**République Algérienne Démocratique et Populaire**
**Ministère de l'Enseignement Supérieur et de la Recherche Scientifique**



**UNIVERSITE YAHIA FARES DE MEDEA**
**Faculté des Sciences et de la Technologie**
**Département de Génie de Procédés et Environnement**

# MEMOIRE DE MASTER

**En vue de l'obtention du Diplôme de Master en Génie des Procédés**
**Option : Procèdes pharmaceutiques**

## Thème

# *Prédiction des paramètres pharmacocinétique des médicaments par les RNA-QSAR*

**Etudié par :**
- ✓ **Mme ALOUAOUI Ilham**
- ✓ **Mlle KHAOUANE Affaf**

**Dirigé par :**
**Pr HANINI Salah**

**Promotion 2015-2016**

الحمد لله كما ينبغي

لجلال وجهه و عظيم سلطانه

# *Dédicaces*

*A mes chers parents,*

*A ma sœur Latifa,*

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

*Affaf.*

# *Dédicaces*

**A** *mon père,*

"*L'épaule solide, l 'œil attentif compréhensive et la personne* la plus digne de mon estime et de mon respect. *Aucune dédicace ne saurait exprimer mes sentiments que ALLAH te préserve dans sa vaste paradis* ",

**A** ma Mère,

"*Tu m ' as donné la vie, la tendresse et le courage pour réussir.Tout ce que je peut t'offrir ne poura éxprimer l'amour et la reconnaissance que je te porte . En témoignage je t'offre ce Modeste travail pour te remercier pour tes sacrifices et pour l'affection dont tu ma toujour entourée ."* ,

**A** *mes frères Ridha, Sofiane, Lotfi et Mohammed,*

**A** *mes soeurs Linda, Ramila et Nesrine,*

**A** *mon marie et mes enfants moncif, Ziad et siradj.*

*ILHAM*

## *Remerciements et gratitude*

*En tout premier lieu, parce que la tradition l'exige certes, mais surtout parce qu'il le mérite, on tient à remercier notre directeur de Mémoire, Monsieur* **Salah HANINI**, *Professeur à l'UYFM, qui nous a guidé avec grande patience tout au long de l'élaboration de ce travail, pour ses aides précieuses qui ont judicieusement éclairé notre chemin vers l'aboutissement et la réussite de la concrétisation de notre mémoire de Master.*

*Nous tenons à exprimer nos plus vifs remerciements à Melle* **Latifa KHAOUANE**, *Maître de Conférences (B) à l'UYFM pour ses aides, son accompagnement et son éclairage techniques. On la doit les remerciements les plus sincères.*

*Nous tenons également à présenter nos profonde gratitude au membre de jury d'avoir acceptés de juger ce modeste travail.*

*Merci à tous ceux qui nous ont aidés de près ou de loin durant ces mois de recherche et d'études. On ne peut conclure sans un mot de remerciement à tous nos enseignants et à tous nos collègues.*

**ملخص** التنبؤ بالمعلمات الدوائية للأدوية باستعمال **QSAR-RNA**

تلعب العلاقة الكمية للنشاط بالهيكل QSAR دورا هاما في العلوم الصيدلية. وكان الهدف الرئيسي من هذا العمل هو تطوير نموذج QSAR باستخدام الشبكة العصبية (RNA) للتنبؤ بارتباط سلسلة متكونة من 277 دواء بالبروتينات البلازمية (*fb*.) انطلاقا من البنية التركيبية للدواء تم إنشاء نماذج مختلفة من واصفات الأدوية (مجموعها 1666). لذلك تم وضع إجراءات اختيارية لاستخراج الواصفات الأكثر أهمية و المتمثلة في عشرة واصفات. ثم تم تصميم النموذج الأمثل (شبكة عصبية مصنعة بعشرة عصبونا في الطبقة الداخلة وطبقة مخفية وعصبون في الطبقة الخارجة) للتنبؤ بارتباط الدواء بالبروتينات البلازمية. قيم REQM من أجل حساب *fb*هي: للتدريب(0.11) للاختبار(0.12) للتنبؤ(0.16). المقارنة بين هذه القيم والقيم الإحصائية الأخرى تظهر تفوق النموذج العصبي على نموذج الانحدار الخطي المتعدد RLM

كلمات المفاتيح: تصميم -الشبكة العصبية (RNA) -الانحدار الخطي المتعدد(RLM)-العلاقة الكمية للنشاط بالهيكل QSAR واصفات- المعلمات الدوائية.

## Résumé : Prédiction des paramètres pharmacocinétique des médicaments par les RNA-QSAR

Les Relations Quantitative Structure Activité (QSAR) sont appelés à jouer un rôle important dans le domaine des sciences pharmaceutiques. L'objectif principal de ce travail a été le développement d'un modèle QSAR utilisant un réseau de neurone feed-forward pour la prédiction de la fraction du médicament liée aux protéines plasmatiques (*fb*) d'une série de 277 médicaments. Différents types de descripteurs (1666 au total) ont été générées à partir des structures des médicaments. Une procédure de sélection a été avancée pour la sélection des descripteurs les plus pertinents et a fourni dix descripteurs. Ensuite, un modèle de réseaux de neurone à une couche caché (10 neurones dans la couche d'entrée, une couche cachée et 1 neurone dans la couche de sortie) a été conçu et optimisé pour la prédiction de $f_b$. Les valeurs de la REQM (Racine de l'Erreur Qudratique Moyenne) pour le calcul de $f_b$ pour la base d'apprentissage, de test et de prédiction sont 0,11, 0,12 et 0,16, respectivement. La comparaison entre ces valeurs et d'autres valeurs statistiques révèlent la supériorité du modèle neuronale sur le modèle RLM (Régression Linéaire Multiple).

**Mots clés :** Modélisation, RNA, MLR, QSAR, $f_b$, Descripteur,pharmacocinétique.

## Summary: Prediction of pharmacokinetic parameters of drugs by RNA-QSAR

Quantitative Structure Activity Relationship (QSAR) models are expected to play an important role in the pharmaceutical sciences. The main aim of the present work was the development of a QSAR model using a feed-forward neural network (FFNN) for the prediction of fraction bound to plasma proteins ($f_b$) of a series of 277 drugs. Different kinds of descriptors (1666 in total) were generated from the drug structures. A feature selection procedure was advanced for variable selection and has provided ten descriptors. Then three layers FFNN (10 neurons in the input layer, 1 hidden layer, and 1 neuron in the output layer) was constructed and optimized for the prediction of $f_b$. The values of *RMSE* (Root Mean Squared Error) for the calculated $f_b$ of training, test, and prediction sets are 0.11, 0.12, and 0.16, respectively. Comparison between these values and other statistical values reveal the superiority of the neuronal model over MLR (Multi-Linear Regression).

**Keywords**: Modélisation, RNA, MLR, QSAR, $f_b$, Descriptors, pharmacokinetics.

# *Table des matières*

Table des matières i
Abréviations iv
Nomenclature et symboles v
Liste des tableaux vii
Liste des figures viii

**Introduction générale** 1

**Chapitre I. Pharmacocinétique et métabolisme des médicaments**
I.1. Définition et but 3
I.2. Absorption d'un médicament 4
I.2.1. Modalité d'absorption 5
*I.3.1.1. Diffusion passive* 5
*I.3.1.2. Transport actif* 5
I.2.2. Les paramètres influençant l'absorption 6
*I.2.2.1. Les caractéristiques du médicament* 6
*I.2.2.2. Les caractéristiques liées à l'individu* 6
I.2.3. Evaluation de l'absorption : la biodisponibilité 6
*I.2.3.1. Le facteur quantitatif F* 7
*I.2.3.2. L'interprétation de la biodisponibilité* 8
*I.2.3.3. Le facteur vitesse Tmax* 8
*I.2.3.4. L'intérêt de l'étude de la biodisponibilité* 9
I.3. Distribution 10
I.3.1. Fixation aux protéines plasmatiques 10
I.3.2. Interprétation de la defixation protéique 11
I.3.3. Diffusion tissulaire 11
I.3.4. Interprétation de la diffusion tissulaire 12
I.3.5. Volume apparent de distribution 12
I.3.6. Facteurs modifiants la distribution 13
I.4. Métabolisme et élimination des médicaments 14
I.4.1. Biotransformation 14
*I.4.1.1. Définition* 14
*I.4.1.2.Réactions de phase I* 15
*I.4.1.3. Réaction de phase II* 15
*I.4.1.4. Comprendre le métabolisme des médicaments* 16
I.4.2. Elimination 16
*I.4.2.1. Elimination hépatique* 16
*I.4.2.2. Elimination rénale* 17
*I.4.2.3. Autres voies d'excrétion* 17
I.4.3. Quantification du métabolisme et de l'élimination 17
*I.4.3.1. Notion de clairance* 17
*I.4.3.2. Notion de coefficient d'extraction* 18
*I.4.3.3. Clairance hépatique* 19
*I.4.3.4. Facteurs influençant la clairance hépatique* 19
*I.4.3.5. Clairance rénale* 20

*I.4.3.6. Facteurs influençant la clairance rénale* 20
*I.4.3.7. Calcul de la clairance* 20
I.4.4. Demi-vie d'un médicament 21

**Chapitre II. Les réseaux de neurones**
II.1. Introduction 23
II.2. Historique 23
II.3. Notions de base sur les réseaux de neurones 24
II.3.1. Du neurone biologique au modèle mathématique 24
*II.3.1.1. Le modèle neurophysiologique* 24
*II.3.1.2 Le modèle mathématique* 26
II.3.2. Architecture des réseaux de neurones 28
*II.3.2.1. Réseau de neurones non bouclé (en anglais feedforward)* 28
*II.3.2.2. Réseau de neurones bouclé (ou récurrent)* 29
II.3.3. Types d'apprentissage des réseaux de neurones 30
*II.3.3.1. Apprentissage supervisé* 30
*II.3.3.2. Apprentissage non supervisé* 30
II.4. Les réseaux de neurones les plus utilisés 31
II.4.1. Le perceptron monocouche (perceptron simple) 31
II.4.2. Le perceptron multicouche (en anglais Multi Layer Perceptron) 31
II.4.3. Les réseaux à fonction radiale 31
II.4.4. Les cartes auto-organisatrices de Kohonen (SOM) 31
II.4.5. Les réseaux de Hopfield 32
II.4.6. Les réseaux ART 32
II.5. Conception d'un modèle neuronal 32
II.5.1. Détermination des entrées/sorties du réseau de neurones 33
II.5.2. Construction de la base de données et choix des échantillons 33
II.5.3. Définition de la structure du réseau 34
*II.5.3.1. Nombre de couches cachées* 34
*II.5.3.2. Nombre de neurones cachés* 34
II.5.4. Apprentissage 35
II.5.5. Evaluation des performances du réseau de neurones 38
II.6. Application des réseaux de neurones en bioprocédés 38
II.6.1. Modélisation et prédiction 39
II.6.2. Capteurs logiciels (en anglais soft sensors) 39
II.6.3. Reconnaissance des formes (en anglais pattern recognition) 40
II.6.4. Optimisation 40
II.6.5. Contrôle 41
II.7. Conclusion 41

**Chapitre III. Modélisation QSPR/QSAR**
III.1. Généralités sur la modélisation QSPR/QSAR 42
III.2. principe des méthodes QSPR/QSAR 43
III.3. Méthodologie générale d'une étude QSPR/QSAR 43
III.3.1. Base de données 44
III.3.2. Descripteurs moléculaires 46
*III.3.2.1. Les descripteurs moléculaires théoriques* 46
*III.3.2.2. Descripteurs moléculaires empiriques* 48
III.3.3. Méthode d'analyse des données 48

*III.3.3.1. Interprétation et validation d'un modèle* QSPR/QSAR 48
*III.3.3.1.1. Validation interne* 49
*III.3.3.1.2. Validation externe* 49
*III.3.3.1.3. Domaine d'applicabilité* 50
III.4. Application des méthodes QSPR/QSAR 51

**Chapitre IV.** ***Modélisation du taux de fixation des médicaments aux protéines plasmatiques***

IV.1. Introduction 53
IV.2. Procédure de développement des modèles prédictives 53
IV.2.1. Collecte des données 54
IV.2.2. Calcul des descripteurs 54
IV.2.3. Sélection des descripteurs les plus influents 56
IV.2.4. Développement des modèles prédictives 57
*IV.2.4.1. La régression linéaire multiple(RLM)* 57
*IV.2.4.2. Les réseaux de neurones artificiels (RNA)* 57
IV.3. Résultats et discussion 58
IV.3.1. Analyse des descripteurs retenus 58
IV.3.2. Performance du modèle neuronale 59
IV.3.3. Comparaison avec le modèle RLM 62
IV.4. Conclusion 65

**Conclusion générale et perspectives** 66
**Références bibliographiques** 68

# *Abréviations*

**DS** **D**éviation **S**tandard

**ESP** **E**rreur **S**tandard de **P**rédiction $$ESP(\%) = \frac{REQM}{Y_e} \times 100$$

Avec $$Y_e = \frac{\sum_{i=1}^{n} Y_{i,\exp}}{n}$$

**PMC** **P**erceptron **M**ulti-**C**ouche

**P-value** **P**robabilité

**QSAR** **Q**uantitative **S**tructure-**A**ctivity **R**elationship

**QSPR** **Q**uantitative **S**tructure-**P**roperty **R**elationship

**REQM** **R**acine de l'**E**rreur **Q**uadratique **M**oyenne $$REQM = \sqrt{\frac{\sum_{i=1}^{n} \left(Y_{i,\exp} - Y_{i,pred}\right)^2}{n}}$$

**RLM** **R**égression **L**inéaire **M**ultiple

**RNA** **R**éseaux de **N**eurones **A**rtificielle

**RNFF** **R**éseau de **N**eurones de type **F**eed **F**orward

**SMILE** **S**implified **M**olecular Input **L**ine **E**ntry **S**pecification

**VIF** **F**acteur d'**I**nflation de la **V**ariance

# *Nomenclatures et symboles*

| Lecture | Désignation | Unité |
|---|---|---|
| $b$ | Biais d'un neurone artificiel | (-) |
| $E_i$ | $i^{ème}$ entrée du modèle | (-) |
| $f_b$ | Taux de liaison des médicaments aux protéines plasmatiques | (%) |
| $L$ | Nombre de couche cachée | (-) |
| $R$ | Coefficient de corrélation | (-) |
| $w$ | Poids synaptique des connexions inter-neuronales | (-) |
| $Z_j$ | $j^{ème}$ sortie de la couche cachée | (-) |

## 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

| Lecture | Désignation |
|---|---|
| *cal* | Calculée |
| *exp* | Expérimentale |
| *H* | Cachée (Hidden) |
| *I* | Entrée (Input) |
| *O* | Sortie (Output) |

## Indices

| Lecture | Désignation |
|---|---|
| *H* | Cachée (Hidden) |
| *i* | Elément i d'un vecteur |
| *i,exp* | Elément i d'un vecteur de valeurs expérimentales |
| *i,cal* | Elément i d'un vecteur de valeurs prédites |
| *ij* | Croisement de la ligne i avec la colonne j d'une matrice |
| *j* | Elément j d'un vecteur |
| *max* | Maximum |
| *min* | Minimum |
| *O* | Sortie (Output) |

# *Liste des tableaux*

## Chapitre II. Les réseaux de neurones


| | | |
|---|---|---|
| **Tableau II.1** | Fonctions de transfert usuelles, *a=f(n)*. | 27 |
| **Tableau II.2** | Résumé des caractéristiques des algorithmes d'apprentissage de rétro-propagation implantés dans MATLAB® | 37 |


## Chapitre III. *Modélisation QSPR/QSAR*


| | | |
|---|---|---|
| **Tableau III.1** | Classification générale des descripteurs moléculaires | 47 |
| **Tableau III.2** | Classification d'ensemble des descripteurs moléculaires empiriques. | 48 |


## Chapitre IV. Modélisation du taux de fixation des médicaments aux protéines plasmatiques


| | | |
|---|---|---|
| **Tableau IV.1** | Nombre de descripteurs calculés. | 56 |
| **Tableau IV.2** | Notation, définition, type et VIF des descripteurs sélectionnés.. | 58 |
| **Tableau IV.3** | Matrice d'autocorelation des descripteurs selectionés | 59 |
| **Tableau IV.4** | Comparaison entre le modèle neuronal (RNA)et le modèle RLM en terms de la *REQM*, l'*ESP* et le *R*. | 64 |

# *Liste des figures*


**Chapitre I. Pharmacocinétique et métabolisme des médicaments**

**Figure I.1** Le trajet du médicament dans l'organisme. 4

**Figure I.2** L'allure générale de la Quantité de médicament présent dans la circulation général en fonction du temps. 8

**Chapitre II. Les réseaux de neurones**

**Figure I.1** Schéma d'un neurone. 25

**Figure I.2** Schéma d'une synapse. 25

**Figure I.3** Neurone formel. 26

**Figure I.4** Réseaux de neurones les plus utilisés. 28

**Figure I.5** Structure d'un réseau de neurones non bouclé. 29

**Figure I.6** Un réseau de neurones bouclé à deux entrées 29

**Chapitre III. *Modélisation QSPR/QSAR***

**Figure III.1** Schéma d'élaboration et validation d'un modèle QSTR 45

**Figure III.2** Représentation schématique du domaine d'applicabilité d'un modèle 51

**Chapitre IV. Modélisation du taux de fixation des médicaments aux protéines plasmatiques**

**Figure IV.1** Schema de la procedure utilisée pour la selection des descripteurs les plus pertinent. 54

**Figure IV.2** Capture d'écran du site intitulé DrugBank version 4.3 55

**Figure IV.3** Comparaison entre les valeurs expérimentales et les valeurs calculées par RNA: **(a)** base d'apprentissage, **(b)** base de test. 60

**Figure IV.4** Representation shématique du réseau de neurones optimise 61

**Figure IV.5** Comparison entre les valeurs experimentale et les valeurs calculées par la RLM : **(a)** Based'apprentissage, **(b)** base de test. 63

**Figure IV.6** Histogrammes de comparaison entre le modèle neuronale (RNA) et le modèle RLM: **(a)***REQM*, **(b)** *ESP*, **(c)***R*. 65

# *Introduction générale*

La liaison d'un médicament aux protéines plasmatiques, principalement à l'albumine sérique et α-acide glycoprotéine, est l'un des nombreux facteurs qui influencent l'ADME (Adsorption, distribution, métabolisme et excrétion) et la toxicité. La liaison d'un médicament aux protéines plasmatiques réduit la fraction libre du médicament dans la circulation sanguine qui seraient autrement disponibles pour la pénétration dans les tissus pour atteindre la cible thérapeutique ou le rein pour l'élimination (Zhang *et al.* 2012). De nombreuses méthodes expérimentales sont disponibles pour mesurer la liaison aux protéines plasmatiques (ultrafiltration, ultracentrifugation, dialyse à l'équilibre, l'analyse frontale haute performance, micro extraction en phase solide (SPME), adsorption sur charbon, et micro dialyse in vivo).

Cependant, ces approches expérimentales pour le développement de nouveaux médicaments candidats sont assez chères et prennent du temps. Cela a entraîné un besoin urgent pour la modélisation computationnelle pour prédire la liaison d'un médicament protéique. Les approches prédictives comprennent des techniques de modélisation quantitatives structure-activité relation (QSAR), qui attirent de plus en plus les chimistes et les scientifiques dans l'industrie pharmaceutique (Helguera *et al.* 2008). Les modèles QSAR sont expecté à jouer un rôle important dans les sciences pharmaceutiques.

L'utilisation de ces méthodes alternatives à l'expérimentation sont devenue d'un grand intérêt et sont même recommandées dans les nouvelles règlementations afin d'obtenir les données nécessaires à l'enregistrement des substances.

L'élaboration des modèles mathématiques QSAR reliant les propriétés physicochimiques et les activités biologiques à la structure moléculaire permet, d'une part, d'expliquer l'origine de ces activités/propriétés et, d'autre part, de les prédire pour des molécules pour lesquelles les données expérimentales ne sont pas disponibles.

Ces méthodes sont à l'heure actuelle principalement utilisées pour le criblage des composés toxiques par exemple. La procédure de criblage consiste à utiliser ces méthodes pour prédire a priori les performances (ou la dangerosité) des substances pour des applications données. Ainsi, il est possible d'identifier les molécules présentant le plus fort potentiel pour la propriété/ l'activité attendue. Les premières applications de ces méthodes, basées sur des analyses statistiques, ont concerné principalement des applications biologiques, toxicologiques ou pharmaceutique (Bellifa 2015).

La présente mémoire s'organise en quatre chapitres :

Le premier chapitre est dédié à la présentation des paramètres pharmacocinétiques

Le deuxième chapitre est une présentation de la théorie des réseaux de neurones. Après avoir exposé un aperçu historique sur les RN, nous détaillons les différents éléments de base de cet outil mathématique. Nous limitons notre présentation à des aspects qui sont utiles pour la suite de ce travail.

Le troisième chapitre est une exposition des modèles *QSAR*. On y parlera des éléments nécessaires à l'élaboration de ces modèles, en particulier la base de données, les descripteurs moléculaires ainsi que la sélection des plus pertinents.

Le quatrième chapitre est consacré, quant à lui, à l'élaboration du modèle *QSAR*, on discutera du volet matériels et méthodes, de la procédure de développement des modèles et des outils statistiques permettant d'analyser leur performance.

Enfin, une conclusion générale qui synthétise l'ensemble des résultats obtenus dans ce travail et des perspectives de recherche pour la continuité de cette mémoire.

# *Chapitre 1*

# *Pharmacocinétique et métabolisme des médicaments*

## I.1.Définition et buts

La pharmacocinétique a pour but d'étudier le devenir d'un médicament dans l'organisme. Elle repose sur l'étude des concentrations des médicaments obtenus suite à l'administration d'une dose du médicament (Lechat 2006).

La pharmacocinétique permet l'étude de la relation dose-concentration en médicament qu'elle soit concentration sanguine ou concentration plasmatique.

La connaissance de la pharmacocinétique d'un médicament permet d'en définir les modalités d'administration à savoir la voie d'administration par exemple IV, PER OS…etc. la forme galénique : comprimé, gélule. La posologie : en fonction du poids, de la fonction hépatique ou de la fonction rénale du patient, la posologie en fonction des médicaments associés et enfin du rythme d'administration(Stanke-Labesque 2012).

On peut distinguer schématiquement 4 étapes dans la pharmacocinétique d'un médicament(Loichot et Grima 2004)

- ✓ Son absorption
- ✓ Sa diffusion dans l'organisme
- ✓ Son métabolisme
- ✓ Son élimination de l'organisme

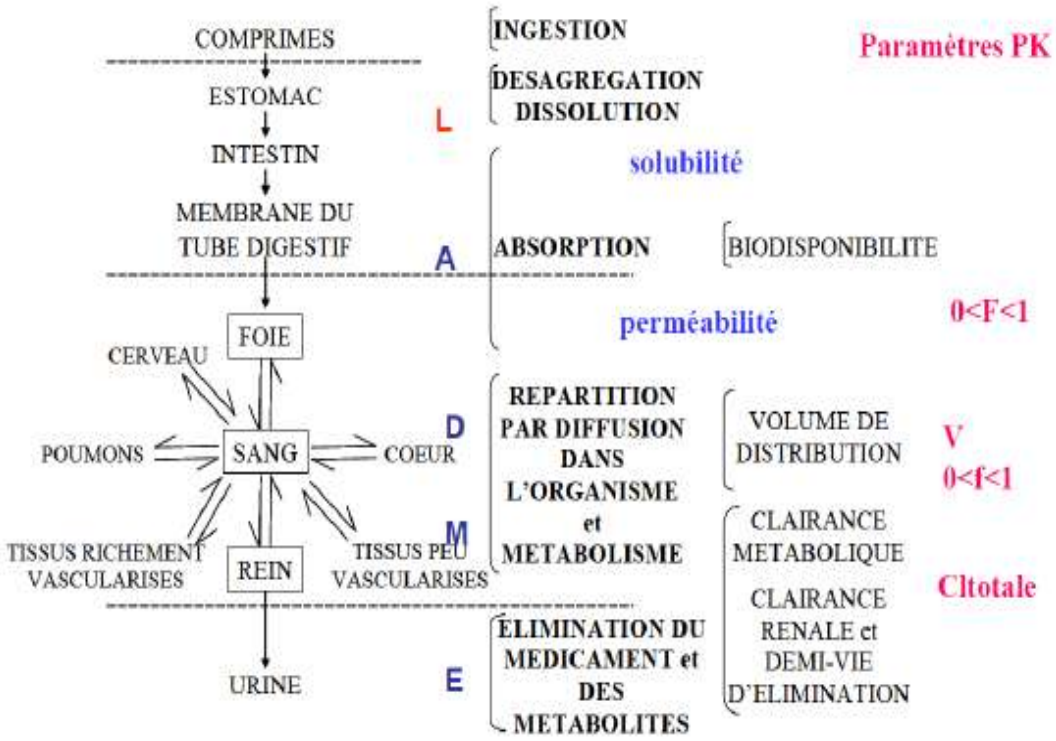



**Figure I.1.** Le trajet du médicament dans l'organisme(Cours 2016).

F : facteur de biodisponibilité

V : volume de dilution

*f:* taux de liaison aux protéines plasmatiques

Ces différentes phases se font à des vitesses plus ou moins rapides selon le médicament étudié et selon les individus traites par ce produit.

## I.2. Absorption d'un médicament

L'absorption est le processus par lequel le médicament inchangé passe de son site d'administration à la circulation générale (site de mesure). La voie d'administration du médicament influence cette première phase : la voie intra-veineuse est la voie de référence puisque par définition, à la différence des autres voies (orale par exemple) toute la dose administrée atteint la circulation générale(Rowe 2012).

### I.2.1. *Modalités d'absorption*

On abordera principalement la voie digestive parce que c'est une voie indirect c'est-à-dire c'est la voie ou les paramètres sont les plus variables et d'autre part parce que c'est la voie d'administration qui est la plus fréquente et qui présente un intérêt d'autant plus générale(Stanke-Labesque 2012).

L'administration d'un médicament par voie digestive nécessite une première étape qui est celle de la libération et de la dissolution du principe actif puis son passage dans la circulation sanguine c'est-à-dire sa résorption. Le médicament doit passer une barrière qui le sépare de la circulation générale (l'épithélium digestif) Parmi les différents mécanismes, 2 sont importants(Stanke-Labesque 2012) :

#### 1.2.1.1. *Diffusion passive*

La diffusion passive est gouvernée par la loi de Fick :

$$\frac{\mathrm{M}}{\mathrm{t}} = \mathrm{Pk} * \mathrm{A}(\mathrm{C1} - \mathrm{C2}) \qquad \text{(I.1)}$$

M/t = flux de médicament qui diffuse (unité de masse/temps)

Pk = coefficient de perméabilité (temps / cm2)

A = section de la surface de diffusion (cm2)

C1 et C2 = concentrations de médicament de part et d'autre de la membrane (unité de masse / par unité de volume).

Cette loi indique simplement que la diffusion passive se fait en suivant le gradient de concentration et jamais contre un gradient de concentration. C'est à dire que le médicament va du milieu le plus concentré vers le milieu le moins concentré.

Ce mécanisme n'est pas saturable et est non spécifique : il n'y a pas de compétition entre les molécules

**1.2.1.2. *Transport actif*** : On appelle transport actif le passage d'une substance à travers une membrane contre un gradient de concentration. Ce système de transport est capable de former un complexe avec la molécule à transporter, laformation de ce complexe se fait sur l'une des faces de

la membrane et sa dissociation sur l'autre, libérant ainsi la molécule transportée. Les principales caractéristiques de ce mécanisme sont(Rosenfeld et Loose 2007) :

- ✓ Il nécessite de l'énergie (par ex. de l'ATP).
- ✓ Il est spécifique d'une substance ou d'un type de substance, (par ex. le transport des acides aminés).
- ✓ Il permet l'entrée ou la sortie d'une molécule dans une cellule. Il est souvent bidirectionnel.
- ✓ C'est un mécanisme saturable.
- ✓ Il peut être inhibé et soumis à une compétition entre les molécules transportées (médicaments et substances endogènes, médicaments entre eux) ce qui entraîne une possibilité d'interactions médicamenteuses.

### 1.2.2. Les paramètres influents sur l'absorption(Lechat 2006)

#### 1.2.2.1. Les caractéristiques du médicament

- ✓ Physico-chimiques : pKa (la forme non ionisée d'un médicament est absorbée plus facilement)
- ✓ Hydro/lipo solubilité
- ✓ Taille et morphologie de la molécule
- ✓ La forme galénique (sirop, comprimé, gélule…) qui détermine la vitesse de dissolutiondu médicament…

#### 1.2.2.2. Les caractéristiques liées à l'individu

- ✓ Le pH digestif
- ✓ La vitesse de vidange gastrique et la mobilité intestinale
- ✓ L'alimentation : repas riche en graisses…
- ✓ La prise associée de médicament (pansements digestifs, modificateurs de vidange gastrique)
- ✓ L'âge.

### I.2.3. *Evaluation de l'absorption : la biodisponibilité*

La biodisponibilité se définit comme étant la fraction de la dose de médicament administré qui atteint la circulation générale et la vitesse à laquelle elle l'atteint(wikipedia 2016)

L'absorption digestive proprement dite, c'est-à-dire la quantité de principe actif atteignant la circulation systémique est difficile à mesurer puisque la circulation porte est d'accès peu aisé. L'approche de cette quantité disponible au niveau systémique se fait donc de manière indirecte à

partir de la quantité de médicament dans le plasma prélevé au niveau périphérique, c'est à dire après le foie(Toutain et Bousquet-Melou 2004).

La quantité de médicament qui atteint la circulation générale (ou systémique) est fonction de la quantité absorbée par l'épithélium digestif (et donc de la dose administrée) mais également, d'autres processus d'élimination pré-systémique(Loichot et Grima 2004) :

- ✓ dégradation dans la lumière intestinale,
- ✓ métabolisme au niveau des anthérocytes (cf. métabolisme),
- ✓ captage hépatique important au premier passage. Lorsque le médicament à une forte affinité pour l'hépatocyte et les enzymes hépatiques, une fraction de la dose absorbée est captée lors du premier passage, c'est à dire avant même d'atteindre la circulation générale. La quantité de médicament retrouvée dans la circulation systémique est alors diminuée. C'est l'effet de premier passage hépatique.

### I.2.3.1. *Le facteur quantitatif F*

Le facteur quantitatif (F) de la biodisponibilité ne peut être apprécié que par rapport à une forme de référence. On distingue ainsi :

- La biodisponibilité absolue: une forme extra-vasculaire est comparée à la forme de référence qui est le médicament administré par voie intraveineuse puisque par définition toute la dose atteint la circulation générale(Lechat 2006).
- La biodisponibilité relative où la forme de référence est administrée par une autre voie que la voie intra-veineuse. Cette forme de référence peut être administrée par la même voie que la forme à tester, mais il s'agit soit d'une autre forme galénique (solution aqueuse, suspension...)Soit d'une autre formulation d'une forme commercialisée depuis longtemps (cas des génériques)(Lechat 2006).

En général la quantification du facteur (F) de biodisponibilité s'effectue par comparaison des surfaces sous la courbe des concentrations en fonction du temps (SSC) après administration de chaque forme séparément. Celles-ci sont en effet proportionnelles à la quantité de médicament présent dans la circulation générale ( figure I.2)(Pharmaomédicale 2016).

F est obtenu selon

$$F = \frac{SSCp0}{SSCiv} \qquad (biodisponibilité\ absolue) \qquad \text{(I.2)}$$

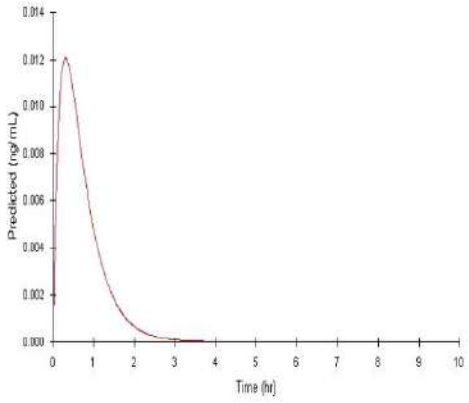



**Figure I.2.** L'allure générale de la Quantité de médicament présent dans la circulation général en fonction du temps(Lechat 2006)

### 1.2.3.2. *Interprétation de la biodisponibilité*

On voit selon l'équation (I.2) que si toute la dose administrée par voie orale est absorbée (comme en intra-veineux) la biodisponibilité absolue de ce produit sera 1.

Une biodisponibilité absolue de 0,5 pour un produit signifie que seule la moitié de la quantité administrée est retrouvée dans la circulation générale.

Ainsi, la dose contenue dans le comprimé ou la gélule ne reflète pas toujours la dose biodisponible

F est donc par définition compris entre 0 et 1(Kwan 1997).

### I.2.3.3. *Le facteur vitesse Tmax*

Le facteur vitesse est apprécié par la constante de vitesse d'absorption Ka ou plus facilement par la concentration maximale (Cmax) et le temps pour atteindre cette concentration (Tmax).

Au même titre que la quantité absorbée, la vitesse d'absorption d'un médicament est un paramètre significatif pour le délai d'action d'un principe actif.

La vitesse de passage est un paramètre prépondérant pour les médicaments destinés à une action rapide (antalgique par exemple) en prise unique ou de courte durée. Pour les traitements chroniques, où une imprégnation constante est recherchée, la notion de Tmax est moins déterminante(Geneviève 2007).

### I.2.3.4. *L'intérêt de l'étude de la biodisponibilité*

1**.** La biodisponibilité absolue est déterminée lors de l'étude d'un nouveau médicament.

La détermination de la biodisponibilité relative est utilisée pour comparer des formes galéniques ; elle est obligatoire pour tout changement de formulation (changement d'excipient...) et avant commercialisation d'un médicament« **générique**». Un exemple classique permet de comprendre la nécessité de telles études est la survenue en 1972 en Angleterre d'intoxications digitaliques chez plusieurs patients liées aune modification dans le procédé de fabrication de la digoxine (Lanoxine®) qui entraîna un doublement de sa biodisponibilité(Lechat 2006).

2. Il ne faut pas assimiler obligatoirement mauvaise biodisponibilité et faible efficacité.

En effet, la mauvaise biodisponibilité peut provenir d'un captage hépatique au 1$^{er}$passage. Il est possible que ce captage aboutisse à la transformation du médicament en métabolite pharmacologiquement actif. Dans ces conditions, malgré une faible biodisponibilité, le médicament administré par voie orale pourrait être aussiactif que par voie intraveineuse. C'est le cas du propranolol dont la biodisponibilité est de 30 % mais qui est métabolisé en 4-OH propranolol dont l'activité bloquante est comparable à celle du propranolol.

A l'inverse, le vérapamil (inhibiteur calcique) avec une biodisponibilité de 15 % est, à dose identique, 7 à 10 fois moins actif par voie orale que par voie intraveineuse: ses métabolites sont beaucoup moins actifs que le produit inchangé(Toutain et Bousquet-Melou 2004).

3. Par définition, les pro-drogues (précurseurs de médicament) ont une biodisponibilité nulle ou très faible puisqu'ils ne sont pas retrouvés dans la circulation générale : ils sont rapidement transformés en molécules responsables de l'activité(Lechat 2006).

4. Une faible biodisponibilité ne serait pas gênante en soi si elle était constante pour un même individu et entre les individus. Ceci n'est pas le cas dans la réalité. Plus la biodisponibilité d'un médicament est faible, plus ses variations auront d'effetsur son profil pharmacocinétique.

Exemple : la biodisponibilité d'un médicament varie au maximum de 5 % entre les individus. Si sa biodisponibilité est faible (10 % en moyenne par exemple) elle peut alors passer de 10 % à 15 % = 50 % d'augmentation relative. A l'inverse si sa biodisponibilité est forte (90 % en moyenne par exemple), une variation de 5 % aura beaucoup d'effet relatif(Lechat 2006).

## I.3. Distribution

Une fois la circulation sanguine atteinte, les médicaments vont se distribuer dans l'organisme. Les caractéristiques physico-chimiques du médicament conditionnent son affinité pour les différents tissus mais d'autres facteurs vont influencer la distribution. On résume donc sous le terme de distribution, le devenir du médicament au niveau sanguin puis sa diffusion au niveau tissulaire (Wikipedia 2016).

### 1.3.1. *Fixation aux protéines plasmatiques (Jacques 2006)*

Dans la circulation générale, le médicament peut se lier aux protéines plasmatiques, présentes en grande quantité, pour former des complexes. Il s'agit le plus souvent d'une liaison réversible et en équilibre tel que :

Ka

Médicament libre + Protéine libre ⇔ Complexe médicament-protéine

Forme libre forme liée

(1-f) *f*

Ka : constante d'association ou d'affinité

*f*: taux de liaison aux protéines plasmatiques

Les substances "acides faibles" se fixent à l'albumine.

Les substances "bases faibles" se fixent préférentiellement a l'AAG alpha-1-glycoprotéine acide.

Il y a une possibilité d'interactions médicamenteuses par déplacement, s'il ya une compétition pour les mêmes sites de liaison.

La forme libre est seule active et diffuse vers les tissus.

La forme liée est une forme de transport ou de stockage.

Différentes protéines plasmatiques et structures cellulaires sont impliquées :

- ✓ Albumine
- ✓ Alpha1 glycoprotéine acide (AAG)
- ✓ Lipoprotéines

- ✓ Gammaglobulines
- ✓ Cellules sanguines (érythrocytes, polynucléaires, lymphocytes, plaquettes)

La fixation aux protéines plasmatiques dépend beaucoup des caractéristiques acido-basiques du médicament.

### 1.3.2. *Interprétation de la défixation protéique*

Le seul pourcentage de fixation aux protéines est insuffisant pour comprendre les conséquences de la fixation sur la pharmacocinétique d'un médicament.

En effet il existe un équilibre entre le plasma, les tissus et les voies d'élimination.

Lorsqu'il est défixé un médicament fortement fixé aux protéines plasmatiques est soit éliminé soit distribué vers les tissus. Si les voies de métabolisme sont efficaces, cette défixation n'aura le plus souvent aucune conséquence.

En pratique, la fixation protéique n'est à considérer que si elle est élevée (> 90 %) et si le médicament a une marge (ou un index) thérapeutique étroite (concentration toxique proche de la concentration efficace)(Médecine et vie 2012).

### 1.3.3. *Diffusion tissulaire*

Généralement, la distribution se fait dans l'espace extracellulaire (volume plasmatique + volume interstitiel) et peut aussi comprendre le volume cellulaire. Pour diffuser les médicaments doivent passer les membranes tissulaires. Dans certains tissus (foie…), la paroi vasculaire est composée de capillaires discontinus permettant une diffusion facile du médicament. A l'opposé dans d'autres organes (cerveau et barrière hémato-encéphalique…) la paroi vasculaire est composée de capillaires continus difficilement franchissables.Les mécanismes du passage transmembranaire du médicament sont identiques à ceux pour l'absorption digestive(pharmacorama 2016).

La diffusion tissulaire est donc dépendante de(Lechat 2006) :

- ✓ Caractéristiques physico-chimiques du médicament (lipophilie)
- ✓ Capacité du médicament à franchir les parois vasculaires et cellulaires
- ✓ La fixation protéique (sanguine et tissulaire)
- ✓ Le débit sanguin tissulaire (très élevé pour le foie et le rein, faible pour l'os et la peau…)

### 1.3.4. *Interprétation de la diffusion tissulaire*

Les molécules fortement fixées aux tissus sont en quelque sorte « aspirées » du compartiment sanguin et les concentrations totales circulantes décroissent fortement, sans entraîner de modification du pourcentage de fixation aux protéines plasmatiques s'il n'y a pas dissociation. A l'inverse, les molécules fortement liées aux protéines plasmatiques se trouveront essentiellement dans l'espace vasculaire (les tests d'exploration pour mesurer les différents volumes liquidiens à l'aide de colorants liés aux protéines sont fondés sur cette propriété) (Lechat 2006).

### *1.3.5. Volume apparent de distribution*

Il est difficile de quantifier la distribution d'un médicament compte-tenu de l'impossibilité de mesurer les concentrations tissulaires, seul le secteur plasmatique étant facilement accessible.

La relation entre la quantité totale de médicament présente dans l'organisme et la concentration observée au niveau plasmatique s'exprime par le Volume de distribution. Connaissant la dose administrée(*D*) et la concentration plasmatique (*C*) extrapolée à l'origine *Co* après injection intraveineuse, il est en effet simple de calculer ce Volume de distribution(Pharmaomédicale 2016)

$$V = \frac{dose}{C0} \qquad \text{(I.3)}$$

Cette méthode est souvent fausse, à cause de l'imprécision régnant sur l'estimation de cette concentration à l'origine, du fait des phénomènes de distribution qui rendent hasardeuse toute extrapolation.

Il est de loin préférable d'utiliser la relation qui existe entre la clairance et la constante d'élimination *k*.

$$V = \frac{Cl}{k} = \frac{dose}{aire\ sous\ la\ courbe * k} \qquad \text{(I.4)}$$

La valeur obtenue par ce calcul est en fait le volume apparent de distribution qui serait atteint en supposant une répartition homogène de la molécule dans un volume tel que la concentration de médicament serait partout identique à celle du plasma. Si une substance est fortement fixée au niveau tissulaire, la concentration plasmatique sera faible et le volume de distribution grand et réciproquement. Le volume de distribution n'est cependant pas une représentation anatomique de la répartition de la molécule Des médicaments comme les antidépresseurs imipraminiques peuvent atteindre des valeurs supérieures à mille litres (au lieu des 40 litres d'eau corporel de l'homme

standard : 3litres plasma + 12 litres de liquide interstitiel + 25 litres de liquide intra-cellulaire) reflétant les très faibles concentrations plasmatiques atteintes après des doses thérapeutiques(UGent 2012).

Une des conséquences cliniques de cette caractéristique pharmacocinétique est qu'en cas d'intoxication par surdosage, il sera vain d'entreprendre une épuration extra-rénale par dialyse pour toutes les molécules à grand volume de distribution : le dialyseur n'ayant accès qu'à des quantités circulantes très faibles, l'efficacité de l'épuration sera insignifiante.

### 1.3.6. *Facteurs modifiant la distribution(Labrecque et Sirois-Labrecque 2003)*

**Volume liquidiens de l'organisme**

- ✓ Age (nourrisson…)
- ✓ Déshydratation

**Rapport masse maigre/tissu adipeux**

- ✓ Obésité
- ✓ Age

**Hémodynamique**

- ✓ Etat de choc
- ✓ Insuffisance cardiaque chronique

**Modifications des protéines plasmatiques**

**Diminution de la concentration d'albumine**

- ✓ Grossesse
- ✓ Syndrome néphrotique
- ✓ Dénutrition
- ✓ Grands brûlés
- ✓ Cirrhose

**Diminution AAG**

- ✓ Grossesse
- ✓ Contraceptifs oraux

- ✓ Age : nouveau-né
- ✓ Cirrhose

**Augmentation de la concentration AAG**

- ✓ Etats inflammatoires
- ✓ Affections rhumatologiques
- ✓ Etats infectieux sévères

## I.4. Métabolisme et élimination des médicaments

L'élimination des médicaments de l'organisme résulte de l'addition de plusieurs processus. Elle comprend la capacité métabolique de différents organes, en premier lieu le foie et l'excrétion sous toutes ses formes, en particulier rénale (urine) mais aussi hépatique (bile) (Lechat 2006).

### I.4.1. *Biotransformations*

#### I.4.1.1. *Définition*

L'analyse du métabolisme d'un principe actif est avant tout une approche descriptive des diverses voies métaboliques et de leur importance relative, ainsi que des conséquences sur l'élimination du médicament.

Le terme de métabolisme fait référence à la transformation, par une réaction enzymatique d'un médicamenten un ou plusieurs autres composés actifs ou inactifs au plan pharmacologique. De nombreux tissus peuvent réaliser cette transformation (peau, poumon, rein, intestin...). Néanmoins le principal site de biotransformation est situé au niveau hépatique, dans les enzymes des microsomes.Ceci est expliqué par le flux sanguin très important du foie, organe épurateur, par rapport aux autres organes : il reçoit environ 1,5 litres de sang par minute (1,2 l par la veine porte et 0,3 l par l'artère hépatique). Les hépatocytes contiennent un grand nombre d'enzymes impliquées dans la transformation des médicaments, en particulier les réactions d'oxydoréduction, les hydroxylations ou la rupture oxydative des liaisons N-C et O-C. L'élément fondamental de ce système enzymatique est le cytochrome P450 comprenant de nombreuses isoenzymes (pharmacocinétique 2005).

Schématiquement on distingue 2 phases de métabolisme selon les processus de transformation induits par ces enzymes : les réactions de phase I et celles de phase II.

### I.4.1.2. *Réactions de phase I*

• Les réactions d'oxydation sont majoritairement localisées dans les microsomes hépatiques.

Elles consomment du NADPH (nicotinamide phosphate réduit), de l'oxygène moléculaire et passent par les cytochromes P450.

• Les réactions de réduction sont beaucoup moins fréquentes et moins bien explorées. La réduction n'intervient pas exclusivement au niveau hépatique mais également dans l'intestin via la flore bactérienne.

• L'hydrolyse enfin est une voie métabolique banale, qui intervient dans le foie, dans différents tissus et même dans le plasma. Les enzymes de type des estérases sont le plus souvent non spécifiques. La réaction d'hydrolyse par clivage d'un ester ou d'un amide, est chez l'homme, très rapide.

L'oxydation, la réduction et l'hydrolyse sont des biotransformations regroupées sous le terme de

« Métabolisme de phase I » qui conduit à des dérivés dont les groupements fonctionnels sont le plus souvent des hydroxyles (-OH), des amines (-NH2) ou des carboxyles (-COOH)(Pharmacie.ma 2006).

### I.4.1.3. *Réactions de phase II*

Les groupements fonctionnels issus des réactions de phase I peuvent être ensuite conjugués. C'est la réaction de phase II.

Les mécanismes de conjugaison chez l'homme font généralement appel à l'acide glucuronique, au glycocolle, au sulfate ou à l'acétyl.

Glucuroconjugaison. La conjugaison avec l'acide glucuronique est la plus fréquente des conjugaisons.Elle est catalysée par le système enzymatique de la glucuronyltransférase et concerne les molécules possédant un groupement hydroxylé, carboxylé ou aminé. Les glucuronides sont très hydrosolubles ce qui explique la facilité avec laquelle ils sont éliminés dans l'urine et la bile. Dans quelques cas, les esters sont instables et après hydrolyse dans l'urine ou le plasma redonnent la molécule mère(Loichot et Grima 2004).

### 1.4.1.4. *comprendre le métabolisme des médicaments (Lechat 2006)*

1. Lorsqu'un médicament est métabolisé, il l'est rarement de façon unique et plusieurs voies métaboliques sont possibles. Tous les métabolites ne sont d'ailleurs pas toujours identifiés (en particulier à cause des problèmes analytiques qui se posent pour les isoler, les caractériser et définir leur structure).

2. Il existe une certaine spécificité pour certains substrats : les différents cytochromes ont en fonction de leur structure protéique une affinité différente pour les divers substrats.

3. Certains substrats modifient l'activité des enzymes responsables des biotransformations

(Augmentation = inducteur ; diminution = inhibiteur)

4. Certaines enzymes des voies de métabolisme du médicament sont soumises à des polymorphismes génétiques qui peuvent modifier leur activité métabolique. On distinguera alors des métaboliseurs lents, intermédiaires, rapides et même ultra-rapides. Ce facteur intervient dans la variabilité entre les individus de réponse à un médicament.

5. Les médicaments qui ont une forte affinité pour les enzymes hépatiques ont après administration orale une faible biodisponibilité due à l'effet de premier passage hépatique.

### I.4.2. *Elimination*

#### I.4.2.1. *Elimination hépatique*

Outre ses capacités métaboliques, le foie participe à l'excrétion des médicaments hors de l'organisme par le biais du système biliaire.

Après excrétion dans la bile, le médicament se retrouve dans la lumière intestinale où il peut être réabsorbé : c'est le cycle entéro-hépatique(Guergouri).

#### I.4.2.2. *Elimination rénale*

La plupart des molécules sont éliminées dans les urines, soit sous forme inchangée, soit sous forme de produits de dégradation. Le plus souvent les médicaments ou leurs métabolites ont une masse moléculaire bien inférieure à 5000 et sont de ce fait filtrés par le glomérule. Seule la partie non fixée est filtrée(pharmacorama 2016).

La réabsorption tubulaire intervient tout au long du néphron. Il s'agit le plus souvent d'un processus passif qui est influencé par le degré d'ionisation du médicament : seule la fraction non

ionisée au Ph urinaire est réabsorbée. Cette propriété est utilisée dans certains surdosages pour accélérer l'élimination du médicament en alcalinisant les urines pour bloquer la réabsorption.

Une sécrétion active est également observée pour quelques molécules, entre autres des cations ou anions qui sont sécrétés dans la lumière du tubule par des systèmes de transport spécifiques, consommant de l'énergie et à capacité saturable. On peut donc observer des phénomènes de compétition(Lechat 2006).

#### I.4.2.3. *Autres voies d'excrétion*

Les autres voies (salivaires, pulmonaire…) sont usuellement négligeables par rapport aux voies rénale et hépatique.

Néanmoins on soulignera l'importance de la voie lactée pouvant donner des risques d'intoxications du nourrisson lors de l'allaitement(Pharmaomédicale 2016).

### I.4.3. *Quantification du métabolisme et de l'élimination*

#### I.4.3.1. *Notion de clairance*

La capacité globale de l'organisme à éliminer une molécule est la clairance, définie comme le volume de plasma totalement épuré par unité de temps ; elle est ainsi habituellement exprimée comme un débit en ml/min.

La clairance totale est égale à la somme des clairances de chaque organe susceptible d'intervenir dans l'élimination du médicament : clairance rénale, hépatique, intestinale, pulmonaire, etc.

La notion de clairance recouvre deux aspects complémentaires l'un de l'autre :

- la biotransformation du composé parent en métabolites dans les différents organes (foie, intestin, peau, etc.),
- l'excrétion du composé inchangé par les voies classiques (rein, voies biliaires, sueur,larmes, etc.).

On peut aussi définir la clairance d'un organe qui traduit la capacité de cet organe à extraire un médicament d'un volume sanguin par unité de temps. On s'intéresse plus particulièrement à la clairance hépatique et la clairance rénale(Loichot et Grima 2004).

#### I.4.3.2. *Notion de coefficient d'extraction*

Pour un organe on a :

$$Clairance = Debit\ sanguin(Q) * Coefficient\ d'extractionde\ l'organe \quad (I.5)$$

$$E = \frac{\text{concentration sanguine arterielle (Ca)} - \text{concentration sanguine veineuse (Cv)}}{\text{concentration sanguine arterielle(Ca)}} \quad (I.6)$$

La capacité d'un organe à éliminer un médicament est ainsi exprimée par la fraction du flux sanguin le traversant qui est complètement épurée du médicament par unité de temps. Cette fraction est définie comme le coefficient d'extraction *E*.

Les médicaments peuvent ainsi être divisés en plusieurs groupes selon leur comportement au niveau de l'organe. On définit classiquement les médicaments en :

- ✓ fortement extraits si $E > 0,7$
- ✓ moyennement extraits quand $0,3 < E < 0,7$
- ✓ faiblement extraits quand $E < 0,3$

Lorsque la molécule est totalement extraite du sang lors de son passage à travers l'organe, la clairance($Cl$) du médicament est égale au débit sanguin à travers ce même organe :

$$E = \frac{\text{Ca}-0}{Ca} = 1 \quad (1.7)$$

Or$Cl = Q * E$donc ici $Cl = Q$

**Exemple** : D'une façon simple, la clairance hépatique indique le volume virtuel de sang perfusant le foie qui est totalement débarrassé du médicament par unité de temps. Le flux sanguin hépatique est normalement de 1,5 l/mn chez l'adulte de 70 kg. Si le médicament est essentiellement éliminé sous forme de métabolites par le foie et que son coefficient d'extraction est de 0,50, la clairance hépatique de ce produit sera donc de 0,75 l/mn pour ce sujet. A contrario, un coefficient d'extraction proche de zéro signifie que la clairance hépatique est très faible et donc qu'il ne participe pas à l'élimination de ce médicament(Lechat 2006).

### I.4.3.3. *Clairance hépatique*

La clairance hépatique se décompose en deux :

- clairance métabolique
- clairance biliaire

**Clairance métabolique**

Elle dépend d'une part de la clairance intrinsèque qui est la capacité du ou des systèmes enzymatiques hépatiques à métaboliser le médicament indépendamment des autres facteurs (débit sanguin par exemple). Elle traduit la fonction brute du foie.

Elle dépend d'autre part de la fraction libre plasmatique du médicament qui est fonction du degré de fixation protéique.On peut alors distinguer 2 situations :

• Lorsque le foie extrait fortement un médicament ($Eh > 0,7$), l'élimination métabolique est dite débit-dépendent. En effet $Eh$ tend vers 1 alors $Clh = Qh \times Eh = Qh$. Tous les facteurs qui influenceront le débit hépatique modifieront donc l'élimination de ce médicament.

• Lorsque le foie extrait faiblement un médicament ($Eh < 0,3$), l'élimination dépend de la clairance métabolique (fraction libre du médicament et clairance intrinsèque)(Moulin et Coquerel 2002)

**Clairance biliaire**

C'est la capacité du système biliaire à éliminer le médicament. Ce système élimine principalement les molécules de forte masse moléculaire. La sécrétion biliaire est le plus souvent active par le biais de transporteurs(FOURCADE *et al.* 2014).

### 1.4.3.4. Facteurs influençant la clairance hépatique

- ✓ Modification du débit sanguin hépatique
- ✓ Modification de la clairance intrinsèque : Induction & Inhibition enzymatique ; polymorphismes génétiques ; insuffisance hépatocellulaire ; âge
- ✓ Modification de la fraction libre
- ✓ Modification de la clairance biliaire : cholestase intra et extra hépatique(Lechat 2006)

### I.4.3.5. *Clairance rénale(Loichot et Grima 2004)*

On a par définition

$$Clr = Qr * Er \tag{1.8}$$

Mais aussi :

$$Cl \text{ rénale} = Cl \text{ filtration} + Cl \text{ sécrétion} - Cl \text{ réabsorption} \tag{1.9}$$

**1.4.3.6. Facteurs influençant la clairance rénale**(Lechat 2006)

- ✓ Modification du débit de filtration glomérulaire : Insuffisance rénale, insuffisance cardiaque, âge
- ✓ Modification de la sécrétion tubulaire : Insuffisance rénale ; insuffisance cardiaque ;âge ; interaction médicamenteuse
- ✓ Modification de la réabsorption tubulaire : pH, débit fraction filtrée, âge
- ✓ Modification de la fraction libre.

**I.4.3.7.** ***Calcul de la clairance(Pharmaomédicale 2016)***

La clairance sanguine totale (ou systémique) est généralement calculée à partir des données sanguinesdu médicament obtenues après injection intraveineuse (bolus) selon la formule :

$$Cl = \frac{\text{dose}}{\text{aire sous la courbe}} \tag{1.10}$$

Si l'on administre le médicament par voie orale, il faut dans le calcul de la clairance tenir comptede la fraction qui atteint réellement la circulation, ce que l'on exprime par :

$$Cl = \frac{\text{F x dose orale}}{\text{aire sous la courbe aprés voie orale}} \tag{1.11}$$

Où *F* est la biodisponibilité

**Remarque**

Si l'on ne pondère pas par la fraction *F*, cela revient à surestimer la clairance ; à l'extrême si la molécule n'est que peu résorbée au niveau du tractus gastro-intestinal, les concentrations sanguines seront très faibles, la surface sous la courbe tendra vers zéro et la clairance sera très élevée, donnant l'illusion d'une grande capacité d'élimination d'une dose de médicament alors qu'en fait ce médicament n'a jamais atteint la circulation.

La clairance rénale est aussi facilement calculable puisque l'on peut doser dans les urines la quantité de médicament éliminé.

Par contre, la clairance hépatique est difficile à quantifier (la clairance biliaire nécessiterait de réaliser des prélèvements biliaires, la clairance intrinsèque est difficile à mesurer…).

On peut alors la déduire en considérant :

Clairance totale = Clairance rénale + Clairance non rénale (= hépatique notamment)(Lechat 2006)

### I.4.4. *Demi-vie d'un médicament*

Un paramètre synthétique est largement utilisé pour exprimer l'élimination d'un médicament de l'organisme : la **demi-vie**. La *t*1/2 correspond au temps nécessaire pour passer d'une concentration plasmatique à sa moitié, quel que soit le niveau de cette concentration. Le facteur de proportionnalité, entre la dose administrée et la concentration plasmatique, est le volume de distribution. Une autre manière d'exprimer la même notion est de considérer la *t*1/2 comme le temps mis pour diminuer de moitié la quantité totale de molécule contenue dans l'organisme quelle que soit la dose(Wikipedia 2016).

Généralement la demi-vie est calculée à partir des concentrations plasmatiques mesurées durant laphase d'élimination.

$$t1/2 = \frac{0.693}{\mathrm{Ke}} \quad et \quad Ke = \frac{\log(\mathrm{C1}/\mathrm{C2})}{t2-t1} \tag{I.12}$$

*C*1 est la concentration au temps 1 et *C*2 la concentration au temps 2.

L'élimination de la molécule inchangée ne peut se faire qu'à partir des organes d'élimination, en contact avec le sang ou le plasma. Comme par ailleurs la quantité de médicament dans le sang ou le plasma à tout moment dépend du volume de distribution, la vitesse de disparition du médicament dépendra à la fois de la clairance et du volume de distribution. D'où :

$$t1/2 = \frac{0.693 \times \mathrm{Vd}}{\mathrm{Cl}} \tag{I.13}$$

Pour une clairance élevée, les molécules à petit volume de distribution, donc à concentrations plasmatiques élevées, seront éliminées avec une demi-vie courte et réciproquement. La *t*1/2 n'est ainsi qu'un résumé de deux paramètres physiologiques caractéristiques de chaque molécule, la clairance et le volume de distribution(Lechat 2006).

La demi-vie n'est donc pas le reflet unique de l'élimination du médicament mais un critère composite lié à la distribution et à l'élimination de celui-ci.

Ainsi, des situations cliniques existent où la *t*1/2 reste constante alors que l'élimination du médicament est altérée. Mais *Cl* et *Vd* ont varié de la même proportion dans le même sens (c'est le cas chez l'insuffisant rénal pour certains médicaments).

Les études pharmacocinétiques chez les malades qui n'utilisent que la demi-vie comme l'indicateur de l'élimination se trouvent parfois faussées car elles négligent le fait qu'une

pathologie donnée puisse modifier la clairance et/ou le volume de distribution : si l'on est dans le cas où *Vd* et *Cl* sont diminués, tous les deux de la même proportion, la *t*1/2 calculée selon l'équation restera inchangée mais les concentrations plasmatiques du médicament auront augmenté et ce de façon potentiellement importante. Les risques de toxicité peuvent augmenter en même temps(Lafaurie *et al.* 2002)

En pratique, la *t*1/2 intervient dans la réflexion pour établir le rythme posologique. Elle permet d'estimer le temps mis pour atteindre le plateau d'équilibre (5*t*1/2 dans tous les cas)et la fluctuation de concentrations entre les prises. Lorsqu'un médicament est administré en chronique, le profil des concentrations dépendra du rapport entre la *t*1/2 et l'intervalle entre les prises(Brion 2003).

Lorsque la dose administrée est totalement éliminée avant la dose suivante, le profil de concentrations sera une succession de pics identiques à une prise unique. Si au contraire, une prise médicamenteuse intervient alors que la dose précédente n'est pas éliminée, cette nouvelle dose vient s'ajouter au reste présent dans l'organisme : il y a accumulation jusqu'à l'obtention du plateau d'équilibre, en 4 à 5 *t*1/2 instant où la quantité apportée par chaque prise compense la quantité éliminée entre deux prises. Plus l'intervalle entre deux administrations est petit en comparaison de la *t*1/2, plus le reliquat auquel vient s'ajouter la nouvelle dose est grand et plus la molécule s'accumule dans l'organisme(Lechat 2006).

# *Chapitre II*

# *Les réseaux de neurones*

## II.1. Introduction

Les réseaux de neurones connaissent depuis quelques années un succès croissant dans divers domaines des sciences de l'ingénieur, celui du génie des procédés n'échappe pas à cette règle. Dans ce chapitre, nous nous intéressons à l'application des réseaux de neurones aux bioprocédés.

Dans un premier temps, nous rappelons les définitions et notations de bases relatives aux réseaux de neurones. Nous poursuivons en exposant les types et la méthodologie d'apprentissage.

Nous présentons aussi, d'une manière générale, les étapes de conception d'un réseau de neurones : le choix des entrées et des sorties, l'élaboration de la base de données et de la structure du réseau…etc.

Nous terminons ce chapitre par la présentation de diverses applications qui sont destinées à montrer que dans le domaine des bioprocédés, les réseaux de neurones sont susceptibles d'apporter des solutions efficaces et élégantes.

## II.2. Historique

En 1943, Mac Culloch et Pitts ont proposé le premier modèle de neurone biologique (Davalo et Naïm 1989). Ce dernier, appelé aussi neurone à seuil, a été inspiré des récentes découvertes en biologie. Ce sont des neurones logiques (0 ou 1). Ces deux physiciens ont montré que les neurones formels peuvent réaliser des fonctions logiques.

En 1949, le psychologue Donald Hebb a introduit, dans son ouvrage "The Organization of Behavior", le terme connexionisme pour parler des modèles massivement parallèles et connectés. Il a proposé de nombreuses règles de mise à jour des poids dont la plus connue à cette époque est la "règle de Hebb" (Hebb 1949, Davalo et Naïm 1989).

En 1958, le psychologue Frank Rosenblatt a développé le modèle du perceptron (Rosenblatt 1958). Il s'agit d'un réseau de neurones, capable d'apprendre à différencier des formes simples et à calculer certaines fonctions logiques. Il est inspiré du système visuel. Il a réussi à l'appliquer dans la reconnaissance des formes. Au début des années 60, les travaux de Rosenblatt ont vécu un grand intérêt dans le milieu scientifique. Mais en 1969, deux scientifiques américains appelés Minsky et Papert (Minsky et Papert 1969, Davalo et Naïm 1989) ont publié un livre dans lequel ont démontré

les limites du perceptron proposé par Rosenblatt. En particulier, son incapacité réside à la résolution des problèmes non linéaires.

Les travaux se sont ralentis considérablement jusqu'aux années 80. En 1982, Hopfield a fini par démontrer l'intérêt des réseaux entièrement connectés (Hopfield 1982, Davalo et Naïm 1989). Parallèlement, Werbos a conçu un mécanisme d'apprentissage pour les réseaux multicouches de type perceptron : la rétro-propagation (en anglais back-propagation). Cet algorithme, qui permet de propager l'erreur vers les couches cachées, sera popularisé en 1986 dans un livre "Parallel Distributed Processing" par Rumelhart et McClelland (Rumelhart et McClelland 1986).

En 1989 Moody et Darken (Moody et Darken 1989) exploitent quelques résultats de l'interpolation multi variables pour proposer le Réseau à Fonctions de base Radiales (RFR), connu sous l'appellation anglophone "Radial Basis Function network" (RBF). Ce type de réseau se distingue des autres types de réseaux de neurones par sa représentation locale (Talbi 2011).

Depuis ces travaux, les applications des réseaux de neurones n'ont cessé de croître. Les applications en sont multiples et variées.

## II.3. Notions de base sur les réseaux de neurones

### II.3.1. Du neurone biologique au modèle mathématique

#### *II.3.1.1. Le modèle neurophysiologique*

Le cerveau humain est composé de cellules distinctes appelées neurones formant un ensemble dense d'environ 10 à 100 milliards d'unités interconnectées. Tel qu'illustré sur la figure II.1, le neurone est une cellule composée d'un corps cellulaire et d'un noyau. Le corps cellulaire se ramifie pour former ce que l'on nomme les dendrites. C'est par les dendrites que l'information est acheminée de l'extérieur vers le soma (corps du neurone).

L'information est traitée alors par le corps cellulaire. Si le potentiel d'action dépasse un certain seuil, le corps cellulaire répond par un stimulus. Le signal transmis par le neurone est cheminé ensuite le long de l'axone (unique) pour être transmis aux autres neurones. La transmission entre deux neurones n'est pas directe. En fait, il existe un espace intercellulaire de quelques dizaines d'Angströms entre l'axone du neurone afférent et les dendrites du neurone efférent. La jonction entre deux neurones, comme schématisé sur la figure II.2, s'appelle synapse (Tzanakou 2000).

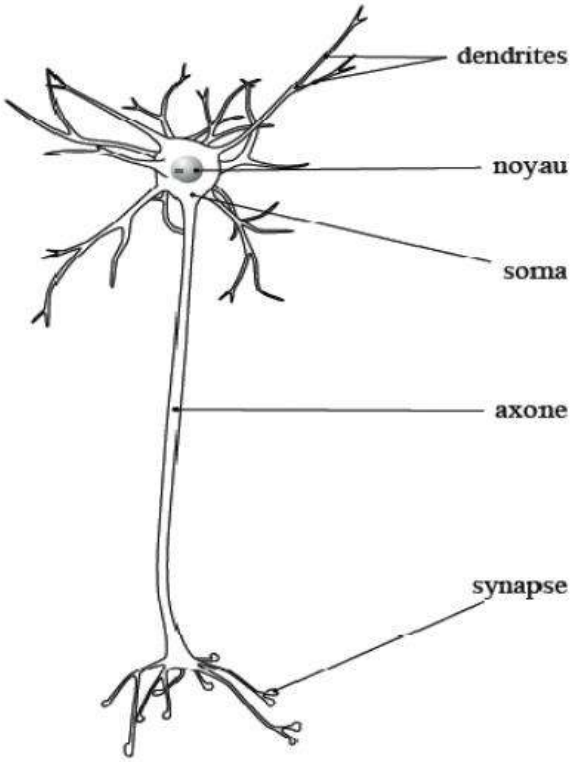



**Figure II.1.** Schéma d'un neurone (Talbi 2011).

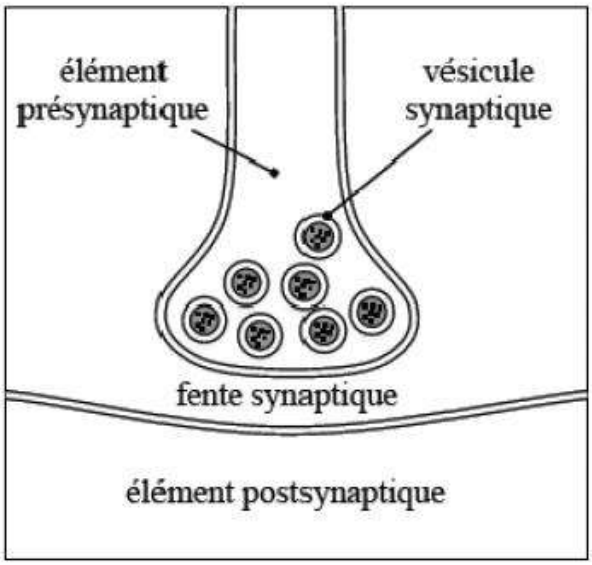



**Figure II.2.** Schéma d'une synapse (Talbi 2011).

### *II.3.1.2 Le modèle mathématique*

Par analogie avec le neurone biologique, chaque neurone artificiel est un processeur élémentaire. Il reçoit un nombre variable d'entrées en provenance de neurones en amont. À chacune de ces entrées est associé un poids *w*, abréviation du mot poids (weight en anglais), représentatif de la force de la connexion. Chaque processeur élémentaire est doté d'une sortie unique qui se ramifie ensuite pour alimenter un nombre variable de neurones en aval tel qu'illustré sur la figure II.3 (Haykin 1999). À chaque connexion est associé un poids.

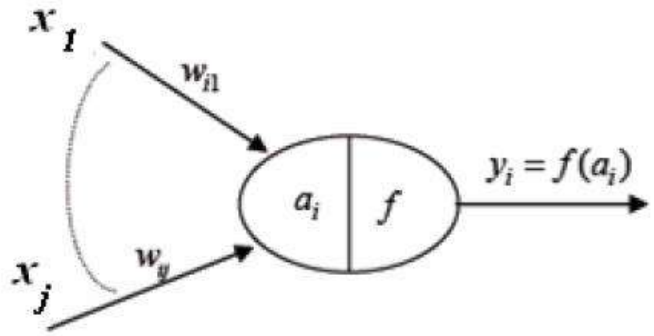



**Figure II.3.** Neurone formel (Talbi 2011).

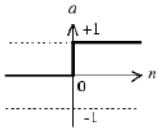
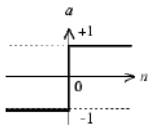
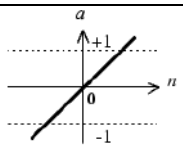
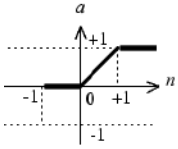
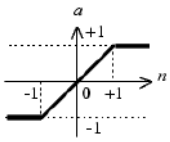
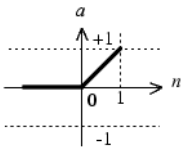
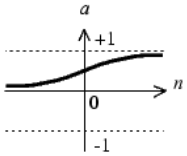
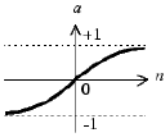
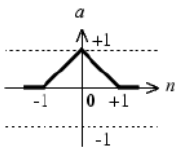
Le comportement du neurone artificiel se compose de deux phases (Haykin 1999) :

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

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

2. Dans la deuxième phase, une fonction de transfert calcule la valeur de l'état interne du neurone. C'est cette valeur qui sera transmise aux neurones en aval. Plusieurs possibilités existent pour spécifier la nature de la fonction d'activation du modèle. Les fonctions mathématiques les plus souvent utilisées sont regroupées dans le tableau 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]. Ainsi, le neurone peut être défini mathématiquement comme étant une fonction algébrique, a priori non linéaire (suivant la fonction d'activation *f*) et bornée, avec des entrées *x* et paramétrée par les poids $w_{ij}$.

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

| **Nom de la fonction** | **Relation entrée / sortie** | **Graphe de la fonction** | **Fonction MATLAB® correspondante** |
|---|---|---|---|
| Seuil | $a = 0$ si $n < 0$<br>$a = 1$ si $n \geq 0$ |  | hardlim |
| Seuil symétrique | $a = -1$ si $n < 0$<br>$a = 1$ si $n \geq 0$ |  | hardlims |
| Linéaire | $a = n$ |  | purelin |
| Linéaire saturée | $a = 0$ si $n < 0$<br>$a = n$ si $0 \leq n \leq 1$<br>$a = 1$ si $n > 1$ |  | satlin |
| Linéaire saturée symétrique | $a = -1$ si $n < -1$<br>$a = n$ si $-1 \leq n \leq 1$<br>$a = 1$ si $n > 1$ |  | satlins |
| Linéaire positive | $a = 0$ si $n < 0$<br>$a = n$ si $n \geq 0$ |  | 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 | $a = 1 - \lvert n \rvert,\ -1 \leq n \leq +1$<br>$a = 0,\ autrement$ |  | tribas |

### II.3.2. Architecture des réseaux de neurones

On distingue deux grandes familles de réseaux de neurones : les réseaux bouclés et les réseaux non bouclés (Figure II.4).

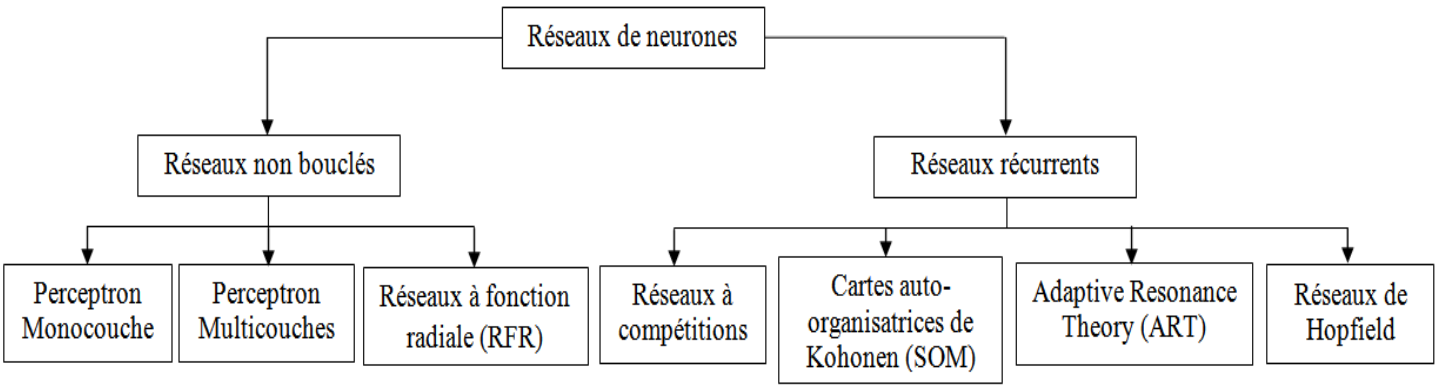



**Figure II.4.** Réseaux de neurones les plus utilisés (Talbi 2011).

#### *II.3.2.1. Réseau de neurones non bouclé (en anglais feedforward)*

Un réseau de neurones non bouclé est donc représenté graphiquement par un ensemble de neurones connectés entre eux, l'information circulant des entrées vers les sorties sans "retour en arrière" (Bishop 1995). On constate que le graphe d'un réseau non bouclé est acyclique. En effet, si on se déplace dans ce type de réseau à partir d'un neurone quelconque en suivant les connexions, on ne peut pas revenir au neurone de départ. Les neurones qui effectuent le dernier calcul de la composition de fonctions sont les neurones de sortie. Ceux qui effectuent des calculs intermédiaires sont les neurones cachés (Figure II.5).

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

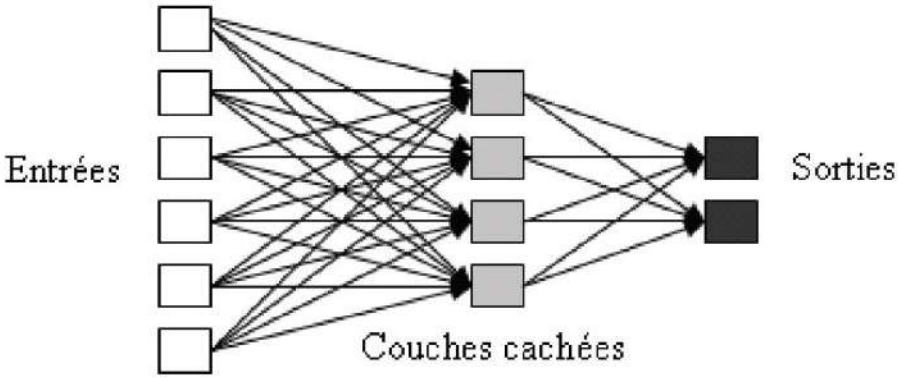



**Figure II.5.** Structure d'un réseau de neurones non bouclé (Chamekh 2007).

### *II.3.2.2. Réseau de neurones bouclé (ou récurrent)*

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

La figure II.6 représente un exemple de réseau de neurones bouclé. Les chiffres dans les carrés indiquent le retard attaché à chaque connexion, exprimé en multiple de l'unité de temps. Ce réseau contient un cycle, qui part du neurone 3 et revient à celui-ci en passant par le neurone 4. La connexion de 4 vers 3 ayant un retard d'une unité de temps.

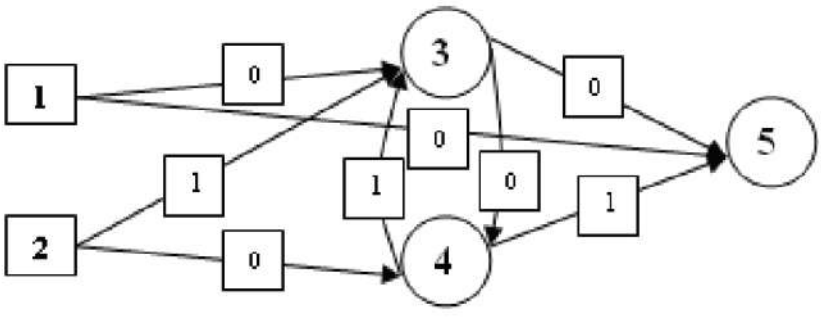



**Figure II.6.** Un réseau de neurones bouclé à deux entrées (Chamekh 2007).

### II.3.3. 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 (Hassoun 1995).

#### *II.3.3.1. 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 (Hassoun 1995).

#### *II.3.3.2. 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 (Hassoun 1995).

## II.4. Les réseaux de neurones les plus utilisés

La Figure II.4 représente les réseaux de neurones les plus utilisés dans la littérature. Dans ce qui suit, nous allons faire une description de chacun d'eux.

### II.4.1. Le 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).

### II.4.2. 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.

### II.4.3. Les 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).

### II.4.4. Les 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).

### II.4.5. Les réseaux de Hopfield

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

### II.4.6. Les réseaux ART

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

## II.5. 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.5.1. Détermination des entrées/sorties du réseau de neurones

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

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

### II.5.2. Construction de la base de données et choix des échantillons

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

Cette étape est cruciale et aide à déterminer le type de réseau le plus approprié pour résoudre un problème donné. En particulier, la façon dont se présente l'échantillon conditionne le type de réseau, le nombre de cellules d'entrée et de sortie et la façon dont il faudra mener l'apprentissage, le test et la validation (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.5.3. Définition de la structure du réseau

La détermination de l'architecture optimale est restée pendant longtemps un problème ouvert. Il existe actuellement diverses méthodes empiriques mettant notamment en jeu des tests statistiques pour une vaste classe de réseaux (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.

#### *II.5.3.1. Nombre de couches cachées*

Mis à part les couches d'entrée et de sortie, il faut décider du nombre de couches intermédiaires ou cachées. Sans couche cachée, le réseau n'offre que de faibles possibilités d'adaptation. Néanmoins, il a été démontré qu'un perceptron multicouches avec une seule couche cachée pourvue d'un nombre suffisant de neurones, peut approximer n'importe quelle fonction avec la précision souhaitée (Hornik 1991).

#### *II.5.3.2. Nombre de neurones cachés*

Chaque neurone peut prendre en compte des profils spécifiques de neurones d'entrée. Un nombre plus important permet donc de mieux "coller" aux données présentées, mais diminue la capacité de généralisation du réseau. Il faut alors trouver le nombre adéquat de neurones cachés nécessaire pour obtenir une approximation satisfaisante. Le nombre de neurones cachés dépend du type de données dont on dispose. En général, ce nombre est déterminé par des règles empiriques. Suivant les auteurs, la taille de la couche cachée doit être (Dames 2005) :

- ✓ Soit égale à celle de la couche d'entrée ;
- ✓ Soit égale à 75 % de celle-ci ;

- ✓ Soit égale à la racine carrée du produit du nombre de neurones dans la couche d'entrée et de sortie.

- ✓ Pour des cas plus sophistiqués, la recherche consisterait soit à :
- ✓ Estimer un réseau comportant de nombreux neurones puis à le simplifier par l'analyse des multi-colinéarités ou par une règle d'apprentissage éliminant les neurones inutiles (Dames 2005) ;

- ✓ Définir une architecture tenant compte de la structure des variables identifiées au préalable par une analyse en composantes principales (Bermejo et Cabestany 1999, Bermejo et Cabestany 2001) ;

- ✓ Augmenter progressivement la complexité du modèle à chaque itération, dite "Cascading", et réévaluer l'efficacité de généralisation à chaque addition d'un nouveau neurone (Dames 2005).

### II.5.4. Apprentissage

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

L'apprentissage correspond donc à la phase du développement du réseau durant laquelle les coefficients synaptiques sont modifiés jusqu'à l'obtention du comportement désiré.

Dans notre étude, comme nous disposons des valeurs de sortie, nous avons choisi l'apprentissage supervisé, mettant en jeu un algorithme d'optimisation. Ce dernier cherche à minimiser, par des méthodes d'optimisation non linéaire, une fonction 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 : le gradient est estimé par une méthode spécifique, dite la règle du gradient de l'erreur (en anglais delta rule). C'est une des règles les plus utilisées pour l'apprentissage supervisé de réseaux de neurones. Initialement développée pour résoudre des problèmes de traitements adaptatifs du signal, elle a conduit ensuite à un algorithme connu sous le nom de l'algorithme de rétro-propagation du gradient de l'erreur (en anglais Back-propagation) (Dames 2005). On distingue deux étapes dans la phase d'apprentissage (Dames 2005) :

- ✓ Une phase de propagation, qui consiste à présenter une configuration d'entrée au réseau, puis à la propager à celle de sortie en passant par les couches cachées.

- ✓ Une phase de rétro-propagation, qui consiste, après le processus de propagation, à minimiser l'erreur commise sur l'ensemble des exemples présentés, erreur considérée comme une fonction des poids synaptiques. Cette erreur représente la somme des différences au carré entre les réponses calculées et celles désirées pour tous les exemples contenus dans l'ensemble de l'apprentissage.

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

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

Le tableau II.2 résume les algorithmes d'apprentissage par rétro-propagation implanté dans 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® (Demuth et Beale 2000).

| **Algorithme** | **Caractéristiques** |
|---|---|
| **Descente du gradient** | |
| Descente du gradient simple (***traingd)*** | Descente du gradient de base. Réponse lente ; peut être utilisée pour apprentissage en mode séquentiel |
| Descente du gradient avec moment (***traingdm)*** | Descente du gradient avec moment. Le moment permet à un réseau de répondre non seulement au gradient local, mais aussi aux tendances récentes dans la surface d'erreur. Sans moment un réseau peut être coincé dans un minimum local peu profond. Plus rapide que ***traingd***, ne peut être utilisée que pour apprentissage en mode séquentiel |
| Descente du gradient avec taux d'apprentissage variable (***traingda)*** | Taux d'apprentissage variable. Plus rapide que ***traingd***, mais ne peut être utilisée que pour apprentissage en mode batch. Utilise des règles heuristiques pour la variation du taux d'apprentissage |
| Descente du gradient avec moment et taux d'apprentissage variable (***traingdx)*** | Descente du gradient avec inertie (momentum) et taux d'apprentissage variable |
| **Descente des gradients conjugués** | |
| Descente des gradients conjugués avec la méthode de mise à jour de Fletcher-Reeves ***(traincgf)*** | Il possède la plus petite mémoire de stockage parmi les algorithmes de gradient conjugués. |
| Descente des gradients conjugués avec la méthode de mise à jour de Polak-Ribiére ***(traincgp)*** | Nécessite une mémoire de stockage légèrement plus grande que ***traincgf***. Convergence plus rapide sur quelques problèmes. |
| Descente des gradients conjugués avec la méthode de reprises Powell-Beale ***(traincgb)*** | Nécessite une mémoire de stockage légèrement plus grande que ***traincgp***. Convergence Généralement plus rapide. |
| Scaled conjugate gradient ***(trainscg)*** | Le seul algorithme de gradients conjugués qui n'exige aucune recherche de ligne. Très bon algorithme d'apprentissage en général. |
| **Quasi-Newton** | |
| BFGS quasi-Newton ***(trainbfg)*** | Il exige le stockage d'une approximation de la matrice Hessiènne et nécessite plus de calcul à chaque itération que les algorithmes de gradients conjugués, mais possède habituellement une convergence plus rapide. |

| | |
|---|---|
| Sécante à un pas ***(trainoss)*** | Cet algorithme exige moins de stockage et calcul par époque que l'algorithme BFGS. On peut considérer cela un compromis entre des algorithmes de quasi-Newton complets et des algorithmes de gradients conjugués. |
| **Levenberg-Marquardt** | |
| Levenberg-Marquardt ***(trainlm)*** | C'est l'algorithme d'apprentissage le plus rapide pour les réseaux de taille modérée. Il a la caractéristique de réduction de mémoire pour l'utilisation quand l'échantillon d'apprentissage est grand. |
| **Régularisation** | |
| Régularisation Bayésienne ***(trainbr)*** | Mise à jour des poids et des biais selon l'optimisation de Levenberg-Marquardt. Il réduit au minimum une combinaison d'erreurs quadratiques et des poids et détermine ensuite la combinaison correcte afin de produire un réseau avec une capacité de généralisation meilleure. Ce processus est appelé la régularisation bayésienne. Il réduit la difficulté de déterminer l'architecture de réseau optimum. |

### II.5.5. Evaluation des performances du réseau de neurones

Enfin, la dernière étape doit permettre d'estimer la qualité du réseau obtenu en lui présentant des exemples qui ne font pas partie de l'ensemble d'apprentissage. Elle permet à la fois d'apprécier les performances du système neuronal et de détecter le type de données qui posent problème. Si les performances ne sont pas satisfaisantes, il faudra, soit modifié l'architecture du réseau, soit modifié la base d'apprentissage (Dames 2005).

## II.6. Application des réseaux de neurones en bioprocédés

Durant les dernières années, la littérature scientifique relative aux réseaux de neurones (RN) s'est enrichie considérablement par la publication des travaux relatant des applications diverses telles que la modélisation, la reconnaissance des formes, l'optimisation et le contrôle des bioprocédés, tous sujet d'intérêt en biotechnologie. Les ouvrages de Baughman et Liu (Baughman et Liu 1995), Montague et Morris (Montague et Morris 1994), Basheer et Hajmeer (Basheer et Hajmeer 2000) et Nicoletti (Nicoletti 2009) donnent des aperçus sur les applications des RN en bioprocédés. Plusieurs applications sont résumées ci-dessous :

### II.6.1. Modélisation et prédiction

Plusieurs exemples, sur la modélisation et la prédiction par les RN, sont cités dans la littérature. À titre d'exemple, l'estimation en ligne des concentrations de la biomasse, du glucose et de l'éthanol par un réseau récurrent ayant pour entrées : les mesures antérieures, les conditions du processus et les entrées de commande du processus. Zhu et ces collaborateurs (Zhu *et al.* 1996) ont utilisé un RN pour décrire la production à grande échelle de lysine et ont prédit avec succès la formation du produit à partir des mesures en ligne et la simulation récursive de la consommation du sucre. Galván et ces co-auteurs (Galván *et al.* 1996) ont utilisé les réseaux de neurones dans l'ajustement de données cinétiques complexes des réactions chimiques, ils ont démontré que cette modélisation s'avère très utile lorsque les mécanismes cinétiques ne sont pas bien saisis et que suffisamment de données sur le processus sont disponibles. Les RN peuvent remplacer les modèles (semi-) mécanistiques, parce que, une fois ajusté, ils sont mathématiquement robuste, facile à utiliser et rapide (Fontaine et Germain 2001).

Dans le travail de Shene et co-auteurs (Shene *et al.* 1999)**,** deux réseaux de neurones différents (un réseau neuronal de type feedforward et un réseau neuronal hybride) ont été conçus pour prédire les variables d'état dans la fermentation batch de l'éthanol en tenant compte de l'effet de la composition du milieu et de la température. Becker et ces collaborateurs (Becker *et al.* 2002) ont appliqué un RN dynamique pour la modélisation de la fermentation de la bière, suivie d'une optimisation des profils de température afin de réduire le temps de fermentation. Saraceno et co-autheurs (Saraceno *et al.* 2009) ont simulé la fermentation de *lactosérum de fromage ricotta* pour la production d'éthanol au moyen d'un modèle neuronal hybride multiple, obtenu par couplage d'un réseau neuronal de type perceptron multicouches avec les équations de bilan de masse pour le lactose, l'éthanol et la biomasse. Nasr et ces collaborateurs (Nasr *et al.* 2013) ont développés un modèle neuronal de type rétro-propagation pour estimer le profil de production d'hydrogène en fonction du temps dans les études batch, les entrées du RN sont : le *pH* initial, les concentrations initiales du substrat et de la biomasse, la température et le temps.

### II.6.2. Capteurs logiciels (en anglais soft sensors)

Les RN sont aussi appliqués pour le développement de capteurs logiciels à des fins d'estimation et de contrôle. Généralement les mesures on line sont reliées aux mesures off line des états du processus (Di Massimo *et al.* 1992, Karim *et al.* 1997, Ye *et al.* 1998, Linko *et al.* 1999). Par exemple, les mesures on line tel que le *pH*, l'ammoniac et la vitesse d'agitation ont été appliqués pour prédire les niveaux de biomasse et de produit dans le travail de Jin et co-auteurs (Jin *et al.* 1996) et

Linko et co-auteurs (Linko *et al.* 1997). Yuan et Vanrolleghem (Yuan et Vanrolleghem 1999) ont appliqué les RN pour la prédiction on line des profils de concentration de pénicilline.

En outre, les informations spécifiques du processus comme la phase du processus et le temps de fermentation peuvent être utilisées comme des inputs additionnels (Di Massimo *et al.* 1992). Les capteurs logiciels peuvent aussi être appliqués pour remplacer les procédures analytiques couteuses et difficiles. Kang et ces collaborateurs (Kang *et al.* 1998) ont développé un capteur logiciel pour la mesure de la concentration de l'acide clavulanique. Une autre nouveauté est le nez électronique (Bachinger et Mandenius 2000).

### II.6.3. Reconnaissance des formes (en anglais pattern recognition)

Les RN sont appliqués pour la détection des irrégularités du processus, pour réduire la variabilité du processus, pour la prédiction on line de la performance du processus et pour détecter les défauts des capteurs (Glassey *et al.* 1997). Un autre exemple est l'application d'un RN à auto-organisation (Kohonen) pour établir un rapport entre les données de la qualité des semences pour prédire la performance de la phase finale de la fermentation (Ignova *et al.* 1999) ; Par ailleurs, la phase finale atypique de la fermentation peut être détectée dès le début. L'application des RN auto-associative pour la détection d'anomalie on line a été proposée par Shimizu et co-auteurs (Shimizu *et al.* 1998) et aussi appliquée dans la production de virginiamycin par Huang et ces collaborateurs (Huang *et al.* 2002).

### II.6.4. Optimisation

Les RN ont été appliqués pour l'optimisation à base de modèle des milieux de culture en mode batch par kennedy et co-auteurs (Kennedy *et al.* 1992, Kennedy et Spooner 1996), qui a abouti à une grande réduction dans les expériences par comparaison à l'approche classique basée sur les plans d'expérience.

La combinaison des RN avec les algorithmes évolutionnaires, tel que les algorithmes génétiques, a donné naissance à une nouvelle approche qui a prouvé et ne cesse de prouver son efficacité et sa robustesse dans l'optimisation des milieux de culture de *Lactobacillus plantarum* (Desai *et al.* 2006), optimisation des constituants du milieu pour la production de lipase (Haider *et al.* 2008), optimisation de la production de scléroglucane (Desai *et al.* 2008), optimisation de la production de L-asparagine (Gurunathan et Sahadevan 2011) et optimisation de la production de lipase (Nelofer *et al.* 2012).

### II.6.5. Contrôle

Les RN sont souvent appliqués au sein des schémas de contrôle en tant que modèles, estimateurs d'état ou optimiseur. Schenker et Agarwal (Schenker et Agarwal 1998) ont étudié différent type de structures neuronales pour le contrôle prédictif des réacteurs chimiques. Kovarova et ces collaborateurs (Kovarova-Kovar *et al.* 2000) ont utilisé un modèle neuronal en combinaison avec une optimisation on line pour l'adaptation de la trajectoire d'alimentation durant la production de riboflavine.

## II.7. Conclusion

Ce chapitre introduit les notions de base sur les réseaux de neurones artificiels (RN). Après un historique sur l'apparition des RN et la présentation du modèle biologique et mathématique du neurone, les différentes architectures des RN ainsi que leurs types d'apprentissage sont présentés. Dans un deuxième temps, nous avons exposé les RN les plus utilisés ainsi que les différentes étapes de conception d'un réseau de neurones. Nous avons terminé ce chapitre par la présentation de différent travaux relatant des applications diverses des RN en bioprocédés notamment dans la modélisation, la prédiction, la reconnaissance des formes, l'optimisation et le contrôle.

# *Chapitre III*

# *Modélisation QSPR/QSAR*

## III.1. Généralités sur la modélisation QSPR/QSAR

Une relation QSPR/QSAR est un modèle ou une formule mathématique qui permet de relier, d'une manière quantitative, la structure d'une molécule à une propriété ou à une activité donnée. Les méthodes QSPR/QSAR sont de plus en plus utilisées, du fait de la croissance des moyens de calculs (machines, logiciels,...). Récemment, on assiste à la mise en place d'un nouveau règlement REACH (*Registration, Evaluation, Authorisation and Restriction of Chemicals*) qui recommande l'utilisation des méthodes alternatives pour limiter le recours à l'expérimentation (Margossian 2007).

En fait, les premiers travaux QSPR/QSAR remontent au 19ème siècle. En effet, dès 1868, Crum-Brown et Fraser ont postulé l'existence de relations entre les activités physiologiques et les structures chimiques en reliant les variations de l'activité biologique à des modifications structurales (Crum Brown *et al.* 1868)**.**

Cependant, à cette époque, les structures moléculaires n'étaient pas encore connues. Une avancée importante vers les modèles QSPR/QSAR proprement dits a été réalisée grâce au développement des équations de Hammett dans lesquelles les constantes σ caractérisent de manière quantitative les vitesses de réactions pour les composés organiques (Hammett 1937).

$$\log \frac{K}{K_{\circ}} = \sigma\rho \qquad \text{(III.1)}$$

Ou K et $K_{\circ}$ sont les constantes respectives de la réaction étudiée et de celle d'une référence et ρ une constante de réaction dépendant du type de réaction (Hammett 1937 ; Brown *et al.* 1868).

Les premiers travaux utilisant la méthodologie QSPR/QSAR telle qu'employée actuellement sont dus à Hansch et Free et Wilson. D'un côté, Hansch a proposé des modèles reliant directement l'activité biologique des composés avec les propriétés hydrophobes, électroniques et stériques à l'échelle moléculaire. D'un autre côté, Free et Wilson ont développé des modèles empiriques, dits de contributions de groupes, pour l'étude de l'activité biologique (Hansch 1995 ; Free *et al.* 1964).

Au cours de ces dernières décennies, l'utilisation des méthodes QSPR/QSAR n'a pas cessé de progresser. Elle est même devenue indispensable en chimie pharmaceutique, en toxicologie et pour la conception de médicaments (Grover et al 2000 ; Katritzky et al 2002).

## III.2. Principe des méthodes QSPR/QSAR

Le principe des méthodes QSPR/QSAR est d'établir une relation mathématique reliant de manière quantitative des propriétés moléculaires, appelées descripteurs, avec une observable macroscopique (activité biologique, toxicité, propriété physico-chimique, etc.), pour une série de composés chimiques similaires à l'aide de méthodes d'analyses de données. La forme générale d'un tel modèle est la suivante (Tropsha *et al.* 2003) :

Propriété/ Activité = f (D1, D2,… Dn…) (III.2)

D1, D2, …Dn sont des descripteurs des structures moléculaires.

L'objectif d'une telle méthode est d'analyser les données structurales afin de détecter les facteurs déterminants pour la propriété /activité mesurée. Pour ce faire, différents types d'outils statistiques peuvent être employés (Tropsha *et al.* 2003) :

- Régressions linéaires simples et multiples,
- Régressions aux moindres carrées partielles (PLS),
- Arbres de décision,
- Réseaux de neurones,
- Algorithmes génétiques, Vecteurs Machines.

Une fois cette relation est établie et validée, elle peut alors être employée pour la prédiction de la propriété /activité de nouvelles molécules, pour lesquelles les valeurs expérimentales ne sont pas disponibles. De tels modèles peuvent être également utilisés pour mieux comprendre les mécanismes et les modes d'action (He 2005).

## III.3. Méthodologie générale d'une étude QSPR/QSAR

La méthodologie générale d'une étude QSAR/QSPR (Voir Figure III.1) est la suivante (Tropsha *et al.* 2003):

a) Constituer une base de données à partir des mesures expérimentales fiables de la propriété ou de l'activité de chaque composé.
b) Sélectionner les descripteurs en relation avec la propriété ou l'activité étudiée.

c) Diviser cette base de données, aléatoirement, en une série d'apprentissage (training set) qui contient généralement les 2/3 de la base de données et une série de test (test set) constituée par le 1/3 restant.
d) Etablir des modèles mathématiques en utilisant la série d'apprentissage.
e) Caractériser les modèles élaborés par leurs indices de validation internes et vérifier leur robustesse par un test de hasardisation (Randomization) de la variable dépendante Y (réponse).
f) Valider les modèles élaborés en utilisant la série de test et calculer leurs paramètres statistiques de validation externe.
g) Explorer et exploiter les modèles validés pour comprendre les mécanismes et les modes d'action.

### *III.3.1. Bases de données*

Un modèle QSPR est très dépendant des données expérimentales de référence. Le choix de la base de données est décisif dans le développement de tel modèle. Pour être de qualité, une base de données doit être composée de données expérimentales aussi fiables que possible obtenues en suivant un protocole unique puisque les erreurs sur celles-ci se propageront dans le modèle final. Il y'a plusieurs éléments à vérifier dans les étapes de nettoyage d'une base de données. Il faut tout d'abord vérifier que les structures sont correctes d'un point de vue chimique (règle de valence, …), des structures erronées entrainent la génération de mauvais descripteurs et donc de mauvais modèles (Tropsha *et al.* 2003).

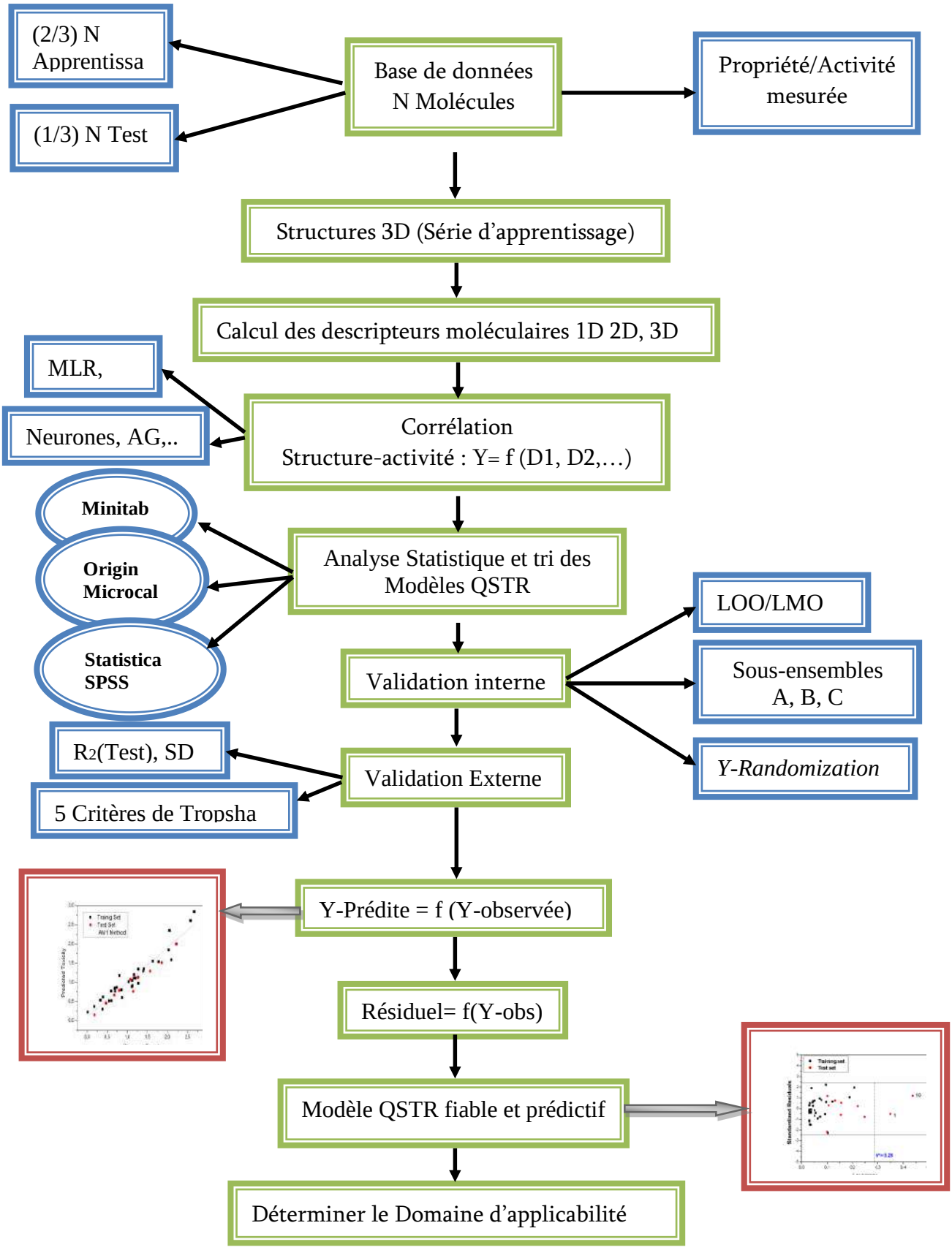



**Figure III.1.** Schéma d'élaboration et validation d'un modèle QSTR (Tropsha *et al.* 2003)

### *III.3.2. Descripteurs moléculaires*

De nombreuses recherches ont été menées, au cours de ces dernières décennies, pour trouver la meilleure façon de représenter l'information contenue dans la structure des molécules et ces structures elles-mêmes en un ensemble de nombres réels appelés descripteurs qui sont soit théoriques ou empiriques (Tropsha *et al.* 2003):

#### *III.3.2.1. Les descripteurs moléculaires théoriques*

***Les descripteurs 1D*** : sont accessibles à partir de la formule brute de la molécule et décrivent des propriétés globales du composé comme le nombre d'atomes et la masse moléculaire,…etc.

Ces descripteurs sont couramment utilisés du fait de leur extrême simplicité. Cependant, ils peuvent poser problème pour une bonne interprétation des mécanismes d'interaction du fait qu'ils ne permettent pas de tenir en compte des effets stériques et d'isomérie.

***Les descripteurs 2D***: sont calculés à partir de la formule développée de la molécule. On distingue :

- *Les indices 2D constitutionnels* : qui caractérisent les différents composants de la molécule. Il s'agit par exemple du nombre de liaisons simples ou multiples, du nombre de cycles,…etc.
- *Les indices 2D 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.

Exemples : indice de Wiener, indice de Randić, indice de connectivité de valence de Kier-Hall, indice de Balaban, …etc. (Wiener1947) (Randić1975) (Kier-Hall1975) (Balaban1982)

Ces descripteurs 2D permettent de prédire les propriétés physiques mais sont insuffisantes pour expliquer certaines propriétés et activités biologiques comme la toxicité.

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

- *Les descripteurs 3D géométriques* : les plus importants sont le volume moléculaire, la surface accessible au solvant, le moment principal d'inertie.
- *Les descripteurs 3D électroniques* : permettent de quantifier les différents types d'interactions inter- et intramoléculaires, de grande influence sur l'activité biologique des molécules. Le calcul de la

plupart de ces descripteurs nécessite la recherche de la géométrie pour laquelle l'énergie est minimale, et fait souvent appel à la chimie quantique.

**Tableau III.1.** Classification générale des descripteurs moléculaires (Karelson 2010).

| Classe | Sous-classe |
|---|---|
| **Descripteurs constitutionnels** | • Dénombrement des atomes ou des liaisons.<br>• Descripteurs basés sur les masses atomiques. |
| **Descripteurs topologiques** | • Indices topologiques (connectivité).<br>• Descripteurs théoriques d'information.<br>• Descripteurs topochimiques. |
| **Descripteurs géométriques** | • Descripteurs liés à la distance.<br>• Descripteurs liés à l'aire de la surface.<br>• Descripteurs liés au volume.<br>• Descripteurs du champ stériquemoléculaire. |
| **Descripteurs liés à la distribution de charge** | • Charges atomiques partielles.<br>• Moments électriques moléculaires<br>• Polarisabilités moléculaires.<br>• Descripteurs du champ électrique moléculaire. |
| **Descripteurs liés aux orbitales moléculaires** | • Energie des OM frontières<br>• Ordres de liaison<br>• Indices de réactivité de Fukui. |
| **Descripteurs température dépendants** | • Fonctions thermodynamiques.<br>• Descripteurs facteurs de Boltzmann pondérés. |
| **Descripteurs température dépendants** | • Fonctions thermodynamiques.<br>• Descripteurs facteurs de Boltzmann pondérés. |
| **Descripteurs de solvatation** | • Energie électrostatique de solvatation.<br>• Energie de dispersion de solvatation.<br>• Enthalpie libre de formation de cavité.<br>• Descripteurs de liaison hydrogène.<br>• Entropie de solvatation.<br>• Descripteurs d'énergie de solvatation Linéaire théoriques. |
| **Descripteurs mixtes** | • Descripteurs topographiques.<br>• Descripteurs électrotopologiques.<br>• Descripteurs de la charge partielle de l'aire de la surface. |

### *III.3.2.2. Descripteurs moléculaires empiriques*

Les descripteurs empiriques peuvent être divisés en deux classes générales (tableau II.2), la première reflète les interactions électroniques intramoléculaires (descripteurs structurels) alors que la seconde tient compte des interactions intermoléculaires dans les milieux condensés tels que les liquides et les solutions (descripteurs de solvatation).

**Tableau III.2.** Classification d'ensemble des descripteurs moléculaires empiriques (Karelson 2010).

| classe | Sous classe |
|---|---|
| **Descripteurs structurels** | • Constantes d'induction<br>• Constantes de résonance<br>• Constantes stériques |
| **Descripteurs de solvatation** | • Echelles de polarité<br>• Echelles de polarisabilité<br>• Echelles d'acidité<br>• Echelles de basicité<br>• Echelles mixtes |

### *III.3.3. Méthodes d'analyse des données*

Pour élaborer un modèle QSPR/QSAR nous avons besoin d'une méthode d'analyse de données, cette méthode permet de quantifier la relation qui existe entre la propriété/Activité et la Structure (descripteurs). Il existe plusieurs méthodes pour construire un modèle et analyser les données statistiques de ce dernier, certaines sont linéaires telles que la régression linéaire multiple (MLR), la régression aux moindres carrés partiels (PLS), d'autres sont non linéaires comme les arbres de décisions, les réseaux de neurones… ces méthodes sont disponibles dans des logiciels tels que, Excel, Origin Microcal, Minitab, Statistica, SPSS et R (Ghasemi et al 2007 ; Duprat *et al.* 1998 ; Gasteiger *et al.* 1993)

### *III.3.3.1. Interprétation et validation d'un modèle QSPR/QSAR*

Une fois développé, le modèle doit être interprété en analysant tous les paramètres statistiques de ce modèle, sa qualité doit être aussi étudiée, cette qualité est vérifiée par ce que l'on appelle validation. Sa robustesse, c'est-à-dire l'influence des composés de la série d'apprentissage sur le modèle, est estimée par des méthodes de validation interne. Afin d'estimer son pouvoir prédictif, des données expérimentales supplémentaires sont nécessaires afin de déterminer la capacité du modèle à prédire ces valeurs c'est ce que l'on appelle validation externe. Enfin, il est important de savoir quel

type de molécules utilisées avec quel modèle. On parle alors de domaine d'applicabilité (Tropsha *et al.* 2003).

#### *III.3.3.1.1. Validation interne*

Dans le passé, la validation interne d'un modèle QSPR/QSAR a été réalisée en utilisant la validation croisée LOO (leave-one out) ou LMO (leave-many out). Ce processus consiste à extraire un certain nombre k de molécules du jeu initial à N molécules et à construire un nouveau modèle avec les (N-k) molécules restantes à l'aide des descripteurs choisis (seules les constantes de la régression changent). Ce processus est ensuite réitéré pour retirer et prédire les valeurs de toutes les molécules de la série d'apprentissage. En fonction du nombre de molécules retirées à chaque itération, on parlera de Leave-One-Out (LOO) ou de Leave-Many-Out (LMO) selon qu'une ou plusieurs molécules est (sont) retirée(s). Dans ces dernières années, d'autres méthodes sont utilisées pour faire la validation interne, tel que la hasardisation de la réponse (Y-Randomization) (Tropsha *et al.* 2003).

**Y-Randomization**

Afin de s'assurer qu'un modèle QSPR est fiable, les tests d'Y-randomization sont une des techniques les plus employées. En effet, il n'est pas rare d'obtenir des corrélations fortuites (Ou chance corrélation), c'est-à-dire un modèle affichant de bons résultats statistiques (R2, SD) pour l'apprentissage, mais impliquant des descripteurs qui ne sont pas reliés à la propriété/activité modélisée. Ces modèles aléatoires peuvent être détectés par la procédure Y-randomization. Elle consiste à mélanger aléatoirement les propriétés/activités expérimentales pour la série d'apprentissage en utilisant les mêmes descripteurs, de nouveaux modèles sont obtenus. Ces derniers doivent avoir des performances très faibles. Cependant, la validation interne est insuffisante pour étudier le pouvoir prédictif d'un modèle. Pour cette raison la validation externe du modèle est devenue une norme et une partie obligatoire dans la modélisation QSPR/QSAR (Tropsha *et al.* 2003).

#### *III.3.3.1.2. Validation externe*

Cette méthode consiste à prédire la propriété/activité d'une série de molécules appelée généralement série de test qui ne sont pas dans la série de développement du modèle, cette validation est caractérisée par les paramètres R2 (test) R2cv (test). Récemment plusieurs études ont montré l'insuffisance des paramètres R2, R2CV pour vérifier le pouvoir prédictif des modèles QSAR. Par conséquent, d'autres paramètres doivent être vérifiés pour cet objectif. Ces paramètres sont connus

sous le nom « critères de validation externe » ou souvent appelés « critères de Trophsa » (Tropsha criteria)

Critères de validation Externe (série de test)

$R^2$>0.7 (III.3)

$\frac{(R^2-R_\circ^2)}{R^2}\langle 0.1$ $0.85 \leq k \leq 1.15$ (III.4)

$\frac{(R^2-R_\circ'^2)}{R^2}\langle 0.1$ $0.85 \leq k' \leq 1.15$ (III.5)

$|R^2 - R_\circ^2|\langle 0.3$ (III.6)

Avec

$R^2$ Coefficient de corrélation pour les molécules de la série de test.

$R_\circ^2$ Coefficient de corrélation entre les valeurs prédites et expérimentales pour la série detest.

$R_\circ'^2$ Coefficient de corrélation entre les valeurs expérimentales et prédites pour la série detest.

k : est la constante de la droite (à l'origine) de corrélation (valeurs prédites en fonction des valeurs expérimentales)

k' : est la constante de la droite (à l'origine) de corrélation (valeurs expérimentales en fonction des valeurs prédites)

### *III.3.3.1.3. Domaine d'applicabilité*

Comme les modèles QSAR sont utilisés pour la prise de décisions relatives aux augmentations de gestion des produits chimiques, la fiabilité des prévisions du modèle est une préoccupation croissante.

Un modèle QSAR/QSPR ne peut pas être considéré comme un modèle universel, parce qu'il est développé sur un nombre limité de composés qui ne couvrent pas tout l'espace chimique. Par conséquent l'activité/propriété prédite d'un composé, chimiquement dissimilaire au jeu d'apprentissage, ne pourra pas être considérée fiable. (Tetko *et al.* 2005 ; Jaworska *et al.* 2005).

Le domaine d'applicabilité (DA) permet de définir la zone dans laquelle un composé pourra être prédit avec confiance. Le DA correspond donc à la région de l'espace chimique incluant les

composés du jeu d'apprentissage et les composés similaires, proches dans ce même espace (Netzeva 2005).

Ils existent plusieurs méthodes pour la détermination de domaine d'applicabilité d'un modèle QSPR/QSAR parmi ces méthodes on trouve la méthode de « leverage » cette méthode est basée sur la variation des résiduels standardisés de la variable dépendante avec « leverage » (la distance entre les valeurs des descripteurs et leurs moyennes). Si un composé a un résiduel et un leverage qui dépasse le seuil h*=3p/n (ou p est le nombre de descripteurs plus 1 et n le nombre d'observations), ce composé est considéré en dehors du domaine d'applicabilité du modèle élaboré.

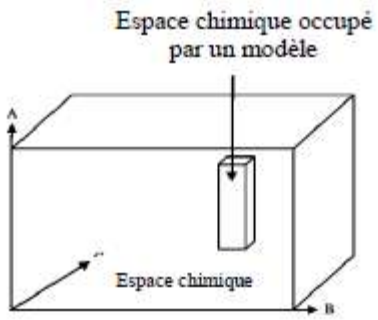



**Figure III.2.** Représentation schématique du domaine d'applicabilité d'un modèle QSPR/QSAR (Netzeva 2005).

## III.4. Applications des méthodes QSPR/QSAR

Les applications des méthodes QSPR/QSAR sont très nombreuses, elles touchent tous les domaines ou la structure chimique intervient, entre autre on peut citer (Tropsha *et al.* 2003) :

- ✓ **Propriétés physico-chimiques** : point d'ébullition, Point de fusion, densité, indice de réfraction, température critique, viscosité, solubilité, pression de vapeur, tension superficielle, Coefficients de partition : eau/octanol, air/eau, huile/air, lait/plasma (Katritzky *et al.* 2010).
- ✓ **Activités biologiques** : Anti VIH, Anti malaria, Anti Diabète, Anti Cancer, Anti-oxydante, Antiinflammatoire… (Selzer *et al.* 1996).
- ✓ **Autres propriétés/activités** :
    - ✓ Prédiction de la toxicité aquatique des composés chimiques vis-à-vis des espèces environnementales (Tetko *et al.* 2008)
    - ✓ Toxicité des nanoparticules

- ✓ Toxicité des pesticides et des colorants
- ✓ Propriétés inhibitrices de corrosion
- ✓ Concentration micellaire critique
- ✓ Prédiction de plusieurs propriétés dangereuses telle que l'explosibilité et l'inflammabilité de certaines familles de molécules chimiques.
- ✓ Conception des médicaments et de nombreux autres produits tels que les agents tensio-actifs, parfums, les colorants et les produits de la chimie fine. …

# *Chapitre IV*

## *Modélisation du taux de fixation des médicaments aux protéines plasmatiques*

## IV.1. Introduction

Ce chapitre traite la modélisation du taux de fixation de certains médicaments aux protéines plasmatiques. Pour ce faire, on commencera par une première partie intitulée "procédure de développement des modèles prédictives". Cette dernière comportera quatre axes. Dans le premier axe, on relatera les diverses étapes de l'élaboration de la base de données nécessaire au fonctionnement des méthodes d'apprentissage statistique qu'on utilisera. Dans notre cas, le paramètre à modéliser (c'est-à-dire la sortie) est le taux de fixation aux protéines plasmatiques pour un ensemble de médicaments. Le second axe sera consacré aux calculs des descripteurs par le logiciel Dragon 1.0. Le troisième axe traitera la méthode permettant de sélectionner les descripteurs pertinents qui constitueront la base de données, elle revêt une importance capitale car elle conditionne la qualité du modèle qui sera obtenu. Le quatrième et dernier axe se portera sur le développement des modèles QSAR. Il sera question de l'ensemble des étapes permettant l'élaboration de deux modèles linéaire (par RLM) et non linéaire (par RNA) de prédiction du taux de fixation aux protéines plasmatiques. Pour estimer les performances des deux modèles, nous utiliserons un certain nombre d'indicateurs statistiques.

La seconde partie de ce chapitre se veut une application à l'élaboration de modèles QSAR pour la prédiction du taux de fixation aux protéines plasmatiques. On y discutera des résultats obtenus.

## IV.2. Procédure de développement des modèles prédictives

La figure IV.1 présente un schéma simplifié de la procédure adopté pour la modélisation prédictive de $f_b$. Cette procédure comprend 6 étapes : Collecte de données, calcul des descripteurs moléculaires, le prétraitement de la base de données et la génération du modèle, qui permet l'établissement de relations statistiquement significatives entre la propriété ciblée (taux de fixation aux protéines plasmatique) et les valeurs des descripteurs.

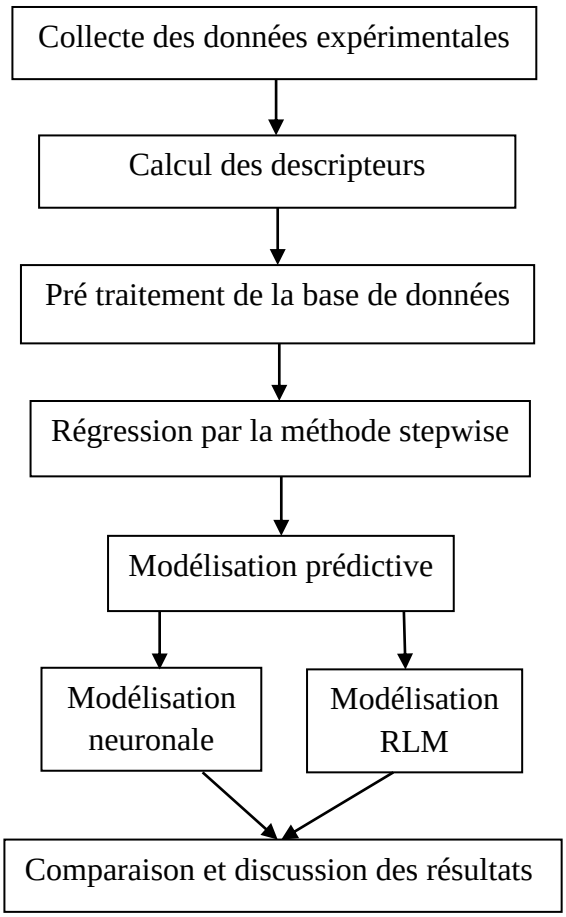



**Figure IV.1.** Schema de la procedure utilisée pour la selection des descripteurs les plus pertinent.

### IV.2.1. *Collecte des données*

La fraction expérimentale du médicament lié aux protéines plasmatiques ($f_b$) d'une série de 277 médicaments a été tiré de la littérature (Alberts DS *et al.* 1979, Atiba *et al.* 1987, Goodman *et al.* 1996, Weber C *et al.* 1999, Anderson *et al.* 2002, Turner *et al.* 2004, Goodman *et al.* 2006, Goodman *et al.* 2011). L'ensemble de données a été divisé au hasard en trois sous-ensembles : une base d'apprentissage (78% : 216 médicaments), une base de test (19% : 54 molécules) et une base pour la validation (3% : 7 composés).

### IV.2.2. *Calcul des descripteurs*

La représentation numérique de la structure moléculaire a été évaluée en termes de descripteurs moléculaires. La notation SMILES nécessaire au calcul des descripteurs a été tirée à partir du site DrugBank database(www.drugbank.ca). Sur les figures IV.2 ci-dessous, nous donnons

la capture d'écran du site utilisé. Ces notations ont été utilisées pour générer 1666 descripteurs répartis en vingt catégories : constitutional (0D), topological (2D), walk and path counts (2D), connectivity indices (2D), information indices (2D), two-dimensional autocorrelations (2D), edge adjacency indices (2D), Burden eigenvalues (2D), topological charge indices (2D), eigenvalue-based indices (2D), Randic molecular profiles (3D), geometrical descriptors (3D), RDF descriptors (3D), 3D-MoRSE descriptors (3D), WHIM descriptors (3D), GETAWAY descriptors (3D), functional group counts (1D), atom-centered fragments (1D), charge descriptors (1D), et molecular properties (1D).

Tous les descripteurs ont été obtenus par le biais du programme en ligne E-Dragon 1,0 (www.vcclab.org). Sur le tableau IV.1 ci-après, est répertorié le nombre de descripteurs obtenus par catégorie.




**Figure IV.2.** Capture d'écran du site intitulé DrugBank version 4.3(www.drugbank.ca)

**Tableau IV.1.** Nombre de descripteurs calculés (www.vcclab.org)

| Catégorie de descripteurs | Nombre par catégorie |
|---|---|
| Constitutional descriptors | 48 |
| Topological descriptors | 119 |
| Walk and pathcounts | 47 |
| Connectivity indices | 33 |
| Information indices | 47 |
| 2D autocorrelations | 96 |
| Edgeadjacency indices | 107 |
| Burdeneigen value descriptors | 64 |
| Topological charge indices | 21 |
| Eigen value based indices | 44 |
| Randic molecular profiles | 41 |
| Geometrical descriptors | 74 |
| RDF descriptors | 150 |
| 3D-morse descriptors | 160 |
| WHIM descriptors | 99 |
| GETAWAY descriptors | 197 |
| Functional group counts | 154 |
| Atomcentred fragments | 120 |
| Charge descriptors | 14 |
| Molecular properties | 31 |
| **Total** | **1666** |

### IV.2.3. *Sélection des descripteurs les plus influents*

Le prétraitement de la base de données a pour finalité d'éliminer les descripteurs qui ne sont pas pertinents afin d'éviter le phénomène de sur-apprentissage. Il s'agit alors de la réduction de la dimension de l'espace des variables (descripteurs) du modèle avec l'élimination des entrées qui n'ont pas ou possèdent peu d'influence sur la sortie (taux de fixation aux protéines plasmatiques).

Une procédure basée sur un pré traitement de la base de données et une régression pas à pas (forword stepwise regression) ont été utilisé pour la sélection des descripteurs les plus pertinents suivants les étapes suivantes (Hamadache *et al.* 2014) :

**1.** On calcule le minimum et le maximum pour chaque descripteur à l'aide d'un programme STATISTICA, puis on enlève les descripteurs qui ont le maximum et le minimum égaux.

**2.** On élimine le descripteur qui possède la même valeur de plus de 75% des échantillons.

**3.** On calcule la déviation standard relative (RSD) pour chaque descripteur et on élimine ceux ayant des valeurs de RSD inférieures à 0.05.

**4.** On obtient alors une matrice qui représente la corrélation entre la $f_b$ et les descripteurs retenus. On classe les descripteurs selon la valeur décroissante du coefficient de corrélation. On prend le descripteur ayant la plus grande corrélation et on le compare aux autres descripteurs de la matrice. Ceux dont la valeur du coefficient de corrélation est supérieure à 0.75 sont éliminés à leur tour. On répète la même chose avec le descripteur classé juste après le premier et ainsi de suite.

**5.** On utilise un programme basé sur la méthode stepwise pour la sélection des descripteurs les plus pertinents parmi ceux obtenus à l'étape 4.

### IV.2.4. *Développement des modèles prédictives*

Les deux méthodes utilisées pour l'établissement des modèles QSAR sont la régression linéaire multiple (RLM) et les réseaux de neurones artificiels (RNA).

#### *IV.2.4.1. La régression linéaire multiple (RLM)*

La régression linéaire multiple est un outil statistique dont l'objectif est d'établir une relation linéaire entre une propriété d'intérêt (taux de fixation aux protéines plasmatique) et un ensemble de variables (descripteurs dans notre cas) et ce suivant l'équation IV.1 suivante :

$$f_b = a_0 + \sum a_i d_i \qquad \text{(IV.1)}$$

Où $d_i$ représente les descripteurs moléculaires, $a_i$ leur coefficient et $a_0$ est une constante (appelé intercept).

#### *IV.2.4.2. Les réseaux de neurones artificiels (RNA)*

Les Réseaux de Neurones Artificiels (RNA) sont particulièrement bien adaptés pour les modèles QSAR / QSPR en raison de leur capacité à extraire des informations non-linéaires présentes dans la matrice de données. Pour trouver un modèle non linéaire entre les descripteurs moléculaires sélectionnés selon la méthode pas à pas et le taux de fixation aux protéines plasmatiques, l'utilisation de la méthode RNA est nécessaire et est devenue une technique de modélisation importante. Elle est largement utilisée pour les études de QSAR non linéaires et est un outil puissant pour construire des modèles prédictifs. Le type de réseaux de neurones retenu pour ce travail est le perceptron multicouche (PMC). Le perceptron multicouche est un des réseaux de neurones les plus utilisés pour des problèmes d'approximation, de classification et de prédiction.

## IV.3. Résultat et discussion

### IV.3.1. *Analyse des descripteurs retenus*

La procédure de sélection des variables a généré dix descripteurs (voir tableau IV.2). Le tableau IV.3 montre le degré d'auto corrélation entre ces descripteurs sélectionnés. La valeur du coefficient de corrélation de chaque paire de descripteurs est inférieur à 0,75 ce qui signifie que les descripteurs sélectionnés sont indépendants. En outre, toutes les variables ont des valeurs du facteur d'inflation de la variance (VIF) dans la gamme 1-5 (voir le tableau IV.2), ce qui indique que le modèle obtenu à une signification statistique évidente (Agrawal et Khadikar 2001) et les descripteurs se sont révélés être relativement orthogonaux.

**Tableau IV.2.** Notation, définition, type et VIF des descripteurs sélectionnés

| No. | Notation | Description | Type | VIF |
|---|---|---|---|---|
| 1 | ESpm14x | Spectral moment 14 from edge adj. matrix weighted by edge degrees. | Edge adjacency indices | 1,4803 |
| 2 | ALOGP | Ghose-Crippenoctanol-water partition coeff. (logP). | Molecular properties | 3,8212 |
| 3 | Mor23p | 3D-MoRSE - signal 23 / weighted by atomic polarizabilities. | 3D-MoRSE descriptors | 1,8831 |
| 4 | Inflammat-80 | Ghose-Viswanadhan-Wendoloskiantiinflammatory at 80% (drug-like index). | Molecular properties | 2,1547 |
| 5 | GATS1v | Geary autocorrelation - lag 1 / weighted by atomic van der Waals volumes. | 2D autocorrelations | 1,2406 |
| 6 | Depressant-80 | Ghose-Viswanadhan-Wendoloski antidepressant at 80% (drug-like index). | Molecular properties | 2,2911 |
| 7 | Neoplastic-50 | Ghose-Viswanadhan-Wendoloski antineoplastic at 50% (drug-like index). | Molecular properties | 1,4583 |
| 8 | TPSA(NO) | topological polar surface area using N,O polar contributions. | Molecular properties | 2,8377 |
| 9 | Mor26m | 3D-MoRSE - signal 26 / weighted by atomic masses. | 3D-MoRSE descriptors | 1,1387 |
| 10 | HATS0u | leverage-weighted autocorrelation of lag 0 / unweighted. | GETAWAY descriptors | 1,4387 |

**Tableau IV.3.** Matrice d'autocorelation des descripteurs selectionés

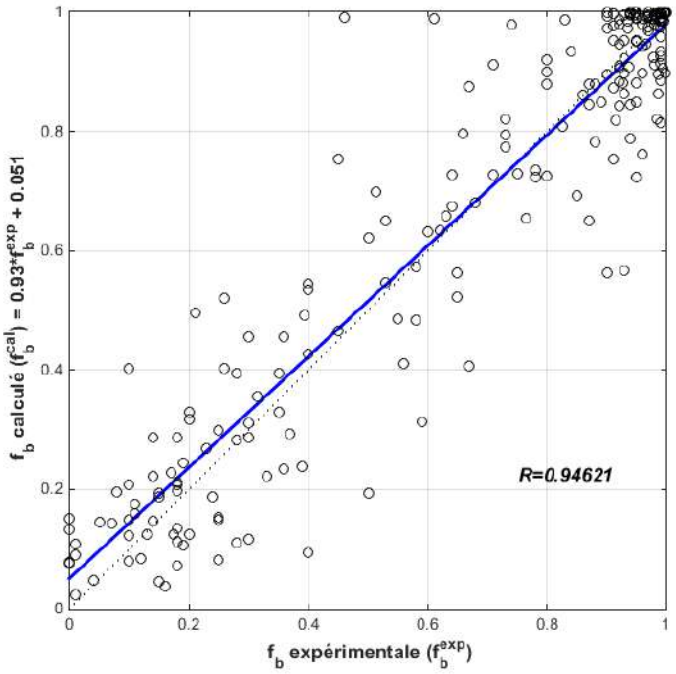
| Descriptors | ESpm14x | ALOGP | Mor23p | Inflammat-80 | GATS1v | Depressant-80 | Neoplastic-50 | TPSA(NO) | Mor26m | HATS0u |
|---|---|---|---|---|---|---|---|---|---|---|
| **ESpm14x** | **1.0000** | 0.1403 | 0.0993 | 0.0300 | 0.1865 | 0.0735 | 0.0123 | 0.3390 | 0.1616 | 0.3790 |
| **ALOGP** | | **1.0000** | 0.6608 | 0.2988 | 0.2941 | 0.2515 | 0.0716 | 0.5784 | 0.0171 | 0.2814 |
| **Mor23p** | | | **1.0000** | 0.2587 | 0.1872 | 0.2080 | 0.0361 | 0.3030 | 0.0065 | 0.1549 |
| **Inflammat-80** | | | | **1.0000** | 0.1349 | 0.6899 | 0.3662 | 0.4079 | 0.0633 | 0.0872 |
| **GATS1v** | | | | | **1.0000** | 0.0218 | 0.0681 | 0.0137 | 0.0953 | 0.2040 |
| **Depressant-80** | | | | | | **1.0000** | 0.4821 | 0.3741 | 0.0209 | 0.0659 |
| **Neoplastic-50** | | | | | | | **1.0000** | 0.0978 | 0.0359 | 0.0835 |
| **TPSA(NO)** | | | | | | | | **1.0000** | 0.1711 | 0.1519 |
| **Mor26m** | | | | | | | | | **1.0000** | 0.2285 |
| **HATS0u** | | | | | | | | | | **1.0000** |

### IV.3.2. *Performance du modèle neuronal*

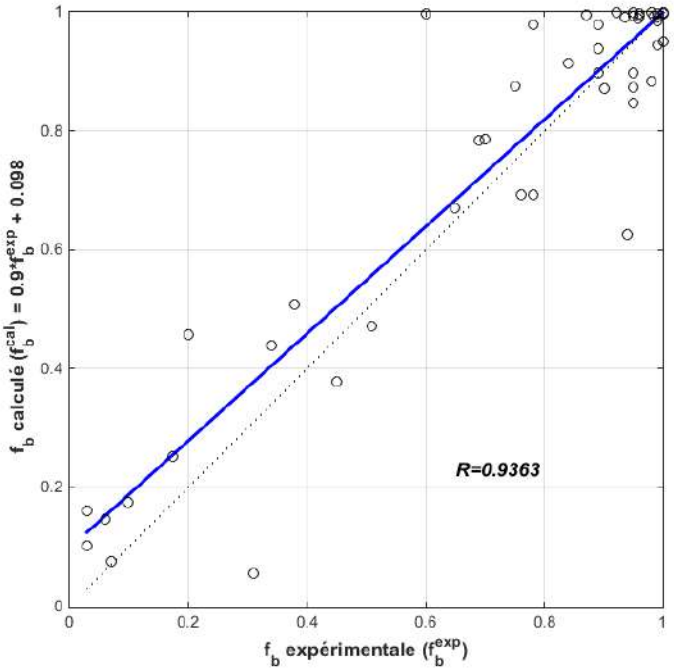
Sur la base de la discussion précédente, un modèle neuronale a été développé dans le but de prédire la fraction de médicament lié aux protéines plasmatiques ($f_b$). Afin d'optimiser son architecture, différentes fonctions de transfert ont été testée et le nombre de neurones dans la couche cachée a été varié de 3 à 15 neurones. L'algorithme d'apprentissage utilisé est le quasi-Newton BFGS.

La meilleure architecture du modèle neuronale obtenu est composée de dix neurones dans la couche d'entrée, six neurones dans la couche cachée ayant "*logsig*" comme fonction de transfert, et un neurone ($f_b$) dans la couche de sortie ayant "*logsig*" comme fonction de transfert (topologie 10-[6] -1).

La figure IV.3 montre la comparaison entre les valeurs expérimentales et les valeurs calculées par le RNA. Il a été prouvé que l'approche neuronale proposée donne des résultats satisfaisants avec des valeurs du vecteur de régression approchant l'idéalité [$\alpha$ = 1 (pente), $\beta$ = 0 (Ordonnée), $R$ = 1 (coefficient de corrélation)].




**(a)**




**(b)**

**Figure IV.3.** Comparaison entre les valeurs expérimentales et les valeurs calculées par RNA:

**(a)** base d'aprentissage, **(b)** base de test.

La figure IV.4 représente le schéma du RNA optimisée. Nous pouvons exprimer $f_b$ par un modèle mathématique qui englobe toutes les entrées $X_i$ (dix descripteurs) comme suit :

La sortie instantanée $Z_j$ de la couche cachée:

$$Z_j = f_H\left[\sum_{i=1}^{10} w_{ji}^I X_i + b_j^H\right] = \frac{1}{1+\exp(-\sum_{i=1}^{10} w_{ji}^I X_i + b_j^H)} \qquad \text{(IV.2)}$$

j=1, 2, … , 6

La sortie $f_b$:

$$f_b = f_0\left[\sum_{j=1}^{6} w_{1j}^H Z_j + b_1^O\right] = \frac{1}{1+\exp(-\sum_{j=1}^{6} w_{1j}^H Z_j + b_1^O)} \qquad \text{(IV.3)}$$

En combinant les équations IV.2 et IV.3, on aura la formule mathématique de $f_b$ en tenant compte de toutes les entrées $X_i$(les dix descripteurs sélectionnés), cette équation peut être facilement appliqué dans la conception assistée par ordinateur des médicaments :

$$f_b = \frac{1}{1+\exp\left(-\sum_{j=1}^{6} w_{1j}^H \left(\frac{1}{1+\exp(-\sum_{i=1}^{10} w_{ji}^I X_i + b_j^H)}\right) + b_1^O\right)} \qquad \text{(IV.4)}$$

où $w_{ji}^I$ est le poids synaptique de la connexion de neurones d'entrée i au neurone caché j, $w_{1j}^H$ est le poids synaptique de la connexion du neurone caché j au neurone de sortie, $b_j^H$ et $b_1^O$ représentent respectivement les biais des neurones cachés j et du neurone de sortie.

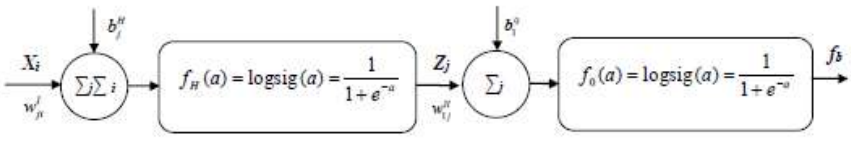



**Figure IV.4.** Representation shématique du réseau de neurones optimise.

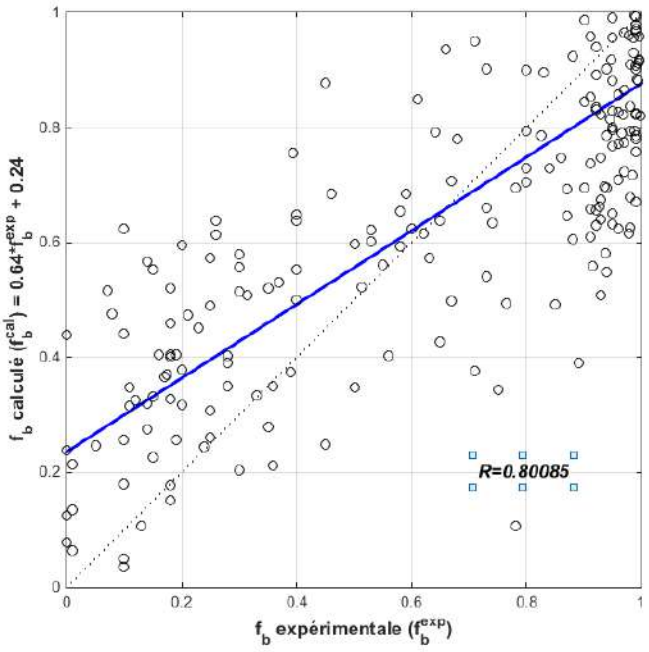
### 4.3.3. *Comparaison avec le modèle RLM*

L'équation du modèle RLM avec les paramètres statistiques obtenus sont comme suit :

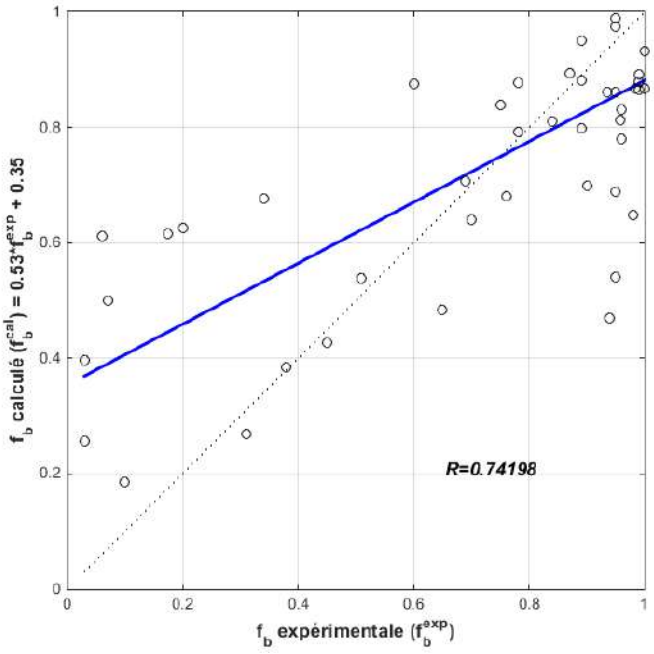
$f_b$=-0.6415+0.0314ESpm14x+0.1136ALOGP-0.2165Mor23p+0.2430Inflammat_80-0.1875GATS1v-0.0299Depressant_80-0.0911Neoplastic_50+0.0026TPSA(NO)-0.1879Mor26m+0.4519HATS0u (IV.5)

*REQM*= 0.211, $R^2$ = 0,641, $R^2_{Ajusté}$= 0,624 et *p*-value = 6.19 $10^{-42}$

La figure IV.5 présente la comparaison entre les valeurs experimental et les valeurs prédites par RLM.




**(a)**




**(b)**

**Figure IV.5.**Comparison entre les valeurs experimentale et les valeures calculés par la RLM :

**(a)** Base d'apprentissage, **(b)** base de test.

Afin d'établir le modèle neuronale développé comme une alternative au modèle RLM, une comparaison entre les deux approches a été faite en termes de l'erreur quadratique moyenne (*REQM*), l'erreur standard de prédiction (*ESP*) et le coefficient de corrélation (*R*) .

$$REQM = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_{i,\exp} - Y_{i,cal}\right)^2}{n}} \qquad \text{(IV.6)}$$

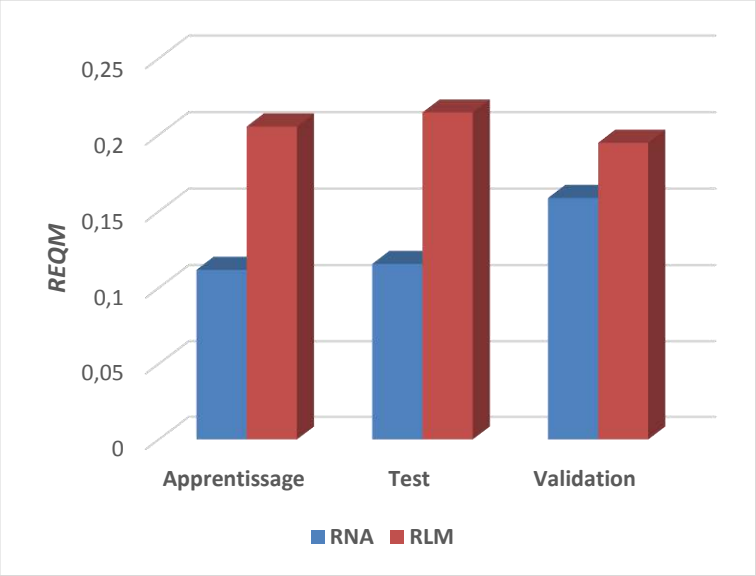
$$ESP(\%) = \frac{RMSE}{Y_e} \times 100 \qquad \text{(IV.7)}$$

Où $Y_{i,exp}$ est la valeur expérimentale du i$^{th}$ médicament; $Y_{i,pred}$ est la valeur prédite correspondant; $Y_e$ est la valeur moyenne des données expérimentales; n est le nombre d'échantillons.

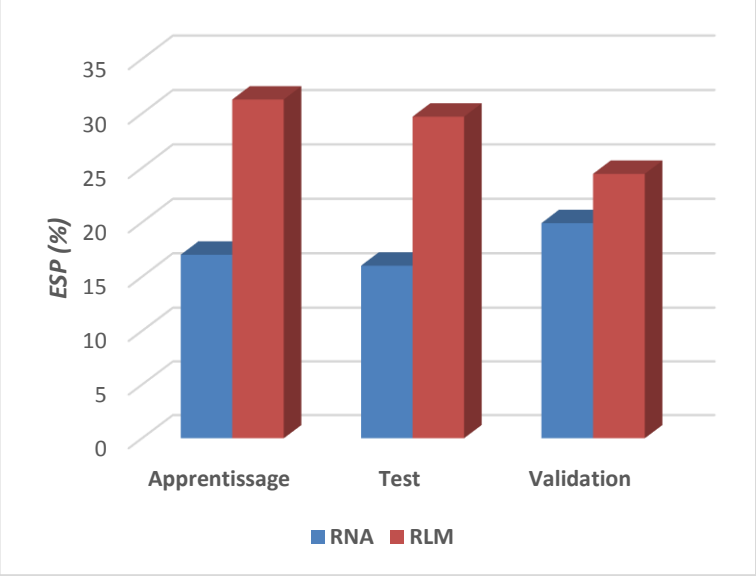
Le tableau IV.4 et la figure IV.6 montrent la comparaison entre le modèle neuronale et le modèle RLM en termes de la *REQM*, l'*ESP*, et le *R*. Le modèle neuronale développé dans ce travail a donné des erreurs plus faibles et *R* plus élevé que le modèle RLM.

**Tableau IV.4.** Comparaison entre le modèle neuronale (RNA) et le modèle RLM en terms de la *REQM*, l'*ESP* et le *R*.

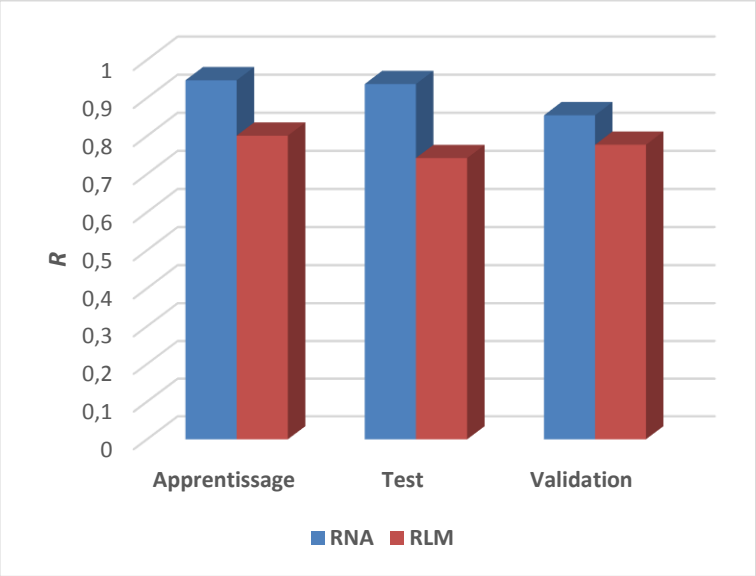
| **Model** | **Phase** | ***REQM*** | ***ESP*(%)** | ***R*** |
|---|---|---|---|---|
| **RNA** | Apprentissage | 0.1118 | 17.0152 | 0.9462 |
| | Test | 0.1157 | 16.0096 | 0.9363 |
| | Validation | 0.1588 | 19.9216 | 0.8544 |
| **RLM** | Apprentissage | 0.2055 | 31.2796 | 0.8009 |
| | Test | 0.2147 | 29.7047 | 0.7420 |
| | Validation | 0.1949 | 24.4463 | 0.7770 |




**(a)**




**(b)**




**(c)**

**Figure IV.6.** Histogrammes de comparaison entre le modèle neuronale (RNA) et le modèle RLM:

**(a)***REQM*, **(b)** *ESP*, **(c)***R*.

## IV.4. Conclusion

Dans ce travail, l'étude quantitative relation structure-activité (QSAR) d'une série de 277 fractions de médicament lié aux protéines plasmatiques ($f_b$) a été réalisée. Différents types de descripteurs ont été calculés à travers la structure moléculaire. Pour sélectionner les descripteurs respectifs, une procédure de sélection basée sur deux étapes (pré-traitement et régression stepwise) a été utilisée et a fourni dix descripteurs principaux.

Ensuite, deux modèles ont été développés le premier utilisant QSAR-RNA (10- [6] -1) ayant pour la validation REQM=0.1588, ESP(%)=19.9216 et R=0.8544. le deuxième QSAR-RLM ayant pour la validation toujours REQM=0.1949, ESP(%)=24.4463 et R=0.7770. Compte tenu de ces résultats, on peut conclure que le premier donne des résultats plus satisfaisants que le deuxième

# *Conclusion générale*

# *et perspectives*

La finalité de la recherche en chimie est la mise sur le marché de nouveaux produits ayant par rapport aux produits existants, au moins un avantage sur le plan de l'efficacité, de la tolérance ou du prix. Comment choisir et arriver à ces nouveaux produits ? En chimie thérapeutique par exemple, malgré des travaux considérables, il ne semble pas exister de procédé permettant de prévoir à coup sûr les actions pharmacologiques d'une substance de structure entièrement nouvelle. Seuls quelques processus généraux d'innovation sont à la base de la découverte de médicaments : le hasard, la chimie combinatoire, le criblage et l'inspiration à partir des structures de substances à activité biologique connue.

Les études des RQSA s'assignent comme objectif de relier l'activité biologique d'une série de composés homologues à leurs structures chimiques ou plutôt à leurs propriétés physico-chimiques. Ils permettent de

- ✓ Connaître les facteurs qui gouvernent l'activité,
- ✓ Prédire l'efficacité des composés synthétisés ou non synthétisés,
- ✓ Orienter la recherche vers de nouveaux composés actifs.

Les QSAR permettent ainsi de réduire le nombre de produits synthétisés, et par conséquent les coûts dépensés dans ce domaine. Elles se servent de certaines méthodes statistiques pour la modélisation des relations entre l'activité et la structure chimique des composés.

L'objectif de cette mémoire était de mettre au point un modèle statistique de prédiction du taux de fixation des médicaments aux protéines plasmatiques. Cette étude s'intègre dans le cadre de la quantification de ce taux sans recourir à l'expérimentation animale. Cette dernière, en plus du fait qu'elle est coûteuse et laborieuse, est décriée par une multitude d'organisations de défense des animaux. Pour atteindre cet objectif, les modèles QSAR ont été établis en utilisant deux méthodes d'apprentissage statistique : la régression linéaire multiple et les réseaux de neurones artificiels à 3 couches (une couche d'entrées, une couche cachée et une couche de sortie). Dans ce contexte, quatre axes de recherche ont été menés.

Le premier axe de notre travail concerne tout d'abord l'élaboration d'une base de données. Nous l'avons élaboré avec comme paramètre à modéliser le taux de fixation aux protéines plasmatiques. Au total, on a recueilli le taux de fixation aux protéines plasmatiques de 277 médicaments. Le choix des variables nécessaires à la description de $f_b$ a été fait sur la base du concept qui stipule que le phénomène à modéliser ($f_b$) est intimement lié à la structure moléculaire. A cette fin, nous avons calculé 1666 descripteurs moléculaires pour chaque médicament.

Le second axe a été l'élaboration d'un modèle de prédiction du taux de fixation aux protéines. 277 médicaments et 10 descripteurs sélectionnés comme étant les plus pertinents (c'est-à-dire ceux qui sont susceptibles de décrire au mieux le phénomène de fixation aux protéines plasmatiques) ont été utilisés pour développer deux modèles QSAR prédictifs, l'un linéaire et l'autre non linéaire. Dans un premier temps, Une modélisation par utilisation des réseaux de neurones (*PMC-RNA*) a été menée à l'effet de développer un modèle non linéaire. Une fois que le choix des conditions de travail (répartition de la base de données en ensemble d'apprentissage et de validation, algorithme d'apprentissage, fonctions d'activation de la couche cachée) eut été fait, un modèle neuronal a été élaboré. Les paramètres statistiques, le domaine d'applicabilité et le calcul de $f_b$ par utilisation de l'équation mathématique du modèle nous a permis de conclure que ce modèle est fiable et possède une capacité prédictive intéressante.

Dans un second temps, le modèle *RLM* obtenu pour la prédiction du taux de fixation aux protéines plasmatiques est un modèle linéaire. La concordance entre les valeurs prédites pour $f_b$ et les valeurs expérimentales, nous amène à conclure que le modèle *RLM* a un bon pouvoir prédictif du taux de fixation aux protéines plasmatiques.

Pour comparer à la fois les performances et la qualité de la prédiction des deux modèles obtenus dans ce travail (*RLM et PMC-RNA*), une évaluation statistique des deux modèles a été faite. Cette comparaison nous a permis de constater que la prédiction déterminée par le modèle *PMC-RNA* est meilleure que celle fournie par le modèle *RLM*. Une amélioration substantielle des paramètres statistiques pour le modèle *PMC-RNA* a été notée.

Aussi, on peut donc conclure que le modèle *PMC-RNA* (ayant pour la phase de validation : REQM =0.1588, ESP(%)=19.9216, R=0.8544) a un meilleur pouvoir prédictif que le modèle *RLM* (ayant pour la phase de validation : REQM=0.1949, ESP(%)=24.4463,R=0.7770) . Cela signifie que le modèle obtenu avec le réseau de neurones permet, dans une large mesure, de confirmer qu'une relation non linéaire relie le taux de fixation aux protéines plasmatiques à leur structure moléculaire.

En conclusion, tout en estimant avoir atteint l'objectif tracé pour cette étude, de nombreuses voies de recherche ont émergé. Ces perspectives concernent plus particulièrement la prédiction d'autres phénomènes biologiques. D'autre part, il faudrait s'intéresser à d'autres méthodes d'apprentissage statistique pour voir s'il est possible d'améliorer le modèle de prédiction du taux de fixation aux protéines plasmatiques.

# *Références bibliographiques*

**-A-**

**Agrawal V. K. et Khadikar P. V. (2001)**. "QSAR prediction of toxicity of nitrobenzenes." *Bioorganic & Medicinal Chemistry***9**(11): 3035-3040.

**Alberts DS, Chang SY, Chen HS, Evans TL et TE. M. (1979)**. "Oral melphalan kinetics." *Clinical Pharmacology & Therapeutics***26**(6): 737-745.

**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.

**Anderson P. O., Knoben J. E. et Troutman W. G. (2002)**. Handbook of Clinical Drug Data, McGraw-Hill Medical Pub. Division.

**Atiba J. O., Taylor G., Pershe R. A. et Blaschke T. F. (1987)**. "Plasma antipyrine half-life can be determined from urine data." *British Journal of Clinical Pharmacology***23**(6): 715-719.

**-B-**

**Bachinger T. et Mandenius C. F. (2000)**. "Searching for process information in the aroma of cell cultures." *Trends in Biotechnology***18**(12): 494-500.

**Balaban A.T. (1982).** " Highly Discriminating Distance-Based Topological Index", *Chem. Phys. Lett.* 89: 399-404.

**Basheer I. A. et Hajmeer M. (2000)**. "Artificial neural networks: Fundamentals, computing, design, and application." *Journal of Microbiological Methods* **43**(1): 3-31.

**Baughman D. R. et Liu Y. A. (1995)**. Neural networks in bioprocessing and chemical engineering. Florida, USA, Academic Press.

**Becker T., Enders T. et Delgado A. (2002)**. "Dynamic neural networks as a tool for the online optimization of industrial fermentation." *Bioprocess and Biosystems Engineering***24**(6): 347-354.

**Bellifa K. (2015)**. Etude des relations quantitatives structure–toxicité des composés chimiques à l'aide des descripteurs moléculaires, Université Abou Beker Belkaid deTlemcen.

**Bermejo S. et Cabestany J. (1999)**. On-line gradient learning algorithms for K-nearest neighbor classifiers. Foundations and tools for neural modeling, *Springer Berlin Heidelberg* **1606:** 546-555.

**Bermejo S. et Cabestany J. (2001)**. "Learning with nearest neighbour classifiers." *Neural Processing Letters***13**(2): 159-181.

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

**Brion F. (2003)**. Les médicaments en périnatologie, Masson.

**Brown A.C. et Fraser T.R. (1868).** "On the connection between chemical constitution and Physiological Action ; with special reference to the physiological action of the salts of the ammonium bases derived from Strychnia, Brucia, Thebaia, Codeia, Morphia, and Nicotia ", *J .Anat. Physiol.* 2: 224-242.

**-C-**

**Carpenter G. A. et Grossberg S. (1987)**. "ART 2: self-organization of stable category recognition codes for analog input patterns." *Applied optics***26**(23): 4919-4930.

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

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

**Cours P. (2016)**. "Relations entre Pharmacocinétique et Pharmacodynamie."

**-D-**

**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.

**Davalo É. et Naïm P. (1989)**. Des réseaux de neurones. Paris, France, Eyrolles.

**Demuth H. et Beale M. (2000)**. Neural network toolbox for use with Matlab. User's Guide version 4. USA, The MathWorks Inc.

**Desai K. M., Akolkar S. K., Badhe Y. P., Tambe S. S. et Lele S. S. (2006)**. "Optimization of fermentation media for exopolysaccharide production from *Lactobacillus plantarum* using artificial intelligence-based techniques." *Process Biochemistry***41**(8): 1842-1848.

**Desai K. M., Survase S. A., Saudagar P. S., Lele S. S. et Singhal R. S. (2008)**. "Comparison of artificial neural network (ANN) and response surface methodology (RSM) in fermentation media optimization: Case study of fermentative production of scleroglucan." *Biochemical Engineering Journal***41**(3): 266-273.

**Di Massimo C., Montague G. A., Willis M. J., Tham M. T. et Morris A. J. (1992)**. "Towards improved penicillin fermentation via artificial neural networks." *Computers & Chemical Engineering* **16**(4): 283-291.

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

**Duprat A.F., Huynh T. et Dreyfus G. (1998).** " Toward a principled methodology for neural network design and performance evaluation in QSAR : Application to the prediction of LogP". *J. Chem. Inf. Comput. Sci.* 38: 586-594.

**--F-**

**Fourcade O., Geeraerts T., Minville V. et Samii K. (2014)**. Traité d'anesthésie et de réanimation (4° Éd.), Lavoisier.

**Fourches D. et Varnek A. (2008).** " Critical assessment of QSAR models of environmental toxicity against Tetrahymena-pyriformis : focusing on applicability domain and overfitting by variable selection ". *J. Chem.Inf. Model.* 48: 1733.

**-G-**

**Galván I. M., Zaldívar J. M., Hernández H. et Molga E. (1996)**. "The use of neural networks for fitting complex kinetic data." *Computers & Chemical Engineering***20**(12): 1451-1465.

**Gasteiger J. et Zupan J. (1993).**" Neural Networks in Chemistry ", Angew. Chem. Int. Ed. Engl. 32, 503-527

**Geneviève U.-S. (2007)**. Pharmacocinétique et vecteur. cours.205.

**Ghasemi J., Saaidpour S. et Brown S.D. (2007).** "QSPR study for estimation of acidity constants of some aromatic acids derivatives using multiple linear regression (MLR) analysis " . J. Mol. Struct. (T heochem) 805, 27-32.

**Glassey J., Ignova M., Ward A. C., Montague G. A. et Morris A. J. (1997)**. "Bioprocess supervision: neural networks and knowledge based systems." *Journal of Biotechnology***52**(3): 201-

**Goodman L. S., Brunton L. L., Blumenthal D. K., Murri N. et Hilal-Dandan R. (2011)**. Goodman & Gilman's the pharmacological basis of therapeutics, 12th Edition. New York, McGraw-Hill Medical.

**Goodman L. S., Gilman A., Brunton L. L., Lazo J. S. et Parker K. L. (2006)**. Goodman & Gilman's the pharmacological basis of therapeutics, 11th Edition. New York, McGraw-Hill.

**Goodman L. S., Gilman A., Hardman J. G., Gilman A. G. et Limbird L. E. (1996)**. Goodman & Gilman's the pharmacological basis of therapeutics, 9th Edition. New York, McGraw-Hill, Health Professions Division.

**Grover M., Singh B., Bakshi M. et Singh S. (2000).** " Quantitative structure–property relationships in pharmaceutical " research–Part 1 Pharm. Sci. Tech. Today 3, 28-35.

**Guergouri F. Z.** Absorption des Medicaments. Département de Médecine. F. d. Médecine, Module de pharmacologie générale.

**Gurunathan B. et Sahadevan R. (2011)**. "Design of experiments and artificial neural network linked genetic algorithm for modeling and optimization of L-asparaginase production by *Aspergillus terreus* MTCC 1782." *Biotechnology and Bioprocess Engineering***16**(1): 50-58.

**-H-**

**Haider M., Pakshirajan K., Singh A. et Chaudhry S. (2008)**. "Artificial neural network-genetic algorithm approach to optimize media constituents for enhancing lipase production by a soil microorganism." *Applied Biochemistry and Biotechnology***144**(3): 225-235.

**Hamadache M., Khaouane L., Benkortbi O., Si Moussa C., Hanini S. et Amrane A. (2014)**. "Prediction of Acute Herbicide Toxicity in Rats from Quantitative Structure–Activity Relationship Modeling." *Environmental Engineering Science***31**(5): 243-252.

**Hammett L.P. (1937).** "The effect of structure upon the reactions of organic compounds. Benzene derivatives ". *J. Am. Chem. Soc,* 59: 96-103.

**Hansch C., Hansch A. et Hoekmann D. (1995).** " Exploring QSAR: hydrophobic, electronic and steric constants ".Washington, DC: Am. Chem. Soc.

**Hassoun M. H. (1995)**. Fundamentals of artificial neural networks. Massachusetts, USA, Mit Press.

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

**Hebb D. O. (1949)**. The organization of behavior: A neuropsychological theory. New York, USA, Wiley.

**He L. et Jurs P.C. (2005).** " Assessing the reliability of a QSAR model's predictions ". *J. Mol. Graph. Model* 23:503-523.

**Helguera A. M., Combes R. D., González M. P. et Cordeiro M. N. D. S. (2008)**. "Applications of 2D descriptors in drug design: a DRAGON tale." *Current Topics in Medicinal Chemistry***8**(18): 1628-1655.

**Hopfield J. J. (1982)**. "Neural networks and physical systems with emergent collective computational abilities." *Proceedings of the National Academy of Sciences***79**(8): 2554-2558.

**Hornik K. (1991)**. "Approximation capabilities of multilayer feedforward networks." *Neural Networks***4**(2): 251-257.

**Huang J., Shimizu H. et Shioya S. (2002)**. "Data preprocessing and output evaluation of an autoassociative neural network model for online fault detection in virginiamycin production." *Journal of Bioscience and Bioengineering***94**(1): 70-77.

**-I-**

**Ignova M., Montague G. A., Ward A. C. et Glassey J. (1999)**. "Fermentation seed quality analysis with self-organising neural networks." *Biotechnology and Bioengineering***64**(1): 82-91.

**-J-**

**Jacques D. (2006)**. P h a rma c o l o g i e g e n e r a l e. Université Victor Segalen Bordeaux 2.

**Jaworska J., Jeliazkova N.N. et Aldenberg T. (2005).** " QSAR applicabilty domain estimation by projection of the training set descriptor space": A review. Altern. Lab. Anim, 33, 445-459.

**Jin S., Ye K., Shimizu K. et Nikawa J. (1996)**. "Application of artificial neural network and fuzzy control for fed-batch cultivation of recombinant *Saccharomyces cerevisiae*." *Journal of Fermentation and Bioengineering***81**(5): 412-421.

**-K-**

**Kang S. G., Lee D. H., Ward A. C. et Lee K. J. (1998)**. "Rapid and quantitative analysis of clavulanic acid production by the combination of pyrolysis mass spectrometry and artificial neural network." *Journal of microbiology and biotechnology* **8**(5): 523-530.

**Karelson M. (2000).**" Molecular Descriptors in Qsar/Qspr". John Wiley & Sons Inc, New York

**Karim M. N., Yoshida T., Rivera S. L., Saucedo V. M., Eikens B. et Oh G. S. (1997)**. "Global and local neural network models in biotechnology: Application to different cultivation processes." *Journal of Fermentation and Bioengineering***83**(1): 1-11.

**Katritzky A.R.D., Fara C., Petrukhin R.O., Tatham D.B., Maran U., Lomaka A. et Karelson M. (2002).** "The present utility and future Potential for medicinal chemistry of QSAR/QSPR with whole molecule descriptors ", Cur r. T o p. Med. Chem 2,1333-1356

**Katritzky A. R., Kuanar M., Slavov S. et Hall C. D. (2010).** "Quantitative Correlation of Physical and Chemical Properties with Chemical Structure: Utility for Prediction, Chem ". Rev. 110, 5714–5789

**Keedwell E. et Narayanan A. (2005)**. Intelligent bioinformatics: The application of artificial intelligence techniques to bioinformatics problems. Chichester, UK, Wiley.

**Kennedy M., Prapulla S. G. et Thakur M. S. (1992)**. "Designing fermentation media: A comparison of neural networks to factorial design." *Biotechnology Techniques***6**(4): 293-298.

**Kennedy M. et Spooner N. (1996)**. "Using fuzzy logic to design fermentation media: A comparison to neural networks and factorial design." *Biotechnology Techniques***10**(1): 47-52.

**Khaouane L., Benkortbi O., Hanini S. et Si-Moussa C. (2013)**. "Modeling of an industrial process of pleuromutilin fermentation using feed-forward neural networks." *Brazilian Journal of Chemical Engineering***30**(1): 105-116.

**Kier L.B., Hall L.H. et Murray W.J. (1975).** " Molecular connectivity I: Relationship to nonspecific local anesthesia ". J. Pharm. Sci.64, 1971–1974

**Kohonen T. (1982)**. "Self-organized formation of topologically correct feature maps." *Biological Cybernetics***43**(1): 59-69.

**Kovarova-Kovar K., Gehlen S., Kunze A., Keller T., von Daniken R., Kolb M. et van Loon A. P. G. M. (2000)**. "Application of model-predictive control based on artificial neural networks to optimize the fed-batch process for riboflavin production." *Journal of Biotechnology***79**(1): 39-52.

**Kwan K. (1997)**. "Oral bioavailability and first-pass effects." *Drug Metab Dispos***25**(12): 1329-1336.

**-L-**

**Labrecque G. et Sirois-Labrecque M. (2003)**. Chronopharmacologie: rythmes biologiques et administration des médicaments, Presses de l'Université de Montréal.

**Lafaurie M., Franche-Comté U. d. et Neuchâtel U. d. (2002)**. Le temps du vivant, Presses Universitaires Franc-Comtoises.

**Lechat P. (2006)**. Pharmacologie. Cours de pharmacologie, Université Pierre et Marie Curie.

**Linko S., Zhu Y. H. et Linko P. (1999)**. "Applying neural networks as software sensors for enzyme engineering." *Trends in Biotechnology***17**(4): 155-162.

**Loichot C. et Grima M. ( 2004)**. Mécanismes d'action des médicaments. cours de Pharmacologie Générale, Faculté de Médecine de Strasbourg.

**-M-**

**Médecine et vie e. (2012)**. "Liaison des médicaments aux protéines plasmatiques."

**Minsky M. L. et Papert S. (1969)**. Perceptrons: An introduction to computational geometry. Massachusetts, USA, MIT Press.

**Montague G. et Morris J. (1994)**. "Neural-network contributions in biotechnology." *Trends in Biotechnology***12**(8): 312-324.

**Moody J. et Darken C. J. (1989)**. "Fast learning in networks of locally-tuned processing units." *Neural computation***1**(2): 281 - 294.

**Moulin M. et Coquerel A. (2002)**. Pharmacologie, Masson.

**-N-**

**N. Margossian. (2007).** Le règlement REACH - La règlementation européenne sur les produits chimiques, Dunod / L'usine Nouvelle, Paris.

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

**Nasr N., Hafez H., El Naggar M. H. et Nakhla G. (2013)**. "Application of artificial neural networks for modeling of biohydrogen production." *International Journal of Hydrogen Energy***38**(8): 3189-3195.

**Nelofer R., Ramanan R., Rahman R., Basri M. et Ariff A. (2012)**. "Comparison of the estimation capabilities of response surface methodology and artificial neural network for the optimization of recombinant lipase production by *E. coli* BL21." *Journal of Industrial Microbiology & Biotechnology***39**(2): 243-254.

**Nicoletti M. C. (2009)**. Computational intelligence techniques for bioprocess modelling, supervision and control. Düsseldorf, Germany, Springer-Verlag.

**-P-**

**Pharmacie.ma. (2006)**. "Le métabolisme des médicaments." cours

**pharmacocinétique c. (2005)**. "Biotransformation des Toxiques."

**pharmacorama. (2016)**. "Le médicament dans le sang." CdM Editions.

**Pharmaomédicale. (2016)**. " cours du site du collège national de pharmacologie médicale."

**-R-**

**Randic M. (1975).** " Characterization of molecular branching ". J. Am. Chem. Soc. 97, 6609–6615, J. Am. Ch em. Soc.97, 6609-6615.

**Rosenfeld G. C. et Loose D. S. (2007)**. Pharmacology, Lippincott Williams & Wilkins.

**Rowe P. (2012)**. Pharmacokinetics, Bookboon.

**Rumelhart D. E. et McClelland J. L. (1986)**. Parallel distributed processing: Explorations in the microstructure of cognition. Foundations. Massachusetts, USA, Mit Press.

**-S-**

**Saraceno A., Curcio S., Calabro V. et Iorio G. (2009)**. "A hybrid neural approach to model batch fermentation of *ricotta cheese whey* to ethanol." *Comput. Chem. Eng.***34**(10): 1590-1596.

**Schenker B. et Agarwal M. (1998)**. "Predictive control of a bench-scale chemical reactor based on neural-network models." *Control Systems Technology, IEEE Transactions on***6**(3): 388-400.

**Shene C., Diez C. et Bravo S. (1999)**. "Neural networks for the prediction of the state of *Zymomonas mobilis* CP4 batch fermentations." *Computers & Chemical Engineering***23**(8): 1097-1108.

**Shimizu H., Yasuoka K., Uchiyama K. et Shioya S. (1998)**. "Bioprocess fault detection by nonlinear multivariate analysis: Application of an artificial autoassociative neural network and wavelet filter bank." *Biotechnology Progress***14**(1): 79-87.

**Si-Moussa C. (2009)**. Modélisation des équilibres de phases a pressions élevées par les réseaux de neurones artificiels, Ecole Nationale Supérieure Polytechnique

**Stanke-Labesque F. (2012)**. Chapitre 2 : Aspects pharmacocinétiques. cours de pharmacocinétique, Université Joseph Fourier de Grenoble.

**Stone M. (1977)**. "An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion." *Journal of the Royal Statistical Society B***39**(1): 44-47.

**-T-**

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

**Tetko I.V., Sushko I., Pandey A.K., Zhu H., Tropsha A., Papa E., Oberg T., Todeschini R.,**

**Tetko I.V., Villa A.E.P. et Livingstone D.J. (1996).** "An Enhancement of Generalization Ability in Cascade Correlation Algorithm by Avoidance of Overfitting/Overtraining", J. Chem. Inf. Comput. Sci.36, 794-803.

**Todeschini R, Consonni V. (2000).** Handbook of molecular descriptors. In : Mannhold R, Kubinyi H, Timmerman H (eds). WILEY VCH, Weinheim.

**Toutain P. L. et Bousquet-Melou A. (2004)**. "La biodisponibilité et son évaluation." *European Medicines Evaluation Agency*.

**Tropsha A., Gramatica P. et Gombar V.K. (2003).** " The importance of being Earnest: Validation is the absolute essential for successful application and interpretation of QSPR models ". QSAR Comb.Sci. 22, 69-77.

**Turner J. V., Maddalena D. J. et Cutler D. J. (2004)**. "Pharmacokinetic parameter prediction from drug structure using artificial neural networks." *International Journal of Pharmaceutics***270**(1–2): 209-219.

**Tzanakou E. M. (2000)**. Supervised and unsupervised pattern recognition: Feature extraction and computational intelligence. London, UK, Taylor & Francis Group.

**-U-**

**UGent M. D. (2012)**. "Manuel de psychopharmacothérapie." *Medicine and Health Sciences*: XXIV, 568 pages.

**Urbani D., Roussel-Ragot P., Personnaz L. et Dreyfus G. (1994)**. The selection of neural models of nonlinear_dynamical systems by statistical tests. Neural Networks for Signal Processing [1994] IV. Proceedings of the 1994 IEEE Workshop.

**-W-**

**Weber C, Gasser R et G. H. (1999)**. "Absorption, excretion, and metabolism of the endothelin receptor antagonist bosentan in healthy male subjects." *Drug Metabolism & Disposition***27**(7): 810-815.

**Wiener H. (1947).** "Structural determination of paraffin boiling points". J. Am. Chem. Soc. 69, 17–20.

**wikipedia (2016)**.
https://fr.wikipedia.org/wiki/Biodisponibilit%C3%A9_%28m%C3%A9decine%29. **2016**. (Consulté le 12/04/2016)

**www.drugbank.ca**. (Consulté le 12/02/2016)

**www.vcclab.org** (Consulté le 20/03/2016)

**-Y-**

**Ye K., Jin S. et Shimizu K. (1998)**. "On the development of an intelligent control system for recombinant cell culture." *International Journal of Intelligent Systems***13**(6): 539-560.

**Yuan J. Q. et Vanrolleghem P. A. (1999)**. "Rolling learning-prediction of product formation in bioprocesses." *Journal of Biotechnology***69**(1): 47-62.

**-Z-**

**Zhang F., Xue J., Shao J. et Jia L. (2012)**. "Compilation of 222 drugs' plasma protein binding data and guidance for study designs." *Drug Discovery Today***17**(9–10): 475-485.