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 **E**nvironnement

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

Filière : Génie des Procédés
Spécialité : Génie Chimique

Modélisation de la rétention des anticancers par nanofiltration

Présenté par :

- M^{elle} BELHADJ Marwa
- M^{elle} BOUDJEMAI Fadhila

Proposé et Dirigé par :

- Dr. AMMI Yamina

Promotion 2023-2024


بسم الله الرحمن الرحيم

Dédicaces

Je dédie ce modeste travail particulièrement à mes chers parents, qui ont consacré existence à bâtir la mienne, pour leur soutien, patience et soucis de tendresse et d'affection pour tout ce qu'ils ont fait pour que je puisse arriver à ce stade.

A ma mère qui m'a encouragé durant tous mes études, et qui sans elle, ma réussite née n'aura pas eu lieu.

A mon père, qui est toujours disponible pour nous, et prêt à nous aider, je lui confirme mon attachement et mon profond respect.

A mes chers frères : Massaud, Rabeh, Yacin et Mahdi.

Et ma très chère sœur : Karima, Hanane, Zohra et Samira.

A toute mes amies : Abir, Hania, Ikram, Sara, wassila et yassmine.

Et a tous ceux que j'ai connus durant mon cycle d'étude.

…..Fadhila

Dédicaces

J'ai le grand honneur de dédier ce modeste travail : A ceux qui m'ont encouragé et soutenu moralement et matériellement pendant les moments les plus difficiles et durant toute ma vie. A mon très cher ***père*** *et ma très chère* ***mère.***

A mon très cher mari ***ACHERAF*** *qui m'a donné la possibilité d'arriver là où je suis, je remercie Allah de vous avoir amené dans ma vie.*

A mes très chers frères ***AHEMAD*** *et* ***SALAH EDDINE.***

A mes adorables sœurs ***FATI*** *et* ***MERIAM.***

A mes oncles et mes tantes et tous mes cousins et cousines.

*Toute la famille «****BELHADJ****» ,«****SALMI****» et «****MERAZI****».*

A mes chères amies ***SOUMIA, MALIKA, FADHILA,*** *et* ***RANIA*** *pour leur amitié sincère et pour tous les bons moments passés ensemble.*

Pour ceux que je n'ai pas cités bien sûr, ne croyez pas que je vous ai oublié, je vous porte toujours dans mon cœur.

A tous les étudiants « science de technologie ***2019/2024 »****.*

.....MARWA

Remerciements ET gratitude

*Nous souhaitons exprimer notre gratitude à notre seigneur, **Allah**, qui est la source de toute connaissance et qui nous a donné la force et la patience pour mener à bien ce travail.*

*En second lieu, Nous voudrions, exprimer nos sincères et ineffables remerciements au Dr. **Yamina AMMI** qui a assuré l'encadrement de ce mémoire et qui a bien voulu nous accueillir au sein de son équipe, sa disponibilité, son écoute et ses conseils. Tout au long de ce mémoire, elle nous n'a cessé de suivre et de s'intéresser à son avancement. Qu'elle trouve ici l'expression de notre reconnaissance et de notre profonde gratitude.*

Nos vifs remerciements vont également aux membres du jury pour l'intérêt qu'ils ont porté à notre recherche en acceptant d'examiner notre travail et de l'enrichir par leurs propositions

Nous souhaitons remercier les professeurs département du génie des procédés de l'environnement pendant les cinq années précédentes.

Merci à tous ceux qui m'ont aidé de près ou de loin durant ces années de recherches et d'études.

.

ملخص

نمذجة الاحتفاظ بمضادات السرطان عن طريق الترشيح النانوي

الهدف من هذه الدراسة هو إنشاء نموذج عصبي مثالي يعتمد على العلاقات الكمية بين خصائص البنية ($QSPR\text{-}RNA_{optimal}$) لنمذجة الاحتفاظ بمضادات السرطان عن طريق الترشيح النانو.

لتوقع وجود بقايا مضادات السرطان، تم بناء نموذج عصبي من نوع feedforward ($QSPR\text{-}RNA_{optimal}$) يتميز ببنية (13 خلية عصبية في طبقة المدخلات، 4 خلايا عصبية في الطبقة المخفية وخلية عصبية واحدة في طبقة الإخراج). تم استخدام قاعدة بيانات مكونة من 554 نقطة لاختبار نموذج $QSPR\text{-}RNA_{optimal}$ ، مع تقسيم البيانات بنسبة 80% للتدريب و20% لاختبار النموذج.

أظهرت مقارنة قيم الاحتفاظ المحسوبة خلال مرحلة الاختبار مع تلك المقاسة بوجود علاقة جيدة (R = 0.9318).

كشف تحليل الحساسية بواسطة طريقة "الوزن" أن المتغيرات الأكثر صلة التي يمكن أن تؤثر على الاحتفاظ بمضادات السرطان في ترشيحها النانوي لها مساهمة كبيرة (IR>5٪).

الكلمات المفتاحية : النمذجة ، الاحتفاظ ، مضادات السرطان ، الترشيح النانوي ، QSPR-RNA.

Abstract

Modeling the retention of anticancer by nanofiltration

The objective of this study is to create an optimal neural model based on quantitative structure-property relationships ($QSPR\text{-}RNA_{optimal}$) to model the retention of anticancer by nanofiltration.

To anticipate the presence of anticancer residues, a feedforward-type neural model ($QSPR\text{-}RNA_{optimal}$) characterized by a structure (13 neurons in the input layer, 4 neurons in the hidden layer, and 1 neuron in the output layer) will be constructed. A database of 554 points was used to test the $QSPR\text{-}RNA_{optimal}$ model, with a data split of 80% for training and 20% for testing the model.

The comparison of the retention values calculated during the test phase with those measured showed a good correlation (R = 0.9318).

The sensitivity analysis by the "weight" method revealed that the most relevant variables that can influence the retention of anticancer their nanofiltration have a significant contribution (IR> 5%).

Keywords: Modeling, Retention, Anticancer, Nanofiltration, QSPR-RNA.

Résumé

Modélisation de la rétention des anticancers par nanofiltration

L'objectif de cette étude est de créer un modèle neuronal optimal basé sur des relations quantitatives structure-propriété ($QSPR\text{-}RNA_{optimal}$) pour modéliser la rétention des anticancers par la nanofiltration.

Pour anticiper la présence de résidus d'anticancer, un modèle neuronal du type feedforward ($QSPR\text{-}RNA_{optimal}$) caractérisé par une structure (13 neurones dans la couche d'entrée, 4 neurones dans la couche cachée et 1 neurone dans la couche de sortie) sera construit. Une base de données de 554 points a été utilisée pour tester le modèle $QSPR\text{-}RNA_{optimal}$, avec une répartition des données en 80 % pour l'apprentissage et 20 % pour le test du modèle.

La comparaison des valeurs de rétention calculées lors de la phase de test avec celles mesurées a montré une bonne corrélation (R = 0.9318).

L'analyse de sensibilité par la méthode "poids" a révélé que les variables les plus pertinentes pouvant influencer la rétention des anticancers leur nanofiltration ont une contribution significative (IR>5 %)

Mots-clés : Modélisation, Rétention, Anticancers, Nanofiltration, QSPR-RNA **.**

Abréviations iii
Nomenclature et symboles iv
Liste des tableaux vi
Liste des figures vii

Introduction générale 1

Chapitre I. Description de la technologie membranaire

I.1 Introduction 04
I.2. Généralités 04
I.2.1. Définition de la membrane 04
I.2.2. Classification des membranes 05
I.2.2.1. Selon leur nature chimique 05
I.2.2.2. Selon leur porosité 05
I.2.2.3 Selon leur morphologie 06
I.2.3 Osmose inverse et nanofiltration 11
I.2.4. Sélectivité 11
I.2.5. Perméabilité 12
I.2.6. Influence des paramètres opératoires 12
I.2.6.1. Pression 12
I.2.6.2. Température 13
I.2.6.3. Taux de récupération "Taux de conversion" 13
I.3. Interactions et propriétés de séparation 13
I.3.1. Encombrement stériques (effet du tamisage) 15
I.3.2. Etat physique de la membrane 15
I.3.3. Taille moléculaire 17
I.3.4. Répulsions électrostatiques (effet de charge) 19
I.3.4.1. Charge de la surface 20
I.3.4.2. Charge du soluté 20
I.3.5. Influence du pH 21
I.3.6. Interaction hydrophobes/adsorption 21
I.3.6.1. Hydrophobicité de la membrane 21
I.3.6.2. Hydrophobicité solutés 22
I.4. Applications des procédés d'OI et NF 23
I.5. Avantages et inconvénients des procédés d'OI et NF 24
I.6.1. Cancer et médicaments anticancéreux 24
I.6.2. Médicaments anticancéreux en milieu aquatique : sources et suivi 27
I.6.3. Effets de toxicité 27
I.6.4. Technologies de traitement explorées pour l'élimination des médicaments anticancéreux 28
I.7. Revue des travaux de la rétention des anticancers lors nanofiltration 29
I.8. Conclusion 29

Chapitre II. QSPR-Réseaux de neurones artificiels

II.1. Introduction 30
II.2. Méthodes d'apprentissages statistiques 31
II.3. Relations quantitatives structure-Activités (QSAR) 31
II.3.1. Outils de QSAR 32

II.3.2. Données biologiques 32
II.3.3. Descripteurs moléculaires 32
II.3.4. Formulation mathématique des QSAR 34
II.4. Réseaux de neurones artificiels (RNA) 35
II.4.1. Histoire des réseaux de neurones artificiels 35
II.4.2. Neurone biologique 36
II.4.3. Neurone formel 37
II.4.4. Architecture d'un réseaux neurones 40
II.4.5. Perceptrons multicouches 41
II.4.6. Types d'apprentissage des réseaux de neurones 42
II.4.7. Réseaux de neurones les plus utilisés 43
II.4.8. Conception d'un modèle neuronal 44
II.4.9. Détermination des entrées/sorties du réseau de neurones 45
II.4.10. Construction de la base de données des échantillons 45
II.4.11. Définition de la structure du réseau 46
II.4.12. Apprentissage 47
II.4.13. Evaluation des performances du réseau de neurones 50
II.5. Conclusion 50

Chapitre III. Modélisation de la rétention des anticancers

III.1. Introduction 51
III.2. Collecte des données, prétraitement et analyse 51
III.3. Développement des modèles QSPR-RNA 57
III.4. Performance des modèles neuronaux 59
III.5 Expression mathématiques de modèle neuronal optimal 65
III.6. Analyse de sensibilité par la méthode de "poids" 67
III.7. Diagnostic des points aberrant et domaine d'applicabilité du modèle neuronal optimisé 69
III.8. Conclusion 72

Conclusion générale et perspectives 71
Références bibliographiques
Annexe A
Annexe B

Symbole	Signification
ART	Adaptive Resonance Theory
b_o	Biais de la couche de sortie
b_h	Biais de la couche cachée
BD	Base de Données
BFGS	Broyden–Fletcher–Goldfarb–Shanno
D	Descripteur moléculaire
DS	Déviation Standard
F	Fonction d'activation
FAV	Fistule artérioveineuse
H*	Levier critique
HB	Molécule est hydrophobe
HL	Molécule est hydrophile
IUPAC	International Union of Pure and Applied Chemistry
MF	Microfiltration
MWCO	Molecular weight cutt off (Seuil de poidsmoléculaire)
NF	Nano filtration
OI	Osmose inverse
PMC	Perceptron Multicouches
QSAR	Relations quantitatives structure-Activités
QSTR	Relations quantitatives structure-proprietés
QSPR	Relations quantitatives structure-proprietés
RFR	Réseaux à fonction de bases radiales
RN	Réseaux de Neurones
RNA	Réseaux de neurones artificiels
SOM	Cartes auto-organisatrices de Kohonen
TFC	Thin-film composite
Tr	Taux de rejet
UF	Ultrafiltration
$W_{j,i}^{I}$	Poids de connexions entrée-cachée
$W_{1,j}^{H}$	Poids de connexions cachée-sortie

Lecture	Désignation	Unité
C	Concentration de la solution	(mol.m^{-3})
Csa	Concentration d'un soluté dans l'alimentation	(kg/m^3)
Csp	Concentration d'un soluté dans le perméat	(kg/m^3)
DCO	Demande Chimique en Oxygène	(mg.L^{-1})
Depth	Profondeur moléculaire	(nm)
EAM	Erreur Absolue Moyenne	(%)
Ei	$I^{éme}$ entrée du modèle	(-)
EQM	Erreur Quadratique Moyenne	(-)
EqWidth	Largeur équivalente moléculaire	(nm)
ESP	Erreur Standard de Prédiction	(%)
Length	Longueur moléculaire	(nm)
log 1/C	Logarithme inverse de l'activité	(-)
$logk_{ow}$	Logarithme du coefficient de partage octanol-eau	(-)
Q_A	Débit volumique d'Alimentation	(m^3/h)
Q_P	Débit volumique de Perméat	(m^3/h)
R	Coefficient de corrélation	(-)
REQM	Racine de l'Erreur Quadratique Moyenne	(%)
T	Température	(k)
W	Poids synaptique de la connexion du neurone	(-)
Width	Largeur moléculaire	(nm)
Y	Taux de récupération	(-)
Zj	Sortie de la couche cachée	(-)
π	Pression osmotique de la solution	(Pa)

Lettres Grecques

Lecture	Désignation	Unité
α	Pente de la droite du meilleur ajustement linéaire	(-)
β	Ordonnée à l'origine de la droite du meilleur ajustement linéaire	(-)

Exposants

Cal	Calculée
Exp	Expérimentale
H	Cachée (Hidden)
I	Entrée (Input)
O	Sortie (Output)

Indices

Lecture	Désignation
Max	Maximum
Min	Minimum
$\bar{x}$	Moyenne

Chapitre I. Description de la technologie membranaire

Tableau I.1.	Classification IUPAC des pores suivants leur taille.	06
Tableau I.2.	Application des procédés d'OI et NF.	23
Tableau I.3.	Avantages et inconvénients de procédés d'OI et NF	24

Chapitre II. QSAR-Réseaux de neurones

Tableau II.1.	Fonctions de transfert usuelles, a=f(n).	38
Tableau II.2.	Résumé des caractéristiques des algorithmes d'apprentissage de rétro-propagation implantés dans MATLAB®	48

Chapitre III. Modélisation de la rétention des anticancers

Tableau III.1.	Taille de la base de données et le nombre des anticancers utilisés dans ce travail.	51
Tableau III.2.	Logiciels utilisés pour calculer LogD, le moment dipolaire, Smin, Smax et Length.	52
Tableau III.3.	Entrées et sortie utilisées pour le développement des modèles QSPR-RNA.	53
Tableau III.4.	Analyse statistique des données d'entrée et de sortie (BD1).	54
Tableau III.5.	Analyse statistique des données d'entrée et de sortie (BD2).	54
Tableau III.6.	Analyse statistique des données d'entrée et de sortie (BD3).	55
Tableau III.7.	Comparaison entres les modèles neuronaux (QSPR-RNA).	60
Tableau III.8.	Structure de modèle RNA optimal QSPR-RNAoptimal.	61
Tableau III.9.	Vecteurs de régression linéaire issue de l'équation linéaire : [équation linéaire : $ycal=\alpha yexp+\beta$, $\boldsymbol{\alpha}$(pente), $\boldsymbol{\beta}$(ordonnée à l'origine), ***R***(coefficient de corrélation)] et l'erreur quadratique moyenne ***(REQM***), l'erreur absolue moyenne ***(EAM)*** et l'erreur de standard de prédiction ***(ESP)***.	64

Annexe A

Tableau 1	Liste de 13 anticancers utilisées pour développer les modèles QSPR-RNA.	

Chapitre I. Description de la technologie membranaire

Figure I.1.	Membrane sélective.	05
Figure I.2.	Représentation schématique des différentes structures membranaires.	06
Figure I.3.	Représentation schématique d'une membrane composite.	07
Figure I.4.	Gamme de rétention de différentes particules par les procédés membranaires.	08
Figure I.5.	Représentation schématique d'un module plan.	09
Figure I.6.	Représentation schématique d'un module tubulaire.	09
Figure I.7.	Représentation schématique d'un module à fibre creuses.	10
Figure I.8.	Représentation schématique d'un module spiralé.	10
Figure I.9.	Molécule de polyamide et ses équilibres acido-basiques.	14
Figure I.10.	Illustration de 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).	16
Figure I.11.	Calcul des dimensions de l'octanol (Length, Width et Depth) de la structure 3D.	19
Figure I.12.	Mesure de l'angle de contact par la méthode de la goutte déposée.	21
Figure I.13.	Nombre de publications de 1995 à 2024 pour les membranes de la nanofiltration.	29

Chapitre II. QSAR-Réseaux de neurones

Figure II.1.	Fonctions de transfert usuelles, a=f(n).	37
Figure II.2.	Neurone formel.	37
Figure II.3.	Analogie entre neurone biologique et neurone formel.	40
Figure II.4.	Réseaux de neurones les plus utilisés.	40
Figure II.5.	Structure d'un réseau de neurones non bouclé.	41

Chapitre III. Modélisation de la rétention des anticancers

Figure III.1.	Rétentions expérimentales en fonction de poids moléculaires : (a) BD1, (b) BD2, (c) BD3.	57
Figure I.1.	Organigramme de développement et l'optimisation de l'architecture d'un QSPR-$RNA_{optimal}$.	58
Figure III.2.	Comparaison entre les valeurs expérimentales et les valeurs calculées pour QSPR-$RNA_{optimal}$: (a) la phase d'apprentissage, (b) la phase de test, (c) la phase totale.	63
Figure III.3.	Performances des modèles neuronaux (QSAR-$RNA_{optimale}$) pour la phase totale.	64
Figure III.4.	Représentation schématique du QSPR-$RNA_{optimal}$.	65
Figure III.5.	Architecture finale du réseau de neurone optimal de type feedforward.	67
Figure III.6.	Présentation des pourcentages de l'importance relative (*IR*) de chaque variable d'entrée sur la rétention des anticancers par nano-filtration.	69
Figure III.7.	Diagnostic des données potentiellement suspectes et domaine.	71

Annexe B

Figure 1	D'applicabilité du modèle neuronal optimal : (a) pour la phase test et (b) pour la phase totale.

Introduction générale

La croissance et le vieillissement de la population, ainsi que les modes de vie actuels, ont entraîné une augmentation des cas de cancer dans le monde. D'ici 2040, le nombre annuel de nouveaux cas de cancer devrait atteindre 30,2 millions, ce qui s'accompagnera d'une augmentation drastique de la consommation de médicaments anticancéreux et de leur rejet conséquent dans le milieu aquatique. Après administration, les médicaments anticancéreux sont excrétés par les patients et rejetés dans les STEP (stations d'épuration des eaux usées). Cependant, la plupart des médicaments anticancéreux ont une biodégradabilité limitée et par conséquent, les stations d'épuration conventionnelles sont probablement inefficaces pour éliminer ces médicaments. Ainsi, des médicaments anticancéreux sont rejetés en permanence dans le milieu aquatique (Cristovao et al. 2022).

Même si les médicaments anticancéreux se trouvent dans l'environnement aquatique à des niveaux de concentration extrêmement faibles (inférieurs à ceux de nombreux autres groupes pharmaceutiques), leur présence pourrait être très préoccupante pour l'environnement. En effet, la plupart des médicaments anticancéreux ont un mécanisme d'action très puissant et nombreux parmi eux manquent de sélectivité. Ainsi, outre les cellules tumorales, ils peuvent attaquer les cellules en croissance normale, provoquant potentiellement des effets néfastes sur la santé des organismes vivants eucaryotes. Récemment, Fonseca et al. ont évalué les effets toxiques d'un mélange de trois médicaments anticancéreux (cisplatine, cyclophosphamide et tamoxifène) sur le polychète marin Nereis diversicolor. Ils ont constaté qu'à des concentrations traces (allant de 0,1 à 10 $ng.L^{-1}$), ces composés étaient capables d'induire des dommages oxydatifs et des effets de génotoxicité après quatorze jours d'exposition. Les mêmes auteurs ont évalué la toxicité individuelle du cyclophosphamide sur le même organisme et ont constaté des dommages à l'ADN à des concentrations de 500 et 1000 $ng.L^{-1}$, après 14 jours d'exposition (Fonseca et al. 2019).

Il est important de commencer par définir quelle est la principale voie d'entrée des produits pharmaceutiques cibles dans l'environnement. En ce qui concerne les médicaments anticancéreux, une étude précédente a conclu que 94 % de la quantité totale de médicaments anticancéreux étaient livrés à des patients ambulatoires, ce qui indique que les effluents domestiques sont la principale source d'entrée de ces médicaments. Par conséquent, l'amélioration du traitement dans les stations d'épuration domestiques devrait plutôt être au centre des préoccupations, plutôt que d'agir dans les hôpitaux.

Diverses technologies de traitement ont été explorées pour l'élimination des médicaments des effluents d'eaux usées, notamment les procédés d'oxydation avancée (par exemple, UV en combinaison avec du peroxyde d'hydrogène ; ozonation) et les procédés de technologie membranaire (par exemple, nanofiltration ; osmose inverse). La nanofiltration est considérée comme une solution prometteuse en raison de sa capacité à produire de l'eau traitée de manière plus durable que l'osmose inverse (consommation d'énergie plus faible) et à obtenir toujours une élimination élevée d'une multitude de micropolluants. En effet, plusieurs études en laboratoire ont déjà démontré l'efficacité des membranes de nanofiltration pour éliminer divers produits pharmaceutiques. Parmi ces études, très peu se sont concentrées sur le rejet des médicaments anticancéreux. Néanmoins, il existe encore peu d'études sur la nanofiltration examinant le rejet des micropolluants provenant de matrices difficiles telles que les effluents d'eaux usées domestiques à l'échelle pilote, dans des conditions réelles et à des niveaux d'occurrence (sans fortifier les composés cibles) (Cristovao et al. 2022).

La littérature met en lumière plusieurs points. D'une part, la nanofiltration 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 anti-cancerns. La compréhension fondamentale et quantitative des mécanismes de transfert et des performances des membranes d'osmose inverse et de nanofiltration sont nécessaires à la sélection, au dimensionnement, au fonctionnement et au contrôle des systèmes industriels. 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 de molécules organiques 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 la rétention des anti-cancers lors de nanofiltration. 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. La révolution de l'informatique et le développement de nouvelles méthodes de calcul numérique basé sur l'intelligence artificielle et les sciences cognitives ont donné naissance durant les dernières décennies à des 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 du génie des procédés. 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. Bien que les réseaux de neurones artificiels (RNA) soient très appliqués dans le domaine du génie des procédés en raison de la précision des résultats qu'ils produisent et de la simplicité relative de la mise en œuvre. Peu de travaux de recherche utilisent les RNA comme moyen de calcul des phénomènes de transports par d'osmose inverse et de nanofiltration lors de la rétention des anti-cancers.

L'objectif de ce travail est l'application des QSPR-RNA de type feedforward et pour modéliser la rétention des anticancers par les membranes nanofiltration. Ainsi, nous avons organisé ce mémoire en trois 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, après la rétention des anticancers lors nanofiltration et une revue des travaux de la rétention des anti-cancers lors nanofiltration.

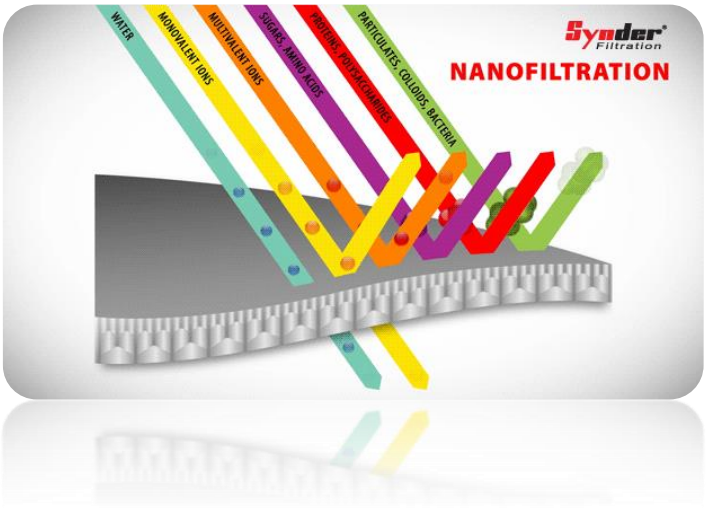
- ✓ Le deuxième chapitre est dédié à la présentation des modèles de réseaux de neurones artificiels basés sur la relation quantitative structure-activité "QSPR-RNA".

- ✓ Nous présentons dans le troisième chapitre la modélisation de la rétention des anti-cancers lors et nanofiltration. Il comptera six parties : la première partie présente la collecte des données, prétraitement et analyse, la deuxième partie sera consacrée au développement des modèles neuronaux, nous discuterons les performances des modèles neuronaux dans la troisième partie, les expressions mathématiques de modèle neuronal optimal seront abordées dans la quatrième partie, nous traiterons dans la cinquième partie l'analyse de sensibilité par la méthode de poids et dans la dernière partie nous discuterons le diagnostic des points aberrants et domaine d'applicabilité du modèle neuronal optimise.

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

Chapitre I

Description de la technologie membranaire



I.1. Introduction

La filtration membranaire est de plus en plus utilisée comme procédé de séparation dans de nombreux domaines notamment dans le cycle de l'eau (potabilisation de l'eau, traitement des effluents, réutilisation de l'eau, adoucissement, dessalement,…). La filtration membranaire est basée sur l'application d'une différence de pression qui permet le transfert du solvant à travers une membrane dont la taille des pores assure la rétention de solutés. Ces opérations, classées selon une taille des pores décroissantes, sont : la microfiltration (MF) l'ultrafiltration (UF), la nanofiltration (NF) et l'osmose inverse (OI).

Les procédés de séparation membranaire ont pris leur essor dans les années 70 en particulier grâce au développement des matériaux et des techniques de fabrication des membranes qui offrent une meilleure rentabilité dans tous les termes de consommation énergétique par apport à des techniques de séparation conventionnelles telle que la distillation, des conditions de fonctionnement tolérables, un traitement souhaitable, et qu'elles favorisent le développement dans toutes les branches de l'industrie (Bouranene et al. 2008). Parmi eux, l'osmose inverse et nanofiltration sont utilisées en traitement de séparation des différentes molécules organiques et inorganiques qui permettent une séparation conforme aux normes actuellement en vigueur.

Dans ce chapitre nous présentons une étude bibliographique, dont la majeure partie porte sur les différents aspects des membranes, à savoir leurs classifications, leur structure et nature, leurs principes de mise en œuvre, les différentes configurations des modules membranaires. Ensuite, nous détaillons les deux procédés membranaires nanofiltration et osmose inverse. Ensuite, nous présenterons la rétention des anticancéreux lors de la nanofiltration. Enfin, nous conclurons ce chapitre par une revue des travaux sur la rétention des anticancéreux par nanofiltration.

I.2. Généralités

I.2.1. Définition de la membrane

La membrane est définie comme étant une barrière sélective séparant deux compartiments et permettant le passage préférentiel d'au moins une espèce parmi les autres sous l'action d'une force de transfert (variation de concentration ou de pression) (Audinos et al. 1986). En générale, les constituantes qui sont plus petits que les pores de la membrane sont capables de passer à travers la membrane tandis que les substances et les molécules de taille plus importante sont retenues.

Les molécules ou particules retenues par la membrane se trouvent dans le retentât et les molécules qui traversent la membrane se récupèrent dans le perméat.

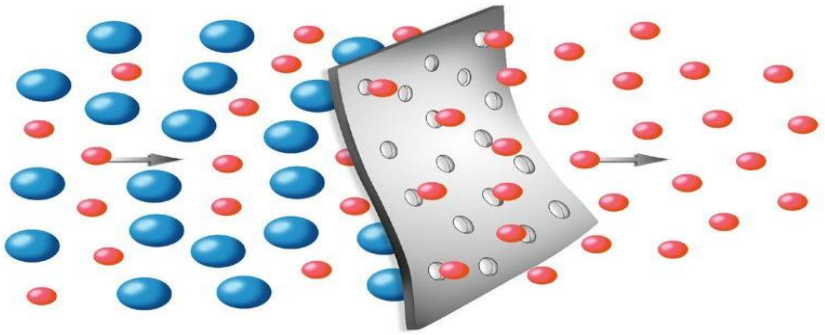


Figure I.1. Membrane sélective (https://www.openpr.com/news/3170003/polymeric-membrane-market-set-to-reach-us-11-153-9-million).

I.2.2. Classification des membranes

I.2.2.1. Selon leur nature chimique

a. Membrane organique : Elles sont fabriquées à partir de polymères organiques (acétate de cellulose , poly sulfones , polyamides , etc.) les qualités de ces matériaux leur confèrent une grande adaptabilité aux différentes applications. Environ 90% des membranes d'ultrafiltration et de microfiltration sont constituées de membranes organique (Remise 2006).

b. Membrane minérale : Elles sont à la base, de silice, de carbone, d'oxyde de graphite et de métaux frittés. Ces matériaux possèdent une grande résistance physicochimique, mécanique et thermique. Leur durée de vie excède de trois fois celle des membranes organiques mais leur cout est sensiblement plus élevé (Barbar 2006).

I.2.2.2. Selon leur porosité

a. Membrane poreuse : Ce type de membrane se compose de pores dont la dimension est supérieure aux dimensions moléculaires (Boukhiar 2016). L'IUPAC (International union of pure and applied chemistry) distingue trois catégories de pores selon leurs tailles (Audinos 2000).

Tableau I.1. Classification IUPAC des pores suivants leur taille.

Dénomination	Micropore	Mésopore	Macropore
Taille de pore	<2nm	2-5nm	>50nm

b. Membrane dense : Ce terme signifie une absence de porosité de dimension microscopique, ces membranes ne présentent aucune porosité. Elle consiste en un film dense à travers lequel le perméat est transporté par diffusion sous l'effet d'un gradient de potentiel électrique, de concentration ou d'une force de pression. Si la solubilité diffère, les membranes denses peuvent séparer des composés de taille voisine (Alfa 2008).

I.2.2.3. Selon leur morphologie

a. Membrane symétrique (isotrope): La membrane symétrique est composée d'un seul matériau ayant une composition sensiblement uniforme dans toute leur épaisseur. Elle peut être dense (absence de pores) ou poreuse. De telles membranes forment la couche active sur toute leur épaisseur (Maurel 2001).

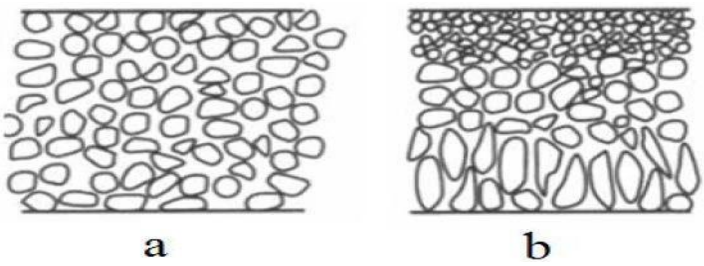


Figure I.2. Représentation schématique des différentes structures membranaires (Baker 2004 ; Chabane 2015).

b. Membrane asymétrique ou anisotrope : la membrane asymétrique ou anisotrope a une porosité variable dans son épaisseur (Baker 2004).

Elle présente deux parties distinctes :

- la peau ou couche active, de très faible épaisseur et porosité, assure la sélectivité de la membrane (Daufin et al. 2004).

- la peau ou couche poreuse ou support, garantit une résistance mécanique : elle représente en général la plus grande partie de l'épaisseur de la membrane et sa porosité interne est très élevée entre 60 et 80 %.

c. Membrane composite : Elle est constituée d'un assemblage de deux ou plusieurs couches de matériaux différents et présentant différentes porosités. La membrane composite comporte deux types de couches : les couches actives (couche filtrant) et le support assurant de bonnes propriétés mécanique (Remigy 2012).

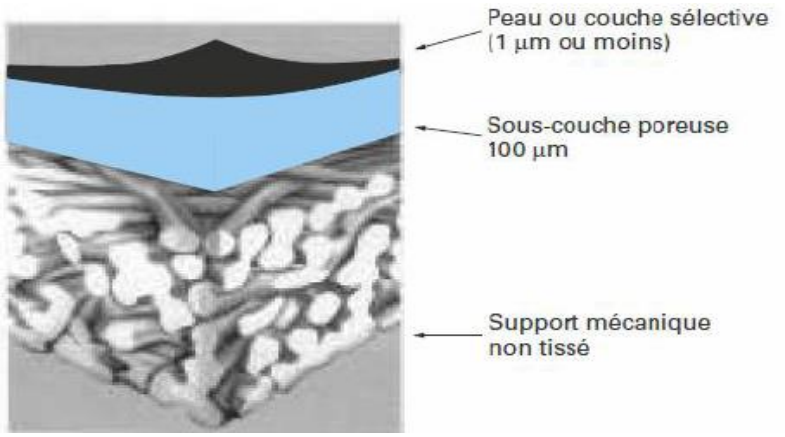


Figure I.3. Représentation schématique des membranes composites (Sara 2009).

d. Membrane à charge électrique : Les membrane chargées électriquement peuvent être denses ou microporeuses, mais elles sont le plus souvent très finement microporeuses. Les parois étant positivement fixées ou des ions chargés négativement. Une membrane à ions fixés à charge positive est appelée membrane échangeuse d'anions car elle lie les anions dans les environs fluide. De même, une membrane contenant des ions chargés négativement est appelées membrane échangeuse de cations. La séparation avec les membranes chargées est obtenue principalement par l'exclusion des ions de la même charge que les ions fixes de la structure de la membrane, et dans une moindre mesure par la taille des pores. La séparation est affectée par la charge et la concentration des ions en solution.

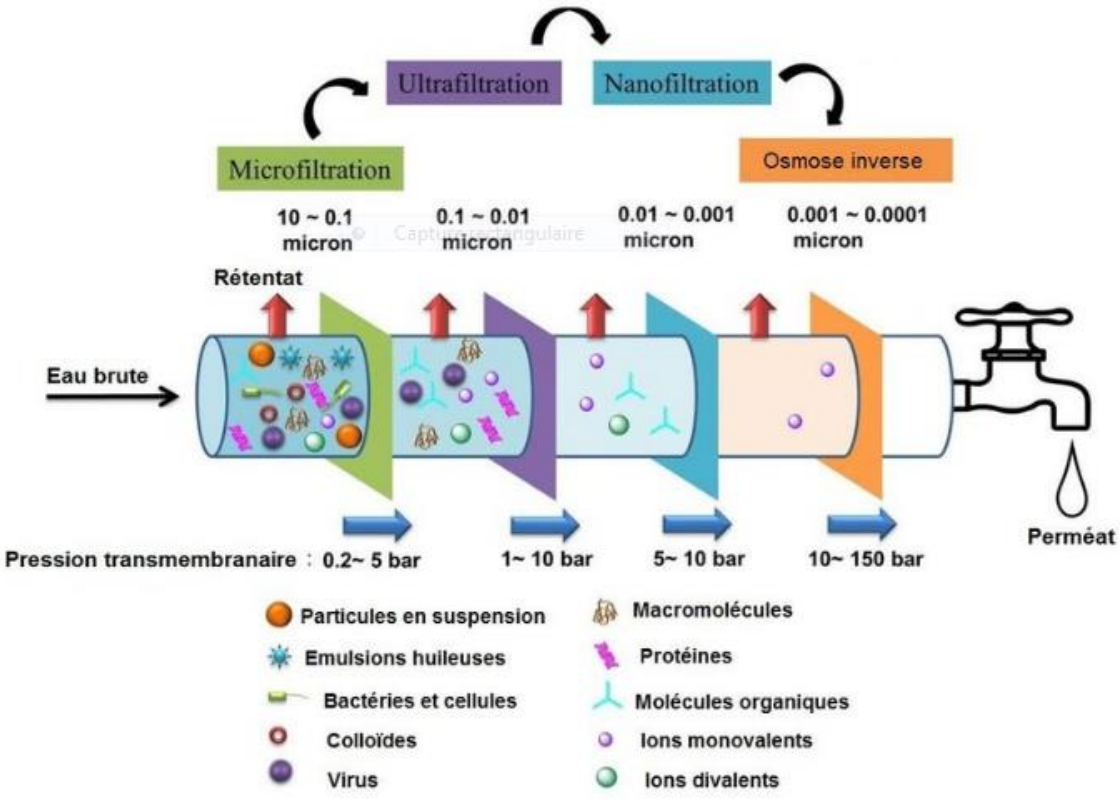


Figure I.4. Gamme de rétention de différentes particules par les procédés membranaires (https://www.researchgate.net/figure/Les-techniques-membranaires-les-plus-utilisees-dans-le-traitement-des-eaux_fig9_334331195).

- **Modules membranaires**

Il existe différents modèles de configuration des membranes permettant d'améliorer la performance globale du processus grâce à des supports sur lesquels sont montées, appelés modules.

Les modules membranaires couramment utilisés sont : le module plan, le module tubulaire, le module spiralé et le module à fibres creuses (Bouchoux 2004).

✓ **Module plan** : Les modules plans sont les plus anciens et qui ont été développés pour des applications commerciales à grande échelle et qui ne sont plus utilisés en raison de leurs prix élevés. Ils sont composés de deux plaques d'extrémité, d'une membrane en feuille plate et des entretoises. La couche de membrane sélective est prise en sandwich entre deux plaques de support, qui fournissent des canaux d'écoulement au fluide des deux côtés de la membrane, L'écoulement du fluide aura lieu donc parallèlement à travers le module. Cette disposition offre une grande flexibilité et facilite le nettoyage et le remplacement de

la membrane. Cependant, ces systèmes ne conviennent pas aux débits élevés, ils occupent trop de surface, ils sont facilement colmatés et nécessitent souvent une préfiltration.

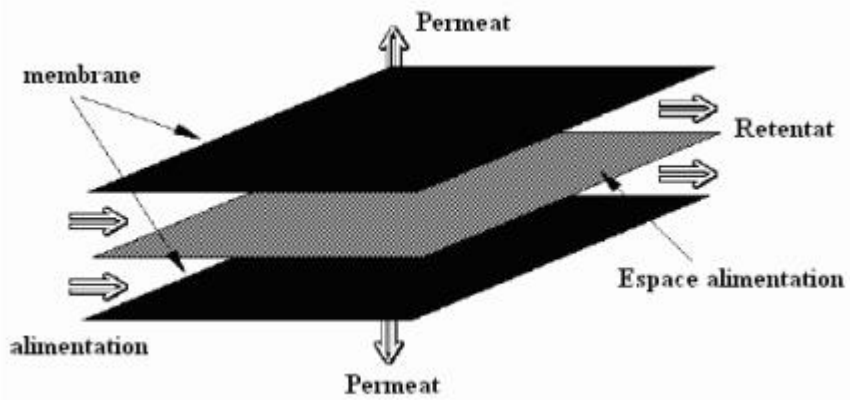


Figure I.5. Module plan.

- ✓ **Module tubulaire :** Dans les modules tubulaires, les membranes se trouvent à l'intérieur d'un tube support. Les tubes sont déposés ensuite, en série ou en parallèle dans une enveloppe cylindrique où l'eau à traiter circule. Le perméat est recueilli l'extérieur donnant naissance à un écoulement turbulent à l'intérieur, les tubes donc constituent des canaux d'écoulement tangentiel. Cette configuration nécessite un grand espace au sol pour fonctionner et des coûts d'investissement élevés à cause de l'énergie importante dépensée mais elle ne nécessite pas de préfiltration.

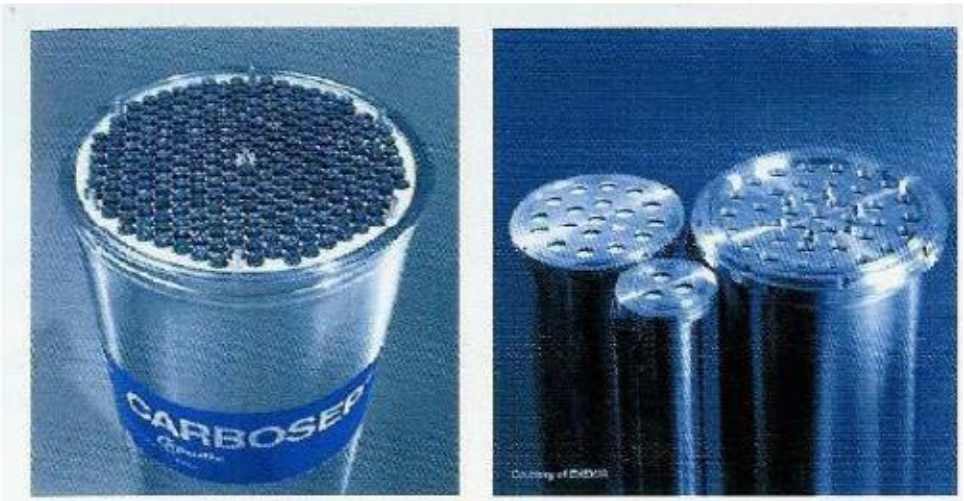


Figure I.6. Membrane tubulaire (Chabane 2015).

- ✓ **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

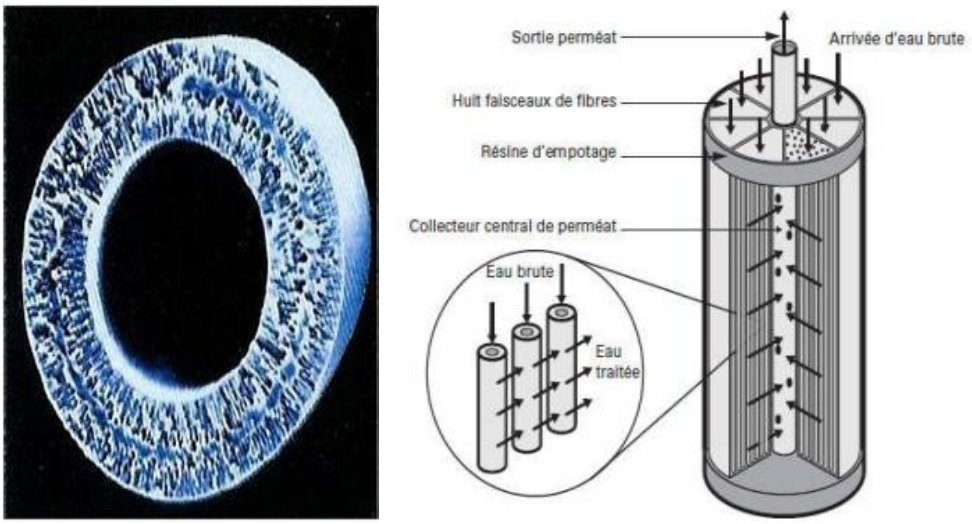


Figure I.7. Représentation schématique d'un module à fibres creuses (Lanteri 2009).

- ✓ **Module spiralé :** Le module le plus populaire dans l'industrie pour les membranes de nanofiltration ou d'osmose inverse est le module spiralé. Au sein de ces modules, une membrane plane est enroulée sur elle-même autour d'un tube collecteur, constituant l'axe central de la membrane, qui recueille le perméat.

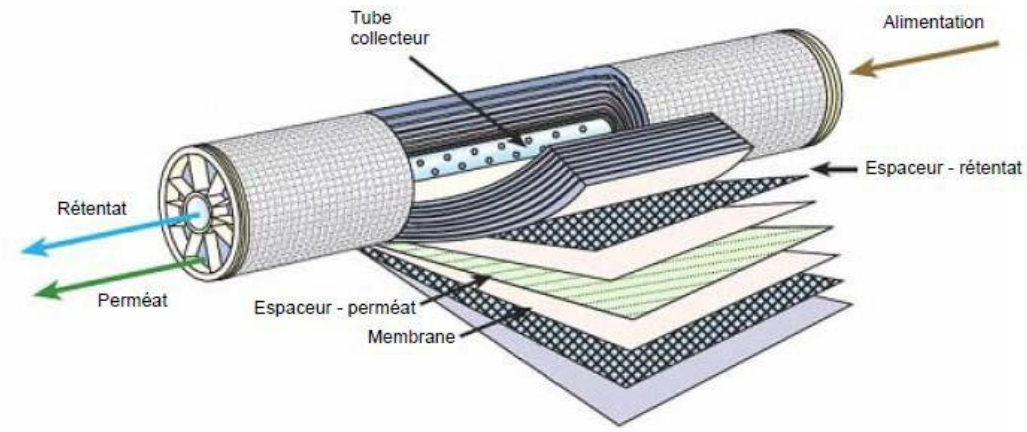


Figure I.8. Module spiralé (Antoine 2004).

I.2.3. Osmose inverse et nanofiltration

L'**osmose** inverse met en œuvre des membranes avec des pores compris entre 0,2- 1 nm. Cette méthode est la moins onéreuse pour éliminer de 90 à 99% des contaminants d'un fluide. Le phénomène d'osmose est un phénomène naturel, notamment observé à travers les membranes cellulaires où la différence de concentration entre les deux côtés d'une membrane semi-perméable crée une différence de pression, connue sous le nom de pression osmotique. Cette dernière force de passage de solvant de la solution la plus diluée vers la plus concentrée. L'osmose inverse est basée sur le principe inverse où le solvant passe de la solution le plus concentrée vers le plus diluée et ceci en appliquant une pression supérieure à la pression osmotique. Ces membranes sont utilisées dans : le dessalement de l'eau de mer, la production d'eau ultra pure, la récupération de matières précieuses, la diminution de la pollution environnementale (Darwiche 2010).

La **nanofiltration** se situe entre l'osmose inverse et l'ultrafiltration. Elle permet la séparation de composants ayant une taille en solution voisine de celle du nanomètre (soit 10 Å) d'où son nom. Les sels ionisés monovalents et les composés organiques non ionisés de masse molaire inférieure à environ 200 - 250 g/mol ne sont pas retenus par ce type de membrane. 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. Les mécanismes de transfert sont intermédiaires entre ceux de l'osmose inverse et ceux de l'ultrafiltration. Cette technique est souvent utilisée pour l'adoucissement des eaux (Jean-Marc, BERLAND et al. 2002). La nanofiltration offre la très intéressante potentialité de séparer des composés de faible poids moléculaire à des pressions qui sont faibles ou moyennes (Darwiche 2010).

I.2.4. Sélectivité

La sélectivité est définie par un taux de rétention T_R, elle dépend essentiellement de la nature et la structure d'une membrane.

$$Tr = \frac{C_{SA} - C_{SP}}{C_{SA}} \qquad (I.1)$$

Où :

C_{SA} : La concentration d'une espèce à retenir dans la solution (kg/m^3) ;

C_{SP}: La concentration de la même espèce dans le perméat (kg/m ^{3}).

Le taux de rétention Tr nous renseigne sur la sélectivité de la membrane, ainsi si : Tr = 0 : Le soluté n'est pas retenu par la membrane ;

Tr = 100 : Le soluté est totalement retenu par la membrane ;

0 < Tr < 100 : Le soluté est partiellement retenu par la membrane.

I.2.5. Perméabilité

Le principe de la perméabilité d'une membrane se définie par son aptitude à laisser passer un fluide à travers la membrane par une différence de pression entre l'entrée et la sortie d'une membrane (Nchad. 2009). La perméabilité des membranes poreuse dépend de la taille des espores alors que pour les membranes denses dépend de la nature chimique d'une membrane (Pinnau 2000). Elle est définie comme suit :

$$\mathrm{A} = \frac{Jp}{\Delta Pm} \qquad (I.2)$$

Où :

A **:** la perméabilité (m/s.Pa) ;

Jp **:** le flux de perméat (m/s) ;

ΔPm **:** la différence de pression transmembranaire (Pa).

Le flux de perméat est le rapport entre le débit volumétrique de perméat et la surface de la membrane :

$$Jp = \frac{Q_P}{S} \qquad (I.3)$$

Tel que :

Q_P : Débit volumétrique de perméat (L/h) ;

S : La surface effective de la membrane (m^2) ;

Si on remplace l'équation de flux de perméat dans la première équation on obtient (Nchad. 2009):

$$A = \frac{Q_P}{S\Delta Pm} \qquad (I.4)$$

I.2.6. Influence des paramètres opératoires

I.2.6.1. Pression

La pression de fonctionnement (Ptm) doit être supérieure à la pression osmotique. L'étude menée par Sridhar et al. 2002 a montré 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 (Padilla et Tavani 1999; Galambos et al. 2004 ; Lin et Yang 2004). De même, le taux de rétention augmente lorsque la pression augmente (Sridhar et al. 2002 ; 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 bar 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.2.6.2. 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 (Bilstad 1995), 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.2.6.3. 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) :

$$y = \frac{Q_P}{Q_A} \qquad (I.5)$$

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 (Cornelissen et al. 2001). 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).

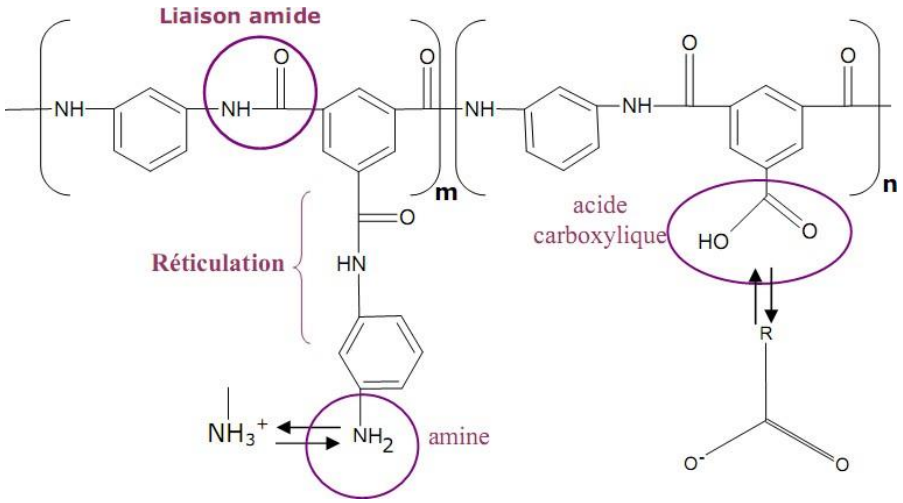


Figure I.9. Molécule de polyamide et ses équilibres acido-basiques (Sagne 2008).

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 (Kiso et al. 2001a ; Ozaki et al. 2002 ; Schafer et al. 2003 ; Kimura et al. 2004 ; Nghiem et al. 2004 ; 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 (Kosutic et 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) (Lanteri et al. 2009).

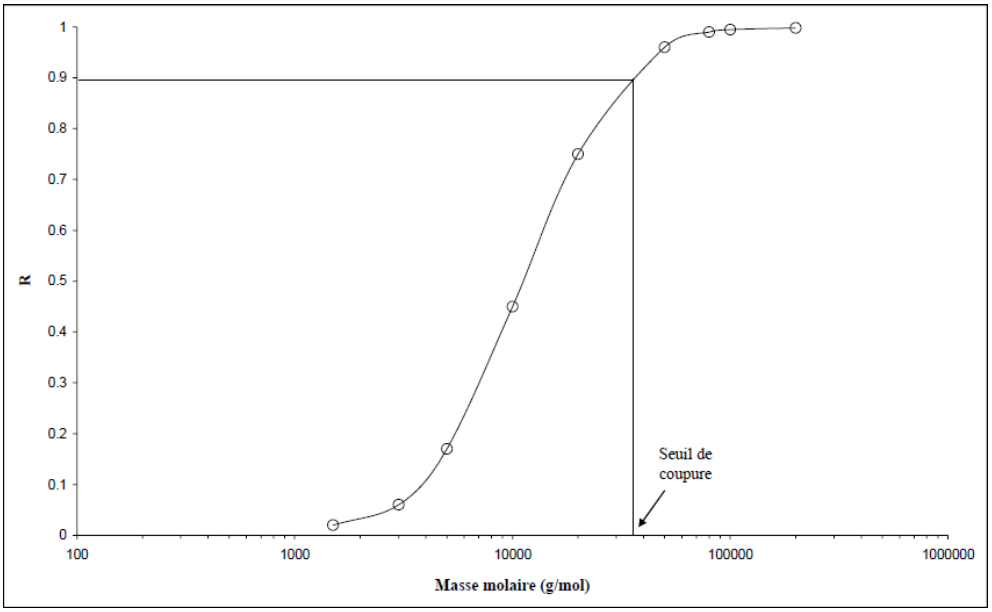


Figure I.10. Illustration du principe de détermination du seuil de coupure d'une membrane, la rétention en fonction de la masse molaire (ici pour un taux de rejet de 0.90) (Lanteri et al. 2009).

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 (Anselme et Jacobs 1996) :

- ✓ 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 (Kiso et al. 2000), de polysaccharides et d'alcools (Kiso et al. 2001a ; Bellona et al. 2004). Étude d'Agenson et al. 2003 sur 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 (Berg et al. 1997).

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 (Van der Bruggen et al. 1998). Il ne peut pas être utilisé pour la modélisation, car il ne représente pas la géométrie de la molécule (Van der Bruggen et al. 1999). Pour les molécules non chargées, peu polaires, la masse molaire peut fournir une première information sur la rétention (Kiso et al. 2001a) avec de bonnes corrélations (Ozaki et Li 2002 ; 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 (Bellona et al. 2004). En effet, des pesticides ayant une section et non une masse molaire plus faible sont moins bien retenus (Berg et al. 1997). Le poids moléculaire est considéré comme une variable de descripteur moins significatif pour la modélisation de la rétention (Yangali-Quintanilla et al. 2008). Par ailleurs, Kratbi et al. 2023; Santos et al. 2006 ont utilisé un autre paramètre pour représenter géométriquement le poids moléculaire, ils ont introduit " diamètre effectif" qui est donné par la relation suivante.

$$d_c = 0.065 M_c^{0.438} \qquad (I.5)$$

A. 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 ; Kiso et al. 1992 et 2001b).

B. 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 (Santos et al. 2006 ; Kiso et al. 1992 et kiso et al. 2010) :

$$Width = \frac{1}{2}\sqrt{Smin} \qquad (I.7)$$

C. 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 (Kiso et al. 1992 ; Santos et al. 2006) :

$$Depth = \frac{1}{2}\sqrt{Smax} \qquad (I.8)$$

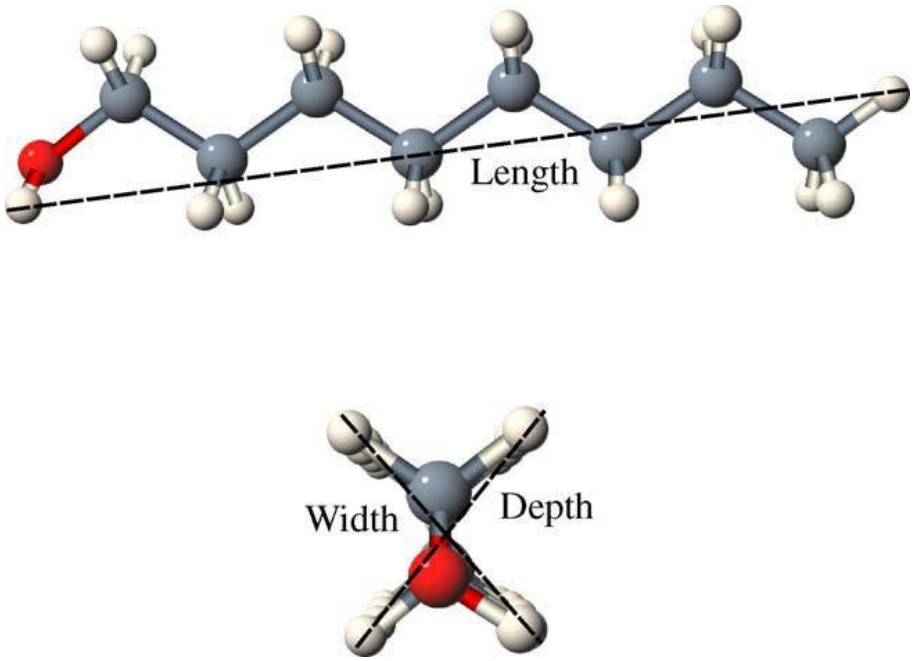


Figure I.11. Calcul des dimensions de l'octanol (*Length*, *Width* et *Depth*) de la structure 3D (Santos et al. 2006).

D. Largeur équivalente moléculaire "*EqWidth*"

La largeur moléculaire équivalente est définie comme étant la moyenne géométrique de la largeur et de la profondeur (Kiso et al. 1992 ; Santos et al. 2006 ; Linares et al. 2011).

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

$$EqWidth = \sqrt{Width ** Depth} \qquad (I.9)$$

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 der Bruggen et 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) (Childress et Elimelech 1996 ; Childress et Desmukh 1998 ; Ozaki et Li 2002 ; Ardhaoui et al. 2003 ; Schäfer et al. 2004 ; Bellona et Drewes 2005).

Le pH de l'alimentation a une influence sur la charge de la membrane, il joue sur la protonation des groupes fonctionnels (Childress et Elimelech 1996). 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 (Ozaki et Li 2002). 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'électroneutralité 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 (Berg et al. 1997 ; Ozaki et Li 2002). 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 (Van der Bruggen et al. 1999 ; Hu et al. 2003). 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 (Childress et Desmukh 1996 ; Childress et Desmukh 1998 ; Schäfer et al. 2004 ; Bellona et Drewes 2005). 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 (Kiso et al. 2001b ; 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 contact α 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.

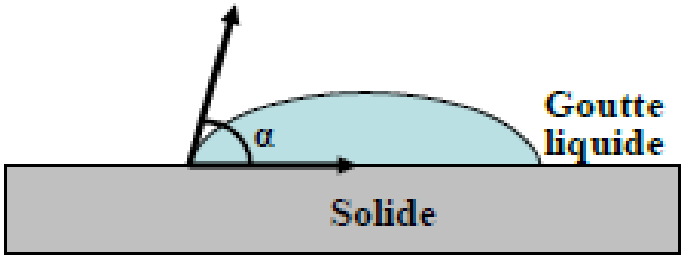


Figure I.12. Mesure de l'angle de contact par la méthode de la goutte déposée (Sagne 2008).

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** (log kow), plus souvent exprimé sous forme logarithmique (log kow, *log P*). S'il est supérieur à 2, la molécule est hydrophobe (HB) ; s'il est inférieur, la molécule est hydrophile (HL) (Connell 1989 et 1990 ; Yangali-Quintanilla et al.2009b).

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

$$logk_{ow} = log\frac{C_o}{C_w} \qquad (I.10)$$

Où C_o et C_w sont les concentrations des molécules dans les couches n-octanol et l'eau (Kiso et al. 2001a ; Yangali-Quintanilla et al. 2010b ; 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, log *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 log kow du soluté (Équations I.11 et I.12) (Verliefde et al. 2008).

Pour les acides, (Verliefde et al. 2008) **:**

$$\log D_{(pH)} = \log kow - log(1 + 10^{pH-pka}) \qquad (I.11)$$

Pour les basses, (Verliefde et al. 2008) **:**

$$\log D_{(pH)} = \log kow - \log\left(1 + 10^{pka-pH}\right) \qquad (I.12)$$

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.2.

Tableau I.2. Applications des procédés d'OI et NF.

Application	Références
Industries diverses : ▪ Récupération des métaux ▪ Production de l'eau ▪ Traitement de surface ▪ Effluents industriels	(Nielsen 1995 ; Kimura et al. 2009 ; Comerton et al. 2008 ; Metaiche 2014 ; Ravikumar et al. 2014 ; Acero et al. 2016).
Industrie agro-alimentaire : ▪ Filière œnologique ▪ Lactosérum ▪ Recyclage des protéines ▪ Effluents agro-alimentaires ▪ Industrie pharmaceutique et biotechnologique : Sérums, principes actifs, antibiotique.	(Alkhatim 1998 ; Yangali-Quintanilla et al. 2010b et 2009b ; Shahmansouri et al. 2013 ; Metaiche 2014 ; Sadmani et al. 2014a et 2014b).
Eau potable, dessalement, saumure : ▪ Élimination nano-polluants organiques, élimination des particules, adoucissement, déminéralisation.	(Alkhatim 1998 ; Marty1999 ; Abdel-Jawad et al. 2002 ; Abbas et El Bastaki 2005 ; Metaiche 2014 ; Ricci et al. 2015; 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.3.

Tableau I.3. 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 surtout 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 Rétention des anticancers lors nanofiltration

I.6.1 Cancer et médicaments anticancéreux

La croissance et le vieillissement de la population, ainsi que les modes de vie actuels adoptés, ont entraîné une augmentation mondiale des cas de cancer. Le cancer est la deuxième cause de décès dans le monde, avec un décès sur six attribué à cette maladie. En 2020, il y a eu 19,3 millions de nouveaux diagnostics de cancer, 50,5 millions de personnes vivant avec un cancer (dans les 5 ans suivant le diagnostic) et 9,96 millions de décès attribués au cancer dans le monde. Au cours des deux prochaines décennies, les décès annuels causés par le cancer et le nombre de nouveaux cas de cancer devraient atteindre respectivement 16,3 millions et 30,2 millions, ce qui indique que la consommation de médicaments anticancéreux et leur rejet conséquent dans l'environnement augmenteront considérablement (Cristavao 2022).

Les médicaments anticancéreux sont un groupe de médicaments thérapeutiques utilisés dans le traitement du cancer, qui sont conçus pour perturber ou prévenir la prolifération cellulaire en ciblant diverses voies moléculaires et métaboliques. Ces médicaments sont classés par l'organisation

mondiale de la Santé (OMS) comme agents antinéoplasiques et immunomodulateurs (classe L). Cette classe peut être divisée en quatre grands groupes : les agents antinéoplasiques (classe L01), la thérapie endocrinienne (classe L02), les immunostimulants (classe L03) et les immunosuppresseurs (classe L04).

Ces groupes peuvent être divisés en différents sous-groupes. Par exemple, le groupe des agents antinéoplasiques, qui comprend la plupart des médicaments anticancéreux utilisés, peut être divisé en agents alkylants (classe L01A), antimétabolites (classe L01B), alcaloïdes végétaux et autres produits naturels (classe L01C), antibiotiques cytotoxiques et substances apparentées (classe L01D), inhibiteurs de protéines kinases (classe L01E) ainsi que d'autres agents antinéoplasiques (classe L01X), tandis que la classe de thérapie endocrinienne peut être divisée en hormones et agents apparentés (classe L02A) et dans les antagonistes hormonaux et agents apparentés (classe L02B).

La division des médicaments anticancéreux en différents groupes et sous-groupes dépend de divers facteurs tels que les structures chimiques, l'utilisation thérapeutique, le mode d'action et l'origine source. Par exemple, la principale action pharmacologique des agents alkylants est de perturber la synthèse de l'ADN et d'empêcher la réplication cellulaire en ajoutant des groupes alkyles à l'ADN. Deux exemples d'agents alkylants sont le cyclophosphamide et l'ifosfamide. Les antimétabolites sont connus pour interférer avec la production de métabolites nucléotidiques majeurs ou agir comme substituts de métabolites naturels, provoquant la perturbation de la synthèse des acides nucléiques et inhibant ainsi la croissance cellulaire. La capécitabine, le 5-fluorouracile et le méthotrexate sont des exemples d'antimétabolites. De plus, les médicaments de thérapie endocrinienne, comme le tamoxifène, sont utilisés pour les cancers dont la croissance dépend des hormones et agissent en affectant la synthèse hormonale ainsi qu'en modifiant l'expression hormonale à l'intérieur de la cellule v (Cristavao 2022).

Les médicaments anticancéreux ont un mécanisme d'action très puissant et même si ces médicaments sont conçus pour tuer les cellules à croissance rapide, telles que celles présentes dans les tumeurs, de nombreux médicaments anticancéreux présentent un manque de sélectivité et, outre les cellules tumorales, ils peuvent également attaquer les cellules en croissance normales, provoquant potentiellement des effets cytotoxiques, génotoxiques, mutagènes, tératogènes et perturbateurs endocriniens dans tout organisme vivant eucaryote. De nombreuses études ont examiné les risques potentiels pour la santé associés à la manipulation de médicaments anticancéreux et une augmentation des taux urinaires de ces médicaments, un risque d'avortement ainsi que des aberrations

chromosomiques structurelles ont été observés chez les travailleurs exposés professionnellement (Cristavao 2022).

La présence de médicaments anticancéreux en milieu aquatique peut donc être très préoccupante et il est donc crucial de surveiller régulièrement la présence de ces médicaments dans les différents compartiments aquatiques.

Néanmoins, il n'est pas possible de surveiller tous les médicaments anticancéreux en raison du nombre et de la diversité élevés des médicaments actuellement utilisés ainsi que des coûts et du temps élevés associés aux méthodes analytiques utilisées pour leur détection. Selon une base de données en libre accès des médicaments anticancéreux autorisés, au 1er juin 2022, il y avait 287 médicaments anticancéreux autorisés, dont 259 étaient approuvés par la Food and Drug Administration (FDA) et 190 par l'Agence Européenne des Médicaments (EMA) (Cristavao 2022). Il est donc extrêmement important de mettre en œuvre des méthodologies permettant de prédire à quelles concentrations les différents médicaments anticancéreux devraient se produire et lesquels de ces médicaments sont les plus susceptibles d'être trouvés dans le milieu aquatique. Une approche utile pourrait être de mettre en œuvre le modèle des concentrations environnementales prédites (CEP) proposé par l'EMA, qui permet d'estimer la concentration de produits pharmaceutiques dans les eaux de surface en tenant compte de différents paramètres, tels que les données de consommation de chaque médicament cible, la fraction d'excrétion, la fraction d'élimination dans la STEP ainsi que le nombre d'habitants de la zone étudiée (Agency 2006).

I.6.2 Médicaments anticancéreux en milieu aquatique : sources et suivi

Au cours des dernières années, il était communément admis que les effluents hospitaliers étaient la principale source d'émission de médicaments anticancéreux dans le milieu aquatique en raison de leur administration élevée dans les hôpitaux. Même si les effluents hospitaliers sont des matrices complexes, dans la plupart des pays, ces effluents sont directement rejetés dans les stations d'épuration sans aucun traitement préalable, ce qui fait des hôpitaux des sources majeures d'intrants de médicaments anticancéreux dans l'environnement. Néanmoins, les effluents ménagers sont également des sources clés d'émissions de médicaments anticancéreux dans le milieu aquatique, car de nombreux patients cancéreux sous chimiothérapie dite quotidienne rentrent chez eux après avoir reçu un traitement à l'hôpital et par conséquent, une grande partie des médicaments administrés finissent par être excrétés à la maison (en fonction de la demi-vie d'élimination du corps humain de chaque médicament). De plus, en raison du développement de nouveaux médicaments et de la tentative d'améliorer le bien-être des personnes sous traitement, il y a une tendance à l'augmentation

du nombre de patients prenant des médicaments anticancéreux à domicile. De plus, les fabricants de médicaments et leurs effluents ainsi que l'élimination des déchets de médicaments peuvent également être des sources potentielles d'intrants de médicaments anticancéreux.

I.6.3 Effets de toxicité

Même si des médicaments anticancéreux ont été trouvés dans l'environnement aquatique à des concentrations extrêmement faibles et ont donc reçu moins d'attention par rapport à d'autres groupes thérapeutiques, le rejet continu de ces médicaments dans l'environnement aquatique peut potentiellement avoir des effets négatifs sur les organismes aquatiques et les humains.

Les essais aigus peuvent sous-estimer la toxicité potentielle réelle des médicaments anticancéreux, car ces médicaments sont continuellement libérés dans l'environnement et par conséquent, les organismes non ciblés sont exposés à long terme aux faibles concentrations environnementales de médicaments anticancéreux. Les dosages chroniques peuvent donc être le moyen le plus pertinent d'évaluer le risque écologique réel des faibles concentrations environnementales de médicaments anticancéreux.

En général, la plupart des concentrations signalées comme induisant des effets de toxicité sont de l'ordre du $mg.L^{-1}$ qui sont substantiellement plus élevés que ceux détectés dans les effluents d'eaux usées et les eaux de surface. Néanmoins, en fonction de l'organisme testé, du critère d'évaluation évalué ainsi que du temps d'exposition, on peut voir que le même médicament anticancéreux peut induire des effets toxiques à une large gamme de concentrations et en effet, de faibles concentrations environnementales ont déjà été signalées pour induire des effets négatifs dans les organismes aquatiques (Cristavao.2022).

I.6.4 Technologies de traitement explorées pour l'élimination des médicaments anticancéreux

Comme le souligne la section présidente, des médicaments anticancéreux ont été détectés dans les effluents de différentes stations d'épuration, ce qui indique que, pour certains de ces médicaments, les traitements conventionnels ne sont pas efficaces pour assurer l'élimination complète de ces micropolluants.

Des études en laboratoire simulant le traitement conventionnel des boues activées ont en effet démontré que certains médicaments anticancéreux ont une dégradabilité biologique limitée. Par exemple, (Buergue et coll 2006) ont rapporté que le cyclophosphamide et l'ifosfamide n'étaient pas dégradés après 24 heures d'exposition aux boues activées. Dans une étude similaire, (Franquet-Griell et al. 2017) ont observé qu'après 8 heures de temps de rétention dans un bioréacteur, 7 des 16

composés cibles étaient encore présents dans l'effluent, à des concentrations allant de 10 à 80% de la concentration initiale. Différentes technologies de traitement ont ainsi été étudiées pour faire face à l'élimination des médicaments anticancéreux.

De plus, même si plusieurs études ont montré l'efficacité de l'osmose inverse (OI) et de la nanofiltration (NF) pour l'élimination des produits pharmaceutiques très peu d'études ont étudié l'élimination des médicaments anticancéreux. Dans ces études, le cyclophosphamide était le composé le plus étudié et, à titre d'exemple, des éliminations supérieures à 90% ont été obtenues pour le cyclophosphamide lors de la filtration d'un dopé eau de surface avec deux membranes NF (Trisep TS-80 et Desal HL). Il est important de souligner que les processus d'adsorption et membranaires permettent l'élimination et la rétention des contaminants cibles, cependant, il n'y a pas de dégradation de ces composés (Cristavao.2022).

Au cours des trois dernières décennies, l'utilisation de procédés de filtration membranaire pour l'élimination d'un large éventail de contaminants de l'eau et des eaux usées a considérablement augmenté. La demande accrue en eau associée à des réglementations plus strictes en matière de qualité de l'eau ainsi qu'à l'intérêt croissant pour la réutilisation de l'eau ont contribué à l'importance croissante des procédés membranaires. Comparés à d'autres technologies de traitement, les procédés membranaires offrent les avantages de produire des effluents d'excellente qualité (garantissant une élimination élevée des contaminants cibles), d'être faciles à mettre à l'échelle et à intégrer à d'autres procédés de traitement (Cristavao 2022).

I.7. Revue des travaux de la rétention des anticancers lors nanofiltration

De nombreux travaux réalisés au niveau international mettent en évidence la rétention des anticancers lors nanofiltration. La base de données consultée est Science Direct. La figure I.13, reporte le nombre de publications de 1995 à 2024 pour les membranes de la nanofiltration.

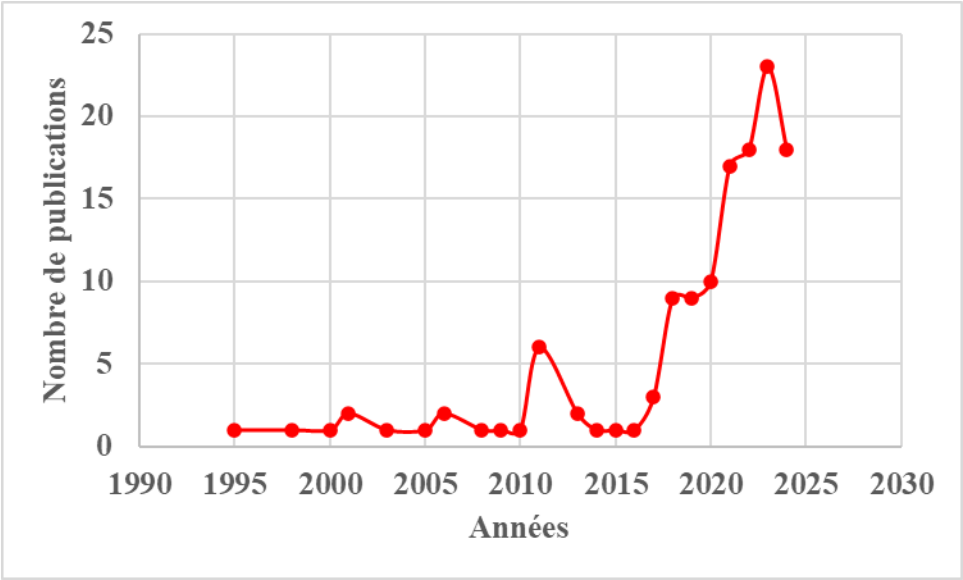


Figure I.13. Nombre de publications de 1995 à 2024 pour les membranes de la nanofiltration.

I.8. Conclusion

Dans cette étude bibliographique, nous nous sommes attachés à poser les bases théoriques de notre travail de manière précise.

La rétention des molécules organiques par les membranes de la nanofiltration et d'osmose 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.

L'usage de la nanofiltration membranaire pour éliminer une large gamme de médicaments anticancéreux des effluents des différentes stations d'épuration a considérablement augmenté.

Chapitre II

OSPR-Réseau de Neurones Artificiels



II.1. Introduction

La Chemoinformatique est un nouveau domaine apparu il y a quelques années. Il concerne le développement, la création, l'organisation, le stockage, la diffusion, l'analyse, la visualisation et l'utilisation de l'information chimique, elle représente un domaine en évolution rapide. Actuellement, les méthodes spécifiques de la chemoinformatique sont devenues indispensables pour le développement de nouveaux composés, matériaux et processus.

La chemoinformatique est une discipline scientifique qui a beaucoup évoluée durant ces 40 dernières années. Cette science est à l'interface entre la chimie et l'informatique. Il a été rendu compte que dans beaucoup de domaines de la chimie moderne, le nombre important de données et d'informations fourni par la recherche en chimie peut seulement être analysé par des méthodes informatiques. De plus, les nouveaux problèmes rencontrés en chimie sont si complexes que l'utilisation de l'informatique a été rendue indispensable. C'est pourquoi des méthodes ont été développées pour prédire les propriétés chimiques, physiques et biologiques des matériaux et des composés chimiques. Ces méthodes ont aussi été développées pour la représentation des médicaments, l'élucidation des structures cristallines, dans la prédiction des réactions chimiques et dans la représentation des molécules organiques. La recherche et le développement en chemoinformatique sont essentiels pour :

- L'amélioration de la compréhension des phénomènes chimiques ;
- Permettre à l'industrie chimique de rester compétitive dans le marché mondial (http://www.chemspider.com/Chemical-Structure.6886.html?rid=0ae6e03a-77d9-40b1a7cd-44816ea08f94&page_num=0).

Ces dernières années, il y a eu plusieurs tentatives pour faire progresser l'utilisation des relations quantitatives structure-propriétés (QSPR) 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'étude sont également appliqué les modèles QSPR 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, Ammi et al. 2015, 2017, 2018, 2020, 2021, 2021, Kratbi 2023).

Ce chapitre a été consacré à des modèles de réseaux de neurones artificiels basés sur la relation quantitative structure-activité "QSAR-RN". On parlera sur la méthode d'apprentissage statistique (relation quantitative structure-activité "QSAR" et les réseaux de neurones artificiels "RNA"). Une

conclusion clôturera ce chapitre et sera une synthèse par laquelle on justifiera le choix du thème de ce mémoire.

II.2. Méthodes d'apprentissages statistiques

La théorie de l'apprentissage statistique est un système d'apprentissage automatique à partir des domaines de la statistique et de l'analyse fonctionnelle. La théorie de l'apprentissage statistique traite du problème de la recherche d'une fonction prédictive basée sur des données. La théorie de l'apprentissage statistique a conduit à des applications dans des domaines tels que la vision par ordinateur, la reconnaissance de la parole, la bioinformatique(https://fr.wikipedia.org/wiki/Bio-informatique).

II.3. Relations quantitatives structure-Activités (QSAR)

Initiées sur la base de relations empiriques durant la seconde partie du 19ème siècle, les QSAR ont été mathématiquement formalisées dans les années 1930 et leur acceptation, en tant que discipline à part entière, a été définitivement acquise au début des années 1960 suites aux travaux de Hansch et Fujita (Devillers, 2013) qui ont proposé un modèle mathématique pour corréler l'activité biologique et la structure chimique. Depuis lors, un grand nombre de modèles a été conçu pour la prédiction de diverses activités d'intérêt biologique.

Il ya de nombreuses utilités dans l'utilisation des méthodes QSAR. Nous donnons ci-dessous, à titre d'exemple, quelques-unes :

- 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 ;
- la prédiction d'une variété de propriétés physico- chimiques des molécules (qu'il s'agisse de produits pharmaceutiques, de pesticides, de produits d'hygiène personnelle, la chimie fine, etc.) (Hamadache 2014).

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.3.2. Données biologiques

Les données biologiques sont habituellement exprimées sur une échelle logarithmique en raison de la relation linéaire entre la réponse et le logarithme de dose dans la région centrale de la courbe de log dose-réponse. Les logarithmes inverses de l'activité (log 1/C) sont également utilisés pour obtenir des valeurs mathématiques plus élevées lorsque les structures sont biologiquement très efficaces. Parmi les données biologiques utilisées dans l'analyse de QSAR, on trouve la constante de vitesse, la constante d'inhibition, … etc (Hamadache 2014).

II.3.3. Descripteurs moléculaires

De nombreuses recherches ont été menées, au cours des dernières décennies, pour trouver la meilleure façon de représenter l'information contenue dans la structure des molécules. Ainsi, les recherches ont abouti à la représentation de la structure moléculaire par un ensemble de valeurs numériques appelées descripteurs moléculaires. On en dénombre aujourd'hui plus de 3000 types, qui quantifient des caractéristiques physico-chimiques ou structurelles de molécules. Ils peuvent être obtenus de manière expérimentale, mais les descripteurs calculés, et non mesurés, sont à privilégier : ils permettent en effet d'effectuer des prédictions sans avoir à synthétiser les molécules, ce qui est un des objectifs 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. Ces derniers sont à nouveaux classés en trois catégories en fonction de leur dimension (Hamadache 2014).

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.

B. **Descripteurs 2D**

Ils sont calculés à partir de la formule développée de la molécule. Ils peuvent être de plusieurs types :

- Les indices constitutionnels caractérisent les différents composants de la molécule. Il s'agit par exemple du nombre de liaisons simples ou multiples, du nombre de cycles…
- Les indices topologiques peuvent être obtenus à partir de la structure 2D de la molécule, et donnent des informations sur sa taille, sa forme globale et ses ramifications.
- L'indice de Wiener par exemple, permet de caractériser le volume moléculaire et la ramification d'une molécule. Un autre indice, nommé "indice de Randic", est l'un des descripteurs les plus utilisés. Il peut être interprété comme une mesure de l'aire de la molécule accessible au solvant.

C. Descripteurs 3D

Ces descripteurs 3D reflètent les propriétés physiques dans la plupart des cas, mais sont insuffisants pour expliquer de façon satisfaisante certaines propriétés ou activités, telles les activités biologiques. Aussi, des descripteurs représentant l'information dérivée de la représentation en trois dimensions des molécules ont pu être calculés grâce au développement des techniques instrumentales. Ils sont évalués à partir des positions relatives de ses atomes dans l'espace, et décrivent des caractéristiques plus complexes. Ces descripteurs apportent davantage d'informations, et sont nécessaires à la modélisation de propriétés ou d'activités qui dépendent de la structure 3D. On distingue plusieurs familles importantes de descripteurs 3D :

- Les descripteurs géométriques les plus importants sont le volume moléculaire, la surface accessible au solvant, le moment principal d'inertie.

- Les descripteurs 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 : auto correlations 2D, descripteurs géométriques, descripteurs RDF, descripteurs 3D-MoRSE, descripteurs WHIM, descripteurs GETAWAY, …etc (Hamadache 2014).

D. Descripteurs 4D

Prend en compte la flexibilité conformationnelle de la molécule au cours du temps.

Les molécules ne sont pas structures statiques. Elles peuvent se plier, se tordre et vibrer de différentes maniéres, et ces différentes conformations peuvent affecter la facon dont elles interagissent avec d'autres molécules, ce qui peut a son tour influencer leur activité.

II.3.4. Formulation mathématique des QSAR

Les équations QSAR prennent la forme générale suivante :

$$QSAR \begin{cases} Activité\ biologique\ =\ fonction\ (Descripteurs) \\ Activité\ thérapeutique\ =\ fonction\ (Descripteurs) \end{cases}$$

$$QSPR : Propriétés = f(Descripteurs)$$

$$QSTR : Toxicités = f(Descripteurs)$$

Dans l'équation, l'activité biologique est normalement exprimée comme log [1/ (symbole de concentration)], où C'est généralement la concentration minimum requise pour causer une réponse biologique définie. La relation mathématique entre un paramètre ou plusieurs paramètres physico-chimiques et l'activité biologique d'un composé peut être exprimée à travers l'équation de Hansch :

$$\log\left(\frac{1}{c}\right) = b_0 + \sum_{i=1}^{j}(b_i D_i) \qquad \text{(II. 1)}$$

Dans cette équation, C'est la concentration de substance nécessaire pour obtenir une réponse biologique, $b_0$est une constante, b_i est un coefficient, $D_I$est un descripteur moléculaire et j est le nombre de descripteurs utilisés.

Les relations QSARs les plus connues ont été introduites par Hansh et Fujita ((Poezevara, 2011). Une des équations les plus utilisées présume que les propriétés physiques responsable de l'activité biologique peuvent être séparées en trois catégories : (i) une composante hydrophobique (π), (ii) une composante électronique (σ) et (iii) une composante stérique (ES, RM) :

$$log\left(\frac{1}{c}\right) = a + b\pi + c\pi^2 + dE_s + eRM + f\sigma \qquad \text{(II. 2)}$$

a, b, c, d, e et f sont des coefficients déterminés durant l'étude QSAR, tandis que π, ES, RM et σ sont des paramètres fonctions des substituants. L'équation de Hansch est utilisée pour prévoir l'activité biologique de composés dont la structure est semblable à ceux utilisés pour former cette équation. Cependant, dans le cas où la valeur prévue est différente de celle déterminée par l'expérimentation, cela peut signifier que l'activité biologique réelle des composés n'est pas prise en compte par les paramètres mis en jeu dans le processus de construction de l'équation (Hamadache 2014).

II.4. Réseaux de neurones artificiels (RNA)

II.4.1. Histoire des réseaux de neurones artificiels

Un réseau de neurones artificiels est un système informatique matériel et / ou logiciel dont le fonctionnement est calqué sur celui des neurones du cerveau humain. Il s'agit là d'une variété de technologie Deep Learning (apprentissage profond), qui fait elle-même partie de la sous-catégorie d'intelligence artificielle du Machine Learning (apprentissage automatique).

Le concept des réseaux de neurones artificiels futinventé en 1943 par deux chercheurs de l'Université de Chicago : le neurophysicien Warren McCullough, et le mathématicien Walter Pitts. Dans un article

publié dans le journal Brain Theory, les deux chercheurs présentent leur théorie selon laquelle l'activation de neurones est l'unité de base de l'activité cérébrale.

En 1957, le Perceptron fut inventé. Il s'agit du plus ancien algorithme de Machine Learning, conçu pour effectuer des tâches de reconnaissance de patterns complexes. C'est cet algorithme qui permettra plus tard aux machines d'apprendre à reconnaître des objets sur des images.

Malheureusement, à l'époque, les réseaux de neurones étaient limités par les ressources techniques. Par exemple, les ordinateurs n'étaient pas assez puissants pour traiter les données nécessaires au fonctionnement des réseaux de neurones. C'est la raison pour laquelle la recherche dans le domaine des Neural Networks est restée en sommeil durant de longues années.

Il aura fallu attendre le début des années 2010, avec l'essor du Big Data et du traitement massivement parallèle, pour que les Data Scientists disposent des données et de la puissance de calcul nécessaires pour exécuter des réseaux de neurones complexes. En 2012, lors d'une compétition organisée par Image Net, un Neural Network est parvenu pour la première fois à surpasser un humain dans la reconnaissance d'image.

C'est la raison pour laquell**e** cette technologie est de nouveau au cœur des préoccupations des scientifiques. A présent, les réseaux de neurones artificiels ne cessent de s'améliorer et d'évoluer de jour en jour (https://www.lebigdata.fr/reseau-de-neurones-artificiels definition#:~:text=(apprentissage%20automatique).,L'histoire%20des%20r%C3%A9seaux%20de%20neurones%20artificiels,et%20le%20math%C3%A9maticien%20Walter%20Pitts).

II.4.2. Neurone biologique

L'idée de réseaux de neurones artificiels a été inspirée de la manière par laquelle les neurones biologiques traitent les informations. Le cerveau humain est composé de cellules distinctes appelées neurones formant une ensemble dense d'environ 10 à 100 milliards d'unités interconnectées. Tel qu'il est illustré sur la figure II.1, le neurone est une cellule composée d'un corps cellulaire et d'un noyau. Le corps cellulaire sera méfié pour former les dendrites. C'est par les dendrites que l'information est acheminée de l'extérieur vers le soma (corps du neurone) (Ammi 2016).

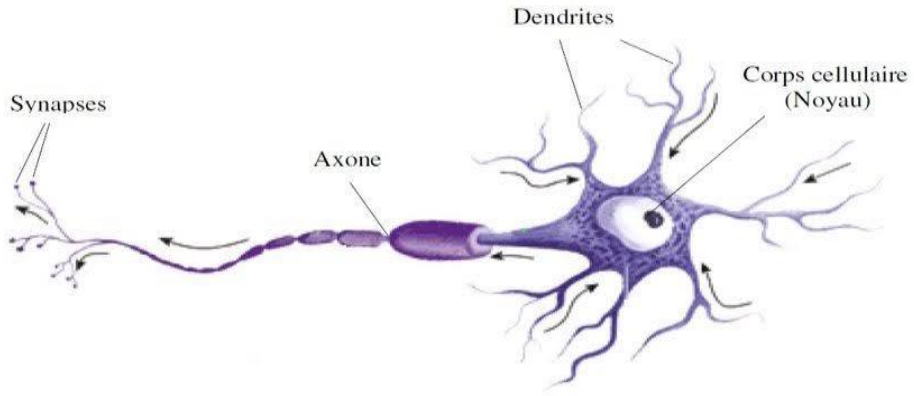


Figure II.1. Représentation d'un neurone biologique par Nicolas Rougier (https://www.researchgate.net/figure/Figure-III1-Representation-schematique-dun-neurone-biologique-Le-systeme-nerveux_fig1_358676241).

Un neurone biologique reçoit des entrées ou signaux transmis par d'autres neurones (interaction dendrites- synapse). Au niveau du corps (soma), le neurone analyse et traite ces signaux en les sommant. Si le résultat obtenu est supérieur au seuil d'activation (ou d'excitabilité), il envoie une décharge alors nommé potentiel d'action le long de son axone vers d'autres neurones biologiques.

II.4.3. Neurone formel

Le neurone formel est l'élément essentiel d'un réseau de neurones. C'est un opérateur mathématique très simple, dont on peut facilement calculer la valeur numérique (Hamadache 2014). Son fonctionnement est schématisé sur la figure II.2.

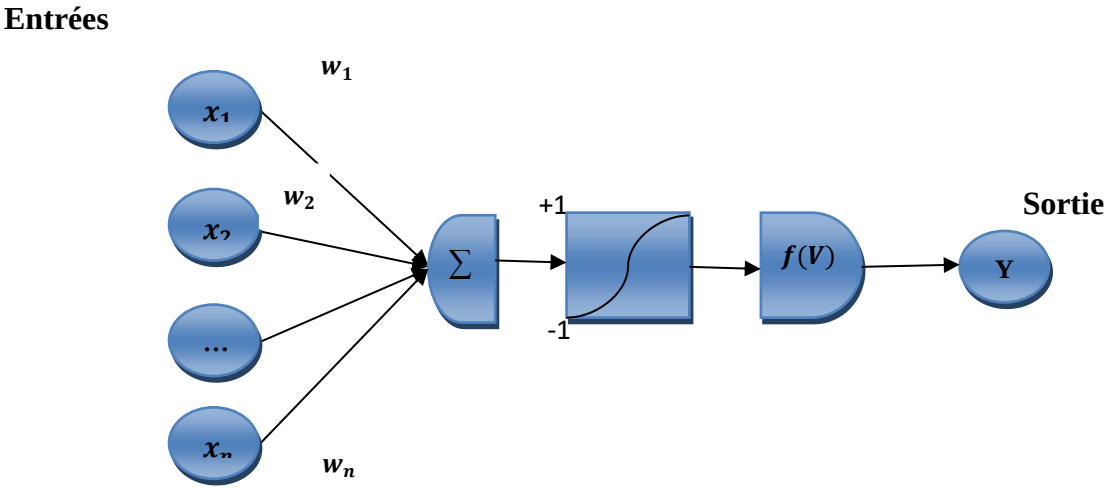


Figure II.2. Neurone formel (Ammi 2016).

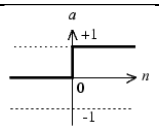
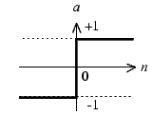
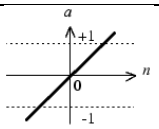
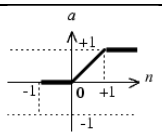
Un neurone formel est une fonction algébrique paramétrée, non linéaire en ses paramètres. Ses entrées peuvent être les sorties d'autres neurones ou des entrées de signaux extérieurs. Sa sortie est une fonction non linéaire *f* d'une combinaison linéaire *v* des entrées (x_i). Le potentiel *v* le plus fréquemment utilisé est la somme pondérée des entréesx_i pondérées par les coefficients (w_i) également appelés poids de connexions :

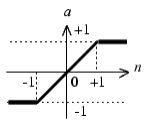
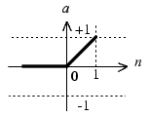
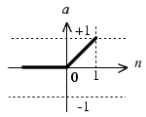
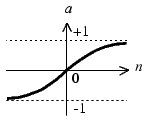
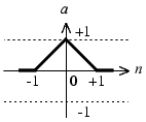
$$v = w_0 + \sum_{i=1}^{n} (w_i x_i) \qquad \text{(II. 3)}$$

Précisons que le poids $w_0$est affecté à une entrée constante appelée "biais". Une fonction *f*, appelée fonction d'activation, est appliquée à ce potentiel *v*. Le choix d'une fonction d'activation se révèle dans certains cas être un élément constitutif important des réseaux de neurones. Les fonctions mathématiques les plus souvent utilisées dans notre travail sont regroupées dans le tableau II.1. À la différence des neurones biologiques dont l'état est binaire, la plupart des fonctions de transfert sont continues, offrant une infinité de valeurs possibles comprises dans l'intervalle [0, +1] ou [-1, +1]. Un neurone formel, réalise donc l'opération suivante (Ammi 2016) :

$$y = f(v) = f\left(w_0 + \sum_{i=1}^{n} (w_i x_i)\right) \qquad \text{(II. 4)}$$

Tableau II.1. Fonctions de transfert usuelles, a=f(n) (Si moussa 2010).

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

Un neurone formel est une représentation artificielle et schématique d'un neurone biologique (Figure II.3) :

- Les synapses sont modélisées par des poids,
- Le soma ou corps cellulaire est modélisé par la fonction de transfert, appelé aussi fonction d'activation
- L'axone par l'élément de sortie (https://scilogs.fr/intelligence-mecanique/reprenons-bases-neurone-artificiel-neurone-biologique/).

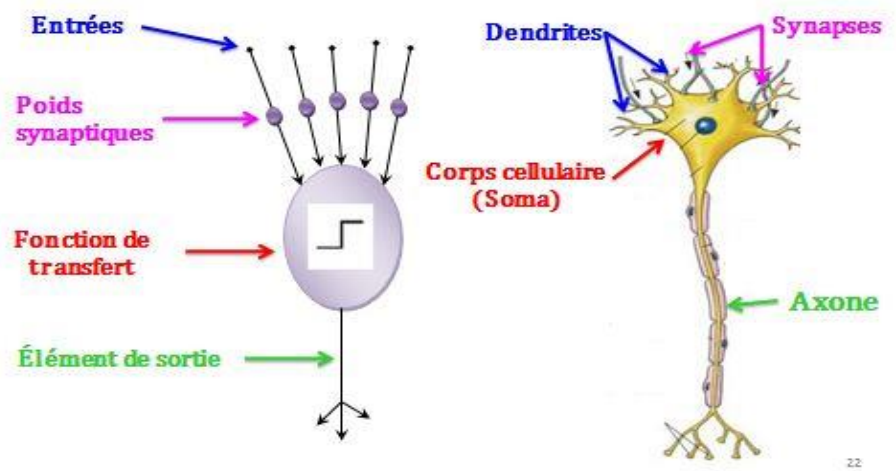


Figure II.3. Analogie entre neurone biologique et neurone formel (Ammi 2021).

II.4.4. Architecture d'un réseau de neurones

Un RNA est formé d'un ensemble de neurones formels interconnectés. Deux grands types d'architectures peuvent être distingués (Figure II.4) :

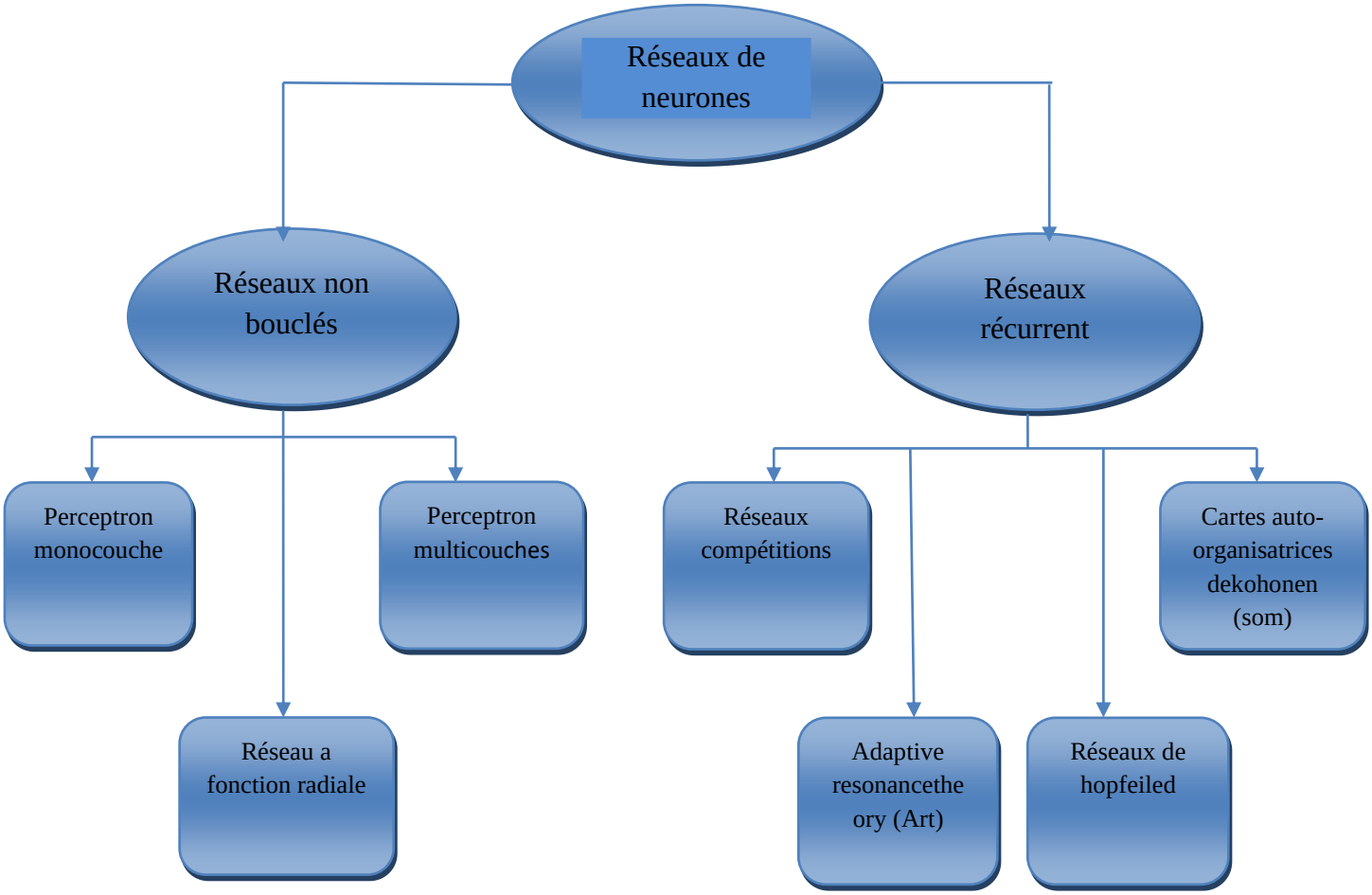


Figure II.4. Réseaux de neurones les plus utilisés.

- **Les réseaux de neurones non bouclés :** Ils réalisent une (ou plusieurs) fonction(s) algébrique(s) des entrées, par composition des fonctions réalisées par chaque neurone. Les réseaux de neurones non bouclés sont des outils statiques utilisés principalement pour effectuer des tâches d'approximation de fonctions non linéaires, de modélisation de processus statiques non linéaires (Figure II.5).

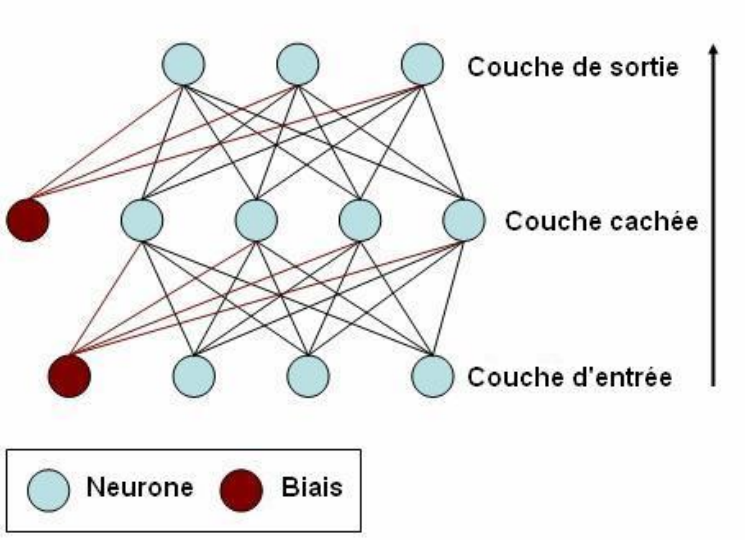


Figure II.5. Structure d'un Réseaux de neurones non bouclés (http://www.statsoft.fr/concepts-statistiques/reseaux-de-neurones-automatises/reseaux-de-neurones-automatises.htm#.YjJNDtXMLIU).

- **Les réseaux de neurones bouclés :** Ils constituent quant à eux un système dynamique temps discret", régi par une (ou plusieurs) équation(s) aux différences non linéaires, résultant de la composition des fonctions réalisées par chacun des neurones et des retards associés à chacune des connexions. Ils sont utilisés pour effectuer des tâches de modélisation et d'adaptation de systèmes dynamiques, de commande de processus ou de filtrage. Dans notre étude, nous avons adopté le modèle le plus couramment utilisé, le Perceptron Multicouches (PMC), illustré par la figure II.6 et dont les principes sont exposés dans ce qui suit (Ammi 2016).

II.4.5. Perceptrons multicouches

Le Perceptron multicouches "PMC" est un réseau de neurones non bouclé caractérisé par une structure bien particulière. La figure II.5 illustre la structure de ce type de réseau. Ces neurones sont organisés en couches successives où les informations circulent dans un seul sens, de la couche d'entrée vers la couche de sortie. Les neurones d'une même couche ne sont pas interconnectés. Un neurone ne peut envoyer son résultat qu'à un neurone situé dans une couche postérieure à la sienne. Le réseau est formé de plusieurs couches, qui s'appellent, par convention :

- couche d'entrée : cette couche représente toujours une couche associée aux entrées du système, composée des cellules d'entrée qui correspondent aux k variables d'entrée.

- couche(s) cachée(s) : chaque couche est composée de p neurones. Ces neurones n'ont aucun lien avec l'extérieur et sont appelés neurones cachés.
- couche de sortie : elle est constituée par la valeur de sortie de la combinaison linéaire des fonctions de transferts de la couche cachée. Les neurones sont reliés entre eux par des connexions pondérées. Ce sont les poids de ces connexions qui gouvernent le fonctionnement du réseau et programment une application de l'espace des entrées vers l'espace des sorties, à l'aide de la transformation non linéaire (Hamadache 2014).

II.4.6. Types d'apprentissage des réseaux de neurones

L'apprentissage est vraisemblablement la propriété la plus intéressante des réseaux de neurones. L'apprentissage est une phase du développement du réseau de neurones durant laquelle on calcule les poids des neurones de telle manière que les sorties du réseau soient aussi proches que possible des sorties désirées. Les procédures d'apprentissage sont divisées, elles aussi, en deux grandes catégories : apprentissage supervisé et apprentissage non supervisé (Khaouane 2013).

A. Apprentissage supervisé

L'apprentissage supervisé implique l'existence d'un professeur qui a pour rôle d'évaluer le succès (ou l'échec) du réseau quand il lui est présenté un stimulus connu (on dit que ce stimulus est un exemple appartenant à la base d'apprentissage). Cette supervision consiste à renvoyer au réseau une information lui permettant de faire évoluer ses connexions (parfois aussi sa propre architecture) afin de faire diminuer son taux d'échec. L'information peut être explicite, sous la forme d'une mesure de l'erreur commise, par exemple, ou globalement sur l'ensemble des exemples de la base. C'est ce qui se passe dans le cas du perceptron. Mais elle peut être plus implicite (apprentissage par renforcement), sous forme d'une simple appréciation (bon ou mauvais, punition ou récompense), sans mesure d'erreur et même être globale, sur l'ensemble des tâches que le réseau doit exécuter. Dans ce dernier cas, il est facile de comprendre que l'apprentissage est le plus difficile, la difficulté majeure consistant pour le réseau à identifier les étapes du processus qui sont responsables de l'échec ou du succès (Khaouane 2013).

B. Apprentissage non supervisé

L'apprentissage non supervisé implique la fourniture à un réseau autonome d'une quantité suffisante d'exemples contenant des répétitions (autrement dit de la redondance), telles que celui-ci

en dégage les régularités automatiquement. Ces réseaux sont souvent appelés auto-organisateurs, ou encore à apprentissage compétitif. Dans l'apprentissage non supervisé, les données ne contiennent pas d'informations sur une sortie désirée, il n'y a pas de superviseur ou expert humain. Il s'agit de déterminer les paramètres du réseau de neurones suivant un critère à définir. Dans ce cas, les exemples présentés à l'entrée provoquent une auto adaptation du réseau afin de produire des valeurs de sortie qui soient proches en réponse pour des valeurs d'entrées similaires. Ce type d'apprentissage possède souvent moins de complexité dans le calcul par rapport à l'apprentissage supervisé. Bien entendu, l'architecture du réseau, préalablement définie par son utilisateur, est une forme de supervision (Khaouane 2013).

II.4.7. Réseaux de neurones les plus utilisés

A. Perceptron monocouche (perceptron simple)

Le perceptron est le premier modèle des réseaux de neurones inventé en 1957 par Frank Rosenblatt (Rosenblatt 1958). C'est un réseau simple, puisqu'il ne se compose que d'une couche d'entrée et d'une couche de sortie. 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).

B. Le perceptron multicouche (en anglais Multi Layer Perceptron)

C'est une extension du précédent, avec une ou plusieurs couches cachées entre l'entrée et la sortie (Keedwell et Narayanan 2005). Chaque neurone dans une couche est connecté à tous les neurones de la couche précédente et de la couche suivante. 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.

C. Réseaux à fonction radiale

Les réseaux à fonction de bases radiales (RFR) ou plus simplement réseaux à bases radiales ont été proposés par Moody et Darken (Moody et Darken 1989). Les RFR seront donc employés dans les mêmes types de problèmes que les perceptrons multicouches à savoir, en classification et en approximation de fonctions, particulièrement. L'apprentissage le plus utilisé pour les RFR est le mode hybride (Talbi 2011).

D. Cartes auto-organisatrices de Kohonen (SOM)

Les cartes topologiques ou cartes auto organisatrices ont été introduites pour la première fois par Kohonen en 1982 (Kohonen 1982). Ce sont des réseaux à apprentissage non-supervisé qui établissent une carte discrète, ordonnée topo-logiquement, en fonction de patterns d'entrée. Le réseau forme ainsi une sorte de treillis dont chaque nœud est un neurone associé à un vecteur de poids. La correspondance entre chaque vecteur de poids est calculée pour chaque entrée. Par la suite, le vecteur de poids ayant la meilleure corrélation, ainsi que certains de ses voisins, vont être modifiés afin d'augmenter encore cette corrélation (Talbi 2011).

E. 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).

F. Réseaux ART

Les réseaux ART (en anglais Adaptive Resonance Theory) sont des réseaux à apprentissage par compétition (Carpenter et Grossberg1987 ; 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).

II.4.8. Conception d'un modèle neuronal

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

- ✓ 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.4.9. 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.4.10. 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 (Bishop 1995).

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 ($\sum(w_i x_i)$) est dans la partie asymptotique de la fonction d'activation (Dames 2005).

II.4.11. 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 (Urbani et al. 1994). 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.

A. 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 (Hornik 1991).

B. 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 (Dames 2005).

II.4.12. 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 2010):

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

Tableau II.2 résume les algorithmes d'apprentissage par rétro-propagation implanté dans la boite à outils des RN du programme MATLAB® (Demuth et Beale 2000).

Tableau II.2. Résumé des caractéristiques des algorithmes d'apprentissage de rétro-propagation implantés dans MATLAB® (Si Moussa 2010).

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 9 variables (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.

II.4.13. 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 (Dames 2005).

II.5. Conclusion

La littérature met en évidence que la modélisation peut permettre de valider les hypothèses simplificatrices émises lors de la formulation mathématique des phénomènes de transfert. Peu de travaux ont appliqué les relations quantitatives structure-activité (QSAR) dans la modélisation des processus membranaires, ils peuvent être considérés comme des outils importants pour la prédiction de la rétention des pesticides lors des membranes d'osmose inverse et de nanofiltration.

En guise de conclusion, cette revue bibliographique nous a servi à fixer l'objectif de ce travail qui consiste en la prédiction de la rétention des pesticides lors d'osmose inverse et de nanofiltration en utilisant des réseaux de neurones artificiels (RNA) et les relations quantitatives structure-propriétés (QSPR). Toutefois, nous avons fait en sorte de combler toutes les lacunes relevées dans les publications antérieures et ce en :

- ✓ choisissant une base la plus large possible ;
- ✓ utilisant une large base de descripteurs moléculaires, caractéristiques de membrane et les conditions opératoires et en déterminer les plus pertinents ;
- ✓ procédant à une optimisation pour le choix de nombre de neurones optimaux dans la couche cachée et pour le choix des fonctions de transfert.

Chapitre III

Modélisation de la rétention des anticancers





III.1 Introduction

Les données expérimentales de la rétention des anticancers lors de nanofiltration ont été utilisées pour appliquer les techniques de modélisation décrites dans le chapitre II. Ce chapitre sera divisé en trois axes : dans le premier axe, nous collecterons les données expérimentales. Ensuite, nous procéderons au prétraitement et l'analyse des données. Par la suite, nous élaborerons les techniques les plus récentes pour créer des modèles de réseaux de neurones artificiels (RNA) basé sur la relation quantitative structure-activité (QSPR dans le deuxième axe. Ensuite, dans le troisième axe, nous aborderons l'analyse de sensibilité à l'aide de la méthode du « poids ». Enfin, nous exposerons le diagnostic des points aberrants et le domaine d'application du modèle neuronal optimisé dans le quatrième axe.

III.2. Collecte des données, prétraitement et analyse

Dans cette étude, nous avons utilisé les données disponibles dans la littérature (Cristóvão et all.2019, 2022, 2024, Gouveia et all.2023, Teresa et all.2023, Verliefde et al. 2007, 2008 et 2009, Wang et all.2009, Yangali-Quintanilla et al. 2009 et 2010). La liste des 11 anticancers est répertoriée dans annexe A. La taille de la base de données et le nombre des anticancers sont présentés dans le tableau III.1.

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

Base de données (BD)	QSPR-RNA	Taille de base des données	Nombre des anticancers
BD1 avec SR(MgSO4)	QSPR-RNA1 QSPR-RNA2 QSPR-RNA3	604	13
BD2 avec SR(NaCl)	QSPR-RNA4 QSPR-RNA5 QSPR-RNA6	554	
BD3 avec SR	QSPR-RNA7 QSPR-RNA8 QSPR-RNA9	604	

Le choix des variables d'entrées et de sortie a été basé sur les interactions entre les propriétés des anticancers, les caractéristiques de la membrane et les conditions opératoires de filtration pour la

rétention des anticancers par les membranes de nanofiltration. Les entrées considérées dans cette étude sont les suivantes : les descripteurs moléculaires (Mw, d_e, LogD, le moment dipolaire, la longueur moléculaire "length", la surface minimale "Smin", la surface maximale "Smax", 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 sulfate de magnésium "SR($MgSO_4$", le rejet de sel de chlorure de sodium "SR(NaCl)", "le masse molaire de rejet de sel "Mw(SR)", la charge de surface de la membrane " potentiel zêta" et l'hydrophobicité de la membrane "angle de contact") et les conditions opératoires (pH, pression, température et le taux de récupération).

Les valeurs de d_e, LogD, moment dipolaire, surface minimale "Smin", surface maximale "Smax" ont été calculées par des logiciels qui sont mentionnés dans le tableau III.2. En annexe B, nous fournissons la capture d'écran d'un site d'internet ainsi que des deux logiciels utilisés.

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

	Logiciels
Log D	Chemspider http://www.chemspider.com/
Moment dipolaire	Hyperchem
Smin	
Smax	
Length	ChemBio3D Ultra (Version 12 ; 2012)

Les valeurs de la largeur équivalente moléculaire (EqWidth) ont été calculées par l'équation I.9 comme indiquer dans le premier chapitre. Le nombre total de variables (les descripteurs moléculaires des anticancers, les caractéristiques membranaires et les conditions d'opératoires) a été réduit lorsqu'il existait une relation entre elles.

Pour chaque base de données (BD1, BD2 et BD3), un modèle QSPR 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-RNA.

	BD01	**BD02**	**BD03**
Entrées	- de - Log D - Moment dipolaire - Length - EqWidth - MWCO - SR (MgSO4) - Zeta potentiel - Angle de contact - pH - Pression - Taux de récupération - Température	- de - Log D - Moment dipolaire - Length - EqWidth - MWCO - SR (NaCl) - Zeta potentiel - Angle de contact - pH - Pression - Taux de récupération - Température	- de - Log D - Moment dipolaire - Length - EqWidth - MWCO - Mw (SR) - SR - Zeta potentiel - Angle de contact - pH - Pression - Taux de récupération - Température
Sortie	Rétention	Rétention	Rétention

L'analyse statistique des données d'entrées et de sortie, comprenant les valeurs minimales "min", les moyennes arithmétiques (Équation III.2) "$\overline{x}$", les valeurs maximales "max" et les écarts-types "DS" (Équation III.3) (Chabane et Koul 2022), est présentée dans les tableaux III.4, III.5 et III.6.

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

$$\mathrm{DS} = \left(\frac{1}{n}\sum_{i=1}^{n}(xi - \overline{x})^{2}\right)^{1/2} \qquad (III.2)$$

Tableau III.4. Analyse statistique des données d'entrée et de sortie (BD1).

	Min	Max	Moyenne	DS
de (g/mol)	0,0829	0,9288	0,1132	0,1472
LogD	-0,9600	5,5100	1,4505	1,6956
Moment dipolaire (Debye)	0,0000	5,8320	1,6942	1,8725
Length (nm)	1,5230	1,7840	1,5949	0,1053
EqWidth	1,0232	1,4915	1,1576	0,1528
MWCO (Dalton)	30,0000	300 ,0000	230,3891	72,8817
SR($MgSO_4$)	0,9070	0,9900	0,9731	0,0221
Zeta potentiel (mv)	-41,3000	1,0300	-20,4043	7,3122
Angle de contact (°)	35,0000	48,0000	40,7669	1,4228
pH	2,70000	10,0100	7,0401	0,5836
Pression (KPa)	500,0000	2500,0000	796,1258	411,3162
Taux de récupération (%)	1,0000	84,0000	49,0377	29,3961
Température (C°)	0,5000	90,0000	21,3386	7,4722
Rétention (%)	20,0000	100,0000	81,8186	22,5392

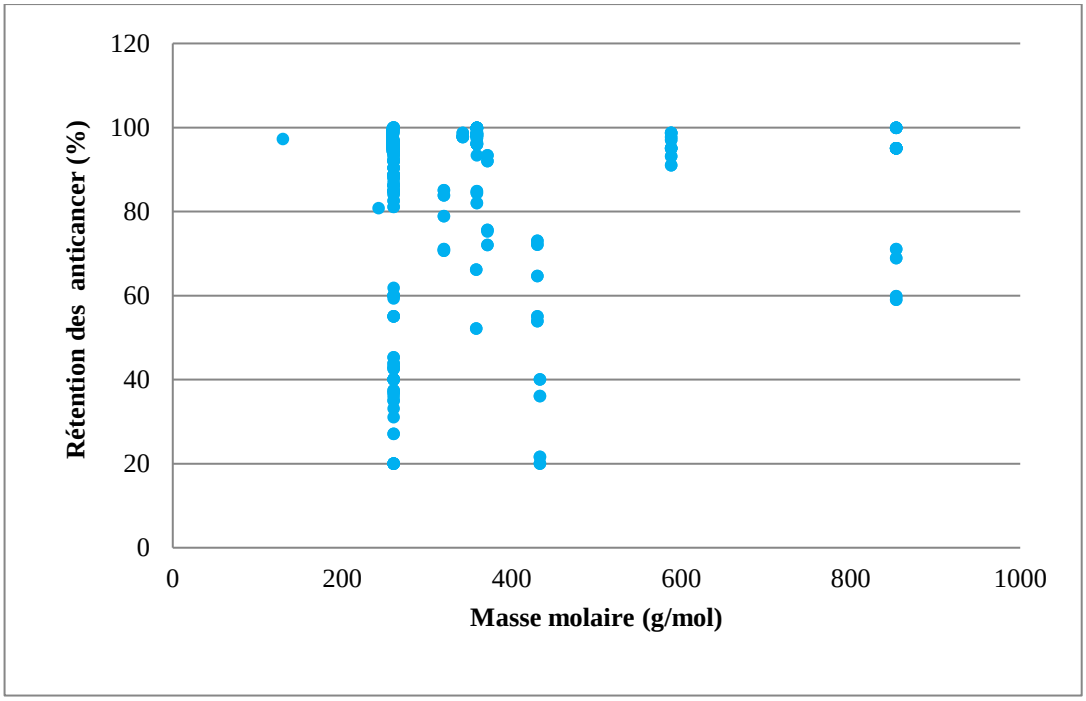
Tableau III.5. Analyse statistique des données d'entrée et de sortie (BD2).

	Min	Max	Moyenne	DS
de (g/mol)	0,0829	0,9288	0,1299	0,1831
LogD	-0,9600	5,5100	1,3439	1,7368
Moment dipolaire (Debye)	0,0000	5,8320	1,6706	1,8699
Length (nm)	1,5230	1,7840	1,5985	0,1099
EqWidth (nm)	1,0232	1,4915	1,1589	0,1469
MWCO (Dalton)	150,0000	300,0000	231,4982	74,7855
SR(NaCl)	0,2053	0,9900	0,9064	0,2423
Zeta potentiel (mv)	-41,3000	-18,5000	-20,9282	7,0396
Angle de contact (°)	40,6000	42,0000	40,7491	0,4223
pH	4,4000	8,3000	6,9839	0,5104
Pression (KPa)	520,0000	2020,0000	733,5740	239,3010
Taux de récupération (%)	2,8000	84,0000	50,6680	29,5646
Température (C°)	20,000	27,1000	21,0487	2,0840
Rétention (%)	20,0000	99,9000	82,1210	22,6411

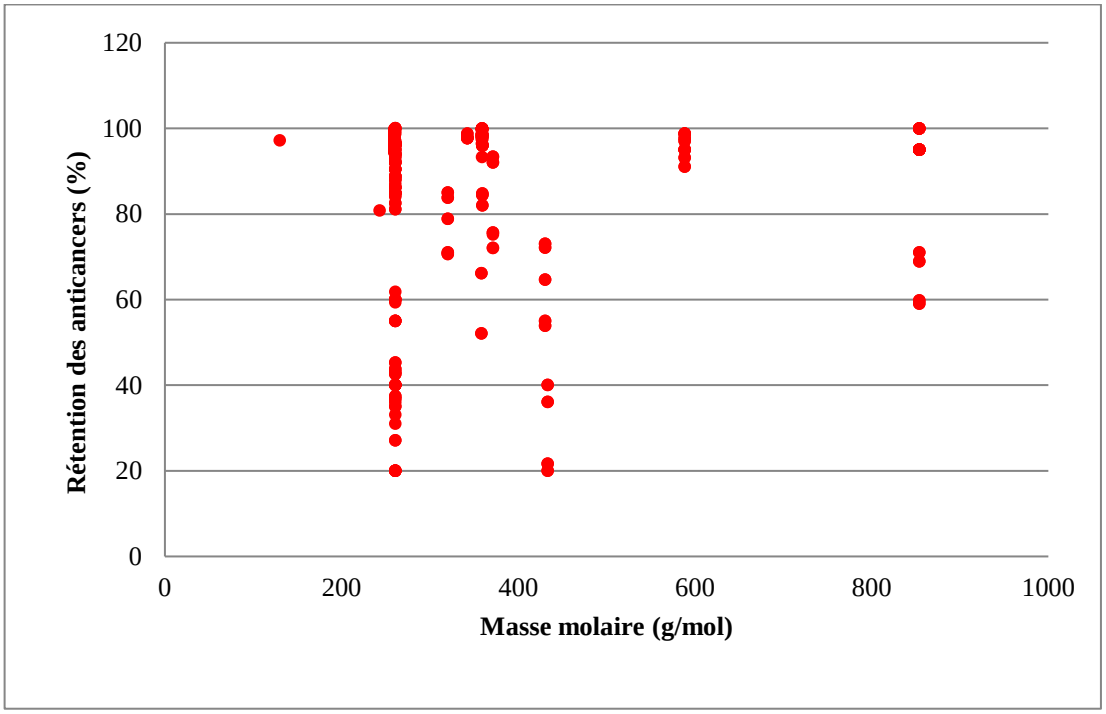
Tableau III.6. Analyse statistique des données d'entrée et de sortie (BD3).

	Min	**Max**	**Moyenne**	**DS**
de (g/mol)	0,0829	0,9288	0,1155	0,1530
LogD	-0,9600	5,5100	1,2782	1,6238
Moment dipolaire (Debye)	0,0000	5,8320	1,5462	1,7592
Length (nm)	1,5230	1,7840	1,5999	0,1074
EqWidth (nm)	1,0232	1,4915	1,1402	0,1397
MWCO (Dalton)	30,0000	300,000	229,5199	73,0048
Mw(SR) (g/mol)	58,4400	120,3660	63,2587	16,6025
SR($MgSO_4$)	0,2053	0,9900	0,9088	0,2381
Zeta potentiel (mv)	-41,3000	1,0300	-20,4817	7,5902
Angle de contact (°)	35,0000	48,0000	40,8129	1,2156
pH	2,7000	10,0100	6,9746	0,7249
Pression (KPa)	500,0000	2500,0000	801,755	426,633
Taux de récupération (%)	1,0000	84,0000	49,2556	29,6902
Température (C°)	0,5000	90,0000	21,2452	7,1538
Rétention (%)	20,0000	100,0000	82,7634	22,1930

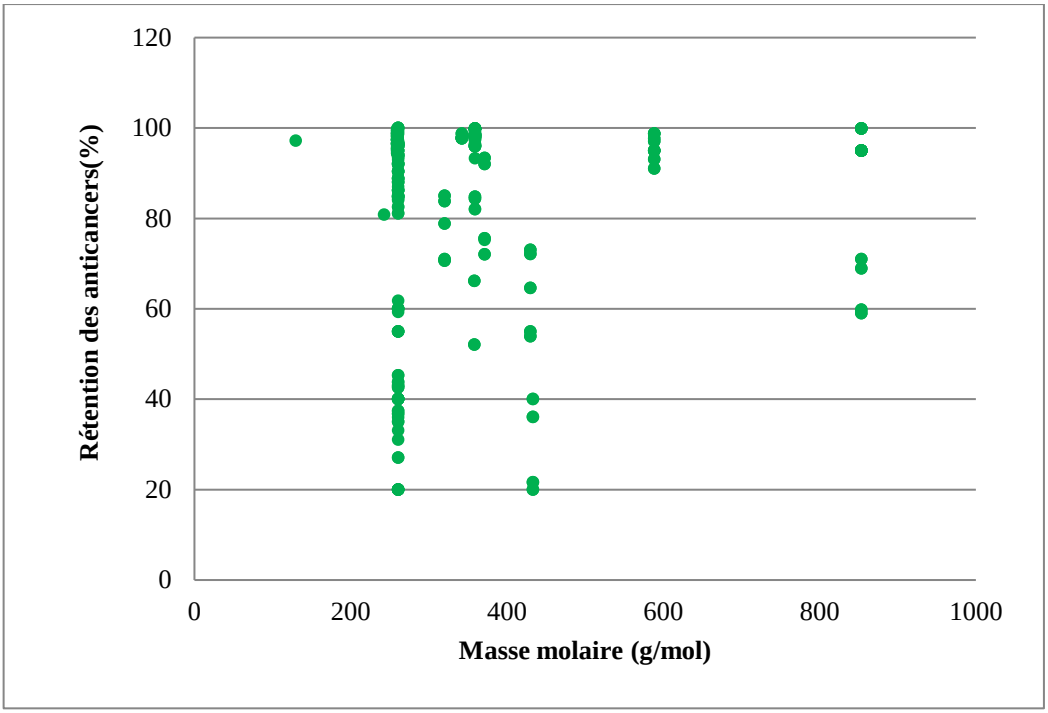
La figure III.1 représente les rétentions expérimentales en fonction de masse molaire "Mw" pour les trois bases de données (les rétentions expérimentales en fonction de la masse molaire "Mw" sont indiquées par des cercles bleu pour BD1 "604 points", des cercles rouges pour BD2 "554" et des cercles verts pour BD3 "604"). Lorsque la masse molaire des agents anticancéreux varie de 300 g/mol à 450 g/mol, les rétentions expérimentales montrent que plus de 60 % d'entre eux s'agglomèrent ensemble.


120
100
80
60
40
20
0
Rétention des anticancer (%)
0
200
400
600
800
1000
Masse molaire (g/mol)

(a)


120
100
80
60
40
20
0
Rétention des anticancers (%)
0
200
400
600
800
1000
Masse molaire (g/mol)

(b)



(c)

Figure III.1. Rétentions expérimentales en fonction de masse molaires des anticancers : (a) BD1, (b) BD2, (c) BD3.

III.3. Développement des modèles QSPR-RNA

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

Les réseaux de neurones artificiels (RNA) sont particulièrement bien adaptés aux modèles QSAR/QSPR en raison de leur capacité à extraire des informations non-linéaires présentes dans la matrice de données (Chabane et Koul 2022). Pour établir un modèle non linéaire entre les descripteurs moléculaires des anticancers, les caractéristiques de la membrane et les conditions opératoires de filtration pour la rétention des anticancers par les membranes de nanofiltration, l'utilisation de la méthode RNA est nécessaire et s'est avérée être une technique de modélisation importante. Effectivement, elle est largement utilisée pour les études de QSPR non linéaires et constitue 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. Cette procédure est basée sur l'organigramme ci-dessous.

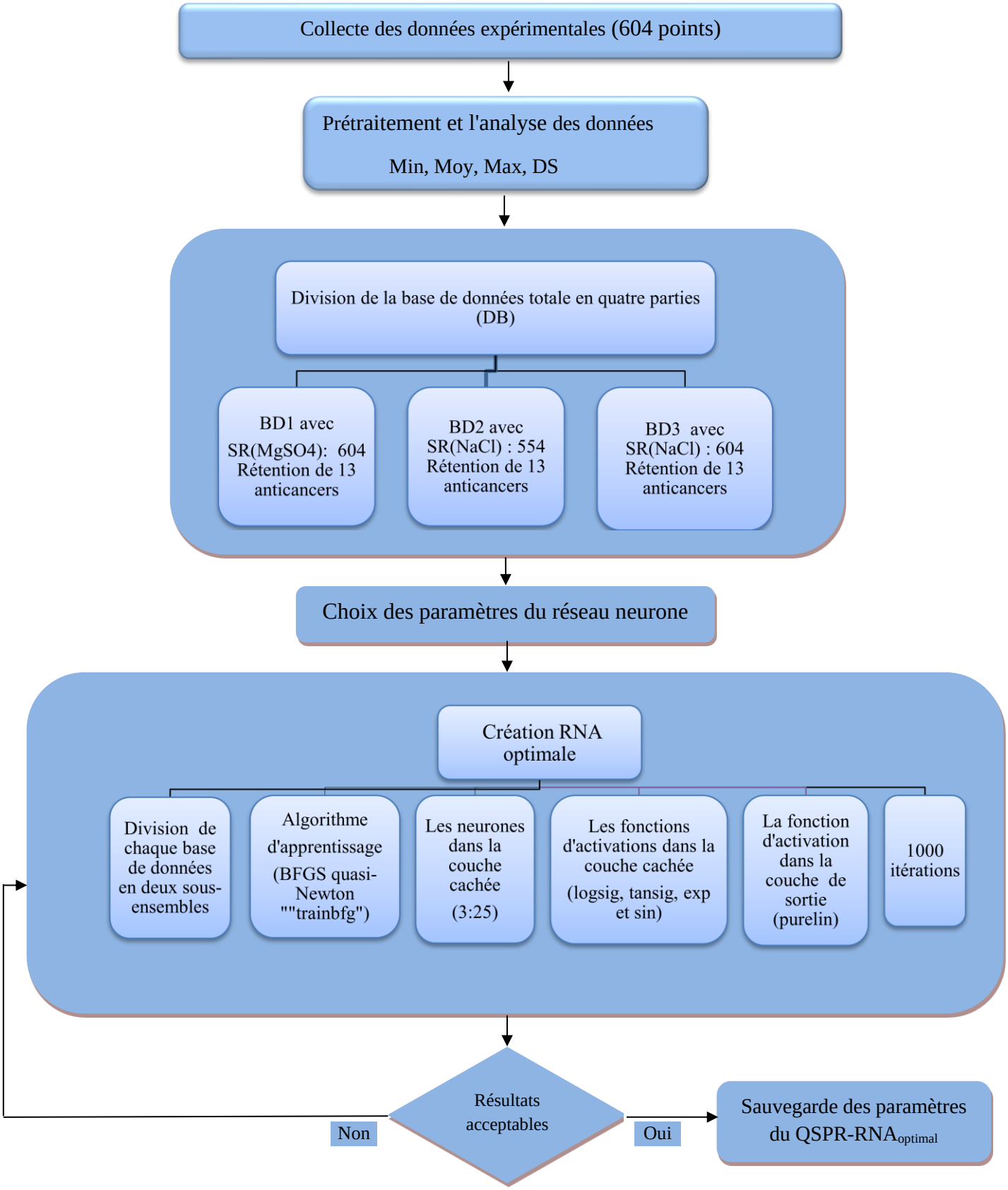


Figure III.2. Organigramme de développement et l'optimisation de l'architecture d'un QSPR-$RNA_{optimal}$.

L'algorithme d'apprentissage utilisé dans ce travail est le BFGS quasi-Newton (trainbfg). Chaque QSPR-RNA 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 QSPR-RNA décrite ci-dessus a été réalisée en utilisant STATISTICA (Annexe C). Ce logiciel est utilisé pour optimiser l'architecture des quatre modèles neuronaux QSPR-RNA employés pour la rétention des anticancers par les membranes de nanofiltration.

III.4 Performance des modèles neuronaux

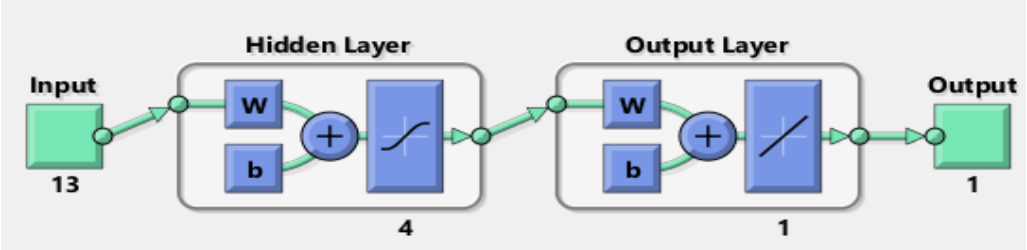
Les résultats de performances de neuf modèles neuronaux "QSPR-RNA" en termes du coefficient de corrélation (R) et l'erreur quadratique moyenne (EQM) sont mentionnés dans le tableau suivant pour chaque division de base de données (division 01 "80% pour la base d'apprentissage et 20% pour la base de test", division 02 "70% pour la base d'apprentissage et 30% pour la base de test" et division 03 "60% pour la base d'apprentissage et 40% pour la base de test"). D'après ce tableau, on constate que le quatrième modèle (QSPR-RNA04) est un modèle optimal (QSPR-RNA$_{optimal}$), il donne des meilleurs résultats par apport des autres modèles (R = 0.9319 et EQM = 34.01035% pour la phase de test).

Tableau III.7. Comparaison entres les modèles neuronaux (QSPR-RNA).

		Pourcentage	**R**	**EQM (%)**
QSPR-RNA01	**Phase d'apprentissage : 484 points**	80%	0.9225	39.0882
	Phase de test : 120 points	20%	0.8820	49.9301
	Phase totale : 604 points	-	0.9152	-
QSPR-RNA02	**Phase d'apprentissage : 423 points**	70%	0.9208	38.1580
	Phase de test : 181 points	30%	0.9018	48.6467
	Phase totale : 604 points	-	0.9149	-
QSPR-RNA03	**Phase d'apprentissage : 363 points**	60%	0.9191	39.7340
	Phase de test : 241 points	40%	0.9041	45.9321
	Phase totale : 604 points	-	0.9131	-
QSPR-RNA04	**Phase d'apprentissage : 444 points**	80%	0.9010	48.85105
	Phase de test : 110 points	20%	0.9319	34.01035
	Phase totale : 554 points	-	0.9097	-
QSPR-RNA05	**Phase d'apprentissage : 388 points**	70%	0.9098	45.8237
	Phase de test : 166 points	30%	0.9287	31.3565
	Phase totale : 554 points	-	0.9152	-
QSPR-RNA06	**Phase d'apprentissage : 333 points**	60%	0.9115	45.9485
	Phase de test : 221 points	40%	0.9139	38.3930
	Phase totale : 554 points	-	0.9124	-
QSPR-RNA07	**Phase d'apprentissage : 484 points**	80%	0.9247	36.3144
	Phase de test : 120 points	20%	0.8223	84.3819
	Phase totale : 604 points	-	0.9029	-
QSPR-RNA08	**Phase d'apprentissage : 423 points**	70%	0.9229	37.1411
	Phase de test : 181 points	30%	0.8376	76.2167
	Phase totale : 604 points	-	0.8962	-
QSPR-RNA09	**Phase d'apprentissage : 363 points**	60%	0.9275	35.6315
	Phase de test : 241 points	40%	0.8476	73.1534
	Phase totale : 604 points	-	0.8944	-

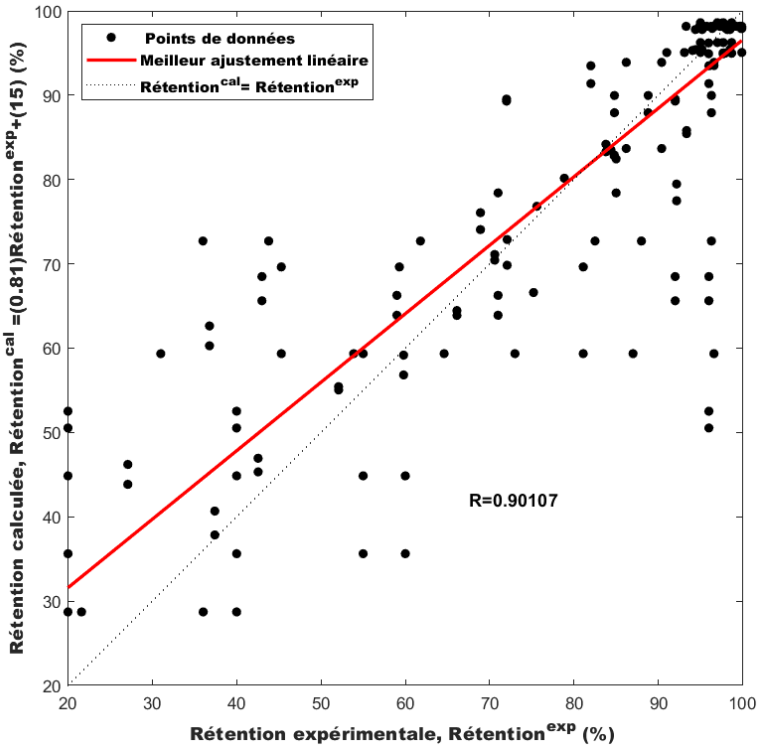
Le tableau III.6 présente la structure de modèle neuronal QSPR-RNA$_{optimal}$ obtenu. Ce modèle comporte quatre neurones dans la couche cachée avec une fonction d'activation tangente hyperbolique.

Table III.8. Structure de modèle QSPR-RNA$_{optimal}$.

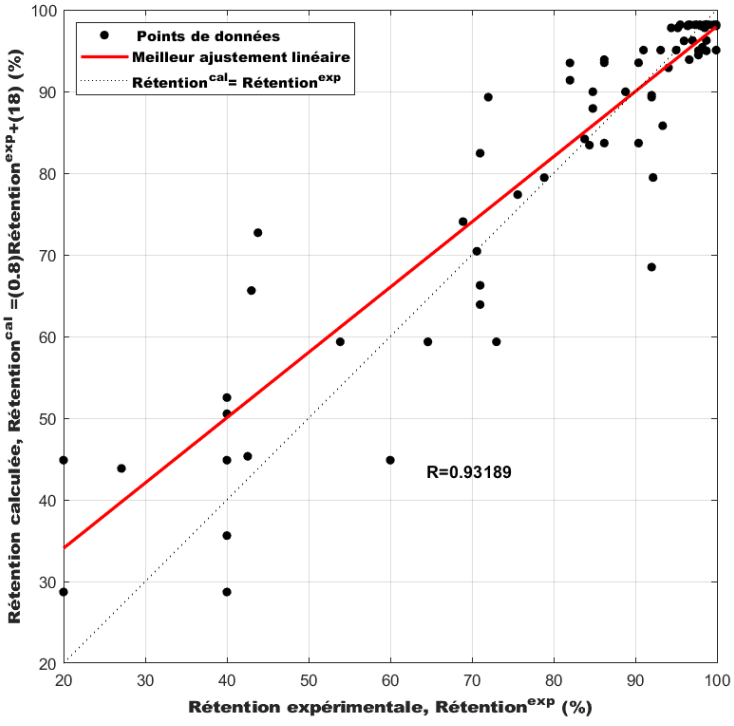
Algorithme d'apprentissage	**Couche entrée**	**Couche cachée**		**Couche sortie**	
	Nombre de neurone	**Nombre de neurone**	**Fonction d'activation**	**Nombre de neurone**	**Fonction d'activation**
BFGS quasi-Newton **(trainbfg)**	13	**4**	**tansig**	1	Purelin
					

Selon la discussion précédente, le modèle neuronal optimal (QSPR-RNA$_{optimal}$) a été développé dans le but de prédire la rétention des anticancers par nanofiltration. L'intrigue et les paramètres de la régression linéaire ont obtenus en utilisant la fonction ***"postreg"*** *de MATLAB*. Un coefficient de corrélation "R" supérieure à 0.5 est généralement considérée comme satisfaisant tandis qu'une valeur supérieure à 0.9 est excellente (Benrkia et Boulenoir 2022).

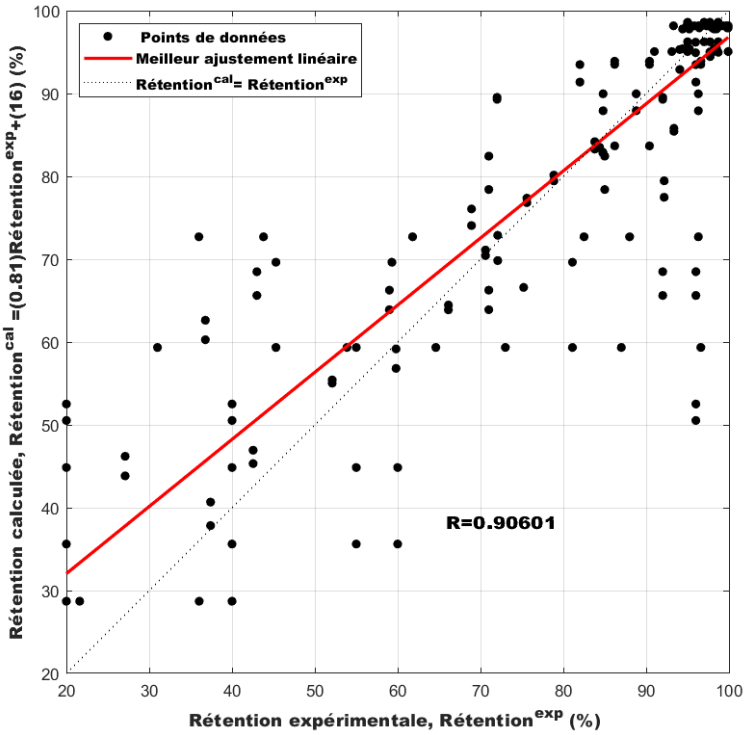
La figure III.3 montre les courbes de régression linéaire des taux de rétentions calculés par le modèle neuronal "QSPR-RNA$_{optimal}$" optimisé avec les taux de rétentions expérimentales avec des vecteurs de régression approchant de l'idéal ([α, β, R] = [0.8211, 14.5328, 0.9052] pour phase d'apprentissage, [α, β, R] = [0.8233, 15.7648, 0.9314] pour la phase de test et [α, β, R] = [0.8218, 14.7569, 0.9097] pour la phase totale). Les coefficients des corrélations sont considérés comme excellentes ($0.90 \leq R \leq 1.00$) pour la phase d'apprentissage, la phase de test et la phase totale, ce qui démontre la robustesse du modèle neuronal "QSAR-RNA$_{optimal}$" établi ainsi que sa capacité à prédire des différents paramètres caractérisant la rétention des anticancers par nanofiltration.


Points de données
Meilleur ajustement linéaire
Rétention^cal= Rétention^exp
R=0.90107
Rétention calculée, Rétention^cal =(0.81)Rétention^exp+(15) (%)
Rétention expérimentale, Rétention^exp (%)

(a)


Points de données
Meilleur ajustement linéaire
Rétention^cal= Rétention^exp
R=0.93189
Rétention calculée, Rétention^cal =(0.8)Rétention^exp+(18) (%)
Rétention expérimentale, Rétention^exp (%)

(b)



(c)

Figure III.3. Comparaison entre les valeurs expérimentales et les valeurs calculées pour QSPR-RNA$_{optimal}$: (a) la phase d'apprentissage, (b) la phase de test, (c) la phase totale.

La performance de modèle neuronal "QSPR-RNA$_{optimal}$" optimisé a été évaluée en termes de coefficient de corrélation (R) et la racine de l'erreur quadratique moyenne (REQM), l'erreur absolue moyenne (EAM) et l'erreur de standard de prédiction (ESP). REQM, EAM et ESP ont été utilisées pour déterminer l'erreur de modélisation entre les valeurs expérimentales et calculées. Les erreurs sont définies comme suit (Chabane et Koul 2022).

$$EAM = \frac{1}{n}\sum_{i=1}^{n}\left|(y_{i,exp} - y_{i,cal})\right| \qquad \text{(III. 3)}$$

$$\mathrm{REQM} = \sqrt{\frac{\sum_{i=1}^{n}(Y_{i,exp} - Y_{i,cal})^2}{n}} \qquad \text{(III. 4)}$$

$$ESP(\%) = \frac{REQM}{Y_e} * 100 \qquad \text{(III. 5)}$$

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

La figure III.4 représente la performance de modèle neuronal optimal. Elle montre que le modèle $RNA_{optimal}$ a donné des erreurs plus faibles. Ce qui indique que le modèle $RNA_{optimal}$ possède une capacité de modélisation très élevée.

Les vecteurs de régression linéaire pour le modèle neuronal "QSPR-$RNA_{optimal}$" optimisé et les erreurs (REQM, EAM, ESP) sont présentés dans le tableau III.9 et la figure III.3. 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 : $ycal=\alpha yexp+\beta$, $\boldsymbol{\alpha}$**=1** (pente), $\boldsymbol{\beta}$**=0** (ordonnée à l'origine), ***R*=1** (coefficient de corrélation)] dans l'ajustement des profils des QSPR-$RNA_{optimal}$. Le modèle $RNA_{optimal}$ a donné des erreurs plus faibles. Ce qui indique que le modèle $RNA_{optimal}$ possède une capacité de modélisation très élevée.

Table III.9. Vecteurs de régression linéaire issue de l'équation linéaire : [équation linéaire : $ycal=\alpha yexp+\beta$, $\boldsymbol{\alpha}$(pente), $\boldsymbol{\beta}$(ordonnée à l'origine), ***R***(coefficient de corrélation)] et l'erreur quadratique moyenne ***(REQM)***, l'erreur absolue moyenne ***(EAM)*** et l'erreur de standard de prédiction ***(ESP)***.

	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	***EAM (%)***	**REQM *(%)***	***ESP (%)***	**R**
Phase d'apprentissage	0.8211	14.5328	5.4732	9.6866	11.8137	0.9052
Phase de test	0.8233	15.7648	5.1382	8.1247	9.8324	0.9314
Phase totale	0.8218	14.7569	5.4067	9.3972	11.4431	0.9097

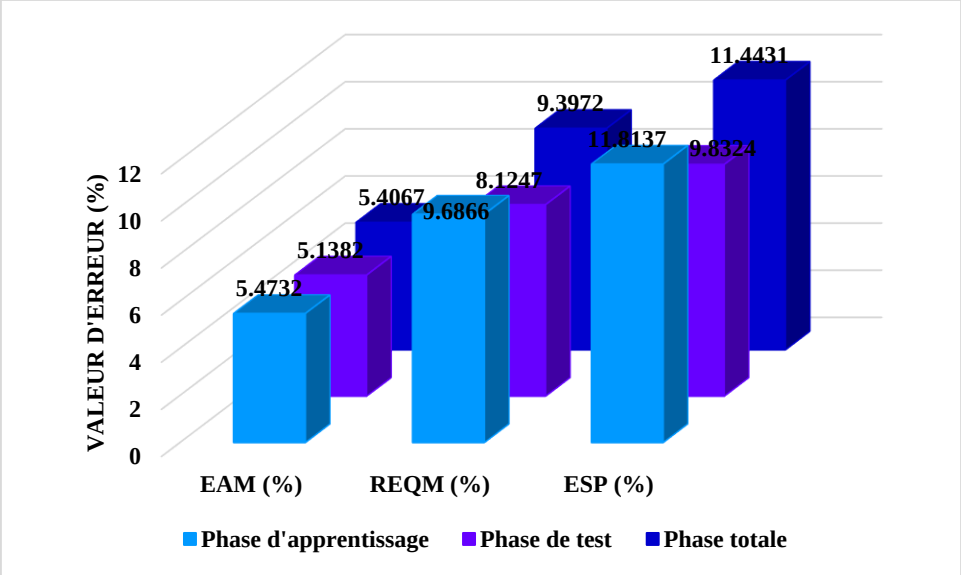


Figure III.4. Performances de modèle neuronal $RNA_{optimal}$ pour la phase d'apprentissage, la phase de test, la phase totale.

III.5 Expressions mathématiques de modèle neuronal optimal

À partir du QSPR-RNA$_{optimal}$ représenté sur la figure III.5, on peut exprimer l'assimilation de la rétention des anticancers lors de nanofiltration par un modèle mathématique y incorporant toutes les entrées *Ei*(de, 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 :

$$Z_j = f_h\left[\sum_{i=1}^{13} w_{ji}^I E_i + b_j^h\right] = \frac{exp\left(\sum_{i=1}^{13} w_{ji}^I E_i + b_j^h\right) - exp\left(-\sum_{i=1}^{13} w_{ji}^I E_i + b_j^h\right)}{exp\left(\sum_{i=1}^{13} w_{ji}^I E_i + b_j^h\right) + exp\left(-\sum_{i=1}^{13} w_{ji}^I E_i + b_j^h\right)} \qquad (III.6)$$

i=1,2,……13

j=1, 2,…. 4

La sortie "Rétention des anticancers" :

$$QSPR - RNA_{optimal} = f_0\left[\sum_{j=1}^{4} w_{1j}^h Z_j + b_1^o\right] = \sum_{j=1}^{4} w_{1j}^h Z_j + b_1^o \qquad (III.7)$$

La combinaison des équations (III.6) et (III.7) mène à la formule mathématique suivante qui décrit l'assimilation de la rétention des anticancers lors de nanofiltration en prenant en compte toutes les entrées *Ei*(de, LogD, moment dipolaire, Length, EqWidth, MWCO, SR(NaCl), angle de contact, potentiel zêta, pH, pression, taux de récupération et température) :

$$QSPR - RNA_{optimal} = \sum_{j=1}^{4} w_{1j}^h \frac{exp\left(\sum_{i=1}^{13} w_{ji}^I E_i + b_j^h\right) - exp\left(-\sum_{i=1}^{13} w_{ji}^I E_i + b_j^h\right)}{exp\left(\sum_{i=1}^{13} w_{ji}^I E_i + b_j^h\right) + exp\left(-\sum_{i=1}^{13} w_{ji}^I E_i + b_j^h\right)} + b_1^o \qquad (III.8)$$

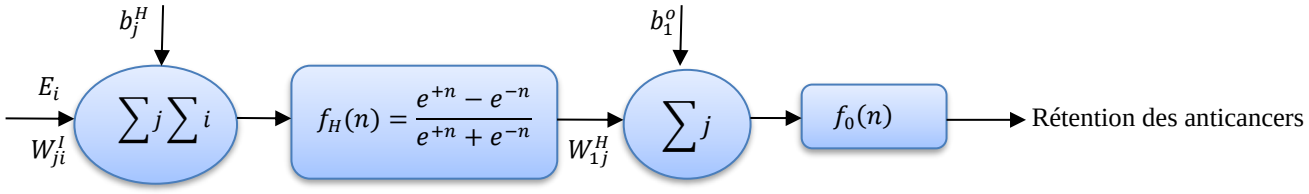


Figure III.5. Représentation schématique du QSPR-RNA$_{optimal}$.

Où :

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 (4 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 (4 lignes x 1colonne).

b_j^h: représente les biais du neurone *j* de la couche cachée (4 lignes).

b_1^o : représente les biais du neurone de la couche de sortie.

Les matrices des poids et biais, sont donnés dans le tableau III.10 :

Tableau III.10. Poids et biais du $RNA_{optimal.}$

Connexions Couche d'entrée-Couche cachée														**Connexions Couche cachée-Couche de sortie**	
Poids													**Biais**	**Poids**	**Biais**
$w_{j,1}^I$	$w_{j,2}^I$	$w_{j,3}^I$	$w_{j,4}^I$	$w_{j,5}^I$	$w_{j,6}^I$	$w_{j,7}^I$	$w_{j,8}^I$	$w_{j,9}^I$	$w_{j,10}^I$	$w_{j,11}^I$	$w_{j,12}^I$	$w_{j,13}^I$	b_j^H	$w_{1,j}^H$	b_1^0
0.077	13.813	14.553	-1.750	-18.055	-14.431	-2.870	41.873	27.792	142.581	-7.288	0.024	-21.342	9.428	0.415	0.544
6.085	-0.072	-3.393	39.720	24.058	10.127	27.790	-40.315	27.817	-96.043	-45.687	2.624	50.621	-4.309	-0.243	
-20.281	-5.118	40.374	75.564	0.185	-20.634	29.675	37.824	6.553	11.223	37.514	2.609	49.132	34.353	-0.091	
7.334	-5.151	91.254	-6.832	-14.442	17.328	31.448	0.308	-174.496	-1.181	10.646	5.350	-1.426	7.921	-0.314	

L'équation mathématique, dérivant le modèle neuronal "QSPR-RNA$_{optimal}$" optimisé, contient juste le degré exigé de complexité et incluent dans leurs formules les caractéristiques pertinentes de la modélisation de la rétention des anticancers lors de nanofiltration. Ils peuvent donc facilement être appliqués dans des boucles de contrôle du processus de la rétention des anticancers. L'architecture des réseaux de neurones optimaux obtenus est donnée sur la figure III.6, ci-après

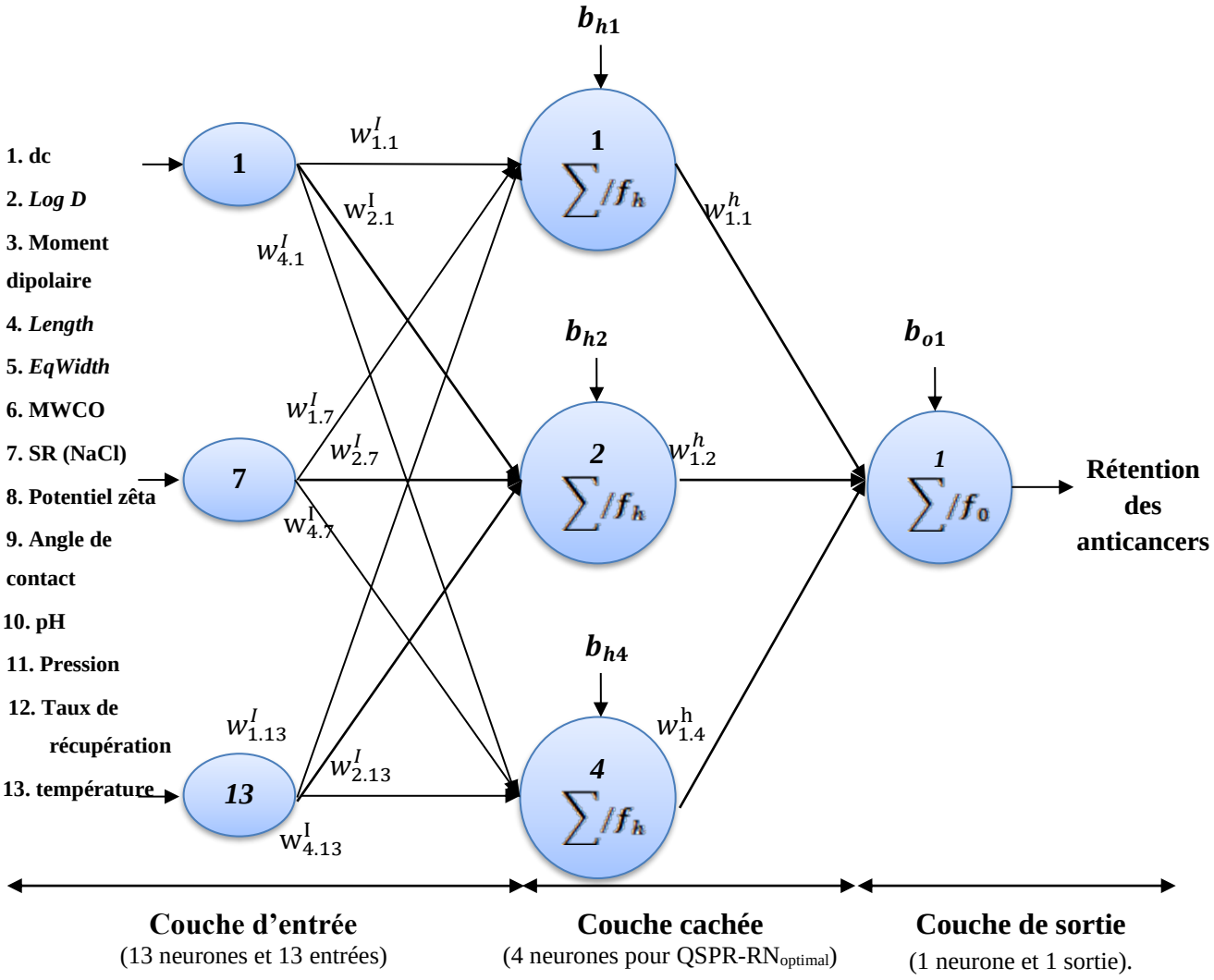


Figure III.6. Architecture finale du réseau de neurone optimal de type feedforward.

III.6. Analyse de sensibilité par la méthode de "poids"

L'influence des variables d'entrée sur la sortie a été évaluée à l'aide d'une analyse de sensibilité utilisant la méthode des "poids".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 RNA (Chabane et Koul2022). 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 mentionnées dans les équations suivant (Chabane et Koul 2022) :

$$P_{ji=}\left|w_{ji}^{I}w_{1j}^{H}\right| \quad \text{(III. 9)}$$

$$Q_{ji} = \frac{P_{ji}}{\sum_{i=1}^{nj} P_{ji}} \quad \text{(III. 10)}$$

$$S_i = \sum_{j=1}^{nj} Q_{ji} \quad \text{(III. 11)}$$

$$IR = 100\frac{S_i}{\sum_{i=1}^{nj} S_i} \quad \text{(III.12)}$$

Pour évaluer la contribution ainsi que le profil de variation de chaque variable d'entrée (de, LogD, Dipole moment, Lenght, EqWidth, MWCO, SR(NaCl), angle de contact, potentiels zêta, pH, pression, taux de récupération et température) sur la sortie de modèle neuronal "QSPR-$RNA_{optimal}$" optimisé (Rétention), nous avons effectué une analyse de sensibilité en utilisant méthode de "poids". Nous avons appliqué cette méthode à modèle neuronal "QSPR-$RNA_{optimal}$" optimisé en utilisant leurs poids respectifs illustrés dans les tableaux III.10. Les résultats de contributions sont montrés sur la figure III.7. Les variables les plus pertinentes qui peuvent influencer la rétention d'anticancers par les membranes de nanofiltration sont : pH, angle contact, moment dipolaire, potentiel zêta, Length, taux de récupération, pression, SR (NaCl). Toutes les entrées exercent des contributions significatives (IR > 5%) sur la rétention des anticancéreux par nanofiltration, tandis que certaines entrées, telles que MWCO, EqWidth, logD, de, et la température, ont des contributions plus faibles (IR < 5%).

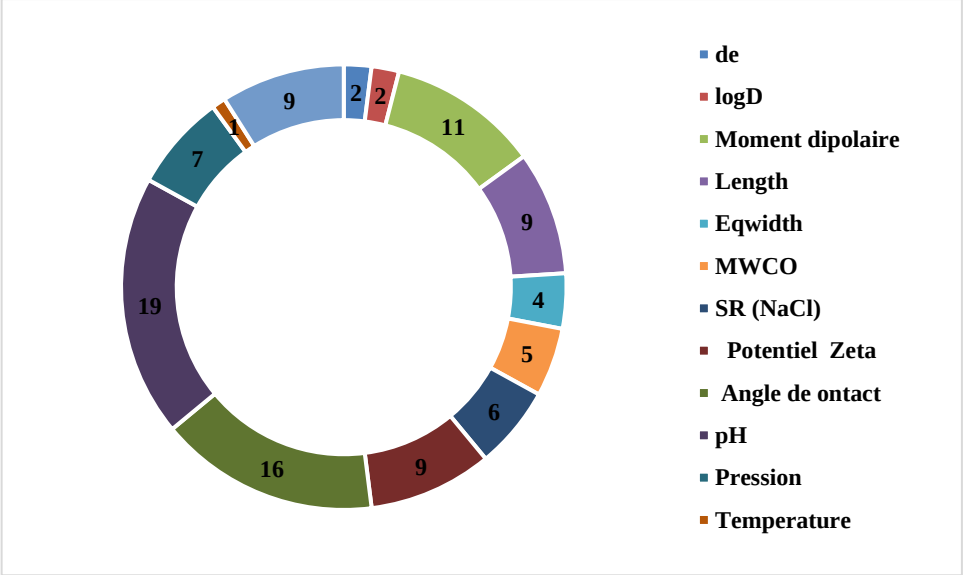


Figure III.7. Présentation des pourcentages de l'importance relative (*IR*) de chaque variable d'entrée sur la rétention des anticancers par nanofiltration.

III.7. 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 (Chabane et Koul 2022).

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 (Chabane et Koul 2022).

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 :

$$H = X(X^tX)^{-1}X^t \qquad (III.13)$$

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.

$$Hat = diagonale(H) \qquad (III.14)$$

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 :

$$H^* = \frac{3(n+1)}{m} \qquad (III.15)$$

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

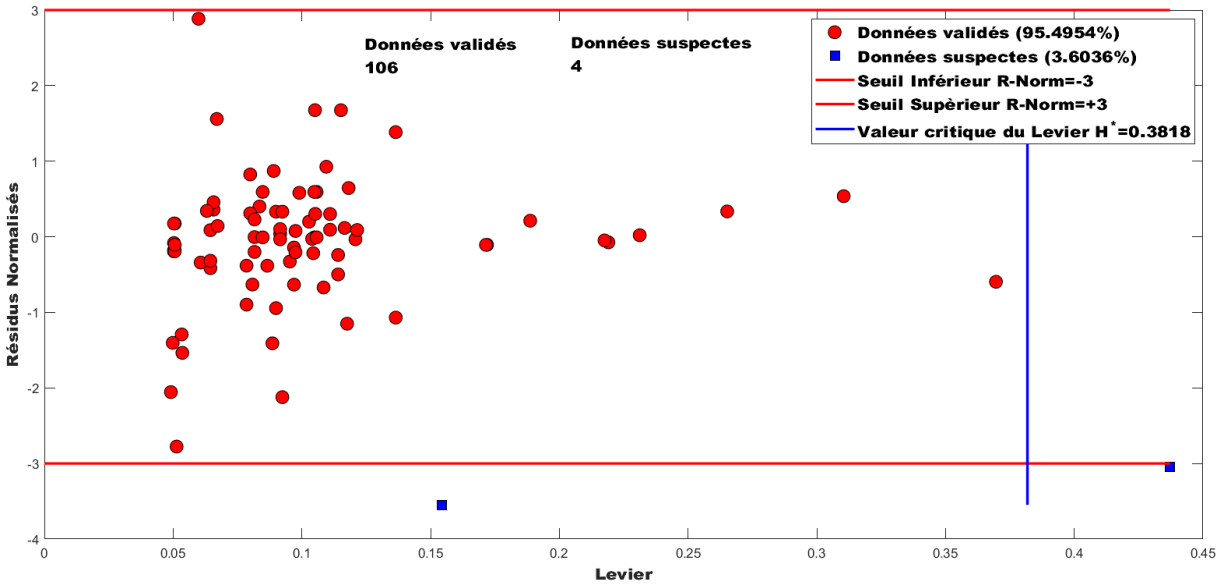
Les résidus normalisés sont calculés à partir des données expérimentales et celle calculée par le modèle.

$$(R_Norm)_i = \frac{\left(y_i^{exp} - y_i^{cal}\right)}{\sqrt{var(y^{exp} - y^{cal})}} \quad i = 1, \dots m \qquad (III.16)$$

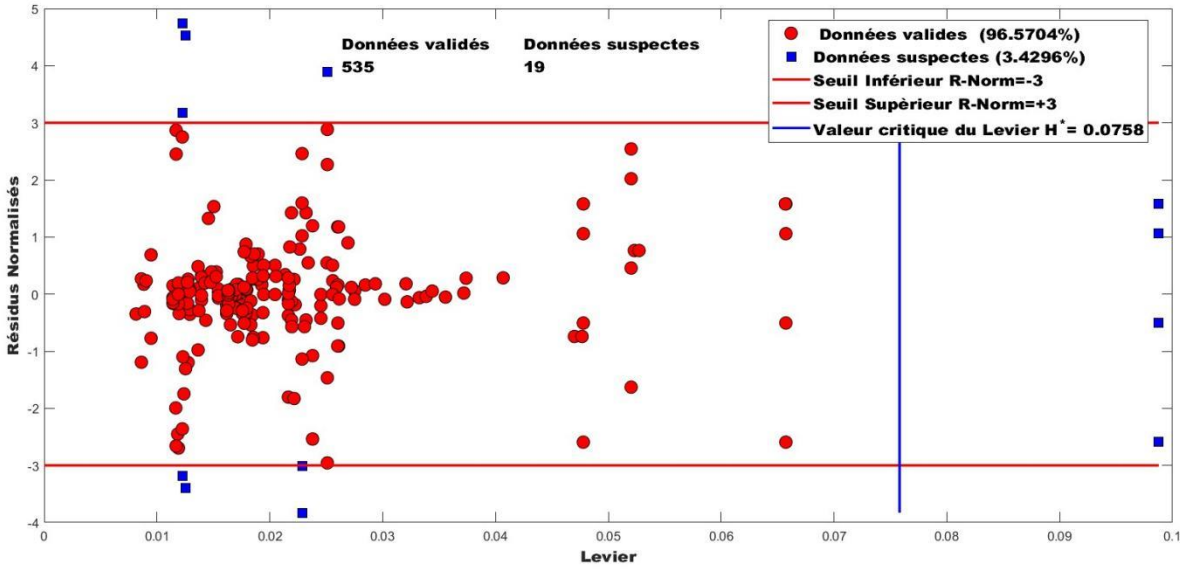
Le résidu normalisé égal à trois est normalement considéré comme une valeur "seuil" pour accepter les points de données dans les écarts types compris sur une plage de ± 3 par rapport à la moyenne (pour couvrir 99% des données normalement distribuées). Si la majorité des points de données se situent dans 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".

La figure III.8 montre le diagramme de William pour le domaine d'applicabilité pour le modèle neuronal optimal "QSPR-$RNA_{optimal}$". Dans ce diagramme l'existence de **106/110 (96.3636%)** de points de données valides et **4/110 (3.6363 %)** de points de données suspectes pour la phase test et **535/554 (96.5704%)** de points de données valides et **19/554(3.4296%)** de points de données suspectes pour la phase totale. La valeur critique de levier est $H^*=3(n+1)/m$=**3(13+1)/110=0.3818** pour la phase de test et $H^*=3(n+1)/m$=**3(13+1)/554=0.0758** pour la phase totale. Cela indique que le

développement du modèle QSPR-RNA$_{optimal}$ 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.



(a)



(b)

Figure III.8. Diagnostic des données potentiellement suspectes et domaine d'applicabilité du modèle neuronal optimal : (a) pour la phase test et (b) pour la phase totale.

III.8. Conclusion

Ce présent chapitre a donné une raison de plus pour l'utilisation généralisée des modèles neuronaux (QSPR-RNA) pour la modélisation de la rétention des anticancers lors de nanofiltration. Le but de la présente étude était de développer un modèle RNA de type feed-forward.

Le modèle QSAR-$RNA_{optimal}$ développé dans ce travail a donné des erreurs faibles. Les erreurs absolues moyennes étaient 14.7569% pour phase totale, 5.4732% pour la phase apprentissage et 5.1382% pour phase test. Les racines des erreurs quadratiques moyennes étaient 9.3972% pour phase totale, 9.6866% pour phase apprentissage et 8.1247% pour la phase test. Les erreurs standards de prédictions étaient 11.4431% pour phase totale, 11.8137% pour phase apprentissage et 9.8324% pour phase test. Ce qui montre une bonne robustesse du modèle neuronal "QSPR-$RNA_{optimal}$" établi et la possibilité de la prédiction des différents paramètres qui caractérisent la rétention des anticancers lors de la nano filtration.

L'analyse de sensibilité par la méthode de "poids" a révélé que les variables les plus pertinentes qui peuvent influencer la rétention d'anticancers par les membranes de nanofiltration sont : pH, angle contact, moment dipolaire, potentiel zêta, Length, taux de récupération, pression, SR (NaCl). Toutes les entrées exercent des contributions significatives (IR > 5%) sur la rétention des anticancéreux par nanofiltration, tandis que certaines entrées, telles que MWCO, EqWidth, logD, de, et la température, ont des contributions plus faibles (IR < 5%).

L'analyse diagnostique des valeurs aberrantes et de la couverture du modèle neuronal optimisé montre que le développement du modèle neuronal optimal QSPR-$RNA_{optimal}$ et ses prédictions s'effectuent dans le cadre de la couverture. Il en résulte un modèle neuronal optimal statistiquement valide. Par conséquent, nous pouvons dire qu'il existe un « bon effet de levier élevé » dans la phase de test et totale.

Conclusion Générale et Perspectives

Les réseaux de neurones possèdent une propriété notable qui ont prouvé leur contribution à la modélisation de processus non linéaires. Les réseaux de neurones sont des approximateurs généraux parcimonieux. Enfin, un modèle des réseaux neurones artificiels basés sur la relation quantitative structure propriété (QSPR-RNA) a été appliqué pour prédire la séparation des anticancers par des membranes de nanofiltration. D'autres chercheurs ont abordé ce problème de manière simplifiée, en réduisant notamment termes la taille de la base de données et les paramètres calculés ont donné des résultats prometteurs.

Le but de ce mémoire est d'étudier l'utilisation des réseaux de neurones artificiels basés sur la relation quantitative structure propriété (QSPR-$RNA_{optimal}$) pour la modélisation du processus de la rétention des anticancers par la nanofiltration. Un modèle neuronal optimal de type feedforward (QSAR-$RNA_{optimal}$) est caractérisé par une structure comprenant 13 neurones dans la couche d'entrée, 4 neurones dans la couche cachée et un neurone dans la couche de sortie.

Le modèle QSAR-$RNA_{optimal}$ développé dans ce travail a donné des erreurs faibles. Les erreurs absolues moyennes étaient 14.7569% pour phase totale, 5.4732% pour la phase apprentissage et 5.1382% pour phase test. Les racines des erreurs quadratiques moyennes étaient 9.3972% pour phase totale, 9.6866% pour phase apprentissage et 8.1247% pour la phase test. Les erreurs standards de prédictions étaient 11.4431% pour phase totale, 11.8137% pour phase apprentissage et 9.8324% pour phase test. Ce qui montre une bonne robustesse du modèle neuronal "QSPR-$RNA_{optimal}$" établi et la possibilité de la prédiction des différents paramètres qui caractérisent la rétention des anticancers lors de la nano filtration.

L'analyse de sensibilité par la méthode de "poids" a révélé que les variables les plus pertinentes qui peuvent influencer la rétention d'anticancers par les membranes de nanofiltration sont : pH, angle contact, moment dipolaire, potentiel zêta, Length, taux de récupération, pression, SR (NaCl). Toutes les entrées exercent des contributions significatives ($IR > 5\%$) sur la rétention des anticancéreux par nanofiltration, tandis que certaines entrées, telles que MWCO, EqWidth, logD, de, et la température, ont des contributions plus faibles ($IR < 5\%$).

L'analyse diagnostique des valeurs aberrantes et de la couverture du modèle neuronal optimisé montre que le développement du modèle neuronal optimal QSPR-$RNA_{optimal}$ et ses prédictions s'effectuent dans le cadre de la couverture. Il en résulte un modèle neuronal optimal

statistiquement valide. Par conséquent, nous pouvons dire qu'il existe un « bon effet de levier élevé » dans la phase de test et totale.

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

- Les réseaux de neurones, comme d'autres techniques statistiques sont des outils statistiques qui nous permettent d'adapter une fonction non linéaire très générale à un ensemble de points.
- 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 de la nanofiltration.
- La modélisation des réseaux de neurones est une méthode basée uniquement sur des données expérimentales et/ou semi-expérimentales (données issues de modèles analytiques, empiriques, numériques, etc). Par conséquent, la qualité et la quantité des données déterminent 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 extraites des modèles neuronaux ont des applications intéressantes dans la prédiction, l'optimisation, et même le contrôle des processus nanofiltration.

Toute en estimant avoir atteint les objectifs tracés pour cette étude, nous envisageons en perspective plusieurs pistes telles que :

- modéliser la rétention des anticancers par la nanofiltration, mais en utilisant d'autres indicateurs de porosité ($MgCl_2$, KCl, Na_2SO_4, …) ;
- prédiction de la rétention des anticancers par nanofiltration à l'aide d'empreintes moléculaires ;
- modéliser d'autres procédés (microfiltration, ultrafiltration, osmose directe et osmose inverse) à des fins prédictives ;
- étendre l'idée de ce travail à d'autres méthodes de l'intelligence artificielle, telles que les machines à vecteurs de support (SVM), les réseaux de neurones multiples, etc.

Références Bibliographiques

- **Abbas A. et Al-Bastaki N. (2005**)."Modeling of an RO water desalination unit using neural networks". Chemical Engineering Journal. 114 (1) : 139–143.
- **Abdel-Jawad M., Al-Shammari S. et Al-Sulaimi J. (2002).** "Non-conventional treatment of treated municipal wastewater for reverse osmosis". Desalination. 142 (1) : 11-18.
- **Acero J. L., Benitez F. J., Francisco J. R. et Fernando T. (2016).**"Micropollutants removal from retentates generated in ultrafiltration and nanofiltration treatments of municipal secondary effluents by means of coagulation, oxidation, and adsorption processes". Chemical Engineering Journal. 289 : 48–58.
- **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.
- **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.
- **Alfa. A .(2008).** "Procédés de séparation membranaire et leur application dans l'industrie alimentaire", Thèse de doctorat, ACER, Québec, 56. Canada.
- **Alkhatim H. S. (1998)**."Treatment of whey effluents from dairy industries by nanofiltration membranes". Desalination. 119 (1-3) : 177-183.
- **Altalyana N., Jonesa B., Bradda J., Nghiemb L. D. et Alyazichia Y. M. (2006)**."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 œuvre de réseaux de neurones pour la modélisation de cinétiques réactionnelles en vue de la transposition batch/continu." Thèse de Doctorat, Institut National Polytechnique de Toulouse. Toulouse, France.
- **Ammi Y. (2016)** "Modélisation des phénomènes de transport lors d'osmose inverse " Thèse de doctorat. Université Yahia Feres. Médéa. Algérie.
- **Ammi Y., HaniniS et Khaouane L. (2021).** "An artificial intelligence approach for modeling the rejection of anti-inflammatory drugs by nanofiltration and reverse osmosis membranes using kernel support vector machine and neural networks" Comptes Rendus. Chimie. 24 (2) : 243-254.
- **Ammi Y., Khaouane L. et Hanini S (2020).** "A Comparison of Models "Neural Networks and Multiple Linear Regressions" to Describe the Rejection of Micropollutants by Membranes" Journal of chemists and chemical engineers of Croatia. 69 (3-4) : 111–127.
- **Ammi Y., Khaouane L. et Hanini S (2021).** "Stacked neural networks for predicting the membranes performance by treating the pharmaceutical active compounds" Neural Computing and Applications. 33 (12429–12444).

- **Ammi Y., Khaouane L. et Hanini S. (2015)**."Prediction of the rejection of organic compounds (neutral and ionic) by nanofiltration and reverse osmosis membranes using neural net works". Korean Journal of Chemical Engineering.32 (11) : 2300-2310.
- **Ammi Y., Khaouane L. et Hanini S. (2018).**"A Model Based on Bootstrapped Neural Networks for Modeling the Removal of Organic Compounds by Nanofiltration and Reverse Osmosis Membranes" Arabian Journal for Science and Engineering. 43 (11): 6271-6284.
- **Anselme C. et Jacobs E. P. (1996)**. "Ultrafiltration, In : Water Treatment Membrane Process". Chapitre 10, McGraw- Hill : 88
- **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. Frânçe.
- **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. et 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.
- **Audinos. (2000).** "Membranes semi- perméables- membranes de microfiltration Tangentielles", Traité constantes physicochimiques.
- **Baker R. W. (2004).** "Memebrane Technology and application". 2nd ed, John Wiley and Sons Chichester : 287-295.
- **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.
- **Bellona C., Drewes J. E., Xu P. et Amy G. (2004).**"Factors affecting the rejection of organic solutes during NF/RO treatment - a literaturereview". Water Research. 38 (12) : 2795-2809.
- **Benrekia S, Boulanouar D. (2022)."**Modélisation de l'effet de l'adsorption sur la rétention des produits pharmaceutiques lors de nanofiltration et d'osmose inverse". Mémoire de master. Université de Médéa. Médéa. Algérie.
- **Berg P., Hagmeyer G. et Gimbel R. (1997).** "Removal of pesticides and other micro pollutants by nanofiltration". Desalination. 113 (2-3) : 205-208.
- **Bermejo S. et Cabestany J. (1999).**"On-line gradient learning algorithms for K-nearest neighbor classifiers." Dans le livre: "Foundations and tools for neural modeling." Springer Berlin Heidelberg. 1606: 546-555.
- **Bermejo S. et Cabestany J. (2001).**"Learning with nearest neigh bour classifiers." Neural Processing Letters 13(2): 159-181.
- **Bilstad T. (1995).** "Nitrogen separation from domestic waste water by reverse osmosis". Journal of Membrane Science. 102 : 93-102.

- **Girard B., Fukumoto L. R. (2000).** "Membrane processing of fruit juices and beverages :areview". Critical Reviews in Biotechnology. 20 (2) : 109-175..
- **Bishop C. M. (1995).** "Neural networks for pattern recognition." Oxford University Press, New York, USA.
- **BOUCHOUX .A . (2004).** Étude de la nanofiltration pour son intégration dans les procèdes de production d'acides organiques. Thèse de doctorat : génie des procédés. Université Paul Sabatier, Toulouse, UFR Physique-chimie-Automatique.
- **BoukhiarS . (2016).** "Caractérisation des membranes d'ultrafiltration. Application au dessalement et à la concentration de jus de fruit", Université Hassiba Ben Bouali, Chlef.
- **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.
- **Carpenter G. A. et Grossberg S. (1987).**"ART 2: Self-organization of stable category recognition codes for analog input patterns." Applied optics 26(23): 4919-30 .
- **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'Abou bekr Belkaid de Tlemcen. Tlemcen. Algérie.
- **Chabane Y, Koul A. (2022).**" La prédiction de la rétention des pesticides lors de nanofiltration et d'osmose inverse par les modèles QSPR-RNA ". Mémoire de master. Université de Médéa. Médéa. Algérie .
- **Childress A. E. et Desmukh S. S. (1998).** "Effects of humic substances and anionic surfactants on the surface charge and performance of reverse osmosis membranes". Desalination. 118 (1-3) : 167-174.
- **Childress A. E. et Elimelech M. (1996).** "Effect of solution chemistry on the surface charge of polymeric reverse osmosis and nanofiltration membranes". Journal of Membrane Science. 119 (2) : 253-268.
- **Comerton A. M., Andrews R. C., Bagley D. M. et Hao C. (2008).** "The rejection of endocrine disrupting and pharmaceutically active compounds by NF and RO membranes as a function of compound and water matrix properties". Journal of Membrane Science. 313 : 323–335.
- **Conell, D. W. (1989).** "Bioaccumulation of Xenobiotic Compounds". CRC Press : 75–110.
- **Cornelissen E. R., Sijbers P., Van den Berkmortel H., Koning J., De Wit A., De Nil F. et Van Impe J. F. (2001).**"Reuse of leachate waste-water using MEMBIOR technology and reverse osmosis". Membrane Technology. 136 : 6-
- **Cristavao s. (2022).**"Anticancer drugs in the aquatic environment: where should we act and is nanofiltration a solution to this problem."These de doctorat .Université Nova lisbon.

- **Dames M. (2005).** "Gestion de procèdes discontinus : Méthodologie de modélisation et d'optimisation d'opérations de dispersion liquide-liquide en cuve agitée." Thèse de Doctorat, Institut National Polytechnique de Toulouse. Toulouse, France.
- **Demuth H. et Beale M. (2000).** Neural network toolbox for use with Matlab. User'sGuideversion 4. USA, The Math Works Inc.
- **Desmukh S. S. et Childress A. E. (2001).** "Zeta potential of commercial RO membranes : influence of source water type and chemistry".
- **Darwiche Ali . (2010).** Membranes zéolithiques de type MFI pour l'extraction et la séparation de l'hydrogène; thèse doctoral. Desalination. 140 (1) : 87-95.
- **Devillers, J. (2013).** "Computational Toxicology: Volume II, Methods in Molecular Biology". Brad Reisfeld and Arthur N. Mayeno (eds.), Springer Science vol. 930, DOI 10.1007/978-1-62703-059-5_1.
- **Dolar 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.
- **Donghan G., Yuhong D. et Ruiliang Y. (2016).** "Performance analysis of an organic Rankine cycle for a reverse osmosis desalination system using zeotropic mixtures". Desalination. 381 : 38–46.
- **Dornier M., Decloux M., Trystram G. et Lebert A. (1995).**"Dynamic modeling of cross flow microfiltration using neural networks". Journal of Membrane Science. 98 (3) : 263–273.
- **Dreyfus G. (2008).** "Réseaux de neurones: Méthodologie et applications." Eyrolles, Paris, France.
- **Daufin .G and Aimar .P. (2004).**"Séparations par membrane dans l'industrie alimentaire", Technique de l'ingénieure,
- **Galambos I., Molina J. M., Jaray P., Vatai G. et Bekassy-Molnar E. (2004).** "High organic content industrial wastewater treatment by membrane filtration". Desalination. .120-117 : 162
- **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.
- **Hamadache M. (2014).** "Modélisation en vue de la prédiction ou de la corrélation de l'activité toxicologique à partir de la structure moléculaire". Thèsededoctorat. Université de Médéa. Médéa. Algérie.
- **Hornik K. (1991).** "Approximation capabilities of multilayer feedforward networks." Neural Networks 4(2): 251-257.
- **http://www.statsoft.fr/concepts-statistiques/reseaux-de-neurones-automatises/reseaux-de-neurones-automatises.htm#.YjJNDtXMLIU.(30/03/2024)**

- **https://scilogs.fr/intelligence-mecanique/reprenons-bases-neurone-artificiel-neurone-biologique/. (30/03/2024).**
- **https://scilogs.fr/intelligence-mecanique/reprenons-bases-neurone-artificiel-neurone-biologique/. (30/03/2024)**
- **https://www.openpr.com/news/3170003/polymeric-membrane-market-set-to-reach-us-11-153-9-million.e-nerveux_fig1_358676241. (30/03/2024).**
- **https://www.researchgate.net/figure/. (30/03/2024).**
- **https://www.researchgate.net/figure/Figure-III1-Representation-schematique-dun-neurone-biologique-Le-system. (30/03/2024).**
- **Les-techniques-membranaires-les-plus-utilisees-dans-le-traitement-des-eaux_fig9_334331195. (30/03/2024).**
- **Hu J. Y., Ong S. L., Shan J. H., Kang J. B. et Ng W. J. (2003).** "Treatability of organic fractions derived from secondary effluent by reverse osmosis". Water Research. 37 (19) : 4801-4809.
- **Pinnau .I, formation. B. (2000).** Formation and modification of polymére membrane overview formation and modification, 2000, v744, p 1-22. K. Kimura, G. Amy, J. Drewes, Y. Watanabe, Adsorption of hydrophobic compounds onto NF/RO membranes: An artifactleading to overestimation of rejection, J. Memb. Sci. 221: 89–101.
- **REMIGY . J. (2012)**."Filtration membranaire (OI, NF, UF) présentation des membranes et modules", Technique de l'ingénieure.
- **BERLAND .J , JUERY. C .(2002).** Les procédés membranaires pour le traitement de l'eau ; Art Décembre : p (9-10-11).
- **Kanda .A, Duval . M ,Sarazin . D and François.J .(1985).** "Theta point of polyacrylamide in aqueous solution and temperature dependence of the molecular dimensions", Polymer, V. 26, n° 3, 406 - 412.
- **Keedwell E. et Narayanan A. (2005).**"Intelligent bioinformatics: The application of artificial intelligence techniques to bioinformatics problems." Wiley, Chichester, UK.
- **Khaouane L. (2013).** "Etude et Modélisation de la Biosynthèse des Antibiotiques à partir de Différentes Souches Productrices – Cas de Pleuro mutiline". Thèse de doctorat. Université de Médéa. Médéa. Algérie.
- **Kimura K., Iwase T., Kita S. et Watanabe Y., (2009).** "Influence of residual organic macromolecules produced in biological wastewater treatment processes on removal of pharmaceuticals by NF/RO membranes". Water Research. 43 (15) : 3751–3758.
- **Kimura K., Toshima S., Amy G. et Watanabe Y. (2004).** "Rejection of neutral endocrine disrupting compounds (EDCs) and pharmaceutical active compounds (PhACs) by RO membranes". Journal of Membrane Science. 245 (1-2) : 71-87.

- **Kiso Y., Kltao T., Jinnob K. et Miyae M. (1992).**"The effects of molecular width on permeation of organic solute through cellulose acetate reverse osmosis membranes". Journal of Membrane Science. 74 : 95-103.
- **Kiso Y., Kon T., Kitao T. et Nishimura K. (2001a).** "Rejection properties of alkyl phthalates with nanofiltration membranes". Journal of Membrane Science. 182 (1-2) : 205-214.
- **Kiso Y., Muroshige K., Oguchi T., Yamada T., Hhirose M., Ohara T. et Shintani T. (2010).** "Effect of molecular shape on rejection of un charge dorganic compounds by nanofiltration membranes and on calculated pore radii". Journal of Membrane Science. 358 (1-2) : 101–113.
- **Kiso Y., Nishimura Y., Kitao T. et Nishimura K. (2000).** "Rejection properties of non-phenylic presticides with nanofiltration membranes". Journal of Membrane Science. 171 (2) : 229-237.
- **Kiso Y., Sugiura Y., Kitao T. et Nishimura K. (2001b).**"Effects of hydrophobicity and molecular size on rejection of aromatic pesticides with nanofiltration membranes". Journal of Membrane Science. 192 : 1–10.
- **Kohonen T. (1982).** "Self-organized formation of topologically correct feature maps." Biological Cybernetics 43(1): 59-69.
- **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.
- **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.
- **Libotean D., Giralt J., Rallo R., Cohen Y., Giralta F., Ridgway H. F., Rodriguez G. et Phipps D., (2008).**"Organic compound passage through RO". Journal of Membrane Science. 313 (1-2) : 23–43.
- **Lin S.H. et Yang C.R.(2004).**"Chemical and physical pretreatment of chemical mechanic al polishing was tewater from semiconductor 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.
- **Marty P. (1999).** "Traitement des effluents par filtration membranaire". Industries Alimentaires et Agricoles. 116 (10) : 50-55.
- **Maurel. (2001).** "dessalement de l'eau de mer et l'eau saumâtres- et autres procédés non conventionnels d'approvisionnement en eau douce", Édition technique et documentation, Paris.
- **Metaiche M. (2014).** "Technologie membranaire". Faculté des sciences et des sciences appliquées de l'université de Bouira, Algérie. Chapitre 1: 4-6.

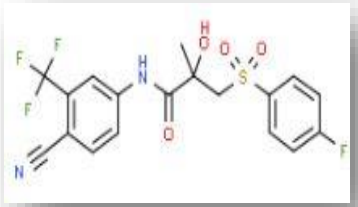
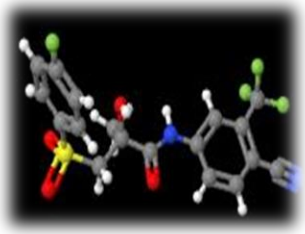
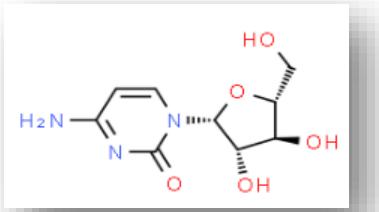
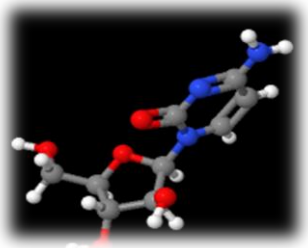
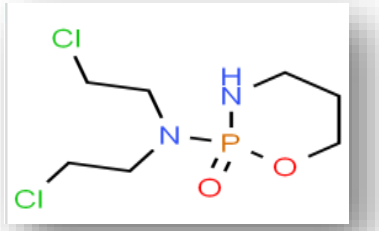
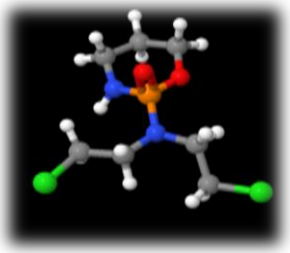
- **Moody J. et Darken C. J. (1989).** "Fast learning in networks of locally-tuned processing units." Neural computation 1(2): 281 - 294.
- **Moulai mostefa M, Tassist R. (2023)."** Modélisation de l'hémodialyse". Mémoire de master. Université de Médéa. Médéa. Algérie .
- **Nascimento C. A. O., Giudici R. et Guardani R. (2000).**"Neural network base dapproach for optimization of industrial chemical processes." Computers &Chemical Engineering-2303 :(10-9)24 .2314
- **Nghiem L. D., Schäfer A. I. et Elimelech M. (2004).** "Removal of natural hormones by nanofiltration membranes : measurement, modelling, and mechanisms". Environment al Science & Technology . .1896–1888 : (6) 38
- **Nielsen C. E., (1995).** "Membrane filtration for strong effluents". Water Management International, Water and Was te water Treatment.77-80.
- **Nechad .A. (2009)**. Élaboration d'une membrane de filtration d'eau à base de polyéthylène basse densité et gypse de Bouzggaia. Mémoire de magister. Université Hassiba Ben Bouali. Chlef.
- **Norberg D., Hong S., Taylor J. et Zhao Y. (2007).**"Surface characterization and performance evaluation of commercial fouling resistant low-pressure RO membranes". Desalination. 202 (1-3) : 45-52.
- **Ozaki H. et Li H. (2002).**"Rejection of organic compounds by ultra-low pressure reverse osmosis membrane". Water Research. 36 (1) : 123-130.
- **P .Remise. (2006).** "étude des rétro lavages assistés par l'air et controle du clomatage résiduel : application a la production d'eau potable en filtration frontale" .
- **Padilla A. P. et Tavani E. L. (1999).**"Treatment of an industrial effluent by reverse osmosis" Desalination. .226-219 : (1) 126
- **Guillaume. P.(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.
- **Audinos R, Audinos and P. Isoard. (1986).** "Glossaire des termes techniques des procédés à membrane", société francaise de filtration: (142)p.
- **Barbar. R . (2006).** Contribution à la mise au point de nouveaux matériaux membranaires pour la filtration de milieux complexes : Etude physico-chimique de membranes d'acétate de cellulose et de méthyl cellulose partiellement alkylée obtenues par immersion- précipitation", Thèse doctorat, Institut National Polytechnique de Lorraine.
- **Baker. R. W. (2004). "**Membrane Technology and Applications", 2nd ed, John Wiley and Sons.
- **Ravikumar Y. V. L., Kalyani S., Satyanarayana S. V. et Sridhar, S. (2014).**"Processing of pharmaceutical effluent condensate by nanofiltration and reverse osmosis membrane techniques". Journal of the Taiwan Institute of Chemical Engineers. 45 (1) : 50-56.

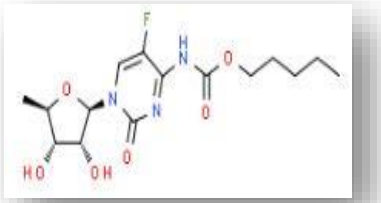
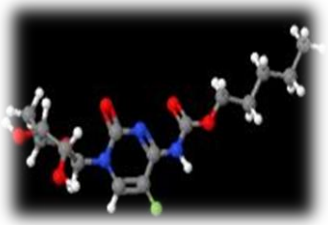
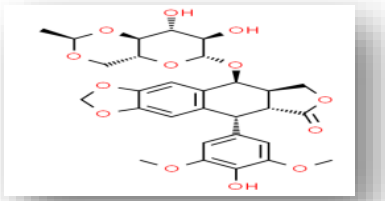
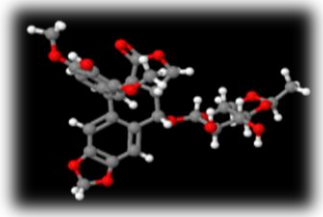
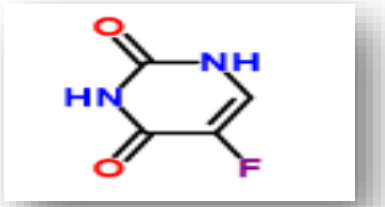
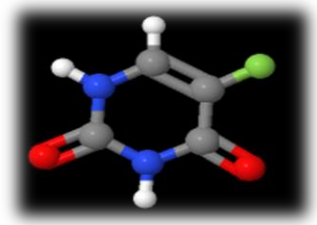
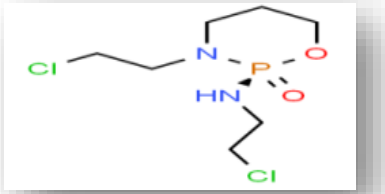
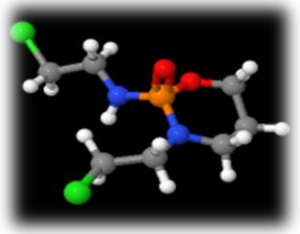
- **Ricci B. C., Ferreira C. D., Aguiar, A. O. et Amaral M. C. S. (2015).**" Integration of nanofiltration and reverse osmosis for metal separation and sulfuric acid recovery from gold mining effluent". Separation and Purification Technology. 154 : 11-21.
- **Sara.S .(2009).** "Étude hydrodynamique et transfert de chaleur et de masse dans un module membranaire", Mémoire de magister, université Djelfa .Algérie.
- **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.
- **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.
- **Sahoo GB. et Ray C. (2006).**"Predicting flux decline in cross flow membranes using artificial neural networks and genetic algorithms". Journal of Membrane Science. 283 (1-2) : 147-157.
- **Santos J. L. C., de Beukelaar P., Vankelecom I. F. J., Velizarov S. et Crespo J. G. (2006)**. "Effect of solute geometry and orientation on the rejection of uncharged compounds by nanofiltration". Separation and Purification Technology. 50 (1) : 122–131.
- **Schäfer A. I., Nghiem L. D. et Waite T. D. (2003).** "Removal of the natural hormone estrone from aqueous solutions using nanofiltration and reverse osmosis". Environmental Science &Technology. 37 (1) : 182-188.
- **Schäfer A. I., Pihlajamäki A., Fane A. G., Waite T. D. et Nyström M. (2004).**"Natural organic matter removal by nanofiltration: effects of solution chemistry on retention of low molar mass acids versus bulk organic matter". Journal of Membrane Science. 242 (1-2) : 73-85.
- **Shahmansouri A. et Christopher B. (2013).**"Application of quantitative structure property relationships (QSPRs) to predict the rejection of organic solutes by nanofiltration". Separation and Purification Technology. 118 : 27–638.
- **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. (2010).** "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 asymptotic equivalence of choice of model by cross-validation and Akaike'scriterion." Journal of the Royal Statistical Society B 39(1): 44-47.

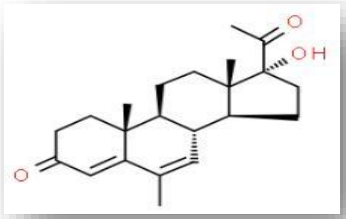
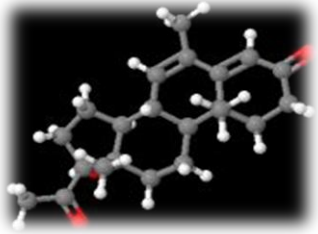
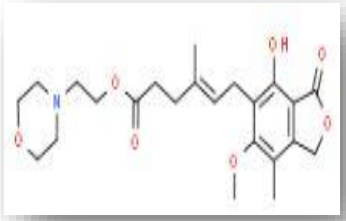
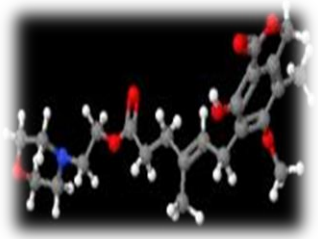
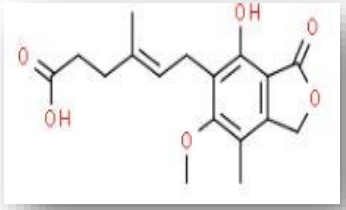
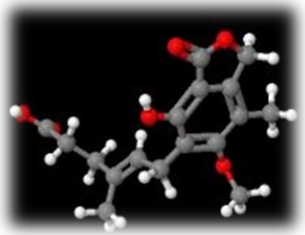
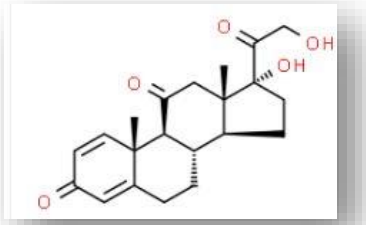
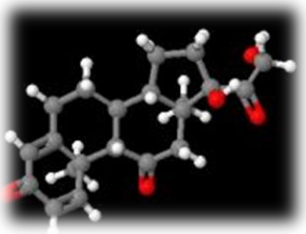
- T**aheran M., Brar S. K., Verma M., Surampalli R. Y., Zhang T. C. et Valero J. R (2016).** "Membrane processes for removal of pharmaceutically active compounds (PhACs) from water and wastewaters". Science of The Total Environment. 547 : 60-77.
- **Talbi M. L. (2011).** "Analyse et traitement du signal électrocardiographique (ECG)." Thèse de Doctorat, Université Mentouri de Constantine. 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.
- **Van der Bruggen B., Schaep J., Wilms D. et Vandecasteele C. (1999).** "Influence of molecular size, polarity and charge on the retention of organic molecules by nanofiltration". Journal of Membrane Science. 156 (1) : 29-41.
- **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 organic solutes in aqueous solutions with nanofiltration". Journal of Membrane Science. 322 (1) : 52–66.
- **Yangali-Quintanilla V., Kim T. U., Kennedy M. et Amy G. (2008).**"Modeling of RO/NF membrane rejection of PhACs and organic compounds :astatistical analysis". Drinking Water Engineering and Science. 1 (1) : 7–15.
- **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 .
- **Yangali-Quintanilla V., Verliefde A., Kime T. U., Sadmani A., Kennedy M. et Amy G. (2009b).** "Artificial neural network models based on QSAR for predicting rejection of neutral organic compounds by polyamide nanofiltration and reverse osmosis membranes". Journal of Membrane Science. 342 (1-2) : 251-262.

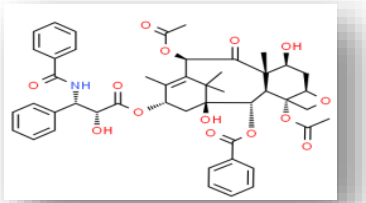
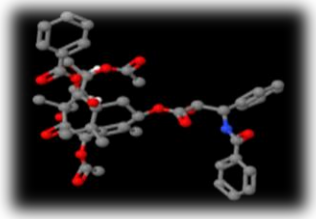
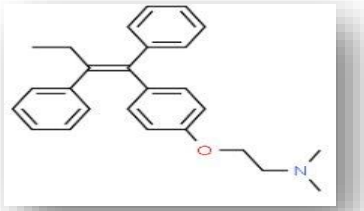
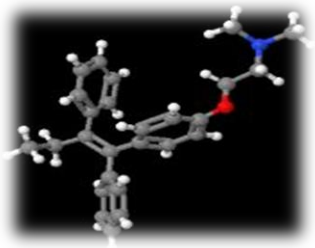
Annexe A

Tableau 1. Liste de 13 anticancers utilisées pour développer les modèles QSPR-RNA.

#	Nom	Formule brute	MW (g/mol)	Structure chimique 2D	Structure chimique 3D
01	**Bicalutamide**	$C_{18}H_{14}F_4N_2O_4S$	430.377		
02	**Cytarabine**	$C_9H_{13}N_3O_5$	243.217		
03	**Cyclophosphamide**	$C_7H_{15}Cl_2N_2O_2P$	261.09		

04	**Capecitabine**	$C_{15}H_{22}FN_3O_6$	359		
05	**Etoposide**	$C_{29}H_{32}O_{13}$	588.57		
06	**Fluorouracil**	$C_4H_3FN_2O_2$	130.1		
07	**Ifosfamide**	$C_7H_{15}Cl_2N_2O_2P$	261.09		

08	**Megestrol**	$C_{22}H_{30}O_3$	343		
09	**Mycophenolate Mofetil**	$C_{23}H_{31}NO_7$	433		
10	**Mycophenolic Acid**	$C_{17}H_{20}O_6$	320		
11	**Prednisone**	$C_{21}H_{26}O_5$	372		

12	**Paclitaxel**	$C_{47}H_{51}NO_{14}$	853.93		
13	**Tamoxifen**	$C_{26}H_{29}NO$	358.428		

Annexe B

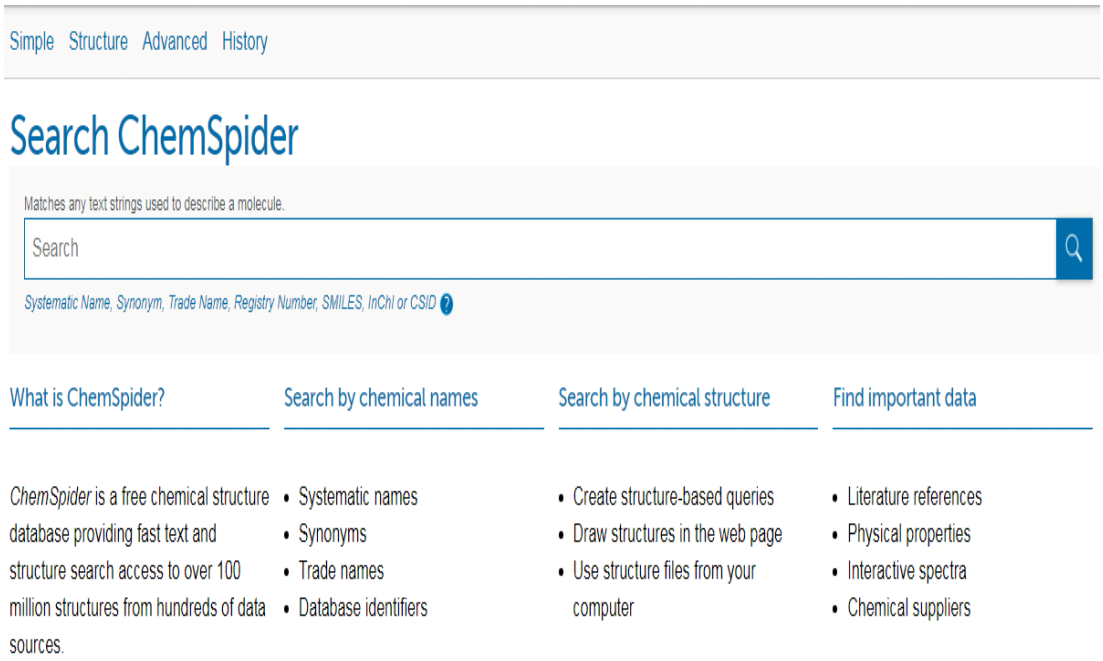
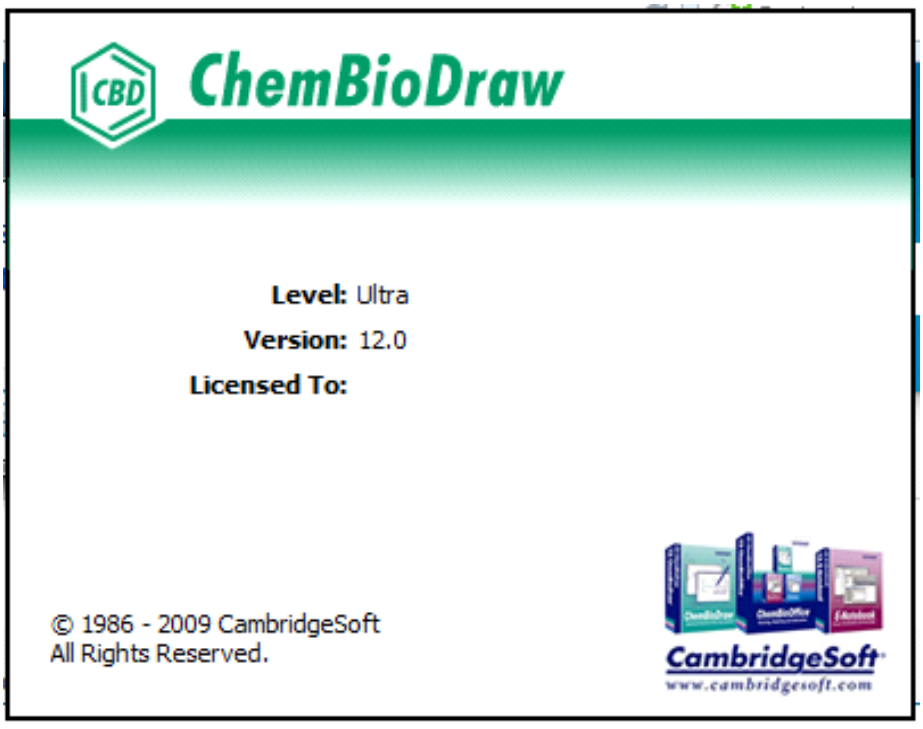
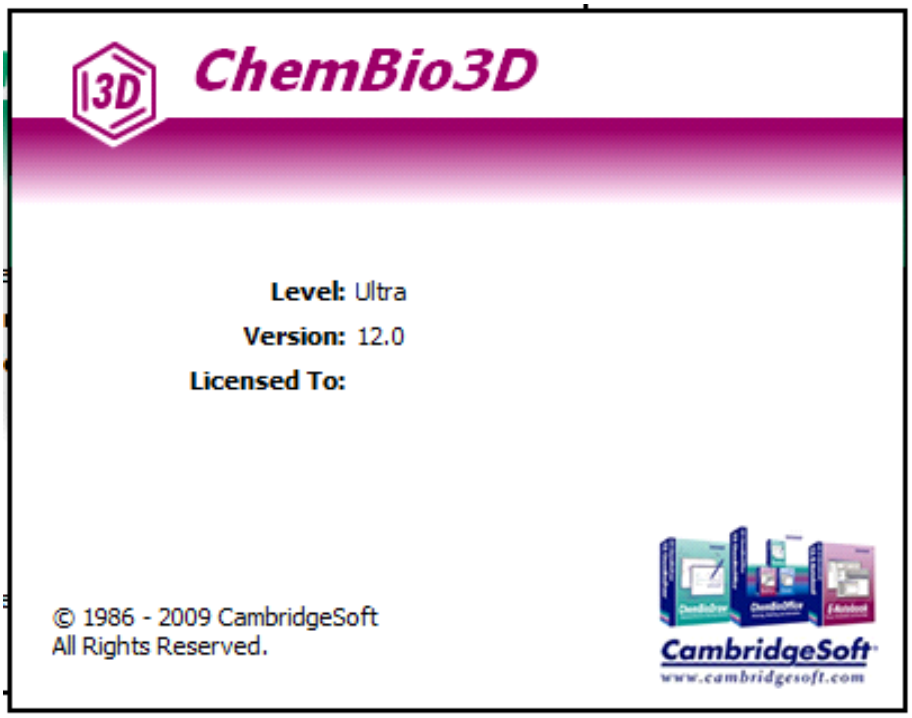


Figure 1. Capture d'écran du site intitulé Chemspider.



(a)



(b)

Figure 2. Capture d'écran des logiciels : (a) ChemBioDraw et (b) ChemBio3D.

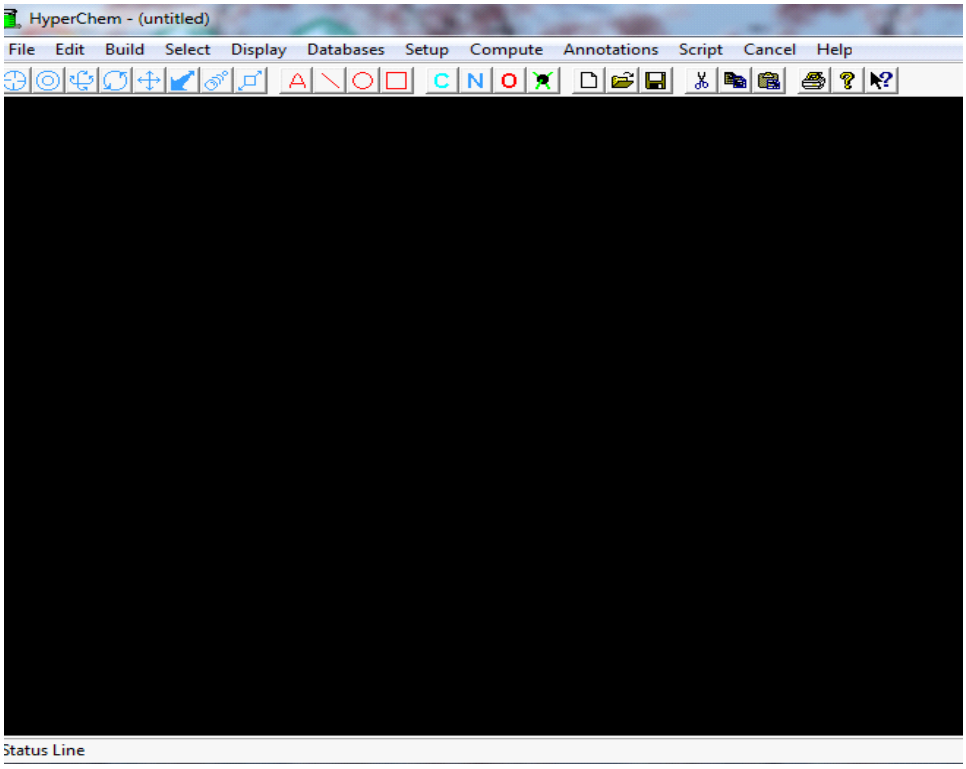


Figure 3. Capture d'écran de logiciel Hyperchem



Figure 4. Capture d'écran de logiciel STATISTICA

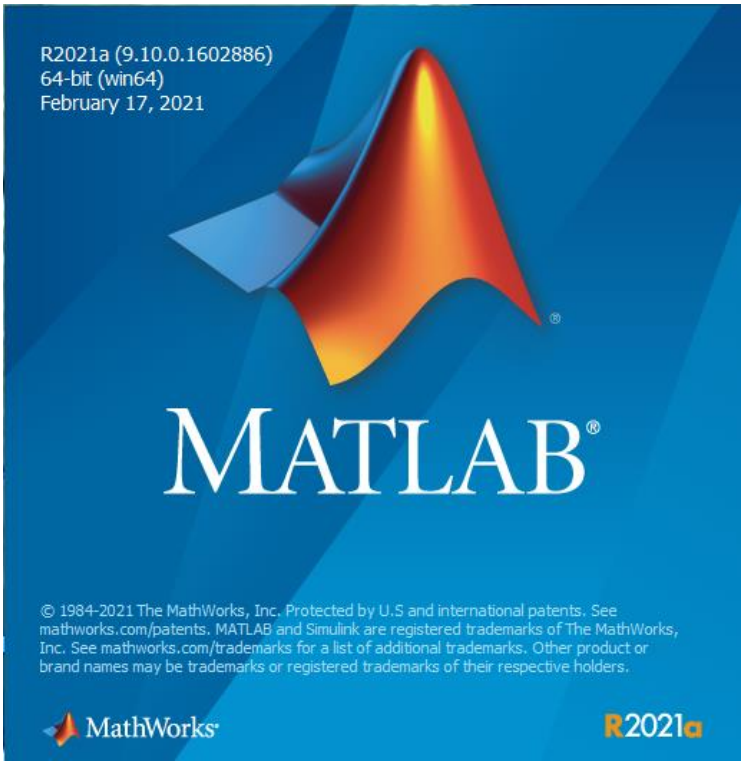


Figure 5. Capture d'écran de logiciel MATLAB.