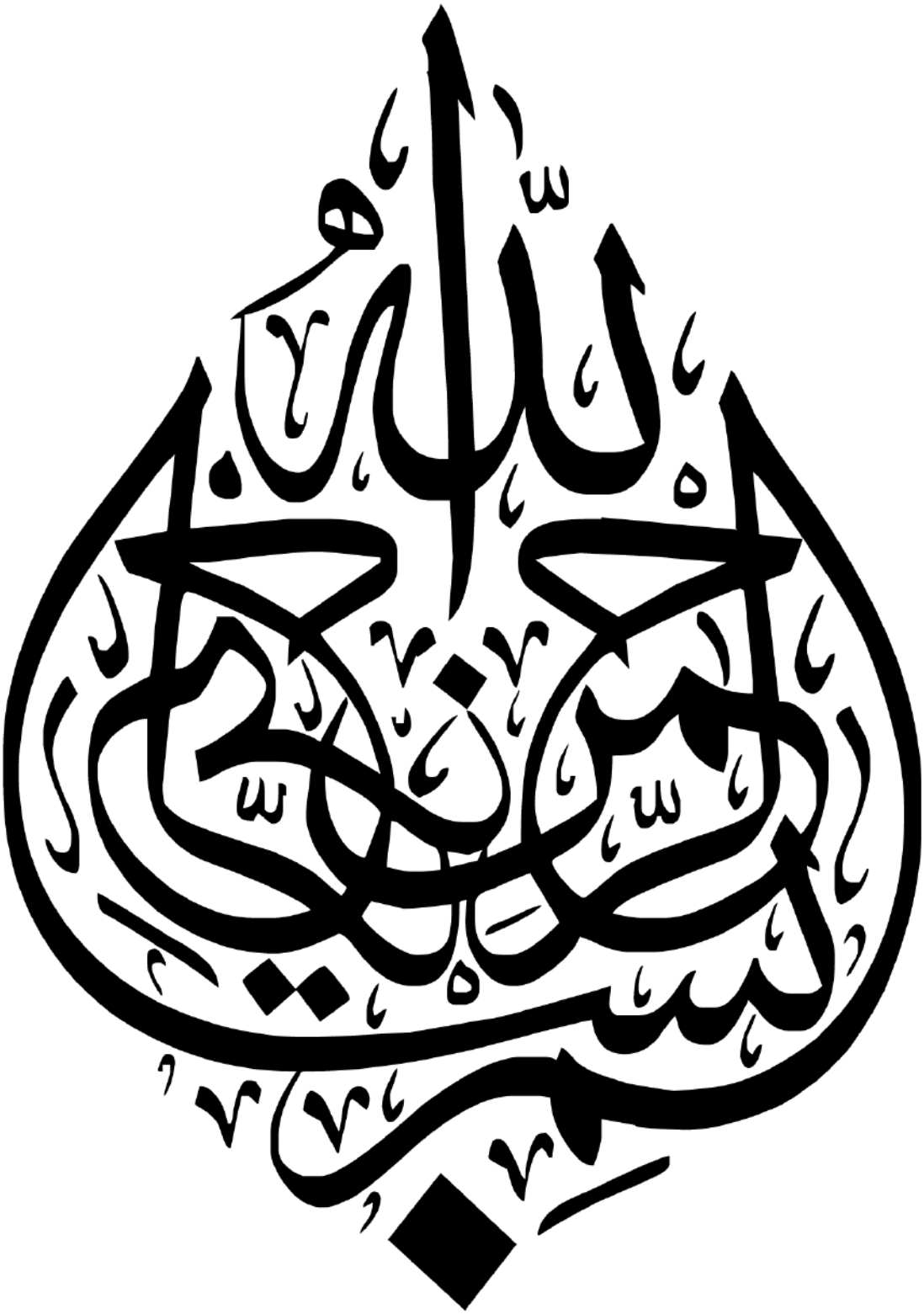
Présenté par :

- M^me^ BENREKIA Sara
- M^elle^ BOULANOUAR Dounia

Devant le jury composé de :

Nom & Prénom	Grade & Etablissement	En Qualité de
Mr. LAIDI Maamar	Professeur, UYFM	Présendent
M^me^. CHABA MOUNA Siham	Maitre de conférence classe B, UYFM	Examinatrice
Mr. HANINI Salah	Professeur, UYFM	Co-promoteur
M^elle^. AMMI Yamina	Maitre de conférences classe A, UYFM	Promotrice

Promotion 2021-2022



Dédicaces

Je dédie ce modeste travail

À mes chers parents,

À mon frère Younes,

À mes sœurs Anfal, Aya, Aroua, Zineb et Fatima,

À mes nièce Lina et Nihal,

À toutes la famille Benrekia

À ma belle binôme Dounia Boulanouar

À toutes mes amies

À tous ceux qui m'ont soutenu de près ou de loin.

Benrekia Sara

Dédicaces

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

A MA CHERE MERE

Aucune dédicace ne saurait exprimer mon respect, mon amour éternel et ma considération pour les sacrifices que tu as consenties pour mon instruction et mon bien être.

Je te remercie pour tout le soutien et l'amour que tu me portes depuis mon enfance et j'espère que ta bénédiction m'accompagne toujours.

Que ce modeste travail soit l'exaucement de votre vœux tant formulé, le fruit de vos innombrables sacrifices.

Puise Allah, le très haut, tu accordes santé, bonheur et longue vie.

A MON CHER PERE

Tu as toujours été à mes côtés pour me soutenir et m'encourager.

Que ce travail traduit ma gratitude et mon affection

UN SPECIAL DEDICACE A MON MARI MAHDI

Qui incarnait l'amour dans toutes ses significations, donc il était mon soutien et mon don, il m'a donné beaucoup d'espoir et de tendresse.

A qui lutte pour mon confort et ma réussite, je ne dirai pas merci, mais je vivrai grâce avec toi

À MON GRAND-PERE BABASIDO

Qui m'a soutenu dans ses prières et qui était une mer limpide coulant d'amour

A MES CHERES GRANDS-PARENTS, MES ONCLES ET MES TANTES

Qu'Allah leur donne une longue et joyeuse vie.

A MES CHERS ZAHRA, RIADH ET RAOUNAK

Qui n'ont pas cessé de me conseiller, encourager et soutenir tout au long de mes études. Qu'Allah protège et leurs offres la chance et le bonheur.

A MES CHERS ANFEL ET DHIAA

Puisse Allah donner santé, bonheur, courage et surtout réussite.

A M'ADORABLE PETITE SŒUR RAGHOUDA

Qui sait toujours comment procures la joie et le bonheur pour toute la famille.

A TOUTE MA FAMILLE BOULANOUAR ET MA BELLE-FAMILLE ELGHERIBI

Je leurs souhaite tout le bonheur du monde.

SANS OUBLIER MON BINOME SARA

Pour son soutien moral, sa patience et sa compréhension tout au long de ce projet, ainsi que sa famille.

Boulanouar Dounia

Remerciement

*Il est agréable d'exprimer notre gratitude premièrement à notre **ALLAH**, la source de toute connaissance qui nous a offert la volonté et la patience pour accomplir ce travail.*

*Nous sommes profondément reconnaissants à notre superviseur le **Dr. AMMI Yamina** pour son encadrement, sa disponibilité et son sens de l'écoute et de l'échange. Et pour nous avoir donné des conseils et un feedback tout au long de ce projet. J'exprime ma gratitude à mon Co-directeur de thème, **Professeur HANINI Salah**.*

Nous voudrions également remercier les membres du jury pour nous avoir fait l'honneur et le plaisir d'accepter et de juger notre travail.

Nous exprimons également notre plus profonde gratitude à tous nos professeurs pour la formation qu'ils nous ont donnée pendant nos études universitaires.

Nous tenons à remercier toutes les personnes qui ont contribué, directement ou indirectement, à la réalisation de ce thème.

Le grand remerciement nous le gardons pour nos parents qui ont toujours veillé sur nous et sur notre avenir.

*Que le tout-puissant **Allah** vous bénisse tous abondamment.*

Résumé

Modélisation de l'effet de l'adsorption sur la rétention des produits pharmaceutiques lors de nanofiltration et d'osmose inverse

L'objectif assigné à ce travail est d'élaborer un modèle QSPR-RN optimal pour modélisation de l'effet de l'adsorption sur la rétention des produits pharmaceutiques lors de nanofiltration et d'osmose inverse.
Un modèle neuronal optimal de type feedforward (QSPR-RN) caractérisé par une structure similaire (treize neurones dans la couche d'entrée, dix-huit neurones dans la couche cachée et un neurone dans la couche de sortie) est construits dans le but de prédire l'effet de l'adsorption sur la rétention des produits pharmaceutiques. Un ensemble de 825 points de données a été utilisé pour tester le réseau de neurones. 80%, 10% et 10% de l'ensemble de données a été utilisée respectivement pour l'apprentissage, la validation et le test. Pour le modèle le plus prometteur de réseau neurone, les valeurs de la rétention calculées de la phase test ont été comparées aux valeurs de la rétention mesurées, de bonnes corrélations ont été trouvées (le coefficient de corrélation "R = 0.9883" et la racine d'erreur quadratique moyenne "REQM = 3.5360%"). L'analyse de sensibilité par la méthode de "poids" a révélé que la rétention des produits pharmaceutiques lors d'osmose inverse et nanofiltration dépend plus précisément des interactions hydrophobe/adsorption que les autres interactions (encombrement strique et interactions électrostatiques).

Mots-clés : Modélisation, Adsorption, Rétention, Produits pharmaceutiques, Nanofiltration, Osmose inverse.

Abstract

Modeling the effect of adsorption on the retention of pharmaceutical actives compounds during the nanofiltration and reverse osmosis

The objective assigned to this work is to develop QSPR-RN a optimal model for modelling the effect of adsorption on the retention of pharmaceutical actives compounds during nanofiltration and reverse osmosis.
An optimal feedforward neuronal model (QSPR-RN) characterized by a similar structure (thirteen neurons in the input layer, eighteen neurons in the hidden layer, and one neuron in the output layer) is constructed in order to predict the effect of adsorption on the retention of pharmaceuticals active compounds. A set of 825 data points was used to test the neural network. 80%, 10%, and 10% of the dataset was used for learning, validation, and testing respectively. For the most promising neural network model, the calculated retention values from of the test phase were compared to the measured retention values, and good correlations were found (the correlation coefficient "R = 0.9883" and the root mean square error "REQM = 3.5360%"). Sensitivity analysis by the "weight" method revealed that the retention of pharmaceutical active compounds during reverse osmosis and nanofiltration depends more precisely on hydrophobic/adsorptive interactions than other interactions (steric hindrances and electrostatic interactions).

Keywords : Modeling, Adsorption, Retention, Pharmaceutical actives compounds, Nanofiltration, Reverse osmosis.

ملخص

نمذجة تاثير الامتزاز على الاحتفاظ بالمنتجات الصيدلانية أثناء الترشيح من فنة النانو و التناضح العكسي

الهدف المحدد لهذا العمل هو تطوير نموذج QSPR-RN الأمثل لنمذجة تأثير الامتزاز على الاحتفاظ بالمنتجات الصيدلانية أثناء الترشيح من فئة النانو والتناضح العكسي.
نموذج عصبي مثالي من نوع feedforward (QSPR-RN)يتميز ببنية مماثلة (ثلاثة عشر خلية عصبية في طبقة المدخلات، ثمانية عشر خلية عصبية في الطبقة المخفية وواحد الخلايا العصبية في طبقة الخرج) من أجل التنبؤ بتأثير الامتزاز على الاحتفاظ بالمنتجات الصيدلانية. تم استخدام مجموعة من 825 نقطة بيانات لاختبار الشبكة العصبية. تم استخدام 80٪ و10٪ و 10٪ من مجموعة البيانات للتعلم والتحقق والاختبار على التوالي. للنموذج الواعد للشبكة العصبية، تمت مقارنة قيم الاحتفاظ المتوقعة لمرحلة الاختبار بقيم الاحتفاظ التجريبية، وتم العثور على ارتباطات جيدة (معامل الارتباط " R = 0.9883" ومتوسط الخطأ التربيعي الجذري " REQM = 3.5360٪ "). كشف تحليل الحساسية من خلال طريقة «الوزن» أن الاحتفاظ بالمنتجات الصيدلانية في التناضح العكسي والترشيح من فئة النانو يعتمد بشكل أكثر دقة على التفاعلات الكارهة للماء/الامتزاز من التفاعلات الأخرى (العوائق الفراغية والتفاعلات الكهروستاتيكية).

الكلمات المفتاحية: النمذجة، الامتزاز، الاحتفاظ، المنتجات الصيدلانية، الترشيح النانو، التناضح العكسي.

Abréviations i
Nomenclature et symboles ii
Liste des tableaux iii
Liste des figures iv

Introduction générale 01

Chapitre I. Rétention des PhACs lors d'OI/NF

I.1. Introduction 04
I.2. Généralités 05
I.2.1. Définition de la membrane 05
I.2.2. Classification des membranes 05
I.2.2.1. Classification d'après la morphologie 07
I.2.2.2. Classification d'après la nature chimique 07
I.2.2.3. Classification d'après la géométrie de préparation des membranes 07
I.2.2.4. Classification d'après le mécanisme de séparation 09
I.2.3. Osmose inverse et nanofiltration 10
I.2.4. Sélectivité d'une membrane 12
I.2.5. Influence des paramètres opératoires 12
I.2.5.1. Pression 12
I.2.5.2. Température 12
I.2.5.3. Taux de récupération "Taux de conversion" 13
I.3. Interactions et propriétés de séparation 13
I.3.1. Encombrement stériques (effet du tamisage) 14
I.3.1.1. État physique de la membrane 14
I.3.1.2. Taille moléculaire 16
I.3.2. Répulsion électrostatique (effet de charge) 18
I.3.2.1. Charge de la surface 18
I.3.2.2. Charge de soluté 19
I.3.2.3. Influence du pH 19
I.3.3. Interaction hydrophobe/adsorption 19
I.3.3.1. Hydrophobicité de la membrane 19
I.3.3.2. Hydrophobicité des solutés 20
I.4. Applications des procédés d'OI et de NF 21
I.5. Avantages et inconvénients des procédés d'OI et de NF 22
I.6. Revue des travaux de la rétention des PhACs lors d'OI/NF 23
I.7. Conclusion 25

Chapitre II. QSAR-Réseaux de neurones

II.1. Introduction 26
II.2. Méthodes d'apprentissage statistiques 26
II.3. Relations quantitatives structure-activité "QSAR" 27
II.3.1. Outils de QSAR 28
II.3.1.1. Activité biologique 28
II.3.1.2. Les descripteurs moléculaires 28
II.3.1.3. Formulation mathématique 30
II.4. Réseaux de neurones artificiels 30
II.4.1. Neurones biologiques 30
II.4.2. Modèle mathématique 31
II.4.3. Architecture des réseaux de neurones 34

II.4.3.1. Réseau de neurones non bouclé (en anglais feedforward) 34
II.4.3.2. Réseau de neurones bouclé (ou récurrent) 35
II.4.4. Modes d'apprentissage 35
II.4.4.1. Apprentissage supervisé 36
II.4.4.2. Apprentissage non supervisé 36
II.4.4.3. Apprentissage renforcé 36
II.5. Réseaux de neurones les plus utilisés 36
II.5.1. Perceptron monocouche (perceptron simple) 36
II.5.2. Perceptron multicouche (en anglais Multi Layer Perceptron) 37
II.5.3. Les réseaux à fonction radiale 37
II.5.4. Les cartes auto-organisatrices de Kohnen (SOM) 38
II.5.5. Les réseaux de Hopfield 39
II.5.6. Les réseau ART 39
II.6. Conception d'un modèle neuronal 40
II.6.1. Détermination des entrées / sorties du réseau de neurones 41
II.6.2. Construction de la base de données et choix des échantillons 41
II.6.3. Définition de la structure du réseau 42
II.6.3.1. Nombre de couches cachées 42
II.6.3.2. Nombre de neurones cachées 42
II.6.4. Apprentissage 43
II.6.5. Evaluation des performances du réseau de neurones 47
II.7. Conclusion 47

Chapitre III. Matériels et méthodes

III.1. Introduction 48
III.2. Collecte des données, prétraitement et analyse 48
III.3. Développement des modèles QSAR 52
III.4. Analyse de sensibilité par la méthode de "poids" 54
III.5. Diagnostic des points aberrants et domaine d'applicabilité du modèle neuronal optimisé 56
III.6. Conclusion 57

Chapitre IV. Résultats et discussion

IV.1. Introduction 58
IV.2. Performance des modèles QSAR-RN 58
IV.2.1. Effet de la division de BD 58
IV.2.2. Performances des modèles QSAR-RN 60
IV.2.3. Expressions mathématiques des modèles neuronaux optimisés 66
IV.3. Analyse de sensibilité par la méthode de "poids" 69
IV.4. Diagnostic des points aberrants et domaine d'applicabilité du modèle neuronal optimisé 71
IV.5. Conclusion 73

Conclusion générale et perspectives 74
Références bibliographiques 76
Annexe A 81
Annexe B 92
Annexe C 96

Symbole	Signification
ART	Adaptive Resonance Theory
BD	Base de Données
BFGS	Broyden–Fletcher–Goldfarb–Shanno
dc	Diamètre effectif
DS	Déviation Standard
FFNN	Feed Forward Neuronal Network
H*	Levier critique
HB	Hydrophobe
HL	Hydrophile
IR	Importance Relative
LM	Levenberg–Marquardt
MF	Microfiltration
MO	Molécule Organique
NF	Nanofiltration
OI	Osmose Inverse
PES	Polyethersulfone
PhACs	Pharmaceuticals Active compounds
PMC	Perceptron Multi-Couches
PS	Polysulfone
QSAR	Quantitative Structure Activity Relationship
QSPR	Quantitative Structure Property Relationship
QSTR	Quantitative Structure toxicity Relationship
RBF	Radial Basis Function Network
RN	Réseaux de Neurones
RNA	Réseaux de Neurones Artificiels
RNFF	Réseaux de Neurones Feedforward
SCG	Scaled Conjugate Gradient
Sc ou MWCO	Seuil de coupure ou Seuil de rétention
TFC	Thin-Film Composite
3D	Trois Dimension
UF	Ultrafiltration

Lecture	Désignation	Unité
$\mathbf{C_{SP}}$	Concentration d'un soluté dans le perméat	(kg/m^3)
$\mathbf{C_{SA}}$	Concentration d'un soluté dans l'alimentation	(kg/m^3)
C	Concentration de la solution	(mol. m^{-3})
$\mathbf{C_0}$	Concentrations des molécules dans la couche de n-octanol	(mol. m^{-3})
$\mathbf{C_w}$	Concentrations des molécules dans la couche de l'eau	(mol. m^{-3})
DCO	Demande Chimique en Oxygène	(mg.L^{-1})
Depth	Profondeur moléculaire	(nm)
EAM	Erreur Absolue Moyenne	(-)
EQM	Erreur Quadratique Moyenne	(-)
EqWidth	Largeur équivalente moléculaire	(nm)
ESP	Erreur Standard de Prédiction	(%)
Length	Longueur moléculaire	(nm)
LogD	Logarithme du coefficient de distribution	(-)
$\mathbf{Logk_{ow}}$ ou **LogP**	Logarithme du coefficient de partage octanol-eau	(-)
MV	Volume molaire	cm^3/mole
MW	Masse molaire	(g.mol^{-1})
π	Pression osmotique de la solution	(Pa)
Ptm	Pression transmembranaire	(Pa)
$\mathbf{Q_A}$	Débit volumique d'Alimentation	(m^3/h)
$\mathbf{Q_P}$	Débit volumique de Permeat	(m^3/h)
REQM	Racine de l'Erreur Quadratique Moyenne	(-)
R	Constante des gaz parfait	(J $mol^{-1}K^{-1}$)
S	Surface	(nm²)
$\mathbf{S_{max}}$	Surface maximale	(nm²)
$\boldsymbol{S_{min}}$	Surface minimale	(nm²)
SR	Rejet de Sel	(-)
T	Température	(K)
$\mathbf{T_r}$	Taux de rétention	(%)
Width	Largeur moléculaire	(nm)
Y	Taux de récupération (Taux de conversion)	(%)
Z	Potentiel zêta	(Mv)

Chapitre I. Rétention des PhACs lors d'OI/NF

Tableau I.1.	Applications des procédés l'OI et NF	22
Tableau I.2.	Avantages et inconvénients de procédés d'OI et NF	23

Chapitre II. QSAR-Réseaux de neurones

Tableau II.1.	Fonctions de transfert usuelles, *f(x)*	33
Tableau II.2.	Les avantages et les inconvenients des quelques architectures des réseaux de neurones	39
Tableau II.3.	Résumé des caractéristiques des algorithmes d'apprentissage de rétro-propagation implantés dans MATLAB®	45

Chapitre III. Matériels et méthodes

Tableau III.1.	Taille de la base de données et le nombre des produits pharmaceutiques utilisés dans ce travail	48
Tableau III.2.	Logiciels utilisés pour calculer *LogD,* le moment dipolaire, S_{min}, S_{max} et Length.	49
Tableau III.3.	Entrées et sortie utilisées pour le développement des modèles QSPR-RNA (QSPR-RNA1, QSPR-RNA2, QSPR-RNA3 et QSPR-RNA4)	50
Tableau III.4.	Analyse statistique des données d'entrée et de sortie (BD)	51

Chapitre IV. Résultats et discussion

Tableau IV.1.	Effet de la division de la base de données (BD1)	58
Tableau IV.2.	Effet de la division de la base de données (BD2)	59
Tableau IV.3.	Effet de la division de la base de données (BD3)	59
Tableau IV.4.	Effet de la division de la base de données (BD4)	60
Tableau IV.5.	Structure des modèles RN optimisés.	60
Tableau IV.6.	Vecteurs de régression linéaire issue de l'équation linéaire : [équation linéaire: $y^{cal} = \alpha y^{exp} + \beta$, $\boldsymbol{\alpha}$(pente), $\boldsymbol{\beta}$(ordonnée à l'origine), $\boldsymbol{R}$(coefficient de corrélation)] et la racine de l'erreur quadratique moyenne ($\boldsymbol{REQM}$).	64
Tableau IV.7.	Poids et biais du QSPR-$RN_{optimal}$.	68

Annexe A

Tableau 1	Liste de 31 produits pharmaceutiques utilisées pour développer les modèles QSPR-RN.	81

Annexe C

Tableau 1	Matrice des coefficients de corrélation pour la base de données (BD)	95

Chapitre I. Rétention des PhACs lors d'OI/NF

Figure I.1.	Membrane sélective.	05
Figure I.2.	Représentation schématique des différentes structures membranaires.	06
Figure I.3.	Représentation schématique d'une membrane multicouche.	06
Figure I.4.	Représentation schématique d'un module tubulaire.	08
Figure I.5.	Représentation schématique d'un module à fibres creuses.	08
Figure I.6.	Représentation schématique d'un module spiralé.	09
Figure I.7.	Gamme de rétention de différentes particules par les procédés membranaires.	10
Figure I.8.	Principe des phénomènes d'osmose inverse.	11
Figure I.9.	Molécule de polyamide et ses équilibres acido-basiques.	13
Figure I.10.	Illustration du principe de détermination du seuil de coupure d'une membrane, la rétention en fonction de la masse molaire (défini ici pour un taux de rejet de 0.90).	15
Figure I.11.	Calcul des dimensions de l'octanol (length, width et depth) de la structure 3D.	18
Figure I.12.	Angle de contact des gouttelettes de liquide sur un matériau solide	20
Figure I.13.	Nombre de publications traitant la rétention des phACspar (a) : nanofiltration (b) osmose inverse	24

Chapitre II. QSAR-Réseaux de neurones

Figure II.1.	Neurone biologique.	31
Figure II.2.	Neurone formel.	32
Figure II.3.	Réseaux de neurones les plus utilisés	34
Figure II.4.	Structure d'un réseau de neurones non bouclé	35
Figure II.5.	Un réseau de neurones bouclé à deux entrées	35
Figure II.6.	L'architecture d'un PMC	37
Figure II.7.	L'architecture d'un RBF	38
Figure II.8.	L'architecture d'une L'architecture d'une carte de Kohonen	38

Chapitre III. Matériels et méthodes

Figure III.1.	Rétentions expérimentales en fonction de diamètre effective "dc".	52
Figure III.2.	Organigramme de développement et l'optimisation de l'architecture d'un QSPR-RNA.	53
Figure III.3.	Organigramme de la méthode de "poids".	55

Chapitre IV. Résultats et discussion

Figure IV.1.	Comparaison entre les valeurs expérimentales et les valeurs calculées pour QSPR-$RN_{optimal}$: (a) la phase totale, (b) la phase d'apprentissage, (c) la phase de validation, (d) la phase de test.	63
Figure IV.2.	Performances des modèles neuronaux QSPR-$RN_{optimal}$ pour : (a) la phase totale, (b) la phase d'apprentissage, (c) la phase de validation, (d) la phase de test.	66
Figure IV.3.	Représentation schématique du QSPR-$RN_{optimal}$	67
Figure IV.4.	Architecture finale du réseau de neurone optimal de type feedforward pour la prédiction de la rétention des produits pharmaceutiques par les membranes d'OI/NF.	69

Figure IV.5. Présentation des pourcentages de l'importance relative (IR) de chaque variable d'entrée sur la rétention : QSPR-$RN_{optimal}$ 70

Figure IV.6. Diagnostic des données potentiellement suspectes et domaine d'applicabilité du modèle du modèle neuronal optimal QSPR-$RN_{optimal}$ pour la phase totale 71

Figure IV.8. Diagnostic des données potentiellement suspectes et domaine d'applicabilité du modèle du modèle neuronal optimal QSPR-$RN_{optimal}$ pour la phase validation 72

Figure IV.9. Diagnostic des données potentiellement suspectes et domaine d'applicabilité du modèle du modèle neuronal optimal QSPR-$RN_{optimal}$ pour la phase test 72

Annexe B

Figure 2. Capture d'écran du site intitulé Chemspider 92

Figure 3. Capture d'écran des logiciels : (a) ChemBioDraw et (b) ChemBio3D 93

Figure 4. Capture d'écran de logiciel Hperchem. 93

Figure 5. Capture d'écran de logiciel STATISTICA 2010 93

Figure 6. Capture d'écran de logiciel MATLAB 2011 94

Abréviations i
Nomenclature et symboles ii
Liste de tableaux iii
Liste de figures iv

Introduction générale 1

Chapitre I. Rétention des PhACs lors d'OI/NF

I.1. Introduction 04
I.2. Généralités 05
I.2.1. Définition de la membrane 05
I.2.2. Classification des membranes 05
I.2.2.1. Classification d'après la morphologie 07
I.2.2.2. Classification d'après la nature chimique 07
I.2.2.3. Classification d'après la géométrie de préparation des membranes 07
I.2.2.4. Classification d'après le mécanisme de séparation 09
I.2.3. Osmose inverse et nanofiltration 10
I.2.4. Sélectivité d'une membrane 12
I.2.5. Influence des paramètres opératoires 12
I.2.5.1. Pression 12
I.2.5.2. Température 12
I.2.5.3. Taux de récupération "Taux de conversion" 13
I.3. Interactions et propriétés de séparation 13
I.3.1. Encombrement stériques (effet du tamisage) 14
I.3.1.1. État physique de la membrane 14
I.3.1.2. Taille moléculaire 16
I.3.2. Répulsion électrostatique (effet de charge) 18
I.3.2.1. Charge de la surface 18
I.3.2.2. Charge de soluté 19
I.3.2.3. Influence du pH 19
I.3.3. Interaction hydrophobe/adsorption 19
I.3.3.1. Hydrophobicité de la membrane 19
I.3.3.2. Hydrophobicité des solutés 20
I.4. Applications des procédés d'OI et de NF 21
I.5. Avantages et inconvénients des procédés d'OI et de NF 22
I.6. Revue des travaux de la rétention des PhACs lors d'OI/NF 23
I.7. Conclusion 25

Chapitre II. QSAR-Réseaux de neurones

II.1. Introduction 26
II.2. Méthodes d'apprentissage statistiques 26
II.3. Relations quantitatives structure-activité "QSAR" 27
II.3.1. Outils de QSAR 28
II.3.1.1. Activité biologique 28
II.3.1.2. Les descripteurs moléculaires 28
II.3.1.3. Formulation mathématique 30
II.4. Réseaux de neurones artificiels 30
II.4.1. Neurones biologiques 30

II.4.2. Modèle mathématique 31
II.4.3. Architecture des réseaux de neurones 34
II.4.3.1. Réseau de neurones non bouclé (en anglais feedforward) 34
II.4.3.2. Réseau de neurones bouclé (ou récurrent) 35
II.4.4. Modes d'apprentissage 35
II.4.4.1. Apprentissage supervisé 36
II.4.4.2. Apprentissage non supervisé 36
II.4.4.3. Apprentissage renforcé 36
II.5. Réseaux de neurones les plus utilisés 36
II.5.1. Perceptron monocouche (perceptron simple) 36
II.5.2. Perceptron multicouche (en anglais Multi Layer Perceptron) 37
II.5.3. Les réseaux à fonction radiale 37
II.5.4. Les cartes auto-organisatrices de Kohnen (SOM) 38
II.5.5. Les réseaux de Hopfield 39
II.5.6. Les réseau ART 39
II.6. Conception d'un modèle neuronal 40
II.6.1. Détermination des entrées / sorties du réseau de neurones 41
II.6.2. Construction de la base de données et choix des échantillons 41
II.6.3. Définition de la structure du réseau 42
II.6.3.1. Nombre de couches cachées 42
II.6.3.2. Nombre de neurones cachées 42
II.6.4. Apprentissage 43
II.6.5. Evaluation des performances du réseau de neurones 47
II.7. Conclusion 47

Chapitre III. Matériels et méthodes

III.1. Introduction 48
III.2. Collecte des données, prétraitement et analyse 48
III.3. Développement des modèles QSAR 52
III.4. Analyse de sensibilité par la méthode de "poids" 54
III.5. Diagnostic des points aberrants et domaine d'applicabilité du modèle neuronal optimisé 56
III.6. Conclusion 57

Chapitre IV. Résultats et discussion

IV.1. Introduction 58
IV.2. Modèles QSAR-RN 58
IV.2.1. Effet de la division de BD 58
IV.2.2. Performances des modèles QSAR-RN 60
IV.2.3. Expressions mathématiques des modèles neuronaux optimisés 66
IV.3. Analyse de sensibilité par la méthode de "poids" 69
IV.4. Diagnostic des points aberrants et domaine d'applicabilité du modèle neuronal optimisé 71
IV.5. Conclusion 73

Conclusion générale et perspectives 74
Références bibliographiques 76
Annexe A 81

Annexe B 92
Annexe C 95

L'utilisation de produits pharmaceutiques dans le monde augmente chaque année. Par conséquent, des produits pharmaceutiques ont été détectés dans le monde entier dans différentes matrices environnementales. De nos jours, les produits pharmaceutiques sont présents dans la gamme ng/L (eaux de surface, de mer et souterraines) comme ainsi que dans la gamme mg/L (effluents hospitaliers et municipaux eaux usées), ce qui semble indiquer un possible impact négatif sur les écosystèmes et une menace pour la santé publique. L'osmose inverse (OI) et la nanofiltration (NF) se sont avérées être efficace dans l'élimination de divers produits pharmaceutiques (Dolar et al. 2016). Dans une application réelle NF/OI, les sites d'adsorption membranaire seront occupés non seulement par de nombreux types de micro-organiques hydrophobes molécules mais aussi par matière organique naturelle en vrac sur un fonctionnement à long terme (Kimura et al. 2003).

Les procédés de séparation membranaires regroupent un grand nombre de techniques permettant de réaliser des séparations en phase liquide ou en phase gazeuse sous l'action de diverses forces de transfert (mécanique, électrique ou chimique). Dans le cas de la filtration en phase liquide réalisée au moyen de procédés baro-membranaires, c'est-à-dire utilisant un gradient de pression comme force motrice, on distingue quatre techniques répertoriées sous les noms de microfiltration (MF), ultrafiltration (UF), nanofiltration (NF) et osmose inverse (OI). La micro-, l'ultra- et la nanofiltration mettent en jeu des membranes possédant une structure poreuse tandis que les membranes d'osmose inverse sont des matériaux denses (Ammi 2016).

La littérature met en lumière plusieurs points. D'une part, la nanofiltration et l'osmose inverse sont des procédés de plus en plus répandus et de nombreuses études indiquent qu'il permet de traiter différents types d'effluents en retenant les produits pharmaceutiques. D'autre part, le rôle complexe et encore mal cerné des nombreuses interactions qui peuvent s'établir entre les solutés, l'eau et la membrane. De nombreux modèles de transferts existent mais ils ne décrivent ou n'expliquent pas toujours correctement le procédé et les résultats obtenus dans le cas du transfert des produits pharmaceutique nécessitent des modifications importantes.

La modélisation est une autre approche permettant de diminuer le nombre d'expériences à effectuer pour déterminer les paramètres optimaux pour l'effet de l'adsorption sur la rétention des produits pharmaceutiques lors de nanofiltration et d'osmose inverse. Si elle représente bien le procédé, les influences des paramètres optimaux pourront être évaluées sans aucune expérience et n'auront qu'à être validés sur le site.

Dans ce contexte s'inscrit ce travail qui consiste d'une part à l'élaboration d'une base de données élaborée à partir de la littérature et d'autre part à modélisation de l'effet de l'adsorption sur la rétention des produits pharmaceutiques lors de nanofiltration et d'osmose inverse, en utilisant les réseaux de neurones.

Pour surmonter ces problèmes, des tentatives ont été faites pour proposer des approches de modélisation permettant de diminuer le nombre d'expériences à effectuer pour déterminer les paramètres optimaux pour le traitement d'un effluent industriel. Si elle représente bien le procédé, les influences des paramètres opératoires pourront être évaluées sans aucune expérience et n'auront qu'à être validées sur le site. De plus, la modélisation peut également permettre de valider les hypothèses simplificatrices émises lors de la formulation mathématique des phénomènes de transfert. Pour cela, il faut chercher d'autres méthodes de modélisation telles que les sciences cognitives (Ammi 2016).

La révolution de l'informatique et le développement de nouvelles méthodes de calcul numérique basés sur l'intelligence artificielle et les sciences cognitives ont donné naissance durant les dernières décennies à de nouvelles méthodes de calcul et de modélisation telles que les réseaux de neurones artificiels "RNA", les algorithmes génétiques et la logique floue. L'efficacité, la fiabilité et la robustesse de ces méthodes par rapport aux difficultés de l'utilisation des approches classiques ont été démontrées par leurs nombreuses applications dans le domaine de génie des procédés (Azizi 2011).

Le principal avantage des RNA, comparé aux modèles de connaissance classique, est leur capacité à modéliser ces procédés sans avoir une connaissance approfondie de la phénoménologie mise en jeu. Un autre avantage important des RNA, comparé à d'autres approximateurs universels tels que les régressions linéaires ou polynomiales, est qu'ils sont des approximateurs non-linéaires parcimonieux, permettant une modélisation de procédés non-linéaires quel que soit le nombre de paramètres indépendants utilisés (Dreyfus et al. 2002).

L'objectif de ce travail est l'application des RN de type feedforward pour modéliser la rétention des produits pharmaceutiques par les membranes d'OI/NF. Ainsi, nous avons organisé ce mémoire en quatre chapitres :

- Le premier chapitre est consacré à la description de la technologie membranaire, on donnera dans un premier temps un aperçu sur les généralités sur les membranes, ensuite les interactions et les propriétés de séparation, puis les applications et les avantages

d'OI/NF, à la fin, une revue des travaux de la rétention des PhACs lors d'OI/NF a été présentée.

- Le deuxième chapitre est dédié à la présentation des réseaux de neurones artificiels basés sur la relation quantitative structure propriétés "QSPR-RNA".

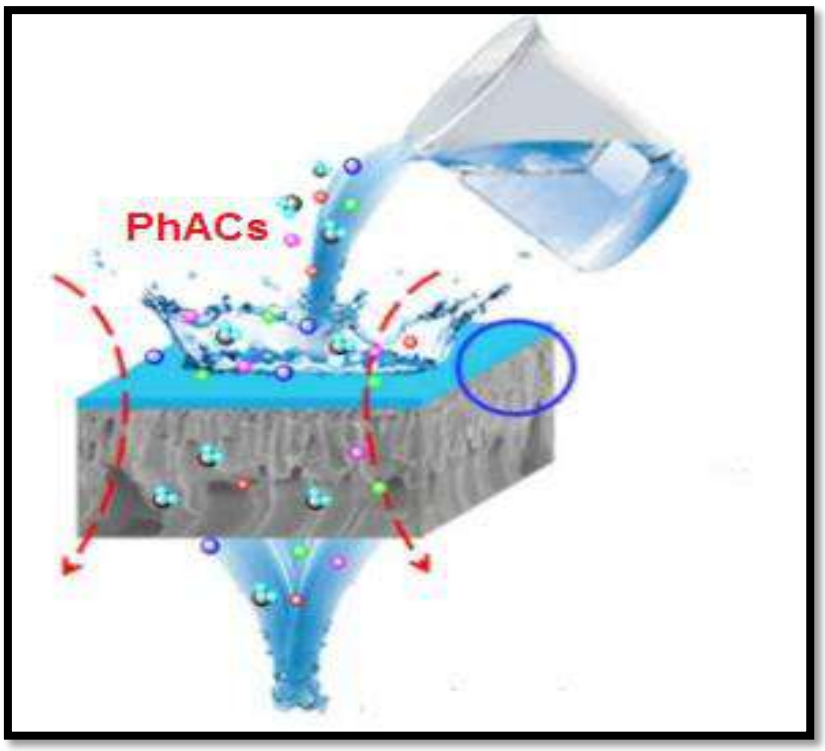
- Nous présentons dans le troisième chapitre les matériels et les différentes procédures de modélisations utilisées. Il comptera quatre axes : dans le premier axe, on examinera les données collectées, prétraitement et analyse. Le second axe sera consacré aux techniques les plus courantes pour établir modèle QSPR, les réseaux de neurones artificiels (RNA). Le troisième axe traitera l'analyse de sensibilité basée la méthode de "poids". Dans le quatrième et dernier axe on présentera le diagnostic des points aberrants et domaine d'applicabilité du modèle neuronal optimisé.

- Le quatrième chapitre est organisé en trois parties : dans la première partie, on discutera l'effet de la division de la base de données "BD", les performances et la formule mathématique de modèle neuronal optimale QSPR-$RN_{optimal}$. Dans la deuxième partie on présentera les résultats de l'analyse de sensibilité par la méthode de "poids". On présentera les résultats de l'analyse diagnostic des points aberrants et domaine d'applicabilité du modèle neuronal optimisé dans la troisième partie.

Les conclusions tirées de ce travail ainsi que les études qui peuvent être envisagées à l'avenir sont présentées à la fin de ce thème.

Chapitre I

Rétention des PhACs lors d'OI/NF



I.1. Introduction

La filtration sur membrane est un procédé de séparation physique se déroulant en phase liquide. Le but est de purifier, fractionner ou concentrer des espèces dissoutes ou en suspension dans un solvant au travers d'une membrane.

La technique de séparation membranaire est connue depuis un certain temps mais son développement reste récent. Depuis les années 70 on retrouve ces techniques principalement dans l'industrie laitière, des boissons, des ovo-produits, des jus de fruits ou du traitement des eaux. Ces procédés consomment relativement peu d'énergie et sont sélectifs.

Lorsque l'on effectue une filtration sur membrane, on obtient :

- Le rétentat (molécules et/ou particules retenues par la membrane) ;
- Le perméat ou filtrat (molécules qui passent à travers la membrane).

Cette technique permet par exemple de séparer par filtration un liquide des micro-organismes qu'il contient pour pouvoir les dénombrer. Cette méthode s'applique à des liquides sans particules solides.

Les membranes sont en général caractérisées par :

- La taille des pores ;
- Le seuil de coupure (Masse molaire critique pour laquelle 90 % des solutés sont retenus par la membrane) ;
- Leur sélectivité ;
- Leur perméabilité(https://fr.wikipedia.org/wiki/Filtration_sur_membrane16/05/2022).

Dans ce premier chapitre nous présentons une étude bibliographique, dont la majeure partie porte sur les différents aspects des membranes, à savoir leurs classifications, leurs principes de mise en œuvre, les différentes configurations des modules membranaires et les procédés bar-membranaires d'une manière générale. Nous étudierons ensuite les interactions solutés-membranes qui gouvernent les propriétés de séparation des composés organiques et nous ferons un bilan des modèles décrivant le transfert de matière au sein des membranes de nanofiltration. Nous terminerons ce chapitre par une revue des travaux de la rétention des PhACs lors d'OI/NF.

I.2. Généralités

I.2.1. Définition de la membrane

Une membrane peut être définie comme étant une couche mince de matière, permettant l'arrêt ou le passage sélectif de substances dissoutes ou non, sous l'action d'une force motrice de transfert (Tamas 2004). Une membranes semi- sélectives est une membrane permettant certains transferts de matière entre deux milieux qu'elle sépare, en interdisant d'autres, ou plus généralement, en favorisant certains par rapport à d'autres (Baker 2004), voir figure I.11. Les membranes d'osmose inverse sont composées d'une couche mince (couche active : peau) de faible épaisseur comprise entre 0,1µm et 1,5µm, composant des micropores. Cette couche active est supportée par une ou plusieurs couches, à la fois plus poreuses et mécaniquement plus résistantes (Audinos 2000).

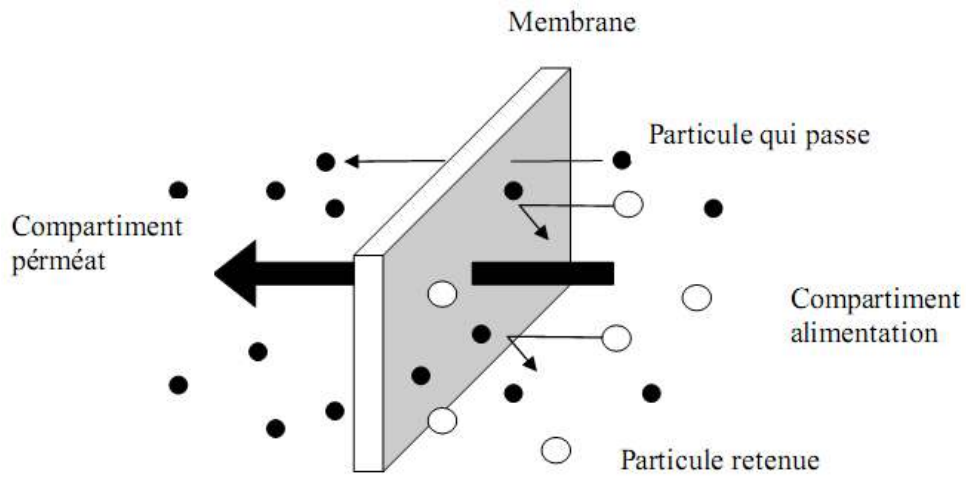


Figure I.1. Membrane sélective (Annane 2015).

I.2.2. Classification des membranes

La classification des membranes se fait sur plusieurs critères :

I.2.2.1. Classification d'après la morphologie

Les membranes peuvent être également classées en fonction de la structure des différents matériaux, on distingue :

- **Membranes à structure symétrique "isotrope" :** ce sont des membranes denses ou poreuses ayant la même structure (structure homogène) sur toute leur épaisseur (Chikhi 2008) (Figure I.2 "a").

- **Membranes à structure asymétrique "anisotrope" :**la structure de la membrane varie d'une couche à une autre (Figure I.2 "b"). Celles-ci présentent une structure hétérogène.

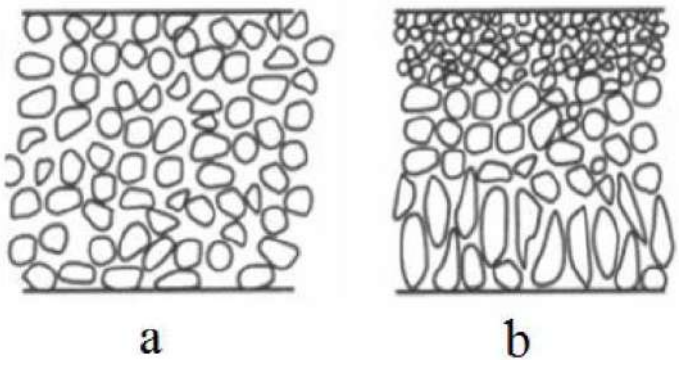


Figure I.2. Représentation schématique des différentes structures membranaires (Chabane2015).

Deux sous-types de membranes asymétriques peuvent été distingués (Annane 2015) **:**

- Membranes préparées à partir du même matériau ;
- Membranes composites : ce sont des membranes constituées principalement de deux couches :
  - **Peau :** une couche de très faible épaisseur en contact direct avec l'eau à traiter. Cette couche donne à la membrane sa sélectivité ;
  - **Couche support :** une couche plus épaisse, d'une perméabilité beaucoup plus grande, qui supporte la peau et confère à la membrane sa résistance mécanique.

La plupart des membranes commerciales de MF, UF et de NF sont des membranes à structure asymétrique, car ce type de structure permet d'avoir des perméabilités plus élevées. Une membrane asymétrique est en général constituée d'une couche support macroporeuse (assurant une bonne résistance mécanique) et d'une couche active (régissant les propriétés de rétention entre lesquelles s'insèrent une ou plusieurs couches intermédiaires dont le rôle est de permettre le dépôt de la couche active et sa tenue en pression (Figure I.3) (Chikhi 2008).

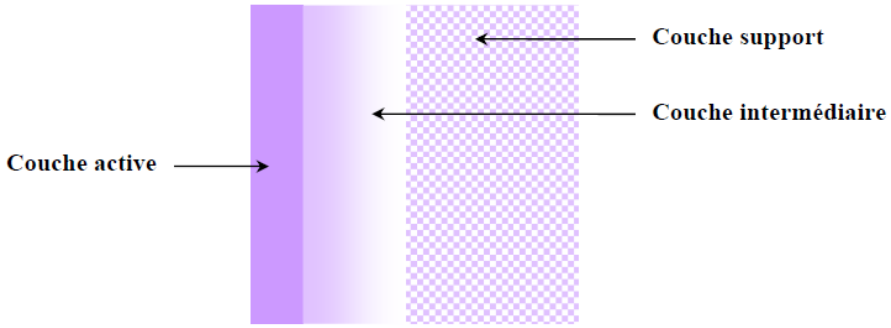


Figure I.3. Représentation schématique d'une membrane multicouche (Bouranene 2008).

I.2.2.2. Classification d'après la nature chimique

Les membranes sont fabriquées à partir des polymères organiques et de matières inorganiques. Les principaux polymères utilisés pour la fabrication des membranes sont (Annane 2015) :

a. **Dérivés de la cellulose** : ces polymères considérés généralement comme plus hydrophiles, peu coûteux et possédant une faible tendance à l'adsorption.
b. **Polysulfone (PS) et polyethersulfone (PES)**, qui ont une tendance à l'adsorption plus élevée que les membranes ci-dessus, mais avec une meilleure stabilité chimique, mécanique et thermique.
c. **Les membranes inorganiques** sont fabriquées à base de céramique (oxydes d'aluminium, de zirconium ou de titane). Ces membranes ont une stabilité chimique, mécanique et thermique supérieure que les membranes organiques (Hamlaoui 2002).

I.2.2.3. Classification d'après la géométrie de préparation des membranes

Les membranes peuvent être classées en fonction de la forme sous laquelle elles sont fabriquées (Chikhi 2008):

- Membranes planes ;
- Membranes spirales ;
- Membranes cylindriques ;
- Membranes tubulaires (ayant le diamètre interne plus grand que 3 mm) ;
- Membranes creuses (ayant le diamètre plus petit que 3 mm).

La forme des membranes conditionne la manière dont elles sont assemblées pour fabriquer des modules.

- **Modules membranaires**

La séparation se déroule dans un module de filtration pouvant prendre plusieurs formes (Lanteri 2009):

✓ **Module plan :** Le système utilise des membranes sous forme de feuilles. Il s'agit du type de configuration la plus simple. La surface filtrante reste toutefois faible par rapport à l'encombrement engendré.
✓ **Module tubulaire :** Il est constitué par des faisceaux de tubes de 4 à 25mm de diamètre intérieur. Ils sont basés sur une technologie simple, une consommation élevée en énergie est nécessaire pour une très faible surface d'échange par unité de volume (compacité réduite) (Figure I.4).

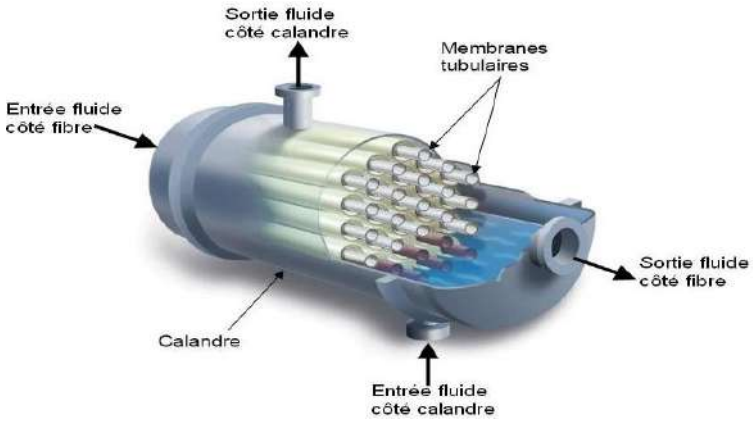


Figure I.4. Représentation schématique d'un module tubulaire(https://www.researchgate.net/figure/Les-modules-tubulaires_fig1_324474113).

- ✓ **Fibre creuse :** Elle rassemble un ensemble de fibres creuses de diamètre externe inférieur à 1 millimètre en un faisceau. Cette configuration leur procure la plus forte densité d'écoulement par module (Figure I.5).

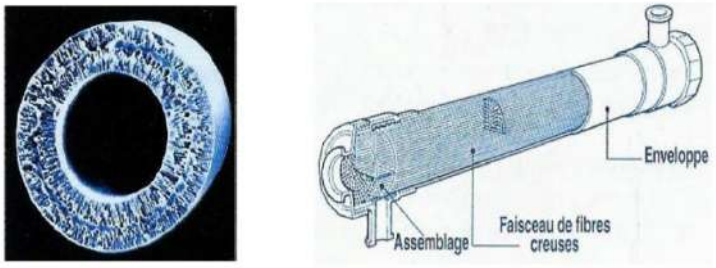


Figure I.5. Représentation schématique d'un module à fibres creuses (Chabane 2015).

- ✓ **Module spiralé :** Une ou plusieurs membranes planes sont enroulées autour d'un tube poreux et creux destiné à collecter le perméat. Ce type de module permet d'obtenir des surfaces filtrantes importantes pour un faible encombrement et un coût peu élevé. C'est une solution typiquement choisie dans l'industrie agro-alimentaire pour le traitement des fluides clarifiés (Figure I.6) (Antoine 2004).

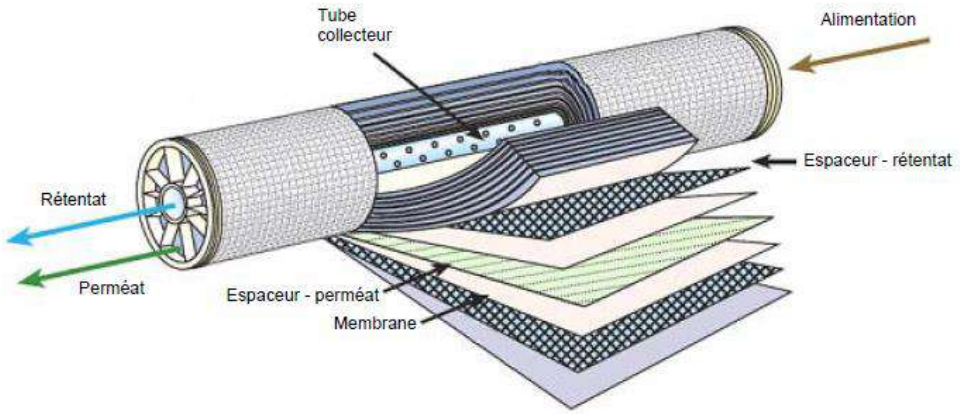


Figure I.6. Représentation schématique d'un module spiralé (Antoine 2004).

I.2.2.4. Classification d'après le mécanisme de séparation

La classification la plus répandue des membranes d'après le mécanisme de séparation est la suivante (Annane 2015) :

a. Membranes poreuses (MF, UF, NF), où les effets de tamisage et de friction jouent des rôles importants, comme en NF, les forces de surface peuvent jouer un rôle important. Les pores peuvent être subdivisés en plusieurs catégories :

- ✓ Macropores-plus grands que 50 nm (MF) ;
- ✓ Mésopores-2 à 50 nm (UF) ;
- ✓ Micropores-plus petits que 2 nm (NF).

b. Membranes denses (perméation de gaz, pervaporation, osmose inverse (OI)) ; ces membranes peuvent être considérées comme des milieux denses où la diffusion des espèces a lieu dans les volumes libres situés entre les chaînes moléculaires du matériel de la membrane ;

c. Membranes échangeuses des ions : Ces membranes sont considérées comme un type spécial de membranes non poreuses, consistant dans des gels denses, ayant une charge positive (échangeuses des anions) ou négative (échangeuses des cations).

Il a noté que la communauté scientifique ne s'entende pas sur cette classification. (Aptel et Buckley 1996) considèrent que les membranes d'OI sont des membranes poreuses. La figure I.7 montre schématiquement la correspondance entre les opérations de filtration par membrane, la taille des pores et les composés retenus.

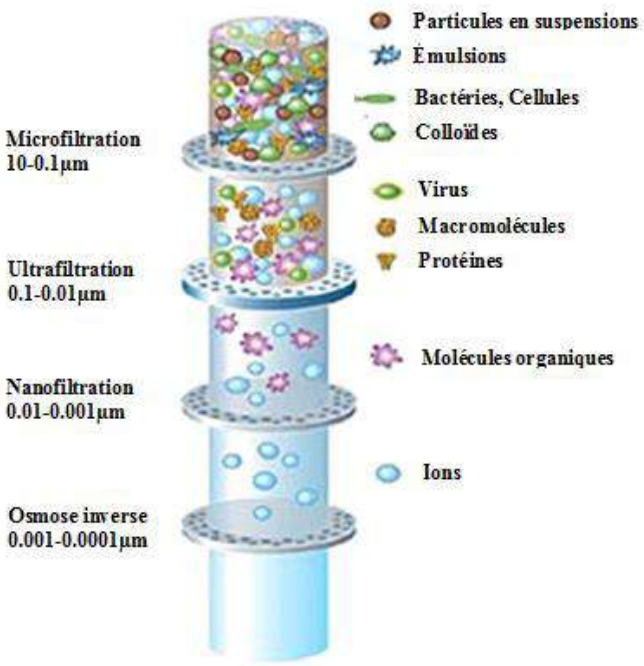


Figure I.7. Gamme de rétention de différentes particules par les procédés membranaires (Metaiche 2014).

I.2.3. Osmose inverse et nanofiltration

L'**osmose inverse** est un procédé de filtration tangentielle qui permet l'extraction d'un solvant, le plus souvent l'eau par perméation sélective à travers une membrane dense sous l'action d'un gradient de pression.

- Elle s'oppose au phénomène naturel d'osmose qui tend à transférer le solvant d'une solution diluée vers une solution concentrée mise en contact par une membrane sélective sous l'action du gradient de concentration.
- Lorsqu'une pression est appliquée sur le compartiment le plus concentré, le flux de solvant diminue et tend vers une valeur nulle pour une pression égale à la pression osmotique de la solution.
- Lorsque la pression appliquée est supérieure à cette pression osmotique, le flux s'inverse : c'est le phénomène d'osmose inverse (Figures I.8) (Ammi 2016).

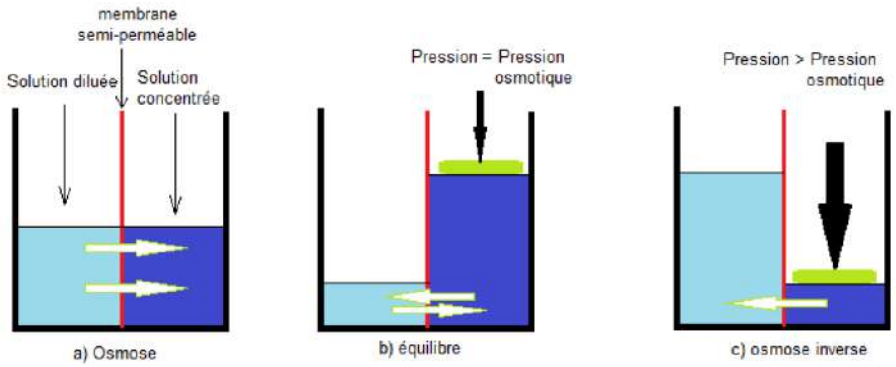


Figure I.8. Principe des phénomènes d'osmose inverse (Triki 2014).

Pour les solutions diluées, considérées comme thermodynamiquement idéales, la pression osmotique est calculée à partir de l'équation de Van't Hoff (Sagne 2008) :

$$\pi = CRT \quad \text{(I.1)}$$

π : pression osmotique de la solution (Pa)

C : concentration de la solution ($mol.m^{-3}$)

R : constante des gaz parfait (8.31 J $mol^{-1}K^{-1}$)

T : température (K).

La **nanofiltration** est une technique intermédiaire entre l'osmose inverse et l'ultrafiltration, elle permet de retenir des molécules de faible poids moléculaire (MW ≥300 g.mol^{-1}) et des ions divalents, sans modifier son affinité vis à vis les ions monovalents. Les caractéristiques d'utilisation sont les suivantes (Hong et Elimelech 1997) :

- Pression de travail 10 à 40 bars,
- Diamètre des pores de la membrane 1 nm,
- Débit volumique 50 à 100l.$h^{-1}m^{-2}$.

Les membranes de nanofiltration sont en fait des membranes classiques greffées par des groupements hydrophiles chargés négativement. Elles ont été testées avec succès en traitement des eaux pour éliminer à la fois les ions divalents (Ca^{+2},Mg^{+2}) et certains pesticides organochlorés. Les sels ionisés multivalents (calcium, magnésium, aluminium, sulfates…) et les composés organiques non ionisés de masse molaire supérieure à environ 250 g/mol sont, par contre, fortement retenus. Ces membranes ont des seuils de coupure entre 300 et 1000 g/mol (Hong et Elimelech 1997).

I.2.4. Sélectivité d'une membrane

La sélectivité d'une membrane pour une substance donnée dépend de sa nature et de sa structure, de l'environnement chimique à proximité de la membrane et des propriétés de la substance à séparer. La sélectivité s'exprime par un taux de rétention *Tr* ou par un facteur de sélectivité (Altalyan et al. 2016) :

$$T_r = \frac{C_{SA}-C_{SP}}{C_{SA}} = 1 - \frac{C_{SP}}{C_{SA}} \quad \text{(I.2)}$$

Où :
C_{SP}: concentration d'un soluté dans le perméat (kg/m^3) ;

C_{SA}: concentration d'un soluté dans l'alimentation (kg/m^3).

La sélectivité d'une membrane est aussi bien liée au seuil de coupure définit comme la limite de la masse moléculaire à retenir exprimer en molecular weight cutoff (MWCO) (Chabane 2015).

I.2.5. Influence des paramètres opératoires

I.2.5.1. Pression

La pression de fonctionnement (Ptm) doit être supérieure à la pression osmotique. L'étude menée par Sridharet al. 2002 a montrée que pour une pression appliquée inférieure à 8.9 bar, aucun perméat n'est produit lors du traitement d'un effluent issu d'industrie d'huile végétale. Ainsi, la pression efficace (Ptm-Π) correspond à la pression qui conduit réellement à la production du perméat. Plus cette pression est élevée, plus le flux du perméat est important (Lin et Yang 2004). De même, le taux de rétention augmente lorsque la pression augmente (Lin et Yang 2004).

D'après Lin et al. 2004, la DCO du perméat passe de 6.7 à 3.1 mg.l^{-1}pour des pressions de 2.1 et 6.3 bars avec un taux de rétention augmente de 96.9% à 98.6%. Sridhar et al. 2002 ont montré que cette augmentation du taux de rétention est due à l'augmentation de la pression transmembranaire ainsi que des flux du solvant et du flux du soluté. La concentration du soluté dans le perméat diminue. Par conséquent, le taux de rétention augmente.

I.2.5.2. Température

L'augmentation de la température engendre une diminution de la viscosité des solutions, et par conséquent une augmentation de la densité de flux du perméat, d'environ 3% par degré Celsius. Le taux de rétention diminue, la diffusivité de l'eau et des solutés augmente avec la température (Girard et Fukumoto 2000). Pour une pression donnée, le compactage de la membrane est d'autant

plus important que la température est élevée. Il est donc important de vérifier que la membrane et le module aient une tenue suffisante à la température de fonctionnement choisie en général, la température maximale recommandée est de l'ordre de 50°C.

I.2.5.3. Taux de récupération "Taux de conversion"

Le taux de récupération "y" d'un système d'OI/NF est le rapport entre le débit de perméat "Q_P" et le débit d'alimentation "Q_A" (Chabane 2015) :

$$y = \frac{Q_P}{Q_A} \tag{I.3}$$

L'augmentation du débit d'alimentation engendre une augmentation de la vitesse de circulation le long de la membrane. Plus elle est importante, plus la concentration aux abords de la membrane sont moins importantes aussi que la densité de flux du perméat au cours du temps. Plus le taux de récupération est élevé, plus le flux diminue, aussi la concentration côté retentât augmente (Sagne 2008).

I.3. Interactions et propriétés de séparation

Dans le cas des molécules organiques, les propriétés de séparation ne sont pas facilement prédictibles. En effet, il a été montré qu'elles résultent de trois interactions qui affectent leur rétention lors d'OI/NF et peuvent être signalées comme suit : encombrements stériques (effet du tamisage) (Altalyan et al. 2016), répulsion électrostatique (effet de charge) et des interactions hydrophobes/adsorption (Sagne 2008).

Rappelons tout d'abord la structure du polyamide, polymère majoritairement utilisé comme couche active des membranes d'osmose inverse et de nanofiltration (Figure I.9).

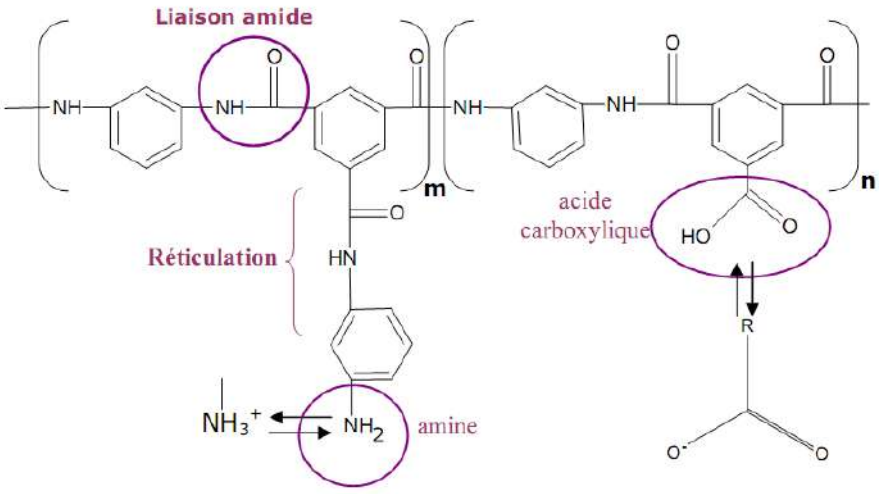


Figure I.9. La molécule de polyamide et ses équilibres acido-basiques (Sagne 2008).

De par cette structure, plusieurs types d'interactions peuvent se produire entre les molécules en solution et la surface de la membrane (Sagne 2008) :

- Les fonctions acides carboxyliques et amines libres (non engagées dans une liaison de réticulation) sont susceptibles de s'ioniser, arborant ainsi respectivement une charge négative et positive et pouvant alors s'engager dans des liaisons électrostatiques fortes.

Elles peuvent également participer à des liaisons hydrogènes.

- Les noyaux aromatiques peuvent développer des interactions π-π avec le nuage électronique d'une autre molécule aromatique, pour former un nuage d'électrons global plus stable.
- Les liaisons amides peuvent développer des interactions de Lifshitz-van der Waals plus faibles, par les doublets d'électrons libres des atomes d'azote et d'oxygène.

I.3.1. Encombrement stériques (effet du tamisage)

La taille moléculaire des molécules organiques et le seuil de coupure (Sc ou MWCO) de membrane ont confirmé l'importance des paramètres dimensionnels pour la rétention des molécules sur les membranes d'OI/NF (Yangali-Quintanilla et al. 2010a ; Dolar et al. 2011). Plusieurs auteurs ont sélectionné les descripteurs de taille moléculaire sachant que la rétention est liée aux mécanismes d'exclusion de taille entre membranes-solutés (Yangali-Quintanilla et al. 2009b-2010b).

I.3.1.1. État physique de la membrane

Les propriétés physiques de la membrane comme le nombre de pores, la forme ainsi que la rugosité de la surface jouent un grand rôle dans la rétention des solutés.

La porosité est exprimée par la taille, la distribution de la taille et le nombre de pores effectifs dans la couche active. Le terme "pore" pour les membranes d'OI/NF correspond à un espace vide dans le polymère par lequel le transport du fluide peut avoir lieu sous l'action d'une force motrice. Dans une membrane dense, les pores sont des vides de moins de 0.5 nm de diamètre qui sont formés naturellement dans les polymères solides. Ces pores, ou défauts de densité, sont causés par des irrégularités de l'entremêlement des chaînes de polymères. Ils peuvent aussi être créés par les contractions locales des polymères cristallisables lors de leur cristallisation. Ces pores peuvent être circulaires ou non, fermés ou ouverts et forment un réseau continu d'interconnections. Ils sont susceptibles d'accueillir de petites molécules qui au cours de leur transport à travers ce pore interagissent fortement avec ses parois (Kosuticet al.2000).

Les indicateurs de porosité de membrane peuvent être le rejet du sel ou le seuil de coupure, attribués par le fabricant en cours de l'expérience.

A. Seuil de coupure ou Seuil de rétention "Sc ou MWCO"

Les membranes d'OI/NF sont caractérisées par leur seuil de coupure. Il s'agit de la masse molaire de la plus petite molécule arrêtée à 90% (ou 95% selon le fournisseur) par la membrane dans des conditions opératoires données. Il est exprimé généralement en Dalton (unité équivalente à des grammes par mole). Le seuil de coupure renseigne l'utilisateur sur la plage d'utilisation de la membrane, mais sa valeur exacte ne doit être considérée qu'à titre indicatif. En effet, les molécules ont des géométries variées et il est possible que des espèces de masse molaire supérieure au seuil de coupure soient retenues par la membrane, de même que des espèces de masse moléculaire légèrement inférieure la traversent. Il est donc important de retenir que le seuil de coupure est défini par rapport à un soluté donné. On peut ainsi caractériser une membrane d'OI/NF, pour une séparation donnée, dans des conditions fixées, par une courbe de variation du taux de rejet d'un soluté en fonction de sa masse moléculaire (Figure I.10) (Lanteriet al.2009).

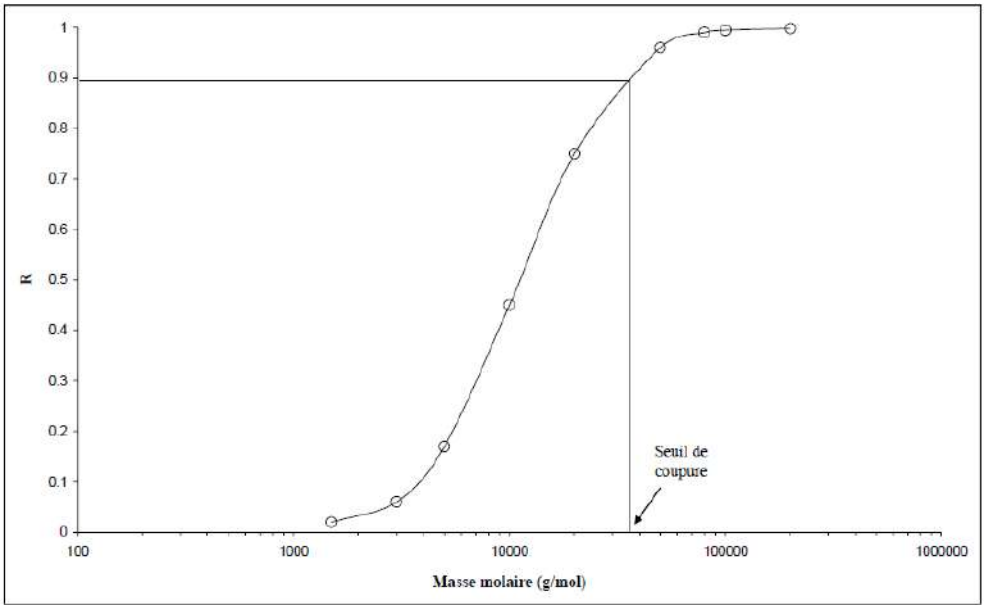


Figure I.10. Illustration du principe de détermination du seuil de coupure d'une membrane, la rétention en fonction de la masse molaire (défini ici pour un taux de rejet de 0.90) (Lanteri et al.2009).

Plus le **Sc** d'une membrane est faible, plus elle peut retenir de petites molécules ou des colloïdes. Le **Sc** est relié principalement à la taille des pores de la membrane, il est influencé par (Ammi 2016) :

- ✓ La forme de la molécule à filtrer ;
- ✓ Sa charge ;
- ✓ Le pH ;

- ✓ La pression d'opération ;
- ✓ Le flux de perméation ;
- ✓ La charge de la membrane.

B. Rejets des sels

- **Rejet du sel de chlorure de sodium SR(NaCl)**

Le premier paramètre donnant une indication sur la "porosité" de la membrane est son taux de rétention en chlorure de sodium (NaCl), mesuré par le fabricant dans des conditions standards, qui lui sont spécifiques. Il a été bien corrélé au taux de rétention de certains pesticides, de polysaccharides et d'alcools (Bellona et al.2004). Étude d'Agensonet al.2003sur 36 composés organiques semi-volatils et volatils d'intérêt environnemental, de masse molaire (MW) comprise entre 70 et 300 g/mol a montré qu'en général les taux de rétention les plus forts sont obtenus pour les membranes ayant un taux de rétention en NaCl le plus élevé.

- **Rejet du sel de sulfate de magnésium SR($MgSO_4$)**

Le rejet du sel de sulfate de magnésium incorpore les encombrements stériques et les effets de répulsion électrostatique qui sont liés à la structure de la membrane et les conditions d'exploitation. L'utilisation du seuil de coupure était acceptable à des fins de modélisation, toute fois les membranes d'OI/NF avec une large gamme de MWCO est difficile à estimer la rétention de contaminants. Le rejet du sel de sulfate de magnésium peut être plus approprié (Yangali-Quintanilla et al.2009b et 2010b).

Kosutic et al. 2000 ont étudié la porosité des membranes commerciales en polyamides et ont rapporté que les différences de performances sont principalement dues par des différences de porosité. Par ailleurs, ils ont indiqué que la variation de la pression du travail a une influence sur cette porosité et donc sur les performances des membranes. Pour certaines membranes, l'augmentation de la pression conduit à un effet de compaction qui fait diminuer la densité de flux du perméat et la taille des pores tout en augmentant leur nombre.

Le pH peut avoir une influence sur la porosité : à pH élevé, la répulsion entre les groupements acides au sein de la membrane peut entrainer une augmentation de la taille des pores (Ammi 2016).

I.3.1.2. Taille moléculaire

La taille moléculaire est un paramètre principal qui détermine la rétention de molécules organiques avec les membranes d'OI/NF. Plusieurs paramètres peuvent être utilisés pour représenter la taille moléculaire : le poids, la longueur, la largeur, la profondeur et la largeur équivalente qui sont

calculés à partir de la structure moléculaire. (Van der Bruggen et al. 1998) ont montré qu'il existe une bonne corrélation entre la rétention et les différents paramètres de taille (Figure I.11).

A. Poids moléculaire "MW"

Le poids moléculaire est un paramètre facilement disponible, il fournit seulement une estimation approximative de la taille moléculaire. Il est utilisé pour avoir une première idée sur la taille relative des molécules étudiées. Il ne peut pas être utilisé pour la modélisation, car il ne représente pas la géométrie de la molécule. Pour les molécules non chargées, peu polaires, la masse molaire peut fournir une première information sur la rétention avec de bonnes corrélations (Sadmani 2014b). Cependant, il ne renseigne pas sur l'encombrement stérique de la molécule. L'évaluation de la taille et de la géométrie de la molécule couplée à la taille de pores peut être un meilleur descripteur que la masse molaire. En effet, des pesticides ayant une section et non une masse molaire plus faible sont moins bien retenus. Le poids moléculaire est considéré comme une variable de descripteur moins significatif pour la modélisation de la rétention (Yangali-Quintanilla et al. 2008).

B. Longueur moléculaire "*Length*"

La longueur moléculaire (*Length*) a été définie comme la distance entre les deux atomes les plus éloignés de la molécule (en tenant compte des forces des Van der Waals). La ligne droite reliant les deux atomes les plus éloignés a été définie comme la L-axe (Santos et al. 2006).

C. Largueur moléculaire "*Width*"

La molécule est projetée sur le plan perpendiculaire à l'axe L. Le rectangle de surface minimale entourant la projection de la molécule est sélectionné. L'aire du rectangle (*S*) est calculée et la largeur moléculaire (*Width*) est définie comme étant la moitié de la racine carrée de S. La racine carrée de S correspond à la longueur du côté du carré dont la superficie est égale à la surface du rectangle (Kiso et al. 2010) :

$$Width = \frac{1}{2}\sqrt{S_{min}} \qquad (I.4)$$

D. Profondeur moléculaire "*Depth*"

La largeur moléculaire et la profondeur moléculaire (la largeur supérieure à la profondeur) sont mesurées par la projection de la molécule sur le plan perpendiculaire à l'axe de la longueur (Santos et al. 2006) :

$$Depth = \frac{1}{2}\sqrt{S_{max}} \quad \text{(I.5)}$$

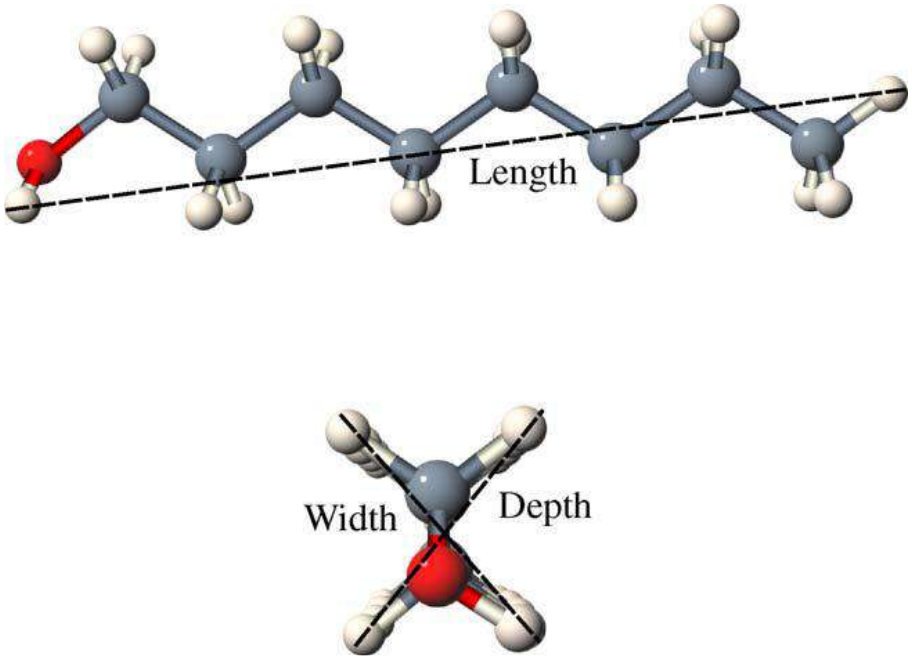


Figure I.11. Calcul des dimensions de l'octanol (*Length*, *Width*et *Depth*) de la structure 3D (Santos et al. 2006).

E. Largeur équivalente moléculaire "*EqWidth*"

La largeur moléculaire équivalente est définie comme étant la moyenne géométrique de la largeur et de la profondeur (Linares et al. 2011).

On peut calculer la largeur équivalente par la formule suivante (Linares et al. 2011) :

$$EqWidth = \sqrt{Width * Depth} \quad \text{(I.6)}$$

I.3.2. Répulsions électrostatiques (effet de charge)

Les interactions électrostatiques ont souvent été rapportées comme participantes de manière importante au mécanisme de rétention. Elles s'établissent entre deux entités chargées (molécules organiques et membranes) (Van derBruggenet al. 1998).

I.3.2.1. Charge de la surface

En général les membranes d'OI/NF thin-film composite (TFC) portent une charge négative (Norberg et al. 2007) qui permet de minimiser l'adsorption des composés colmatant chargés

négativement et d'augmenter la rétention des sels ionisés dans l'eau. La charge de la membrane est généralement quantifiée par mesure du potentiel zêta (ζ) (Ammi 2016).

Le pH de l'alimentation a une influence sur la charge de la membrane, il joue sur la pronation des groupes fonctionnels. A pH élevé, la membrane est chargée négativement du fait de la dissociation du groupement carboxylique et à pH faible elle est chargée positivement du fait de l'ionisation du groupement amines. Le pH pour lequel le potentiel zêta est nul est appelé point isoélectrique ou pH d'ionisation de la membrane (Desmukh et Childress 2001).

I.3.2.2. Charge du soluté

Comme nous venons de voir, les membranes d'osmose inverse et de nanofiltration présentent une charge généralement négative. Il est donc clair qu'un soluté chargé négativement sera mieux retenu, par répulsion électrostatique, qu'un soluté neutre. Le respect de l'électro neutralité implique que les composés chargés positivement le seront également (Shim et al. 2002). Les molécules ionisables sont donc en général mieux retenues sous leur forme ionisée. Cet effet est d'autant plus marqué que la molécule est petite, car sa densité de charge est alors plus élevée.

Pour des molécules de masses molaires proches, celles de même signe que la membrane sont mieux retenues que les neutres, elles-mêmes mieux retenues que celles de charges opposées à celles de la membrane. Pour ces dernières, l'attraction électrostatique favorise leur passage. Pour les plus grosses molécules les effets stériques restent prédominants (Van der Bruggen et al. 1999).

I.3.2.3. Influence du pH

Comme nous l'avons vu précédemment, la charge des membranes dépend du pH de la solution avec laquelle elles sont en contact. Suite à cela, il peut se créer des interactions électrostatiques entre la membrane et les solutés chargés, ionisés ou non, présents en solution. Le positionnement de ceux-ci près de la surface modifie alors la charge apparente de la membrane (Sagne 2008). Par ailleurs, si le pH n'a pas d'influence sur le taux de rétention des molécules non chargées (Agenson et al. 2003), alors il joue un grand rôle sur celui des molécules ionisables puisqu'il gouverne leur ionisation. En effet, la charge des espèces acido-basiques faibles varie avec le pH.

I.3.3. Interaction hydrophobes/adsorption

L'adsorption est fortement liée à l'hydrophobicité des molécules, à leur capacité à former des liaisons hydrogènes et à leur polarité.

I.3.3.1. Hydrophobicité de la membrane

La mesure de l'angle de contactα que forme l'eau sur une membrane, permet d'avoir une idée de l'hydrophobicité de cette membrane. Cette dernière est supposée gouverner la mouillabilité de la membrane en particulier par l'eau et donc avoir une influence sur ses performances (Ardhaoui et al. 2003). Il existe trois méthodes de mesure de l'angle de contact : la bulle captive, la technique de la lame de Wilhelmy et la goutte déposée (Gumi et al. 2003). Par l'utilisation de la méthode de la goutte déposée sur une surface solide (Figure I.12), deux cas peuvent être présentés (Sagne 2008) :

- ✓ α est supérieur à 90°, la membrane est dite hydrophobe ;
- ✓ α est inférieur à 90°, la membrane est dite hydrophile.

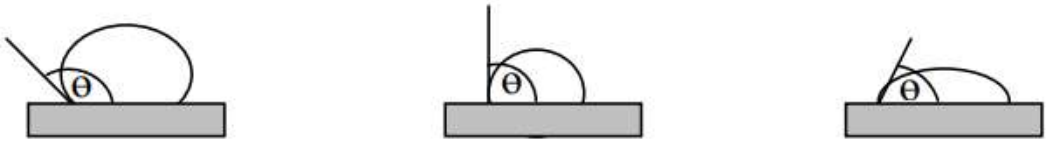


Figure I.12. Angle de contact des gouttelettes de liquide sur un matériau solide (Yangali Quintanilla 2010a).

La plupart des membranes d'OI/NF sont hydrophiles possédant des valeurs de l'angle de contact généralement inférieur à 90°.

I.3.3.2. Hydrophobicité des solutés

Le caractère hydrophobe ou hydrophile des solutés va donc jouer un rôle sur leur rétention. Il est évalué par **le logarithme du coefficient de partage octanol-eau** ($Logk_{ow}$), plus souvent exprimé sous forme logarithmique ($logk_{ow}$, *logP*). S'il est supérieur à 2, la molécule est hydrophobe (HB) ; s'il est inférieur, la molécule est hydrophile (HL) (Yangali-Quintanilla et al.2009b).

Le logarithme du coefficient de partagen-octanol/eau, *LogP*, a été utilisé comme un indice hydrophobe est défini comme suit (Équation I.7) (Taheran et al. 2016):

$$Logk_{ow} = \log(\frac{C_0}{C_W}) \qquad \text{(I.7)}$$

Où C_0et $C_W$sont lesconcentrationsdes molécules dans les couchesn-octanol et l'eau (Taheran et al. 2016).

Pour les solutés ionisables, le caractère hydrophobe est fonction du pourcentage des chargé contre les espèces non chargées, il dépend donc du pH. Par conséquent, un coefficient de partage

apparent à un pH donné est calculé pour les solutés ionisables. Le coefficient de distribution octanol-eau dépend du pH ,*D* ,est le rapport entre les concentrations d'équilibres de toutes les espèces(ionisé et non) d'une molécule dans l'octanol à la même espèce dans la phase aqueuse. Le logarithme du coefficient de distribution peut être calculé à partir du *PKa* et $Logk_{ow}$du soluté (Équations I.8 et I.9) (Verliefde et al. 2008).

Pour les acides, (Verliefdeet al. 2008)

$$\mathrm{LogD_{(pH)}} = \mathrm{logk_{ow}} - \log(1 + 10^{(\mathrm{pH-pKa})}) \qquad \text{(I.8)}$$

Pour les basses, (Verliefdeet al. 2008)**:**

$$\mathrm{LogD_{(pH)}} = \mathrm{logk_{ow}} - \log(1 + 10^{(\mathrm{pKa-pH})}) \qquad \text{(I.9)}$$

On parle également des molécules polaires/apolaires. Les molécules hydrophobes sont en général apolaires (Sagne 2008). Le dipôle est dirigé vers la membrane de telle sorte que le côté du dipôle de charge opposée est plus proche de la membrane. Cette orientation n'est pas statique, afin elle doit être considérée comme une tendance statistique des molécules en mouvement rapide d'avoir cette orientation préférentielle. Le dipôle dirigé est donc attiré dans les pores et entre ainsi plus facilement dans la structure de la membrane (Van der Bruggen et al. 1998).

Les molécules ayant un moment dipolaire élevé ont des rétentions inférieures par rapport à des molécules non polaires. Cet effet, est indépendant du signe de la charge de la membrane (positif ou négatif) (Van der Bruggen et al. 1999).

I.4. Applications des procédés d'OI et NF

Les différentes applications envisagées pour les procédés OI et NF sont résumées dans le tableau I.1.

Tableau I.1. Applications des procédés l'OI et NF.

Application	Références
Industries diverses : ▪ Récupération des métaux ▪ Production de l'eau ▪ Traitement de surface ▪ Effluents industriels	(Metaiche 2014).
Industrie agro-alimentaire : ▪ Filière œnologique ▪ Lactosérum ▪ Recyclage des protéines ▪ Effluents agro-alimentaires ▪ Industrie pharmaceutique et biotechnologique : Sérums, principes actifs, antibiotique.	(Sadmani et al. 2014a et 2014b).
Eau potable, dessalement, saumure : ▪ Élimination nano-polluants organiques, élimination des particules, adoucissement, déminéralisation.	(Donghan et al. 2016).

I.5. Avantages et inconvénients des procédés d'OI et NF

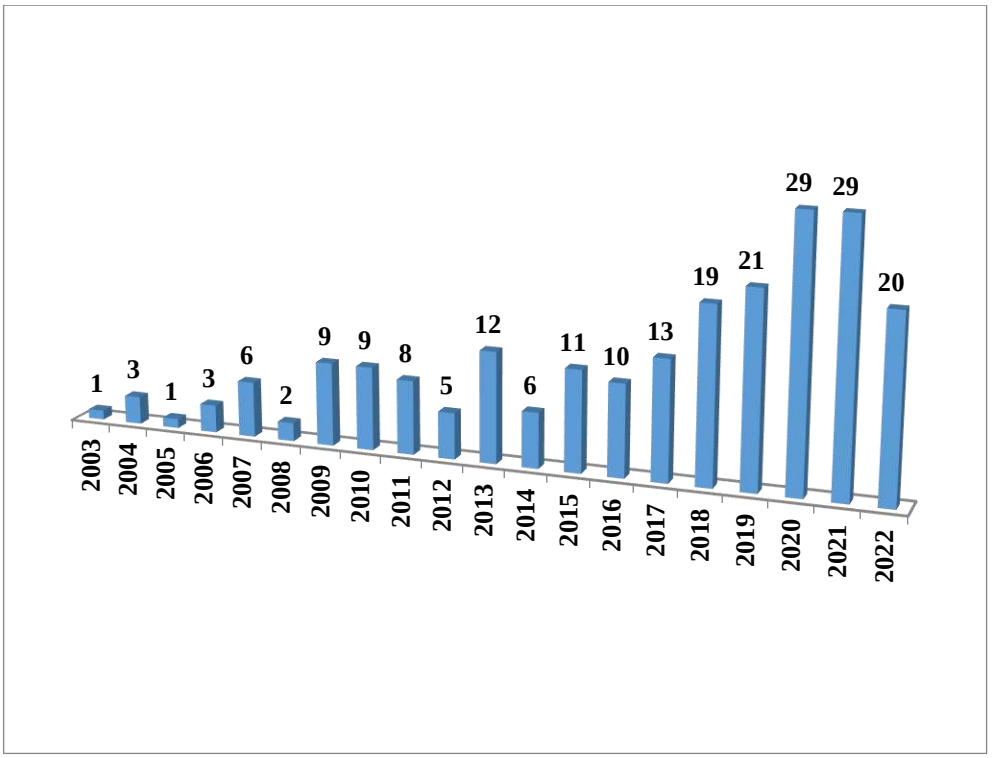
Les avantages et les inconvénients des procédés d'OI et NF sont regroupés dans le tableau I.2.

Tableau I.2. Avantages et inconvénients de procédés d'OI et NF (Aptel et Buckley 1996).

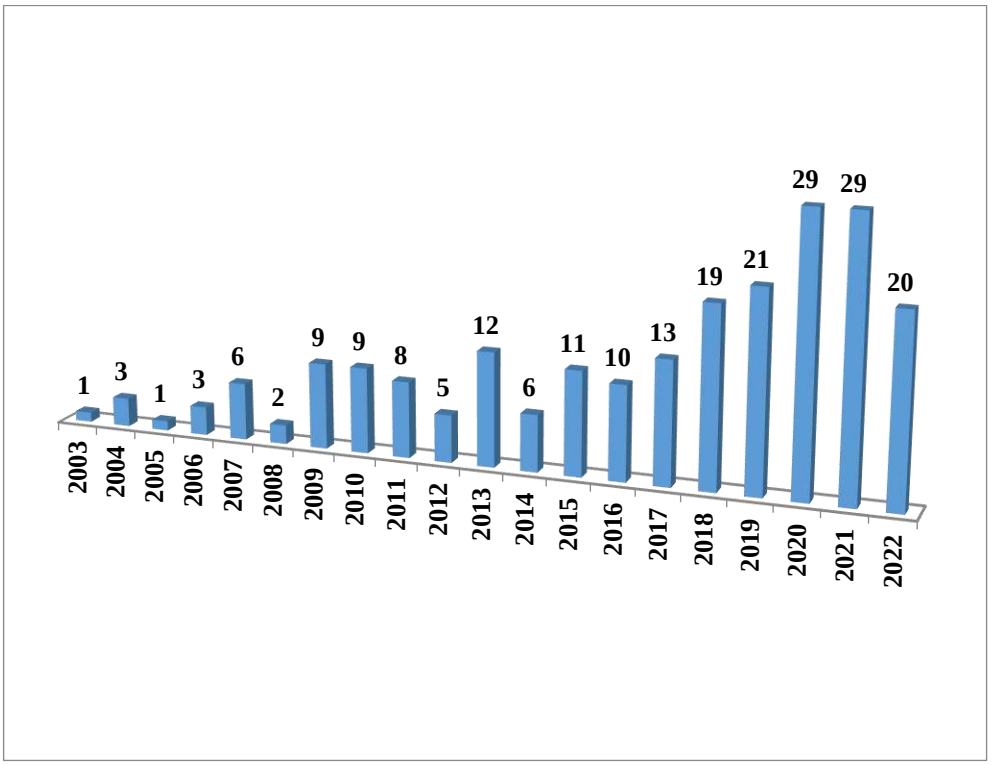
Avantages	Inconvénients
▪ Un taux de conversion élevé (supérieur à 55%). ▪ Une consommation spécifique d'énergie basse (3 à 5 Kwh/m^3). ▪ Un coût d'investissement relativement bas. ▪ Une durée de mise en œuvre réduite. ▪ Opération à température ambiante ce qui évite la dégradation de molécules fragiles dans le domaine agro-alimentaire. ▪ Pas d'intervention de réactifs chimiques comme des agents d'extraction qui sont des sources de pollution.	▪ Sensibilité à la variation de la qualité d'eau brute et surtout aux pollutions. ▪ Nécessité d'un personnel qualifié. ▪ Exploitation relativement difficile sur tout au niveau du prétraitement. ▪ Un coût de maintenance élevé dû aux frais de renouvellement des membranes d'une durée de vie ne dépassant pas 7 ans à présent. ▪ Baisse de la perméabilité et modification de la sélectivité en cas de colmatage des membranes.

I.6. Revue des travaux de la rétention des PhACs lors d'OI/NF

De nombreux travaux réalisés au niveau international mettent en évidence la rétention des produits pharmaceutiques par d'OI/NF. La base de données consultée est Science Direct. La figure I.13, reporte le nombre de publications de 2003 à 2022 pour les membranes de la nanofiltration et d'osmose inverse.



(a)



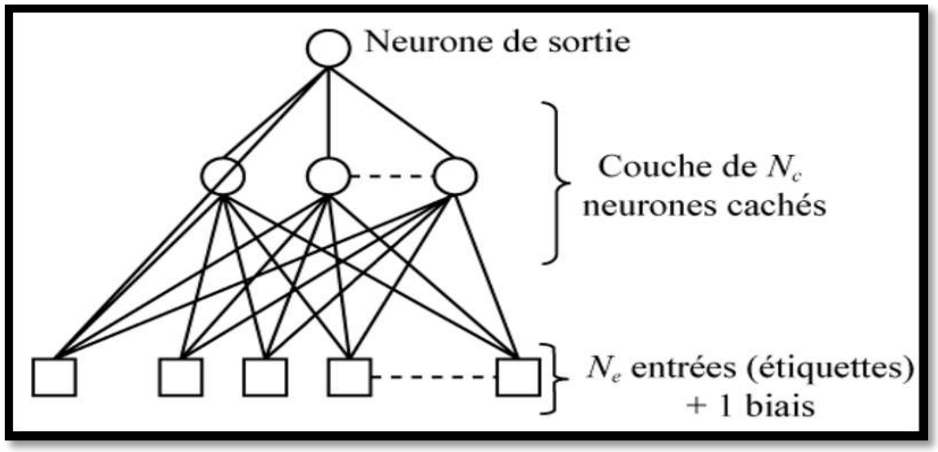
(b)

Figure I.13. Nombre de publications traitant la rétention des phACs par (a) : nanofiltration (b) osmose inverse, (ScienceDirect le 20/05/2022).

I.7. Conclusion

La rétention des molécules organiques par les membranes de la nanofiltration et d'osmose inverse dépend d'un grand nombre de facteurs liés à la fois à la membrane (structure physique, charge, hydrophobicité), à la molécule (taille, structure, charge, polarité ou hydrophobicité) et aux conditions expérimentales (pH, pression, taux de récupération et température). Les interactions mises en place sont à la fois d'ordre stériques, électrostatiques et dues au caractère hydrophobe/hydrophile.

Chapitre II
QSPR-Réseaux de neurones



II.1. Introduction

La chémoinformatique est la discipline dont le but est l'étude et la résolution des problèmes relatifs à la chimie en appliquant des méthodes informatiques. Cette discipline est devenue une véritable interface entre la chimie et l'informatique (Ammi 2016).

Ces dernières années, il y a eu plusieurs tentatives pour faire progresser l'utilisation des relations quantitatives structure-activité (QSAR) comme une approche viable pour développer des modèles pilotés par les données pour décrire la performance des procédés membranaires (Dornier et al. 1995 ; Sahoo et al. 2006 ; Libotean et al. 2009). Peu d'études ont également appliqué les modèles QSAR à la modélisation de la rétention des composés organiques par les membranes d'osmose inverse (OI) et de nanofiltration (NF) (Libotean et al. 2008 ; Yangali-Quintanilla et al. 2008, 2009b et 2010b ; Shahmansouri et al. 2013 ; Sadmani et al. 2014a ; Sadmani et al. 2014b). Les réseaux de neurones ont gagné en popularité au fil des ans dans divers domaines. Dans le domaine des sciences de l'ingénieur, le génie des procédés ne fait pas exception de cette règle.

Ce chapitre a été consacré à une revue bibliographique a été scindé en quatre parties. Dans la première partie, on parlera sur les méthodes d'apprentissage statistiques**.** On passera dans la deuxième partie les relations quantitatives structure-activité "QSAR". La troisième partie a été consacrée à une revue les réseaux de neurones artificiels. Dans la quatrième partie nous présentons aussi, d'une manière générale, les étapes de conception d'un réseau de neurones : le choix des entrées et des sorties, l'élaboration de la base de données et de la structure du réseau…etc. Une conclusion clôturera ce chapitre et sera une synthèse par laquelle on justifiera le choix du thème.

II.2. Méthodes d'apprentissage statistiques

Au cours de la dernière décennie, les algorithmes d'apprentissage statistique ont suscité beaucoup d'intérêt dans le milieu académique et au sein d'entreprises de diverses industries. Ils ont été implantés avec succès pour l'accomplissement de tâches prédictives reliées à des processus statistiques observés pour lesquels on peut identifier plusieurs variables explicatives. Cette étude se concentre sur une classe particulière de ces algorithmes :

- La relation quantitative structure-activité "QSAR";
- Les réseaux de neurones artificiels "RNA".

II.3. Relations quantitatives structure-activité "QSAR"

Depuis le milieu du 19ème siècle, et, à partir de la fin de 20ème siècle, une appréciation que l'activité biologique est liée à la structure chimique surtout aux propriétés physico- chimiques (Wayne Schultz et al .2003).

Les premiers travaux utilisant la méthode (QSAR) ont commencé au début des années 60 avec Hansch d'une part et de Free et Wilson d'autre part, qui ont proposé un modèle mathématique pour corréler l'activité biologique et la structure chimique. Les développements de cette étude (QSAR) sont très anciens, à partir de 1868, lorsqu'Alexander Crum-Brown et Thomas Fraser montrent l'existence d'une relation entre l'activité physiologiques et la structure chimique, suivis, avec Richet qui établit une relation entre la toxicité et les propriétés physicochimiques. De manière indépendante, Meyer et Overton ont décrit une corrélation linéaire entre la lipophilie (coefficient de partage huile-eau) et les effets biologiques (narcotiques) (Hmamouchi et al.2016). Nous donnons ci-dessous, à titre d'exemple, quelques-unes (Hamadache 2014) :

- ✓ Prédire l'activité biologique et les propriétés physico-chimiques.
- ✓ Comprendre et rationaliser les mécanismes d'action dans une série de produits chimiques.
- ✓ Économiser le coût de développement de produits (par exemple dans l'industrie pharmaceutique, pesticides, produits d'hygiène personnelle, …etc.).
- ✓ Réduire ou même remplacer les tests longs et coûteux sur les animaux.

La capacité à prédire une activité biologique est utile dans un certain nombre de situations. Il existe un grand nombre d'applications de ces modèles QSAR au sein de l'industrie, du milieu universitaire et des organismes gouvernementaux chargés de la réglementation. Nous résumons ci-après quelques cas d'utilisations de ces modèles :

- ✓ L'optimisation de l'activité pharmacologique, biocide ou pesticide.
- ✓ La conception rationnelle de nombreux autres produits tels que des agents tensio-actifs, des parfums, des colorants ou des produits chimiques.
- ✓ L'identification des composés dangereux à des stades précoces de développement.
- ✓ La conception de la toxicité et des effets secondaires pour les nouveaux composés.
- ✓ La prédiction de la toxicité pour les humains pour les multiples expositions (délibérée, occasionnelle ou professionnelle) et de l'écotoxicité.

II.3.1. Outils de QSAR

Les modèles QSAR permettent de quantifier la relation entre la structure et l'activité biologique et d'apprécier l'impact de diverses propriétés physico-chimiques sur cette activité. L'élaboration de tels modèles QSAR repose sur l'utilisation de trois outils : une ou plusieurs données biologiques à prédire, des descripteurs moléculaires pour représenter les propriétés des molécules et des méthodes d'apprentissage statistique qu'on introduira dans une partie à part (Hamadache 2014).

II.3.1.1. Activité biologique

La construction du modèle QSAR est très dépendante des données expérimentales faire référence à. Par conséquent, le choix de la base de données est le point clé de son développement. Dans la plupart des cas, les données expérimentales sont issues de la littérature.

En général, les données d'activité à modéliser peuvent être considérées comme un effet propriétés biologiques (activité pharmacologique ou toxique…) ou physico-chimiques (coefficient de partage, point de fusion, etc.). Les effets biologiques impliquent un certain type des interactions entre les composés et les cellules qui produisent des réponses spécifiques (p. ex., perturbation de la fonction cellulaire, inhibition des enzymes, expression différences génétiques, …etc.) (Hmamouchi et al. 2016).

II.3.1.2. Descripteurs moléculaires

Les descripteurs sont généralement des représentations numériques de caractéristiques moléculaires spécifiques. Il existe un grand nombre de descripteurs moléculaires disponibles pour l'analyse QSAR afin de prédire les événements biologiques. En règle générale, les composés dans l'analyse QSAR sont décrits par leurs propriétés physicochimiques ou une certaine forme de propriétés structurelles.

Ils permettent de faire des prédictions sans synthétiser de molécules, ce qui est l'un des buts de la modélisation. Les descripteurs moléculaires peuvent être divisés en deux grandes catégories (Poezevara 2011) : les descripteurs expérimentaux (par exemple, la lipophilie [logP], la polarisabilité, la densité électronique, le potentiel d'ionisation) et les descripteurs moléculaires théoriques. Ceux-ci sont ensuite divisés en trois catégories en fonction de leur dimension.

A. Descripteurs 1D

Ils sont accessibles à partir de la formule brute de la molécule et décrivent des propriétés globales du composé. Il s'agit par exemple de sa composition, c'est-à-dire les atomes qui le constituent, ou de sa masse molaire (Hamadache 2014).

B. Descripteurs 2D

Les descripteurs 2D, sont des propriétés numériques qui peuvent être calculées à partir de la table de connectivité d'une molécule ou d'une représentation planaire (2D) de la structure. Ils sont basés sur les éléments présents, les charges partielles, la nature des liaisons, le nombre de cycles. La mesure la plus couramment utilisée est le coefficient de partage octanol-eau P, normalement utilisé dans sa forme logarithmique (log P) (Hmamouchi 2016).

C. Descripteurs 3D

Les descripteurs de ce type nécessitent la conformation 3D de la molécule ; ils sont évalués et décrire les positions relatives de leurs atomes dans l'espace caractéristiques plus complexes ; par conséquent, leur calcul nécessite souvent de connaître, la géométrie 3D d'une molécule par modélisation moléculaire empirique la plupart de ces descripteurs sont relativement coûteux en calcul, mais fournir plus d'informations et pour modéliser les propriétés ou activités qui reposent sur des structures 3D.

Il existe plusieurs familles de descripteurs 3D (Hamadache2014):

- ✓ **Les descripteurs géométriques** les plus importants sont le volume moléculaire, la surface accessible au solvant, le moment principal d'inertie
- ✓ **Les descripteurs électroniques** permettent de quantifier différents types d'interactions inter et intramoléculaires, de grande influence sur l'activité biologique de molécules. Par exemple, les énergies de la plus haute orbitale moléculaire occupée et de la plus basse vacante sont des descripteurs fréquemment sélectionnés. Par ailleurs, le moment dipolaire, le potentiel d'ionisation, et différentes énergies relatives à la molécule sont d'autres paramètres importants.
- ✓ **Les descripteurs spectroscopiques** : les molécules peuvent être caractérisées par des mesures spectroscopiques, par exemples par leurs fonctions d'onde vibrationnelles. En effet, les vibrations d'une molécule dépendent de la masse des atomes et des forces d'interaction entre ceux-ci ; ces vibrations fournissent donc des informations sur la structure de la molécule

et sur sa conformation. Les spectres infrarouges peuvent être obtenus soit de manière expérimentale, soit par calcul théorique, après recherche de la géométrie optimale de la molécule. Par exemple, les descripteurs de type MoRSE sont calculés à partir d'une simulation du spectre infrarouge.

Par exemple, le programme appelé DRAGON permet de calculer différents groupes de descripteurs : autocorrélations 2D, descripteurs géométriques, descripteurs RDF, descripteurs 3D-MoRSE, descripteurs WHIM, descripteurs GETAWAY, …etc.

D. Descripteurs 4D

Ils correspondent à la mesure des propriétés 3D (potentiel électrostatique, d'hydrophobicité, de liaison hydrogène...) d'une molécule en tout point de l'espace. Ils permettent d'avoir l'information sur la structure de la cible (protéine). On pourra ainsi distinguer les descripteurs 4D qui nécessitent un alignement de la molécule guidé par l'étude des complexes ligand-cible (ou, au moins, par des contraintes visant d'optimiser le recouvrement spatial des champs électriques et stériques des ligands, faute d'information sur le vrai mode de fixation dans la cible) avant d'être calculés. Ces descripteurs sont obtenus par le calcul des champs d'interactions moléculaires (CoMFA, CoMSIA) entre une molécule et une sonde représentée par une autre molécule (eau, amide, ...) (Chtita 2017).

II.3.1.3. Formulation mathématique

Les équations prennent la forme générale suivante :

$$\boldsymbol{QSAR}\begin{cases}[Activité\ biologique] = f[Descripteurs] \\ [Activité\ thérapeutique] = f[Descripteurs]\end{cases}$$

$$\boldsymbol{QSPR}[Propriété] = f\,[Descripteurs]$$

$$\boldsymbol{QSTR}[Toxicité] = f\,[Descripteurs]$$

Dans ce mémoire, l'équation représente la propriété de la rétention des produits pharmaceutique lors des membranes de nanofiltration et d'osmose inverse.

II.4. Réseaux de neurones artificiels

II.4.1. Neurones biologiques

Le cerveau humain est constitué d'un très grand nombre de cellules nerveuses appelées neurones, environs 100 milliards, avec 1000 à 10000 synapses (connexions) par neurone (Chtita 2017). Le

neurone biologique (Figure II.1) est une cellule nerveuse spécialisée dans le traitement de l'information (signaux électriques). Il est constitué de trois composantes principales :

- ✓ **Les dendrites** : fines prolongations du corps cellulaire entourant celui-ci en une sorte de filet qui capte les oscillations et les informations issues d'autres cellules nerveuses et les transmettent au corps cellulaire.
- ✓ **Le corps cellulaire** : qui a pour fonction de recevoir les excitations, les intégrer et les transmettre ou non. Il contient également le noyau qui assure la vie du neurone.
- ✓ **L'axone** : Les axones conduisent les signaux électriques de la sortie d'un neurone vers l'entrée à un autre neurone. Le point de contact entre l'axone d'un neurone et la dendrite d'un autre neurone s'appelle la synapse (Chtita 2017)

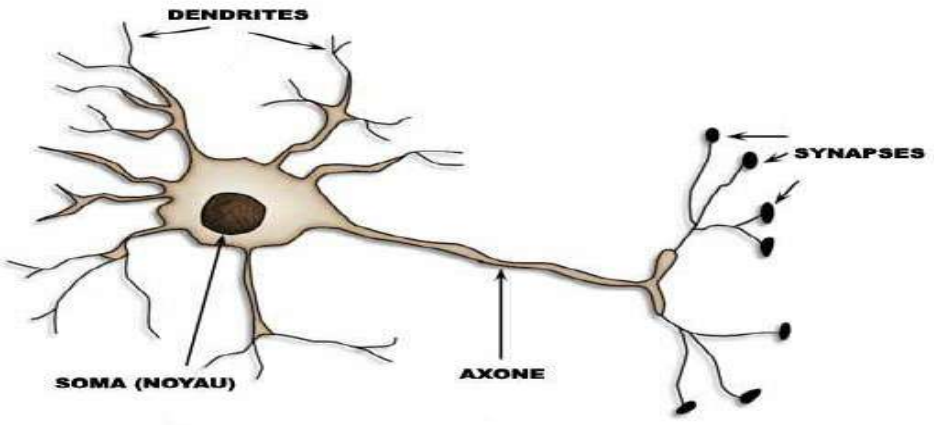


Figure II.1. Le neurone biologique (https://www.researchgate.net/figure/Representation-type-dune-cellule-nerveuse-neurone-3_fig1_275642235 10/04/2022).

Au niveau du neurone se produit une intégration (sommation) des signaux reçus et si cette somme dépasse un certain seuil le neurone émet à son tour un signal électrique vers d'autres neurones. Ce signal peut renforcer ou diminuer l'activité des neurones qui le reçoivent selon que les synapses soient excitatrices ou inhibitrices (Chabaa et al . 2011).

II.4.2. Modèle mathématique

Par analogie avec le neurone biologique, chaque neurone artificiel est un processeur élémentaire. Il reçoit un nombre variable d'entrées en provenance de neurones en amont. À chacune de ces entrées est associé un poids *w*, abréviation du mot poids (weight en anglais), représentatif de la force de la connexion. Chaque processeur élémentaire est doté d'une sortie unique qui se ramifie ensuite pour alimenter un nombre variable de neurones en aval tel qu'illustré sur la figure II.2 (Haykin 1999). À chaque connexion est associé un poids.

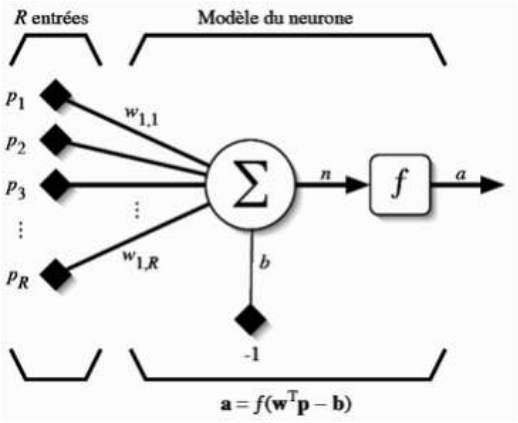


Figure I.2. : Neurone formel.http://informatique.coursgratuits.net/methodes-numeriques/reseaux-de-neurones-formels.php 10/04/2022

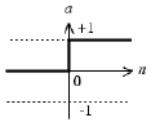
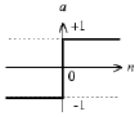
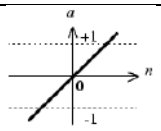
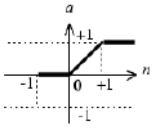
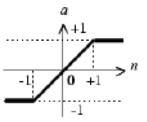
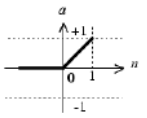
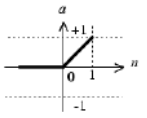
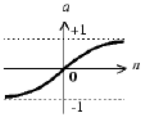
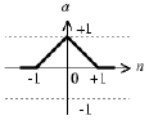
Le comportement du neurone artificiel se compose de deux phases (Haykin 1999) :

1. La première phase représente le prétraitement des données reçues en calculant la somme pondérée des entrées selon l'expression suivante :

$$a_i = \sum_j w_{ij}.x_j \qquad (II.1)$$

2. Dans la deuxième phase, une fonction de transfert calcule la valeur de l'état interne du neurone. C'est cette valeur qui sera transmise aux neurones en aval. Plusieurs possibilités existent pour spécifier la nature de la fonction d'activation du modèle. Les fonctions mathématiques les plus souvent utilisées sont regroupées dans le tableau I.1. À la différence des neurones biologiques dont l'état est binaire, la plupart des fonctions de transfert sont continués, offrant une infinité de valeurs possibles comprises dans l'intervalle [0, +1] ou [-1, +1]. Ainsi, le neurone peut être défini mathématiquement comme étant une fonction algébrique, a priori non linéaire (suivant la fonction d'activation *f*) et bornée, avec des entrées *x* et paramétrée par les poids w_{ij}.

Tableau II.1. Fonctions de transfert usuelles, *a=f(n)* (Si-Moussa 2009).

Nom de la fonction	**Relation entrée / sortie**	**Graphe de la fonction**	**Fonction MATLAB® Correspondante**
Seuil	a= 0 si $n < 0$ a=1 si $n \geq 0$		Hardlim
Seuil symétrique	a= −1 si $n < 0$ a=1 si $n \geq 0$		Hardlims
Linéaire	a= n		Purelin
Linéaire saturée	a= 0 si n <0 a= n si $0 \leq n \leq 1$ a=1 si $n > 1$		Satlin
Linéaire saturée Symétrique	a= −1 si $n < -1$ a= n si $-1 \leq n \leq 1$ a=1 si $n >1$		Satlins
Linéaire positive	a= n si $n \geq 0$ a= 0 si $n < 0$		Poslin
Logistique (Sigmoïde Logarithmique)	$a = \frac{1}{1+e^{-n}}$		Logsig
Tangente hyperbolique	$a = \frac{e^{+n}-e^{-n}}{e^{+n}+e^{-n}}$		Tansig
Triangulaire de base	$a = 1 - \lvert n \rvert, -1 \leq n \leq 1$ a =0 *autrement*		Tribas

II.4.3. Architecture des réseaux de neurones

On distingue deux grandes familles de réseaux de neurones : les réseaux bouclés et les réseaux non bouclés (Figure I.3).

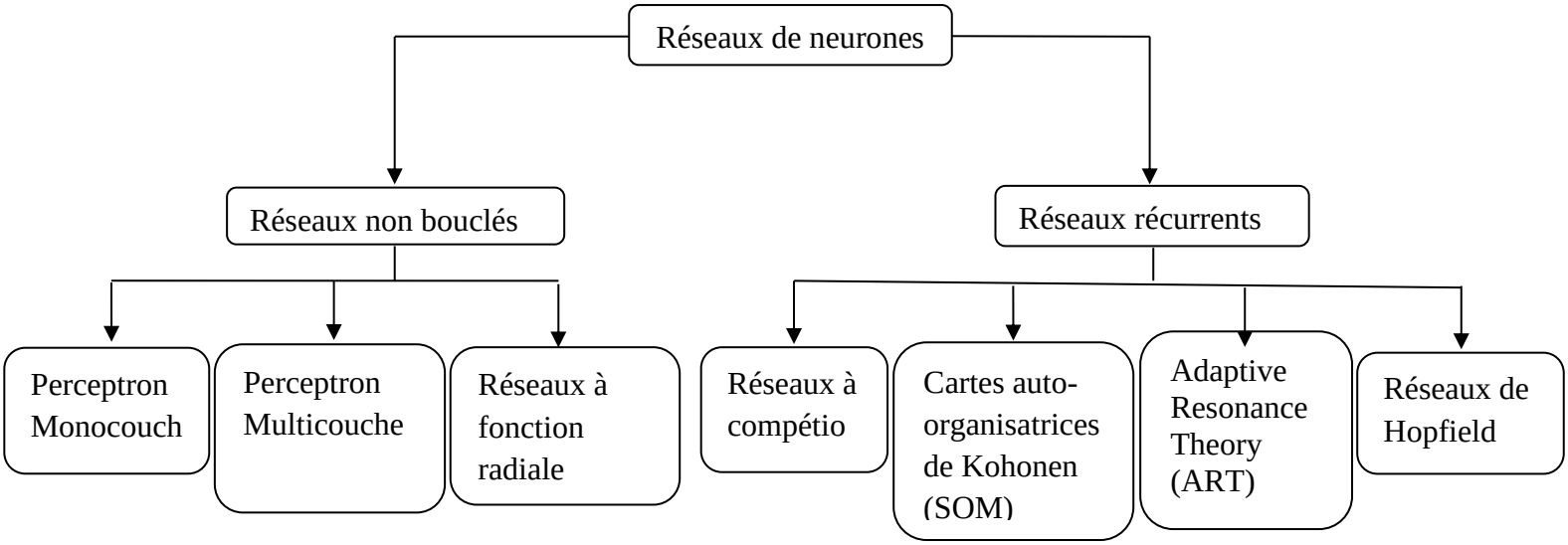


Figure I.3. Réseaux de neurones les plus utilisés.

II.4.3.1. Réseau de neurones non bouclé (en anglais feedforward)

Un réseau de neurones non bouclé est donc représenté graphiquement par un ensemble de neurones connectés entre eux, l'information circulant des entrées vers les sorties sans "retour en arrière" (Bishop 1995). On constate que le graphe d'un réseau non bouclé est acyclique. En effet, si on se déplace dans ce type de réseau à partir d'un neurone quelconque en suivant les connexions, on ne peut pas revenir au neurone de départ. Les neurones qui effectuent le dernier calcul de la composition de fonctions sont les neurones de sortie. Ceux qui effectuent des calculs intermédiaires sont les neurones cachés (Figure II.4).

Les réseaux non bouclés à couche sont structurés tel que les neurones qui appartiennent à une même couche ne soient pas connectés entre eux, chacune des couches recevant des signaux de la couche précédente et transmettant le résultat de ses traitements à la couche suivante. Les deux couches extrêmes correspondent à la couche d'entrée qui reçoit ses entrées du milieu extérieur d'une part et à la couche de sortie qui fournit le résultat des traitements effectués d'autre part. Les couches intermédiaires sont appelées couches cachées, leur nombre est variable (Chamekh 2007).

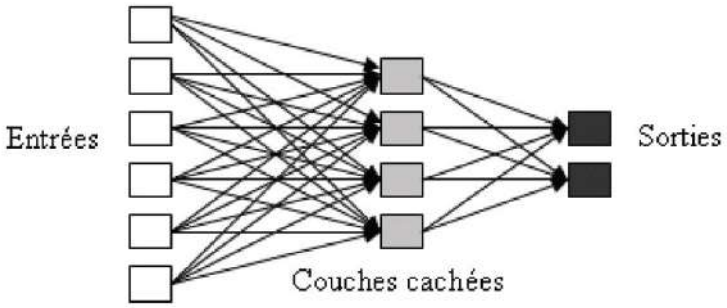


Figure II.4. Structure d'un réseau de neurones non bouclé (Chamekh 2007).

II.4.3.2. Réseau de neurones bouclé (ou récurrent)

Un réseau de neurones bouclé est schématisé par un graphe de connexions qui est cyclique. Lorsqu'on se déplace dans le réseau en suivant le sens des connexions, il est possible de trouver au moins un chemin qui revient à son point de départ (un tel chemin est désigné sous le terme de "cycle"). Ainsi, un retard entier multiple de l'unité de temps choisie est attaché à chaque connexion d'un réseau de neurones bouclé (ou à chaque arrête de son graphe). Une grandeur à un instant donné ne peut pas être fonction de sa propre valeur au même instant. Tout cycle du graphe du réseau doit avoir un retard non nul (Chamekh 2007).

La figure II.5 représente un exemple de réseau de neurones bouclé. Les chiffres dans les carrés indiquent le retard attaché à chaque connexion, exprimé en multiple de l'unité de temps. Ce réseau contient un cycle, qui part du neurone 3 et revient à celui-ci en passant par le neurone 4. La connexion de 4 vers 3 ayant un retard d'une unité de temps.

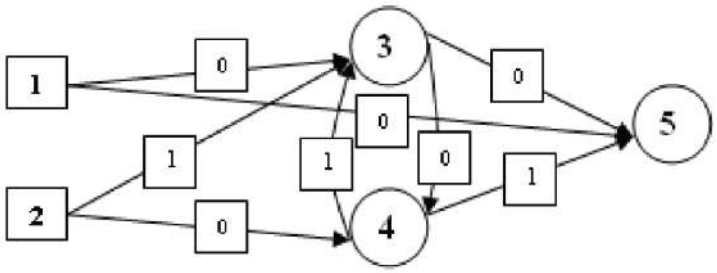


Figure II.5. Un réseau de neurones bouclé à deux entrées (Chamekh 2007).

II.4.4. Modes d'apprentissage

Nous distinguons trois modes d'apprentissage : l'apprentissage supervisé, non supervisé et renforcé.

II.4.4.1. Apprentissage supervisé

Dans ce mode, un professeur qui connaît parfaitement la sortie désirée ou correcte guide le réseau en lui apprenant à chaque étape le bon résultat. Donc l'apprentissage ici consiste à comparer le résultat obtenu avec le résultat désiré, puis à ajuster les poids des connexions pour minimiser la différence entre les deux (Samelides et al. 1991).

II.4.4.2. Apprentissage non supervisé

Dans l'apprentissage non supervisé, le réseau modifie ses paramètres en tenant compte seulement des informations locales. Ces méthodes n'ont pas besoins de sorties désirées préétablies. Les réseaux utilisant cette technique sont appelés réseaux à dynamique autonome et sont considérés comme des détecteurs de régularité, car le réseau apprend en détectant les régularités dans la structure des motifs d'entrée et produit la sortie la plus satisfaisante (Davalo et al .1993).

II.4.4.3. Apprentissage renforcé

Il est utilisé quand une information en retour sur la qualité de la performance est fournie, mais que la conduite souhaitée du réseau n'est pas complètement spécifiée par un professeur. Donc l'apprentissage est moins dirigé que l'apprentissage supervisé. Contrairement à l'apprentissage non supervisé où aucun signal de retour n'est donné, le réseau à apprentissage renforcé peut utiliser le signal de renforcement pour trouver les poids les plus désirables quand c'est nécessaire (Samelides et al. 1991).

II.5. Réseaux de neurones les plus utilisés

La figure I.4 représente les réseaux de neurones les plus couramment utilisés dans la littérature. à ceci Ensuite, nous décrirons chacun d'eux.

II.5.1. Perceptron monocouche (perceptron simple)

Un perceptron simple ou perceptron monocouche est un réseau de neurones développé par Rosenblatt (Rosenblatt1958). Ce réseau est constitué d'un seul neurone, Il peut aussi être utilisé pour faire de la classification et pour résoudre des opérations logiques simples, il suit généralement un apprentissage supervisé (Talbi 2011).

II.5.2. Perceptron multicouche (en anglais Multi Layer Perceptron)

Le perceptron multicouche(PMC) (McClelland et al. 1989)est un ensemble de perceptrons répartis dans des couches successives (Figure II.6), à savoir : une couche d'entrée, un nombre des couches cachées et une couche de sortie. Il peut résoudre des problèmes non linéairement séparables et des problèmes logiques plus compliqués (Talbi 2011).Il suit un apprentissage supervisé, ce réseau a été choisi dans la suite de nos travaux.

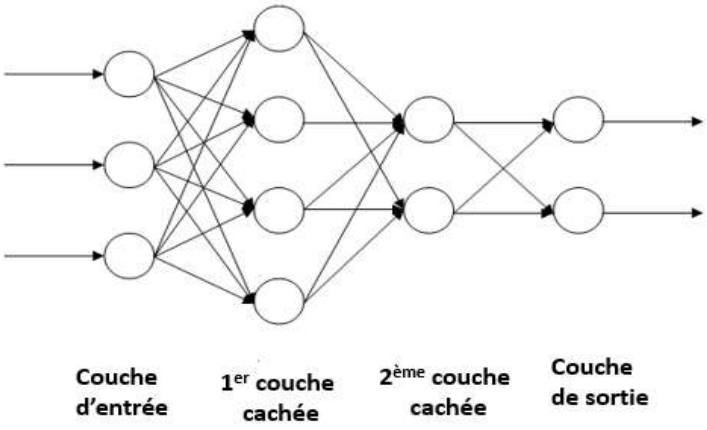


Figure II.6. L'architecture d'un PMC (MSAAF et al 2015).

II.5.3. Réseaux à fonction radiale

Le RN à base de fonction radiale (RBF) (Broomhead et al. 1988, Moodyet al. 1989)est un réseau de neurones (Figure II.7) à trois couches une couche d'entrée, une cachée et une couche de sortie. La particularité de ce réseau réside dans sa couche cachée (deuxième couche), Ce réseau est capable de fournir une représentation locale de l'espace, en classification et en approximation de fonctions, particulièrement. L'apprentissage le plus utilisé pour les RFR est le mode hybride (Talbi 2011).

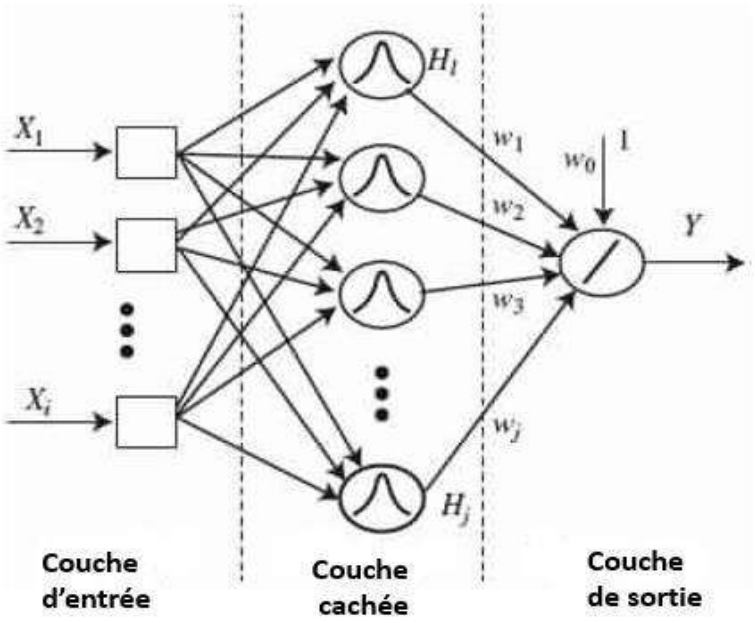


Figure II.7. L'architecture d'un RBF (Msaaf et all 2015).

II.5.4. Cartes auto-organisatrices de Kohonen (SOM)

La carte de Kohonen (Kohonen 1982) est une carte auto-organisatrice en deux couches (Figure II.8), une couche d'entrée et une autre de sortie. La couche de sortie appelée aussi couche compétitive est en deux dimensions. Chaque neurone d'entrée est connecté à l'ensemble des neurones de sortie par des poids W_{ij}. La carte de Kohonen est une carte à apprentissage non supervisé, à partir d'une base d'apprentissage constitué seulement des éléments d'entrées, elle permet de regrouper ces entrées en classes.

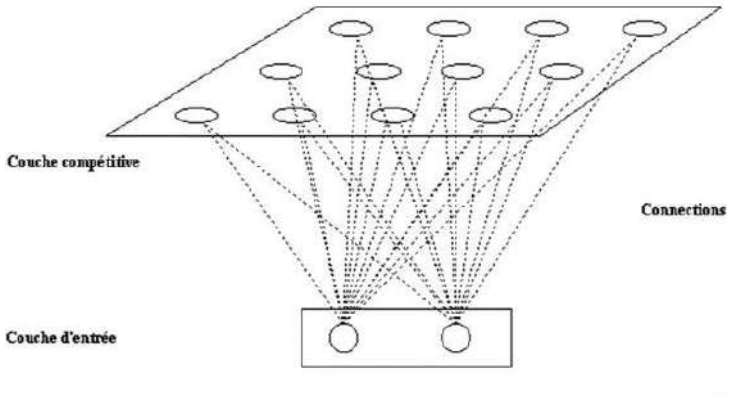


Figure II.7. L'architecture d'une L'architecture d'une carte de Kohonen (Msaaf et all 2015).

II.5.5. Réseaux de Hopfield

Les réseaux de Hopfield (Hopfield 1982) sont des réseaux récurrents et entièrement connectés. Dans ce type de réseau, chaque neurone est connecté à chaque autre neurone et il n'y a aucune différence entre les neurones d'entrée et de sortie. Ils fonctionnent comme une mémoire associative non-linéaire et sont capables de trouver un objet stocké en fonction de représentations partielles ou bruitées. L'application principale des réseaux de Hopfield est l'entrepôt de connaissances, mais aussi la résolution de problèmes d'optimisation. Le mode d'apprentissage utilisé ici est le mode non-supervisé (Talbi 2011).

II.5.6. Réseaux ART

Les réseaux ART (en anglais Adaptive Resonance Theory) sont des réseaux à apprentissage par compétition (Carpenter et Grossberg 1987; Carpenter et Grossberg 1988). Dans ce genre de réseau, les vecteurs de poids ne seront adaptés que si l'entrée fournie est suffisamment proche d'un prototype déjà connu par le réseau. Il est à noter qu'il existe deux principaux types de réseaux ART les ART-1 pour des entrées binaires et les ART-2 pour des entrées continues. Le mode d'apprentissage des ART peut être supervisé ou non (Talbi 2011).

Tableau II.2 Les avantages et les inconvénients des quelques architectures des réseaux de neurones

RN	Description	Apprentissage	Avantages	Inconvénients
Perceptron simple	Un seul neurone	**Supervisé**	-Architecture simple	-N'accepte pas les données bruitées et la classification non linéaire

PMC	-Plusieurs neurones organisés en couches -Sorties des neurones : sigmoïde	**Supervisé**	- Accepte les données bruitées et la classification non-linéaire - Représentation globale de l'espace -Architecture simple	- Classe les éléments qui n'appartiennent à aucune classe - Le nombre de couches caché et des neurones dans des couches caches est indéfinit
RVFLNN	-Réseau en trois couches -Couche d'entrée reliée à la couche de sortie	**Supervisé** (seulement les poids reliant la couche d'entrée avec celle de la sortie)	-Accepte les données bruitées et la classification non-linéaire - Représentation globale de l'espace - Des entrées optimisées	-Architecture plus complexe que le PMC Moins de précision que le PMC
RBF	-Plusieurs neurones organisés en 3 couches -Couche cachée : gaussienne	**Non supervisé (μ, σ) + Supervisé**	- Accepte les données bruitées et la classification non-linéaire - Représentation globale de l'espace - Une grande précision - Apprentissage rapide	-Apprentissage complexe - Nécessite une grande capacité de calcule
Carte de Kohonen	-Couche de sortie sous forme d'une carte (2D)	**Non supervisé**	-Classifier les données sans connaissance préalable de procédé. -regroupement automatique des données	-Parfois les résultats ne correspondent pas à celle souhaitée

I.6. Conception d'un modèle neuronal

Les réseaux de neurones réalisent des fonctions non linéaires paramétrées. Leurs mises en œuvre nécessitent (Khaouane 2013) :

- La détermination des entrées et des sorties pertinentes, c'est-à-dire les grandeurs qui ont une influence significative sur le phénomène que l'on cherche à modéliser ;
- La construction de la base de données et le choix des échantillons ;
- La détermination du nombre de neurones cachés nécessaires pour obtenir une approximation satisfaisante ;
- La réalisation de l'apprentissage ;
- L'évaluation des performances du réseau de neurones à l'issue de l'apprentissage.

II.6.1 Détermination des entrées / sorties du réseau de neurones

Pour toute conception de modèle, la sélection des entrées doit prendre en compte deux points essentiels (Ammar 2007) :

- Premièrement, la dimension intrinsèque du vecteur des entrées doit être aussi petite que possible, en d'autres termes, la représentation des entrées doit être la plus compacte possible, tout en conservant pour l'essentiel la même quantité d'information et en gardant à l'esprit que les différentes entrées doivent être indépendantes.
- En second lieu, toutes les informations présentées dans les entrées doivent être pertinentes pour la grandeur que l'on cherche à modéliser : elles doivent donc avoir une influence réelle sur la valeur de la sortie.

II.6.2. Construction de la base de données et choix des échantillons

Le processus d'élaboration d'un réseau de neurones commence par la construction d'une base de données et le choix des échantillons. Il faut donc déterminer les grandeurs qui ont une influence significative sur le phénomène que l'on cherche à modéliser.

Cette étape est cruciale et aide à déterminer le type de réseau le plus approprié pour résoudre un problème donné. En particulier, la façon dont se présente l'échantillon conditionne le type de réseau, le nombre de cellules d'entrée et de sortie et la façon dont il faudra mener l'apprentissage, le test et la validation (Khaouane 2013).

Afin de développer une application à base de réseaux de neurones, il est nécessaire de disposer de deux bases de données, une pour effectuer l'apprentissage et l'autre pour tester le réseau obtenu et déterminer ses performances.

Pour mieux contrôler la phase d'apprentissage, il est souvent préférable de posséder une troisième base de données appelée "base de validation croisée" (Stone 1977). Les avantages liés à l'utilisation de cette troisième base de données seront exposés dans les paragraphes suivants.

Notons qu'il n'y pas de règle pour déterminer ce partage d'une manière quantitative, néanmoins chaque base doit satisfaire aux contraintes de représentativité de chaque classe de données et doit généralement refléter la distribution réelle, c'est-à-dire la probabilité d'occurrence des diverses classes (Nascimento et al. 2000).

Généralement, les bases de données subissent un prétraitement qui consiste à effectuer une normalisation appropriée tenant compte de l'amplitude des valeurs acceptées par le réseau. Les valeurs d'entrées et de sortie sont normalisées dans un intervalle spécifique afin de donner à chaque paramètre la même influence statistique. De plus, un problème additionnel peut survenir à cause d'une possible saturation de la sortie, si la somme pondérée des entrées du perceptron multicouches ($\sum w_{ij}.x_j$) est dans la partie asymptotique de la fonction d'activation (Dames 2005).

II.6.3. Définition de la structure du réseau

La détermination de l'architecture optimale est restée pendant longtemps un problème ouvert. Il existe actuellement diverses méthodes empiriques mettant notamment en jeu des tests statistiques pour une vaste classe de réseaux (Khaouane 2013).
Comme nous l'avons mentionné, nous avons retenu le perceptron multicouche comme base du modèle. Nous structurons ce réseau en précisant le nombre de couches et de neurones cachés pour que le réseau soit en mesure de reproduire ce qui est déterministe dans les données.

II.6.3.1. Nombre de couches cachées

Mis à part les couches d'entrée et de sortie, il faut décider du nombre de couches intermédiaires ou cachées. Sans couche cachée, le réseau n'offre que de faibles possibilités d'adaptation. Néanmoins, il a été démontré qu'un perceptron multicouche avec une seule couche cachée pourvue d'un nombre suffisant de neurones, peut approximer n'importe quelle fonction avec la précision souhaitée (Khaouane 2013).

II.6.3.2. Nombre de neurones cachés

Chaque neurone peut prendre en compte des profils spécifiques de neurones d'entrée. Un nombre plus important permet donc de mieux "coller" aux données présentées, mais diminue la capacité de généralisation du réseau. Il faut alors trouver le nombre adéquat de neurones cachés

nécessaire pour obtenir une approximation satisfaisante. Le nombre de neurones cachés dépend du type de données dont on dispose. En général, ce nombre est déterminé par des règles empiriques. Suivant les auteurs, la taille de la couche cachée doit être (Dames 2005) :

- Soit égale à celle de la couche d'entrée ;
- Soit égale à 75 % de celle-ci ;
- Soit égale à la racine carrée du produit du nombre de neurones dans la couche d'entrée et de sortie.

Pour des cas plus sophistiqués, la recherche consisterait soit à :

- Estimer un réseau comportant de nombreux neurones puis à le simplifier par l'analyse des multi-colinéarités ou par une règle d'apprentissage éliminant les neurones inutiles (Dames 2005) ;
- Définir une architecture tenant compte de la structure des variables identifiées au préalable par une analyse en composantes principales (Bermejo et Cabestany 1999 ; Bermejo et Cabestany 2001) ;
- Augmenter progressivement la complexité du modèle à chaque itération, dite "Cascading", et réévaluer l'efficacité de généralisation à chaque addition d'un nouveau neurone (Dames2005).

II.6.4. Apprentissage

Une fois l'architecture choisie, elle doit subir une phase d'apprentissage. Cette dernière consiste à calculer les coefficients synaptiques de telle manière que les sorties du réseau de neurones soient, pour les exemples utilisés lors de l'apprentissage, aussi proches que possibles des sorties "désirées". Celles-ci correspondent à la valeur de la fonction que l'on veut approcher ou à la sortie du processus que l'on veut modéliser (Dreyfus 2008).

L'apprentissage correspond donc à la phase du développement du réseau durant laquelle les coefficients synaptiques sont modifiés jusqu'à l'obtention du comportement désiré. Dans notre étude, comme nous disposons des valeurs de sortie, nous avons choisi l'apprentissage supervisé, mettant en jeu un algorithme d'optimisation. Ce dernier cherche à minimiser, par des méthodes d'optimisation non linéaire, une fonction de coût qui constitue une mesure de l'écart entre les réponses réelles du réseau et ses réponses désirées (Dreyfus 2008). Cette optimisation se fait de manière itérative, en modifiant les poids en fonction du gradient de la fonction de coût : le gradient est estimé par une méthode spécifique, dite la règle du gradient

de l'erreur (en anglais delta rule). C'est une des règles les plus utilisées pour l'apprentissage supervisé de réseaux de neurones. Initialement développée pour résoudre des problèmes de traitements adaptatifs du signal, elle a conduit ensuite à un algorithme connu sous le nom de l'algorithme de rétro-propagation du gradient de l'erreur (en anglais Back-propagation) (Dames 2005). On distingue deux étapes dans la phase d'apprentissage (Dames 2005) :

- **Une phase de propagation**, qui consiste à présenter une configuration d'entrée au réseau, puis à la propager à celle de sortie en passant par les couches cachées.
- **Une phase de rétro-propagation**, qui consiste, après le processus de propagation, à minimiser l'erreur commise sur l'ensemble des exemples présentés, erreur considérée comme une fonction des poids synaptiques. Cette erreur représente la somme des différences au carré entre les réponses calculées et celles désirées pour tous les exemples contenus dans l'ensemble de l'apprentissage.

  Les algorithmes de rétro-propagation implantés actuellement dans les logiciels de RNA appartiennent à une des cinq classes suivantes (Si-Moussa 2009):

  - Descente du gradient ;
  - Descente des gradients conjugués ;
  - Quasi-Newton ;
  - Levenberg-Marquardt ;
  - Régularisation.

Le tableau II.2 résume les algorithmes d'apprentissage par rétro-propagation implanté dans laboite à outils des RN du programme MATLAB® (Khaouane 2013).

Tableau II.3. Résumé des caractéristiques des algorithmes d'apprentissage de rétro propagations implantées dans MATLAB® (Khaouane 2013).

Algorithme	**Caractéristiques**
Descente du gradient	
Descente du gradient simple ***(traingd)***.	Descente du gradient de base. Réponse lente ; peut être utilisée pour apprentissage en mode séquentiel.
Descente du gradient avec moment (***traingdm)***.	Descente du gradient avec moment. Le moment permet à un réseau de répondre non seulement au gradient local, mais aussi aux tendances récentes dans la surface d'erreur. Sans moment un réseau peut être coincé dans un minimum local peu profond. Plus rapide que ***traingd***, ne peut être utilisée que pour apprentissage en mode séquentiel.
Descente du gradient avec taux d'apprentissage variable ***(traingda)***	Taux d'apprentissage variable. Plus rapide que ***traingd***, mais ne peut être utilisée que pour apprentissage en mode batch. Utilise des règles heuristiques pour la variation du taux d'apprentissage
Descente du gradient avec moment et taux d'apprentissage variable **(traingdx).**	Descente du gradient avec inertie (Momentum) et taux d'apprentissage variable.
Descente des gradients conjugués	
Descente des gradients conjugués avec la méthode de mise à jour de Fletcher-Reeves ***(traincgf)***.	Il possède la plus petite mémoire de stockage parmi les algorithmes de gradient conjugués.
Descente des gradients conjugués avec la méthode de mise à jour de Polak-Ribiére ***(traincgp)***.	Nécessite une mémoire de stockage légèrement plus grande que ***traincgf*** . Convergence plus rapide sur quelques problèmes.

Descente des gradients conjugués avec la méthode de reprises Powell-Beale ***(traincgb)***.	Nécessite une mémoire de stockage légèrement plus grande que ***traincgp***. Convergence généralement plus rapide.
Scaledconjugate gradient ***(trainscg)***.	Le seul algorithme de gradients conjugués qui n'exige aucune recherche de ligne. Très bon algorithme d'apprentissage en général.
Quasi-Newton	
BFGS quasi-Newton ***(trainbfg)***.	Il exige le stockage d'une approximation de la matrice Hessiènne et nécessite plus de calcul à chaque itération que les algorithmes de gradients conjugués, mais possède habituellement une convergence plus rapide.
Sécante à un pas (***trainoss***)	Cet algorithme exige moins de stockage et calcul par époque que l'algorithme BFGS. On peut considérer cela un compromis entre des algorithmes de quasi-Newton complets et des algorithmes de gradients conjugués.
Levenberg-Marquardt	
Levenberg-Marquardt***(trainlm)***.	C'est l'algorithme d'apprentissage le plus rapide pour les réseaux de taille modérée. Il a la caractéristique de réduction de mémoire pour l'utilisation quand l'échantillon d'apprentissage est grand.
Régularisation	
Régularisation Bayésienne ***(trainbr)***.	Mise à jour des poids et des biais selon l'optimisation de Levenberg-Marquardt. Il réduit au minimum une combinaison d'erreurs quadratiques et des poids et détermine ensuite la combinaison correcte afin de produire un réseau avec une capacité de généralisation meilleure. Ce processus est appelé la régularisation bayésienne. Il réduit la difficulté de déterminer l'architecture de réseau optimum.

II.6.5. Evaluation des performances du réseau de neurones

Enfin, la dernière étape doit permettre d'estimer la qualité du réseau obtenu en lui présentant des exemples qui ne font pas partie de l'ensemble d'apprentissage. Elle permet à la fois d'apprécier les performances du système neuronal et de détecter le type de données qui posent problème. Si les performances ne sont pas satisfaisantes, il faudra, soit modifié l'architecture du réseau, soit modifié la base d'apprentissage (Khaouane 2013).

II.7. Conclusion

Ce chapitre introduit le concept fondamental sur lequel la littérature se concentre, à savoir que la modélisation peut vérifier des hypothèses simplificatrices faites dans la formulation mathématique. Peu de travaux ont appliqué la relation quantitative structure-activité (QSAR) à la modélisation des processus membranaires, et ils peuvent être considérés comme un outil important pour prédire la rétention des produits pharmaceutique dans les processus membranaires d'osmose inverse et de nanofiltration.

Dans la seconde, nous présentons les réseaux de neurones de manière générale, en montrant des modèles biologiques et mathématiques de neurones, en montrant diverses architectures de RN et leurs types de formation, ainsi que les RN les plus couramment utilisés et les différentes conceptions de réseaux de neurones.

Chapitre III

Matériels et Méthodes



III.1. Introduction

Les réseaux de neurones artificiels (RNA) ont acquis une grande popularité dans la dernière décennie en raison de leur parallélisme élevé, de leur robustesse de leur capacité inhérente à extraire à partir des données expérimentales, les relations fortement non linéaires et complexes entre les variables du problème sans aucune connaissance détaillée du système (Si-Moussa 2010).

Ce chapitre comportera quatre axes. Dans le premier axe, on examinera les données collectées, prétraitement et analyse. Le second axe sera consacré aux techniques les plus récentes à savoir les réseaux de neurones artificiels (RNA) et les modèles QSPR. Le troisième axe traitera l'analyse de la sensibilité basée sur la méthode de "poids". Le diagnostic des points aberrants et domaine d'applicabilité du modèle neuronal optimisé sera présenté dans le quatrième axe.

III.2. Collecte des données, prétraitement et analyse

Dans cette étude, nous avons utilisé les données disponibles dans la littérature (Bellona et al. 2005; Tanga et al. 2009; Yangali-Quintanilla et al. 2009; Huang et al. 2011; Dolar et al. 2011; 2012; 2013; 2013b; 2016; Garcia et al. 2017; Licona et al. 2018; Yang et al. 2018b; Xu et al. 2019; Couto et al. 2020; Renata et al. 2022). La taille de la base de données et le nombre de produits pharmaceutiques sont montrés dans le tableau III.1.

Tableau III.1. Taille de la base de données et le nombre des produits pharmaceutiques utilisés dans ce travail.

Base de Données (BD)	**Taille de base de données**	**Nombre des produits pharmaceutiques**
BD	825	31

La liste des 31 produits pharmaceutiques sont répertoriés dans l'annexe A. Le choix des variables d'entrées et de sortie a été basé sur les interactions entre les propriétés des produits pharmaceutiques, les caractéristiques de la membrane et les conditions opératoires de filtration pour la rétention des produits pharmaceutiques par les membranes d'OI/NF. Les entrées considérées dans cette étude sont : les descripteurs moléculaires (MW , diamètre effectif "dc", LogD, le moment dipolaire, la longueur moléculaire "length", la surface minimale "S_{min}", la surface maximale "S_{max}", la largeur moléculaire "Width", la profondeur moléculaire "Depth" et la largeur équivalente moléculaire "Eqwidth"), les

caractéristiques de la membrane (le poids moléculaire de coupure "MWCO", le rejet de sel de sodium chloride "SR(NaCl)", la charge de surface de la membrane "potentiel zêta" et l'hydrophobicité de la membrane "angle de contact") et les conditions opératoires (le pH, la pression, la température et le taux de récupération).

Les valeurs de LogD, moment dipolaire, surface minimale "S_{min}", surface maximale "S_{max}" ont été calculées par les logiciels qui sont mentionnés dans le tableau III.2. Sur l'annexe B, nous donnons la capture d'écran de deux sites et deux logiciels utilisés.

Tableau III.2. Logiciels utilisés pour calculer LogD, le moment dipolaire, Smin, Smax et Length .

	Logiciels
LogD	Chemspider http://www.chemspider.com/
Moment dipolaire	Hyperchem - (unitiled)
S_{min}	
S_{max}	
Length	ChemBio3D Ultra (Version 12 ; 2012)

Les valeurs de la largeur équivalente moléculaire (EqWidth) ont été calculées par l'équation I.6 (voir le premier chapitre). Le nombre total de variables (les propriétés des produits pharmaceutique, les caractéristiques membranaires et les conditions d'opératoires) a été réduit en utilisant des matrices de corrélation (voir annexe C).

Chaque matrice regroupe l'intensité de la liaison qui existe entre ces paramètres. Cette intensité de liaison a été obtenue par le calcul des coefficients de corrélation de Pearson (R) (Équation III.1) (Ammi 2016) :

$$R(Z,Y) = \frac{\sum_i (Z_i - \bar{z})(y_i - \bar{y})}{\sqrt{\sum_i (z_i - \bar{z})^2}\sqrt{\sum_i (y_i - \bar{y})^2}} \qquad \text{(III. 1)}$$

Où Z_i est la $i^{\text{ème}}$ valeur du vecteur Z (d_c, LogD, Length, Dipole moment, S_{min} ,S_{max}, Depth, width, EqWidth, MWCO, SR (NaCl), angle de contact, potentiel zêta, pH, pression, température, taux de récupération et rétention) ; yi est la $i^{\text{ème}}$ valeurdu vecteur Y (d_c, LogD, Length, Dipole moment, S_{min}, S_{max}, Depth, width, EqWidth, MWCO, SR (MgSO4), angle de contact, potentiel zêta, pH, pression,

température, taux de récupération et rétention) ; $\bar{z}$ et $\bar{y}$sont respectivement les moyennesarithmétiques des vecteurs *Z* et *Y*. Le degré de coefficient de corrélation de Pearson doit être modéré (R<0.75) (Ammi 2016), On élimine l'entrée qui possède une relation entre eux et les l'entrée qui ont la même valeur pour plus de 0.75 des échantillons.

Pour chaque base de données (BD1, BD2, BD3 et BD4), un modèle QSPR-RN a été élaboré et un ensemble de variables d'entrées et de sortie a été identifié (Tableau III.3).

Tableau III.3. Entrées et sortie utilisées pour le développement des modèles QSPR-RN (QSPR-RN1, QSPR-RN2, QSPR-RN3 et QSPR-RN4)

	QSPR-RN1	**QSPR-RN2**	**QSPR-RN3**	**QSPR-RN4**
Entrées	-d_c -LogD -Dipole moment -Length -S_{max} -MWCO -SR(NaCl) -Zeta potentiel -Contacte angle -PH -Pression -Taux de récupération -Température	-d_c -LogD -Dipole moment -Length -S_{min} -MWCO -SR(NaCl) -Zeta potentiel -Contacte angle -PH -Pression -Taux de récupération -Température	-d_c -LogD -Dipole moment -Length -Width -Depth -MWCO -SR(NaCl) -Zeta potentiel -Contacte angle -PH -Pression -Taux de récupération -Température	-d_c -LogD -Dipole moment -Length -Eqwidth -MWCO -SR(NaCl) -Zeta potentiel -Contacte angle -PH -Pression -Taux de récupération -Température
Sortie	Rétention	Rétention	Rétention	Rétention

L'analyse statistique des données d'entrées et de sortie, en termes du minimum "min", de la moyenne arithmétique (Équation III.2) "$\bar{x}$", du maximum "max" et de la déviation standard (écart type) "DS" (Équation III.3) (Khaouane 2013), sont présentées dans les tableaux III.4.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \text{(III. 2)}$$

$$DS = \left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{1/2} \qquad \text{(III. 3)}$$

Tableau III.4. Analyse statistique des données d'entrée et de sortie (BD).

	Min	**Moyenne**	**Max**	**DS**
dc (g/mol)	0,562791	0,938969	0,749531	0,093045
LogD	-9,62000	3,72000	-0,00968	2,26300
Moment dipolaire (Debye)	0,00000	16,73000	4,89191	3,70784
***Length* (nm)**	0,150000	1,997000	1,256159	0,420348
***EqWidth* (nm)**	2,750000	6,440000	4,629188	0,815818
MWCO (Dalton)	96,0000	405,0000	192,4242	90,4085
SR (NaCl) (%)	0,12100	75,00000	0,91063	3,11162
Potentiel zêta (mv)	-80,0000	3,1000	-30,1389	19,9710
Angle de contact (°)	15,12000	80,00000	48,23612	16,38807
pH	2,750000	9,500000	6,317915	1,391140
Pression (Kpa)	280,000	4100,000	1037,976	465,608
Taux de récupération (%)	0,0300	100,0000	22,7935	33,0178
Température (°C)	20,00000	45,00000	28,52727	8,01134
Rétention (%)	1,2900	100,0000	83,1138	21,8418

La figure III.1 représente les rétentions expérimentales en fonction de poids moléculaires pour la base de données totale.

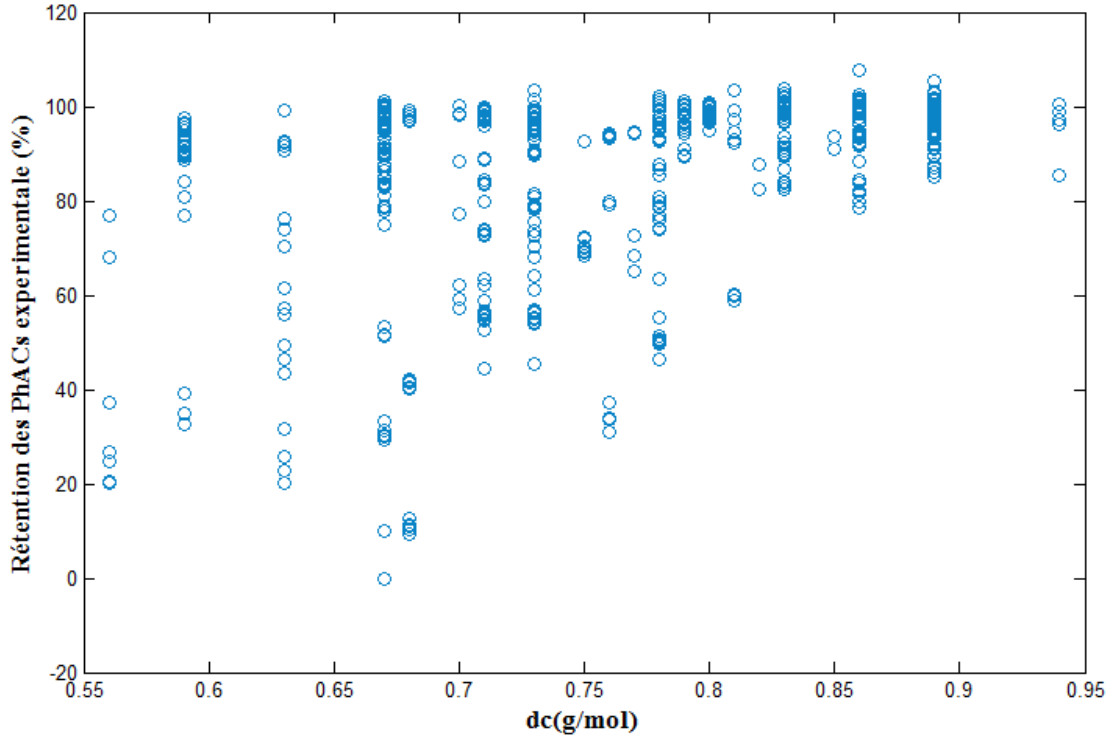


Figure III.1. Rétentions expérimentales en fonction de diamètre effective "dc".

III.3. Développement des modèles QSPR-RNA

Cette étape concerne la création des modèles de réseaux de neurones artificiels basées sur la relation quantitative structure propriété "QSPR-RNA".

Les réseaux de neurones artificiels (RNA) sont particulièrement bien adaptés pour les modèles QSAR/QSPR en raison de leur capacité à extraire des informations non-linéaires présentes dans la matrice de données (Yangjeh et Jenagharad 2009). Pour trouver un modèle non linéaire entre les propriétés des produits pharmaceutiques, les caractéristiques de la membrane et les conditions opératoires de filtration pour la rétention des produits pharmaceutiques par les membranes d'OI/NF, l'utilisation de la méthode RNA est nécessaire et est devenue une technique de modélisation importante. Elle est largement utilisée pour les études de QSPR non linéaires et est un outil puissant pour construire des modèles prédictifs.

Dans ce travail, une procédure basée sur le développement et l'optimisation de l'architecture d'un QSPR-RNA est décrite. Elle est basée, comme décrite sur l'organigramme ci-dessous.

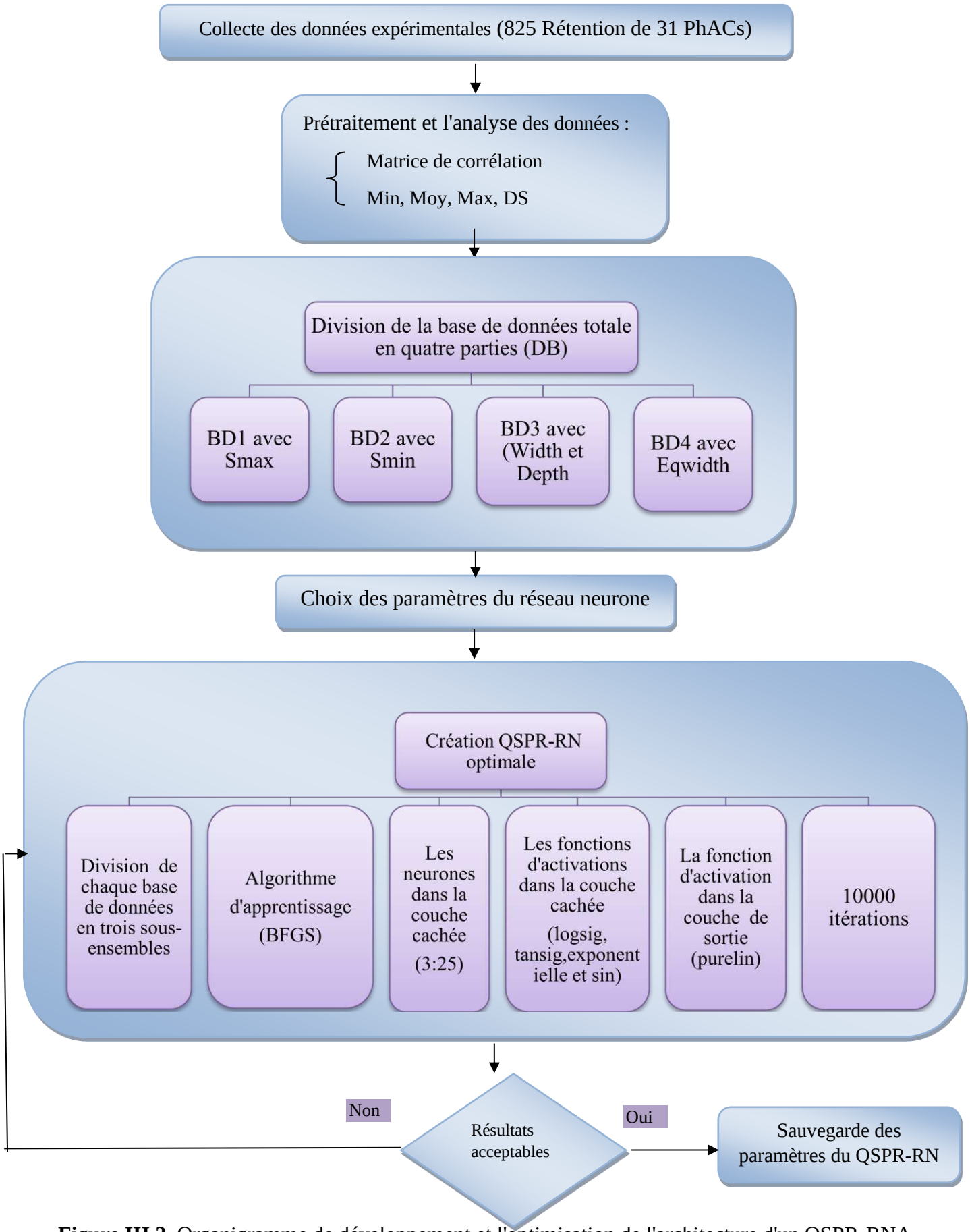


Figure III.2. Organigramme de développement et l'optimisation de l'architecture d'un QSPR-RNA.

L'algorithme d'apprentissage utilisé dans ce travail est le BFGS quasi-Newton (trainbfg). Chaque QSPR-RN contient trois couches de neurones ou des nœuds : une couche d'entrée avec 13 neurones dans la couche d'entrée, une couche cachée avec un nombre de neurones actifs optimisés lors de l'apprentissage et une couche de sortie avec une unité générant la valeur estimée de la rétention. Le nombre de neurones dans la couche cachée a été varié de 3 à 25 neurones. Les fonctions d'activations « La tangente hyperbolique sigmoïde (tansig), la sigmoïde logarithmique (logsig), l'exponentielle et sin » ont été utilisées dans la couche cachée. La fonction de transfert linéaire (purelin) a été utilisée dans la couche de sortie.

La procédure d'élaboration du QSAR-RN décrite ci-dessus a été réalisée en utilisant STATISTICA (Annexe B). Ce logiciel est utilisé pour optimiser l'architecture des quatre modèles neuronaux QSAR-RN employés pour la rétention des produits pharmaceutiques par les membranes d'OI/NF.

III.4. Analyse de sensibilité par la méthode de "poids"

La contribution des variables d'entrées (dc, *LogD*, Moment dipolaire**,** *Lenght*, *EqWidth*, MWCO, SR(NaCl), angle de contact, potentiels zêta, pH, pression, taux de récupération et température) sur la sortie a été déterminée par une analyse de sensibilité en utilisant la méthode de "poids" pour chaque réseau de neurone QSAR-RN

Cette méthode, proposée d'abord par Garson (Garson 1991) et reprise par Goh (Goh 1995), donne une quantification de l'importance relative (*IR*) des différentes entrées sur la sortie du RN (Gevrey et al. 2003). Elle se base sur le partitionnement des poids de connexions en :

- Poids de connexions entrée-cachée (w_{ji}^{I}) ;
- Poids de connexions cachée-sortie (w_{1j}^{H}).

Le processus de calcul, compose de quatre étapes, est schématisé sur l'organigramme ci-dessous (Khaouane 2013).

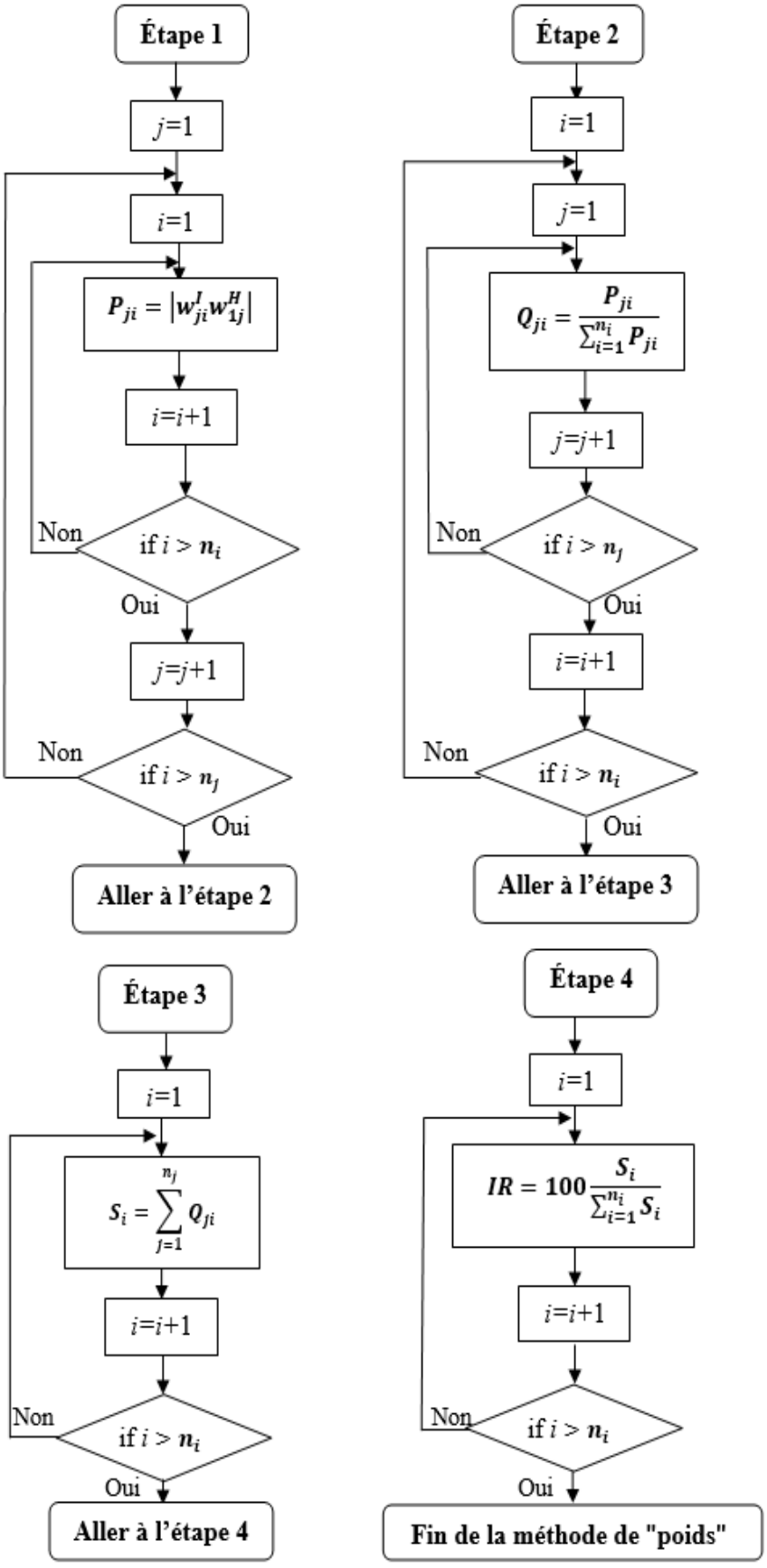


Figure III.3. Organigramme de la méthode de "poids" (Khaouane 2013).

III.5. Diagnostic des points aberrants et domaine d'applicabilité du modèle neuronal optimisé

Les diagrammes de dispersions des modèles optimisés ont montré un certain nombre des points éloignés de la première bissectrice que nous avons attribué à l'existence probable de points aberrants. Ceci nous a motivé à réaliser un diagnostic des données potentiellement suspectes selon la méthode décrite dans les deux références suivantes (Hosseinzadeh and Hemmati-Sarapardeh 2014) ; (Baghban, Mohammadi et al. 2017).

Les valeurs aberrantes sont bien connues en tant que points de données distincts de la tendance générale des principaux points de données. La détection des valeurs aberrantes, qui joue un rôle important dans la proposition de modèles, est utilisée pour reconnaître des données individuelles (ou des groupes de données) pouvant différer de la masse des données constituant un ensemble de données. À cette fin, il est essentiel de trouver des méthodes rigoureuses pour détecter les valeurs potentiellement suspectes afin d'éliminer des données expérimentales inexactes et d'améliorer la précision des modèles (Hosseinzadeh and Hemmati-Sarapardeh 2014) ; (Baghban, Mohammadi et al. 2017). Dans cette étude, la technique mathématique "Leverage" a été utilisée pour rechercher les valeurs aberrantes. Dans cette technique, les valeurs résiduelles sont d'abord calculées. Ensuite une matrice H de points de données d'entrée est créée sur la base de la formule suivante :

$$H = X(X^tX)^{-1}X^t \qquad (III.4)$$

Où X fait référence à la matrice m × n (m et n représentent respectivement le nombre d'échantillons et les paramètres (variables d'entrée) du modèle). Les valeurs de Hat sont obtenues à partir de la diagonale principale de la matrice de H.

$$Hat = diagonale(H) \qquad (III.5)$$

Le diagramme de William qui représente les résidus normalisés en fonction du "Leverage" levier (Hat) peut détecter graphiquement les valeurs aberrantes. Dans ce graphique, le levier critique (H^*) (seuil) est généralement fixé à la valeur donnée par l'équation suivante :

$$H^* = \frac{3(n+1)}{m} \qquad (III.6)$$

Ou m et n représentent respectivement le nombre d'échantillons et les paramètres (variables d'entrée) du modèle.

Les résidus normalisés sont calculés à partir des données de la sortie expérimentale et celle calculée par le modèle.

$$(R_Norm)_i = \frac{(y_i^{exp} - y_i^{cal})}{\sqrt{var(y^{exp} - y^{cal})}} \quad i = 1, \dots m \qquad (III.7)$$

Le résidu normalisé égal à trois est normalement considéré comme une valeur "seuil" pour accepter les points de données dans les écarts types compris sur une plage de ± 3 par rapport à la moyenne (pour couvrir 99% des données normalement distribuées). Si la majorité des points de données se situent dans les plages de $0 \leq$ Hat $\leq H^*$ et de $-3 \leq$ R_Norm ≤ 3, cela indique que le développement du modèle et ses prédictions sont effectués dans le domaine d'applicabilité, ce qui conduit à un modèle statistiquement valide. Ainsi on peut affirmer qu'il existe des points ‘’Good High Leverage’’ dans le domaine de $0 \leq$ Hat $\leq H^*$ et $-3 \leq$ R_Norm ≤ 3. Les points situés dans la plage de R_Norm <-3 ou R_Norm>3 (qu’ils soient plus grands ou plus petits que la valeur H^*) sont désignés comme points aberrants du modèle ou comme points "Bad High Leverage".

III.6. Conclusion

Nous avons décrit ici les outils nécessaires à la modélisation de la rétention des produits pharmaceutiques lors d’osmose inverse et de nanofiltration par les modèles QSPR-RN. Dans cette étape, nous avons présenté les données collectées, prétraitement et analyse. Après nous avons présenté la technique la plus récente pour établir des modèles QSPR, les réseaux de neurones artificiels" RNA". Aussi une analyse de sensibilité basée sur la méthode de "poids". Nous avons terminé ce chapitre par diagnostic des points aberrants et domaine d’applicabilité du modèle neuronal optimisé.

Le chapitre suivant sera consacré aux résultats et discussion pour mieux comprendre le phénomène étudié.

Chapitre IV

Résultats et discussions



IV.1. Introduction

Dans ce chapitre, les résultats obtenus sont discutés, il est organisé en trois parties :

- Dans la première partie, on présentera l'effet de la division de BD, les performances des modèles QSPR-RN et la formule mathématique de modèle optimal.
- La deuxième partie on discutera les résultats de l'analyse de sensibilité par la méthode de "poids".
- La troisième partie présentera les résultats de l'analyse diagnostic des points aberrants et domaine d'applicabilité du modèle neuronal optimisé.

IV.2. Performance des modèles QSPR-RN

IV.2.1. Effet de la division de BD

Afin de développer le modèle QSPR-RNA, la base de données finale est divisée en 04 bases de données, chaque base de données est divisée en 3 bases : base d'apprentissage, base de test, base validation et base totale ; Les résultats de performances en termes de la racine de l'erreur quadratique moyenne (REQM) et en termes du coefficient de corrélation (R) sont résumés dans les tableaux IV.1, IV.2, IV.3 et IV.4. Les résultats montrent que la division 01 avec QSPR-RNA4 donne des meilleur résultats (R =0.9883 et REQM =3.5360 pour la phase de test). Ceci implique que le modèle QSPR-RNA4 est le réseau de neurones optimal (QSPR-RNA$_{optimal}$).

Table IV.1. Effet de la division de la base de données (BD1).

	QSPR-RN1 avec (S_{min})	**Coefficient de corrélation "R"**	**RMSE (%)**
Division 01	Phase totale "825 points"	0.9754	4.8147
	Phase d'apprentissage 80% "661 points"	0.9760	4.8358
	Phase de validations 10% "82points"	0.9520	5.5796
	Phase de test 10% "82 points"	0.9863	3.6862
Division 02	Phase totale "825 points"	0.9319	7.9284
	Phase d'apprentissage70% "579 points"	0.9339	7.7349
	Phase de validations 15% "123 points"	0.9358	7.0097
	Phase de test 15% "123 points"	0.9232	9.5316
Division 03	Phase totale "825 points"	0.9449	7.1570
	Phase d'apprentissage 60% "495 points"	0.9594	6.1060

	Phase de validations20% "165 points"	0.9432	6.7636
	Phase de test20% "165 points"	0.9097	9.9254

Table IV.2. Effet de la division de la base de données (BD2).

	QSPR-RN2 avec (S_{max})	**Coefficient de corrélation "R"**	**RMSE (%)**
Division 01	Phase totale "825 points"	0.9794	4.4089
	Phase d'apprentissage 80% "661 points"	0.9804	4.3717
	Phase de validations 10% "82points"	0.9571	5.3480
	Phase de test 10% "82 points"	0.9871	3.5923
Division 02	Phase totale "825 points"	0.9363	7.6695
	Phase d'apprentissage70% "579 points"	0.9435	7.1692
	Phase de validations 15% "123 points"	0.9152	8.0062
	Phase de test 15% "123 points"	0.9230	9.4066
Division 03	Phase totale "825 points"	0.9443	7.1881
	Phase d'apprentissage 60% "495 points"	0.9625	5.8774
	Phase de validations20% "165 points"	0.9398	7.1141
	Phase de test20% "165 points"	0.9044	10.2028

Table IV.3. Effet de la division de la base de données (BD3).

	QSPR-RN3 avec (Depth width)	**Coefficient de corrélation "R"**	**RMSE (%)**
Division 01	Phase totale "825 points"	0.9749	4.8613
	Phase d'apprentissage 80% "661 points"	0.9754	4.8922
	Phase de validations 10% "82points"	0.9506	5.6343
	Phase de test 10% "82 points"	0.9868	3.6180
Division 02	Phase totale "825 points"	0.9522	6.6701
	Phase d'apprentissage70% "579 points"	0.9601	6.0524
	Phase de validations 15% "123 points"	0.9417	6.7034
	Phase de test 15% "123 points"	0.9314	9.0019
Division 03	Phase totale "825 points"	0.9342	7.8065
	Phase d'apprentissage 60% "495 points"	0.9499	6.7723
	Phase de validations20% "165 points"	0.9246	7.9030
	Phase de test20% "165 points"	0.9026	10.2300

Table IV.4. Effet de la division de la base de données (BD4).

	QSPR-RN4 avec (Eqwidth)	Coefficient de corrélation "R"	REQM (%)
Division 01	Phase totale "825 points"	0.9746	4.8948
	Phase d'apprentissage 80% "661 points"	0.9771	4.7254
	Phase de validations 10% "82points"	0.9249	6.9678
	Phase de test 10% "82 points"	0.9883	3.5360
Division 02	Phase totale "825 points"	0.9369	7.6333
	Phase d'apprentissage70% "579 points"	0.9460	7.0095
	Phase de validations 15% "123 points"	0.9163	8.0292
	Phase de test 15% "123 points"	0.9175	9.7497
Division 03	Phase totale "825 points"	0.9449	7.1528
	Phase d'apprentissage 60% "495 points"	0.9622	5.8993
	Phase de validations20% "165 points"	0.9299	7.4765
	Phase de test20% "165 points"	0.9124	9.7727

IV.2.2. Performances des modèles QSAR-RN

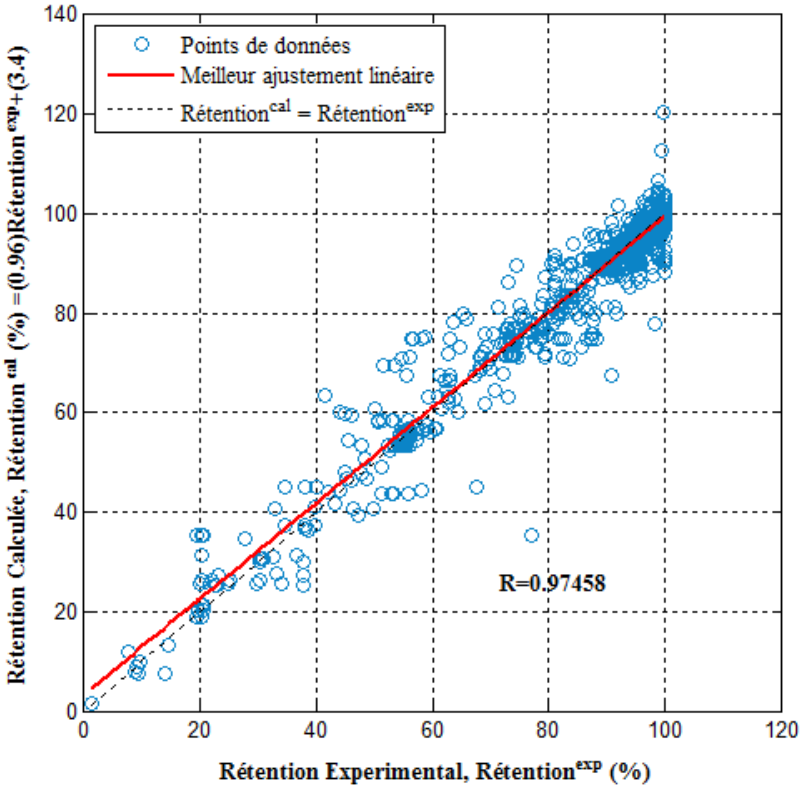
Le tableau IV.1 montre la structure de modèle neuronal QSPR-RN$_{optimal}$.

Table IV.5. Structure de modèle QSPR-RN optimal.

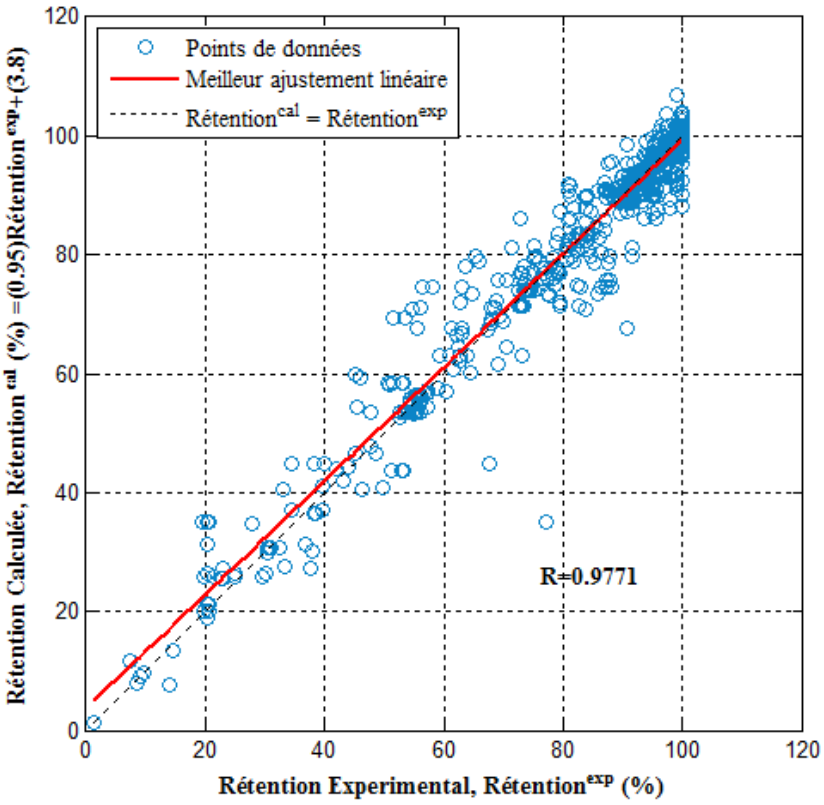
Algorithme d'apprentissage	Couche entrée	Couche cachée		Couche sortie	
	Nombre de neurone	Nombre de neurone	Fonction d'activation	Nombre de neurone	Fonction d'activation
BFGS quasi-Newton ***(trainbfg)***	13	18	Tanh	1	Identity

Selon la discussion précédente, un modèle neuronal QSPR-RN$_{optimal}$ a été développé dans le but de prédire la rétention des produits pharmaceutiques par les membranes d'OI/NF. L'intrigue et les paramètres de la régression linéaire sont franchement obtenus en utilisant la fonction MATLAB "***postreg***". Le pouvoir de prédiction d'un tel modèle peut être estimé par le coefficient de corrélation *R*, une valeur de *R* supérieure à 0.5 est généralement considérée comme satisfaisante, une valeur supérieure à 0.9 est excellente (Ammi 2016).

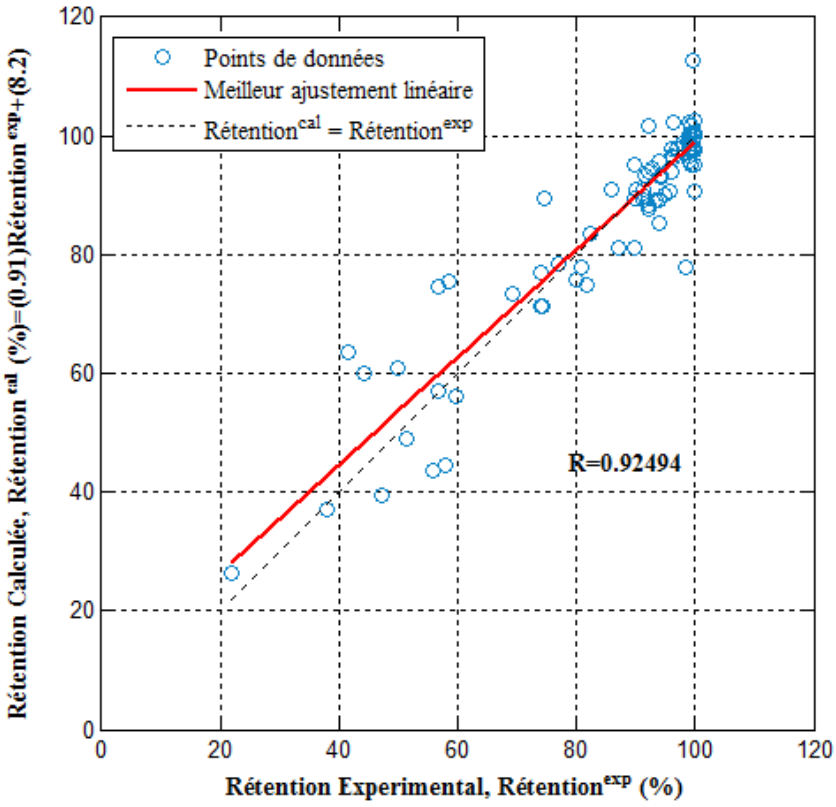
La figure IV.1 montre les courbes de régression linéaire des taux de rétentions calculés par modèle neuronal "QSAR-RN$_{optimal}$" optimisé avec les taux de rétentions expérimentales avec des vecteurs de régression approchant de l'idéal ([α, β, R] = [0.9591, 3.4385, 0.9746] "phase totale" ; [α, β, R] = [0.9564, 3.7688, 0.9771] "phase d'apprentissage" ; [α, β, R] = [0.9079, 8.2275, 0.9249] "phase de validation", et [α, β, R] = [1.0290, -2.2945, 0.9883] "phase de test"). Le coefficient de corrélation est couramment considéré comme excellente ($0.95 \leq R < 0.90$) pour toutes les phases, ce qui montre une bonne robustesse du modèle neuronal "QSAR-RN$_{optimal}$" établi et la possibilité de la prédiction des différents paramètres qui caractérisent la rétention des produits pharmaceutiques lors d'OI/NF.



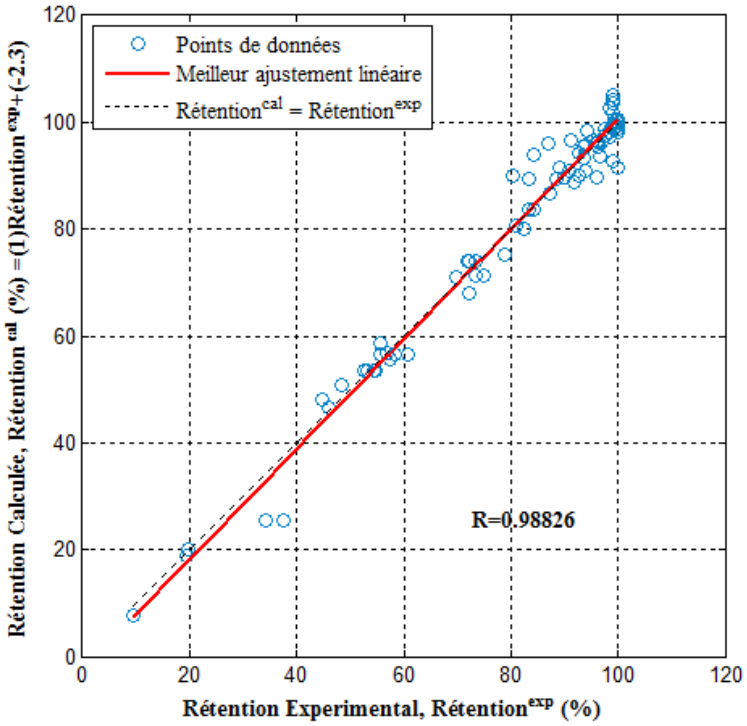
(a)


Points de données
Meilleur ajustement linéaire
Rétention^cal = Rétention^exp
R=0.9771
Rétention Calculée, Rétention^cal (%) =(0.95)Rétention^exp+(3.8)
Rétention Experimental, Rétention^exp (%)
0
20
40
60
80
100
120

(b)


Points de données
Meilleur ajustement linéaire
Rétention^cal = Rétention^exp
R=0.92494
Rétention Calculée, Rétention^cal (%)=(0.91)Rétention^exp+(8.2)
Rétention Experimental, Rétention^exp (%)
0
20
40
60
80
100
120

(c)



(d)

Figure IV.1. Comparaison entre les valeurs expérimentales et les valeurs calculées pour QSPR-$RN_{optimal}$: (a) la phase totale, (b) la phase d'apprentissage, (c) la phase de validation, (d) la phase de test.

La performance de modèle neuronal QSPR-$RN_{optimal}$ a été évaluée en termes de coefficient de corrélation (R), l'erreur absolue moyenne *(EAM*), la racine de l'erreur quadratique moyenne (*REQM*) et l'erreur de standard de prédiction (*ESP*) (Ammi et al. 2016) :

$$\mathrm{REQM} = \sqrt{\frac{\sum_{i=1}^{n}(Y_{i,exp} - Y_{i,cal})^2}{n}} \qquad (\mathrm{IV.1})$$

$$EAM = \frac{1}{n}\sum_{i=1}^{n}\left|(y_{i,exp} - y_{i,cal})\right| \qquad (\mathrm{IV.2})$$

$$ESP(\%) = \frac{REQM}{Y_e} * 100 \qquad (\mathrm{IV.3})$$

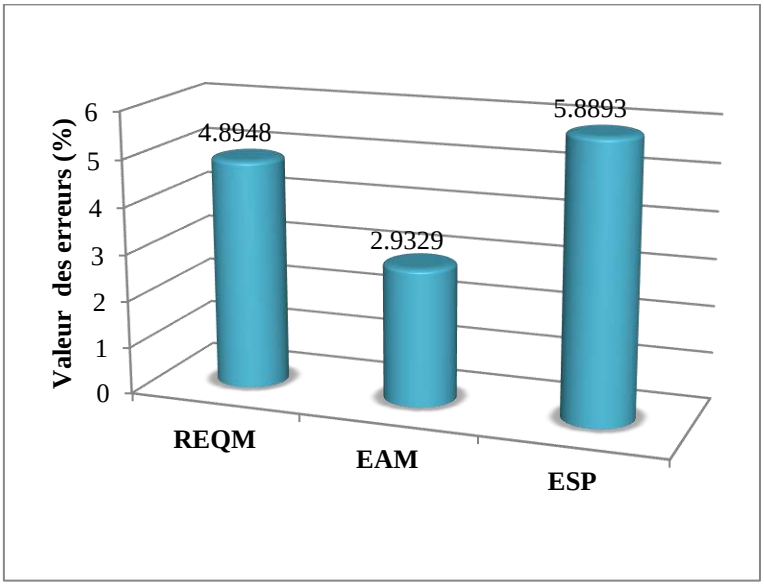
Où n est le nombre total de données, $y_{i,exp}$ est la valeur expérimentale, $y_{i,cal}$ représente la valeur calculée à partir du modèle QSPR-$RN_{optimal}$ et Y_e est la valeur moyenne des données expérimentales.

Les vecteurs de régression linéaire pour le modèle neuronal QSPR-RN$_{optimal}$ sont présentés dans le tableau IV.2. Il est prouvé que l'approche neuronale proposée donne des résultats excellents avec des valeurs de vecteur de régression approchant à l'idéal [équation linéaire : $y^{cal} = \alpha y^{exp} + \beta$, ***α*=1** (pente), ***β*=0** (ordonnée à l'origine), ***R*=1** (coefficient de corrélation)] dans l'ajustement des profils des QSPR-RN$_{optimal}$.

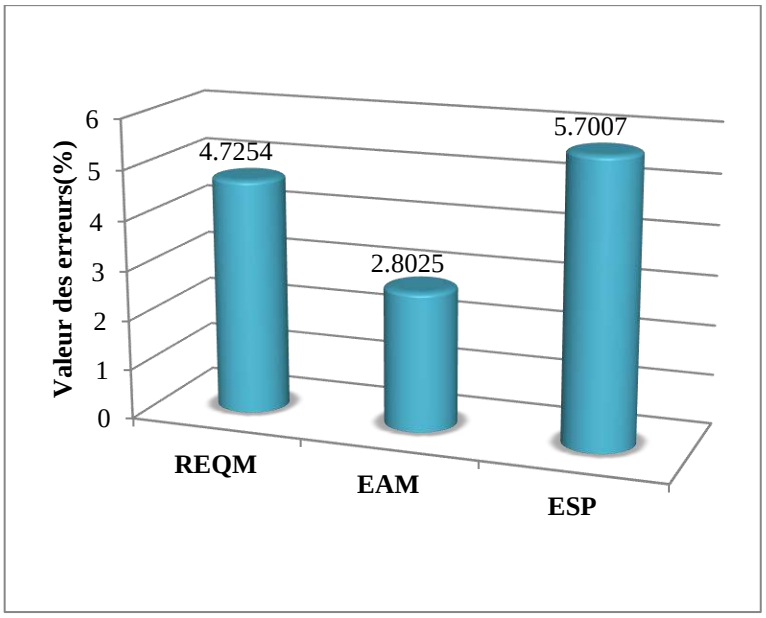
Table IV.6. Vecteurs de régression linéaire issue de l'équation linéaire : [équation linéaire : $y^{cal} = \alpha y^{exp} + \beta$, ***α***(pente), ***β***(ordonnée à l'origine), ***R***(coefficient de corrélation)] et la racine de l'erreur quadratique moyenne (***REQM***).

	α	***β***	***R***	***REQM (%)***
Phase d'apprentissage	0.9564	3.7688	0.9771	4.7254
Phase de validation	0.9079	8.2275	0.9246	6.9678
Phase de test	1.0290	-2.2945	0.9883	3.5360
Totale	0.9591	3.4385	0.9746	4.8948

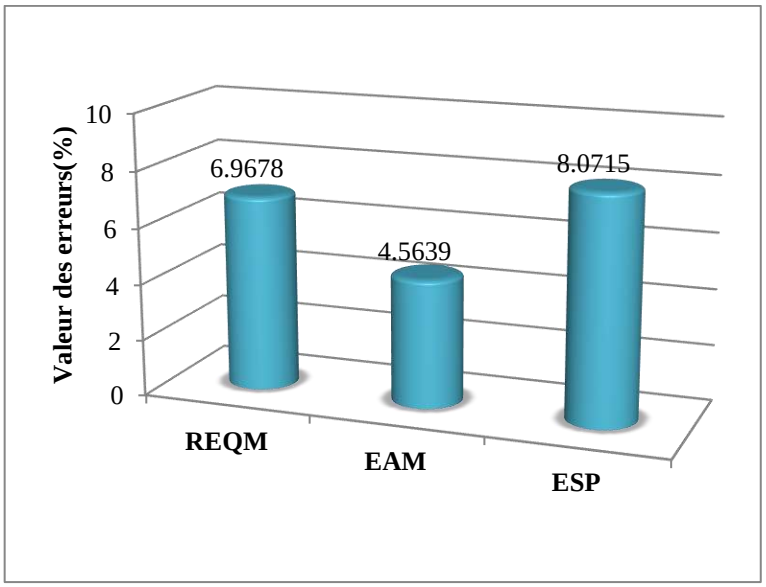
La figure IV.2 montre que le modèle neuronal " QSPR-RN$_{optimal}$" a donné des erreurs plus faibles. Ce qui indique que le modèle neuronal " QSPR-RN$_{optimal}$" possède une capacité de modélisation très élevée par rapport aux modèles.



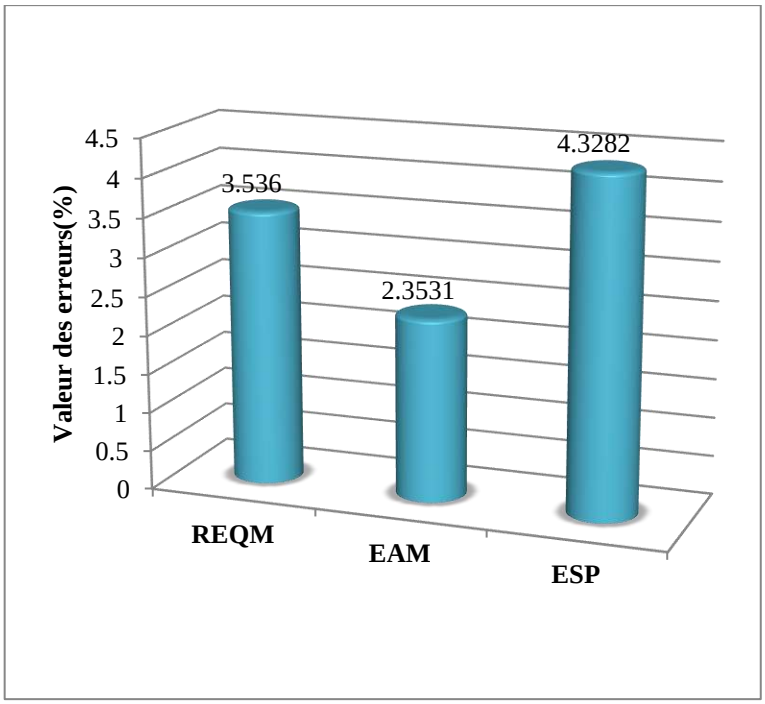
(a)


Valeur des erreurs(%)
0
1
2
3
4
5
6
4.7254
2.8025
5.7007
REQM
EAM
ESP

(b)


Valeur des erreurs(%)
0
2
4
6
8
10
6.9678
4.5639
8.0715
REQM
EAM
ESP

(c)



(d)

Figure IV.2. Performances de modèle neuronal QSPR-RN$_{optimal}$ pour : (a) la phase totale, (b) la phase d'apprentissage, (c) la phase de validation, (d) la phase de test.

IV.2.3. Expressions mathématiques de modèle neuronal optimal

À partir du QSPR-RN$_{optimal}$, représenté sur la figure IV.3, on peut exprimer l'assimilation de la rétention des produits pharmaceutiques par les membranes d'OI/NF par un modèle mathématique y (Rétention) incorporant toutes les entrées *Ei* (dc, *LogD*, moment dipolaire, *Length*, *EqWidth*, *MWCO*, SR(NaCl), angle de contact, potentiel zêta, pH, pression, taux de récupération et température) comme suit :

Les sorties Z_j de la couche cachée :

$$Z_j = f_H\left[\sum_{i=1}^{13} w_{ji}^I E_i + b_j^H\right] = \frac{exp\left(\sum_{i=1}^{13} w_{ji}^I E_i + b_j^H\right) - exp\left(-\sum_{i=1}^{13} w_{ji}^I E_i + b_j^H\right)}{exp\left(\sum_{i=1}^{13} w_{ji}^I E_i + b_j^H\right) + exp\left(-\sum_{i=1}^{13} w_{ji}^I E_i + b_j^H\right)} \qquad IV.4$$

j=1, 2, …,18

La sortie "rétention" :

$$Rétention\ de\ \text{PhACs}\ (QSPR - RN) = f_0\left[\sum_{j=1}^{18} w_{1j}^H Z_j + b_1^o\right] = \sum_{j=1}^{18} w_{1j}^H Z_j + b_1^o \qquad (IV.5)$$

La combinaison des équations (VI.4) et (VI.5) mène à la formule mathématique suivante qui décrit l'assimilation de rétention des produits pharmaceutiques par les membranes d'OI/NF en prenant en compte toutes les entrées *Ei* (dc, *LogD*, moment dipolaire, *Length*, *EqWidth*, *MWCO*, SR(NaCl), angle de contact, potentiel zêta, *pH*, pression, taux de récupération et température) :

$$Rétention\ des\ PhACs = \sum_{j=1}^{18} w_{1j}^{H} \frac{exp\left(\sum_{i=1}^{13} w_{ji}^{I} E_i + b_j^{H}\right) - exp\left(-\sum_{i=1}^{13} w_{ji}^{I} E_i + b_j^{H}\right)}{exp\left(\sum_{i=1}^{13} w_{ji}^{I} E_i + b_j^{H}\right) + exp\left(-\sum_{i=1}^{13} w_{ji}^{I} E_i + b_j^{H}\right)} + b_1^{o} \qquad (IV.6)$$

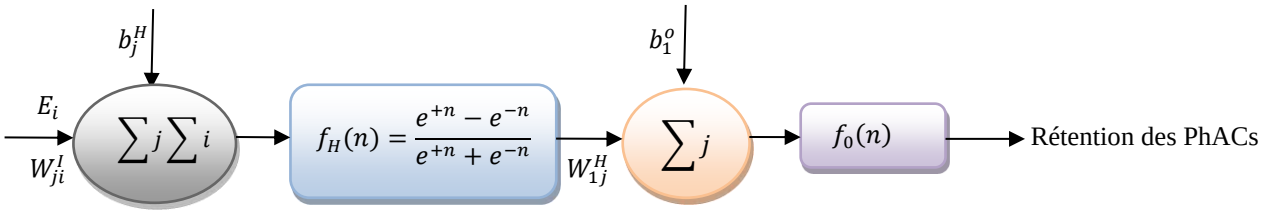


Figure IV.3. Représentation schématique du QSPR-RN$_{optimal}$.

Ou w_{ji}^{I} représente le poids synaptique de la connexion du neurone *i* de la couche d'entrée au neurone *j* de la couche cachée (18 lignes x 13 colonnes) ; w_{1j}^{H} représente le poids synaptique de la connexion du neurone *j* de la couche cachée au neurone de la couche de sortie (18 lignes x 1) ; b_j^{H} représente les biais du neurone *j* de la couche cachée (18 lignes) et b_1^{o} représente les biais du neurone *j* de la couche de sortie (1 ligne). Les matrices des poids et biais, sont donnés dans le tableau IV.5.

Tableau IV.7. Poids et biais du QSPR-RN$_{optimal}$.

Connexions Couche d'entrée-Couche cachée														**Connexions Couche cachée-Couche de sortie**	
Poids													**Biais**	**Poids**	**Biais**
$w_{j,1}^{I}$	$w_{j,2}^{I}$	$w_{j,3}^{I}$	$w_{j,4}^{I}$	$w_{j,5}^{I}$	$w_{j,6}^{I}$	$w_{j,7}^{I}$	$w_{j,8}^{I}$	$w_{j,9}^{I}$	$w_{j,10}^{I}$	$w_{j,11}^{I}$	$w_{j,121}^{I}$	$w_{j,131}^{I}$	b_j^H	$w_{I,j}^{H}$	b_1^0
-8,4817	42,0189	0,1339	-4,2694	-6,5117	-2,4078	15,9538	-9,0965	-31,6147	5,0037	-2,5141	-0,1607	3,1471	-6,6347	-0,1363	
13,0771	19,6317	5,6951	-3,3036	-1,8360	-6,9062	-3,5682	8,9484	-30,4789	-4,8484	0,0485	5,1489	-0,5849	-8,0011	2,5922	
2,5974	20,0007	-8,8248	2,0600	-14,5796	5,3340	-7,6499	-0,3614	44,0877	-4,3782	5,0230	6,4081	3,0025	-0,3187	-0,1143	
-13,2317	4,1193	-8,0739	-22,3344	0,0190	-4,5015	-0,8004	4,5893	24,7260	-3,5277	-3,3017	3,9659	-9,2759	-0,2884	0,0500	
4,8545	39,6750	-9,1102	-7,8421	9,3121	0,2869	2,2693	1,5786	14,3541	6,3155	-4,5255	2,0958	4,1987	18,6996	-0,1788	
-2,3199	0,2799	-2,6103	17,3527	9,2389	0,0495	20,1904	0,0092	-20,6337	-0,4726	-3,8090	0,0454	9,1887	-13,3202	0,1010	
-0,7764	14,7822	-2,4722	-29,1342	0,4486	21,0290	-6,3083	0,7620	-11,0010	1,3427	-9,9468	6,5593	-6,5992	-0,4076	-0,0724	
3,9694	6,5148	-6,6285	15,6351	0,4149	5,8396	9,0815	-3,6992	5,9047	-1,3619	-15,0288	4,3320	-2,8423	-11,4492	0,1117	
8,6375	1,7545	9,7696	-8,5591	-0,6323	6,6483	-2,4395	1,6111	-34,8984	3,3231	-6,8604	-1,2449	2,9527	16,4443	3,2194	0,2746
1,1693	0,5698	-3,1615	-5,4485	-1,8882	-1,5030	10,6576	-7,8963	19,4126	-3,5547	-1,1319	5,4023	-4,9818	12,5637	0,1058	
9,0033	-0,8966	-7,1364	3,9655	-8,0584	4,1722	-18,4372	-5,3372	24,9614	-5,7248	18,0635	1,2917	-2,1044	4,1265	0,2751	
5,0167	10,0555	7,4594	2,0575	-4,4566	0,0895	-5,8630	-0,5952	-18,5345	9,4290	0,2591	-2,8384	0,3300	0,7583	-0,2887	
-1,3138	-2,9279	10,9089	7,3427	-7,0461	5,1704	8,8451	1,2599	6,3896	3,5326	16,8742	12,6341	-0,8159	4,6160	0,0492	
13,5616	-2,0607	-0,3323	-0,2546	7,3927	-9,3906	-2,7868	5,3792	3,7164	-0,1131	6,9993	-18,8425	1,5148	3,4472	0,0051	
-12,0240	-0,5205	9,5201	-2,6629	-12,7560	3,2237	-15,5048	-5,1268	-0,4645	-10,4032	-21,6141	-1,9214	12,6654	5,4928	0,1672	
-29,4546	-8,4473	15,6767	11,3981	9,4208	5,1491	2,8176	-3,3666	-5,6721	-6,5377	5,8810	-5,6111	-0,9846	3,9192	0,5596	
-8,8183	6,9875	3,0282	0,0765	2,1588	5,7349	6,4129	1,6300	-0,8205	16,4637	8,7588	-5,8664	-2,1016	21,5159	-5,6665	
4,6117	-2,3964	-6,0852	1,6750	-5,2648	0,6501	-2,4504	-38,9743	-6,5475	3,3389	0,4515	3,4425	16,1341	-1,1692	-0,2569	

L'architecture des réseaux de neurones optimaux obtenus est donnée sur la figure IV.4, ci-après.

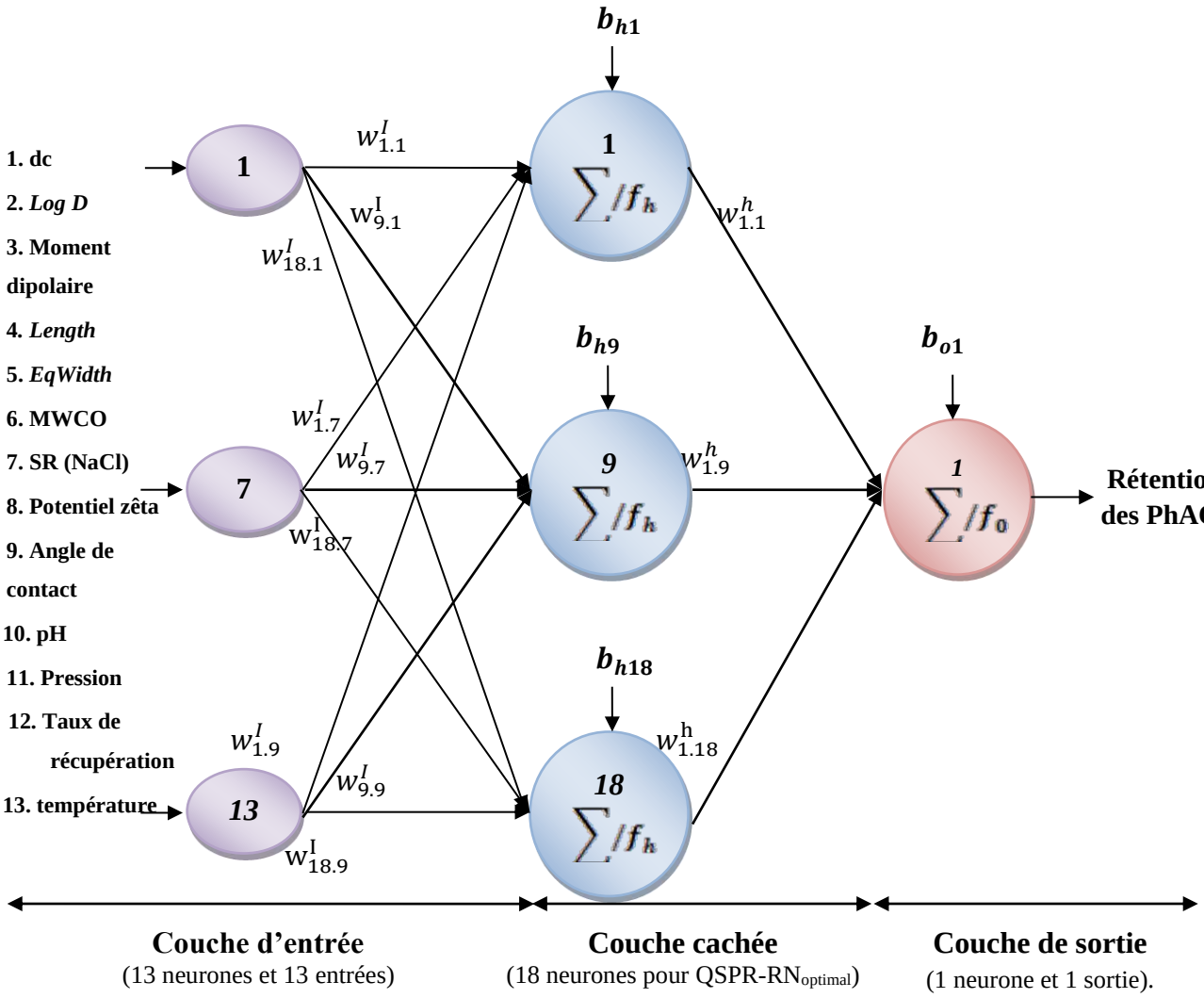


Figure IV.4. Architecture finale du réseau de neurone optimal de type feedforward pour la prédiction de la rétention des produits pharmaceutiques par les membranes d'OI/NF

IV.3. Analyse de sensibilité par la méthode de "poids"

Pour voir la contribution ainsi que le profil de variation de chaque variable d'entrée (dc, *LogD*, *Lenght*, *EqWidth*, MWCO, SR(NaCl), angle de contact, potentiels zêta, pH, pression, taux de récupération et température) sur la sortie de réseau QSPR-RN$_{optimal}$. On a procédé à une analyse de sensibilité par méthode de "poids".

Nous avons appliqué cette méthode au QSPR-RN$_{optimal}$ en utilisant leurs poids illustrés dans le tableau IV.7. Les résultats de contributions sont montrés sur la figure IV.5. Les variables les plus pertinentes

qui peuvent influer sur la rétention des produits pharmaceutiques par les membranes d'OI/NF sont :(dc, *LogD*, *Lenght*, *EqWidth*, MWCO, SR(NaCl), angle de contact, potentiels zêta, pH, pression, taux de récupération et température). Toutes les entrées exercent les contributions importantes sur la rétention des produits pharmaceutiques lors d'osmose inverse et nanofiltration.

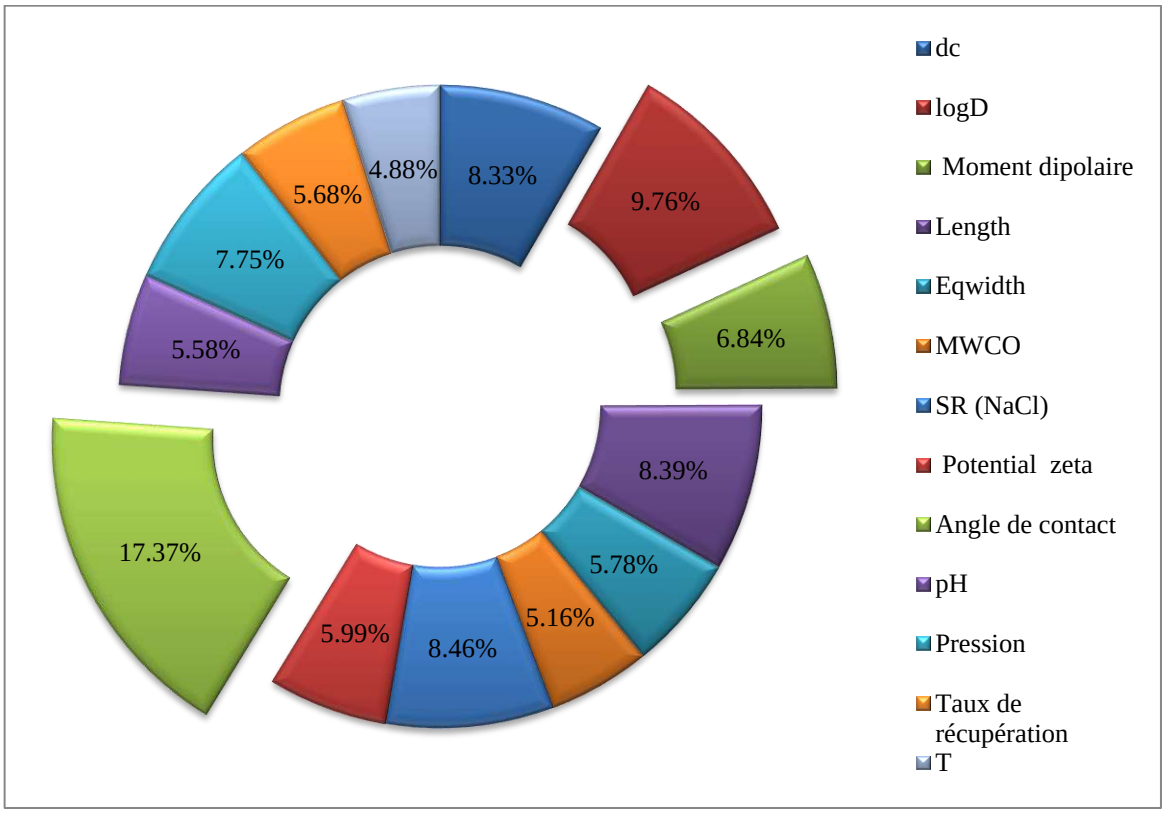


Figure IV.5. Présentation des pourcentages de l'importance relative (*IR*) de chaque variable d'entrée sur la rétention : QSPR-RN$_{optimal}$

D'après la figure IV.5, il est évident que le paramètre stérique/taille MWCO est moins efficace que le paramètre stérique/taille SR(NaCl) pour prédire la rétention des produits pharmaceutiques lors d'OI/NF (*IR* (MWCO) = 5.16%, *IR* (SR (NaCl)) = 8.46%). Les résultats de l'analyse de sensibilité par la méthode "poids" montrent qu'il est plus simple de caractériser une membrane en termes de SR(NaCl) plutôt que MWCO. Ils sont en accord avec les conclusions d'Ammi et al. 2017.

Ainsi, la figure IV.5 montre que les descripteurs moléculaires logD et moment dipolaire ont des importances relatives supérieur (*IR* (logD) = 9.75%, *IR* (moment dipolaire) = 6.84%) avec le paramètre de l'hydrophobicité de la membrane "angle de contact" (*IR* (angle de contact) = 17.37%). Ce ci indique que le moment dipolaire, logD et l'angle de contact sont des entrées plus pertinentes sur la rétention des PhACs lors d'OI/NF. On constate que la rétention des produits pharmaceutiques lors d'osmose inverse et nanofiltration dépend plus précisément des interactions

hydrophobe/adsorption que les autres interactions (encombrement strique et interactions électrostatiques).

IV.5. Diagnostic des points aberrants et domaine d'applicabilité du modèle neuronal optimisé

La figure IV.6. Montre le diagramme de William pour le domaine d'applicabilité pour le modèle neuronal optimal "QSPR-RN$_{optimal}$" (Phase totale). Dans ce diagramme l'existence de **798/825 (96.73%)** de points de données valides, 27/825 (3.27%) de points de données suspectes. La valeur critique de levier $H^* = \frac{3(n+1)}{m} = \frac{3(13+1)}{825} = 0.05$. Cela indique que le développement du modèle QSPR-RN$_{optimal}$ et ses prédictions sont effectués dans le domaine d'applicabilité Ce qui conduit à un modèle neural optimal statistiquement valide. Ainsi on peut affirmer qu'il existe des points "Good High Leverage" pour la phase totale.

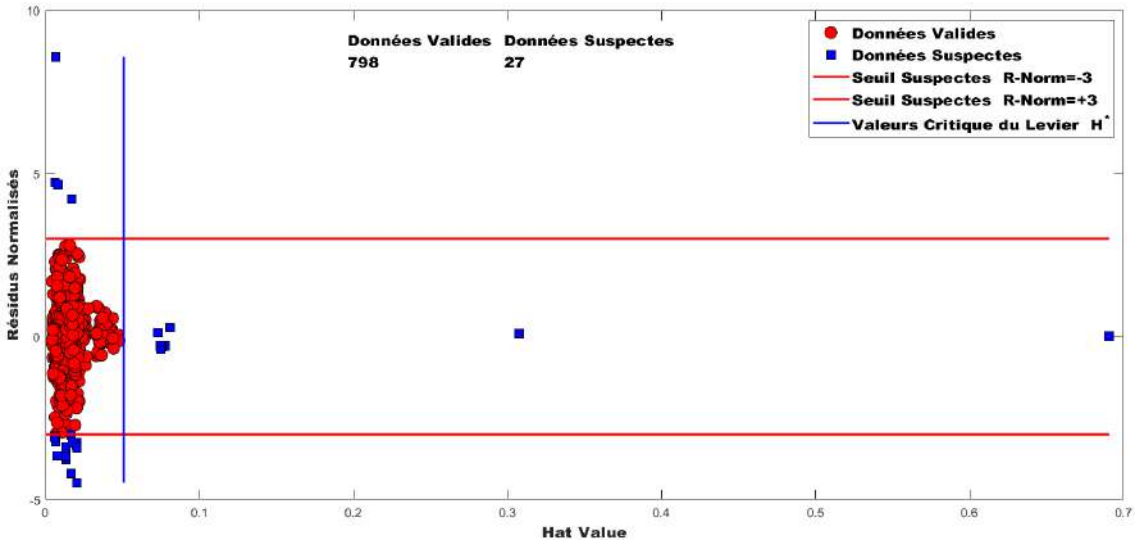


Figure IV.6. Diagnostic des données potentiellement suspectes et domaine d'applicabilité du modèle neuronal optimal "QSPR-RN$_{optimal}$" pour la phase totale .

La figure III.7. Montre le diagramme de William pour le domaine d'applicabilité pour le modèle neuronal optimal "QSPR-RN$_{optimal}$" (Phase de validation). Dans ce diagramme l'existence de **81/83 (97.59%)** de points de données valides, 2/83 (2.40%) de points de données suspectes. La valeur critique de levier est $H^* = \frac{3(n+1)}{m} = \frac{3(13+1)}{83} = 0.506$. Cela indique que le développement du modèle QSPR-RN$_{optimal}$ et ses prédictions sont effectués dans le domaine d'applicabilité. Ce qui conduit à un modèle neural optimal statistiquement valide. Ainsi on peut affirmer qu'il existe des points "Good High Leverage" pour la phase validation.

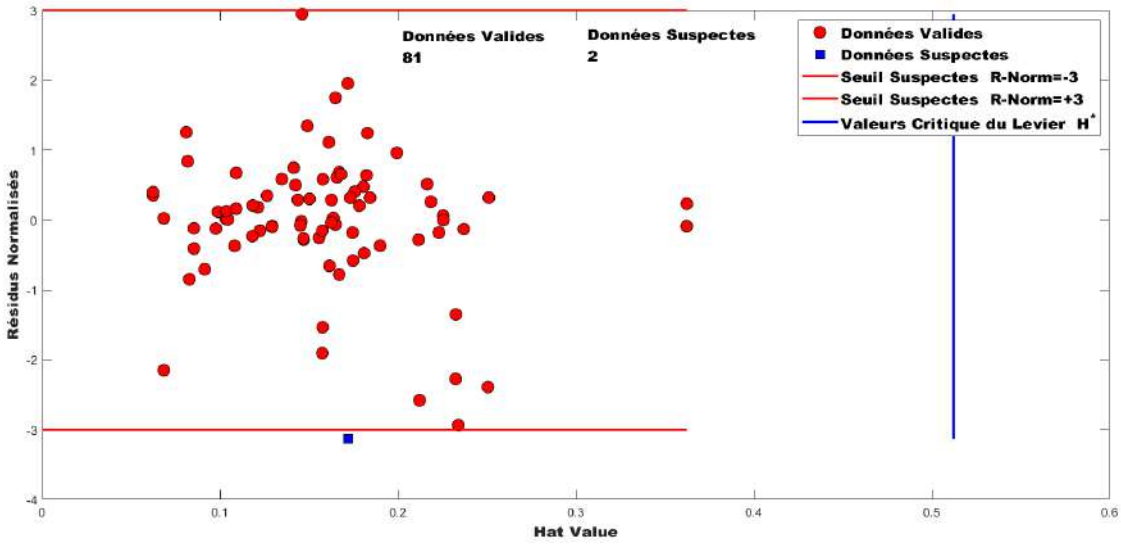


Figure IV.7. Diagnostic des données potentiellement suspectes et domaine d'applicabilité du modèle du modèle neuronal optimal "QSPR-RN$_{optimal}$" pour la phase validation.

La figure III.9. Montre le diagramme de William pour le domaine d'applicabilité pour le modèle neuronal optimal "QSPR-RN$_{optimal}$" (Phase de test). Dans ce diagramme l'existence de **80/82 (97.56%)** de points de données valides, 2/82 (2.43%) de points de données suspectes. La valeur critique de levier $H^* = \frac{3(n+1)}{m} = \frac{3(13+1)}{82} = 0.51$. Cela indique que le développement du modèle QSPR-RN$_{optimal}$ et ses prédictions sont effectués dans le domaine d'applicabilité. Ce qui conduit à un modèle neural optimal statistiquement valide. Ainsi on peut affirmer qu'il existe des points "Good High Leverage" pour la phase de test.

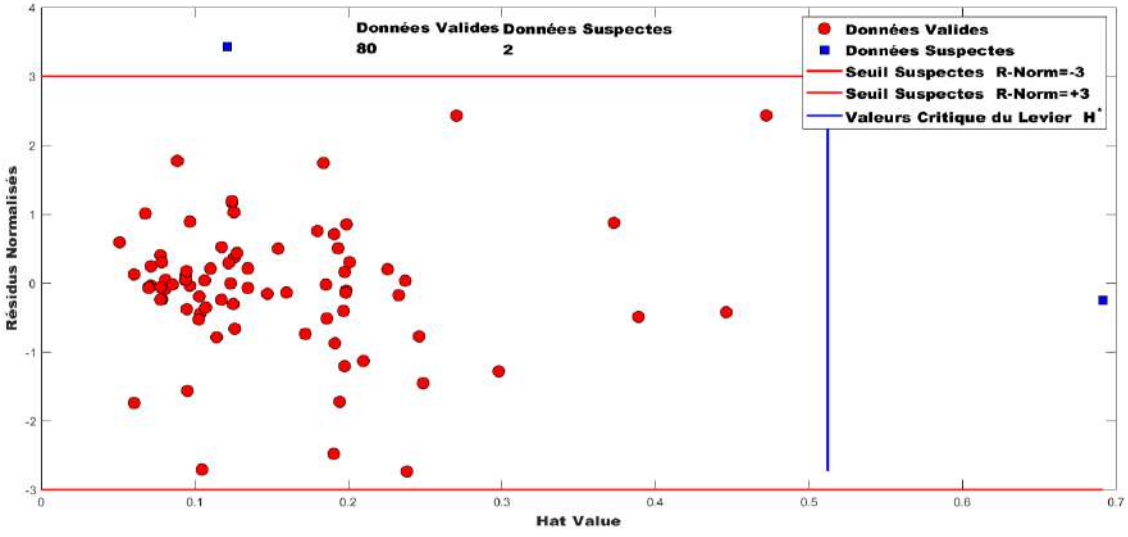


Figure IV.8. Diagnostic des données potentiellement suspectes et domaine d'applicabilité du modèle du modèle neuronal optimal "QSPR-RN$_{optimal}$" pour la phase de test.

IV.6. Conclusion

Ce présent chapitre a donné une raison de plus pour l'utilisation généralisée des modèles QSPR-RN pour la modélisation de la rétention des produits pharmaceutiques par les membranes d'osmose inverse et de nanofiltration. Le but de la présente étude était de développer quatre modèles QSPR-RN de type feed-forward capables de résumer les interactions entre les caractéristiques de la membrane, les conditions d'opératoire d'adsorption et les propriétés physico-chimiques des produits pharmaceutiques.

Le modèle QSPR-$RN_{optimal}$ développé dans ce travail a donné des erreurs plus faibles. Les erreurs absolues moyennes étaient 2.9329% pour phase totale, 2.8025% pour la phase apprentissage, 4.5639% pour phase validation et 2.3531% pour phase test. Les racines des erreurs quadratiques moyennes étaient 4.8948% pour phase totale, 4.7254% pour phase apprentissage, 6.9678% pour la phase validation et 3.5360% pour la phase test. Les erreurs standards de prédictions étaient 5.8893% pour phase totale, 5.7007% pour phase apprentissage, 8.0715% pour phase validation et 4.3282% pour phase test.

L'analyse de sensibilité par la méthode de "poids" a révélé que les variables les plus importantes qui influent sur la prédiction de la rétention des produits pharmaceutiques sont : dc, *LogD*, moment dipolaire, *Length*, *Eqwidth*, MWCO, SR (NaCl), potentiel zêta, angle de contact, pH, pression, taux de récupération et température. Le moment dipolaire, logD et l'angle de contact sont des entrées plus pertinentes sur la rétention des PhACs lors d'OI/NF. La rétention des produits pharmaceutiques lors d'osmose inverse et nanofiltration dépend plus précisément des interactions hydrophobe/adsorption que les autres interactions (encombrement strique et interactions électrostatiques).

Analyse diagnostic des points aberrants et domaine d'applicabilité du modèle neuronal optimisé a montré que le développement du modèle QSPR-$RN_{optimal}$ et ses prédictions sont effectués dans le domaine d'applicabilité. Ce qui conduit à un modèle neural optimal statistiquement valide. Ainsi on peut affirmer qu'il existe des points "Good High Leverage" pour la phase de test.

Conclusion générale et perspectives

Les réseaux de neurones possèdent une propriété remarquable dont l'apport est manifesté pour la modélisation des processus non linéaires : ce sont des approximateurs universels parcimonieux. Les modèles des réseaux neurones artificiels (RNA) ont finalement été appliqués pour la prévision de séparation des substances organiques par les membranes d'osmose inverse et de nanofiltration qui ont été abordés par d'autres chercheurs d'une façon réductible dont la taille de la base de données et les paramètres calculés ont donné des résultats encourageants.

L'objectif de ce mémoire a porté sur l'utilisation des réseaux de neurones pour la modélisation du processus de la rétention des produits pharmaceutiques par les membranes d'osmose inverse et de nanofiltration. Pour cela quatre modèles QSPR neuronaux de type feed-forward. Il est capable de résumer les interactions entre les caractéristiques de la membrane, les conditions opératoires de filtration et les propriétés des produits pharmaceutiques. Les quatre modèles neuronaux de type feedforward (QSPR-RN) caractérisés par une structure similaire (treize neurones dans la couche d'entrée, dix-huit neurones dans la couche cachée et un neurone dans la couche de sortie) sont construits dans le but de prédire la rétention des produits pharmaceutiques. Une bonne concordance, entre les valeurs expérimentales et les valeurs prédites, pour chaque modèle neuronal a été obtenue (la racine d'erreur quadratique moyenne pour la base de données totale étaient 4.8948%.

D'autre part, l'analyse de sensibilité par la méthode de "poids" a révélé que les variables les plus importantes qui influent sur la prédiction de la rétention des produits pharmaceutiques sont : dc, *LogD*, moment dipolaire, *Length*, *Eqwidth*, MWCO, SR (NaCl), potentiel zêta, angle de contact, pH, pression, taux de récupération et température. Le moment dipolaire, logD et l'angle de contact sont des entrées plus pertinentes sur la rétention des PhACs lors d'OI/NF. La rétention des produits pharmaceutiques lors d'osmose inverse et nanofiltration dépend plus précisément des interactions hydrophobe/adsorption que les autres interactions (encombrement strique et interactions électrostatiques).

Analyse diagnostique des points aberrants et domaine d'applicabilité du modèle neuronal optimisé a montré que le développement du modèle QSPR-$RN_{optimal}$ et ses prédictions sont effectués dans le domaine d'applicabilité. Ce qui conduit à un modèle neural optimal statistiquement valide. Ainsi on peut affirmer qu'il existe des points "Good High Leverage" pour la phase de test.

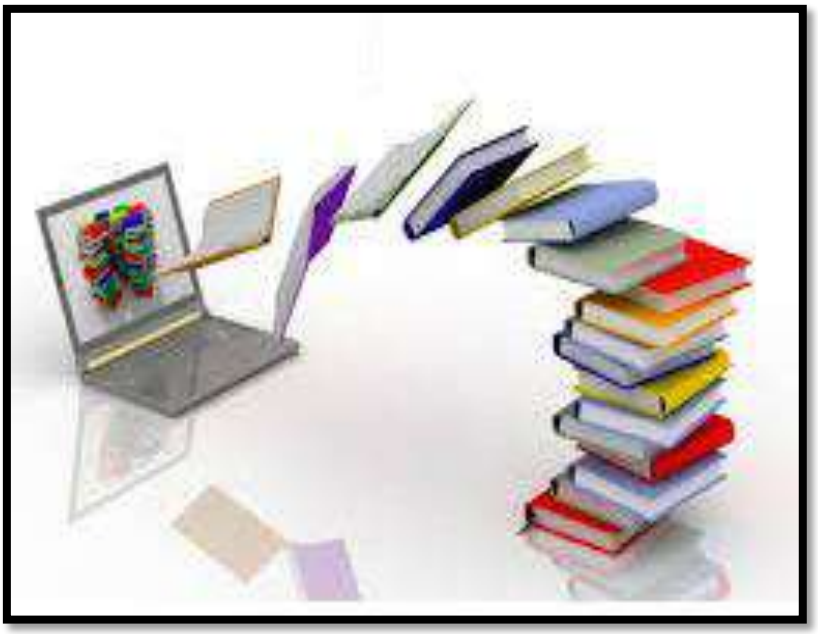
À la lumière des résultats obtenus, on peut tirer les conclusions suivantes :

- Les réseaux de neurones sont des outils statistiques, qui permettent d'ajuster des fonctions non linéaires très générales à des ensembles de points comme toute méthode statistique ;
- Les réseaux de neurones fonctionnent comme une boîte noire et les paramètres du réseau (poids et biais) n'ont pas de signification physique. L'application des réseaux neuronaux ne développe pas directement de la théorie de description du système, mais rend faisable la modélisation et la simulation des procédés d'osmose inverse et de nanofiltration ;
- La modélisation par réseau de neurones est une méthode qui est basée seulement sur des données expérimentales et/ou semi-expérimentales (déduites des modèles analytiques, empiriques, numériques, …), donc la qualité et la quantité de données décident de l'exactitude et la robustesse du modèle ;
- Les réseaux de neurones peuvent être très robustes et performants même en présence de peu de données expérimentales à condition que ces dernières soient statistiquement bien réparties dans le domaine d'entrée du modèle neuronal, ce qui est le cas avec le plan composite centré ;
- Les formules mathématiques extraites des modèles neuronaux peuvent trouver des applications intéressantes dans la prédiction, l'optimisation et même le contrôle des procédés d'osmose inverse et de nanofiltration.

De ce travail de mémoire ont émergé de nombreuses pistes de recherche qui feront l'objet de futurs travaux.

- ✓ Les perspectives concernent plus particulièrement la modélisation de la rétention des produits pharmaceutiques par les membranes d'OI/NF, mais en utilisant d'autres indicateurs de porosité, on propose de faire une étude avec) le rejet de sel SR($CaCl_2$).
- ✓ Modéliser d'effet d'interactions sur la rétention des produits pharmaceutiques lors d'OI/NF : encombrement stérique et répulsion électrostatique.
- ✓ Les perspectives concernent plus particulièrement la modélisation d'autres processus (microfiltration, ultrafiltration) pour des fins de prédiction.
- ✓ Nous pouvons aussi étendre l'idée de ce travail à d'autres méthodes de l'intelligence artificielle (Support vecteur machine, Réseaux de neurones multiples,…).

Références Bibliographiques



Agenson K. O., Oh J. I. et Urase T. (2003). "Retention of a wide variety of organic pollutants by different nanofiltration /reverse osmosis membranes : controlling parameters of process". Journal of Membrane Science. **225** (1-2) : 91-103.

Audinos Rémy.(2000)."Membranes semi-perméables: Membranes d'osmose inverse, Techniques de l'Ingénieur", traité constantes physico-chimiques K362, 2000.

Altalyana N., Jonesa B., Bradda J., Nghiemb L. D. et Alyazichia Y. M. (2016)."Removal of volatile organic compounds (VOCs) from groundwater by reverse osmosis and nanofiltration". Journal of Water Process Engineering. 9 : 9-21.

Ammar M. Y. (2007). "Mise en oeuvre de réseaux de neurones pour la modélisation de cinétiques réactionnelles en vuede la transposition batch/continu." Thèse de Doctorat, Institut National Polytechnique de Toulouse. Toulouse,France.

Ammi. Y (2016). "Modélisation des phénomènes de transport des phénomènes de transport lors d'osmose inverse" Thèse de doctorat. Université de Médéa. Médéa. Algérie.

Annane K. (2015). "Extraction et transport des ions de cadium (II) par membranes d'affinité dans des dispositifs micro fluidiques". Thèse de doctorat. Université de Mouloud Mammeri. Tizi-Ouzou. Algérie.

Antoine B. (2004). "Etude de la nanofiltration pour son intégration dans les procédés dans de production d'acides organiques". Thèse de doctorat. Université Paul Sabatier. Toulouse. Françe.

Aptel Ph. et Buckley C. A. (1996)."Categories of membrane operations, In : Water Treatment Membrane Process" Chapitre 2, McGraw-Hill : 24.

Ardhaoui M., Cherichi H., Ognier S., Ghoul M.(2003)."Caractérisation des propriétés de surface des membranes de nanofiltration: développement d'un outil d'aide à la décision." Récents Progrès en Génie des Procédés 90 (SFGP 2003, St Nazaire). 32–39 (2003).

Broomhead D. S. and Lowe D.(1988)."Radial basis functions, multi-variable functional interpolation and adaptive networks"," DTIC Document1988.

Baker R. W. (2004). "Memebrane Technology and application". 2nd ed, John Wiley and Sons Chichester : 287-295.

Bellona C., Drewes J. E., Xu P. et Amy G. (2004). " Factors Affecting the rejection of organic solutesduring NF/RO treatment - a literature review ". Water Research. 38 (12) : 2795-2809.

Bellona C. et Drewes J. E. (2005)."The role of membrane surface charge and solute physico-chemical properties in the rejection of organic acids by NF membranes". Journal of Membrane Science. **249** (1) : 227-234.

Bermejo S. et Cabestany J. (2001). "Learning with nearest neighbour classifiers." Neural ProcessingLetters13(2):159-181.

Bishop C. M. (1995). "Neural networks for pattern recognition." Oxford University Press, New York, USA.

Bouranene, S. (2008). "Etude de la rétention de soluté neutres et ioniques par des membranes de nanofiltration". Thèse de doctorat. Université de Franche-Comté. École Doctorale Louis Pasteur. Besançon. France.

Coutoa C.F, Santosa A.V,Míriam Cristina Santos Amarala , LiséteCelina Langea , Laura Hamdan de Andradeb , Ana Flávia Souza Foureauxa , Bárbara Stella Fernandesa.(2020). "Assessing Potential of nanofiltration, reverse osmosis and membrane distillation drinking water treatment for pharmaceutical active compounds (PhACs) removal" Journal of Water Process Engineering 33 (2020) 101029.

Carpenter G. A. et Grossberg S. (1988). "The ART of adaptive pattern recognition by a self-organizing neural network." Computer 21(3): 77-88.

Chabane M. (2015)."Étude de la résistance au colmatage et de la rétention des composes toxiques par membranes". Thèse de doctorat. Université d'Aboubekr Belkaid de Tlemcen. Tlemcen. Algérie.

ChabaaS.(2011)." Identification des Systèmes non Linéaires en Utilisant les Techniques d'Intelligences Artificielles et les Bases de Fonctions de Laguerre pour la Modélisation des Données du Trafic dans les Réseaux Internet."Thèse de Doctorat, Université Cadi Ayyad, Faculté des Sciences Semlalia -Marrakech, 2011

Chamekh A. (2007). "Optimisation des procédés de mise en forme par les réseaux de neurones artificiels." Thèse de Doctorat, Université d'Angers. Angers, France.

Chikhi M. (2008). "Etude de la compléxation des cations métalliques en vue de leur séparation par un procédé membranaire". Thèse de doctorat. Université Mentouri. Constantine. Algérie.

Chtita S. (2017). Bouachrine M, Lakhli T " Modélisation de molécules organiques hétérocycliques biologiquement actives par des méthodes QSAR/QSPR." Recherche de nouveaux médicaments Chimie. Faculte des sciences, Meknes 2017. Francais

Dames M. (2005). "Gestion de procèdes discontinus : Méthodologie de modélisation et d'optimisation d'opérations dedispersion liquide-liquide en cuve agitée." Thèse de Doctorat, Institut National Polytechnique de Toulouse. Toulouse, France.

Davor D., Ana V., Danijela A. et Kresimir., K. (2011). "Effect of water matrices on removal of veterinary pharmaceuticals by nanofiltration and reverse osmosis membranes". Journal of Environmental Sciences. 23 (8) : 1299–1307.

Davor D., IgnjaticZ. T., Kosutic K., Asperger D. et MutavdzicPavlovic D. (2012). " RO/NF membrane treatment of veterinary pharmaceutical wastewater:comparison of results obtained on a laboratory and a pilot scale Environmental science and pollution". Research International. 19 (4) :1033-1042.

Davor D., Zokić T. I., Košutić K., Periša M. et Babić S. (2013)." Photolysis of enrofloxacin and removal of its photodegradation products from water by reverse osmosis and nanofiltration membranes". Separation and Purification Technology. 115 (13) : 1-8.

Davor D, Nina D. S, Kre_simirKo_sutic, Irena _Skoric ,DanijelaA_sperger.(2016)." Adsorption of hydrophilic and hydrophobic pharmaceuticals on RO/NF membranes: Identification of interactions using FTIR.

Desmukh S. S. et Childress A. E. (2001). "Zeta potential of commercial RO membranes : influence of source water type and chemistry". Desalination. **140** (1) : 87-95.

Donghan G., Yuhong D. et Ruiliang Y. (2016). "Performance analysis of an organic Rankine cycle for a reverse osmosis desalination system using azeotropic mixtures ". Desalination. 381 : 38–46.

Dreyfus G. (2008). "Réseaux de neurones: Méthodologie et applications." Eyrolles, Paris, France.

Davalo E., NaïmP.(1993)."Des Réseaux de neurones" Edition Eyrolles 1993.

Garcia-Ivars J, Martella L, Massella M, Carbonell C -Alcaina, Miranda M.I.A, Iborra-Clar M.I.(2017)." Nanofiltration as tertiary treatment method for removing trace pharmaceutically active compounds in wastewater from wastewater treatment plants"

Garson G. D. (1991). "Interpreting neural-network connection weights". Artificial Intelligence Expert. 6 (4) : 46-51.

Gevrey M., Dimopoulos I. et Lek S. (2003). "Review and comparison of methods to study the contribution of variables in artificial neural network models". EcologicalModelling. 160 (3) : 249-264.

Girard B., Fukumoto L. R. (2000). "Membrane processing of fruit juices and beverages : areview". Critical Reviews in Biotechnology. 20 (2) : 109-175.

Gumi T., Valiente M., Khulbe K. C., Palet C. et Matsuura T. (2003). "Characterization of activated composite membranes by solute transport, contact angle measurement, AFM and ESR". Journal of Membrane Science **212** (1-2) : 123-134.

Goh A. T. C. (1995). "Back-propagation neural networks for modeling complex systems ". Artificial Intelligence in Engineering. 9 (3) : 143-151.

Hamadache M. (2014). "Modélisation en vue de la prédiction ou de la corrélation de l'activité toxicologique à partir de la structure moléculaire". Thèse de doctorat. Université de Médéa. Médéa. Algérie.

Hamlaoui M. L., Kherrat R., Marrakchi M., Jaffrezic-Renault N. et Walcarius A. (2002). "Development of an ammonium ISFET sensorwith a polymeric membrane includingzeolite". Materials Science and Engineering 21 (1-2) : 25-28.

Haykin S. S. (1999). "Neural networks: A comprehensive foundation." Prentice Hall, New Jersey, USA.

Hong S. et Elimelech M. (1997). "Chemical and physical aspects of natural organic matter (NOM) fouling of nanofiltration membranes". Journal of Membrane Science 132 (2) : 159-181.

Hosseinzadeh, M. and A. Hemmati-Sarapardeh (2014). "Toward a predictive model for estimating viscosity of ternary mixtures containing ionic liquids." Journal of MolecularLiquids200: 340-348.

Huang H., Cho H., Schwab K. et Jacangelo J. G. (2011). " Effects of feed water pretreatment on the removal of organic microconstituents by a lowfouling reverse osmosis membrane". Desalination. 281: 446–454.

Hmamouchi Rachid, Mohammed Bouachrine2 et Tahar Lakhlifi1(2016). "Tentative Pratique du Relation Quantitatives Structure-Activité/Propriété (QSAR/QSPR)" . Revue Interdisciplinaire

Khaouane L. (2013). "Etude et Modélisation de la Biosynthèse des Antibiotiques à partir de Différentes Souches Productrices – Cas de Pleuromutiline". Thèse de doctorat. Université de Médéa. Médéa. Algérie.

Kiso Y., Muroshige K., Oguchi T., Yamada T., Hhirose M., Ohara T. et Shintani T. (2010). "Effect of molecularshape on rejection of unchargedorganic compounds by nanofiltration membranes and on calculated pore radii". Journal of Membrane Science. 358 (1-2) : 101–113.

Kosutic K., Kastelan-Kunst L. et Kunst B. (2000). "Porosity of some commercial reverse osmosis and nanofiltration polyamide thin-film composite membranes". Journal of Membrane Science. 168 (1-2) : 101-108.

Kohonen T.(1982). "Self-organized formation of topologically correct featuremaps," Biologicalcybernetics, vol. 43, pp. 59-69, 1982.

Lanteri Y. (2009). "Transport a travers des membranes de nanofiltration : caractérisation des propriétés électronique". Thèse de doctorat. Université de Franche-COMTE. Besançon. France.

Libotean D., Giralt J., Giralt F., Rallo R., Wolfe T. et Cohen Y. (2009). " Neural network approach for modeling the performance of reverse osmosis membrane desalting". Journal of Membrane Science. 326 (2) : 408-419.

Liconaa K.P.M., GeaquintobL.R. de O., NicolinicJ.V., FigueiredobN.G., ChiapettabS.C.,HabertcA.C., YokoyamaaL.(2018)." Assessing Potential of nanofiltration and reverse osmosis for removal of toxic pharmaceuticals from water".Journal of Water Process Engineering".Journal of Water Process Engineering 25 (2018) 195–204.

Lin S. H. et Yang C. R. (2004). ""Chemical and physicalpretreatment of chemicalmechanicalpolishingwastewaterfromsemiconductor fabrication". Journal of Hazardous Materials. 108 (1-2) : 103-109.

Linares R.V., Yangali-Quintanilla V., Li Z. et Amy G. (2011). " Rejection of micropollutants by clean and fouled forward osmosis membrane ". Water Research. 45 (20) : 6737 -6744.

McClelland J. L., RumelhartD. E., and GroupP. R. (1986). "Parallel distributed processing.(1986)." Explorations in the microstructure of cognition, vol. 2, 1986

MSAAF M, Balmajdoub F.(2015)." L'application des réseaux de neurones de type « feedforward » dans le diagnostic statique." Laboratoire des techniques industriel Faculté des sciences et techniques Université Sidi Mohammed Ben Abdellah Fès, Maroc.

Metaiche M. (2014). "Technologie membranaire". Faculté des sciences et des sciences appliquées de l'université de Bouira, Algérie. Chapitre 1: 4-6.

Nascimento C. A. O., Giudici R. et Guardani R. (2000). "Neural network based approach for optimization of industrial chemical processes." Computers & Chemical Engineering 24(9-10): 2303-2314.

Poezevara G. (2011)."Fouille de graphes pour la découverte de contrastes entre classes : application à l'estimation de la toxicité des molécules". Spécialité Informatique et applications, Thèse de doctorat, École Doctorale SIMEN , Université de Caen Basse-Normandie.

Renata Z.Y., Magdalena F.C., Irena K.S., Marcin K., Jacek B., Stanisław L.(2022). "Impact of Polymer Membrane Properties on the Removal"Membranes 2022, 12, 150

Rosenblatt F.(1958). "The perceptron: a probabilistic model for information storage and organization in the brain," Psychology Review, vol.65, p. 386, 1958.

.

Sadmani A. H. M. A., Andrews R. C. et Bagley M. D. (2014a). "Impact of natural water colloids and cations on the rejection of pharmaceutically active and endocrine disrupting compounds by nanofiltaration". Journal of Membrane Science. 450 : 272-281.

Sadmani A. H. M. A., Andrews R.C. et Bagley M. D. (2014b)."Nanofiltration of pharmaceutically active and endocrine disrupting compounds as a function of compound interactions with DOM fractions and cations in natural water". Separation and Purification Technology. 122 : 462–471.

Samelides H., BouretP., ReggiaJ.(1991). Réseaux neuronaux. Une approche connexionniste de l'intelligence artificielle. Édition TEKNEA 1991.

Sagne C. (2008). "Etude des mécanismes de transfert de molécules organiques en osmose inverse". Thèse de doctorat. École nationale supérieure de chimie de Montpellier. Montpellier. France.

Santos J. L. C., de Beukelaar P., Vankelecom I. F. J., Velizarov S. et Crespo J. G. (2006). "Effect of solutegeometry and orientation on the rejection of uncharged compounds by nanofiltration". Separation and Purification Technology. 50 (1) : 122–131.

Shim Y., Lee H. J., Lee S., Moon S. H. et Cho J. (2002). "Effects of natural organic matter and ion species on membrane surface charge". Environnemental Science and Technology. **36** (17) : 3864- 3871.

Si-Moussa C. (2009). "Modélisation des équilibres de phases a pressions élevées par les réseaux de neurones artificiels". Thèse de doctorat. École nationale supérieure polytechnique. Alger. Algérie.

Sridhar S., Kale, A., Khan A. A. (2002). "Reverse osmosis of edible vegetable oil industry effluent". Journal of Membrane Science. 205 (1-2) : 83-90.

Stone M. (1977). "An asymptoticequivalence of choice of model by cross-validation and Akaike scriterion." Journal ofthe Royal Statistical Society B 39(1): 44-47.

Taheran, M., Brar, S.K.;Verma, M.,Surampalli, R., Zhang, T.,Valero, J.(2016). "Membrane processes for removal of pharmaceutically active compounds (PhACs) from water and wastewaters." Sci. Total Environ. 2016, 547, 60–77.

Tamas Adrian Paul.(2004). Etude comparée du colmatage en Nano- filtration et en ultrafiltration d'eau de surface, Mémoire de Maîtrise, Université Laval, 2004.

Talbi M. L. (2011). "Analyse et traitement du signal électrocardiographique (ECG)." Thèse de Doctorat, Université Mentouri de Constantine. Constantine, Algérie.

Triki Z. (2014). "Études, Analyses et Optimisation de la Consommation Énergétique des Unités de Dessalement pour les Sites Isolés". Thèse de doctorat, Université Constantine 1. Constantine. Algérie.

Van der Bruggen B., Schaep J., Maes W., Wilms D. et Vandecasteele C. (1998). " Nanofiltration as a treatment method for the removal of pesticides from ground waters". Desalination. 117: 139-147.

Verliefde A. R. D., Cornelissen E. R., Heijman S. G. J., VerberkJ .Q. J. C., Amy G. L., Van der Bruggen B. et van Dijk J. C.(2008). "The role of electrostatic interactions on the rejection of organicsolutes in aqueous solutions with nanofiltration". Journal of Membrane Science. 322 (1) : 52–66.

Wayne Schultz T, Cronin, M.T.D., Walker, J.D., and Aptula, A.O. (2003). "Quantitative structure–activity relationships (QSARs) in toxicology: ahistorical perspective", JournalofMolecular Structure (Theochem), 622, 1–22

Xu R., Pan Z., Qian W., Xiaomao W., Kaichang Y, Tao X., Xianghua W.(2019)." Influences of multi influent matrices on the retention of PPCPs by nanofiltration membranes. Separation and Purification Technology 212 (2019) 299–306

Yangali-Quintanilla V., Sadmani A., McConville M., Kennedy M. et Amy G. (2009a). "Rejection of pharmaceutically active compounds and endocrine disrupting compounds by clean and fouled nanofiltration membranes". Water Research. 43 (9) : 2349 – 2362.

Yangali-Quintanilla V., Verliefde A., Kime T. U., Sadmani A., Kennedy M. et Amy G. (2009b). "Artificial neural network modelsbased on QSAR for predicting rejection of neutral organic compounds by polyamide nanofiltration and reverse osmosis.

Yangali Quintanilla V. (2010a). "Rejection of emerging organic contaminants by nanofiltration and reverse osmosis membranes effects of fouling, modelling and water reuse". Thèse de doctorat. Université de Delft de technologie. Delft. Netherlands

Yangali-Quintanilla V., Sadmani A., McConville M., Kennedy M. et Amy G. (2010b). "A QSAR model for predicting rejection of emerging contaminants (pharmaceuticals, endocrine disruptors) by nanofiltration membranes". Water Research. 44 (2) : 373 – 384.

Yan-ling L, Xiao-mao W., Hong-wei Y., Yuefeng F. X.(2018). "Quantifying the influence of solute-membrane interactions on adsorption and rejection of pharmaceuticals by NF/RO membranes." Journal of Membrane Science 551 (2018) 37–46.

https://fr.wikipedia.org/wiki/Filtration_sur_membrane (Consulté le 16/05/2022).

https://www.researchgate.net/figure/Les-modules-tubulaires_fig1_32447411 (Consulté16/05/2022).

ScienceDirect :
https://www.sciencedirect.com//search?qs=retention%20of%20pharmaceutical%20actif%20compounds%20by%20nanofiltration (Consulté le 20/05/2022).
https://www.sciencedirect.com//search?qs=retention%20of%20pharmaceutical%20actif%20compounds%20by%2020reverse%20osmosis (Consulté le 20/05/2022).

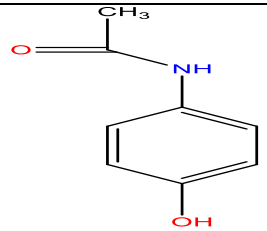
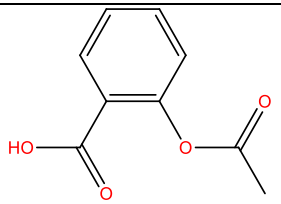
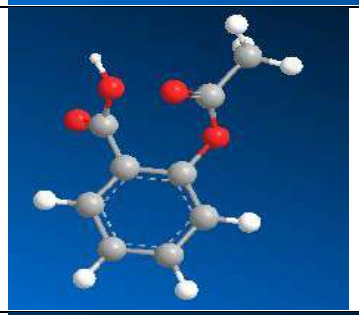
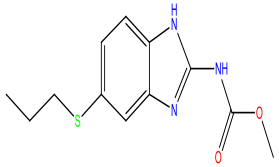
(https://www.researchgate.net/figure/Representation-type-dune-cellule-nerveuse-

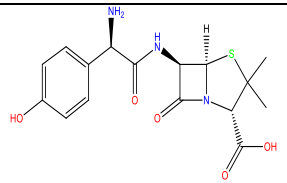
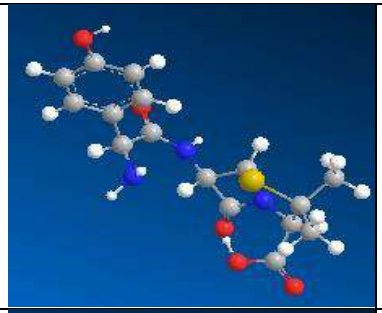
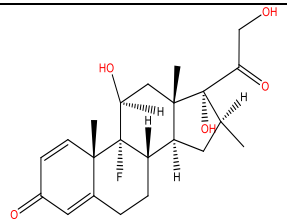
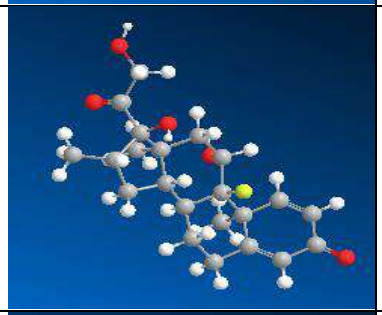
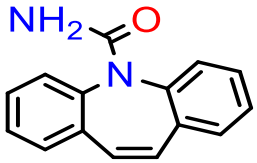
neurone3_fig1_275642235(Consulté10/04/2022).

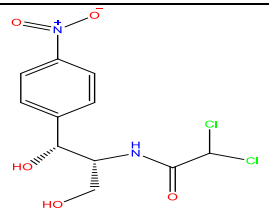
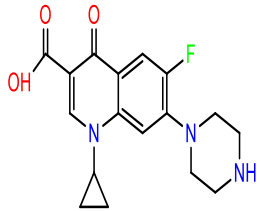
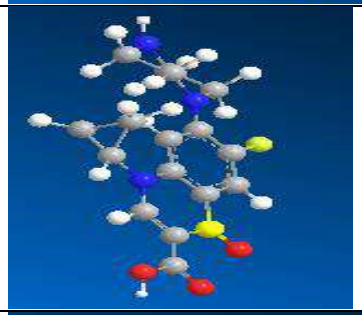
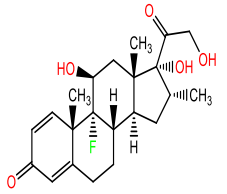
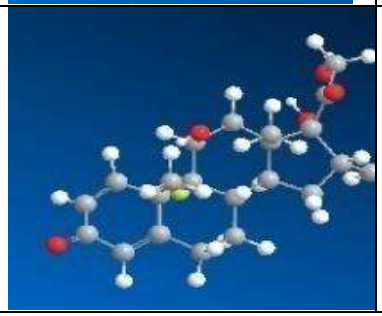
http://informatique.coursgratuits.net/methodes-numeriques/reseaux-de-neurones-formels.php(Consulté10/04/2022).

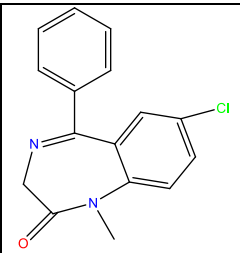
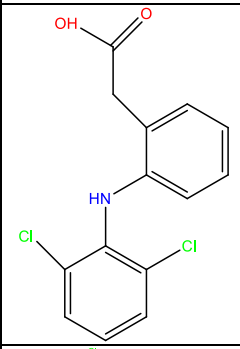
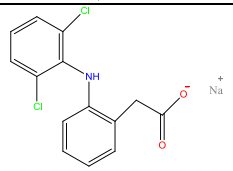
http://www.chemspider.com/(Consulté 10/04/2022).

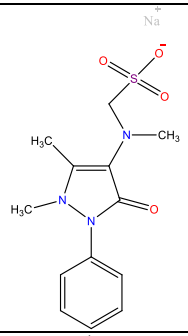
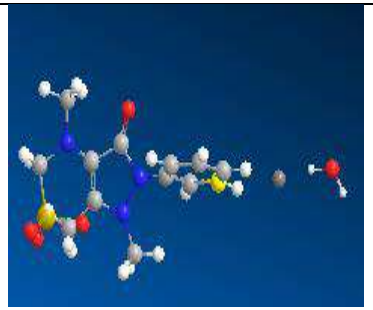
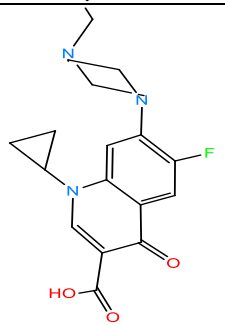
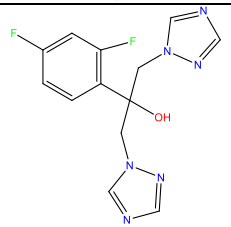
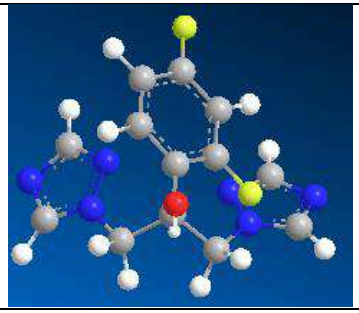
Tableau 1. Liste de 31 produits pharmaceutiques utilisées pour développer les modèles QSPR-RN.

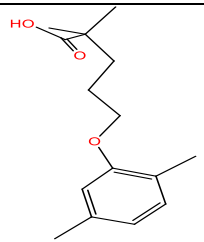
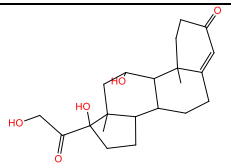
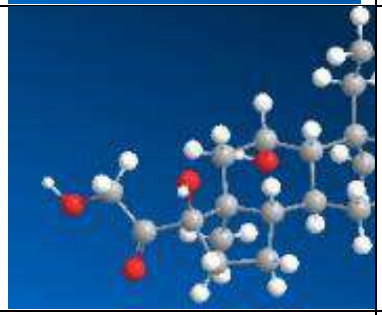
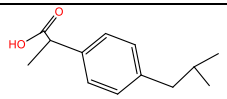
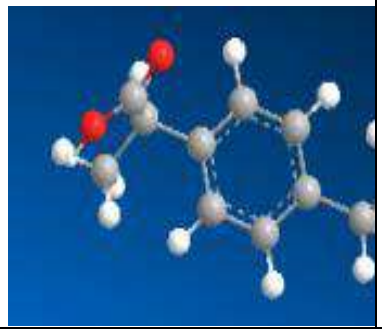
#	Nom	Formule brute	MW (g/mol)	Structure chimique 2D	Structure chimique 3D	Classe thérapeutique
01	**Acetaminophen**	$C_8H_9NO_2$	151.17			Analgesic Anti-inflammatory drug Antipyretic
02	**Acetylsalicylic Acid**	$C_9H_8O_4$	180.16			Analgesic Anti-inflammatory drug Antipyretic Antirheumatic drug
03	**Albendazole**	$C_{12}H_{15}N_3O_2S$	265.33			Anthelminthic drug

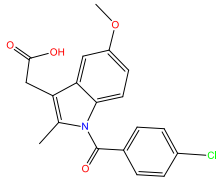
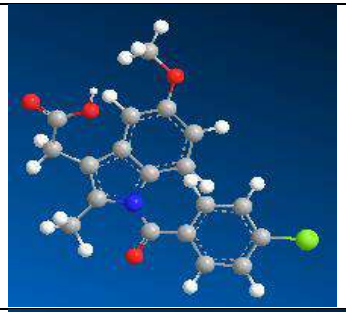
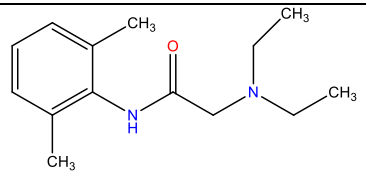
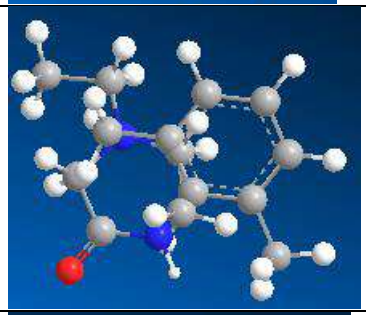
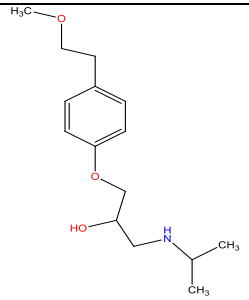
04	**Amoxicillin**	$C_{16}H_{19}N_3O_5S$	365.40			Antibacterial drug Beta-lactam antibiotic
05	**Betamethasone**	$C_{22}H_{29}FO_5$	392.46			Glucocorticoid
06	**Carbamazepine**	$C_{15}H_{12}N_2O$	236.269			Analgesic Anticonvulsant

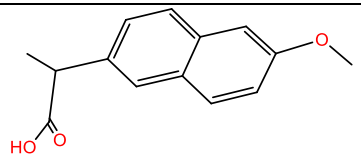
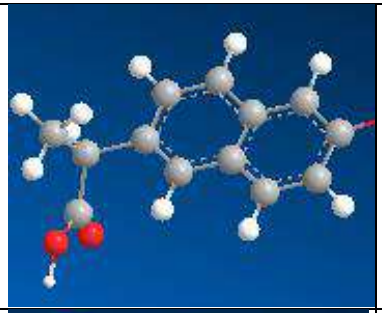
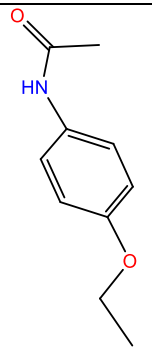
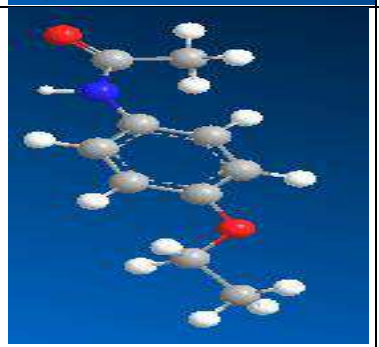
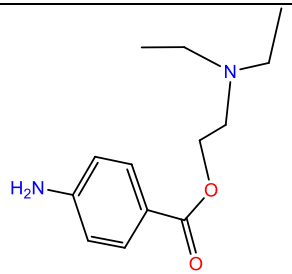
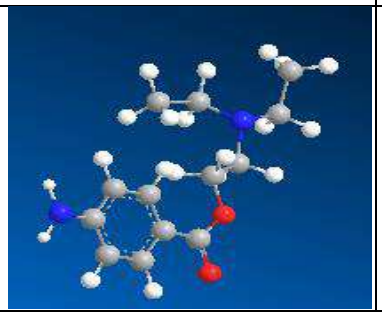
07	**Chloramphenicol**	$C_{11}H_{12}Cl_2N_2O_5$	323.129			Antibacterial drug
08	**Ciprofloxacin**	$C_{17}H_{18}FN_3O_3$	331.341			Antibacterial drug Antibiotic Antiinfective agent
09	**Dexamethasone**	$C_{22}H_{29}FO_5$	392.461			Anti-asthmatic agent Anti-inflammatory drug Antineoplastic agent Glucocorticoid

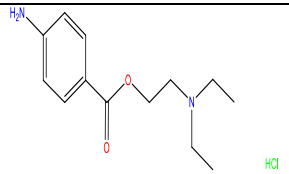
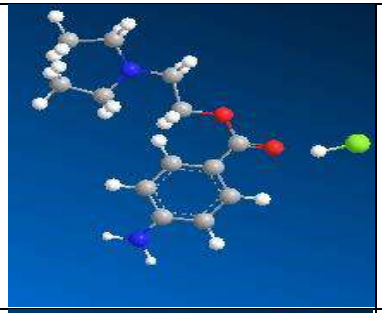
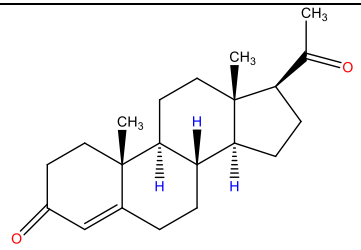
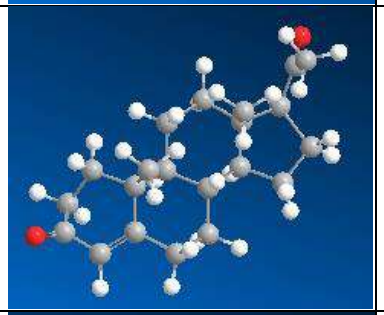
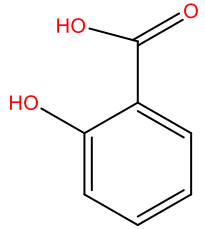
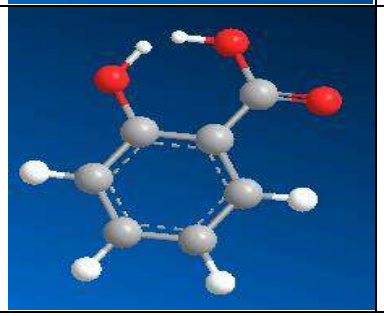
10	**Diazepam**	$C_{16}H_{13}ClN_2O$	284.74			Anxiolytic drug Muscle relaxant Sedative
11	**Diclofenac**	$C_{14}H_{11}Cl_2NO_2$	296.149			Anti-inflammatory drug
12	**Diclofenac Sodium**	$C_{14}H_{10}Cl_2NNaO_2$	318.13			Anti-inflammatory drug

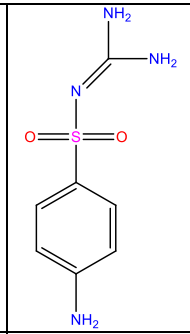
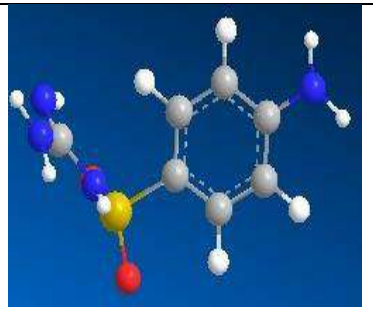
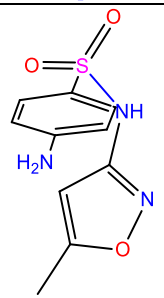
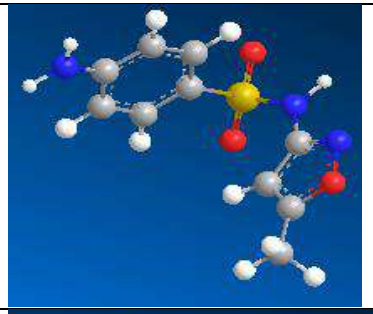
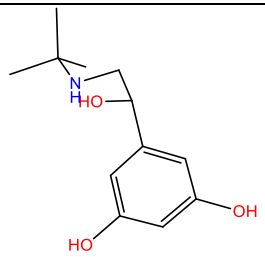
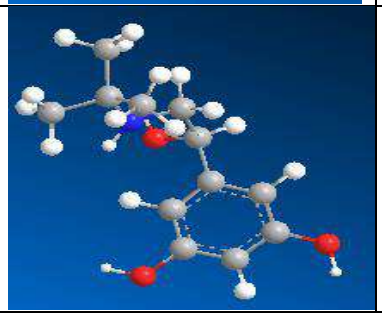
13	**Dipyrone**	$C_{13}H_{18}N_3NaO_5S$	351.35			Analgesic Antipyretic
14	**Enrofloxacin**	$C_{19}H_{22}FN_3O_3$	359.39			Antibacterial drug
15	**Fluconazole**	$C_{13}H_{12}F_2N_6O$	306.10			Antifungal agent

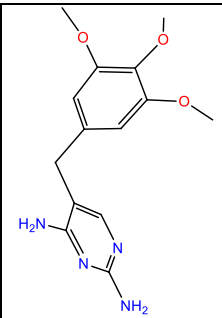
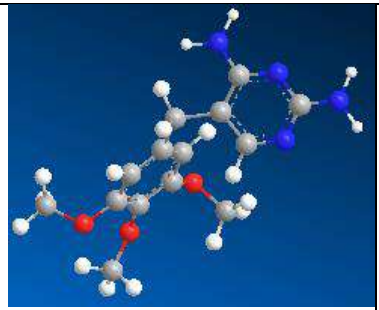
16	**Gemfibrozil**	$C_{15}H_{22}O_3$	250.33			Antilipemic drug
17	**Hydrocortisone**	$C_{21}H_{30}O_5$	362.46			Anti-inflammatory drug Glucocorticoid
18	**Ibuprofen**	$C_{13}H_{18}O_2$	206.28			Anti-inflammatory drug

19	**Indomethacin**	$C_{19}H_{16}ClNO_4$	357.78			Analgesic Anti-inflammatory drug Antipyretic Tocolytic agent
20	**Lidocaine**	$C_{14}H_{22}N_2O$	234,34			Local anaesthetic
21	**Lopressor**	$C_{34}H_{56}N_2O_{12}$	267,36			Antihypertensive agent

22	**Naproxen**	$C_{14}H_{14}O_3$	230,26			Analgesic Anti-inflammatory drug Antipyretic
23	**Phenacetin**	$C_{10}H_{13}NO_2$	179.216			Analgesic Antipyretic
24	**Procaine**	$C_{13}H_{20}N_2O_2$	236.310			Local anaesthetic

25	**Procaine hydrochloride**	$C_{13}H_{21}ClN_2O_2$	272.771			Local anaesthetic
26	**Progesterone**	$C_{21}H_{30}O_2$	314,46			Progestogen
27	**Salicylic acid**	$C_7H_6O_3$	138.121			Analgesic Anti-inflammatory drug Antipyretic Keratolytic drug

28	**Sulfaguanidine**	$C_7H_{10}N_4O_2S$	214.245			Antibacterial drug Antibiotic Antiinfective agent
29	**Sulfamethoxazole**	$C_{10}H_{11}N_3O_3S$	253.278			Antibacterial drug Antibiotic Antiinfective agent
30	**Terbutaline**	$C_{12}H_{19}NO_3$	225,28			Beta-adrenergic agonist Bronchodilator agent Ec 3.1.1.8 (cholinesterase) inhibitor Tocolytic agent

31	**Trimethoprim**	$C_{14}H_{18}N_4O_3$	290.318			Antibacterial drug Antimalarial

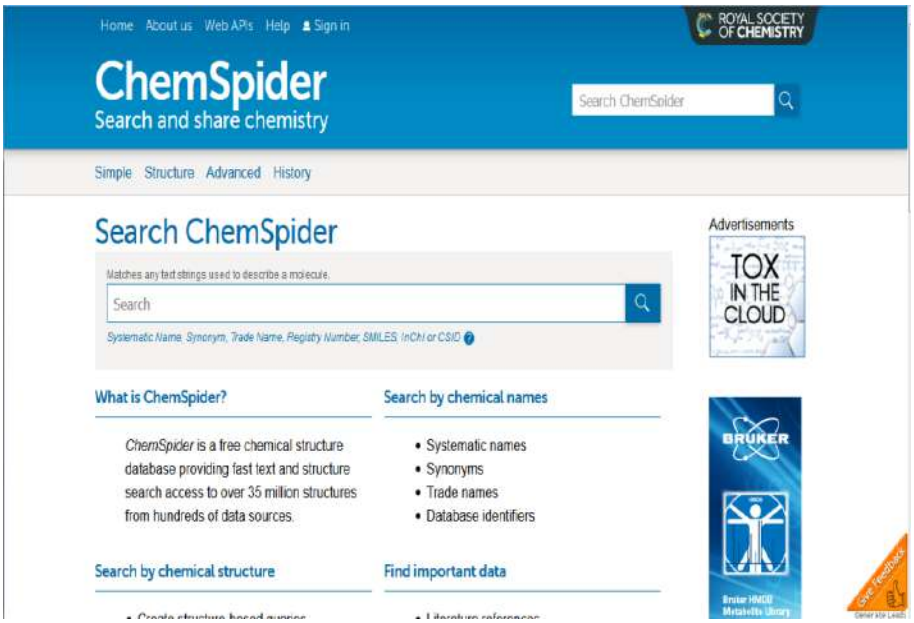
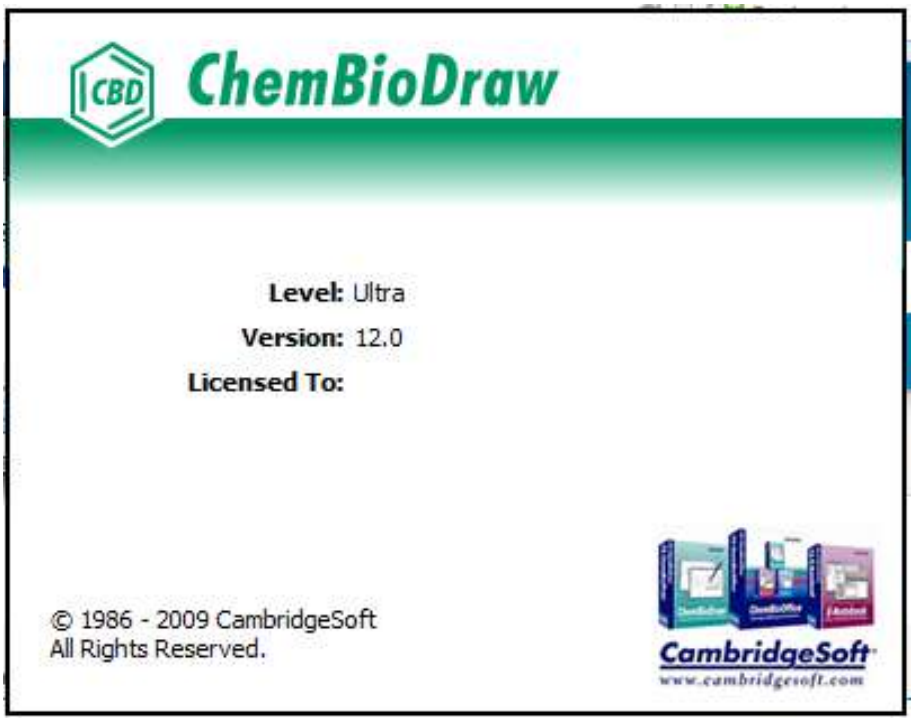


Figure 1. Capture d'écran du site intitulé Chemspider.



(a)



(b)

Figure 3. Capture d'écran des logiciels : (a) ChemBioDraw et (b) ChemBio3D

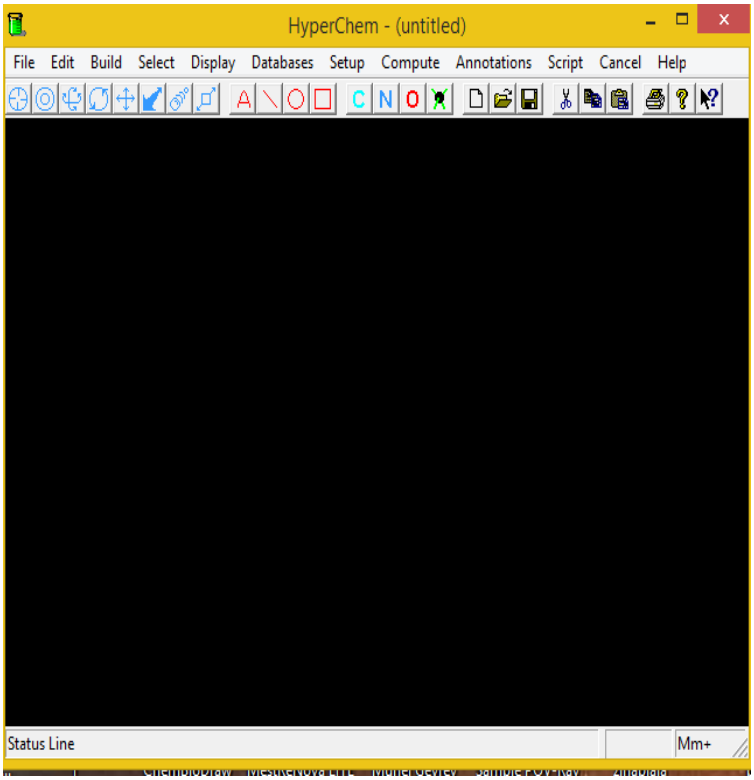


Figure 4. Capture d'écran de logiciel Hyperchem.



Figure 5. Capture d'écran de logiciel STATISTICA.

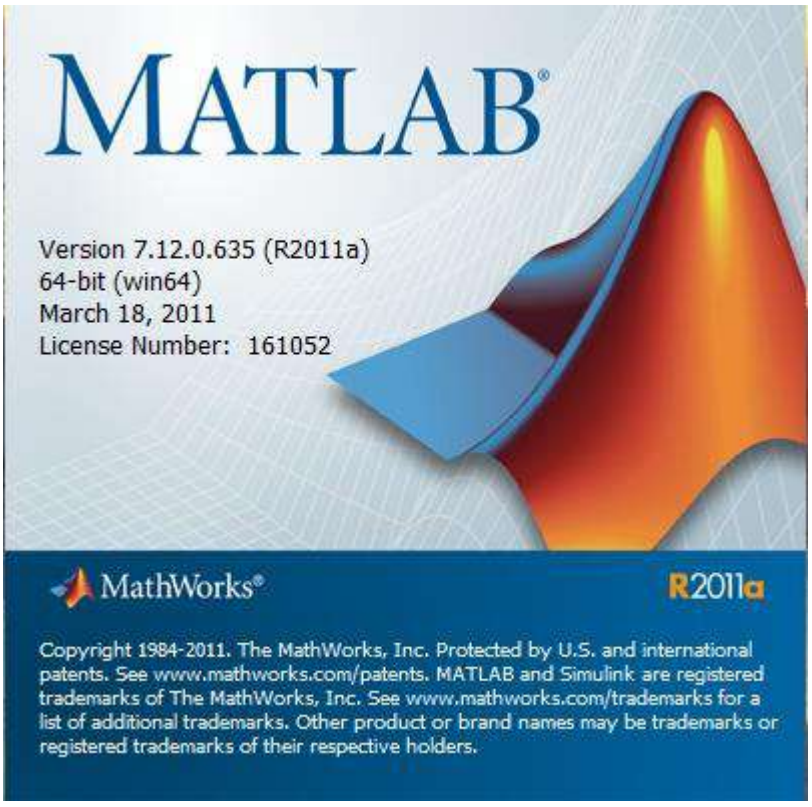


Figure 6. Capture d'écran de logiciel MATLAB 2011.

Tableau 1. Matrice des coefficients de corrélation pour la première base de données (BD)

	dc(g/mol)	Logd	Dipole moment (Debye)	Length (nm)	S_{max} (nm^2)	S_{min} (nm^2)	Width (nm)	Depth (nm)	Eqwidth (nm)	MWCO (Dalton)	SR (NaCl) (%)	Zeta potential (mv)	Contact angle (°)	pH	Pression (kpa)	Taux de récupération (%)	T(C°)	Rétention (%)
dc(g/mol)	1.0000	-0.0362	0.1986	-0.1798	0.8917	0.5907	0.5989	0.8920	0.7962	-0.1172	-0.1000	0.0938	0.0308	0.1261	0.0740	0.1920	-0.1930	0.4708
Logd	-0.0362	1.0000	-0.6070	0.1665	-0.1710	0.2004	0.1992	-0.1395	0.0214	-0.1505	-0.0107	0.0487	0.1248	0.2956	-0.1983	0.1102	-0.3643	0.0378
Dipole moment (Debye)	0.1986	-0.6070	1.0000	-0.0329	0.2832	0.0665	0.0751	0.2703	0.1875	-0.0134	-0.0348	-0.0036	-0.1804	-0.0713	0.2070	-0.0568	0.0921	0.0476
Length (nm)	-0.1798	0.1665	-0.0329	1.0000	-0.1890	0.0978	0.0837	-0.1821	-0.0592	0.1632	-0.0133	-0.0473	-0.2981	-0.1976	0.0557	0.0786	0.1021	-0.2033
S_{max} (nm^2)	0.8917	-0.1710	0.2832	-0.1890	1.0000	0.7860	0.7906	0.9982	0.9475	-0.1196	-0.1105	0.1982	0.0255	0.1032	0.1116	0.2099	-0.0913	0.4722
S_{min} (nm^2)	0.5907	0.2004	0.0665	0.0978	0.7860	1.0000	0.9987	0.8047	0.9400	-0.1393	-0.1289	0.1094	-0.0440	0.1530	0.0029	0.1305	-0.2293	0.3985
Width (nm)	0.5989	0.1992	0.0751	0.0837	0.7906	0.9987	1.0000	0.8108	0.9441	-0.1417	-0.1375	0.1070	-0.0411	0.1668	0.0025	0.1285	-0.2294	0.4117
Depth (nm)	0.8920	-0.1395	0.2703	-0.1821	0.9982	0.8047	0.8108	1.0000	0.9585	-0.1258	-0.1199	0.1954	0.0295	0.1273	0.1009	0.2157	-0.1096	0.4813
Eqwidth (nm)	0.7962	0.0214	0.1875	-0.0592	0.9475	0.9400	0.9441	0.9585	1.0000	-0.1396	-0.1346	0.1619	-0.0040	0.1536	0.0577	0.1854	-0.1751	0.4717
MWCO (Dalton)	-0.1172	-0.1505	-0.0134	0.1632	-0.1196	-0.1393	-0.1417	-0.1258	-0.1396	1.0000	-0.0559	0.0058	-0.1560	-0.1212	-0.1721	0.0886	0.2903	-0.4314
SR (NaCl) (%)	-0.1000	-0.0107	-0.0348	-0.0133	-0.1105	-0.1289	-0.1375	-0.1199	-0.1346	-0.0559	1.0000	-0.0197	0.0051	-0.0923	-0.0029	-0.0521	0.0039	-0.0017
Zeta potential (mv)	0.0938	0.0487	-0.0036	-0.0473	0.1982	0.1094	0.1070	0.1954	0.1619	0.0058	-0.0197	1.0000	0.2955	0.0659	0.1676	0.4785	0.3467	0.0765
Contact angle (°)	0.0308	0.1248	-0.1804	-0.2981	0.0255	-0.0440	-0.0411	0.0295	-0.0040	-0.1560	0.0051	0.2955	1.0000	0.2473	-0.0982	0.2187	0.1142	0.0459
pH	0.1261	0.2956	-0.0713	-0.1976	0.1032	0.1530	0.1668	0.1273	0.1536	-0.1212	-0.0923	0.0659	0.2473	1.0000	-0.1088	0.2509	-0.1887	0.1920
Pression (kpa)	0.0740	-0.1983	0.2070	0.0557	0.1116	0.0029	0.0025	0.1009	0.0577	-0.1721	-0.0029	0.1676	-0.0982	-0.1088	1.0000	0.1353	0.3082	0.1984
Taux de récupération (%)	0.1920	0.1102	-0.0568	0.0786	0.2099	0.1305	0.1285	0.2157	0.1854	0.0886	-0.0521	0.4785	0.2187	0.2509	0.1353	1.0000	0.3426	-0.1016
T(C°)	-0.1930	-0.3643	0.0921	0.1021	-0.0913	-0.2293	-0.2294	-0.1096	-0.1751	0.2903	0.0039	0.3467	0.1142	-0.1887	0.3082	0.3426	1.0000	-0.0987
Rétention (%)	0.4708	0.0378	0.0476	-0.2033	0.4722	0.3985	0.4117	0.4813	0.4717	-0.4314	-0.0017	0.0765	0.0459	0.1920	0.1984	-0.1016	-0.0987	1.0000