

الجمهورية الجزائرية الديمقراطية الشعبية

République Algérienne Démocratique et Populaire

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR

ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITÉ ABOU BEKR BELKAÏD
FACULTÉ DE MÉDECINE

DR. B. BENZERDJEB - TLEM CEN



وزارة التعليم العالي
والبحوث العلمي

جامعة أبو بكر بلقايد
كلية الطب

د. ب. بن زرجب - تلمسان

DÉPARTEMENT DE PHARMACIE

MÉMOIRE DE FIN D'ÉTUDES POUR
L'OBTENTION DU DIPLÔME DE DOCTEUR EN PHARMACIE

THÈME :

Modélisation Moléculaire : Classification Et Intérêt Dans La Conception Des
Médicaments

Présenté par :

Madame HAMMOUMI Ferial

Soutenu le 16-10-2021

Devant le jury composé de :

Président :

Pr BENSAID Okkacha

Professeur en chimie Faculté de science
Université de Tlemcen

Membres :

Dr HAMZI Imane

Maitre de conférences B en chimie organique
Université de Tlemcen

Dr LOUZIM Habiba

Maitre assistante hospitalo-universitaire en chimie thérapeutique
Université de Tlemcen

Encadreur :

Dr BEGHADADI Sara El Mansouria

Maitre assistante hospitalo-universitaire en chimie thérapeutique
Université de Tlemcen

Remerciement

Avant tout je remercie notre grand Dieu tout puissant pour exprimer ma reconnaissance envers sa grande générosité. Dieu m'a donné la volonté, la patience, la santé et la confiance durant toutes mes années d'études, ce qui ma donné le pouvoir de réaliser ce modeste travail.

Je tiens à exprimer ma profonde gratitude et sincères remerciements à Madame Dr BEGHDAI Sara El Mansouria, qui m'a encadré, tout au long de ce mémoire. Je vous remercie pour votre précieuse présence assistance, votre disponibilité et l'intérêt que vous avez manifesté pour ce modeste travail. Je vous remercie pour vos orientations et votre enthousiasme envers mon travail. Les judicieux conseils et rigueur que vous m'avez prodigué tout au long de ce travail. J'ai pris un grand plaisir à travailler avec vous.

J'adresse tout particulièrement mes remerciements à Monsieur BENSALD Okkacha, pour tous les efforts déployés pour notre formation pendant la troisième année. Je vous remercie très chaleureusement pour votre gentillesse d'avoir accepté de présider le jury qui examinera ce mémoire.

Je tiens également à exprimer mes vifs remerciements à Madame HAMZI Imane et Madame LOUZIM Habiba, pour sa contribution à ma formation et son acceptation de discuter et d'évaluer ce modeste travail.

Enfin, je ne saurai oublier de remercier tous mes enseignants du département de Pharmacie, qui m'ont accompagnés et aidés à m'améliorer durant mon cursus de formation.

Dédicace

C'est avec une grande émotion et un immense plaisir que je dédie ce modeste travail :

A la mémoire de mon cher père

A ma chère mère pour son encouragement et son soutien et l'aide qu'elle m'a apporté tout au long de ce travail.

A mes très chères sœurs Fedwa et Meriem et mon frère Ibrahim

A Ibtissem, ma chère tante qui m'a aidé énormément pour réaliser ce travail

A tout les membres de ma famille, petits et grands, cousins et cousines

Aux membres de ma promotion, qui de près ou de loin m'ont soutenue moralement durant ce mémoire.

A mon cher père

رحمه الله و تخفر له و أسكنه الجنة

Liste des figures

CHAPITRE 1

Figure 1.1 – Modélisation moléculaire comme outil informatique.(5).	04
Figure 1.2 – La représentation de Lewis de l'atome d'oxygène.(7).	06
Figure 1.3 – Différentes phases du processus de découverte de médicaments. (24).	14
Figure 1.4 –Processus (simplifié) de recherche de médicaments découpé en 4 grandes étapes.(22)	14
Figure 1.5 – Aventure du médicament.(30).	17
Figure 1.6 –Evolution et comparaison internationale des coûts en R&D pharmaceutiques.(13).	19
Figure 1.7 – Schéma récapitulatif des méthodes de modélisation insérées dans la phase de R&D pharmaceutique. (1).	20
Figure 1.8 - Organigramme des processus CADD. (2).	21
Figure 1.9 – Domaines d'impact de la modélisation biologique dans le processus de R&D pharmaceutique.(40)	22

CHAPITRE 2

Figure 2.1 - Illustration de la CADD.(46).	27
Figure 2.2 - Premiers exemples de la SBDD.(52).	29
Figure 2.3 - Diagramme de flux de travail du processus de conception de médicaments basée sur la structure (SBDD).(57).	35
Figure 2.4 – Evolution dans le temps du nombre de séquences de protéines disponibles et de structures protéiques résolues durant les 20 dernières années.(61).	36
Figure 2.5 – Algorithme représentant les étapes de réalisation de la modélisation homologue.(4).	39
Figure 2.6 – Principales étapes de la modélisation homologue des protéines.(50).	40
Figure 2.7 - Aperçu du processus de modélisation homologue et ses applications dans la découverte de médicaments.(71).	44
Figure 2.8 - Différence entre la modélisation par homologie et la modélisation de novo.(76).	48
Figure 2.9 - Aperçu du processus de Docking moléculaire. (77).	49

LISTE DES FIGURES

Figure 2.10 - Deux modèles de Docking moléculaire. (A) Un modèle de type "lock-and-key" (B) Modèle d'ajustement induit. (79)	51
Figure 2.11 – Principe général du Docking moléculaire avec les différents logiciels utilisés.(81).	51
Figure 2.12 - Vue d'ensemble schématique de méthode de Docking inverse. (89).	55
Figure 2.13 - Diagrammes schématiques comparant (A) Le Docking conventionnel et (B) Le Docking inverse.(92).	57
Figure 2.14 - Principales applications du Docking/Docking inverse moléculaire dans la découverte actuelle de médicaments. (91).	58
Figure 2.15 - Evolution des simulations de dynamique moléculaire en biologie structurale.(96)..	60
Figure 2.16 -Dynamique moléculaire et ses applications.(95).	63
Figure 2.17 -Exemple graphique de différentes représentations moléculaires d'une même structure (ibuprofène, ici représentée comme une structure 2D)(113).	72
Figure 2.18 - Représentation des descripteurs moléculaires utilisés dans la modélisation QSAR.(116).	76
Figure 2.19 - Principales étapes d'une étude QSAR en décrivant l'importance cruciale des descripteurs moléculaires.(112).	76
Figure 2.20 - Flux de travail typique des méthodes QSAR.(117).	78
Figure 2.21 -Principales étapes de la mise au point et de l'utilisation de QSAR. (113) . . .	79
Figure 2.22 - Quelques médicaments découverts à l'aide de la QSAR. (a) Zolmitriptan; (b) Norfloxacin; (c) Losartan.(86).	84
Figure 2.23 -Flux de travail pour la modélisation des pharmacophores.(130)	89
Figure 2.24 -Importance de la modélisation de pharmacophore dans le processus du Drug Design.(127).	90
Figure 2.25 - Aperçu chronologique du nombre de publications parues au cours des dernières années, correspondant aux requêtes "pharmacophore" et "criblage virtuel".(123)..	91
Figure 2.26 - Empreintes digitales 2D comme descripteurs de ligands. (128).	94

CHAPITRE 3

Figure 3.1- Evolution du nombre de publications utilisant différentes méthodes d'intelligence artificielle dans découverte de médicaments pendant la période [2008-2018].(138).....	100
Figure 3.2- Collaboration de l'esprit et la machine dans la découverte des médicaments. (135).....	100
Figure 3.3- Exemples d'applications de "Cloud computing" pour des projets de modélisation moléculaire.(146).....	108
Figure 3.4- Représentation des intérêts des « Cloud computing » dans la découverte de médicaments.(145).....	109

Liste des tableaux

CHAPITRE 2

Tableau II.1 – Exemples de médicaments découverts par les méthodes SBDD.(57)..... **31**

Liste des abréviations

2D : Représentation bidimensionnelle

3D : Représentation tridimensionnelle

ADMET : Absorption, Distribution, Métabolisme, Excrétion et la toxicité

ADN : Acide désoxyribonucléique

AGP : Analyse en grappes

AMD: Découverte Moléculaire Accélérée

ARN : Acide ribonucléique

BLAST : Basic Local Alignment Search Tool : outil de recherche d'alignement local de base

CADD : Computer aided drug design : Conception de médicaments assistée par ordinateur

CK1δ: Casein Kinase I isoform delta

CCR5 : Corécepteur Cellulaire 5 : chimiokine

CPK : les modèles "bâtons" de Dreiding ou les modèles "remplissant l'espace" de Corey, Pauling et Koltun

CoMFA: Comparative Molecular Field Analysis

CoMSIA: Comparative Molecular Similarity Indices Analysis

CoMMA : Analyse Comparative Des Moments Moléculaires

DARPA: Defense Advanced Research Projects Agency

DL: Deep Learning

DM: Dynamique moléculaire

DYLOMMS: Dynamic Lattice Oriented Molecular Modeling System

FASTA: Fast Alignement

FDA: Food and Drug Administration

Fs: Femtoseconde

GOLD: Genetic Optimization for Ligand Docking

GPR30 : Récepteur couplé à la protéine G 30

GPU: Unité de Traitement Graphique

GRIND : descripteurs indépendants de la grille

HDAC8: Histone désacétylase 8

HPC : Calcul Parallèle A Haute Performance

HTS: High throughput Screening: Criblage à haut débit

HTD: High Throuthput Docking: Docking à haut debit

HRV2: Human rhino-virus de type 2

LISTE DES ABREVIATIONS

I3C : indole-3-carbinol

IA: Intelligence artificielle

IA/M2 : Influenza A / Protéine de matrice M2

InhA : Inibine alpha

INVDOCK : Procédure de Docking inverse

IPP : Interaction protéine-protéine

IT: Indices topologiques

IUPAC: International Union of Pure and Applied Chemistry

LBDD : Ligand based drug design : Conception de médicaments basée sur le ligand

ML: Machine Learning

MLR : Régression Linéaire Multiple

MM : Mécanique moléculaire

Mpro : Main Protease : Protéase principale

Ns : Nanoseconde

PCA: Analyse en Composantes Principales

PCR : Régression en Composantes Principales

PDB : Protein data bank : Banque de données des protéines

PDTD: Potential Drug Target Database

PFKFB3: 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3

PR : Protéase rétrovirale

QSAR : Quantitative structure-activity Relationship : relation quantitative de structure-activité

QSPR : relation quantitative structure-propriété

QSTR : relation quantitative structure-toxicité

R&D : Recherche et développement

RCPG : Récepteurs couplés à la protéine G

RD : Reverse Docking

RMN : Résonance magnétique nucléaire

SAR : Relation structure-activité

SARS-Cov : Syndrome respiratoire aigu sévère- coronavirus

SAVES : Serveur d'analyse et de vérification structurelle

SBDD : Structure based drug design : Conception de médicament basée sur la structure

SBVS : Criblage virtuel basé sur la structure

TBM : Template-based modeling : modélisation basée sur un modèle (par homologie)

TGEV : Virus de la gastro-entérite transmissible

LISTE DES ABREVIATIONS

TTD: Therapeutic Target Database

VIH-1 : Virus de l'immunodéficience humaine type 1

VS: Virtual screening: criblage virtuel

vs: Versus (comparaison)

Modélisation moléculaire : Classification et intérêt dans la
conception des médicaments

PLAN

Introduction générale.....	1
I. CHAPITRE 1: Généralités	3
I.1 Introduction	4
I.2 Définition.....	5
I.3 Impact de l'évolution de l'informatique sur la modélisation	9
I.3.1 Naissance de la chimie-informatique	9
I.3.2 Développement de la modélisation moléculaire	9
I.3.3 Evolution actuelle.....	10
I.3.4 L'ordinateur apporte-t-il une dimension supplémentaire pour la modélisation moléculaire ?.....	11
I.4 Importance dans le processus de recherche pharmaceutique	13
I.4.1 Médicament: Définition	13
I.4.2 Recherche de médicaments : « Drug design »ou « conception de médicaments » 13	
I.4.2.1 Recherche et Développement (R&D).....	15
I.4.2.1.1 Identification et validation de cibles.....	15
I.4.2.1.2 Génération de Hits et de Leads.....	15
I.4.2.1.3 Optimisation de Leads	16
I.4.2.2 Essais précliniques, cliniques et commercialisation.....	16
I.4.3 Coût de l'innovation thérapeutique.....	18
I.4.4 Insertion de la modélisation moléculaire durant le processus de recherches pharmaceutiques (Drug design)	19
I.4.5 Etudes encore plus récentes... De la modélisation moléculaire vers la modélisation biologique.....	22
I.4.5.1 Modèles de signalisation cellulaire.....	23
I.4.5.2 Modèles de signalisation et de réponse cellulaire	23
I.4.5.3 Modèles physiologiques	24
I.5 Conclusion	25

II. CHAPITRE 2: Classification simple des méthodes26

II.1	Introduction	27
II.2	« STRUCTURE –BASED DRUG DESIGN » (SBDD)	28
II.2.1	Historique	29
II.2.2	Développement de l’approche.....	30
II.2.3	Aperçu du processus.....	32
II.2.4	Principales méthodes utilisées.....	36
II.2.4.1	Prédiction de structure de la protéine.....	36
II.2.4.1.1	Modélisation par homologie.....	38
II.2.4.1.1.1	Concept d« homologie entre les protéines ».....	39
II.2.4.1.1.2	Principales étapes	40
II.2.4.1.1.3	Applications dans la découverte des médicaments.....	43
II.2.4.1.1.4	Exemple d’application sur le SARS-Cov	44
II.2.4.1.2	Modélisation de novo : prédiction « ab initio »	45
II.2.4.1.3	Principe général.....	45
II.2.4.1.3.1	Exigences pour une modélisation de Novo réussie	45
II.2.4.1.3.2	Logiciels et méthodes	46
II.2.4.1.3.3	Limites des méthodes « ab initio »	47
II.2.4.1.4	« ab initio » vs Homologie	48
II.2.4.2	« DOKING » Moléculaire.....	49
II.2.4.2.1	Définition	50
II.2.4.2.2	Théorie de base.....	50
II.2.4.2.3	Logiciels et algorithmes	52
II.2.4.2.4	Fonctions de scores	53
II.2.4.2.5	Classification des méthodes utilisées	53
II.2.4.2.5.1	Docking rigide : « Lock and Key »	54
II.2.4.2.5.2	Docking semi-flexible	54
II.2.4.2.5.3	Docking flexible: Modèle d'ajustement induit.....	54
II.2.4.2.6	Docking inverse ou « Reverse Docking »(RD).....	55
II.2.4.2.7	Docking conventionnel vs Docking inverse.....	56
II.2.4.2.8	Applications du Docking / RD dans la découverte des médicaments.....	57
II.2.4.3	Dynamique moléculaire (DM).....	60
II.2.4.3.1	Histoire de la DM.....	60

II.2.4.3.2	Qu'est-ce que la DM ?	61
II.2.4.3.3	Principe général.....	61
II.2.4.3.4	Calcul de la DM	63
II.2.4.3.5	Simulations combinées de Docking et de la DM	64
II.2.4.3.6	Limites de la DM.....	64
II.2.4.3.7	Comment la DM peut-elle contribuer à la découverte des médicaments et la SBDD ?	65
II.3	« LIGAND –BASED DRUG DESIGN » (LBDD)	67
II.3.1	Principales méthodes utilisées.....	68
II.3.1.1	Approche quantitative de la relation structure-activité (QSAR).....	68
II.3.1.1.1	Historique de la QSAR/ QSPR.....	68
II.3.1.1.2	Définition	69
II.3.1.1.3	Hypothèse de base de QSAR	69
II.3.1.1.4	Principe général.....	69
II.3.1.1.4.1	Données biologiques.....	70
II.3.1.1.4.2	Descripteurs moléculaires.....	70
II.3.1.1.4.3	Modélisation	77
II.3.1.1.5	Méthodologie	77
II.3.1.1.6	Classification.....	80
II.3.1.1.6.1	2D QSAR (Classique)	80
II.3.1.1.6.2	3D QSAR.....	81
II.3.1.1.6.3	2D QSAR vs. 3D QSAR.....	82
II.3.1.1.7	Applications dans la conception des médicaments	83
II.3.1.2	Modélisation du pharmacophore.....	85
II.3.1.2.1	Historique	85
II.3.1.2.2	Définition du pharmacophore.....	86
II.3.1.2.3	Principe de base.....	86
II.3.1.2.4	Génération du modèle pharmacophore	87
II.3.1.2.5	Limites des méthodes pharmacophores.....	89
II.3.1.2.6	Applications dans la découverte des médicaments	90
II.3.1.3	Recherche de similarité.....	92
II.3.1.3.1	Principe général.....	92
II.3.1.3.2	Notion d'empreintes digitales	93
II.3.1.3.3	Applications dans la découverte des médicaments	94

II.4	Conclusion	95
III.	CHAPITRE 3: Développement actuel des recherches	96
III.1	Introduction	97
III.2	L'intelligence artificielle (IA)	98
III.2.1	Qu'est-ce que l'IA ?	98
III.2.2	Contexte historique de L'IA en conception moléculaire	98
III.2.3	Application de l'IA dans la conception des médicaments	100
III.2.4	L'apprentissage automatique « Machine learning » (ML).....	102
III.2.4.1	Définition	102
III.2.4.2	Application dans la découverte de médicaments	102
III.2.5	L'apprentissage profond « Deep learning » (DL)	104
III.2.5.1	Définition	104
III.2.5.2	Application dans la découverte de médicaments	104
III.3	Autres nouvelles méthodes pour améliorer la modélisation moléculaire.....	105
III.3.1	« Cloud computing » ou calcul en nuage	106
III.3.1.1	Définition	106
III.3.1.2	Pourquoi envisager « Cloud computing » pour les projets de modélisation moléculaires ?.....	106
III.3.1.3	Applications dans la découverte des médicaments	108
III.3.2	« Grid computing » : le calcul en grille.....	109
III.3.2.1	Définition	109
III.3.2.2	Application dans les entreprises pharmaceutiques.....	109
III.4	Conclusion	111
	Conclusion générale et perspectives	112
	Références bibliographiques	115

Introduction générale

INTRODUCTION GENERALE

"Toute science est une science informatique". Lorsqu'un article du New York Times publié le 25 mars 2001 a utilisé cette phrase dans son titre, le grand public était conscient que l'introduction des ordinateurs a changé la façon dont les sciences expérimentales ont été menées jusqu'à présent. La chimie, associée à la physique, est le meilleur exemple d'une telle nouvelle façon de d'étudier de la science.(1)

De nombreuses propriétés et comportements physiques et chimiques des molécules ne peuvent être prédits et compris que si les structures moléculaires et électroniques de ces espèces étaient conçues et manipulées dans des modèles tridimensionnels (3D). La gestion de la modélisation par des techniques informatiques est appelée « chimie-informatique ».

Ce domaine est, non seulement, récent, en pleine émergence mais aussi s'actualise d'une manière potentiellement rapide avec la révolution de l'informatique. Il présente un large éventail d'applications dans diverses disciplines, particulièrement, la science biomédicale et le Drug Design. De ce fait, une confusion, entre ses nombreuses méthodes et leurs utilisations, pourrait avoir lieu au cours de la conception des nouvelles molécules.

Ce travail représente une revue de la littérature dans ce domaine. Son objectif est de fournir une vue d'ensemble sur les approches de la chimie-informatique les plus importantes utilisées dans le domaine de conception de médicaments assistée par ordinateur (CADD), des plus anciennes aux plus récentes. Certaines applications de ces techniques sont illustrées afin de donner un aperçu sur les capacités de la modélisation moléculaire dans ce processus.

D'abord, on aborde dans **le premier chapitre** des généralités sur la modélisation moléculaire ; la définition, l'histoire de développement au fur et à mesure de l'évolution spectaculaire de l'informatique, et son intérêt dans le processus de Drug Design.

Ensuite, **le deuxième chapitre** évoque une classification simple et pratique des principales méthodes. Celles de conception basées sur la structure de cible (SBDD) et autres basées sur le ligand (LBDD), constituant l'épine dorsale des processus modernes de CADD.

Concernant la SBDD, les techniques de prédiction de la structure de protéine, que ce soit par homologie ou bien de novo, sont citées en général. Plus en détail, le Docking moléculaire, et la dynamique moléculaire (DM) sont expliquées.

En outre, la LBDD sera développée plus précisément par l'étude de QSAR, les modèles pharmacophores, et brièvement la similarité moléculaire.

Pour **le troisième chapitre**, les voix de développement, avec l'intégration de l'intelligence artificielle en plus des systèmes de parallélisation, sont éclaircies puisque cette discipline est en amélioration continue.

Enfin, on achève ce travail avec une conclusion avec des perspectives dans ce domaine.

I. CHAPITRE 1: Généralités

I.1 Introduction

Au fil des ans, la modélisation des biomolécules est devenue un outil important dans le domaine des biosciences moléculaires. Initialement, réservée aux spécialistes possédant des connaissances approfondies en physique et en informatique et ayant accès aux superordinateurs.

Actuellement, la modélisation moléculaire est de plus en plus utilisée par les pharmaciens, les biologistes et chimistes qui s'intéressent à l'étude de problèmes biologiques et cherchent à les minimiser par la conception des nouveaux médicaments. (3)

Les systèmes de modélisation moléculaire fournissent des outils puissants pour construire, visualiser, d'analyser et de stocker des modèles de systèmes moléculaires complexes qui peuvent aider à interpréter la relation structure-activité comme le montre la figure 1.1.(4)

L'évolution de ces systèmes a été favorisée par l'amélioration du matériel informatique, qui facilite les recherches sans accroître les exigences du matériel ou le temps de calcul. (3)

D'abord, la première question à débattre est la définition du sujet. Qu'est-ce que la modélisation moléculaire ? Et pourquoi s'en soucier ? Et comment elle s'est évoluée ? Et puis, quel est son rôle dans les recherches pharmaceutiques ?

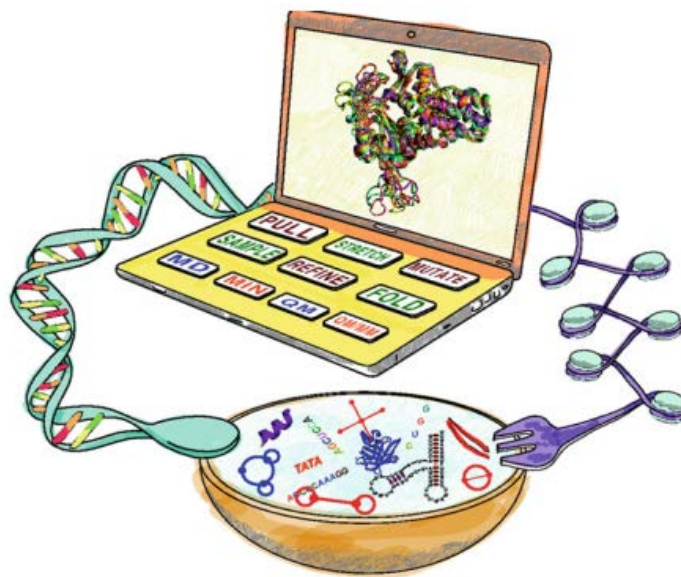


Figure 1.1 –Modélisation moléculaire comme outil informatique.(5)

I.2 Définition

Qu'est-ce que la modélisation moléculaire ?

Une méthode scientifique pour expliquer quoi que ce soit implique une hypothèse, une théorie et des lois.

Une hypothèse n'est qu'une supposition éclairée ou une conclusion logique à partir de faits connus. L'hypothèse est ensuite comparée à toutes les données disponibles et les détails sont développés. Si l'hypothèse s'avère cohérente avec les faits connus, on l'appelle une théorie et elle est généralement publiée.

La plupart des théories expliquent des phénomènes observés, prédisent les résultats de futures expériences, et peuvent être présentées sous forme mathématique. Lorsqu'une théorie s'avère être toujours correcte pendant une longue période, elle est finalement appelée une loi scientifique. (6)

Ce processus est très utile; néanmoins, nous utilisons souvent certaines constructions, qui ne rentrent pas dans le schéma de la méthode scientifique. Cependant, une construction est un outil très avantageux, et peut être pratiqué pour communiquer en science. L'une des constructions les plus couramment utilisés est le modèle.(7)

Par définition, un modèle vient du latin *modulus*, diminutif de *modus*, un terme qui désigne la mesure arbitraire servant à établir les rapports de proportion entre les parties d'un ouvrage d'architecture.(8)

L'activité humaine a toujours été caractérisée par l'usage d'outils. L'activité intellectuelle, qui en la forme la plus évoluée, ne déroge pas à la règle. Au sein de celle-là, les modèles au même titre que les langages sont les principaux instruments méthodologiques au service de la Connaissance. Ils participent à sa recherche, sa découverte, sa validation, son usage, sa communication, son apprentissage et son évolution.(9)

Un modèle est une manière simple de décrire et de prédire des résultats scientifiques. Les modèles peuvent être de simples descriptions mathématiques ou des visuels complètement non-mathématiques.

Les modèles sont très bénéfiques car ils nous permettent de prédire et de comprendre des phénomènes sans effectuer les manipulations mathématiques complexes dictées par une théorie rigoureuse.(7)

En chimie, un modèle simple, que nous considérons à un niveau élémentaire, est la représentation du point de Lewis (point électronique). Par exemple, la structure en points de Lewis de l'atome d'oxygène est donnée à la figure 1.2

CHAPITRE 1: GENERALITES

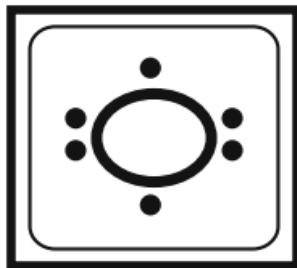


Figure 1.2– La représentation de Lewis de l'atome d'oxygène.(7)

La formule du point d'électrons (également appelée formule du point de Lewis) cherche à désigner l'atome comme un symbole représentant ce que l'on appelle le "noyau" qui comprend la partie de l'atome autre que les électrons de valence. (7)

En général, le dictionnaire anglais Oxford définit "model" comme "une description simplifiée ou idéalisée d'un système ou d'un processus, souvent en termes mathématiques, conçue pour faciliter les calculs et les prévisions". Le terme "molecular" implique clairement un lien avec les molécules.(10)

La modélisation moléculaire semble donc s'intéresser aux moyens d'imiter le comportement des molécules et des systèmes moléculaires. Elle ne doit pas se limiter à fournir des informations, mais elle devrait également permettre de suggérer de nouvelles expériences, c'est-à-dire de nouvelles structures adaptées pour avoir l'activité biologique désirée.(4)

Aujourd'hui, la modélisation (ou simulation) moléculaire est invariablement associée à la modélisation informatique, mais il est tout à fait possible d'effectuer quelques études simples de modélisation de molécules en utilisant des modèles mécaniques ou un crayon, du papier et une calculatrice manuelle. (10)

Néanmoins, les techniques de calcul ont révolutionné cette dernière, au point que la plupart des calculs ne pourraient pas être effectués sans l'aide d'un ordinateur. Cela ne veut pas dire qu'un modèle plus sophistiqué est nécessairement meilleur qu'un modèle simple, mais les ordinateurs ont certainement élargi la gamme des modèles qui peuvent être considérés et les systèmes auxquels ils peuvent être appliqués.(11)

Les "modèles" que la plupart des chimistes rencontrent en premier lieu sont des modèles moléculaires tels que les modèles "bâtons" de Dreiding ou les modèles "remplissant l'espace" de Corey, Pauling et Koltun (communément appelés modèles CPK). Ces modèles permettent de construire des représentations tridimensionnelles des structures des molécules.

Ces modèles structuraux continuent à jouer un rôle important tant dans l'enseignement que dans la recherche, mais la modélisation moléculaire concerne également des modèles plus abstraits, dont beaucoup ont une histoire remarquable. Un exemple évident est la mécanique quantique, dont les bases ont été jetées bien des années avant que les premiers ordinateurs ne soient construits. (10)

Il y a beaucoup de confusion sur la signification des termes "**chimie théorique**", "**chimie informatique**" et "**modélisation moléculaire**". En effet, de nombreux praticiens utilisent ces trois étiquettes pour décrire certains aspects de leur recherche, comme l'exige l'occasion ! La

CHAPITRE 1: GENERALITES

« Chimie théorique » est souvent considérée comme synonyme de mécanique quantique, alors que la chimie computationnelle (informatique) englobe non seulement la mécanique quantique, mais aussi la mécanique moléculaire, les simulations, l'analyse conformationnelle et d'autres méthodes informatiques pour comprendre et prédire le comportement des systèmes moléculaires.(10)

Les méthodes de mécanique quantique sont basées sur la résolution de l'équation de Schrödinger. Cette approche fondamentale est attrayante puisque les structures 3D, les énergies moléculaires et de nombreuses propriétés associées peuvent être calculées sur la base de principes physiques fondamentaux, à savoir les structures électroniques et nucléaires des atomes et des molécules.(5)

Le terme de mécanique moléculaire a été inventé, dans les années 1970, pour décrire l'application de la mécanique classique à la détermination des structures d'équilibre des molécules. La méthode connue auparavant sous au moins deux noms différents : la méthode de Westheimer et la méthode du champ de force. Le nom et l'acronyme MM sont maintenant bien établis.(12)

Les méthodes informatiques de mécanique moléculaire peuvent être définies comme étant des logiciels utilisés pour prédire qualitativement les affinités de liaison protéine/ligand.(13)

La simulation informatique, ou simulation numérique, est une série de calculs effectués sur un ordinateur et reproduisant un phénomène physique. Elle aboutit à la description du résultat de ce phénomène, comme s'il s'était réellement déroulé. Cette représentation peut être une série de données, une image ou même un film vidéo.(14)

L'analyse conformationnelle a pour objectif d'étudier et de déterminer les différentes conformations possibles adoptées par une molécule. Dans les années 50, c'est le chimiste anglais Derek Harold Richard Barton qui fonde les premiers travaux en travaillant sur les stéroïdes.(15)

La modélisation moléculaire utilise toutes ces méthodes, on la considère comme toute technique théorique ou informatique qui permet de comprendre le comportement des systèmes moléculaires. (5)

La modélisation moléculaire peut être considérée simplement comme une gamme de techniques informatisées basées sur des méthodes de chimie théorique et des données expérimentales qui peuvent être utilisées soit pour analyser les systèmes moléculaires ou pour prédire les propriétés moléculaires et biologiques. La construction de modèles peut être aussi simple que des gabarits en plastique ou des tiges métalliques, ou aussi sophistiquée que des stéréographies interactives en couleur et des sculptures en bois réalisées au laser.(16)

Les calculs mathématiques utilisés en informatique moléculaire sont basés sur des théories chimiques et physiques bien établies et confirmés par les études les plus récentes.(17)

Les calculs englobent la mécanique quantique (ab initio), semi-empirique, la mécanique empirique (moléculaire), la dynamique moléculaire, les méthodes de Docking, les relations structure/activité (SAR), les informations et les bases de données chimiques et biochimiques,

CHAPITRE 1: GENERALITES

ainsi que de nombreuses autres procédures établies que seront détaillés au fur et à mesure qu'on avance dans cette thèse.

Le raffinement des données expérimentales, telles que celles obtenues par résonance magnétique nucléaire (RMN) ou cristallographie aux rayons X, est également une composante de la modélisation biomoléculaire. (5)

I.3 Impact de l'évolution de l'informatique sur la modélisation

Après l'invention de l'ordinateur et le développement des logiciels informatiques, la modélisation moléculaire, comme toutes les branches de science qui tendent à utiliser ces techniques pour mieux améliorer et faciliter les recherches, avait passé par une histoire intéressante pour arriver à son stade moderne.

I.3.1 Naissance de la chimie-informatique

Il y a quelques décennies, on a réalisé que l'énorme quantité d'informations accumulées en chimie ne peut être gérée et accessible à la communauté scientifique que sous forme électronique. Cependant, la nouvelle discipline émergeant du stockage, de manipulation et du traitement de l'information chimique n'a pas reçu de nom propre. Les scientifiques travaillant dans ce domaine déclaraient qu'ils traitaient avec "l'information chimique".

Et pour établir une distinction claire entre la bibliothéconomie et le développement de méthodes informatiques, le terme "chimie-informatique" a pris forme il y a quelques années, et son usage s'est rapidement généralisé. (6)

C'est de cette discipline là, les chimistes extraient le domaine de la modélisation moléculaire.

I.3.2 Développement de la modélisation moléculaire

Les racines à partir desquelles les méthodes de la modélisation moléculaire moderne se sont développées remontent au début de notre siècle. Les premières représentations réussies des structures moléculaires étant étroitement liées aux développements rapides de la physique nucléaire. (18)

Les concepts scientifiques qui sous-tendent cette approche ont été compris depuis des générations, mais leur application pratique était hors de portée de la technologie existante. L'existence des récepteurs et les concepts de "lock and key " actuellement considérés dans la conception des médicaments ont été formulés par P. Ehrlich (1909) et E. Fischer (1894).

Ce n'est que dans les années soixante-dix qu'il a été possible de comprendre certaines des subtilités des mécanismes impliqués dans les processus vitaux. Des échantillons purs de protéines cibles ont été isolés et la cristallographie aux rayons X a révélé leur architecture moléculaire.(16)

La cristallographie a été l'axe décisif du développement de la modélisation moléculaire. Les connaissances de la complexité des structures cristallines ont augmenté très rapidement, mais leur résolution nécessitait encore d'énormes dépenses arithmétiques pour ne produire qu'une représentation bidimensionnelle (2D). Et après l'utilisation de kits moléculaires, il est devenu possible d'obtenir une impression en 3D de la structure cristalline. (18)

CHAPITRE 1: GENERALITES

Afin de poursuivre ces progrès, une approche rationnelle de la découverte de médicaments a émergé dans l'industrie pharmaceutique et a contribué au développement rapide de la modélisation moléculaire en tant que discipline à part entière.(16)

En 1958, André Dreiding a inventé un modèle moléculaire utilisé principalement en stéréochimie, nommé le stéréo-modèle de Dreiding. Il s'agissait d'un modèle dit squelettique. Ces modèles sont devenus célèbres car ils contenaient toutes les connaissances de la chimie de la structure à l'époque.

Les dimensions ont été traduites linéairement à partir de l'aire de l'Angstrom. Les obstacles stériques des substituants, les interactions des liaisons hydrogène... étaient très bien représentés par ces modèles. (18)

Une qualité de modélisation similaire, quoique moins précise - mais remplissant l'espace - était fournie par les modèles de Stuart-Briegleb ou de CPK.

Watson et Crick ont décrit leurs tâtonnements avec de tels kits moléculaires... d'abord pour modéliser l'appariement des bases, puis pour dessiner l'hélice d'ADN. (18)

Les champs de force empiriques pour les propriétés moléculaires ont été proposés et utilisés pour la première fois par Westheimer (1946), puis par Warshel (1968) et Boyd (1970).(19)

Le domaine s'est développé rapidement depuis les années 1980. Un certain nombre d'avancées spectaculaires ont été réalisées en biologie moléculaire, en chimie structurale expérimentale et théorique, ainsi que dans les technologies informatiques. Elles constituent toutes des éléments importants du cadre de la modélisation moléculaire.(16)

Dans un article publié en 1982, Cox décrit les applications des calculs de mécanique moléculaire et la façon dont ils sont utilisés pour étudier les conformations moléculaires, propriétés thermodynamiques, et les spectres vibrationnels.

Casanova (1993) a présenté une vue d'ensemble historique de la modélisation moléculaire par ordinateur en chimie et attribue au développement de la mécanique moléculaire la force qui a permis de faire de la modélisation moléculaire un outil de routine.(19)

I.3.3 Evolution actuelle

La chimie computationnelle est devenue une discipline passionnante et en pleine expansion qui traite la modélisation informatique de systèmes tels que les biomolécules, les médicaments.

Depuis son avènement, la chimie computationnelle s'est développée jusqu'à l'état où elle se trouve aujourd'hui et est devenue populaire grâce aux améliorations considérables apportées au matériel et aux logiciels informatiques au cours des dernières décennies.(5)

Grâce à une puissance de calcul élevée utilisant des grilles de calcul et des algorithmes numériques à haute performance de la plupart des laboratoires universitaires et industriels,

plus rapides et efficaces, la chimie computationnelle a donné naissance à une modélisation moléculaire très améliorée.

La modélisation moléculaire est devenue un domaine de recherche appliquée très efficacement pour résoudre des problèmes chimiques et biologiques complexes au cours des phases initiales des recherches pharmaceutiques.(18)

Avec les progrès expérimentaux, algorithmiques et informatiques, des problèmes et des approches qui étaient insurmontables il y a quelques années, sont désormais possibles.

Les techniques actuellement disponibles permettent de mieux comprendre les caractéristiques moléculaires précises qui sont responsables de la régulation des processus biologiques : géométries moléculaires, aspects atomiques et électroniques moléculaires et forces hydrophobes. Toutes ces caractéristiques structurelles sont d'une importance capitale pour la compréhension des relations structure-activité et la conception rationnelle des médicaments.(16)

Un nombre croissant de revues se concentrent désormais sur la modélisation moléculaire : *Journal of Computational Chemistry*, *Computers in Chemistry*, *Journal of Computer-Aided Molecular Design*, *Journal of Molecular Graphics*, *Molecular Simulations*, and *Tetrahedron Computer Methodology*(4)

Plusieurs projets récents, des textes et des revues, décrivent les progrès de la modélisation moléculaire dans la recherche et les applications.

I.3.4 L'ordinateur apporte-t-il une dimension supplémentaire pour la modélisation moléculaire ?

La modélisation moléculaire n'est pas a priori une science informatique, mais effectivement, le développement de l'ordinateur s'est fait en synergie, des processeurs de plus en plus rapides ayant répété les nécessaires en des temps de plus en plus courts, de sorte qu'il est facile aujourd'hui de traiter des protéines contenant des milliers d'atomes.

Pour la première fois, dans les années 1970, la description pseudo 3D d'une molécule, codée par couleur et pouvant tourner, était possible sur un écran d'ordinateur.

Des "modèles virtuels de Dreiding" avaient été créés. Sans la technologie informatique, le flux de données émergeant d'une structure complexe telle qu'une protéine aurait dépassé les limites de saturation de l'efficacité humaine.

Les protéines n'auraient pas pu être mesurées par des méthodes telles que l'analyse de la structure par rayons X et la résonance magnétique nucléaire sans la technologie informatique correspondante. En effet, c'est la technologie informatique qui a fait de ces méthodes ce qu'elles sont aujourd'hui. (18)

CHAPITRE 1: GENERALITES

Des progrès considérables ont été réalisés dans le domaine de la simulation informatique des systèmes moléculaires depuis la publication des algorithmes de base et des premières applications aux systèmes liquides simples ont été publiées.

La complexité, la taille des systèmes étudiés et le nombre d'applications publiées chaque année ont énormément augmenté au cours de la dernière décennie. Cette expansion est en corrélation avec les améliorations du matériel informatique et la plus grande disponibilité des logiciels de simulation.

Les projets qui n'étaient pas réalisables dans les années 1970 sont devenus réalisables à l'aide de superordinateurs dans les années 1980 et sont devenus des applications de routine sur des stations de travail à haut débit.(20)

Boyd et Lipkowitz (1982) ont pris la parole lors des ateliers d'échange du programme de chimie quantique sur les applications pratiques de ses méthodes où ils ont constaté qu'une grande majorité de scientifiques industriels utilisaient couramment la mécanique moléculaire par des logiciels informatiques.

Deux études, réalisées en 1991, l'une par Sauers et l'autre par Rosenfeld ont incorporé la modélisation moléculaire en chimie organique et ils ont valorisé l'utilisation systématique des programmes computationnelles en matière de recherches.(19)

Ainsi, nous pouvons désormais réserver les superordinateurs actuels aux applications de modélisation les plus sophistiquées et, espérons-le, les plus réalistes, à la pointe de la science informatique.(20)

I.4 Importance dans le processus de recherche pharmaceutique

Les grandes firmes pharmaceutiques s'orientent de nos jours très tôt vers une analyse de diversité chimique disponible dans des chimiothèques (librairies ou banques de molécules) ou bien de nouvelles molécules pour, à la fin, arriver à la commercialisation d'un nouveau médicament.(21)

Dans la section suivante, on commence d'abord par une simple signification du médicament, ensuite, une description du processus de recherche et développement de médicament, puis, une estimation du coût de cette recherche pour achever, enfin, sur les domaines d'application de la modélisation moléculaire dans la recherche pharmaceutique.

I.4.1 Médicament: Définition

Le médicament ou la molécule thérapeutique est généralement une petite molécule activant ou inhibant la fonction d'une biomolécule telle que les protéines et entraînant un bénéfice thérapeutique pour le patient. Fondamentalement, la conception de médicament inclut la préparation de petites molécules, de formes et charges complémentaires à la cible moléculaire, avec laquelle elles interagiront.(13)

I.4.2 Recherche de médicaments : « Drug design »ou « conception de médicaments »

Appelé aussi **Drug design rationnel**, il peut être défini comme un ensemble de processus inventifs capables de trouver de nouveaux médicaments, basé sur les connaissances de cibles biologiques. C'est un processus multidisciplinaire extrêmement long (10-20 ans) et coûteux.(22)

Le processus de découverte d'un médicament comporte sept étapes : sélection de la maladie, hypothèse de la cible, identification des hits puis Leads (criblage), optimisation des Leads, essai préclinique, essai clinique et optimisation pharmacogénomique. Traditionnellement, ces étapes sont exécutées de manière séquentielle, et si l'une d'entre elles est lente, elle ralentit l'ensemble du processus (Cf. Figure 1.3)(23)

CHAPITRE 1: GENERALITES

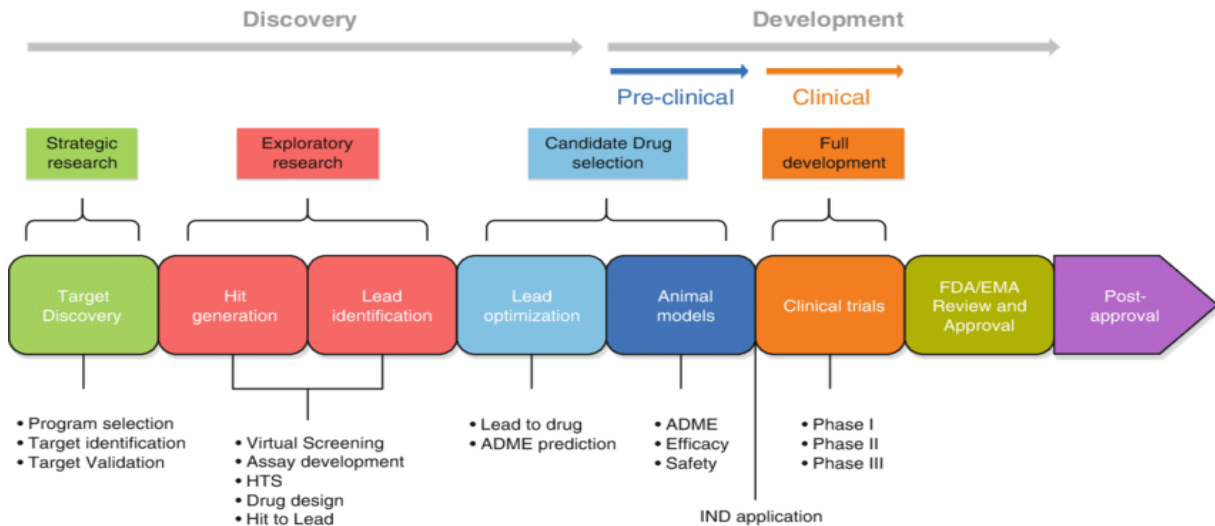


Figure 1.3– Différentes phases du processus de découverte de médicaments.(24)

L'identification et la découverte, aussi tôt et de manière aussi fiable que possible, de nouvelles molécules susceptibles de devenir des médicaments, représente un objectif principal dans le domaine de la recherche pharmaceutique, elle constitue un enjeu majeur pour les années à venir. Le coût, le temps nécessaire et même la disponibilité des laboratoires équipés pour la réalisation des synthèses et des tests rendent le processus particulièrement difficile.(25)

Les acteurs de la recherche pharmaceutique tentent ainsi constamment d'optimiser chacune des étapes menant à la commercialisation d'un médicament. Les entreprises pharmaceutiques font notamment partie de celles qui investissent le plus au monde, avec en moyenne 10 à 20 % de leur chiffre d'affaires réinvesti dans **la recherche et le développement (R&D)**. L'ensemble de ce processus peut se diviser en quatre grandes étapes présentées dans la figure 1.4.(22) Nous allons les décrire brièvement dans les sections qui suivent.

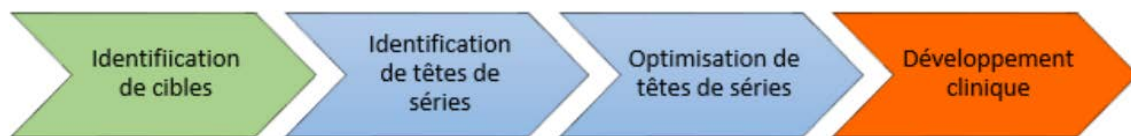


Figure 1.4– Processus (simplifié) de recherche de médicaments découpé en 4 grandes étapes.(22)

Identification et validation de cibles
 Génération de Hits et de Leads
 Optimisation de Lead
 Essais cliniques et commercialisation

} (R&D)

I.4.2.1 Recherche et Développement (R&D)

L'un des principaux objectifs de la phase R&D est de délivrer un ou plusieurs candidats médicaments actifs et innovants, présentant le moins de toxicité possible et ayant ainsi le plus de chance de passer avec succès les étapes qui suivront.

I.4.2.1.1 Identification et validation de cibles

La première étape consiste à identifier une ou plusieurs entités biologiques (généralement des protéines) dont la modulation de l'activité permettraient d'obtenir un effet bénéfique par rapport à une maladie ciblée.

Cette étape est essentielle, en ce sens que toutes les recherches qui suivront se baseront sur l'hypothèse que ces cibles sont effectivement liées à la maladie et que l'action visée par les médicaments aura un effet positif sur l'homme. Elle est d'autant plus délicate que la pertinence de la cible par rapport à la maladie ciblée doit être mise en balance avec les effets secondaires qui pourraient apparaître en conséquence de la modification de son activité. (11)

I.4.2.1.2 Génération de Hits et de Leads

Une fois la ou les cible(s) identifiée(s), une très grande majorité des projets de recherche de médicaments se poursuivent par une étape de criblage. Celle-ci a pour but d'identifier un premier ensemble de molécules actives, plus communément appelées hits.(26)

Un hit est une molécule ayant montré une activité modérée ou forte lors du test expérimental. Il existe différentes méthodes expérimentales permettant l'identification de hits. Aujourd'hui, le criblage expérimental à haut débit - High Throughput Screening (HTS) - est probablement la méthode la plus utilisée dans l'industrie pharmaceutique.(22)

Une fois identifiés, les hits doivent être confirmés à l'aide de tests plus poussés auront pour objectif d'évaluer plus précisément l'activité des molécules, mais également d'établir les caractéristiques physico-chimiques de celles-ci (solubilité, lipophilie, stabilité métabolique,...).La prochaine étape sera alors de transformer les hits les plus prometteurs en tête de série, plus communément appelée Lead. (27)

On définit un Lead (tête de série) comme une molécule (ou structure de base d'un ensemble de molécules)qui non seulement présente une activité significative pour la cible (hit), mais qui en plus est sélective pour celle-ci lors d'un test expérimental.(28)

La sélection de Leads est une étape importante étant donné que la suite du projet se focalisera généralement sur les quelques molécules ainsi obtenues, qui seront amenées vers l'étape d'optimisation et qui représentent autant de molécules possibles pour l'obtention d'un candidat médicament.(27)

I.4.2.1.3 Optimisation de Leads

Durant cette étape, les chimistes médicaux démarreront un processus itératif durant lequel des modifications chimiques seront opérées autour de quelques molécules obtenues lors des criblages afin d'optimiser l'activité et les propriétés des futurs candidats médicaments.

L'objectif de cette étape est d'obtenir un nombre limité de molécules ayant, d'une part, une activité importante (à nouveau le seuil d'activité dépend du projet et de l'objectif, l'idée étant de maximiser le ratio activité / concentration) et d'autre part, des propriétés physico-chimiques, biologiques et toxicologiques optimales. (13)

Cette étape est, de ce fait, certainement la plus difficile, puisqu'elle nécessite d'optimiser en parallèle à la fois l'activité, mais également les autres propriétés qui feront de la molécule un médicament à la fois efficace et peu (ou pas) toxique. On parle généralement d'optimisation multi-objectifs et de nombreuses recherches sont menées dans cette direction, notamment en chimie-informatique.(29)

I.4.2.2 Essais précliniques, cliniques et commercialisation

Lorsqu'un médicament est considéré comme suffisamment actif, il peut entrer en phase de test sur l'animal puis l'homme. Les essais précliniques, cliniques sont alors portés sur un nombre très limité de molécules issues de l'étape d'optimisation. C'est la phase la plus longue et la plus coûteuse du processus de recherche et c'est également celle où le taux d'échec est le plus élevé. Elle dure entre 6 et 8 ans et a pour objectif de tester l'efficacité sur l'animal puis l'homme du ou des candidat(s) médicament. (29)

Le but principal des essais précliniques (sur l'animal) est d'évaluer le niveau de toxicité que présente le ou les candidat(s) médicament.

On distingue trois grandes étapes durant les essais cliniques :

La **première phase** permet d'évaluer sur un petit nombre de volontaires (sains ou malades), la tolérance au médicament et l'absence d'effets secondaires notables.

La **deuxième phase** s'effectue sur un nombre croissant de patients malades, afin de mieux estimer les effets secondaires, l'efficacité thérapeutique et les doses nécessaires pour obtenir l'effet désiré.

La **troisième phase** est une étude comparative à très grande échelle sur plusieurs groupes de patients (plusieurs milliers), durant laquelle l'effet du médicament est comparé à un traitement de référence et / ou à un placebo.

CHAPITRE 1: GENERALITES

Une fois les essais cliniques passés avec succès, l'ensemble des résultats obtenus sont soumis aux autorités de régulation afin d'officialiser le statut de médicament de la molécule et permettre ainsi sa commercialisation. Le processus de pharmacovigilance débute alors, durant lequel chaque médicament sera surveillé pour ses éventuels effets indésirables à court ou long termes.(27) La figure 1.5 suivante éclaire le chemin que passe le médicament depuis la recherche jusqu'à commercialisation et utilisation.

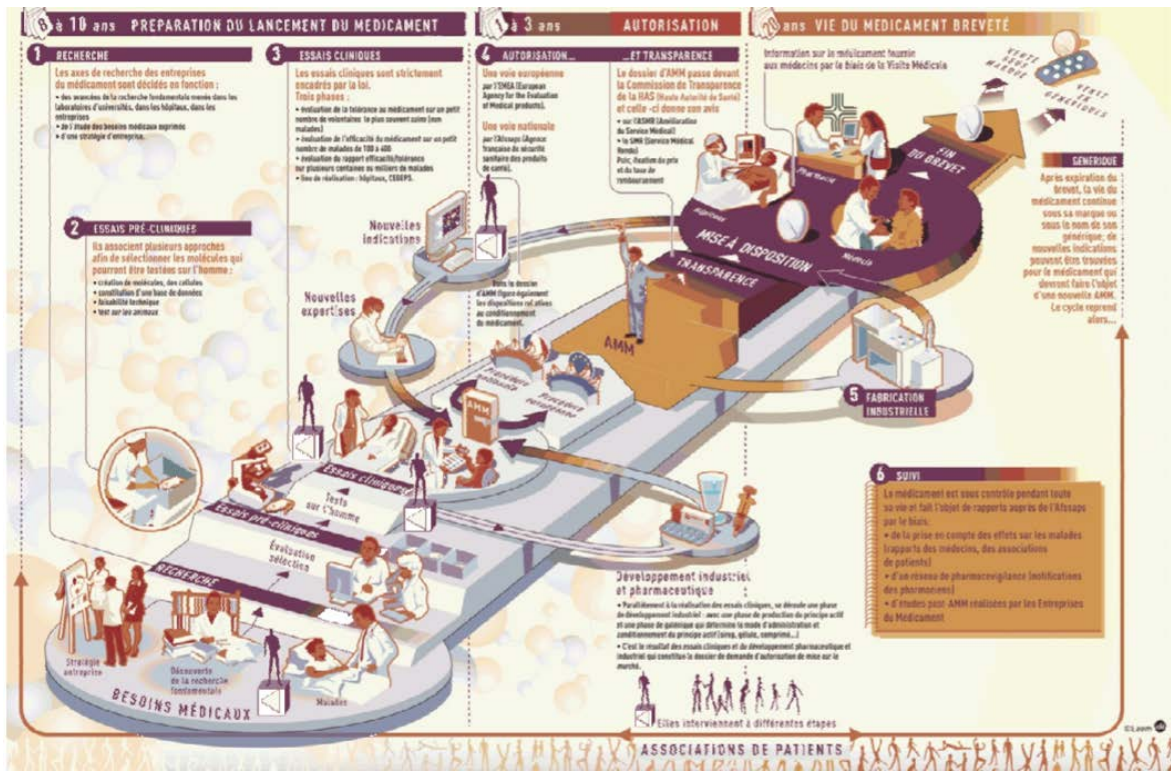


Figure 1.5– Aventure du médicament.(30)

Ce processus n'aboutit toutefois que rarement au développement d'un nouveau médicament : « Le fort taux d'échec de la recherche pharmaceutique est intimement lié à son caractère empirique et séquentiel » [Barbosa & Revah 2004, 18]. Notons entre autres que les méthodes de criblage fonctionnent par essais et erreurs, que le choix des modèles d'animaux utilisés est souvent problématique ainsi que le passage de ces modèles aux tests cliniques humains, notamment du fait de la complexité des facteurs efficacité thérapeutique, toxicité et pharmacocinétique.(26)

En moyenne, pour 10 000 molécules synthétisées et testées, une molécule qui arrive sur le marché en tant que médicament innovant. De plus, le développement d'un médicament demande généralement entre 10 et 15 ans de recherche. Il s'agit en effet de trouver une molécule qui doit à la fois présenter des propriétés thérapeutiques particulières, et posséder le minimum d'effets secondaires indésirables.

Le prix de revient d'un médicament est essentiellement dû à ces synthèses longues, coûteuses et finalement inutiles. Par conséquent, l'industrie pharmaceutique s'oriente vers de nouvelles

méthodes de recherche, qui consistent à prédire les propriétés et les activités des molécules avant même que celles-ci ne soient synthétisées(25)

I.4.3 Coût de l'innovation thérapeutique

Un autre aspect non négligeable du développement de solutions thérapeutiques est l'argent mis en jeu. On trouve dans la littérature des estimations assez divergentes concernant son coût global, évalué entre 300 millions et plus de 1,7 milliard de dollars. L'étude de DiMasi, l'une des plus citées, estime ce coût à environ 802 millions de dollars, tandis qu'une étude plus récente le situe autour de 1,2 milliard.(22)

En effet, d'après le dernier rapport des comptes nationaux de la santé datant de septembre 2010, le montant des dépenses courantes de santé a atteint 223.1 milliards d'euros, pour l'ensemble des Français, dont les quatre cinquièmes représentent la consommation de soins et de biens médicaux. Cette dernière consommation de 175.7 milliards d'euros se traduit par une dépense moyenne en soins et biens médicaux équivalente à 2724 euros par habitant en 2009, dont plus de 500 euros consacrés aux médicaments.(31)

En amont de tout cela viennent les investissements de plus en plus importants dans les secteurs de la recherche et du développement du médicament. En 2008, c'est environ 26,6 milliards d'euros qui ont été investis en Europe contre 24,2 milliards d'euros aux Etats-Unis. Une étude comparative a été effectuée sur les dépenses des trois plus grands ensembles de pays investisseurs dans le domaine de la recherche pharmaceutique.(13,32)

Le graphique illustré dans la figure 1.6 montre bien une croissance mondiale d'investissements dans les institutions de recherche et développement (R&D) des firmes pharmaceutiques qui a explosé durant la dernière décennie. À titre illustratif, durant l'année 2006, 956 autorisations de mise sur le marché ont été délivrées en France. (13)

Cependant, malgré le fort taux d'investissement dans la recherche et développement des médicaments, le nombre de molécules innovantes, quant à lui, tend à diminuer. Aux Etats Unis, par exemple, depuis 1994 une moyenne de 20 nouvelles molécules voient le jour par an.(33)

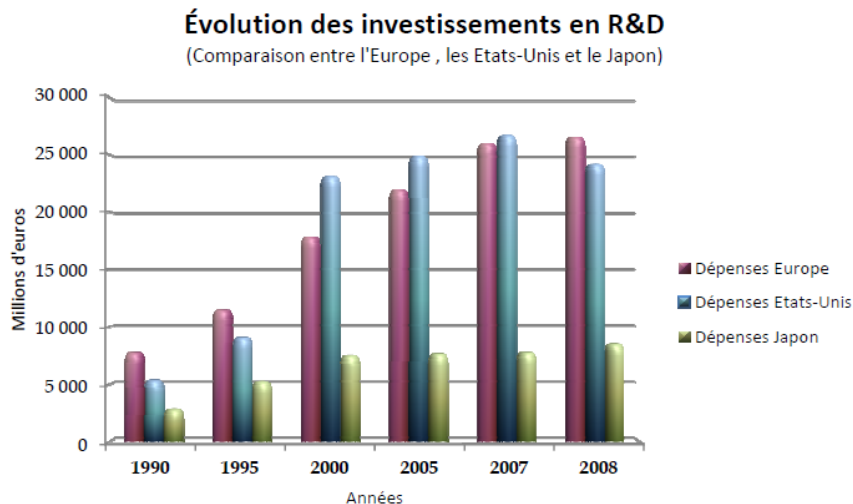


Figure 1.6– Evolution et comparaison internationale des coûts en R&D pharmaceutiques.(13)

Les avancées dans le domaine des découvertes des structures tridimensionnelles des protéines, de la résolution du séquençage du génome humain ainsi que dans le développement des connaissances des maladies et tout leur système de signalisation ont contribué à la croissance des bases de données des molécules à vertus thérapeutiques.(31)

Ce progrès constant des moyens utilisés en R&D du secteur pharmaceutique, contribue à rendre ce domaine vorace en argent et en temps d'investissement. Pour toutes ces raisons, l'avènement des approches « in silico » dans le domaine de la R&D, à la fin des années soixante, apparaît comme une option de plus en plus prisée, dans le milieu pharmaceutique en particulier.(28)

Différents outils théoriques informatisés sont aujourd'hui employés aux différentes étapes de la recherche pré-clinique.

I.4.4 Insertion de la modélisation moléculaire durant le processus de recherches pharmaceutiques (Drug design)

Le domaine du « **Drug design** » ou « **conception de médicaments** » peut être exploré par la modélisation de nouvelles molécules à l'aide des outils informatiques.

La conception de médicaments repose fréquemment sur les techniques de modélisation par ordinateur dites aussi « in silico ». Ce type de modélisation est connu sous le nom de conception de médicaments assistée par ordinateur (CADD).

L'utilisation des méthodes informatiques dans la conception pré-clinique de molécules thérapeutiques s'est développée à partir de 1970 dans le cadre d'une stratégie dite de « Rational Drug Design ».(13)

CHAPITRE 1: GENERALITES

Ces méthodes, développées pour accompagner les différentes étapes de la conception pré-clinique d'un médicament, sont toutes liées à des banques de données, appelés aussi des « Chimiothèques ».(26)

Le schéma suivant (Figure 1.7) représente une vue d'ensemble des méthodes de modélisation qui participe à la phase de R&D pharmaceutique.

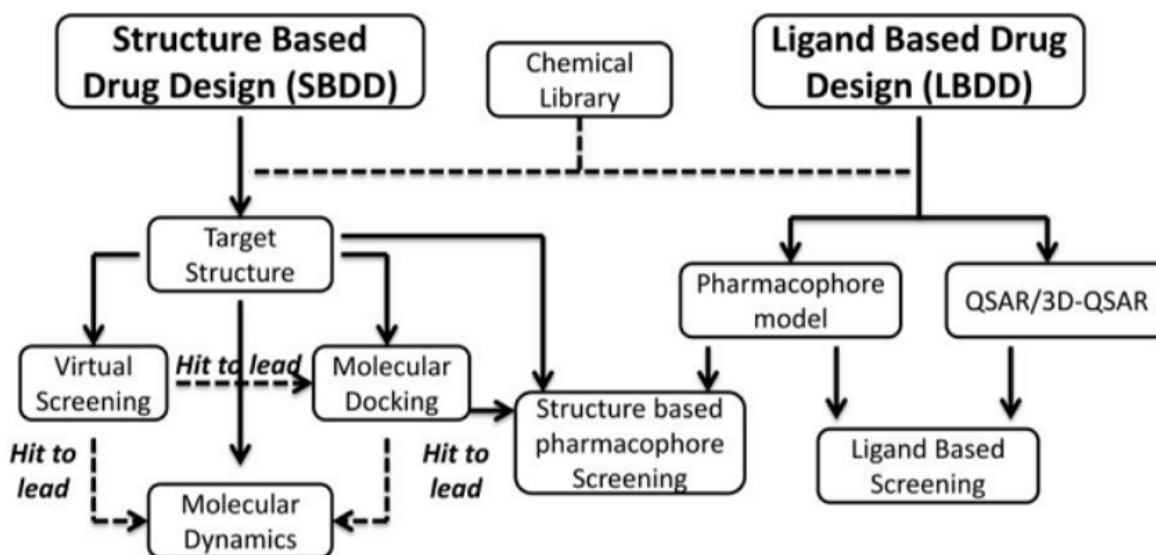


Figure 1.7 – Schéma récapitulatif des méthodes de modélisation insérées dans la phase de R&D pharmaceutique.(34)

D'abord, au niveau du stade de **l'identification d'un ligand**, des méthodes informatiques sont employées afin d'améliorer parfaitement les méthodes de criblage. Ainsi, l'approche dite de « **structure-based drug design** » se propose pour concevoir des ligands spécifiques en utilisant les connaissances structurales de la cible. L'approche de « **ligand –based drug design** » est appliquée en cas d'absence de structure de la cible.

Le « **Docking** » (amarrage) est toutefois également employé dans des pratiques plus récentes de criblage virtuel. Le développement de celles-ci est lié à l'avènement de la chimie-informatique combinée à l'infographie virtuelle. Le ligand et la cible sont modélisés sur ordinateur dans le but d'ajuster, par des modifications chimiques virtuelles du ligand, les interactions entre ces deux entités.

Ces nombreux ligands virtuels sont ensuite soumis à un criblage sur ordinateur (**Génération des hits et de Leads**), soit en utilisant une procédure de « Docking », soit en sélectionnant dans la chimiothèque un profil de propriétés physico-chimiques correspondant à celui d'une famille de molécules ayant montré expérimentalement une certaine activité par la méthode de **recherche de similarité** ou bien le **criblage virtuel à base de pharmacophore**.(13)

Enfin, certaines méthodes de calcul, aujourd'hui informatisées, accompagnent la troisième étape évoquée de la conception d'un médicament (**Optimisation des Leads**) : les molécules

CHAPITRE 1: GENERALITES

sélectionnées lors de criblages sont modifiées en utilisant des méthodes dites **QSAR** «Quantitative Structure-Activity Relationship», basées sur un principe de corrélation statistique. Elles ont pour but de corréler des propriétés mesurées expérimentalement (affinité, solubilité...) à des descripteurs moléculaires calculés in silico.

Toutes ces méthodes entre dans le cadre de conception de médicament assistée par ordinateur (**computer aided drug design** « **CADD** ») récapitulé dans la figure 1.8.

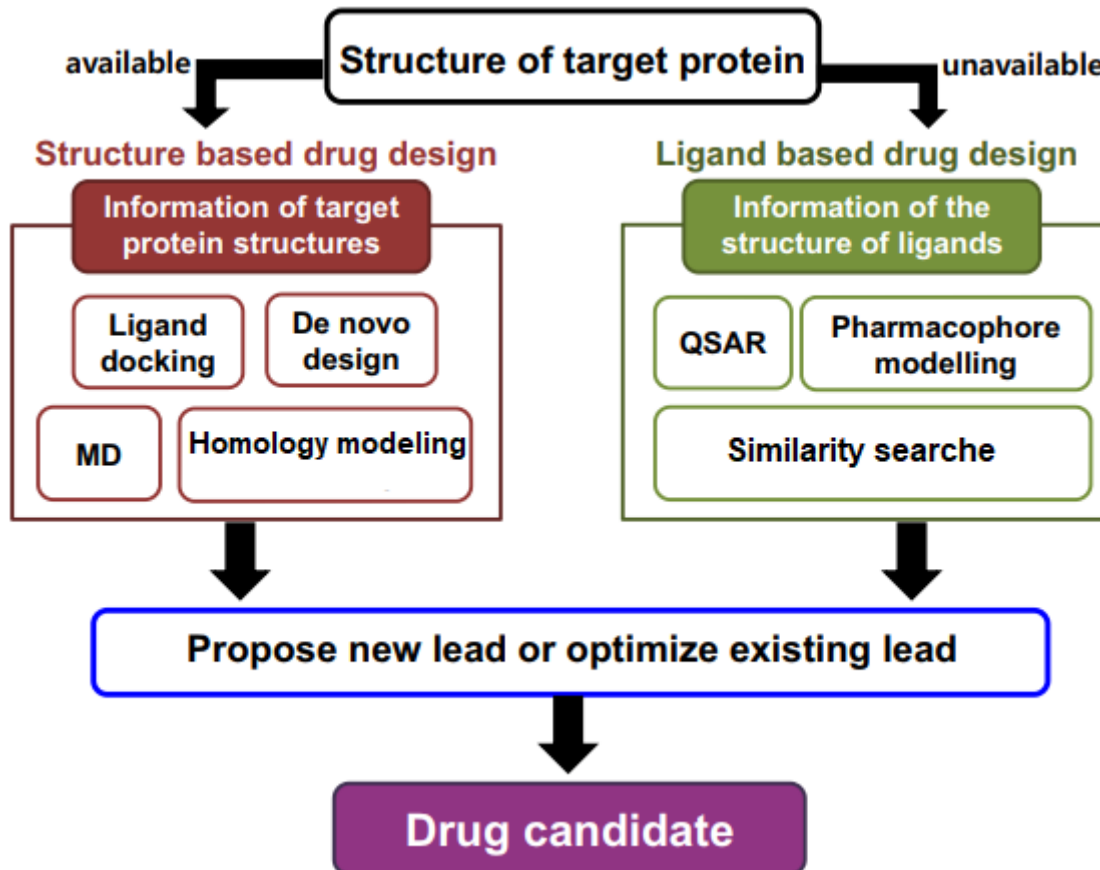


Figure 1.8- Organigramme récapitulatif de CADD.(35)

Les techniques de modélisation tentent de prédire l'affinité entre le ligand et sa cible et, à ce jour, les résultats sont globalement satisfaisants. Cependant, il existe d'autres propriétés qui doivent être optimisées pour que le ligand devienne sûr et efficace autant que puisse l'être un médicament. Ces caractéristiques moléculaires que sont l'absence d'effets secondaires, la biodisponibilité, la demi vie métabolique, la toxicité ... sont par ailleurs assez difficiles à optimiser avec les techniques de conception de médicaments rationnelle.(36)

I.4.5 Etudes encore plus récentes... De la modélisation moléculaire vers la modélisation biologique

Que se soit par comparaison et alignement de séquences biologiques, par recherche de motifs, par assemblage de fragments d'ADN, par cartographie génétique, ou bien par structure 3D des protéines et ARN ; les méthodes computationnelles se sont avérées essentielles pour accomplir ces tâches. Ceci a enfin donné aux biologistes une liste plus complète des composantes biologiques du vivant et de leurs caractéristiques.(37)

La biologie computationnelle est une science interdisciplinaire relativement nouvelle, soutenant de manière cruciale le mouvement vers la "biologie à grande échelle", en particulier la génomique.

La biologie devenant de plus en plus une discipline à forte intensité d'information avec une liste exhaustive des composantes et de leurs caractéristiques, l'application de méthodes informatiques devient non seulement indispensable à la gestion, à la compréhension, à la présentation des données, mais également imbriquée dans le tissu du domaine dans son ensemble.(38)

En conséquence, un travail d'adaptation important de différents langages de modélisation pour des applications biologiques a été accompli dans la dernière décennie et les outils informatiques de **modélisation biologique** se sont multipliés. Les biologistes peuvent maintenant modéliser et simuler, facilement, ce qui va y avoir un impact, de plus en plus important, sur le processus de R&D pharmaceutique.(39)

La figure 1.9 représente un diagramme de la filière de recherche et développement pharmaceutique. On a identifié trois domaines potentiels dans lesquels la modélisation biologique peut avoir un impact substantiel sur l'efficacité et le développement du processus.

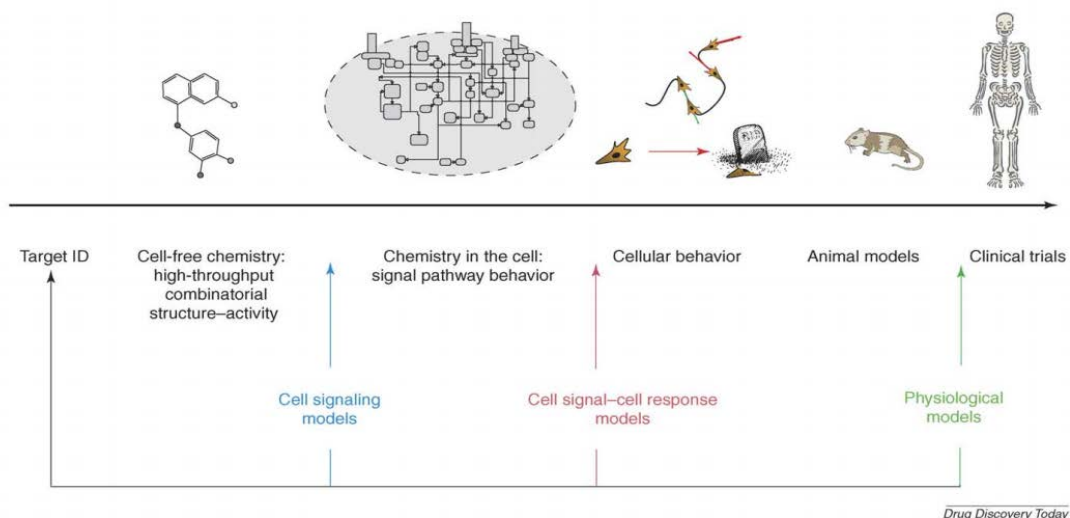


Figure 1.9– Domaines d'impact de la modélisation biologique dans le processus de R&D pharmaceutique.(40)

CHAPITRE 1: GENERALITES

Le premier domaine est le comportement du signal cellulaire, où l'application de modèles caractérise la façon dont les composés principaux affectent la signalisation intracellulaire. Ils prédisent les effets des médicaments sur elle.

Le deuxième domaine est le comportement signal-réponse, où les modèles prédisent le phénotype cellulaire à partir des informations de signalisation. Ils établissent une corrélation entre les signaux intracellulaires et les comportements cellulaires (réponses, tels que la migration et l'apoptose). Ils prédisent les effets des médicaments sur le comportement cellulaire.

Le troisième domaine est la physiologie, dans lequel les modèles sont utilisés pour simuler des résultats cliniques. Les modèles physiologiques simulent le comportement au niveau des organes et prédisent la réponse humaine aux médicaments.

Chaque catégorie de modèle peut aider à identifier de nouvelles cibles médicamenteuses. On aborde chaque domaine d'application séparément dans l'industrie pharmaceutique.(40)

I.4.5.1 Modèles de signalisation cellulaire

A cause de sa complexité et l'hétérogénéité des processus biologiques qui la composent, la signalisation cellulaire a été étudiée par de nombreuses approches de modélisation.(41)

Les modèles informatiques qui sont utilisés pour décrire les voies de signalisation pertinentes pour les maladies, appelés les modèles de signalisation cellulaire. Ils sont importants dans la recherche pharmaceutique pour trois raisons principales :

- 1.** Ils saisissent souvent le comportement non intuitif des signaux et identifient une nouvelle fonction moléculaire.
- 2.** Ils permettent aux chercheurs d'expérimenter *in silico* dans un large éventail de conditions (par exemple, le nombre de récepteurs, les concentrations de ligand et les taux de phosphorylation), ce qui permet d'économiser des ressources expérimentales et d'identifier d'autres expériences importantes.
- 3.** Ils servent de base de données pour la plupart des informations connues sur une voie particulière.(40)

I.4.5.2 Modèles de signalisation et de réponse cellulaire

Par ailleurs, Il existe aussi des modèles informatiques pour comprendre le comportement cellulaire envers la signalisation cellulaire. Les modèles de ce type sont très utiles pour l'industrie pharmaceutique, car ils permettent aux scientifiques de modifier les signaux et de prédire ensuite comment le comportement cellulaire change *in silico*.(41)

CHAPITRE 1: GENERALITES

Pour corriger un comportement cellulaire aberrant à l'aide de médicaments, il faut avoir des connaissances quantitatives sur de multiples protéines de signalisation (c'est-à-dire des ensembles de données multi-variées).(40)

Il existe très peu de modèles qui peuvent relier, avec précision, les voies de signalisation au comportement cellulaire à ce niveau de description mathématique. Le problème, par conséquence, nécessite l'utilisation des modèles de signalisation plus abstraits.

Les modèles abstraits identifient les relations statistiques entre les signaux et le comportement, ce qui suggère des relations causales entre le signal et le comportement. Elles peuvent être approfondies par des approches de biologie moléculaire ou de génétique. Les prédictions faites sur la base de ces modèles peuvent révéler comment un médicament, ou une classe de médicaments, affectera un comportement cellulaire donné.(39)

I.4.5.3 Modèles physiologiques

La modélisation physiologique définit le développement de modèles mathématiques qui reproduisent le comportement des organes et systèmes biomédicaux. Cette branche du génie biomédical vise à mieux comprendre les interactions fonctionnelles entre les principaux composants multi-échelles du corps humain, notamment les gènes, les protéines, les cellules, les organes et les systèmes.

Elle permet également de quantifier et de prédire comment ces interactions évoluent dans des états pathologiques ou après l'application de thérapies.(42)

Ces modèles sont caractérisés par une structure mathématique (c'est-à-dire un ensemble d'équations, de lois et de corrélations utilisées) et par un ensemble de paramètres de modèle à estimer à partir des données cliniques.(40)

Les modèles physiologiques fiables doivent être, physiologiquement, cohérents (c'est-à-dire basés sur les mécanismes physiologiques et les voies métaboliques observés à l'intérieur de l'organisme) et fournissent une variance minimale des prédictions, grâce à une estimation, des paramètres constitutifs du modèle, statistiquement précise.(43)

I.5 Conclusion

On peut définir la modélisation moléculaire comme une discipline qui s'appuie sur les lois fondamentales de la mécanique quantique ou de la mécanique statistique pour décrire le comportement ou les propriétés d'un système en représentant explicitement sa géométrie à l'échelle moléculaire ou atomique et son comportement vis-à-vis des enzymes.

Le développement de cette discipline durant plusieurs décennies s'est accompagné par l'évolution de nouvelles technologies informatiques très intéressantes concernant la chimie des protéines mais aussi la physique nucléaire. Actuellement, les ordinateurs sont beaucoup plus améliorés donnant naissance à des techniques plus performantes de modélisation.

Concernant la découverte des médicaments, l'approche traditionnelle repose sur la synthèse et le criblage (Screening) d'un grand nombre de composés afin d'optimiser les profils d'activité, ce qui est extrêmement long et coûteux.

Compte tenu de la nécessité de réduire le coût et d'accélérer les processus de découverte de médicaments, tous les outils qui peuvent améliorer l'efficacité de cette démarche sont hautement souhaitables. La modélisation moléculaire est l'un de ces outils qui peut réduire de manière significative le coût, le travail et le temps, et augmenter l'efficacité du processus de découverte de médicaments.

II. CHAPITRE 2: Classification simple des méthodes

II.1 Introduction

La découverte et le développement d'un nouveau médicament est un processus long, complexe, coûteux et très risqué qui a peu d'équivalents dans le monde commercial. C'est pourquoi les approches de conception de médicaments assistée par ordinateur (**CADD**) sont largement utilisées dans l'industrie pharmaceutique pour accélérer le processus et contribuer à réduire le temps et le coût des recherches de découverte.(44)

La conception de médicaments basée sur la structure (**SBDD**) et la conception de médicaments basée sur le ligand (**LBDD**) sont des domaines de recherche actifs dans le monde universitaire et commercial appartenant à la CADD (Cf. Figure 2.1).(45)

Les sujets abordés sont principalement la modélisation des protéines par homologie, le Docking et la dynamique moléculaire concernant la SBDD ; la similarité moléculaire, la relation quantitative structure/activité, les méthodes pharmacophores notamment pour l'étude de la LBDD.

Ce chapitre couvre la SBDD et la LBDD, et fournit les informations les plus récentes sur un large éventail de sujets à l'intention des praticiens de la chimie computationnelle, de génie génétique et surtout de la chimie médicinale.

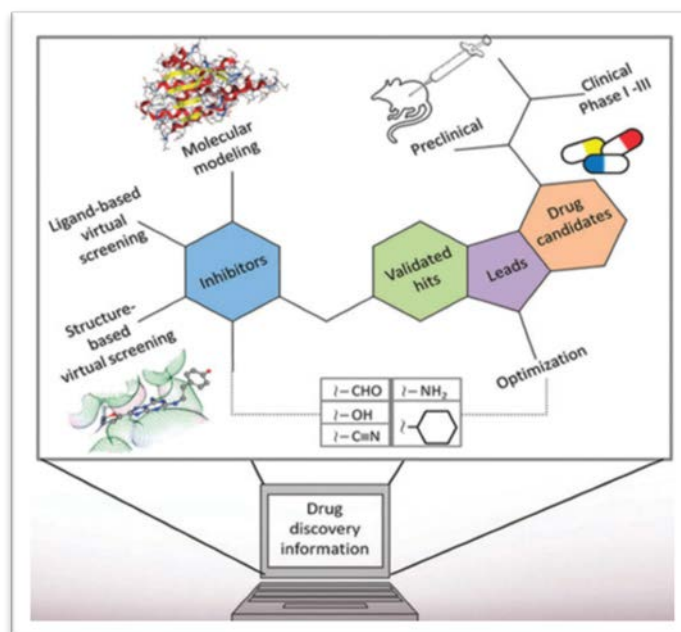


Figure 2.1- Illustration de la CADD.(46)

II.2 « STRUCTURE -BASED DRUG DESIGN » (SBDD)

La conception de médicament basée sur la structure de la cible (enzyme)

La plupart des médicaments ont été découverts dans des cribles aléatoires (screenings) ou en exploitant des informations sur les récepteurs macromoléculaires. L'une des sources de ces informations se trouve dans les structures des protéines et des acides nucléiques.

L'explosion des informations génomiques, protéomiques, et structurelles a fourni des centaines de nouvelles cibles et opportunités pour la découverte des nouveaux médicaments. L'approche de la conception basée sur la structure (structure-based) associe ces informations à des programmes informatiques spécialisés pour proposer de nouveaux inhibiteurs d'enzymes et d'autres agents thérapeutiques.(47)

La conception de médicaments basée sur la structure (SBDD) est une méthode qui dépend de la connaissance des structures 3D des cibles biologiques. Elle se développe rapidement avec l'invention de nouvelles technologies pour la recherche de moyens potentiels de combattre les maladies. La dernière décennie a permis de tripler le nombre de logiciels et d'outils disponibles dans les référentiels en ligne.(48)

Il existe maintenant de nombreux exemples réussis de conception de nouveaux ligands basés sur la connaissance de la structure des protéines cibles. Dans la plupart des cas, ces ligands ne conviennent pas comme médicaments en raison de problèmes de toxicité, de stabilité ou de biodisponibilité.(49)

Quelques 37.000 structures de protéines sont aujourd'hui disponibles dans la Protein Data Bank(PDB) mais de nombreuses familles ne sont pas représentées. Cependant, lorsqu'une structure fait défaut, il est parfois possible d'en construire un modèle, grâce aux techniques de construction par homologie.

Lorsque l'on dispose d'un modèle tridimensionnel d'une protéine cible, il est alors possible d'étudier les interactions de ce récepteur avec de petites molécules organiques, telles que le substrat ou le ligand naturel, des activateurs ou des inhibiteurs.(50)

Des logiciels de Docking recherchent les positions et les orientations les plus probables de ces petites molécules en interaction avec la cible et évaluent l'énergie d'interaction de chaque complexe. Ces modèles aident à comprendre le mode d'action des molécules bioactives et à concevoir de nouvelles structures capables d'exploiter au mieux la spécificité du site de liaison et le potentiel d'interaction existant.(51)

II.2.1 Historique

Le premier exemple de conception basée sur la structure a été rapporté par le groupe de Beddell et Goodford en 1973 aux Wellcome Laboratories au Royaume-Uni. L'hémoglobine a été choisie comme cible, ce qui, à l'époque, était le seul exemple de pertinence pharmacologique avec une structure cristalline connue.

L'objectif de ces études était de développer un ligand qui agisse de manière similaire à l'effecteur allostérique naturel, le diphosphoglycérate(1). Ce ligand endogène se lie à l'hémoglobine et régule son affinité pour l'oxygène. En prenant cette molécule comme référence, le groupe Wellcome a conçu des dérivés de dialdéhyde (2) et des adduits bisulfites apparentés (3) qui, comme prévu, modifient l'affinité de l'oxygène pour l'hémoglobine (Cf. Figure 2.2(A)).

Plusieurs années plus tard, l'antihypertenseur Captopril (4), qui inhibe l'enzyme de conversion de l'angiotensine, a été introduit sur le marché. Il s'agissait du premier médicament développé à partir d'informations structurales. Bien qu'une molécule peptidique fût déjà connue à partir d'un poison de serpent, l'importante percée n'a été réalisée qu'au milieu des années 1970 par Cushman et al après avoir modéliser le site actif de l'enzyme (Cf. Figure 2.2(B)).

Comme référence structurale, le site de liaison de la carboxy-peptidase (l'enzyme de conversion de l'angiotensine) a été utilisé. Peu avant, la structure cristalline de cette enzyme était déterminée. Notant, en plus, que c'est une protéinase à base de zinc.(52)

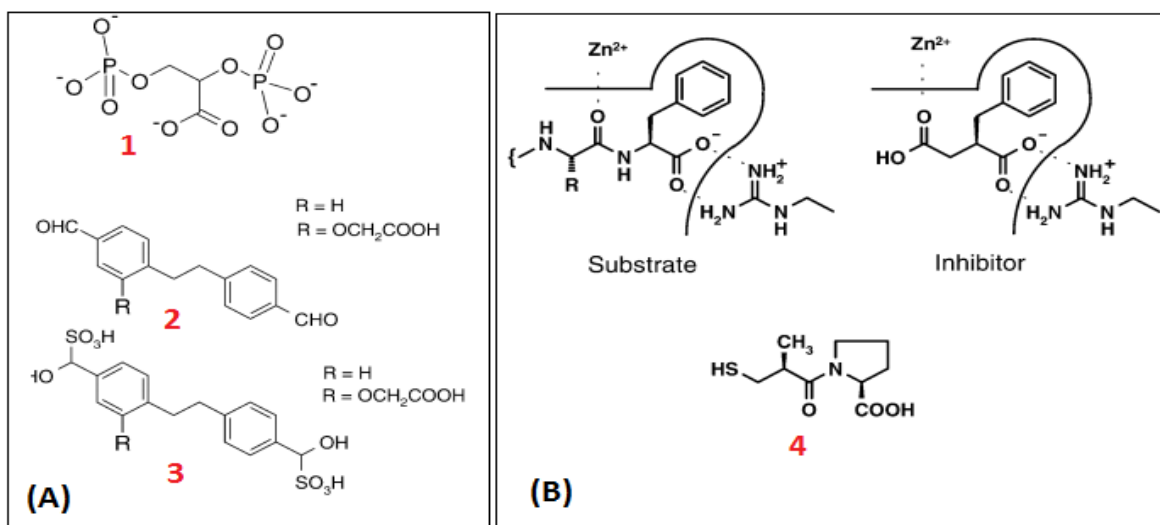


Figure 2.2- Premiers exemples de la SBDD.(52)

II.2.2 Développement de l'approche

L'achèvement du projet du génome humain, le début des révolutions de la protéomique et de la génomique structurale, ainsi que les développements des technologies de l'informatique offrent une opportunité encore plus grande pour cette approche de participer à la réussite de la découverte de nouveaux médicaments. (53)

D'excellentes cibles médicamenteuses sont identifiées à un rythme accéléré grâce aux progrès de la bioinformatique. Les gènes de ces cibles peuvent être clonés rapidement, et la protéine exprimée et purifiée jusqu'à homogénéité. Les progrès de la cristallographie à haut débit, tels que l'automatisation de toutes les étapes, un rayonnement synchrotron plus intense et de nouveaux développements dans la détermination de la phase, ont raccourci les délais de détermination des structures. (54)

La résonance magnétique nucléaire (RMN) a apporté une aide considérable à cela, au cours des dernières années. En plus, on a assisté à une forte augmentation des volumes de logiciels qui peuvent aider à réaliser efficacement les différentes phases de la SBDD de manière efficace.(48)

Les 20 dernières années de conception de médicaments ont été marquées par la caractérisation structurale d'un nombre considérable de cibles, thérapeutiquement, importantes. Le nombre croissant d'applications réussies, de la conception de médicaments à base de structure, a conduit à la découverte de nouvelles thérapies (Cf. Tableau II.1). (52)

Comme exemple, la protéase rétrovirale (PR) du virus de l'immunodéficience humaine de type 1 (VIH-1) a été identifiée il y a plus de dix ans comme une cible potentielle pour la conception de médicaments basés sur la structure. Cet effort a été couronné de succès. Quatre médicaments sont déjà approuvés, et d'autres sont en cours d'essais cliniques. Le (VIH) a démontré, de manière convaincante, l'impact et la pertinence des approches basées sur la structure pour le développement de nouveaux médicaments.(55)

Aujourd'hui, même s'il reste encore beaucoup à faire pour perfectionner le processus, la SBDD fait partie intégrante de la plupart des programmes industriels de découverte de médicaments et constitue le principal sujet de recherche de nombreux laboratoires universitaires. Il est finalement devenu un défi de choisir des combinaisons réussies de stratégies et d'outils pour une découverte efficace de Leads.(56)

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

Tableau II.1– Exemples de médicaments découverts par les méthodes SBDD.(57)

Médicament	Cible du médicament	Maladie ciblée	Technique
Raltitrexed	Thymidylate synthase	VIH	SBDD
Amprénavir	Protéase antiretrovirale	VIH	Modélisation des protéines et dynamique moléculaire (DM)
Isoniazide	InhA	Tuberculose	Criblage virtuel basé sur la structure (SBVS) et modélisation de pharmacophore
Inhibiteur de Pim-1 Kinase	Pim-1 Kinase	Cancer	Criblage virtuel hiérarchique multistage
Epalrestat	Aldose Reductase	Neuropathie diabétique	DM et SBVS
Flurbiprofène	Cyclooxygenase-2	Polyarthrite rhumatoïde, Ostéoarthrite	Docking moléculaire

II.2.3 Aperçu du processus

Dans l'ensemble du paradigme de la découverte de médicaments, la SBDD est le processus le plus puissant et le plus efficace.

C'est une procédure itérative et passe par de multiples cycles pour aboutir à un candidat médicament optimisé qui subit des essais cliniques puis commercialisé.

1. Identification de la cible

Tout d'abord, l'étape fondamentale implique le clonage du gène cible, suivi de l'extraction, purification et, par la suite, la détermination de la structure 3D de la protéine. (57)

Les informations structurales de toutes les cibles sont généralement obtenues par cristallographie aux rayons X ou par RMN. Cependant, dans le cas de : cible dont la structure n'a pas été déterminée expérimentalement, plusieurs approches computationnelles telles que la modélisation « ab initio », ou la modélisation comparative peuvent être utilisées pour prédire les structures 3D.(48)

Cette structure nécessite l'existence de la structure cristalline d'un homologue étroitement apparenté à la protéine cible, en cas de modélisation par homologie. Si on ne dispose pas une structure de séquence similaire de la cible, on utilise la prédiction « ab initio » ou dite « De novo ».(52)

Il existe de nombreux programmes automatisés qui peuvent faire des prédictions de structure protéique de haute qualité. Comme les structures 3D sont les exigences fondamentales pour commencer la SBDD, la plupart des progiciels de conception de médicaments comprennent des utilitaires de prédiction de structure.(48)

2. Validation de la structure 3D

Lorsque la structure cible est prédite, quelques étapes de validation sont essentielles pour confirmer la qualité stéréochimique de la structure prédite avant d'aller plus loin dans la SBDD.(54)

Le serveur d'analyse et de vérification structurale (SAVES)est un serveur de validation de structures qui offre une combinaison d'outils d'évaluation de structures attribuant différents paramètres structurels.(48)

3. Reconnaissance du site de liaison

Si on trouve une structure homologue de la cible complexé avec un ligand, c'est l'idéal. Ce complexe permet de démêler le mode de liaison et la conformation du ligand étudié et indique les aspects essentiels déterminant son affinité de liaison.

Il est ensuite utilisé pour générer de nouvelles idées sur les moyens d'améliorer un ligand existant ou de développer de nouvelles squelettes de liaison alternatifs. Des méthodes de calcul complétées par des graphiques moléculaires sont appliquées pour faciliter cette étape de génération d'hypothèses. (52)

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

Les méthodes actuelles de SBDD prennent en compte les caractéristiques clés de la cavité de liaison de la cible thérapeutique pour concevoir des ligands efficaces.

Ces molécules sont classées selon un système de scoring basé sur les interactions électrostatiques et stériques avec le site de liaison. Une étude approfondie de ces interactions, y compris la présence de cavités, de fentes et de poches allostériques, peut être réalisée à l'aide d'une structure 3D de la molécule cible.(57)

4. Virtuals creening (VS): criblage virtuel

Les caractéristiques de la poche de liaison de la protéine peuvent être traduites en requêtes utilisées pour le criblage virtuel (VS) par ordinateur de grandes chimiothèques de composés ou pour concevoir de nouveaux ligands de novo. Ces ligands initiaux (hits) doivent être confirmés expérimentalement donnant naissance à des Leads. (52)

5. Optimisation et synthèse des Leadsj

A ce stade, les meilleurs Leads sont optimisés pour obtenir une plus grande affinité et une meilleure sélectivité.

Ce dernier aspect est de haute importance pour définir et contrôler le profil pharmacologique d'un ligand. Une condition préalable à l'adap tation de la sélectivité par la conception rationnelle est une compréhension détaillée des paramètres moléculaires qui déterminent la sélectivité.(52)

Ensuite, ils sont synthétisés. En outre, les composés les mieux classés, présentant une affinité élevée et une sélectivité pour la protéine cible, sont testés in vitro avec des essais biochimiques.

Ces ligands interfèrent avec des voies cellulaires cruciales, conduisant ainsi au développement de médicaments ayant un effet thérapeutique et pharmacologique souhaité.(48)

Les propriétés biologiques telles que l'efficacité, l'affinité et la puissance des composés sélectionnés sont évaluées par des méthodes expérimentales.

L'étape suivante consiste à déterminer la structure 3D de la protéine cible en complexe avec le ligand prometteur obtenu dans l'étape précédente. La structure 3D fournit des informations détaillées sur les caractéristiques intermoléculaires qui aident au processus de reconnaissance moléculaire et de liaison du ligand.(56)

La connaissance de la structure du complexe ligand-protéine aide à l'analyse de diverses conformations de liaison, l'identification de liaisons inconnues et les interactions ligand-protéine, l'évaluation de l'efficacité du ligand.(54)

Par la suite, de multiples itérations permettent d'accroître l'efficacité et la spécificité de la molécule. La troisième phase comprend les essais cliniques des composés principaux. Les composés qui les réussissent passent à la quatrième phase au cours de laquelle le médicament est distribué sur le marché.(48)

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

La figure 2.3 résume le processus de développement de médicament à base de SBDD en quatre panneaux. **Le premier panneau** montre le séquençage du génome humain suivi de l'extraction et la purification des protéines cibles.

Le deuxième panneau représente la détermination de la structure des protéines importantes sur le plan thérapeutique de son site de liaison à l'aide d'approches de biologie structurale intégrative et même des approches de modélisation computationnelle.

Le troisième panneau représente la préparation de la base de données des composés actifs. Ensuite, Ils passeront au crible et s'insèrent dans la cavité de liaison de la protéine cible.

Dans **le dernier panneau**, l'identification du composé principal puissant est présentée. Les composés les plus performants obtenus à la suite du criblage virtuel et de Docking sont synthétisés et testés in vitro. D'autres modifications peuvent être effectuées pour optimiser le composé principal.(57)

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

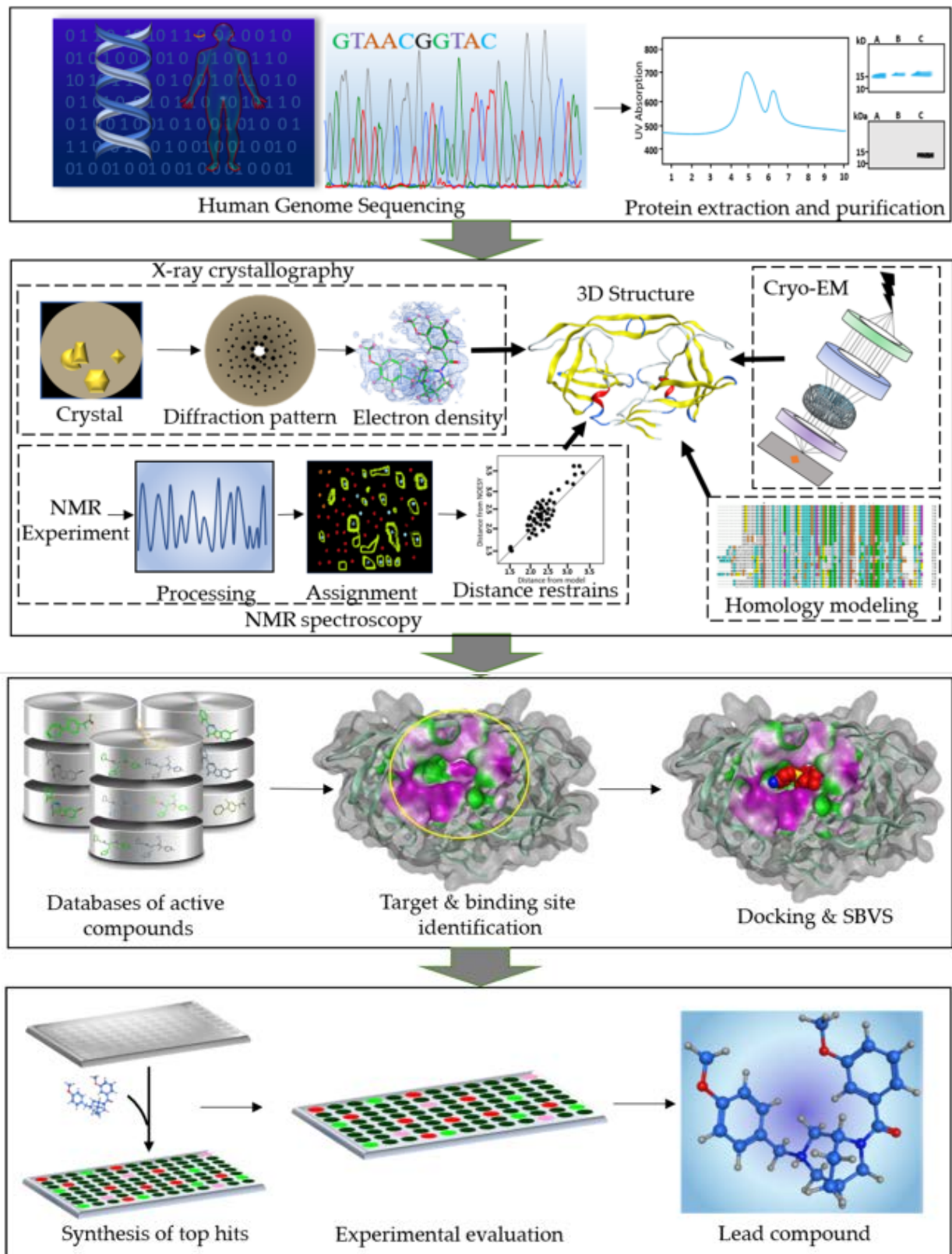


Figure 2.3-Diagramme de flux de travail du processus de conception de médicaments basée sur la structure (SBDD).(57)

II.2.4 Principales méthodes utilisées

II.2.4.1 Prédiction de structure de la protéine

Avec le succès d'un nombre croissant de projets de séquençage du génome, le nombre de séquences de protéines connues a augmenté de façon exponentielle. Cependant, les séquences ne permettent pas à elles seules de savoir ce que fait chaque protéine dans la cellule. (58)

Bien que les informations sur la structure des protéines soient essentielles, pour comprendre la fonction, la vitesse de détermination de la structure des protéines est très en retard sur l'augmentation des séquences.(59)

Cet écart ne pourra pas être comblé, car les techniques expérimentales de détermination de structure (diffraction de rayons X et résonance magnétique nucléaire RMN) sont coûteuses, laborieuses, et lentes (un an de travail en moyenne pour une seule protéine).(60)

À la fin de l'année 2015, environ 90 millions de séquences de protéines ont été déposées dans la base de données UniProtKB (<http://www.uniprot.org/>). Cependant, le nombre correspondant de structures protéiques dans la (PDB) (<http://www.rcsb.org>) n'est que d'environ 100 000. L'écart se creuse rapidement, comme l'indique la figure 2.4, où le rapport entre les séquences et les structures est passé de moins d'une magnitude à environ trois magnitudes au cours des deux dernières décennies.(61)

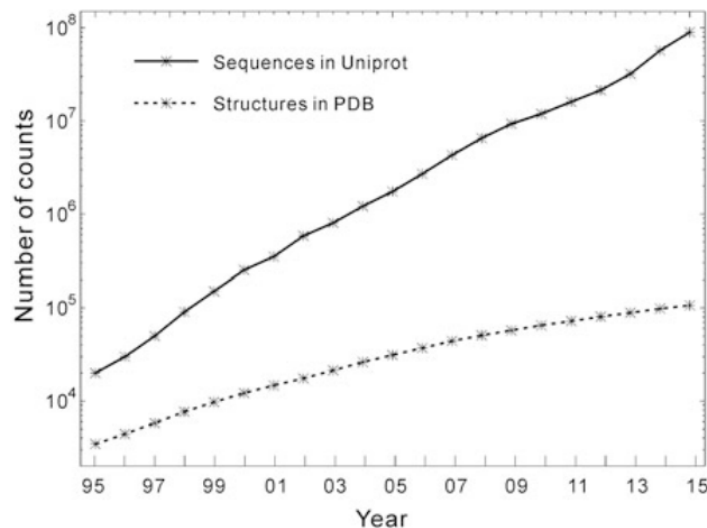


Figure 2.4—Evolution dans le temps du nombre de séquences de protéines disponibles et de structures protéiques résolues durant les 20 dernières années.(61)

Heureusement, le nombre de familles protéiques différents existant dans la nature semble être limité, et au sein d'une famille de protéines, la similarité structurale entre deux protéines homologues peut être déduite de la similarité des séquences.

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

Ainsi, le développement d'algorithmes informatiques efficaces, qui peuvent générer des prédictions de structure 3D à haute résolution, devient probablement la seule voie pour combler cette lacune.(50)

Selon que des protéines similaires ont été, expérimentalement, résolues ou non, les méthodes de prédiction de la structure des protéines peuvent être regroupées en deux catégories.

Si des protéines de structure similaire sont identifiées à partir de la bibliothèque PDB, le modèle cible peut être construit en copiant et en affinant le cadre des protéines résolues (modèles). Cette procédure est appelée "template-based modeling"(TBM) ou, plus fréquemment, « modélisation de la structure d'une protéine par homologie ». Elle sera abordée dans la partie suivante.

Par contre, si les modèles de protéines ne sont pas disponibles, nous devons construire les modèles 3D à partir du zéro. Cette procédure a reçu différents noms, par exemple la modélisation « ab initio »; « modélisation de novo »; « modélisation basée sur la physique » ; ou encore « la modélisation libre ».(61)

Dans ce chapitre, le terme de modélisation « ab initio » est uniformément utilisé pour éviter toute confusion. Contrairement à la modélisation basée sur des modèles, une procédure de modélisation « ab initio » réussie pourrait aider à répondre aux questions fondamentales sur la manière et la raison pour laquelle une protéine adopte une structure spécifique parmi de nombreuses possibilités.(62)

Cependant, malgré les progrès récents dans ce domaine, les prédictions de novo sont limitées à des protéines relativement petites et manquent de précision par rapport aux modèles comparatifs.(50)

II.2.4.1.1 Modélisation par homologie

La connaissance de la structure (3D) des protéines est une information capitale. Un moyen, rapide et disant facile, d'obtenir cette structure est de la prédire par des moyens bioinformatiques. La technique de prédiction la plus précise actuellement est la modélisation par homologie. Celle-ci est basée sur la similitude de structure entre deux protéines de séquences similaires.

La modélisation comparative homologique prédit la structure 3D d'une séquence protéique donnée (cible) en se basant principalement sur son alignement avec une ou plusieurs protéines connues (modèles). C'est l'étape critique de cette méthode (Cf. Figure 2.5).(60)

La modélisation par homologie est un type spécifique de modélisation comparative dans lequel la même origine évolutive est partagée par les protéines modèles et les protéines cibles. Les programmes informatiques et les serveurs web les plus utilisés pour le processus automatisé de modélisation de l'homologie sont les suivants : PSI-PRED et MODELER.(63)

La banque de données des protéines utilisé est la PDB. C'est un répertoire mondial de dépôt d'informations sur la structure 3D des protéines et des acides nucléiques. Ces molécules proviennent de l'ensemble des règnes biologiques. Les structures 3D sont issues principalement d'analyses par diffraction des rayons X, les autres d'analyses par RMN ou de modélisations moléculaires.

La PDB est gratuitement accessible par Internet et contient un grand nombre d'informations complémentaires comme la séquence ou la phylogénie des macromolécules. Les séquences d'acides aminés sont codées selon une procédure qui fait correspondre une lettre de l'alphabet à chaque acide aminé.(64)

Cette approche est couramment utilisée dans de nombreuses applications biologiques et recherches médicales pour générer des modèles structurels fiables. Généralement, l'effort de calcul pour un projet de modélisation est inférieur à 2 heures. Cependant, cela ne comprend pas le temps nécessaire à la visualisation et à l'interprétation du modèle, qui peut varier en fonction de l'expérience personnelle de travail avec les structures protéiques.(50)

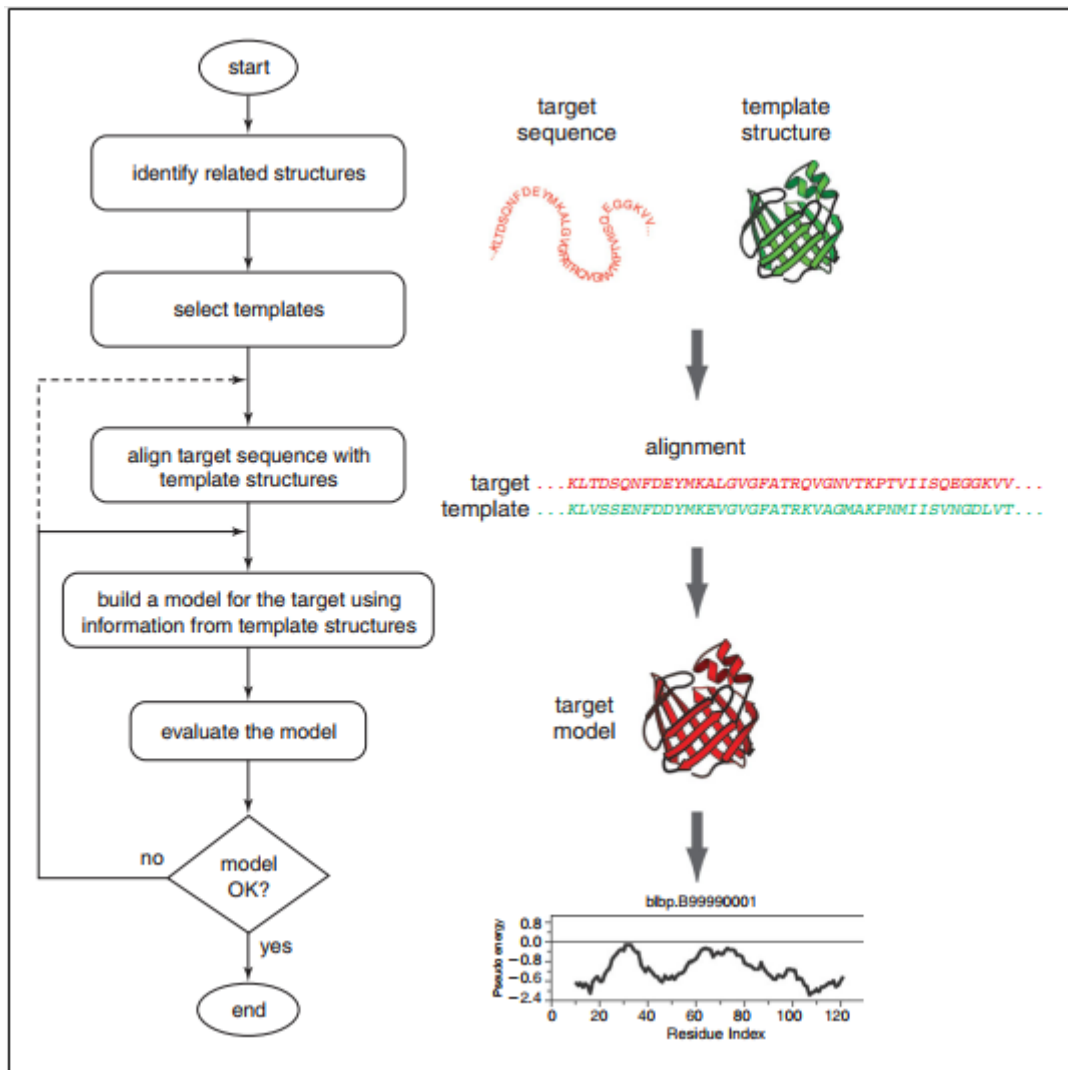


Figure 2.5– Algorithme représentant les étapes de réalisation de la modélisation homologue.(60)

II.2.4.1.1.1 Concept d« homologie entre les protéines »

Deux protéines sont homologues si elles sont liées par des processus évolutifs naturels. De nombreuses protéines homologues sont suffisamment proches pour que leurs séquences d'acides aminés puissent être alignées de telle sorte que le nombre de paires d'acides aminés similaires ou identiques aux positions alignées soit supérieur à celui attendu par hasard.

Une similarité de séquence globale suffisante entre protéines implique une homologie et une similarité de conformation. Cette idée est à la base des méthodes de modélisation des protéines par homologie dont on va discuter. (20)

Des similitudes étroites entre les séquences de protéines permettent de regrouper en familles des protéines de fonctions différentes ou provenant de différents organismes. Ce regroupement fournit preuve d'homologie, indiquant que tous les membres d'une famille sont apparentés et dérivés, du point de vue de l'évolution, d'un ancêtre commun.(3)

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

De nombreuses informations sur la structure primaire et secondaire des protéines sont stockées dans diverses bases de données. La détermination de la séquence des protéines est maintenant un travail de routine dans les laboratoires de biologie moléculaire.

Par conséquent, le taux de publication des données primaires a augmenté de façon spectaculaire au cours des dernières années. Les séquences de plus de 100000 protéines sont désormais disponibles.(18)

II.2.4.1.1.2 Principales étapes

En pratique, la modélisation homologue est un processus à plusieurs étapes qui peut être résumée en quatre parties (Cf. Figure 2.6) :

1. Sélection de modèles
2. Alignement cible-modèle
3. Construction de modèle
4. Evaluation de la qualité du modèle

Les deux premières étapes sont souvent fusionnées, car les méthodes les plus courantes d'identification de modèles dépendent de la production des alignements de séquences d'acides aminés. Ces processus peuvent être effectués par itération afin d'améliorer la qualité du modèle final.(50)

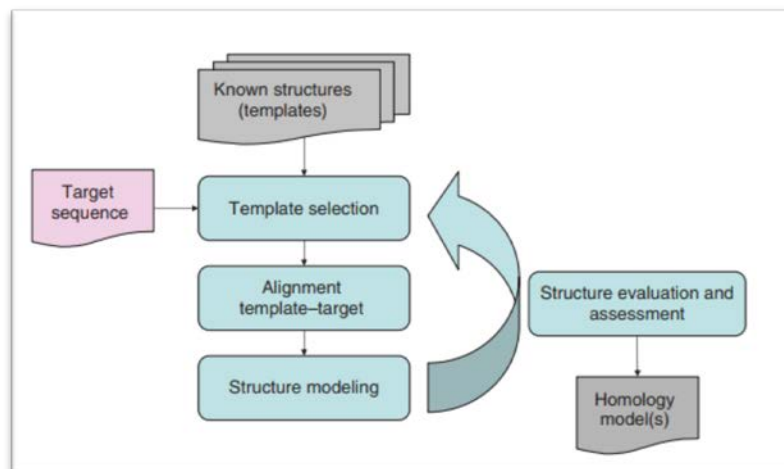


Figure 2.6 – Principales étapes de la modélisation homologue des protéines.(50)

II.2.4.1.1.2.1 Sélection des modèles

La recherche de protéines homologues est basée sur la comparaison des séquences d'acides aminés des protéines de structures connues avec la séquence d'acides aminés de la protéine que l'on veut construire. Cette recherche peut se faire par différentes méthodes telles que celle de Needleman&Wunsch, FASTA (Fast Alignment) ou BLAST (Basic Local Alignment Search Tool).(65)

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

Les algorithmes FASTA et BLAST permettent de comparer en un temps très court une ou plusieurs séquences cibles aux milliers de séquences contenues dans les bases de données telles que la PDB.

Ils procèdent par comparaison de paires d'acides aminés. Chaque comparaison obtient un score qui reflète l'identité ou le degré de similarité entre les séquences comparées. Plus le score est élevé, plus le degré de similarité est important. Le résultat final se présente sous la forme d'un alignement des séquences trouvées accompagnées de scores d'identité et de similarité. (64)

Cependant, ces programmes ne fonctionnent bien que pour l'alignement de séquences présentant de fortes similitudes. Des méthodes telles que PSI-BLAST et ScanPS ont récemment augmenté la possibilité de détecter des homologues éloignés.(50)

II.2.4.1.1.2.2 Alignement cible-modèle

Définition :

Un alignement de séquences est une table à deux dimensions dans laquelle les lignes représentent les séquences protéiques ou nucléotidiques et les colonnes représentent les positions des résidus. Les séquences sont placées sur cette grille de telle manière que:

(a) la position absolue des résidus, c'est-à-dire leur position dans l'ordre de la séquence non alignée, soit préservée.

(b) les résidus similaires dans toutes les séquences soient placés dans la même colonne en introduisant, si nécessaire, des cellules vides appelées *gap* (trou).(66)

Méthodes :

De très nombreuses méthodes ont été développées pour aligner des séquences protéiques. On peut regrouper les algorithmes d'alignement suivant deux stratégies différentes. La première recherche à établir une similarité entre des séquences prises dans leur intégralité (alignement global) et la seconde s'intéresse seulement à des régions similaires (alignement local). Comme exemple, les programmes issus de l'algorithme BLAST comparent les séquences par des alignements locaux.(67)

Ces stratégies correspondent à des contextes biologiques différents. L'alignement global ne se justifie que si l'on considère des séquences supposées largement similaires sur toute leur longueur.

La recherche de similarité locale se base sur le fait que les sites fonctionnels (sites catalytiques des enzymes, par exemple) sont localisés dans des régions relativement courtes et bien conservées malgré des insertions, mutations ou délétions dans les autres régions de la séquence. Néanmoins, il faut noter que de courts segments de séquences similaires n'ont pas nécessairement la même structure 3D.(66)

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

Les méthodes d'alignement local ont l'avantage d'être plus sensibles que celles d'alignement global. En effet, elles permettent de mettre en évidence l'homologie de séquences dont le pourcentage d'identité est faible.

Une fois qu'une famille d'homologues est identifiée, pouvant ne comporter qu'un seul membre, les fichiers de coordonnées atomiques de ces protéines sont insérés dans le module PRODAT de SYBYL. PRODAT est la banque de structures locale du logiciel à partir de laquelle le module COMPOSER nous permet de construire des modèles.(64)

Un alignement peut contenir des lacunes car, au cours de l'évolution d'une séquence protéique, des événements de mutation et d'insertion/délétion peuvent se produire dans l'ADN codant. La formation d'un alignement précis des séquences d'acides aminés est absolument essentielle pour la construction d'un modèle utile.(20)

Ainsi, obtenir un alignement de qualité est un préalable essentiel à la modélisation par homologie. Si pas moins de 40% des résidus identiques dans un alignement optimal, il est probable que les méthodes standard d'alignement de séquences permettent d'obtenir des résultats satisfaisants.

Lorsque les séquences ont divergé beaucoup plus loin que ce seuil, il est impossible de déterminer l'alignement correct à partir de paires de séquences seulement, mais il peut être possible de déterminer un alignement correct à partir d'une analyse de l'ADN ou d'une comparaison des structures, à condition bien sûr qu'elles soient disponibles.(68)

De même, les alignements de séquences multiples sont beaucoup plus informatifs que les alignements d'une seule paire de séquences. (20)

II.2.4.1.1.2.3 Construction du modèle

Lorsque l'alignement est prêt, la construction du modèle proprement dite peut commencer. Il suffit de copier les coordonnées des résidus du modèle qui apparaissent dans l'alignement avec la séquence de cible. Si les deux résidus alignés diffèrent, seules les coordonnées du squelette peuvent être copiées. S'ils sont identiques, on peut également inclure la chaîne latérale.(68)

Bien que la théorie derrière la construction d'un modèle d'homologie de protéine soit compliquée, l'utilisation des programmes disponibles est relativement facile. Plusieurs programmes de modélisation sont disponibles, utilisant différentes méthodes pour construire les structures 3D, tels que WHAT IF, Des serveurs Web tels que SWISS MODEL et le serveur ROSETTA permettent de générer encore plus facilement un modèle.(65)

Dans les méthodes de mise en correspondance par segment, la cible est divisée en courts segments, et l'alignement se fait sur les segments plutôt que sur la protéine entière.

Un problème régulièrement rencontré dans la modélisation de l'homologie est celui des boucles. Une directive générale est que toute insertion/boucle plus longue que d'environ cinq résidus doit être omise. Il existe des programmes qui tentent de modéliser les boucles, comme MODELLER et son serveur de modélisation de boucles plus spécialisé, MODLOOP.(69)

II.2.4.1.1.2.4 Evaluation de la qualité du modèle

Tout modèle homologue contient des erreurs. Une étape essentielle du processus de modélisation homologue est donc la vérification du modèle.

Le nombre d'erreurs, pour une méthode donnée, dépend principalement de deux valeurs :

1. Le pourcentage d'identité de séquence entre le modèle et la cible : s'il est supérieur à 90%, la précision du modèle peut être comparée aux structures déterminées par la cristallographie, à l'exception de quelques chaînes latérales individuelles.

De 50 % à 90 % d'identité, les erreurs locales dans les coordonnées modélisées sont considérablement plus importantes.

Si l'identité de la séquence tombe à 25%, l'alignement s'avère être le principal goulot d'étranglement pour la modélisation de l'homologie conduisant souvent à des erreurs très importantes.

2. Le nombre d'erreurs dans le modèle : Les erreurs dans un modèle deviennent moins problématiques si elles peuvent être localisées. Par exemple, il est peu important qu'une boucle éloignée du site actif d'une enzyme soit placée incorrectement. (68)

La plupart des méthodes utilisées pour la vérification des modèles peuvent également être appliquées aux structures expérimentales (et donc aux modèles utilisés pour la construction des modèles). Une vérification détaillée est essentielle lorsque l'on tente de tirer de nouvelles informations du modèle, soit pour interpréter ou prédire des résultats expérimentaux ou planifier de nouvelles expériences.(67)

A presque toutes les étapes, des choix doivent être faits. Le modélisateur ne peut jamais être sûr de faire les meilleurs, et donc une grande partie du processus de modélisation consiste à réfléchir sérieusement à la façon de jouer entre plusieurs choix apparemment similaires.(70)

II.2.4.1.1.3 Applications dans la découverte des médicaments

En l'absence de structures expérimentales, la modélisation homologue « in silico » constitue une alternative fiable et rentable pour générer des modèles raisonnablement précis pour la découverte de médicaments basés sur la structure.

À ce jour, la modélisation homologue a été utilisée avec succès pour identifier des résultats positifs en utilisant le Docking à haut débit (High Throuthput Docking HTD), suggérer des modes de liaison précis et des interactions ligand-cible, faciliter les expériences de mutagenèse, rationaliser les données SAR et guider l'optimisation vers des ligands plus puissants (Cf. Figure 2.7).(71)tous ces éléments peuvent être utilisés pour rationaliser, accroître l'efficacité, la rapidité et la rentabilité du processus de découverte de médicaments.

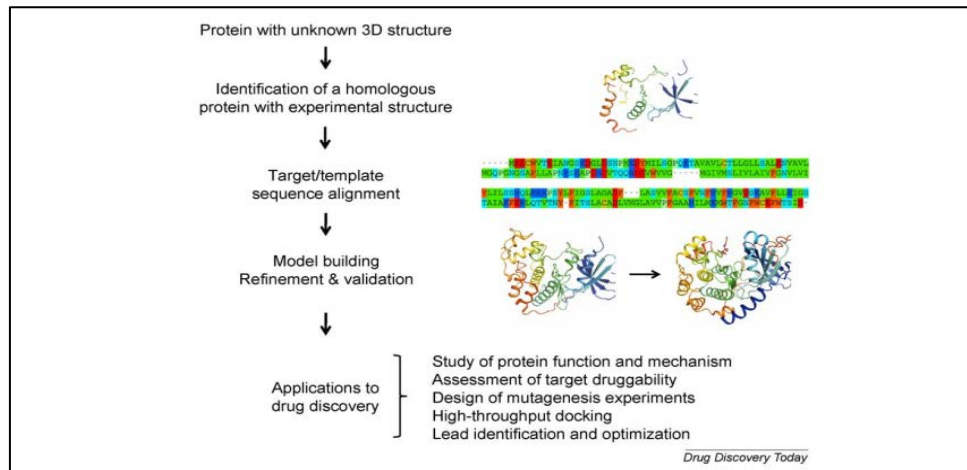


Figure 2.7 - Aperçu du processus de modélisation homologique et ses applications dans la découverte de médicaments.(71)

II.2.4.1.1.4 Exemple d'application sur le SARS-Cov

L'analyse de la structure et de la fonction d'une protéine a généralement de nombreuses applications, depuis les expériences de mutagenèse de base jusqu'aux différentes étapes du processus de découverte de médicaments.

On donne ici un seul exemple d'une percée dans la conception de médicaments qui utilise la modélisation homologique. Le syndrome respiratoire aigu sévère (SARS) a été identifié en Chine en 2002, et s'est rapidement propagé à d'autres pays. La cause en était un nouveau coronavirus (CoV). Peu après, les génomes entiers de différentes souches de SARS-CoV ont été résolus. La protéase principale (Mpro), qui joue un rôle important dans la réplication du virus, est devenue une cible médicamenteuse immédiate.

La séquence de CoV-Mpro présente une identité de 40 % et 46 % avec celle du coronavirus de la gastro-entérite transmissible (TGEV-Mpro) et au coronavirus humain 229E, et des structures aux rayons X étaient déjà disponibles.(72)

Plusieurs groupes ont publié le modèle d'homologie de la protéase en mai 2003. Une comparaison de l'inhibiteur complexé avec le TGEV-Mpro avec les complexes d'inhibiteurs disponibles dans PDB a donné un mode de liaison d'inhibiteur similaire dans le complexe de humanrhino-virus de type 2 (HRV2) protéinase 3Cavec AG7088.

A l'époque, l'AG7088 était en essais cliniques pour le traitement du rhino-virus humain qui cause le rhume. L'AG7088 s'est arrimé au site de liaison au substrat du modèle SARS-CoVMpro, ce qui indique qu'il serait un bon point de départ pour la conception d'un médicament anti-SARS. Peu après, il a été démontré que l'AG7088 a également une activité anti-SARS in vitro. (65)

II.2.4.1.2 Modélisation de novo : prédiction « ab initio »

La modélisation par homologie est limitée par le besoin absolu d'une structure template (modèle). Pour près de la moitié des protéines, un template correspondant ne peut être détecté ou n'est pas encore connu. Dans ces cas, les méthodes de prédiction de novo qui tentent de prédire la structure 3D à partir de la seule séquence, sont la seule alternative.(66)

II.2.4.1.3 Principe général

Selon l'hypothèse thermodynamique d'Anfinsen (1973), les protéines ne sont pas assemblées dans leur structure native par un processus biologique, mais le repliement est un processus purement physique qui ne dépend que de la séquence spécifique d'acides aminés de la protéine et du solvant environnant.(59)

L'héritage d'Anfinsen suggère qu'une protéine dans son environnement natif se replie *in cellulo* vers une conformation thermodynamiquement stable (minimum d'énergie libre de Gibbs) dans une échelle de temps allant de la milliseconde à plusieurs heures.

Cette observation est à l'origine même des méthodes de prédiction dites « *ab initio* » où on cherche à reproduire la forme repliée d'une protéine uniquement par mécanique moléculaire. La structure prédite étant la forme associée au minimum d'énergie potentielle des forces appliquées au système.(73)

Cette idée définit bien le problème de la prédiction de la structure des protéines, car elle permet de déduire la structure macroscopique de nombreuses protéines à partir de quelques types d'interactions microscopiques entre les constituants de la protéine.

D'un autre côté, la prédiction de la structure des protéines reste extrêmement complexe, car même de courtes séquences d'acides aminés peuvent former un nombre abondant de structures géométriques parmi lesquelles il faut identifier le minimum d'énergie libre.

Du coup, de nombreuses approches de la prédiction computationnelle de la structure des protéines utilisant les premiers principes ont été développées au cours de la dernière décennie, basées sur l'hypothèse thermodynamique d'Anfinsen.(59)

II.2.4.1.3.1 Exigences pour une modélisation de Novo réussie

Généralement, la modélisation « ab initio » effectue une recherche de conformation sous la direction d'une fonction d'énergie conçue. Cette procédure génère généralement un certain nombre de conformations possibles, ensuite, des modèles finaux sont sélectionnés parmi eux.(62)

Par conséquent, une modélisation « ab initio » réussie dépend de trois facteurs :

(1) une fonction énergétique précise avec laquelle la structure native d'une protéine correspond à l'état le plus stable sur le plan thermodynamique, par rapport à toutes les structures leurres possibles.

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

(2) une méthode de recherche efficace qui peut identifier rapidement les états de faible énergie par une recherche conformationnelle.

(3) une stratégie permettant de sélectionner des modèles quasi-natifs à partir d'un ensemble de structures leurres.(61)

II.2.4.1.3.2 Logiciels et méthodes

La prédiction « ab initio » peut potentiellement fournir un grand nombre de modèles de structure en investissant beaucoup moins de temps, d'argent et d'efforts humains que les approches expérimentales. Les récentes expériences de prédiction de structure CASP3 et CASP4 montrent que Rosetta est probablement la meilleure méthode actuelle pour la prédiction de structures des protéines.

La méthode remportant le plus de succès depuis plusieurs années est ROSETTA développée par le groupe de Baker. Les auteurs considèrent des fragments de 3 à 9 résidus de long et réalisent une exploration conformationnelle très large grâce à de très fortes capacités de calculs.(74)

Rosetta est basé sur une image du repliement des protéines dans laquelle les segments de séquence locaux alternent rapidement entre différentes structures locales possibles, et le repliement se produit lorsque les conformations et les orientations relatives de ces segments locaux se combinent pour former des structures globales à faible énergie.(62)

Deux types de méthodes sont à distinguer : les méthodes *ab initio* « pures » et les méthodes « *de novo* ».

Les méthodes *ab initio* « pures » reposent uniquement sur les propriétés physicochimiques des protéines et sur la recherche de la structure de plus basse énergie parmi l'ensemble des structures possibles.

Les méthodes les plus récentes utilisent une approche hiérarchique comme dans ASTRO-FOLD dans laquelle les hélices sont prédites en premier, puis les brins et la topologie globale des feuilletts. Les boucles sont modélisées à part.

L'ensemble des contraintes issues de ces prédictions isolées sont enfin utilisées pour la prédiction d'une structure 3D. Ces méthodes nécessitent une exploration de l'espace conformationnel très importante. Elles restent donc pour le moment applicables uniquement à de petites protéines de moins d'une centaine de résidus.(70)

Les méthodes dites « *de novo* » exploitent les informations obtenues de l'analyse des structures 3D connues.

Actuellement, les approches les plus performantes sont les méthodes d'assemblage de fragments. Ces méthodes reposent sur l'hypothèse selon laquelle, même si nous n'avons pas encore observé tous les repliements possibles, nous avons probablement vu presque toutes les sous-structures. (75)

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

Étant donné que la relation séquence-structure n'est pas suffisamment forte pour déterminer avec certitude la structure de fragments de séquence, la première étape consiste à réduire le nombre de conformations possibles pour un segment de séquence donné. Ces conformations sont extraites d'une banque des structures locales connues. Les fragments prédits sont ensuite assemblés pour générer de nombreux modèles 3D possibles.(62)

II.2.4.1.3.3 Limites des méthodes « ab initio »

La modélisation « ab initio » pure est encore très coûteuse et inefficace mais les méthodes hybrides (homologie + « ab initio ») telles que l'assemblage de fragments sont plus performantes.

L'une des principales limites des méthodes de prédiction de novo des protéines est l'extraordinaire quantité de temps informatique nécessaire pour résoudre avec succès la confirmation native d'une protéine.

Les méthodes distribuées, telles que Rosetta, ont tenté de remédier à ce problème en recrutant des personnes qui se portent volontaires pour consacrer du temps libre à l'ordinateur de leur domicile afin de traiter les données.

Cependant, le temps et le nombre d'ordinateurs nécessaires pour réaliser cet exploit sont énormes, près de deux ans et environ 70 000 ordinateurs personnels, respectivement.(75)

II.2.4.1.4 « ab initio » vs Homologie

La modélisation de la structure des protéines de novo se distingue de la modélisation par homologie par le fait qu'aucun homologue résolu de la protéine d'intérêt n'est connu, ce qui rend extrêmement difficile la prédiction de la structure d'une protéine à partir de la séquence d'acides aminés.

Les méthodes basées sur des modèles (par homologie) commencent par une séquence, prédisent la structure secondaire et tentent de trouver une structure modèle et/ou des fragments de structures existantes dans la PDB qui se plieront de manière similaire à la séquence cible (Cf. Figure 2.8).

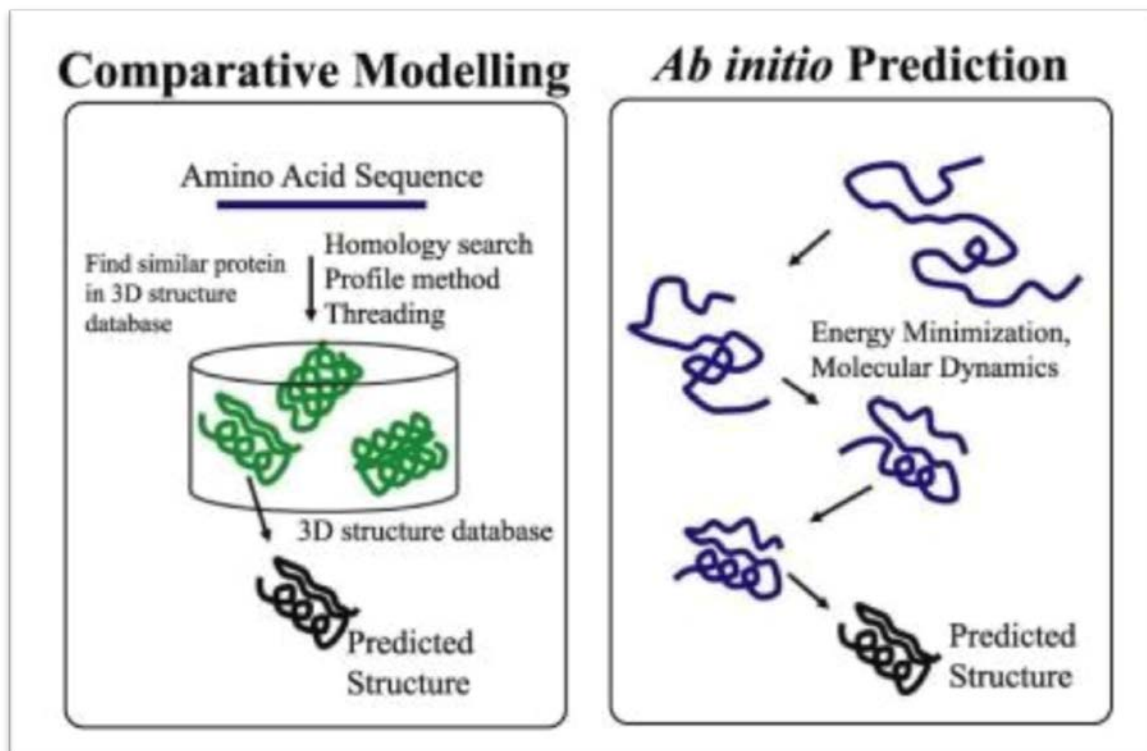


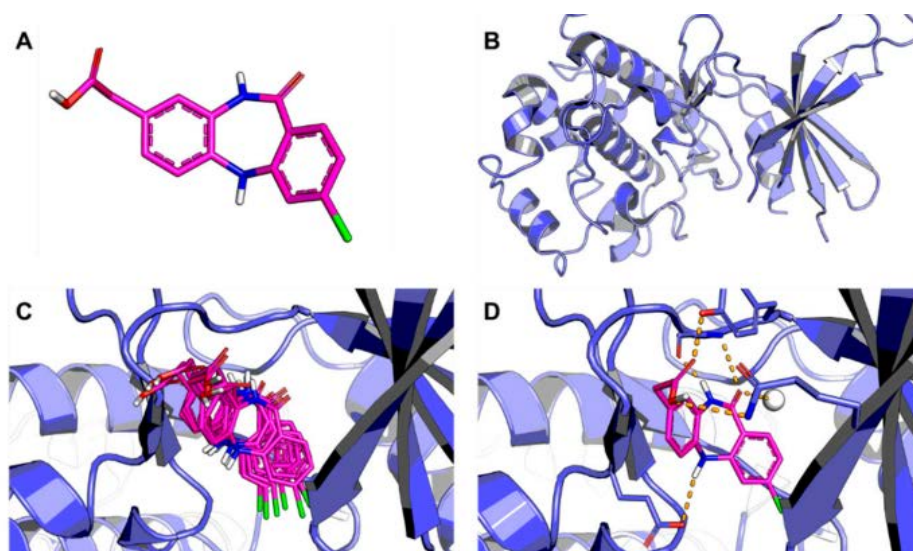
Figure 2.8- Différence entre la modélisation par homologie et la modélisation de novo.(76)

II.2.4.2 « DOKING » Moléculaire

Le **Docking moléculaire** est l'une des méthodes les plus fréquemment utilisées dans le cadre de la SBDD en raison de sa capacité à prédire, avec un degré de précision substantiel, la conformation des ligands dans le site de liaison cible approprié (CF. Figure 2.9).

Après la mise au point des premiers algorithmes dans les années 1980, le Docking moléculaire est devenu un élément essentiel de la stratégie de découverte de médicaments. Par exemple, les enquêtes impliquant des événements moléculaires cruciaux, y compris les modes de liaison des ligands et les interactions intermoléculaires correspondantes qui stabilisent le complexe ligand-récepteur, peuvent être facilement réalisés.

En outre, les algorithmes de Docking moléculaire effectuent des prédictions quantitatives de l'énergie de liaison, fournissant des classements des composés insérés en fonction de l'affinité de liaison des complexes ligand-récepteur.(77)



- (A) Structure tridimensionnelle du ligand
- (B) Structure tridimensionnelle de la cible
- (C) Le ligand est inséré dans la cavité de liaison de la cible et les éventuelles conformations sont explorées
- (D) La conformation de liaison la plus probable et les interactions intermoléculaires correspondantes sont identifiées.

Le squelette de la protéine est représenté sous forme de dessin animé. Le ligand (carbone en magenta) et les résidus du site actif (carbone en bleu) sont représentés par des bâtons. L'eau est représentée par une sphère blanche et les liaisons hydrogène sont indiquées par des lignes pointillées.

Figure 2.9 - Aperçu du processus de Docking moléculaire.(77)

II.2.4.2.1 Définition

Le Docking (amarrage) moléculaire est une méthode empirique qui permet de prédire l'affinité entre deux molécules (qui sont généralement le ligand et la protéine), entre autre, la position et l'orientation la plus favorable pour un ligand venant en interaction avec une protéine cible. Cet outil est actuellement très sollicité dans la recherche de nouvelles molécules à visée thérapeutique assistée par ordinateur.

Il est utilisé pour faire des criblages très rapides et peu coûteux afin d'identifier une molécule pouvant avoir une forte activité sur une protéine cible ou au contraire d'identifier la cible d'une molécule ayant des activités biologiques intéressantes. D'autres parts, lorsque la structure 3D d'une protéine cible est définie, le Docking permet de mettre en évidence les régions et les résidus à explorer pour optimiser l'affinité d'un ligand avec cette cible.

Ainsi, il est également utilisé pour optimiser la sélectivité d'une molécule entre deux ou plusieurs protéines.(51)

Les techniques de Docking moléculaire visent à prédire le meilleur mode de liaison d'un ligand à une protéine cible. Elles consistent à générer un certain nombre de conformations/orientations possibles, c'est-à-dire de poses, du ligand dans le site de liaison de la protéine.

Pour cette raison, la disponibilité de la structure tridimensionnelle de la cible moléculaire est une condition nécessaire ; il peut s'agir d'une structure résolue de manière expérimentale (par exemple par la méthode des rayons X ou RMN, par exemple) ou une structure obtenue par des techniques informatiques (telles que la modélisation par homologie).(78)

II.2.4.2.2 Théorie de base

Le Docking moléculaire consiste à simuler la conformation optimale selon la complémentarité, ce qui permet de prédire et d'obtenir l'affinité de liaison et le mode d'interaction entre le ligand et la cible.

La figure 2.10 (A) montre le premier modèle proposé « lock-and-key », qui se réfère à l'arrimage rigide des récepteurs et des ligands pour trouver l'orientation correcte de la "clé" afin d'ouvrir la « serrure ». Ce modèle souligne l'importance de la complémentarité géométrique.

Cependant, le processus réel de Docking est si flexible que les récepteurs et les ligands doivent changer leur conformation pour bien s'adapter les uns aux autres. Nous avons donc développé un « modèle d'ajustement induit »(Cf. Figure 2.10 (B)). (79)

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

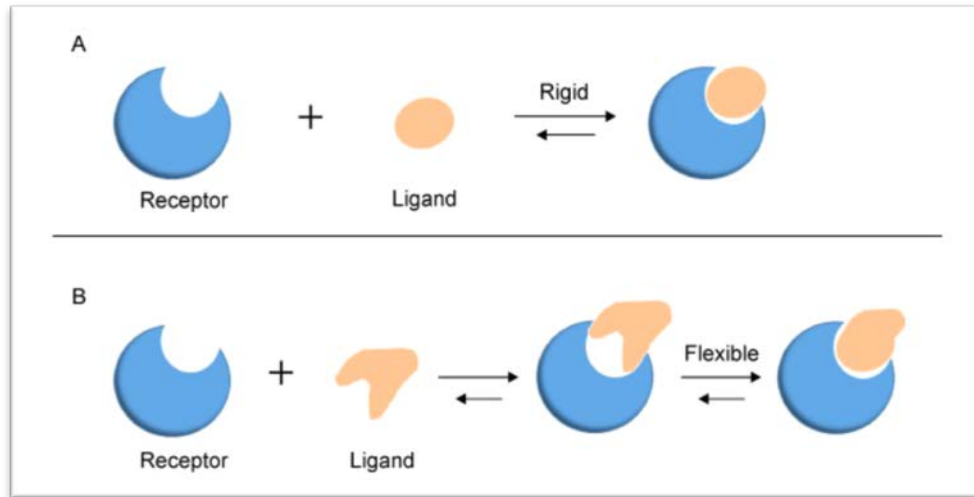


Figure 2.10 -Deux modèles de Docking moléculaire. (A) Un modèle de type "lock-and-key". (B) Modèle d'ajustement induit. (79)

Sur la base de la complémentarité géométrique et énergétique, on garantit que les récepteurs et les ligands obtiendront la structure la plus stable de manière à minimiser l'énergie libre.(80)

Comme le montre la figure 2.11, un logiciel de Docking moléculaire peut nous aider à trouver la conformation et l'orientation optimale en fonction de la complémentarité avec un algorithme spécifique (Docking proprement dit), puis d'appliquer une fonction de score pour prédire l'affinité de liaison et le mode interactif entre le ligand et la cible (Scoring) en utilisant plusieurs logiciels informatiques. (79)

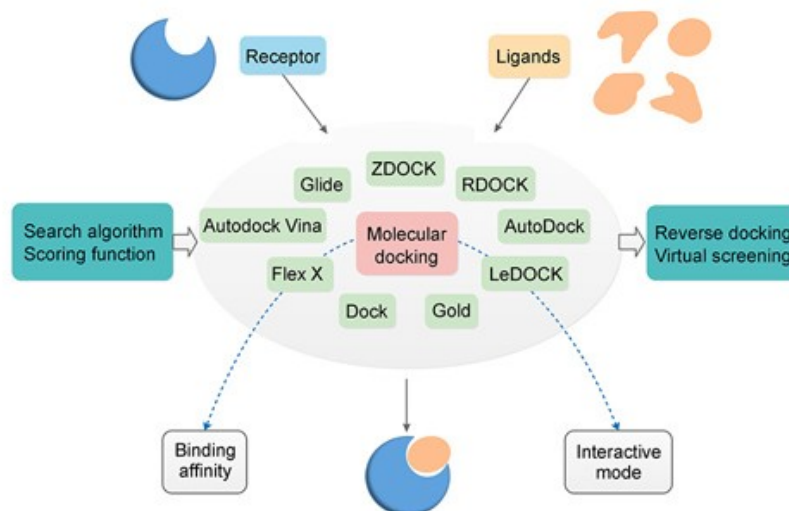


Figure 2.11 – Principe général du Docking moléculaire avec les différents logiciels utilisés.(81)

Le processus du Docking est itératif et chaque passe de calcul s'articule en deux étapes. Tout d'abord, une portion limitée de l'espace des conformations du complexe protéine-ligand est explorée afin d'améliorer la pose (orientation et/ou conformation) du ligand dans le site actif

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

de la protéine. L'étape suivante fait intervenir une fonction de score qui évalue la qualité de la pose générée.

Cette estimation « in silico » de l'affinité du ligand pour la cible est basée sur un examen simplifié des interactions entre les deux partenaires. La répétition du cycle de recherche associé à l'estimation du score guidant l'exploration de l'espace, doit assurer la convergence de l'algorithme vers un état représentatif du minimum global de l'énergie libre d'association.(13)

II.2.4.2.3 Logiciels et algorithmes

En principe, un Docking peut être fait de façon manuelle, directement par le modélisateur, en plaçant le ligand dans le site actif de la protéine à l'aide d'une interface graphique. Ensuite, la géométrie de l'ensemble est optimisée de manière à corriger les problèmes stériques et obtenir un complexe énergétiquement stable. Cette approche est appliquée quand on a une idée précise du mode d'interaction réel du ligand.

Néanmoins, le plus souvent, le mode d'interaction réel n'est pas connu. Dans ce cas, tester manuellement toutes les conformations et orientations des ligands s'avère impossible d'un point de vue pratique, même en considérant la protéine comme un corps rigide.

Pour contourner cette difficulté, les algorithmes de Docking ont été conçus pour rechercher de façon objective, rapide et efficace les modes d'association protéine-ligand les plus favorables et de faire correspondre les distances entre chacun des points pharmacophoriques du ligand et la protéine.(82)

Un grand nombre de programmes (logiciels) de Docking moléculaire ont été publiés et utilisent des algorithmes spécifiques dont un certain nombre a également été développé pour utiliser les informations provenant des protéines et des ligands. (83)

L'un des critères de classification des algorithmes est la façon dont les ligands sont traités pendant le Docking. Dans certains de ces algorithmes, le ligand est construit de manière incrémentielle, à partir d'un "fragment de base" arrimé. Les programmes qui suivent cette approche comprennent Hammerhead, DOCK et FlexX.

Dans d'autres programmes, tels qu'AutoDock, Genetic Optimization for Ligand Docking (GOLD), ICM-Dock et QXP, le ligand est traité dans son intégralité.

En plus de la flexibilité du ligand, il peut être souhaitable de garder au moins une partie du récepteur flexible afin de permettre les changements de conformation qui sont nécessaires pour le ligand, un phénomène appelé "ajustement induit".

En raison de leur coût de calcul élevé, peu de programmes de Docking permettent la flexibilité des protéines. Les exceptions notables sont les dernières versions d'AutoDock , FlexE, QXP, et la dernière version de ICM-Dock.(84)

II.2.4.2.4 Fonctions de scores

Après la génération de milliers d'orientations des ligands et son échantillonnage, la seconde étape de l'amarrage moléculaire est de définir une fonction objective pour évaluer l'interaction : la fonction de score. Le score est une donnée numérique qui permet une approximation de l'énergie libre résultant du passage de la forme libre de la protéine et du ligand à la formation du complexe.(2)

La fonction de score est une approximation de l'affinité entre la protéine et le ligand. Les fonctions de score sont spécifiquement conçues pour compléter ces algorithmes de Docking car elles évaluent les interactions entre les composés et les cibles potentielles, prédisant ainsi leur activité biologique.(63)

Actuellement, les fonctions de score sont regroupées en trois grands types : champ de force, empirique et knowledge-based.

Les fonctions basées sur le champ de force : consistent en une somme de termes énergétiques. L'énergie potentielle tient généralement compte des termes liés (longueur de la liaison, angle, dièdres) et non liés (Van der Waals, électrostatique).Ce type de fonction néglige généralement les effets du solvant et les entropies.

Les fonctions de score empiriques : sont dérivées des relations quantitatives structure-activité qui ont été idéalisées pour la première fois par Hansh et Fujita. L'objectif est de prédire l'affinité de liaison avec une grande précision en utilisant les données expérimentales connues.(83)

Les fonctions dérivées d'informations descriptives (knowledge-based) : utilisent des méthodes statistiques pour extraire des caractéristiques sur les interactions atomiques à partir de complexes protéine-ligand déterminés expérimentalement. Par exemple, PMF, Drug Score et SMOG sont des programmes appliquant cette classe de fonctions de score.(85)

Des études comparatives qui évaluent les fonctions de score des méthodes de Docking utilisent des caractères d'évaluation tels que la pose de la liaison, l'affinité de liaison et le classement des véritables liants.

Wang et al. ont évalué les performances de quatorze fonctions de score différentes en utilisant 800 complexes dans la base de données PDB. La performance a été évaluée par les affinités de liaison prédites des complexes protéine-ligand par les différentes fonctions.(86)

II.2.4.2.5 Classification des méthodes utilisées

Les méthodes de Docking sont divisées, selon le degré de flexibilité des molécules impliquées dans le calcul, en trois catégories, rigide, semi-flexible et flexible.

II.2.4.2.5.1 Docking rigide : « Lock and Key »

Le fait que les événements de reconnaissance moléculaire puissent être des interactions hautement spécifiques n'est pas nouveau pour les chimistes médicaux. Des analogies avec un concept de "serrure et clé" pour décrire ces processus ont été proposées pour la première fois il y a un siècle par Fischer (1894) et par Ehrlich (1909).(87)

Cette approximation est analogue au modèle de liaison "lock and key" et est principalement utilisée pour le Docking protéine-protéine, où le nombre de degrés de liberté conformationnels est trop élevé pour être échantillonné. (78)

Ce modèle implique un appariement précis de composants immuables, cela implique que le ligand et la protéine sont considérés comme des entités rigides et seuls les trois degrés de liberté en translation et trois degrés de liberté en rotation sont considérés pendant l'échantillonnage. Cela simplifie grandement le problème du Docking en réduisant le nombre de degrés de liberté de plusieurs milliers à seulement six. (78,87)

II.2.4.2.5.2 Docking semi-flexible

Une seule des molécules, le ligand, est flexible, tandis que la protéine est rigide. Ainsi, les degrés de liberté conformationnels du ligand sont échantillonnés, en plus des six degrés de translation et de rotation. Ces méthodes supposent qu'une conformation fixe d'une protéine peut correspondre à celle capable de reconnaître les ligands à docker. Cette hypothèse n'est pas toujours vérifiée.(78)

II.2.4.2.5.3 Docking flexible: Modèle d'ajustement induit

Le modèle "lock-and-key" pour la liaison protéine-ligand ne peut pas expliquer tous les aspects de la spécificité des enzymes (Koshland, 1994). Par exemple, la fonction des ligands qui modulent l'activité enzymatique mais ne participent pas directement à la catalyse n'a pu être définie (Koshland, 1971).

Des observations comme celles-ci ont conduit à la proposition d'une théorie modifiée, la théorie de l'ajustement induit (Koshland, 1958), qui soutient que les ligands induisent des changements dans la structure de la protéine avant qu'un ajustement approprié puisse se produire et qu'une protéine n'est pas une entité passive et rigide pendant la liaison et considère le ligand et la protéine comme des homologues flexibles.(78,87)

Les idées d'effets d'ajustement induit et de plasticité macromoléculaire trouvent un soutien croissant au fur et à mesure que les détails structurels des événements de reconnaissance moléculaire sont élucidés.(87)

Différentes méthodes ont été introduites au fil des ans, certaines reposant sur le modèle de liaison à ajustement induit et d'autres sur la sélection conformationnelle.(78)

II.2.4.2.6 Docking inverse ou « Reverse Docking »(RD)

Le Docking inverse (ou pêche à la cible) est un effort relativement nouveau visant à identifier et valider la ou les cibles potentielles d'un ou de plusieurs composés actifs.(88)

Le Docking inverse peut être utilisé pour découvrir de nouvelles cibles pour les médicaments existants et les composés naturels, expliquer la polypharmacologie et le mécanisme moléculaire d'une substance, trouver des indications alternatives aux médicaments par le repositionnement des médicaments, et détecter les effets indésirables et la toxicité des médicaments.

Les éléments nécessaires pour le processus sont similaires à ceux des méthodes d'arrimage directes : préparation des ensembles de données, recherche de poses de ligands, scoring et classement des structures complexes (Cf. Figure 2.12). Cependant, plusieurs problèmes, notamment un coût de calcul élevé et un biais de score inter-protéines, rendent le processus assez complexe. (89)

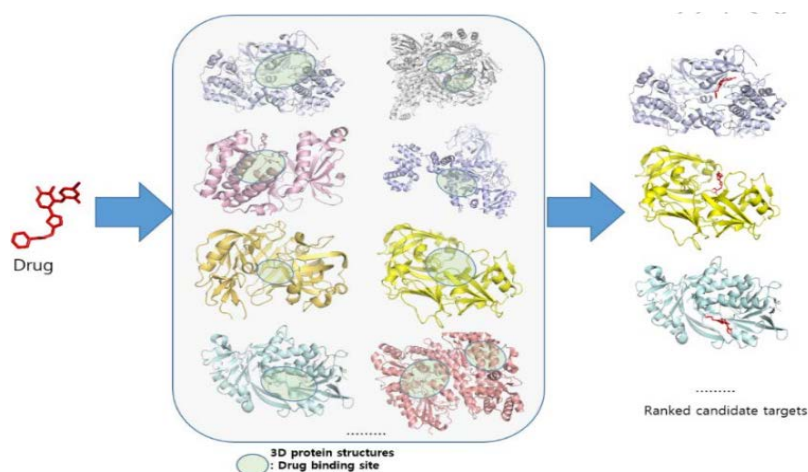


Figure 2.12- Vue d'ensemble schématique de méthode de Docking inverse.(89)

Une procédure de Docking inverse appelée INVDOCK a été introduite pour effectuer des recherches de RD informatisées dans la PDB afin d'identifier les cibles protéiques potentielles d'une petite molécule.(90)

Les programmes d'accostage moléculaire utilisés dans les processus d'arrimage inverse sont similaires à ceux des méthodes conventionnelles de Docking direct. De nombreux programmes populaires tels que GOLD, DOCK, FlexX ou Glide ont été utilisés avec quelques modifications pour rendre l'ensemble du processus plus efficace et plus précis sur le plan informatique. Le processus d'arrimage inverse est plus exigeant sur le plan informatique car il doit traiter un grand nombre de cibles protéiques.

Le bruit des scores de Docking produit des faux positifs au cours de l'identification des cibles. De nombreuses tentatives ont été faites pour surmonter le bruit inter-protéines des scores de Docking en développant un certain type de normalisation de ceux ci.(89)

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

Une nouvelle méthode de scoring qui effectue une analyse compétitive de la liaison en plus de l'évaluation de l'énergie d'interaction ligand-protéine en mécanique moléculaire. La procédure est testée sur un certain nombre de complexes ligand-protéine provenant de la PDB.(90)

Cependant, l'application de ces approches nécessite des bibliothèques de cibles appropriées. En effet, plusieurs bases de données sont actuellement disponibles pour faciliter le dépistage des RD. L'une des bases de données les plus connues pour faciliter l'identification computationnelle de cibles est Potential Drug Target Database (PDTD), qui fournit des informations sur les structures des protéines, les maladies, les fonctions biologiques et les médicaments.

En outre, des bibliothèques de cibles peuvent, également, être constituées manuellement à partir de bases de données publiques de structures cristallines et de poches de liaison, telles que la PDB, Pocketome et Therapeutic Target Database (TTD).(91)

En particulier, les bases de données PDB et TTD représentent des réservoirs d'informations bien connus, développées pour faciliter la biologie computationnelle, moléculaire et structurale, et pour fournir des données sur les cibles et les maladies, respectivement.

Le Docking inverse est un développement relativement nouveau et nécessite encore une évaluation expérimentale appropriée. Si elle réussit, elle devrait fournir une nouvelle approche de la validation des cibles.(88)

II.2.4.2.7 Docking conventionnel vs Docking inverse

Le Docking moléculaire est une méthode permettant de prédire le(s) mode(s) de liaison prédominant(s) d'un ligand avec une protéine dont la structure 3D est connue. Cette méthode est couramment utilisée dans la découverte de médicaments basée sur la structure pour l'identification des Hits et l'optimisation des Leads.

Le Docking inverse, opposé au processus conventionnel d'application du docking, par exemple dans la VS basée sur la structure, consiste à docker une petite molécule donnée dans les sites de liaison prédéfinis d'un pool de structures protéiques. La protéine identifiée (Hit) peut alors servir de cible potentielle pour la validation expérimentale. Une comparaison de ces deux paradigmes est présentée dans la figure 2.13.(92)

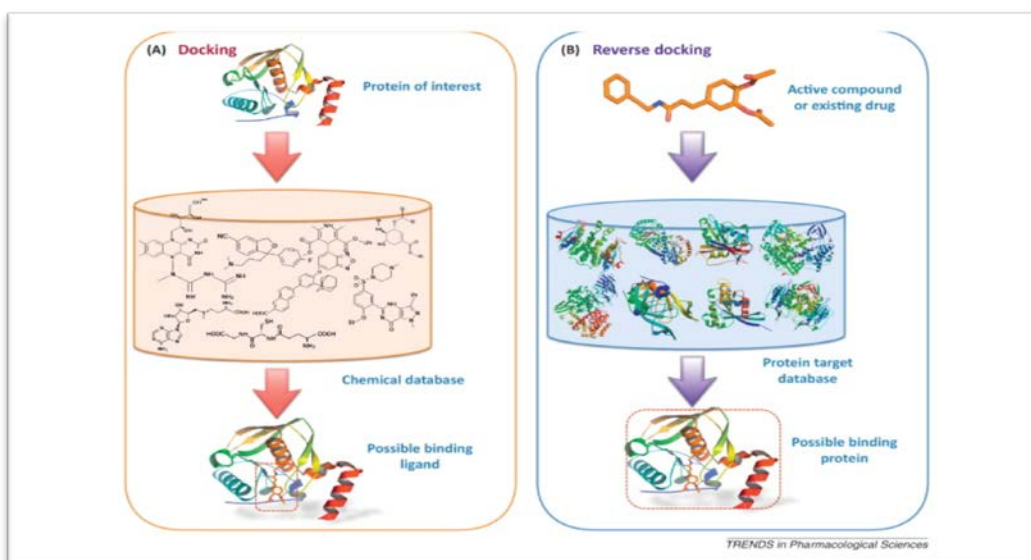


Figure 2.13 -Diagrammes schématiques comparant (A) Le Docking conventionnel et (B) Le Docking inverse.(92)

Alors que le Docking conventionnel est utilisé pour cribler des bibliothèques de composés contre une cible médicamenteuse potentielle, le Docking inverse est utilisé pour dockeur un composé donné dans les sites de liaison prédéfinis d'un groupe de cibles médicamenteuses.(92)

II.2.4.2.8 Applications du Docking / RD dans la découverte des médicaments

Parmi les premières études, et les plus intéressantes, sur l'utilisation du Docking dans la découverte de médicaments et en biologie, on trouve celle de Kuntz et al. Au début des années 1980. Dans cette étude, les auteurs ont décrit une méthode de calcul permettant l'exploration d'alignements ligand-récepteur géométriquement réalisables pour les structures connues de l'hème-myoglobine/metmyoglobine et de la thyroxine/préalbumine.

Actuellement, le Docking moléculaire est employé pour aider à rationaliser l'activité des ligands vers une cible d'intérêt et pour effectuer des campagnes de criblage virtuel basées sur la structure.

Outre ces applications, elle peut être également utilisée pour identifier des séries de cibles pour lesquelles les ligands présentent une bonne complémentarité (target fishing and profiling), certaines d'entre elles étant potentiellement responsables des effets indésirables inattendus des médicaments (off-targets prediction), c'est le Docking inverse.

De plus, le Docking est aussi utilisé pour l'identification de ligands qui se lient simultanément à un ensemble de cibles d'intérêt sélectionnées (polypharmacology) et pour identifier de nouvelles utilisations de composés chimiques dont le profil de sécurité est déjà optimisé (drug repositioning)(Cf. Figure 2.14)(91)

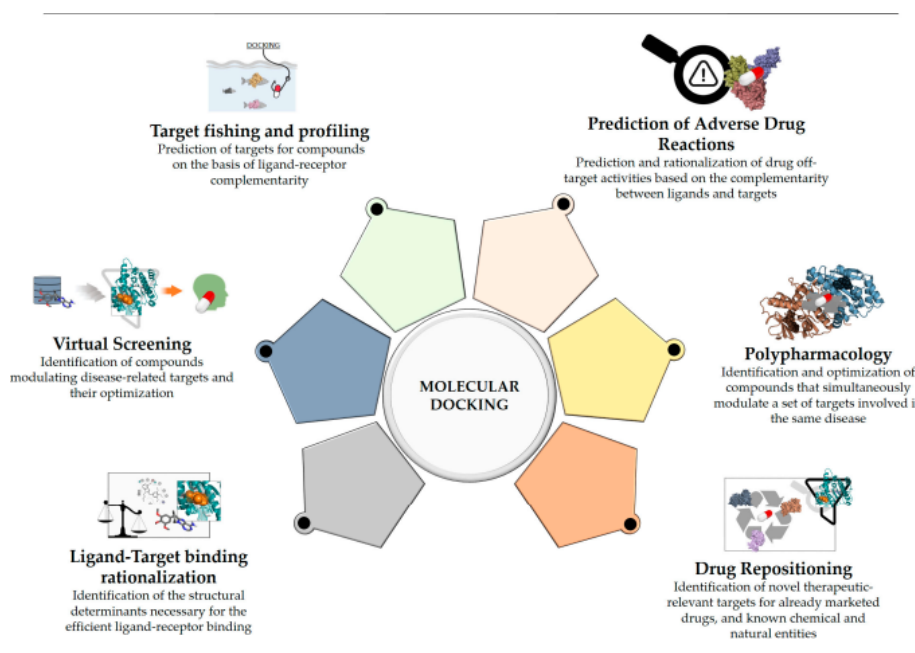


Figure 2.14- Principales applications du Docking/Docking inverse moléculaire dans la découverte actuelle de médicaments.(91)

La combinaison de la modélisation des protéines et des études de Docking (modélisation-Docking) est l'une des méthodologies les plus importantes dans la conception de médicaments basée sur la structure.(80)

En outre, ces études ont aidé à la conception de médicaments pour la cible des 6-phosphofructo-2-kinases (PFKFB3) dans le flux glycolytique et la croissance tumorale ; la découverte de composés principaux pour la cible des CK1 δ dans la maladie d'Alzheimer ; et la conception de médicaments pour la chimiokine CCR5, cible de l'infection par le VIH.

La méthodologie de modélisation et de Docking a également facilité la conception de médicaments dérivés de la génomique. Par exemple, Elle a permis d'accélérer la découverte de composés à succès pour la protéinase 3C-like du coronavirus du syndrome respiratoire aigu sévère (SARS), qui est considérée comme une cible potentielle pour la conception de médicaments destinés au traitement du SRAS. (85)

D'autre part, les méthodes de Docking inverse ont permis l'élaboration de thérapies diverses.

La capacité de MDock, combinée à ITScore, a été testée pour identifier la ou les cibles potentielles de PRIMA-1, un composé qui restaure la fonction de suppression de tumeur de la p53 mutante et entraîne l'apoptose de plusieurs lignées cellulaires cancéreuses.(93)

Zheng et al. ont utilisé AutoDock et Tarfisdock pour tenter d'identifier les protéines cibles potentielles pour les polyphénols du thé vert avec une large activité antinéoplasique. Leur système de Docking inverse a permis de générer un ensemble de protéines exploitables pour une validation expérimentale future.

Ward a étudié la faisabilité de Docking inverse pour les fragments de protéines. Il a utilisé un ensemble d'essai de 950 ligands de tailles diverses contre un ensemble de 13 cibles (31

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

complexes). En utilisant le Docking avec Glide (XP et SP), il a démontré que sa méthode était capable de classer par ordre de priorité les cibles les plus susceptibles de renvoyer des molécules de type Lead et peut être capable d'identifier la ou les poches les plus accessibles sur une cible.(88)

II.2.4.3 Dynamique moléculaire (DM)

Le développement de thérapies modernes ciblées visant à perturber les voies de la maladie adopte souvent de nombreuses approches, y compris la liaison spécifique des récepteurs avec de petites molécules, la liaison des marqueurs de surface cellulaire avec des anticorps monoclonaux, et/ou l'inhibition compétitive de métabolites pathogènes.

Les simulations de dynamique moléculaire (DM) des protéines au niveau atomique sont un outil éprouvé pour caractériser à la fois le comportement des protéines et les interactions protéine-ligand impliquées dans la signalisation cellulaire pour les processus pathologiques.

L'une des propriétés du comportement des protéines pour laquelle la DM est particulièrement bien adaptée à l'investigation, est le comportement des molécules sous les diverses contraintes mécaniques qui se produisent dans un environnement cellulaire. Par exemple, le changement de conformation d'une protéine sous l'effet de l'application d'une force externe peut permettre ou restreindre l'accès à un site de liaison. (94)

II.2.4.3.1 Histoire de la DM

Les simulations DM ne sont pas nouvelles. Les premières, étant de gaz simples, ont été réalisées à la fin des années 1950.

La première simulation DM d'une protéine a été réalisée à la fin des années 1970, et les travaux de base qui ont permis ces simulations font partie des réalisations reconnues par le prix Nobel de chimie 2013.(95)

Les simulations DM sont toutefois devenues beaucoup plus populaires et visibles ces dernières années, notamment du point de vue des biologistes moléculaires expérimentaux (Cf. Figure 2.15). Les simulations ont commencé à apparaître fréquemment dans les articles de biologie structurale, où elles sont utilisées pour interpréter les résultats expérimentaux et pour guider le travail.(96)

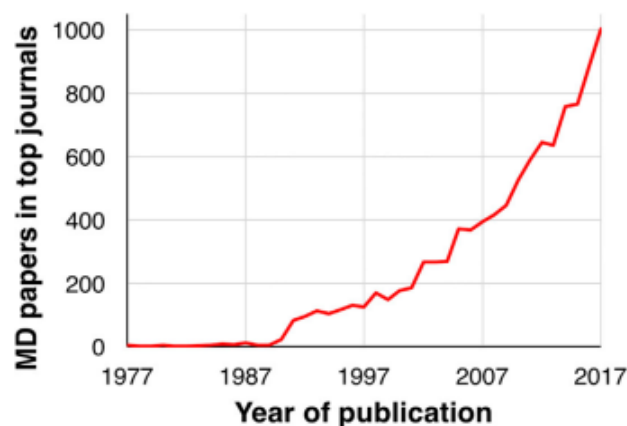


Figure 2.15-Evolution des simulations de dynamique moléculaire en biologie structurale.(96)

II.2.4.3.2 Qu'est-ce que la DM ?

La dynamique moléculaire est la science de la simulation des mouvements d'un système de particules. Elle a été appliquée à des systèmes aussi petits qu'un atome et une molécule diatomique subissant une réaction chimique, et aussi grands qu'une galaxie.

Dans tous les cas, les éléments essentiels d'une simulation de dynamique moléculaire sont la connaissance du potentiel d'interaction des particules, à partir duquel les forces peuvent être calculées, et des équations du mouvement régissant la dynamique des particules. Le potentiel d'interaction peut varier de la simple interaction gravitationnelle entre étoiles aux forces complexes à plusieurs corps entre atomes et molécules. (97)

II.2.4.3.3 Principe général

Les simulations de la DM peuvent fournir non seulement d'abondantes informations structurales dynamiques sur les bio-macromolécules, mais aussi une grande quantité d'informations énergétiques sur les interactions entre protéines et ligands. Ces informations sont très importantes pour comprendre la relation structure-activité de la cible et l'essence des interactions protéine-ligand et pour guider le processus de découverte et de conception de médicaments.(98)

Le DM est utile non seulement pour rationaliser les propriétés mesurées expérimentalement au niveau moléculaire, mais il est bien connu que la plupart des structures déterminées par des méthodes de rayons X ou de RMN ont été affinées par des méthodes de DM.

Par conséquent, l'interaction entre les techniques informatiques et expérimentales dans le domaine des simulations DM est ancienne.(99)

La DM consiste à étudier la trajectoire d'une molécule en lui appliquant les lois de la mécanique classique newtonienne. Une molécule est considérée comme un assemblage d'atomes dont les positions spatiales varient au cours du temps.

Chaque atome est considéré comme une masse ponctuelle dont le mouvement est déterminé par l'ensemble des forces qui s'exercent sur elle. Il en résulte un mouvement atomique correspondant à des oscillations autour d'un minimum d'énergie ou au passage d'un de ces minima à un autre.(78)

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

La DM possède ainsi la faculté d'extraire la molécule d'un minimum local. Cette technique repose sur l'intégration des équations de mouvements de Newton pour chaque atome i :

$$\vec{F}_i(t) = m_i \cdot \vec{a}_i(t)$$

Où :

m_i : Masse de l'atome i

\vec{a}_i : Accélération de l'atome i

\vec{F}_i : Force s'exerçant sur l'atome i

Pour obtenir la trajectoire des atomes en fonction du temps, on va réaliser une intégration numérique sur un court intervalle de temps des équations de mouvements. Ces dernières sont déterminées. Il faut, pour réaliser une simulation de DM, partir d'une structure raisonnable du système (expérimentale ou modèle) que l'on veut étudier et on déterminera les coordonnées et vitesses initiales en procédant ainsi :

- Une minimisation d'énergie du système fournira le jeu de coordonnées de départ de la dynamique. Cette étape de minimisation nécessaire permet d'éviter qu'une énergie mécanique trop forte ne détériore l'assemblage moléculaire dès le début de la production de dynamique moléculaire.
- Les vitesses de départ des atomes sont attribuées aléatoirement à partir d'une distribution gaussienne de maxwell-Boltzman correspondant à la température T .(82)

L'augmentation de la puissance informatique a permis d'allonger le temps de simulation qui est passé, en une vingtaine d'années, de la pico seconde à la nano seconde et parfois même à la micro seconde. De meilleurs champs de forces ont, par ailleurs, vu le jour, impliquant un meilleur traitement des interactions électrostatiques à longue distance.

Les programmes couramment utilisés dans la simulation des biomolécules par dynamique moléculaire sont AMBER, CHARMM, GROMOS et NAMD.(100)

En utilisant des ordinateurs de plus en plus améliorés avec des programmes de grande performance, on peut développer des techniques de DM très perfectionnées avec plusieurs applications dans divers domaines (Cf. Figure 2.16).

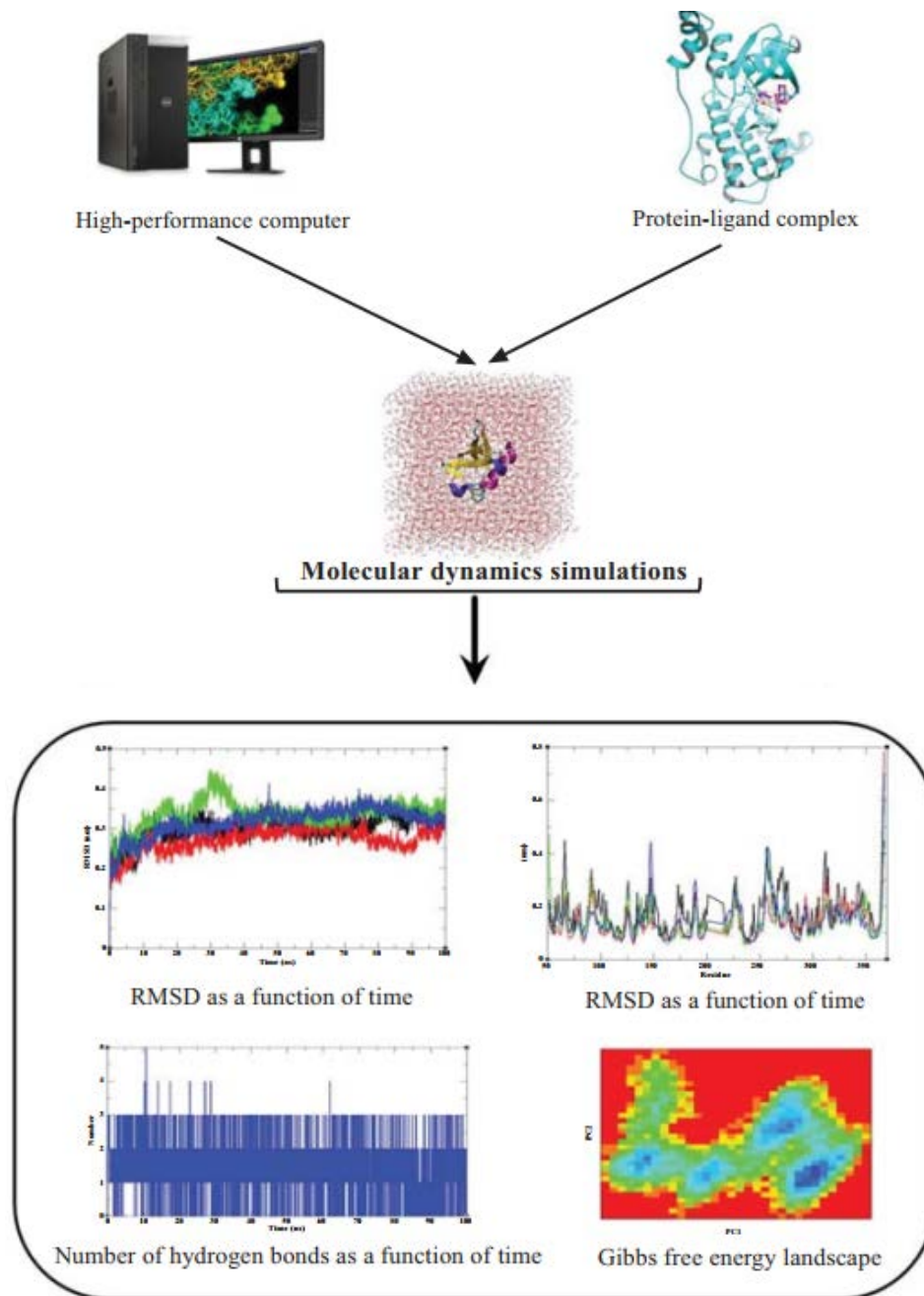


Figure 2.16-Dynamique moléculaire et ses applications.(95)

II.2.4.3.4 Calcul de la DM

En connaissant l'énergie, on peut calculer la vitesse et donc la position de chaque atome. Le calcul de la trajectoire va se faire en intégrant numériquement les n (nombre entier) équations à résoudre pour les n atomes du système à étudier. L'algorithme d'intégration, choisi parmi les nombreux qui existent (velocityverlet, leapfrog, etc...), recalcule à chaque incrément de temps δt les nouvelles positions et vitesses à partir des précédentes.

On obtient donc une succession de clichés instantanés du système. Pour obtenir des calculs stables, la valeur du pas d'intégration doit être extrêmement faible ($1 \text{ fs} = 10^{-15} \text{ s}$) de manière à prendre en compte les vibrations atomiques les plus rapides (ex : liaison C-H).(2)

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

C'est ce qui explique pourquoi les simulations en biologie dépassent rarement les 10 à 100 Ns sans l'aide de supercalculateurs. La DM est tributaire de la qualité des structures expérimentales, structures qui représentent un ensemble de conformères responsables de l'activité biologique de la protéine étudiée.(2,99)

II.2.4.3.5 Simulations combinées de Docking et de la DM

La flexibilité du site de liaison de la cible est un aspect essentiel, mais souvent négligé, à prendre en considération dans le Docking moléculaire. Les enzymes et les récepteurs peuvent subir des changements de conformation pendant le processus de reconnaissance moléculaire.

Dans certains cas, ces réarrangements structuraux sont faibles et le ligand s'insère dans un site de liaison avec une faible mobilité. Dans le cas contraire, certaines protéines subissent des changements conformationnels importants qui peuvent impliquer des éléments de la structure secondaire et tertiaire. (100)

Les protocoles de Docking rapides et peu coûteux peuvent être combinés avec des techniques DM précises mais plus coûteuses, afin de prédire des complexes protéine-ligand plus fiables. La force de cette combinaison réside dans leurs forces et faiblesses complémentaires.

D'une part, les techniques de Docking sont utilisées pour explorer le vaste espace conformationnel des ligands en peu de temps, ce qui permet d'examiner de grandes bibliothèques de composés de type médicamenteux à un coût raisonnable. Les principaux inconvénients sont l'absence, ou la faible flexibilité de la protéine, qui n'est pas autorisée à ajuster sa conformation lors de la liaison du ligand, et l'absence d'une fonction de score unique et largement applicable, nécessaire pour générer un classement fiable des complexes finaux.

D'un autre côté, les simulations DM peuvent traiter à la fois le ligand et la protéine d'une manière flexible, permettant un ajustement induit du site de liaison du récepteur autour du ligand nouvellement introduit. En outre, l'effet des molécules d'eau explicites peut être étudié directement et des énergies libres de liaison très précises peuvent être obtenues.(99)

Par conséquent, le couplage des deux puissantes techniques que sont le Docking et la dynamique moléculaire doit théoriquement augmenter le pouvoir prédictif de notre modèle.(100)

Aujourd'hui, la modélisation moléculaire et la relation structure-activité constituent les besoins fondamentaux des industries pharmaceutiques pour la découverte et le développement de médicaments et la simulation de DM et le Docking sont deux méthodes fondamentales pour faciliter la découverte de médicaments assistée par ordinateur, à la fois rapide, économique et efficace.(95)

II.2.4.3.6 Limites de la DM

Indépendamment de son utilité, la DM a ses limites. Parmi celles-ci, nous pouvons souligner le coût de calcul élevé exigé par la simulation de grands systèmes, qui sont généralement

constitués de milliers d'atomes lorsque des complexes ligand-récepteur sont étudiés. Certains des changements de conformation entrepris par les récepteurs pendant la reconnaissance moléculaire se produisent sur des échelles de temps dépassant la capacité de calcul disponible.

Malgré ses limites, la DM est en mesure d'apporter des contributions importantes à la SBDD, en particulier lorsqu'il est combiné à d'autres méthodes de modélisation moléculaire comme le Docking moléculaire.(77)

II.2.4.3.7 Comment la DM peut-elle contribuer à la découverte des médicaments et la SBDD ?

Les approches de la DM sont une composante importante des efforts en matière de SBDD. L'essence de la SBDD repose principalement sur la compréhension des interactions médicament-protéine. Au début de l'ère de la SBDD, le ligand et le récepteur étaient considérés comme des corps rigides avec seulement un minimum de degrés de liberté conformationnelle.

Néanmoins, une telle approximation peut ne pas refléter le scénario réel, où le médicament et la protéine se trouvent dans un environnement dynamique.

La structure tridimensionnelle, qui est obtenue soit par des méthodes expérimentales (RMN et cristallographie aux rayons X), soit par des approches de modélisation *in silico*, est généralement une représentation statique unique de la structure de la protéine. De telles informations sont utiles pour obtenir des connaissances structurelles sur la dimension donnée de la structure de la protéine. Cependant, l'information ne peut être complète sans comprendre les mouvements moléculaires des protéines. (35)

Par exemple, l'histone désacétylase8 (HDAC8) est l'une des cibles importantes du cancer, dont la structure présente des différences significatives dans leurs conformations. Jusqu'à récemment, au moins 21 structures cristallines aux rayons X de l'enzyme HDAC8 disponibles dans la base de données PDB affinées par la DM.(101)

Dans une autre étude, Wang et al. (2011) ont démontré l'efficacité de la méthode de la DM pour guider la conception des inhibiteurs pour le type sauvage et les états mutés du virus de l'influenza A M2 (IA/M2).

L'IA/M2 participe à la formation d'un canal sélectif de protons homo-tétramérique dans la membrane virale et reste une importante cible antivirale importante. Les états mutants de l'IA/M2, tels que V27A et L26F, que l'on trouve généralement chez l'homme, les oiseaux et les porcs, sont connus pour provoquer une résistance élevée aux médicaments. Cette résistance aux médicaments des formes mutantes entrave souvent la thérapie antivirale. Wang et al. ont utilisé des simulations DM classiques à l'échelle de la nanoseconde pour explorer les mécanismes de liaison de l'amantadine, un antiviral bien connu, contre l'IA/M2 afin de concevoir des inhibiteurs puissants. (102)

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

Par conséquent, l'étude décrit la puissance des simulations DM pour comprendre les modes de liaison des médicaments connus, ce qui guidera la conception de nouveaux inhibiteurs, même pour les cibles impossibles à traiter.

D'autre part, la DM est également capable de révéler des informations importantes sur les protéines, qui parfois ne sont pas accessibles à partir d'expériences. Par exemple, dans un article marquant de McCammon et ses collègues, les auteurs ont découvert une nouvelle région de liaison dans l'enzyme intégrase du VIH, qui n'était pas observable dans la structure cristalline aux rayons X.(101)

Les progrès récents de la biologie structurale ont permis d'obtenir les structures de nombreuses cibles clés de la découverte de médicaments en neurosciences (par exemple, les RCPG, les canaux ioniques, transporteurs, etc.). Pour exploiter pleinement la puissance de la conception de ces cibles et d'autres, il faut prendre en compte les propriétés dynamiques de ces protéines.(96)

Tous ces exemples montrent l'importance de la simulation DM dans le succès des approches SBDD.

II.3 « LIGAND -BASED DRUG DESIGN » (LBDD)

La conception de médicament basée sur la structure de ligand

La principale alternative à la SBDD est la LBDD.

Lorsque les ligands et les données sur ses activités biologiques sont les seules informations disponibles pour le développement de médicaments, et dans le cas où la cible potentielle d'un médicament est inconnue et que la prédiction de cette structure à l'aide de méthodes telles que la modélisation homologique ou la prédiction de structure ab initio est difficile, la conception de médicaments basée sur les ligands (LBDD) assistée par ordinateur est une méthode efficace pour étendre les connaissances sur les ligands connus pour concevoir des composés présentant une meilleure activité biologique.(103,104)

Bien que le nombre de structures protéiques 3D résolues augmente chaque année, il y a encore beaucoup de cibles qui ne sont pas encore disponibles pour la modélisation basée sur la structure. Par conséquent, les approches de modélisation basées sur les ligands restent donc intéressantes pour les chercheurs.(105)

Les méthodes de LBDD peuvent être utilisées pour la découverte de composés principaux et l'optimisation de ligands déjà connus, comme dans la SBDD, il est préférable de commencer par l'exploration des bases de données moléculaires.(104)

La conception de médicaments basée sur des ligands, ou conception indirecte de médicaments, s'appuie sur la connaissance d'autres molécules actives connues ayant un potentiel contre des cibles biologiques d'intérêt.(35)

Le QSAR, la modélisation des pharmacophores et les approches de similarité moléculaire (relation quantitative structure-activité) sont des approches populaires de la LBDD. Ils peuvent fournir des modèles prédictifs adaptés à l'identification et à l'optimisation des Leads.

Le **QSAR** est une méthode de calcul qui modélise la relation entre les caractéristiques structurelles des ligands qui se fixent à une cible et l'effet de l'activité biologique correspondante déterminée expérimentalement. Ces corrélations QSAR prédites peuvent à leur tour être utilisées pour prédire l'activité de nouveaux analogues.

Par ailleurs, dans la **modélisation basée sur des ligands pharmacophore**, Des modèles de pharmacophores sont dérivés de ces molécules connues pour définir les caractéristiques structurelles nécessaires pour permettre la liaison avec la cible biologique.

Dans les méthodes de **similarité moléculaire**, l'empreinte moléculaire des ligands connus qui se lie à une cible est utilisée pour trouver des molécules avec des empreintes similaires par le criblage de bibliothèques moléculaires. (86)

L'importance de la LBDD est soulignée par le fait que plus de 50 % des médicaments, approuvés par la FDA, ciblent des protéines membranaires, telles que les récepteurs couplés

aux protéines G (RCPG), les récepteurs nucléaires et les transporteurs, pour lesquels les structures tridimensionnelles (3D) ne sont souvent pas disponibles.

Compte tenu des difficultés à déterminer les structures 3D des protéines associées à la membrane, les méthodologies LBDD devraient continuer à avoir un impact significatif sur le développement de médicaments dans un avenir prévisible.(103)

II.3.1 Principales méthodes utilisées

II.3.1.1 Approche quantitative de la relation structure-activité (QSAR)

Les relations structure-activité (RAS) sont utilisées universellement en chimie médicinale et en biologie pharmaceutique car l'analyse de ces relations facilite l'identification des groupes fonctionnels ou chimiques responsables de la suggestion d'un effet biologique ciblé. Ces connaissances peuvent ensuite être utilisées pour diriger les modifications chimiques d'un candidat principal afin d'améliorer la puissance biochimique, la sélectivité, les propriétés pharmacocinétiques ou pharmacodynamiques et la toxicité. (106)

II.3.1.1.1 Historique de la QSAR/ QSPR

Il y a plus d'un siècle et demi, en 1863, Croso observé que le point d'ébullition et le point de fusion des alcanes augmente avec le nombre d'atomes de carbone et la masse moléculaire. Il a observé également une diminution de la solubilité dans l'eau des alcools avec l'augmentation du nombre d'atomes de carbone et la masse moléculaire, cela est considéré comme la première formulation générale en QSPR.

Cinq ans après, en 1868, Crum-Brown et Fraser postulèrent que « l'activité biologique d'une molécule est une fonction de sa constitution chimique ».

Quelques décennies plus tard, en 1893, Richeta montré que la cytotoxicité de certains composés organiques était inversement proportionnelle à leur solubilité dans l'eau.

A la fin du 19^{ème} siècle, Meyer en 1899 et Overton en 1901, ont indépendamment observé « une relation linéaire entre l'activité des narcotiques et leur coefficient de partage huile-eau ».(107)

En 1962, Hansen a montré l'existence d'une corrélation entre la toxicité des acides benzoïques substitués et les constantes électronique de ses substituants.

L'année 1964 est considérée comme le début des méthodes QSAR classiques (2D-QSAR). Hansch et Fujita ont établi les premières corrélations entre les propriétés physico-chimiques (log P, pKa, paramètres stériques et électroniques) et l'activité biologique (activité

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

enzymatique, pharmacologique), Ces méthodes seront appelées par la suite l'analyse de Hansch et l'analyse de Free Wilson.

Maintenant, des méthodes 3D-QSAR comme l'étude CoMFA (Comparative Molecular Field Analysis) et CoMSIA (Comparative Molecular Similarity Indices Analysis) permettent de traiter les relations structure-activité en trois dimensions.(25)

II.3.1.1.2 Définition

QSAR est le nom donné aux méthodes qui mettent en corrélation la structure moléculaire avec des propriétés telles que l'activité biologique in vitro ou in vivo.

Parfois, les méthodes QSAR prennent des noms plus spécifiques comme : QSPR (relation quantitative structure-propriété) ou QSTR (relation quantitative structure-toxicité). Les QSPR sont utilisées pour les propriétés physicochimiques, comme le point d'ébullition, la solubilité, logP.(1)

Par définition, Une QSAR est un modèle mathématique qui associe un ou plusieurs paramètres dérivés de la structure chimique, à une mesure quantitative d'une propriété ou d'une activité.(25)

II.3.1.1.3 Hypothèse de base de QSAR

QSAR est une méthode qui relie la structure chimique à l'activité biologique ou chimique à l'aide de modèles mathématiques. Si l'activité d'un ensemble de ligands peut être déterminée, un modèle peut être construit pour décrire cette relation.(108)

Les méthodes QSAR sont basées sur des statistiques qui mettent en corrélation les activités des cibles avec divers descripteurs moléculaires. La base de la méthode QSAR est le fait que les molécules structurellement similaires ont tendance à présenter une activité biologique similaire. Ces modèles décrivent mathématiquement comment la réponse de l'activité d'une cible, qui se lie à un ligand, varie avec les caractéristiques structurelles du ligand.

Les relations QSAR peuvent être utilisées pour prédire l'activité de nouvelles molécules de médicaments analogues.(86)

Les progrès récents du QSAR et leurs applications dans le processus de découverte de médicaments ont été joliment décrites par Eleni et al.(2003) et Dudek et al. (2006).(109)

II.3.1.1.4 Principe général

Le QSAR se rapporte à la construction de modèles prédictifs d'activités biologiques, en fonction de l'information de structure moléculaire d'une série de composés. C'est un procédé qui relie de manière quantitative les propriétés moléculaires aussi bien électroniques que géométriques, appelées descripteurs, avec une activité biologique, suivant un modèle mathématique :

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

$$[\text{Activité biologique}] = f [\text{Descripteurs}]$$

Dans l'équation, l'activité biologique est normalement exprimée comme $\log [1/C]$, où C est généralement la concentration minimum requise pour causer une réponse biologique définie.(110)

Au départ, un groupe d'entités chimiques ou de molécules sont identifiées qui présentent l'activité biologique souhaitée. Une relation quantitative est établie entre les caractéristiques physico-chimiques des molécules actives et l'activité biologique. (103)

Le modèle QSAR développé est ensuite utilisé pour optimiser les composés actifs afin de maximiser l'activité biologique pertinente. Les composés prédits sont ensuite testés expérimentalement pour l'activité souhaitée.

La méthode QSAR peut donc être utilisée comme un outil de guide pour l'identification des modifications de composés présentant une meilleure activité.(108)

Avant de construire un modèle par la QSAR, il faut, à l'avance, avoir des connaissances multidisciplinaires de base qui interviennent dans ce processus. Ils se divisent en 3 piliers principaux, à savoir les données biologiques, les descripteurs moléculaires et la modélisation.

II.3.1.1.4.1 Données biologiques

Les données biologiques utilisées dans le QSAR proviennent d'essais en laboratoire sur des systèmes vivants (animaux ou végétaux). Ils sont mesurables et quantifiables (concentration d'enzymes ou d'hormones spécifiques, distribution phénotypique de gènes spécifiques dans une population, etc.) et servent d'indices pour les évaluations physiologiques liées à la santé, au risque de maladie, aux troubles psychiatriques, à l'exposition environnementale et à ses effets, aux processus métaboliques, à la toxicomanie, etc.(111)

Les données biologiques sont généralement exprimées sur une échelle logarithmique. De plus, le logarithme inverse de l'activité ($\log 1/C$) est également utilisé pour obtenir des valeurs mathématiques plus élevées lorsque les structures sont biologiquement très efficaces.(44)

II.3.1.1.4.2 Descripteurs moléculaires

Les propriétés chimiques, telles que la réactivité des substances, nous disent que plus nous avons de substances plus il y a de nouveaux produits chimiques qui pourraient être développés.(111)

Les descripteurs moléculaires représentent la forme informatisée des propriétés chimiques des molécules. Ils transforment la structure du composé en un ensemble de valeurs numériques ou binaires représentant diverses propriétés moléculaires qui sont jugées importantes pour expliquer l'activité.(112)

Chaque descripteur moléculaire ne prend en compte qu'une petite partie de l'ensemble des informations chimiques contenues dans la molécule réelle et, par conséquent, le nombre de

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

descripteurs ne cesse d'augmenter avec la demande croissante d'études approfondies sur les systèmes chimiques et biologiques.

Les différents descripteurs ont des méthodes ou des perspectives indépendantes pour visualiser une molécule, en tenant compte des diverses caractéristiques de la structure chimique. (44)

Les descripteurs moléculaires font désormais partie des variables les plus importantes utilisées dans la modélisation moléculaire et, par conséquent, gérées par les statistiques et la chimie-informatique.(111)

II.3.1.1.4.2.1 Classification des représentations moléculaires

Les informations saisies par les descripteurs moléculaires peuvent varier de simples propriétés de masse à des définitions tridimensionnelles complexes ou à la fréquence de sous-structure. En particulier, différents niveaux de complexité (également appelés "dimensionnalité") peuvent être utilisés pour représenter une molécule donnée (Cf. Figure 2.17), comme suit :

0-Dimensionnelle (0D) : La représentation moléculaire la plus simple est la formule chimique, c'est-à-dire la spécification des éléments chimiques et leur présence dans une molécule. Par exemple, la formule chimique de 2,3,7,8-Tétrachlorodibenzodioxine (un contaminant connu pour sa toxicité pour l'homme et les écosystèmes) est la suivante : $C_{12}H_4Cl_4O_2$.(113)

1-Dimensionnelle (1D) : Selon cette représentation, les molécules sont perçues comme un ensemble de sous-structures, telles que des groupes fonctionnels ou des fragments centrés sur les atomes. Cette représentation ne nécessite pas la connaissance complète des structures moléculaires.(114)

2-Dimensionnel (2D) : C'est la représentation topologique qui est la plus populaire. Il s'agit d'une représentation bidimensionnelle qui, en plus de la composition atomique, inclut les informations sur la connectivité des atomes dans la molécule sous la forme d'un graphe moléculaire.(114)

À partir d'une représentation graphique, plusieurs quantificateurs numériques de la topologie moléculaire sont dérivés mathématiquement de manière directe et sans ambiguïté. Ils sont communément appelés comme des indices topologiques (IT)(113)

3-Dimensionnel (3D) : Un niveau supplémentaire de complexité en percevant la molécule non seulement en termes de type d'atomes, de connectivité et d'adjacence, mais aussi en la considérant comme un objet géométrique dans l'espace, caractérisé par la configuration spatiale des atomes.(113)

En d'autres termes, chaque atome de la molécule est associé à un ensemble de coordonnées tridimensionnelles x-y-z déterminant sa propre position dans l'espace. Cette représentation est désignée sous le nom de représentation géométrique d'une molécule où elle est considérée comme une structure géométrique rigide.(114)

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

4-Dimensionnel (4D) : En plus de la géométrie moléculaire, une "quatrième dimension" peut également être introduite, visant généralement à identifier et de caractériser quantitativement les interactions entre la ou les molécules et le ou les sites actifs d'un récepteur biologique.(113)

Elle prend également en compte les propriétés moléculaires découlant de la distribution des électrons et les interactions des molécules avec les sondes. (114)

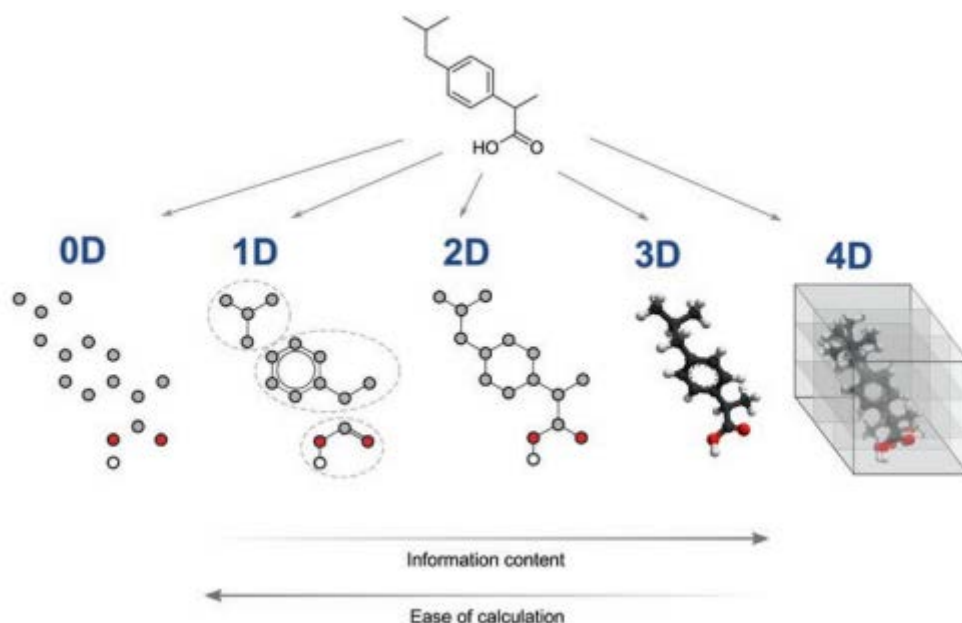


Figure 2.17- Exemple graphique de différentes représentations moléculaires d'une même structure (ibuprofène, ici représentée comme une structure 2D)(113)

Des dimensions plus élevées que la quatrième (c'est-à-dire 5D-QSAR et 6D-QSAR) peuvent être ajoutées au problème de modélisation, par exemple, en considérant différents modèles d'ajustement induit et/ou différents scénarios de solvation.(113)

II.3.1.1.4.2.2 Calcul des descripteurs moléculaires

Le but est de calculer les propriétés des molécules qui servent de descriptions numériques ou de caractérisations des molécules dans d'autres calculs tels que le QSAR, l'analyse de la diversité ou la conception de bibliothèques combinatoires.

En principe, comme n'importe quelle propriété moléculaire peut être utilisée comme descripteur moléculaire, il n'existe pas de stratégie unique pour calculer et/ou sélectionner les descripteurs moléculaires.(115)

Une fois que toutes les structures moléculaires ont été vérifiées et conservées, les descripteurs moléculaires sont calculés à partir de la représentation moléculaire choisie, en utilisant un ou plusieurs outils logiciels disponibles.

Les descripteurs à utiliser sont généralement choisis sur la base de connaissances a priori sur la propriété à modéliser ou sur leurs performances précédentes pour le problème analysé (si

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

disponible). Ils deviennent alors les nouvelles variables indépendantes à utilisées pour développer le(s) modèle(s) d'intérêt.(113)

II.3.1.1.4.2.3 Classification des descripteurs moléculaires

Il existe de nombreux descripteurs différents et deux types de classification, celle basée sur sa dimension et celle basée sur l'effet (stérique, hydrophobe, électronique, etc.).

1. 4.2.3.a. Classification basée sur la dimension

On trouve quatre types de descripteurs moléculaires : 1D, 2D, i3D et x3D .(115)

Par ailleurs, on peut trouver les descripteurs 0D , 4D et 5D.(113)

Descripteurs 0D

Ils sont très simples à calculer et à interpréter, mais ils présentent un faible contenu informatif et un degré de dégénérescence élevé, c'est-à-dire qu'ils peuvent avoir des valeurs égales pour différentes molécules. Quelques exemples sont le nombre d'atomes (par exemple, le nombre d'atomes de carbone), le poids moléculaire et la somme ou la moyenne des propriétés atomiques (par exemple, les volumes atomiques de van der Waals).(113)

Descripteurs 1D

Ces descripteurs décrivent des propriétés globales du composé, et ne permettent pas de distinguer les isomères de constitution et ne permettent pas d'élaborer des modèles plus complexes. Il est nécessaire d'ajouter d'autres types de descripteurs.(110)

Ils sont généralement binaires, codant pour la présence/absence de sous-structures donnée, par exemple la présence et/ou l'absence des cycle aromatiques ou bien une fonction amide attachée à la chaîne principale.(114)

Descripteurs 2D

Les descripteurs 2D, appelés aussi IT, sont des propriétés numériques qui peuvent être calculées à partir de la table de connectivité d'une molécule ou d'une représentation planaire (2D) de la structure. (110)

Les IT codent les propriétés topologiques (par exemple, la contiguïté, la connectivité) et sont généralement sensibles aux caractéristiques structurales telles que la taille, la forme, la symétrie, la ramification et la cyclicité. Souvent, des propriétés chimiques spécifiques des atomes sont également considérées, par exemple la masse et la polarisabilité, ou la présence de donneurs/accepteurs de liaisons hydrogène.(113)

La mesure la plus couramment utilisée est le coefficient de partage octanol-eau P, normalement utilisé dans sa forme logarithmique (log P). (110)

Descripteurs 3D

Les descripteurs dérivant de la représentation 3D ont un contenu d'information élevé et peuvent être particulièrement utiles pour modéliser les propriétés pharmaceutiques et biologiques. Lorsqu'ils traitent avec la représentation 3D, les utilisateurs doivent garder à l'esprit plusieurs problèmes liés à l'optimisation géométrique des molécules, tels que :

- (1) l'influence de la méthode d'optimisation sur les valeurs des coordonnées.
- (2) la présence de plus d'un conformère similaire à énergie minimale pour les molécules très flexibles .
- (3) la différence entre la géométrie bioactive et la géométrie de la molécule, le degré de déformation dépendant de ce qui suit le nombre de liaisons librement rotatives dans la molécule.

Pour ces raisons, le rapport coût/bénéfice de l'utilisation des descripteurs 3D dépend du cas et doit être soigneusement évalué.(113)

Ces descripteurs s'avèrent donc relativement coûteux en temps de calcul, mais apportent davantage d'informations, et sont nécessaires à la modélisation de propriétés ou d'activités qui dépendent de la structure 3D. (110)

On peut distinguer deux types :

i3D : les descripteurs internes-3D utilisent les coordonnées 3D des informations sur chaque molécule, restant cependant invariants aux rotations et translations de la conformation ;

x3D : les descripteurs 3D externes utilisent également des informations de coordonnées 3D mais nécessitent également un cadre de référence absolu (par exemple, des molécules ancrées dans le même récepteur).

1. 4.2.3.b. Classification basée sur l'effet

On trouve dans la littérature un autre type de classification qui ne prend en compte que des caractéristiques illustrés par le descripteur (Cf. Figure 2.18). Ces descripteurs vont de simples mesures d'entités constituant la molécule, en passant par ses propriétés topologiques et géométriques jusqu'aux descripteurs électrostatiques et quantiques calculés ou des méthodes avancées de comptage des fragments.(112)

1. Descripteurs constitutionnels :

En fonction du nombre et du type d'atomes, de liaisons et de groupes fonctionnels, ils saisissent les propriétés de la molécule qui sont liées aux éléments qui constituent sa structure. Ces descripteurs sont rapides et faciles à calculer. Voici quelques exemples, le nombre total d'atomes dans la molécule et le nombre d'atomes d'identité différente. De même, un certain nombre de propriétés relatives aux liaisons sont utilisées, notamment le nombre total de liaisons de type simple, double, triple ou aromatique, ainsi que le nombre de cycles aromatiques.(1,115)

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

2. Descripteurs géométriques :

Ils donnent la surface et le volume des molécules obtenues à partir des surfaces de Van der Waals atomiques et de leur chevauchement, les moments d'inertie, les projections de zones d'ombre, et les indices gravitationnels.

3. Descripteurs topologiques :

Ils sont basés sur la topologie des molécules. Seules les informations structurelles sont utilisées pour générer la description. Ils traitent la structure du composé comme un graphe, avec les atomes comme sommets et les liaisons comme des arêtes.(1,112)

4. Descripteurs physico-chimiques :

Ils tentent d'estimer les propriétés physiques des molécules. La masse moléculaire, les accepteurs ou les donneurs de liaisons hydrogène, et les coefficients de partage, sous la forme de logP qui indique la lipophilie générale (ou l'hydrophobie) de la substance.

5. Descripteurs électrostatiques :

capturer des informations sur la nature électronique de la molécule c'est à dire des informations sur les charges atomiques nettes et partielles.(112)

6. Descripteurs de chimie quantique :

Ils sont liés à l'orbite moléculaire et à leurs propriétés.

7. Empreinte digitale :

C'est une chaîne binaire codant une molécule, où le 1 ou le 0 d'une position signifie que la sous-structure de cette position dans le dictionnaire est présente ou non.(1)



Figure 2.18- Représentation des descripteurs moléculaires utilisés dans la modélisation QSAR.(116)

La structure moléculaire est codée à l'aide de descripteurs numériques. L'ensemble des descripteurs est élagué pour sélectionner les plus informatifs. L'activité est dérivée en fonction des descripteurs sélectionnés (Cf. Figure 2.19)(112)

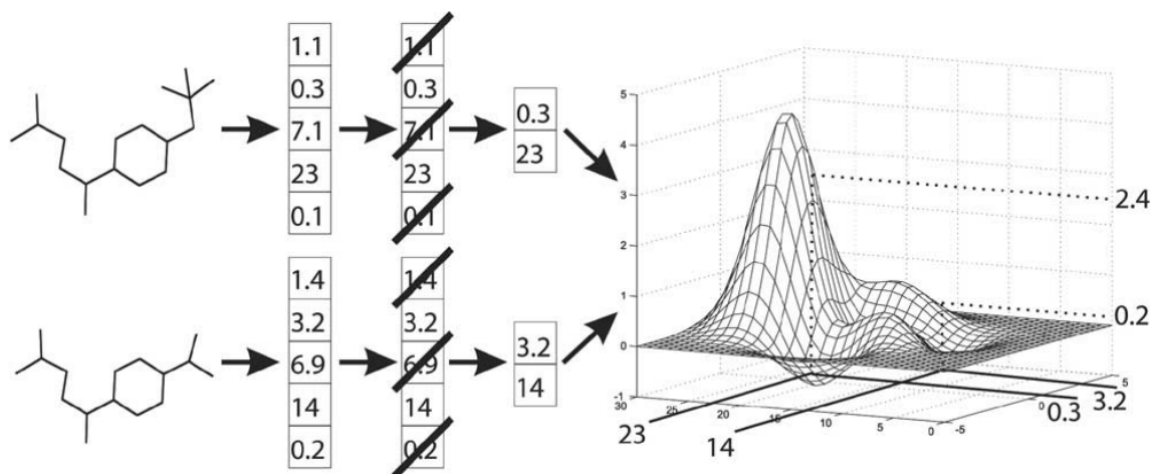


Figure 2.19- Principales étapes d'une étude QSAR en décrivant l'importance cruciale des descripteurs moléculaires (formation des landscapes).(112)

Le succès de tout modèle QSAR dépend grandement du choix des descripteurs moléculaires et de la capacité à générer la relation mathématique appropriée entre les descripteurs et l'activité biologique en question. Depuis les premiers jours de QSAR, il était clair que la définition des descripteurs moléculaires est la partie cruciale de la méthode. Les récents développements des logiciels permettent maintenant de générer un grand nombre de descripteurs moléculaires qui peuvent être utilisés pour les méthodes QSAR.(117)

II.3.1.1.4.3 Modélisation

La modélisation est le dernier pilier, mais non le moindre, du QSAR. En gros, il y a deux grands courants de modélisation : la modélisation des données et la modélisation algorithmique.

La modélisation des données est la voie communément développée par les statisticiens : à partir de l'analyse des données, ils postulent le type de relation entre les données et la réponse, et utilisent un large ensemble d'outils mathématiques pour dériver le modèle.

La modélisation algorithmique a été développée plus récemment, à partir du milieu des années 1980, lorsque de nouveaux algorithmes puissants d'ajustement des données sont devenus disponibles.(111)

II.3.1.1.5 Méthodologie

La méthodologie générale de l'approche QSAR repose sur une série d'étapes consécutives (Cf. Figure 2.20) :

1. Les molécules actives qui se fixent à la cible médicamenteuse souhaitée et leurs activités sont identifiées par une recherche dans une base de données, une recherche documentaire ou des expériences HTS.
2. Identification des caractéristiques moléculaires structurales ou physicochimiques (descripteurs) affectant l'activité biologique (par ex. liaison, atome, nombre de groupes fonctionnels, surface, etc.)
3. Construction d'un modèle QSAR entre l'activité biologique et les caractéristiques identifiées de médicaments.
4. Validation du pouvoir prédictif de l'activité biologique du QSAR.(86)

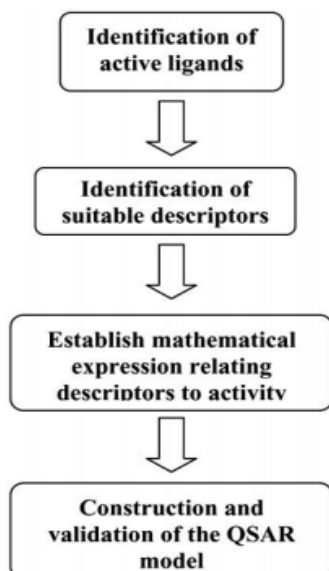


Figure 2.20 - Flux de travail typique des méthodes QSAR.(117)

L'ensemble de processus peut être résumé comme suit (Cf. Figure 2.21) :

A partir d'un ensemble de molécules avec des propriétés expérimentales annotées (par exemple, physicochimiques, toxicologiques et biologiques), plusieurs types de descripteurs moléculaires peuvent être calculés.

L'ensemble des données obtenues (descripteurs moléculaires + propriétés expérimentales) est ensuite utilisé dans la phase d'extraction d'informations, pour obtenir un modèle QSAR fiable et validé.

Ce modèle peut ensuite être appliqué pour prédire les propriétés de molécules non testées, pour obtenir des informations mécanistiques par l'interprétation des descripteurs moléculaires ou pour concevoir de nouvelles molécules.(113)

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

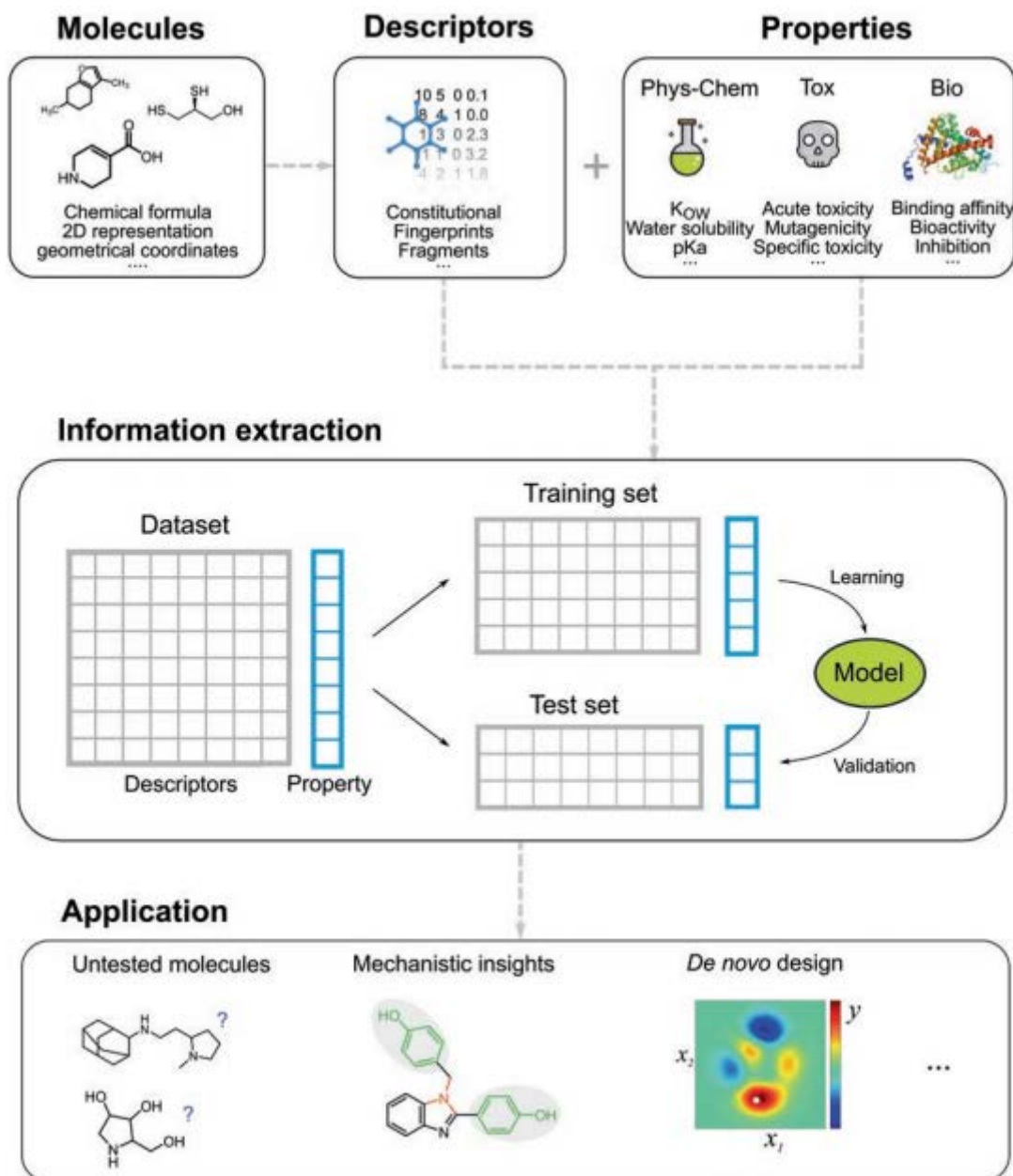


Figure 2.21- Principales étapes de la mise au point et de l'utilisation de QSAR. (113)

Aucun modèle QSAR ne peut remplacer les essais expérimentaux, bien que les techniques expérimentales ne soient pas non plus exemptes d'erreurs.

En raison de nombreux problèmes évidents de simulation des situations réelles, tous les paramètres *in vivo* ne peuvent pas être inclus dans la modélisation QSAR. Cependant, tout est mis en œuvre pour développer un modèle aussi proche que possible de la réalité. (118)

Au fil des ans, les stratégies d'exécution des étapes citées précédemment ont évolué pour faire de la technique QSAR une partie essentielle du processus d'optimisation des médicaments. Les progrès de la méthodologie QSAR ont principalement porté sur l'éventail des descripteurs moléculaires et la façon dont ils sont liés à l'activité..(117)

II.3.1.1.6 Classification

En fonction de la nature des descripteurs utilisés, les méthodes QSAR peuvent être divisées majoritairement en deux classes, à savoir 2D QSAR et 3D QSAR.

II.3.1.1.6.1 2D QSAR (Classique)

Dans la méthode QSAR classique, également appelée approche Hansch-Fujita, diverses caractéristiques électroniques, hydrophobes et stériques sont corrélées avec l'activité biologique d'une série de composés congéniques. Elle implique généralement des descripteurs moléculaires 2D. Elle est principalement basée sur des fragments moléculaires spécialisés qui constituent le composé chimique.(119)

En 1964, Free-Wilson a développé indépendamment un modèle mathématique reliant la présence de divers substituant chimiques à l'activité biologique. Chaque type de groupe chimique a reçu une contribution d'activité, selon le type et l'emplacement d'un substituant (tel que méta ou para) et l'impact connexe sur l'activité biologique.

Les approches Hansch et Free-Wilson ont servi d'outils prédictifs dans les études QSAR classiques pendant de nombreuses années. Plus tard, une méthode combinée Hansch / FreeWilson a été développée.(120)

La force du QSAR classique est qu'en utilisant des relations mathématiques très simplistes impliquant diverses propriétés physico-chimiques et des substituants chimiques, il est capable d'expliquer et de prédire l'activité biologique d'une série de molécules similaires.(121)

1. f.1.1. Limites des méthodologies 2D QSAR classiques

Les méthodes QSAR classiques sont beaucoup plus simples, plus rapides et plus faciles à automatiser que les approches 3D-QSAR. Elles incluent des descripteurs physico-chimiques clairement définis et sont les mieux adaptées à l'analyse d'un grand nombre de composés et au criblage computationnel de bases de données moléculaires.(121)

Bien qu'elles soient utilisées depuis des dizaines d'années pour corrélérer et prédire l'activité des molécules, elles souffrent de sérieuses limitations dans certaines situations, dont certaines sont les suivantes :

- Seules les structures 2D sont prises en compte
- Indisponibilité de paramètres physico-chimiques appropriés (par exemple, des descripteurs numériques) pour des substituants nouveaux ou inhabituels rend le composé impropre à l'inclusion dans l'analyse QSAR.
- Paramètres insuffisants pour décrire les interactions médicament-récepteur
- Limitée à seulement quelques substitutions dans une structure de référence commune (simple variation des substituants aromatiques) et fonctionne mieux avec une série de congénères.

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

- Pas de représentation de la stéréochimie ou de la structure 3D des molécules, quelle que soit leur disponibilité.
- Risque élevé de corrélations fortuites.
- Risque élevé d'échec dû à des prédictions " trop éloignées ".
- Pas de résultats graphique, ce qui rend l'interprétation en termes chimiques familiers, souvent difficile, voire impossible.
- Nécessite une connaissance considérable des constantes des substituants en chimie organique physique pour concevoir une molécule, puisque l'équation classique du QSAR ne suggère pas directement de nouveaux composés à synthétiser.(118)

II.3.1.1.6.2 3D QSAR

Au début des années 1980, une nouvelle approche de l'étude de la relation structure-activité a été proposée, qui comprend l'étude des propriétés moléculaires des composés chimiques dans une grille 3D. Ces propriétés calculées sont ensuite corrélées avec l'activité biologique, à l'aide d'une technique appelée DYLOMMS (dynamic lattice oriented molecular modeling system).

Les méthodes 3D QSAR se concentrent largement sur toutes les propriétés des atomes d'un composé qui sont représentées en tant que descripteur et qui correspond principalement à la représentation spatiale d'une molécule. (119)

Comme son nom l'indique, la méthode 3D QSAR comprend des descripteurs qui décrivent les caractéristiques 3D d'une molécule pour développer un modèle QSAR.

Diverses caractéristiques géométriques, physiques et des descripteurs de chimie quantique peuvent être utilisés pour décrire les propriétés 3D des ligands dans la méthode QSAR. Ces descripteurs moléculaires sont ensuite combinés pour créer un pharmacophore qui peut expliquer l'activité biologique des ligands.(117)

1. f.2.1. Méthodes 3D QSAR

Puisque la méthodologie 3D-QSAR est beaucoup plus complexe sur le plan informatique que l'approche 2D-QSAR, les descripteurs correspondants sont aussi plus complexes.

En général, elle implique plusieurs étapes pour obtenir des descripteurs numériques de la structure du composé. Tout d'abord, la conformation du composé doit être déterminée soit à partir de données expérimentales, soit à partir de la mécanique moléculaire, puis affinée en minimisant l'énergie. Ensuite, les conformères de la base de données doivent être alignés uniformément dans l'espace. Enfin, l'espace avec le conformère immergé est sondé par calcul pour divers descripteurs. Certaines méthodes indépendantes de l'alignement des composés ont également été développées.(112)

Méthodes 3D QSAR dépendants de l'alignement :

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

Il existe plusieurs méthodes qui se concentrent sur l'alignement moléculaire avant le calcul des descripteurs 3D. Ces méthodes calculent le descripteur en mettant en correspondance les atomes du récepteur, les atomes du ligand ou les complexes d'atomes du récepteur et du ligand.

Parmi les différentes méthodes dépendants de l'alignement, citons l'analyse comparative des champs moléculaires (**CoMFA**), l'analyse comparative des indices de similarité moléculaire (**CoMSIA**), la modélisation des récepteurs génétiquement évolués (**GERM**), l'analyse comparative de l'énergie de liaison (**CoMBINE**), Adaptation des champs pour la comparaison moléculaire (**AFMoC**).⁽¹¹⁹⁾

Méthodes 3D QSAR indépendants de l'alignement :

Les méthodes conventionnelles basées sur l'approche d'alignement ont de nombreuses limitations comme ils prennent du temps, peuvent affecter la sensibilité du modèle résultant.

Pour surmonter toutes ces limitations, un autre groupe de descripteurs 3D est développé. Ce sont des descripteurs invariants à la rotation et à la translation de la molécule dans l'espace. Ainsi, aucune superposition de composés n'est nécessaire.⁽¹¹²⁾

Les différentes méthodes appartenant à cette catégorie comprennent l'analyse comparative des moments moléculaires (**CoMMA**), **COMPASS**, Holo-QSAR (**HQSAR**), les descripteurs moléculaires invariants holistiques pondérés (**WHIM**), l'analyse spectrale comparative (**CoSA**) et les descripteurs indépendants de la grille (**GRIND**).⁽¹¹⁹⁾

1. f.2.2. 3D QSAR indirecte

La prédiction de l'activité biologique d'un composé chimique se fait principalement sur la base d'informations détaillées sur le récepteur et le ligand.

Cependant, dans de nombreux cas, l'information 3D des molécules réceptrices n'étaient pas connues, et dans ces cas, la méthode indirecte du 3D QSAR est la plus souvent suivie. L'approche indirecte de 3D QSAR est basée sur les informations de la molécule du ligand telles que l'alignement moléculaire des atomes, les pharmacophores, le volume, ou des champs pour générer un récepteur virtuel.⁽¹⁰⁷⁾

II.3.1.1.6.3 2D QSAR vs. 3D QSAR

Les 2D QSAR classiques sont basés sur l'hypothèse de Hansh selon laquelle trois propriétés moléculaires principales sont nécessaires pour expliquer les variations de composés congénères : les propriétés électroniques, hydrophobes et stériques. L'objectif de ces premiers QSAR n'était pas prédictif mais explicatif, afin d'améliorer la compréhension des propriétés biochimiques considérées.⁽¹²²⁾

Les premières équations QSAR établissaient une relation linéaire entre les trois descripteurs et le logarithme de l'activité. Mais des modèles non linéaires ont rapidement été introduits pour expliquer certaines propriétés, par exemple la lipophilie.

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

La disponibilité d'ensembles de données plus importants et les besoins des applications réelles ont progressivement étendu le QSAR classique pour prendre en compte d'autres propriétés moléculaires et des composés non congénères, en appliquant toutes les connaissances statistiques et algorithmiques modernes à l'analyse de ces propriétés pour créer des modèles plus sophistiqués.(111)

Ces QSAR non classiques (3D QSAR) sont aujourd'hui les plus utilisés à des fins plus prédictives qu'explicatives.

La principale différence entre le 2D et le 3D QSAR réside dans l'utilisation des méthodes statistiquement plus robustes pour la sélection des descripteurs moléculaires, comme la régression linéaire simple, la régression linéaire multiple (MLR), l'analyse en composantes principales (PCA), la régression en composantes principales (PCR), l'analyse en grappes (AGP), les réseaux neuronaux artificiels et la méthode du k-voisin le plus proche, qui aident à la prédiction quantitative d'un ensemble diversifié de propriétés 3D de composés chimiques.(119)

II.3.1.1.7 Applications dans la conception des médicaments

La conception de médicaments consiste à trouver de petites molécules dont la forme (stérique) et la charge (électrostatique) sont complémentaires à une cible biomoléculaire. Une petite modification d'un groupe fonctionnel entraînera une déviation de la bio-activité.

Le QSAR est une technique précise qui peut être appliquée dans la conception de médicaments pour trouver des composés potentiellement actifs. Cette méthode peut être utilisée uniquement avec des valeurs d'activité connues, en divisant initialement un ensemble de données en un ensemble d'apprentissage et un ensemble de test. L'ensemble d'apprentissage peut être utilisé pour construire modèle QSAR et l'ensemble de test peut être utilisé pour évaluer la performance du modèle.(106)

Certains des médicaments qui sont sur le marché avec l'aide de la découverte de médicaments basée sur les ligands sont le Zolmitriptan, la Norfloxacin et le Losartan.

La Norfloxacin est un médicament utilisé dans les infections des voies urinaires. Il a été développé à l'aide d'un modèle QSAR et approuvé par la FDA en 1986. Le losartan est utilisé pour l'hypertension et le zolmitriptan est utilisé comme traitement de la migraine (Cf. Figure 2.22)(86)

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

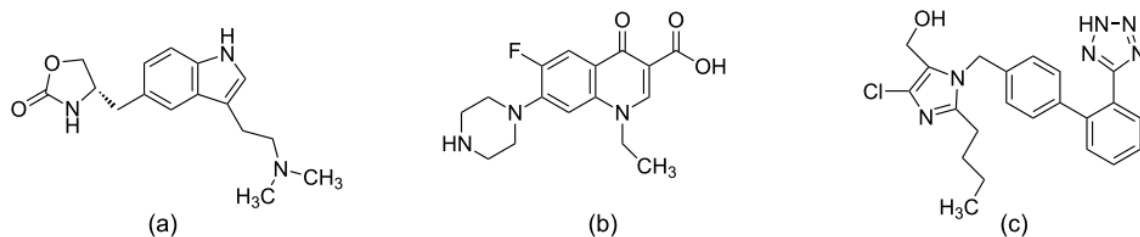


Figure 2.22-Quelques médicaments découverts à l'aide de la QSAR. **(a)**Zolmitriptan; **(b)**Norfloxacin; **(c)** Losartan.(86)

Dans une étude menée par Obiol-Pardo et al, deux modèles QSAR ont été incorporés dans un système de simulation multi-échelle conçu pour prédire la cardiotoxicité induite par des médicaments suite à des interactions avec deux canaux potassiques, hERG et KCNQ1.

Un autre exemple de ce type de méthode est l'analyse QSAR des inhibiteurs de CCR2, par Saghaie et al. CCR2 est une chimiokine couplée à une protéine G, un récepteur important dans certaines inflammations.(108)

II.3.1.2 Modélisation du pharmacophore

Les modèles pharmacophores se sont révélés utiles pour l'identification des Leads et son optimisation, l'acquisition de données sur la relation structure-sélectivité, le profil de l'activité des ligands in silico, la prédiction des sites de métabolisme, des interactions médicamenteuses, des effets secondaires et de la toxicité et, enfin, la conception de novo des médicaments d'une manière efficace sur le plan informatique.(123)

Récemment, avec l'apparition de médicaments multi-cibles, le modèle pharmacophore a un grand potentiel pour prédire sa polypharmacologie et participe même à sa découverte.

II.3.1.2.1 Historique

Le mérite de la première utilisation du concept de pharmacophore revient à Paul Ehrlich. Il a donné la première définition d'un pharmacophore en 1890. Il a développé les premières explorations sur la relation entre la structure chimique et l'activité biologique, ce qui conduit à l'élaboration de la théorie des récepteurs de médicaments. L'un des principes de cette théorie était que certaines fonctionnalités des molécules actives, appelées pharmacophores, déterminaient l'activité.(124,125)

Le concept de pharmacophore n'a pu atteindre sa pleine utilité qu'avec le développement de logiciels de recherche de bases de données 3D dans les années 1990. Le premier programme informatique, MOLPAT pour reconnaître les modèles de pharmacophore, a été développé par Günd, Wipke et Langridge à l'université de Princeton en 1974.(125)

Ces dernières années, plusieurs méthodes ont été développées et publiées afin de générer des ensembles de conformations diverses et pharmacologiquement pertinentes qui peuvent être utilisées dans des protocoles de recherche de pharmacophores en 3D pour augmenter le nombre de résultats significatifs de ces expériences.(126)

Après un siècle de développement, le concept de base du pharmacophore reste inchangé, mais sa signification intentionnelle et son champ d'application ont été considérablement élargis.(127)

Ce concept, dans la recherche moderne des médicaments, revêt une grande importance et favorise le développement de l'industrie pharmaceutique, comme l'indique le nombre de publications disponibles.

L'ensemble du développement et l'histoire du concept de pharmacophore au cours du dernier siècle a été examiné par Gund et Wermuth.(127)

II.3.1.2.2 Définition du pharmacophore

Le concept de pharmacophore a été introduit pour la première fois en 1909 par Ehrlich, qui a défini le pharmacophore comme "un cadre moléculaire qui porte (phoros) les caractéristiques essentielles responsables de l'activité biologique (pharmacon) d'un médicament". (127)

Il découle de cette définition qu'un pharmacophore définit les caractéristiques nécessaires qu'un ligand actif doit posséder. En général, les caractéristiques principales, la position et la direction d'un ligand actif, sont encodés dans un modèle de pharmacophore, ainsi que les éventuelles contraintes stériques.(108)

Selon la définition très récente de l'UIPAC (International Union of Pure and Applied Chemistry), un modèle pharmacophore est "un ensemble de caractéristiques stériques et électroniques nécessaire pour assurer les interactions supramoléculaires optimales avec une cible biologique spécifique et pour déclencher (ou bloquer) sa réponse biologique". (103)

En dehors de cette définition officielle, d'autres définitions similaires, ainsi que des remarques, ont été décrites dans la littérature.(127)

Le modèle pharmacophore représente un ensemble de points dans l'espace avec certaines propriétés et distances entre eux, qui définissent la liaison d'un groupe donné de ligands avec la cible.

Les points du pharmacophore peuvent être les atomes chargés positivement et négativement, des groupes cycliques, des atomes donneurs ou accepteurs. Un tel modèle est conçu par l'alignement d'un ensemble de ligands connus à révéler le meilleur regroupement de points pharmacophores similaires de différentes molécules dans une même zone spatiale.(104)

Il existe deux types de pharmacophores : les pharmacophores basés sur la structure, directement dérivés des structures aux rayons X de complexes protéine-ligand, et les pharmacophores basés sur le ligand dérivés des structures de composés actifs connus.(105)

Le pharmacophore basé sur le ligand est habituellement défini comme un modèle relationnel de caractéristiques qui sont communes aux ligands qui interagissent avec le même récepteur.(128)

Cependant, il est presque impossible de trouver des composés dans une nouvelle classe de composés chimiques, généralement les composés sélectionnés représentent des analogues de structures connues .(104)

II.3.1.2.3 Principe de base

Cette approche est basée sur les caractéristiques des ligands et les conformations 3D de ligand-récepteur et les règles stéréochimiques qui déterminent le mode de liaison préféré.

Les caractéristiques susmentionnées comprennent généralement des donneurs et des accepteurs de liaison hydrogène, les groupes chargés et les cycles aromatiques. Les relations

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

spatiales des caractéristiques pharmacophoriques peuvent ensuite être facilement transformées en coordonnées 3D et en distances, qui peuvent servir de base à des recherches de similarité.(128)

Un modèle de pharmacophore peut être généré à partir d'un ensemble de ligands connus. Cependant, les données relatives à la structure protéique tridimensionnelle ou aux complexes protéine-ligand, combinées à des informations sur les sites actifs, peuvent également être utilisées pour modéliser un pharmacophore en étudiant le site de liaison et les interactions possibles.(129)

En l'absence d'informations sur les récepteurs en 3D et d'un ensemble de ligands actifs, il est possible de créer des modèles pharmacophores 3D dérivés de séquences nucléotidiques des cibles.(122)

Cependant, la modélisation de pharmacophore est recommandée lorsque l'on dispose d'informations sur un ensemble de ligands pour une cible donnée sans structure 3D résolue. Dans un tel cas, les détails sur un site de liaison à la cible doivent être récupérés à partir de ligands connus et utilisés dans le criblage pharmacophore.(124)

Les stratégies basées sur les pharmacophores sont largement appliquées dans les projets de criblage et dans certains cas, elles surpassent les résultats du Docking.(128)

Elles sont étroitement liées au 3D QSAR ou 4D QSAR et souvent réalisées lors de l'étape d'alignement moléculaire des études de modélisation QSAR.

Plusieurs programmes couramment utilisés pour la génération automatique de pharmacophores incluent Discovery Studio, PHASE, LigandScout, MOE, GASP et Catalyst. Il a été rapporté que GASP et Catalyst sont plus performants que DISCO.(112)

En conséquence, la génération d'un modèle, avec une sensibilité et une spécificité bien équilibrées, sert à réduire les résultats faux négatifs et faux positifs, respectivement.(122)

II.3.1.2.4 Génération du modèle pharmacophore

Le déroulement détaillé de la construction d'un modèle pharmacophore dépendra du logiciel que l'on utilise. Cependant, la procédure générale peut être résumée comme suit (Cf. Figure 2.23)

1. La première étape consiste à assembler un ensemble de composés actifs, structurellement différents, généralement à partir de recherches bibliographiques et d'interrogations de bases de données moléculaires. Un seuil cohérent doit être appliqué lors de la recherche de composés actifs provenant de sources multiples.(108)

Pour la construction du modèle pharmacophore, l'ensemble de données sur les ligands doit être sélectionné avec beaucoup de soin. Le type de ligands, la taille de l'ensemble de données et sa diversité chimique affectent le modèle pharmacophore final.(125)

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

Les niveaux d'activité des composés ne sont pas exigés s'il n'est pas nécessaire de construire un modèle QSAR basé sur le pharmacophore. Comme le modèle devrait être capable de discriminer entre des molécules avec et sans bio-activité, l'ensemble des molécules devrait inclure à la fois des composés actifs et inactifs.(108,120)

2. Analyse conformationnelle :

(a) Pour un modèle pharmacophore 2D : les types d'atomes essentiels et leur connectivité sont définis

(b) Pour un modèle pharmacophore 3D : les conformations sont définies en utilisant nomenclature de l'IUPAC.(112)

L'ensemble de conformations, à basse énergie, généré est susceptible de contenir la conformation bioactive pour chacune des molécules sélectionnées.(120)

Selon l'IUPAC, les conformations sont définies comme la disposition spatiale des atomes permettant de distinguer les stéréo-isomères.(108)

3. L'alignement ou superposition moléculaire: Superposer ("ajuster") toutes les combinaisons des conformations à basse énergie des molécules. Des groupes fonctionnels (bioisostériques) similaires communs à toutes les molécules de l'ensemble peuvent être adaptés (par exemple, des cycles phényle ou des groupes acide carboxylique). L'ensemble des conformations (une conformation de chaque molécule active) qui donne le meilleur ajustement est présumé être la conformation active.(120)

4. Abstraction : Transformer les molécules superposées en une représentation abstraite. Par exemple, des noyaux phényle superposés pourraient être désignés de manière plus conceptuelle comme un élément pharmacophore 'à cycle aromatique'. De même, les groupes hydroxy peuvent être désignés comme élément pharmacophore «donneur / accepteur de liaison hydrogène».(120)

Les modèles de pharmacophore sont générés par un algorithme d'élucidation des pharmacophores. Cette procédure permet souvent d'obtenir plus d'un modèle de sortie, et la sélection des meilleurs modèles parmi les résultats est une tâche importante.

Cette tâche est généralement assistée par la fonction de score contenue dans le logiciel de construction de pharmacophores. Généralement, le modèle avec le score le plus élevé est le membre le plus représentatif de l'ensemble des modèles obtenus.(108)

5. Validation : Un modèle de pharmacophore est une hypothèse qui tient compte des activités biologiques observées d'un ensemble de molécules qui se lient à une cible biologique commune. Le modèle n'est valable que dans la mesure où il est capable de prendre en compte les différences d'activité biologique d'une gamme de molécules.

Au fur et à mesure que les activités biologiques des nouvelles molécules deviennent disponibles, le modèle pharmacophore peut être mis à jour pour l'affiner davantage.(120)

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

La plupart des modélisateurs de pharmacophores conduisent l'ensemble du flux de travail à l'aide d'un logiciel de chimie-informatique, car tous les outils nécessaires se trouvent dans ce type de logiciel.(108)

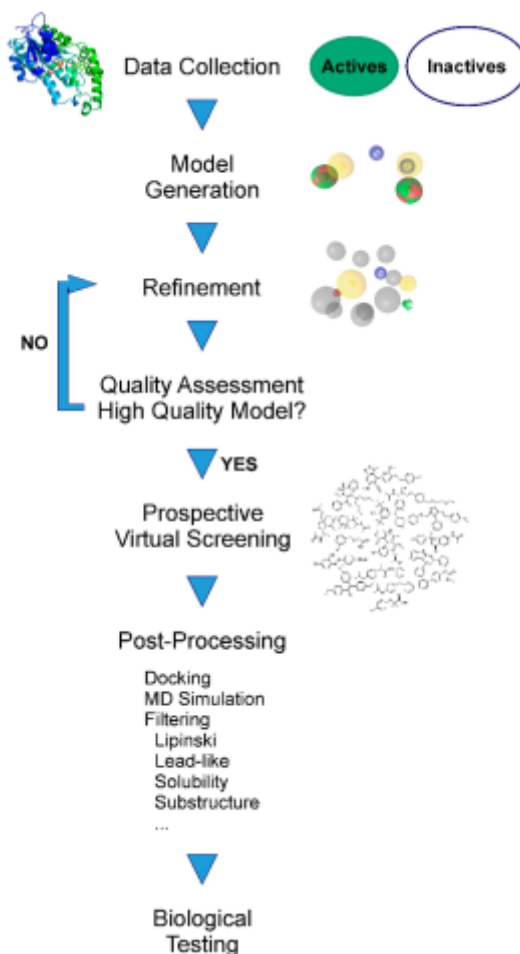


Figure 2.23-Flux de travail pour la modélisation des pharmacophores.(130)

II.3.1.2.5 Limites des méthodes pharmacophores

Comme pour toute méthode, malgré l'abondance de cas réussis de conception de médicaments reposant sur la modélisation du pharmacophore, elle n'est pas toujours fiable. Il faut donc être prudent quant aux limites de cette technique.

La principale, étant l'absence de bonnes mesures de scoring. Alors que les simulations de Docking sont basées sur des fonctions de scoring essayant de prédire l'affinité.(120)

La deuxième est la dépendance du logiciel de génération des pharmacophores sur une base de données de conformation pré-calculée. Ces bases de données ne contiennent qu'un nombre limité de conformations de faible énergie par molécule. Il est donc possible qu'une molécule active ne puisse pas être identifiée car sa conformation est absente. Ceci est particulièrement le cas pour les nombreuses conformations différentes des liaisons rotatives de petites fonctionnalités moléculaires telles que les groupes hydroxyle. (129)

II.3.1.2.6 Applications dans la découverte des médicaments

Les modèles de pharmacophore sont largement utilisés pour obtenir des inhibiteurs spécifiques de protéines liées à des maladies, notamment les récepteurs couplés aux protéines G, les enzymes et les canaux ioniques. Ils sont également utilisés avec d'autres méthodes de découverte de médicaments (Cf. Figure 2.24)

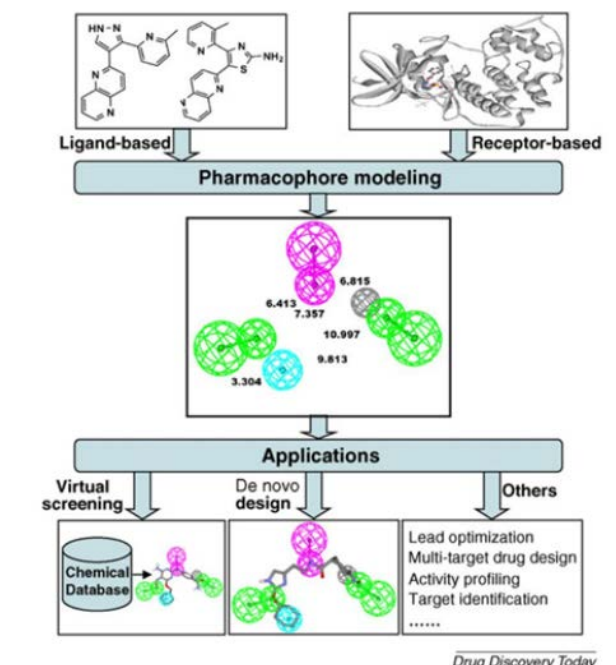


Figure 2.24–Importance de la modélisation de pharmacophore dans le processus du Drug Design.(127)

Les modèles pharmacophores sont souvent utilisés dans les processus de criblage virtuel. Par exemple, Mustata et al. ont découvert de nouveaux perturbateurs de l'hétérodimère Myc-Max en utilisant un modèle de pharmacophore généré en utilisant des perturbateurs connus. Dans cette étude, les auteurs ont effectué le criblage virtuel avec le modèle de pharmacophore après qu'il ait été validé en utilisant un ensemble de molécules sans entraînement. Des essais biologiques ont ensuite été menés pour vérifier la puissance des molécules criblées.(108)

La figure 2.25 présente une recherche de base sur le Web of Knowledge (Thomson Reuters, <http://www.webofknowledge.com>) sur les requêtes " pharmacophore " et " criblage virtuel ". Une augmentation marquée des publications au cours des dernières années démontre l'intérêt croissant des recherches dans le criblage virtuel à base des pharmacophores pour la découverte des médicaments.(123)

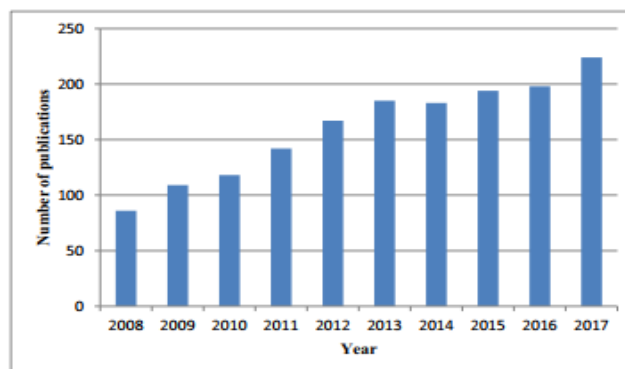


Figure 2.25 -Aperçu chronologique du nombre de publications parues au cours des dernières années, correspondant aux requêtes "pharmacophore" et "criblage virtuel".(123)

En plus des utilisations décrites ci-dessus, les modèles de pharmacophores peuvent être utilisés comme filtres pour réduire les faux positifs, soit pendant le prétraitement d'une base de données moléculaires avant de la soumettre à une autre procédure de criblage virtuel, soit pour filtrer les résultats d'autres processus de criblage.

Les modèles de pharmacophore ne sont pas seulement appliqués sur les enzymes, mais aussi sur les récepteurs membranaires (tels que les RCPG), les facteurs de transcription, les canaux ioniques, les transporteurs, ainsi que les acides nucléiques. L'interaction protéine-protéine (IPP) a représenté un trésor de nouvelles cibles potentielles pour les médicaments, étant donné qu'elle est cruciale et omniprésente dans la fonction cellulaire.

Les méthodes d'identification rationnelle de petites molécules inhibitrices de ces interactions restent limitées, la modélisation pharmacophore offre une méthode efficace pour les chercheurs afin d'obtenir rapidement les premiers succès pour une optimisation ultérieure.(123)

Une molécule anticancéreuse d'origine naturelle identifiée à l'aide du QSAR est l'I3C (indole-3-carbinol). Cependant, cette molécule n'a jamais dépassé les essais cliniques en raison de sa faible puissance. Ce composé actif a été optimisé à l'aide d'une modélisation pharmacophore basée sur les ligands pour développer l'analogue SR13668 qui s'est révélé très puissant contre plusieurs types de cancer.(112)

II.3.1.3 Recherche de similarité

Le nombre dramatiquement croissant de composés qui deviennent disponibles pour une évaluation biologique représente un défi important pour la conception, la gestion et l'exploration.

Les approches informatiques pour le criblage, le profilage, ou le filtrage de grandes collections de composés sont maintenant largement utilisées dans la recherche pharmaceutique.

Parmi les techniques populaires de classification des composés et d'exploration de bases de données de plus en plus volumineuses, les méthodes de similarité moléculaire sont très efficaces sur le plan informatique, car elles ne dépendent pas des comparaisons par paires de composés pour évaluer la similarité ou la diversité moléculaire.(115)

II.3.1.3.1 Principe général

L'idée principale des approches basées sur la similarité ou l'empreinte digitale est de sélectionner de nouveaux composés sur la base de la similarité chimique et physique avec des médicaments connus pour la cible.

Les méthodes de recherche de la similarité des ligands sont des approches simples mais efficaces, basées sur la théorie selon laquelle les molécules structurellement similaires ont tendance à avoir des propriétés de liaison similaires et augmentent les chances de partager un profil bioactif commun. (86)

Ainsi, ce concept peut être utilisé lors de la création de bibliothèques " ciblées " contenant des composés similaires à au moins un véritable ligand à son récepteur spécifique, par exemple un substrat, un inhibiteur ou un cofacteur. Sur la base d'une recherche de similarité structurelle parmi les ligands, il est possible de retrouver des composés contenant des sous-structures identiques qui partagent l'affinité pour les mêmes récepteurs.

Néanmoins, il est important de garder à l'esprit que certaines structures se retrouvent de manière répétée dans les bases de données des médicaments connus. (117)

En général, lors de l'analyse de systèmes biologiques, on ne connaît ni les composés actifs ni le mode de liaison des cofacteurs naturels. S'il n'y a pas d'informations structurelles à utiliser dans les recherches de similarité, il est possible de définir les modes de liaison préférés des ligands, en criblant des bibliothèques de composés vastes et structurellement diversifiées.

Conformément à la stratégie de l'entonnoir, une grande quantité de données chimiques est d'abord étudiée pour extraire un ensemble plus restreint de composés. Ensuite, cet ensemble est analysé plus en détail. En utilisant ces bibliothèques, des résultats plus variés peuvent être obtenus, montrant des structures chimiques nouvelles et innovantes.

Des scaffolds structurellement diversifiés, sélectionnés par cette méthode doivent être analysés et optimisés pour l'identification de structures chimiques intéressantes.(128)

CHAPITRE 2: CLASSIFICATION SIMPLE DES METHODES

Dans la recherche de similarité chimique, les propriétés 3D des composés peuvent également être incluses. En utilisant des descripteurs 3D, il est possible d'appliquer des informations sur la forme d'un ligand et d'obtenir des informations utiles concernant les interactions possibles.

Goldman et Wipke ont présenté cette nouvelle approche de la recherche de similarité moléculaire basée sur la forme. La méthode introduite par les auteurs est capable de superposer des molécules dissemblables en utilisant des descripteurs de surface moléculaire géométriquement invariants.(115)

II.3.1.3.2 Notion d'empreintes digitales

Fondamentalement, les recherches de similarité utilisent les propriétés moléculaires 2D (empreintes moléculaires, descripteurs topologiques, propriétés moléculaires) ou 3D (pharmacophore) d'un ou plusieurs composés actifs pour calculer un indice de similarité (en général, coefficients de Tanimoto, Dice ou Tversky, voire des scores d'ajustement spatial, dans le cas de pharmacophores 3D) qui est utilisé pour classer les composés dans une base de données de composés inconnus.(131)

Un aspect essentiel du filtrage basé sur la recherche de similarité est la sélection des "empreintes" digitales appropriées, qui peuvent être définies comme une représentation mathématique des molécules.(128)

Les empreintes moléculaires (ou fingerprints) sont une classe de descripteurs très utilisée en chimie-informatique et notamment pour la recherche par similarité et pour les études de diversité.

Un fingerprint est un ensemble fini et non ordonné de descripteurs, généralement représenté sous la forme d'un vecteur de bits (les éléments du vecteur prenant leurs valeurs dans 0, 1). Chaque élément du vecteur représente une propriété, typiquement la présence ou non d'un certain motif (fragment, atome, liaison, etc.)(22)

En d'autres termes, les informations chimiques du composé sont traduites en données qui peuvent être facilement comprises et traitées par les méthodes de calcul. La principale force de cette approche réside dans sa capacité à comparer plusieurs empreintes digitales et à calculer leur similarité en utilisant par exemple le coefficient de Tanimoto, ce qui facilite grandement les recherches de similarité moléculaire.

Les empreintes digitales peuvent être classées en fonction de leur dimensionnalité allant de l'unidimensionnel (1D) au (3D)(Cf. Figure 2.26).

Parmi toutes celles qui sont couramment utilisées, les plus populaires et les plus efficaces sont les empreintes digitales 2D (plus que la forme 3D ou les méthodes de Docking) et ceci grâce à leurs facilité pour retrouver des composés actifs.(128)

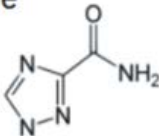
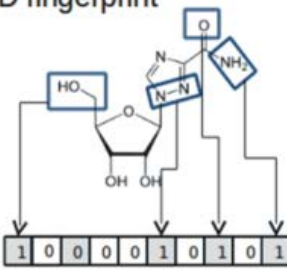
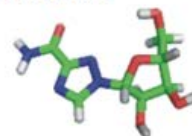
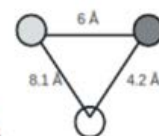
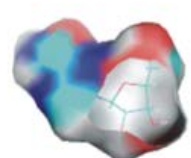
1D	2D	3D
Chemical formula Molecular weight Number of atoms Number of bonds Number of rings Polar surface area LogP	Substructure  2D fingerprint 	Structure  Pharmacophores  Shape 

Figure 2.26- Empreintes digitales 2D comme descripteurs de ligands. (128)

II.3.1.3.3 Applications dans la découverte des médicaments

Ce type de méthodologie est très populaire en raison de son faible coût de calcul et de sa facilité d'utilisation par rapport aux approches basées sur la structure. Le protocole de recherche de similarité utilisé dans le criblage virtuel repose sur le principe selon lequel les composés qui présentent une similarité structurelle ont également des activités biologiques similaires.

En général, cette hypothèse est valable et il existe plusieurs cas réussis de recherche de similarité dans l'identification de composés bioactifs. Cependant, dans certains cas, la similarité structurelle est un paramètre utilisé pour exclure des composés similaires lors de criblages virtuels afin d'obtenir des scaffolds innovants de composés bioactifs pour le système étudié.(131)

Comme exemple, un agoniste spécifique, qui active GPR30 (Récepteur couplée à la protéine G 30), a été développé en utilisant des recherches de similarité. Le score de similarité final qui a été utilisé comprenait un score 2D et un composant de similarité de structure 3D.(86)

II.4 Conclusion

L'objet de ce chapitre est de faire une étude détaillée des méthodes de modélisation appliquées à l'analyse des mécanismes de reconnaissance moléculaire et à la conception de nouveaux composés bioactifs à l'aide des techniques informatiques. La faisabilité et l'efficacité de ces méthodes sont démontrées par leur application à de nombreux cas.

Les méthodes étudiées sont représentatives des deux grands axes de conception de molécules par modélisation : la conception basée sur la structure des récepteurs SBDD et la conception basée sur la structure des ligands LBDD.

Dans le cadre du **premier axe SBDD**, on a étudié les méthodes de construction de protéine par homologie ou bien « ab initio » et la méthode d'amarrage moléculaire « DOCKING », ensuite la DM.

Dans le cadre du **second axe LBDD**, on a détaillé les descripteurs chimiques permettant de construire des modèles QSAR des composés biologiquement actifs en décrivant le processus en premier lieu, puis on a abordé la technique de modélisation des pharmacophres en citant ses applications majeures et ses limites, et on a terminé par quelques notes sur les recherches de similarité. (45)

**III. CHAPITRE 3:
Développement actuel des
recherches**

III.1 Introduction

Dans la conception des médicaments, l'intelligence artificielle (IA) connaît une vague d'effervescence que l'on n'avait pas vue depuis l'émergence de la chimie computationnelle à la fin des années 1980 et au début des années 1990.

Apparemment, les leçons de l'histoire récente n'ont pas été retenues, on nous promet des solutions imminentes et omniprésentes aux maux de la conception de médicaments et des augmentations significatives de la productivité dans notre quête de nouvelles thérapies innovantes.

Cependant, des questions importantes restent-elles à résoudre avant que l'IA n'entre dans la boîte à outils quotidienne du chimiste médicinal en exercice ?(132)

Il existe également d'autres méthodes purement informatiques en cours d'évolution qui s'occupent du stockage des données énormes appartenant à la chimie médicinale. Elles seront traitées ultérieurement.

III.2 L'intelligence artificielle (IA)

Gartner est une société de recherche et de conseil qui publie chaque année le "cycle de l'effervescence technologique". Si vous passez un peu de temps dans le monde de la santé numérique, vous verrez que le terme "cycle d'effervescence" ou « Hype cycle » est fréquemment utilisé. Ce concept décrit essentiellement ce qui se passe lorsque de nouvelles technologies entrent dans la sphère publique et génèrent de l'excitation et l'incertitude.

Actuellement, les technologies qui sont au sommet du Hype Cycle de Gartner sont toutes, fondamentalement, associées à l'intelligence artificielle.(133)

III.2.1 Qu'est-ce que l'IA ?

Il existe de multiples définitions. Selon l'une d'entre elles, il s'agit de l'intelligence humaine manifestée par des machines. En informatique, le domaine de la recherche sur l'IA la définit comme l'étude des "agents intelligents", qui sont "des dispositifs qui perçoivent leur environnement et prennent des mesures pour maximiser leurs chances de réussite dans un but donné".(133)

L'intelligence artificielle est définie, en termes très généraux, comme un domaine qui traite de la conception et de l'application des algorithmes d'analyse, d'apprentissage et d'interprétation des données.

Ainsi, l'IA au sens large englobe de nombreux domaines tels que l'apprentissage statistique et automatique, de la reconnaissance des formes, le regroupement, les méthodes basées sur la similarité, la logique et la théorie des probabilités ainsi que des approches à motivation biologique, telles que les réseaux neuronaux, l'informatique évolutionniste ou la modélisation floue, collectivement décrites comme "l'intelligence informatique".(134)

Les différentes définitions et interprétations de ce terme s'accordent sur trois capacités essentielles de l'IA (se référant le plus souvent à un ordinateur ou à une machine) :

1. La résolution de problèmes.
2. L'apprentissage par l'expérience (mémoire et adaptation)
3. L'adaptation à de nouveaux défis (généralisation).(135)

III.2.2 Contexte historique de L'IA en conception moléculaire

L'IA avec le ML (machine learning) ne sont pas nouveaux pour les chercheurs en conception moléculaire assistée par ordinateur. La DL (deep learning) trouve ses racines dans les années 1960, du moins dans sa forme théorique, avec les travaux d'Ivakhnenko et Lapa. L'IA trouve ses racines encore plus loin, à un atelier organisé au Dartmouth College en 1956.

Même en tenant compte de la longue histoire de l'IA, et généralement plus longue que beaucoup ne l'imaginent, le domaine a connu un certain nombre d'hivers où les attentes ne

CHAPITRE 3: DEVELOPPEMENT ACTUEL DES RECHERCHES

correspondant pas à la réalité. Cela a entraîné un certain nombre de revers pour le domaine et il a fallu du temps pour s'en remettre.(136)

En 2015, les avancées informatiques, telles que l'intégration généralisée de GPU dans les cadres informatiques modernes et l'augmentation de la quantité de RAM disponible, la formation de réseaux neuronaux plus grands et plus profonds est devenue possible. Lors du célèbre défi Kaggle, une équipe de Toronto, a utilisé un réseau neuronal profond pour remporter un défi de SAR lancé par Merck.

Cette compétition est communément perçue comme un tournant dans lequel une méthode complexe de DL de l'IA a surpassé d'autres approches de ML et est donc devenue un outil utile pour la conception moléculaire computationnelle. (137)

Ainsi, La puissante capacité d'exploration de données de la technologie de l'IA a donné une nouvelle vitalité à la conception de médicaments assistée par ordinateur, ce qui favorise et accélère fortement le processus.(138)

Bien qu'il existe aujourd'hui de multiples applications prometteuses de l'IA pour dériver des descripteurs moléculaires et comprendre leur relation avec les propriétés biologiques, ces méthodes sont, intrinsèquement, liées au big data.

Ces algorithmes sont généralement très gourmands en données avant de pouvoir fournir des solutions utiles, en prime, ils offrent des possibilités sans naviguer préalablement dans de grands ensembles de données.(136)

Au cours de la dernière décennie, le nombre d'articles publiés en utilisant certains algorithmes d'IA dans la découverte de médicaments est illustré dans la figure 3.1, dont les données sont générées à partir de la base de données de "SciFinder". Bien que le nombre de modèles bayésiens utilisés soit important, les modèles DL présentent une croissance quasi exponentielle. (138)

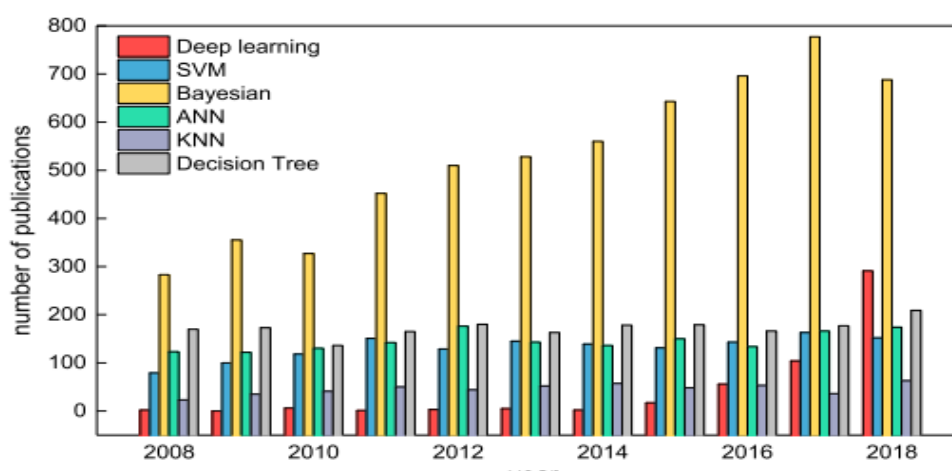


Figure 3.1- Evolution du nombre de publications utilisant différentes méthodes d'intelligence artificielle dans découverte de médicaments pendant la période [2008-2018].(138)

CHAPITRE 3: DEVELOPPEMENT ACTUEL DES RECHERCHES

L'intelligence artificielle et l'automatisation des laboratoires augmentent la prise de décision humaine pour la synthèse chimique et les tests biologiques dans le cycle conception-synthèse-test-analyse.

On s'attend à ce que cette intelligence collaborative issue de la combinaison de "l'esprit et de la machine" permettra de prendre de meilleures décisions dans la découverte de médicaments. (Cf. Figure 3.2) (135)

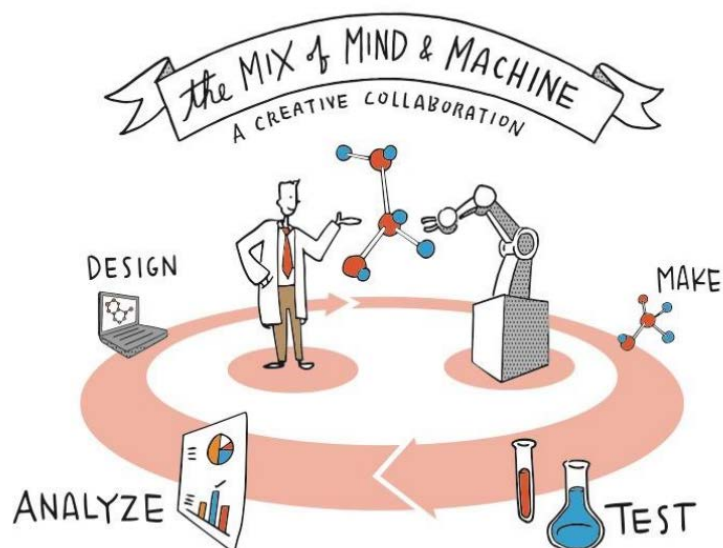


Figure 3.2- Collaboration de l'esprit et la machine dans la découverte des médicaments.(135)

Pour réussir à long terme, la conception de médicaments à l'aide de l'IA doit apporter des solutions à plusieurs questions. Elles peuvent être regroupées en cinq " grands défis "

1. Générer et obtenir des ensembles de données appropriés.
2. Générer de nouvelles hypothèses.
3. Optimisation dans un mode multi-objectif.
4. Réduire les temps de cycle.
5. Changer la culture de la recherche et créer un état d'esprit approprié.

Dans le but de faire la lumière sur ces défis et sur les approches permettant de les relever, un groupe d'experts internationaux issus de divers milieux et institutions scientifiques s'est réuni à San Francisco, en décembre 2018, pour un atelier organisé par l'ETH Zurich, afin de repenser la conception de médicaments avec l'IA.(135)

III.2.3 Application de l'IA dans la conception des médicaments

Avec l'apparition successive des propriétés de Google (d'AlphaGo, AlphaGoZero et AlphaZero) et leurs remarquables performances, la technologie de l'intelligence artificielle est devenue le centre d'attention du monde entier, encore et encore.

Avec l'émergence des ordinateurs, en particulier des grappes de calcul parallèle à haute performance (HPC), la puissance de calcul s'est rapidement améliorée avec le développement

CHAPITRE 3: DEVELOPPEMENT ACTUEL DES RECHERCHES

de l'unité de traitement graphique (GPU) et l'accumulation explosive de données informatiques, une technologie d'IA dotée d'une forte capacité de généralisation et d'extraction de caractéristiques émerge dans la conception de médicaments.(138)

En octobre 2018, l'Agence pour les projets de recherche avancée de la défense (Defense Advanced Research Projects Agency, DARPA) a annoncé le lancement du programme de découverte moléculaire accélérée (AMD), qui vise à développer de nouvelles approches systématiques basées sur l'IA pour accélérer la découverte et l'optimisation de molécules, y compris des molécules de médicaments. (132)

En outre, des entreprises pharmaceutiques internationales, comme Merck, Sanofi, Genentech et Takeda, ont lancé des efforts de coopération pertinents avec des entreprises d'IA. Tous ces exemples démontrent pleinement l'énorme potentiel d'application de la technologie de l'IA dans le domaine de le Drug Design.(134)

L'apprentissage machine (ML) est un sous-domaine important de l'IA. Par rapport aux algorithmes d'apprentissage traditionnels, il ne s'appuie pas sur les progrès de la physique et la chimie complexes, mais comprend d'énormes données biomédicales et les convertit en connaissances réutilisables.

L'apprentissage profond (DL) est le dernier développement de la technologie de l'IA à partir de ML. C'est une méthode de traitement des données prometteuse et efficace, il peut produire des résultats fiables à faible coût et dans un délai court. En tant que nouveau paradigme de ML, il se concentre sur le modèle hiérarchique profond des données pour leur apprentissage. (138)

Les ML traditionnels nécessitent la création et l'extraction manuelles de caractéristiques à partir de données brutes, tandis que le DL contient plusieurs couches cachées et des neurones, qui peuvent extraire automatiquement des caractéristiques abstraites de haut niveau à partir de données brutes volumineuses, hétérogènes et hautement dimensionnelles en utilisant un processus commun.

Ce processus ne nécessite pratiquement aucune intervention manuelle et a une faible erreur de généralisation, ce qui permet d'obtenir de meilleurs résultats lors des tests de référence ou des tests concurrentiels.(133)

III.2.4 L'apprentissage automatique « Machine learning » (ML)

III.2.4.1 Définition

C'est la science (et l'art) de programmer des ordinateurs pour qu'ils puissent apprendre à partir de données, c'est aussi une branche de l'intelligence artificielle qui se concentre sur une tâche parmi plusieurs d'autres : la construction de modèles.(135)

L'apprentissage automatique (ML) est une technique d'exploration et de création des modèles à partir de données. Il nous permet d'extraire des corrélations complexes et souvent cachées (et donc idéalement des idées, des modèles, des règles et des conseils) à partir d'ensembles de données et de les encapsuler sous forme mathématique.(139)

Le ML est généralement classé en quatre catégories : supervisé, semi-supervisé, non supervisé et par renforcement. La principale différence entre ces types est essentiellement la quantité d'informations (c'est-à-dire l'étiquetage, le contexte) qui est disponible pour la variable cible servant de base pour l'apprentissage de l'algorithme ML.

Bien que tous les types de ML aient trouvé une application dans la recherche chimique, l'apprentissage supervisé a été jusqu'à présent le plus couramment utilisé.(140)

III.2.4.2 Application dans la découverte de médicaments

L'application de la ML à l'exploration de l'espace chimique et à la création de nouveaux composés (allant des petites molécules aux polymères et aux matériaux) peut être divisée en deux approches distinctes :

1. La découverte : le ML est utilisé pour générer des modèles de prédiction rapides pour les propriétés d'intérêt, avec lesquels des enquêtes à grande échelle de l'espace chimique identifient les composés présentant les profils de propriétés souhaités.

2. La conception : le ML est utilisé pour développer une compréhension quantitative d'intérêt de SPR, qui sert à suivre la conception rationnelle ciblée (ou l'ingénierie inverse) de composés ayant des propriétés particulières.(139)

L'apprentissage machine (ML) est devenu un élément crucial de la découverte précoce de médicaments. Ce domaine de recherche a été alimenté par deux facteurs principaux :

1. L'augmentation rapide de la disponibilité de données expérimentales pertinentes. Ces ensembles de données sont par exemple ceux qui contiennent les bioactivités de molécules de structure chimique connue contre une cible non moléculaire (par exemple, une lignée de cellules cancéreuses), les affinités de liaison de ces molécules contre une cible moléculaire (par exemple, une kinase particulière validée pour un type de cancer spécifique) ou les structures cristallines aux rayons X d'une cible moléculaire.

CHAPITRE 3: DEVELOPPEMENT ACTUEL DES RECHERCHES

Ce facteur a été stimulé par le développement de ressources communautaires, telles que ChEMBL, PubChem, NCI-60 ou PDBbind, qui conservent et facilitent la réutilisation de ces ensembles de données pour la modélisation prédictive.

2. L'accès facile à des implémentations de haute qualité et bien documentées d'une gamme d'algorithmes ML, y compris ceux des avancées récentes telles que XGBoost, l'apprentissage profond « Deep learning » (DL), ou la prédiction conforme.

En conséquence, un nombre croissant de modèles ML basés sur les données a été proposé et s'est avéré, d'une certaine manière, avantageux pour identifier de nouveaux points de départ pour le processus de découverte de médicaments.(141)

III.2.5 L'apprentissage profond « Deep learning » (DL)

III.2.5.1 Définition

C'est un ensemble de techniques d'apprentissage automatique utilisant des réseaux neuronaux multicouches pour dériver des relations à partir de données. (135)

III.2.5.2 Application dans la découverte de médicaments

La technologie de l'IA, en particulier la méthode DL, a montré son application potentielle dans la découverte de nouveaux médicaments à l'ère du big data. Elle peut apprendre automatiquement des connaissances pharmaceutiques pertinentes et extraire des caractéristiques abstraites, de haut niveau, à partir d'une grande quantité de données pharmaceutiques. Ces données peuvent être utilisées pour découvrir et concevoir des molécules ayant les propriétés souhaitées et optimiser et améliorer le taux de réussite de l'approbation de nouveaux produits chimiques.

Plus important encore, les méthodes DL sont capables de traiter des tâches complexes, avec des ensembles de données volumineux, hétérogènes ou bien à haute dimension sans aucune intervention manuelle, ce qui s'est avéré utile dans la littérature et les applications commerciales.

La combinaison de l'analyse ML, en particulier de la DL, avec l'expertise et l'expérience humaines peut être le seul moyen d'intégrer complètement de nombreux bases de données.(138)

Les dernières techniques de DL sont également largement utilisées pour la découverte de médicaments. De nombreuses recherches ont été menées dans ce domaine. Par exemple, Erik Gawehn et al. ont décrit les modèles actuels de DL utilisés pour le calcul des modèles de QSAR, et mentionne brièvement comment ils pourraient être utilisés dans la découverte de médicaments.(142)

La méthode de DL sélectionne automatiquement des représentations à partir des données originales et des données à haute dimension lorsque les descripteurs moléculaires ne le sont pas. Les réseaux de neurones convolutifs (CNN) sont les plus couramment utilisés, ils ont fait de grands progrès dans la communauté de la vision par ordinateur, et ont été appliqués dans les domaines de la conception de médicaments, notamment l'identification de novo de molécules, l'ingénierie des protéines et l'analyse de l'expression génétique.(137)

III.3 Autres nouvelles méthodes pour améliorer la modélisation moléculaire

De nos jours, la modélisation moléculaire fait partie intégrante de nombreux projets de découverte de médicaments.

Comprendre les préférences conformationnelles d'une molécule médicamenteuse potentielle et comment elle pourrait interagir avec une protéine hôte est crucial pour prendre des décisions de conception éclairées.(143)

Les efforts de l'industrie et des universités en matière de découverte de médicaments sont généralement soutenus par des techniques de modélisation informatique.

Ces techniques de calcul utilisées vont du criblage virtuel à haut débit, basé sur la structure du ligand ou de la cible, de milliers ou de millions de composés, à l'étude plus détaillée par des simulations ou des calculs de structure électronique.

Si certaines de ces méthodes sont surtout exigeantes en termes de puissance d'unité centrale de traitement (UC), d'autres, comme les simulations et les méthodes de criblage à haut débit, requièrent en outre de grandes quantités de stockage sur disque.(144)

Cette croissance significative de la capacité à mesurer et à collecter des données en quantités auparavant inimaginables entraîne de nouvelles exigences en matière de développement et l'application de nouvelles méthodes et technologies de l'informatique.

En fonction de la stratégie de parallélisation applicable à la méthode concernée, les technologies récentes basées sur les unités centrales de traitement, par exemple, le « **Cloud computing** » et le « **Grid computing** », peuvent être utilisées pour accélérer considérablement leur temps d'exécution et diminuer le coût.(143)

Elles sont aussi capables de stocker, de transmettre et d'analyser d'avantage les données, ou potentiellement capables les d'utiliser comme entrée pour la modélisation mathématique. Cela permet au modélisateur moléculaire d'examiner des ensembles de données plus importants, ou d'utiliser des méthodes plus précises.(143,144)

La parallélisation de la construction du modèle sur un cluster d'ordinateurs est une façon, si possible, de diviser un problème en plusieurs parties et de les exécuter en parallèle, mais ce processus peut nécessiter une plus grande expertise informatique et des coûts initiaux importants. En outre, tout le monde n'a pas accès à un cluster d'ordinateurs.(145)

III.3.1 « Cloud computing » ou calcul en nuage

III.3.1.1 Définition

« Cloud computing » est un type de système parallèle et distribué constitué d'un ensemble d'ordinateurs interconnectés et virtualisés qui sont approvisionnés de manière dynamique et présentés comme une ou plusieurs ressources informatiques unifiées sur la base d'un système de gestion des données. (143)

Le « Cloud computing » est apparu relativement récemment comme une approche permettant de fournir à la demande une infrastructure de calcul à grande échelle (selon Google Trends, l'intérêt commence à se manifester aux alentours de 2007).

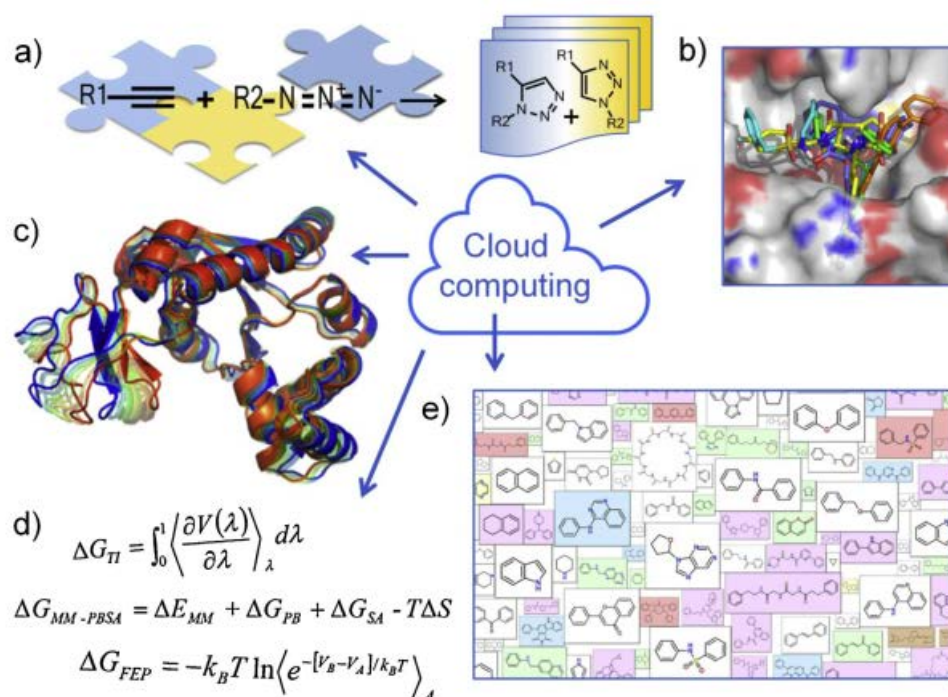
Les caractéristiques spécifiques du « Cloud computing », et qui le distinguent d'autres approches de calcul distribué et de haute performance sont toutefois nébuleuses.(146)

III.3.1.2 Pourquoi envisager « Cloud computing » pour les projets de modélisation moléculaires ?

Certaines tâches de modélisation, comme les simulations de DM, les calculs de mécanique quantique ou la construction de bibliothèques virtuelles en 3D sont très gourmandes en ressources informatiques et souvent associées à la génération ou au traitement de quantités massives de données (Cf. Figure 3.3).(145)

« Cloud computing » peut apporter une contribution majeure non seulement en termes de disponibilité de la puissance de calcul à la demande, mais aussi en termes d'innovation dans le développement de nouvelles approches qui utilisent cette capacité de manière plus efficace.(146)

CHAPITRE 3: DEVELOPPEMENT ACTUEL DES RECHERCHES



- (a) Génération de grandes bases de données de composés
- (b) Criblage basé sur la structure
- (c) Simulations de dynamique moléculaire
- (d) Des calculs d'énergie libre
- (e) L'analyse d'énormes ensembles de données générés par exemple par des projets de chimiogénomique.

Figure 3.3- Exemples d'applications du "Cloud computing" pour des projets de modélisation moléculaire.(146)

Très souvent, surtout dans les petites et moyennes entreprises, les exigences en matière de données et de calcul varient dans le temps. Trois principales raisons pour lesquelles le "Cloud Computing" pourrait être plus convaincant pour ces types de projets de modélisation moléculaire que les installations informatiques internes plus traditionnelles : l'évolutivité, la fiabilité et le moindre coût.

Les technologies émergentes du « Cloud computing » permettent aux fournisseurs de services de déployer, de migrer ou de dupliquer facilement des serveurs sans devoir investir dans une nouvelle infrastructure matérielle, et permettent d'appliquer des stratégies, d'équilibrage de charge, adaptatives basées sur des images de cloud.(147)

L'importance du « Cloud computing » pour les applications à grande échelle et à forte intensité de données est de plus en plus reconnue. On peut s'attendre à une croissance rapide de disponibilité de nouvelles applications et une baisse du coûts.(146)

III.3.1.3 Applications dans la découverte des médicaments

La découverte de médicaments est confrontée à une croissance de données similaire aux autres domaines scientifiques, et de nombreuses analyses sont exigeantes en calcul et nécessitent des calculs à haute performance (HPC).

C'est le cas, par exemple, lorsque des modèles prédictifs de relation structure-activité sont construits à partir de données *in vitro*. Pour les organisations qui ont un processus continu de réception de données expérimentales entrantes provenant d'essais *in vitro*, comme AstraZeneca R&D, il est important de fournir des modèles prédictifs actualisés, formés sur les données les plus récentes.

Dans ces cas, le temps de modélisation peut devenir un goulot d'étranglement, et il est important que la modélisation soit non seulement très précise mais aussi avec un temps raisonnable. (144)

Le «Cloud computing» est une technologie intéressante pour fournir des ressources informatiques au développement de médicaments pour de nombreuses raisons.

Tout d'abord, elle ne nécessite qu'un faible investissement matériel si l'on utilise un fournisseur de cloud public.

Ensuite, les utilisateurs ne paient que pour les ressources utilisées pendant l'analyse et peuvent augmenter ou diminuer le nombre de ressources en fonction des besoins. Pour la modélisation prédictive, cela est attrayant car la plupart de la puissance de calcul est nécessaire pendant un temps limité, et il est également facile de quantifier le coût de la construction d'un modèle (Cf. Figure 3.4).

Pour les créateurs de modèles qui n'ont pas accès à des clusters d'ordinateurs, le «Cloud computing» peut être une technologie très intéressante pour accélérer la construction de modèles. (145)

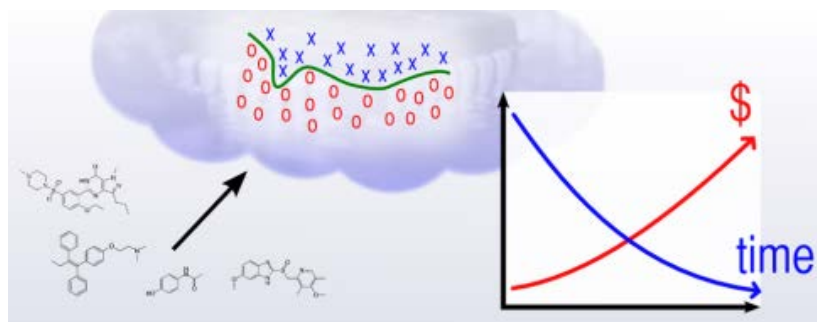


Figure 3.4- Représentation des intérêts des « Cloud computing » dans la découverte de médicaments.(145)

III.3.2 « Grid computing » : le calcul en grille

III.3.2.1 Définition

Plusieurs définitions différentes ont été proposées dans la littérature. L'une d'entre elles, une définition simple proposée par le professeur Coveney est la suivante : « Le calcul en grille est un calcul distribué effectué de manière transparente à travers plusieurs domaines administratifs ».(148)

Le terme "Grid computing" fait référence au paradigme consistant à distribuer une tâche logicielle sur un réseau d'ordinateurs ou de nœuds.(143)

III.3.2.2 Application dans les entreprises pharmaceutiques

Historiquement, l'un des problèmes qui limitent l'utilisation généralisée des ressources de Grid dans la modélisation moléculaire est l'ensemble limité d'applications appropriées se prêtant à la parallélisation à gros grains.(147)

La plupart des grandes entreprises pharmaceutiques ont utilisé le « Grid computing » pour extraire plus efficacement la valeur de leurs ordinateurs de bureaux existants et pour étendre leurs ressources de calcul, avec l'intention de minimiser les dépenses financières supplémentaires. La plupart des applications de modélisation moléculaire utilisant le « Grid computing » sont limitées à des problèmes qui nécessitent une communication interprocessus minimale.(148)

Le criblage virtuel, dans lequel un grand nombre de composés sont traités indépendamment sur plusieurs machines de calcul, est un bon exemple de problème parallèle à gros grain. Les applications de criblage virtuel sont probablement les outils les mieux connus de modélisation moléculaire qu'utilisent le « Grid computing ».

Une application est apparue dans un rapport récent utilisant FlexX et AutoDock pour cribler les inhibiteurs de la plasmepsine, une cible potentielle pour le traitement du paludisme. Les auteurs rapportent l'identification de plusieurs nouvelles classes de composés pour l'inhibition de la plasmepsine, ainsi que l'identification de composés dans plusieurs chimiotypes connus. Bien que ces résultats *in silico* n'aient pas encore été confirmés *in vitro*, l'application démontre l'utilité potentielle de "Grid computing" pour le VS.(143)

Bien qu'utiles et importantes, il y a eu une pénurie d'autres applications de grille dans la découverte pharmaceutique.

CHAPITRE 3: DEVELOPPEMENT ACTUEL DES RECHERCHES

Les progrès récents de la technologie des infrastructures de Grid, associées aux progrès de la recherche et de la conception des applications, permettront de résoudre des problèmes parallèles à grain fin, tels que la mécanique quantique et la dynamique moléculaire, qui étaient auparavant inaccessibles à l'environnement de Grid. Cela permettra de réaliser de nouvelles recherches scientifiques et d'accroître la flexibilité des ressources pour équilibrer les charges et programmer les charges de travail existantes.(147)

III.4 Conclusion

Avec le développement spectaculaire du calcul à haute performance, l'apparition de meilleurs algorithmes et l'accumulation de grandes quantités de données chimiques et biologiques, la technologie de conception de médicaments assistée par ordinateur joue un rôle de plus en plus important dans la découverte et le développement grâce à ses avantages : rapidité, faible coût et grande efficacité.

Ces dernières années, grâce au développement constant de la théorie de l'apprentissage automatique ML et de l'intelligence artificielle IA, une technologie puissante d'exploration des données a été largement utilisée à différents stades de la conception des médicaments.

Plus récemment, la conception de médicaments est entrée dans l'ère du big data, les méthodes ML ont progressivement évolué vers une méthode d'apprentissage profond DL avec une généralisation plus forte.

La DL, avec une plus grande capacité de généralisation et un traitement plus efficace des données volumineuses, favorise la combinaison de la technologie de l'IA avec celle de la conception de médicaments assistée par ordinateur, facilitant ainsi le processus.(138)

Conclusion générale et perspectives

CONCLUSION GENERALE ET PERSPECTIVES

Pour conclure notre travail, la modélisation moléculaire englobe toutes les méthodes théoriques et les techniques de calcul utilisées pour modéliser ou imiter le comportement des molécules. Ces techniques sont utilisées dans les domaines de la chimie-informatique, la conception de médicaments, la biologie computationnelle et la science des matériaux pour étudier des systèmes moléculaires, allant des petits composés chimiques aux grandes molécules biologiques jusqu'aux assemblages de matériaux.

Les calculs les plus simples peuvent être effectués à la main, mais inévitablement, des ordinateurs sont nécessaires pour effectuer la modélisation moléculaire de tout système de taille raisonnable. La caractéristique commune de ses techniques est la description, au niveau atomique, des systèmes moléculaires. (80)

L'augmentation inexorable de la puissance des ordinateurs, les nouvelles méthodes de calcul améliorées nous permettent d'intégrer la modélisation dans le processus de découverte de médicaments ce qui n'était pas possible il y a encore quelques dizaines années.

Depuis les trois dernières décennies, le rôle de la modélisation dans ce secteur s'est considérablement accru ; du coup, la découverte de médicaments s'est énormément développée. Il y a eu des progrès scientifiques remarquables sur des travaux réalisés dans les domaines expérimentaux et informatiques qui sont à la base de la conception moderne de médicaments assistée par ordinateur (Computer aided drug design CADD).

Globalement, les techniques modernes de la CADD pour la conception et le développement de médicaments sont basées sur la structure de cible (SBDD) et de ligand (LBDD). Elles constituent l'épine dorsale des processus modernes de CADD.

On a discuté plusieurs applications de ces techniques ainsi que leurs limites. Un concept clair et une connaissance avancée des méthodes de CADD amélioreront la qualité de la recherche et faciliteront l'identification de nouvelles entités chimiques, conduisant au développement de nouveaux médicaments.(149)

Outre l'amélioration des méthodes, les groupes universitaires et industriels ont acquis une expérience significative dans l'application de ces approches aux problèmes de découverte de médicaments. Les structures protéiques, le Docking, la recherche de pharmacophores, le QSAR...ect sont devenus des éléments essentiels dans le processus et sont, presque universellement, appliqués par les entreprises pharmaceutiques.

Sans oublier le rôle important de la biologie dans les premières phases de R&D pharmaceutiques, on suspecte une augmentation des travaux de modélisation biologique avec développement de nouvelles approches ou bien des logiciels plus performants afin de mieux simuler les systèmes réels.

Actuellement, l'industrie pharmaceutique a commencé à adopter, avec prudence, les technologies liées à l'IA, notamment diverses méthodes ML et DL. Une approche curieuse mais prudente est recommandée, étant donné la nature réglementée du développement des médicaments et le passage de la science traditionnelle à une science incluant et valorisant la contribution de L'IA. (135)

CONCLUSION GENERALE ET PERSPECTIVES

Dans un avenir proche, avec l'accumulation de données médicales et le développement avancé d'algorithmes d'IA, la technologie de l'intelligence artificielle devrait couvrir tous les domaines de la découverte et conception de nouveaux médicaments.

Couplée au suivi synchrone de l'automatisation, une plateforme intelligente de développement de médicaments, intégrant le big data (le modèle de prédiction de l'IA) et la synthèse automatique, est susceptible d'apparaître. Elle devrait modifier la situation actuelle caractérisée par un long cycle de développement du médicament, un coût élevé et un taux d'échec important.(138)

En outre, il y a peu de temps qu'on a assisté au développement d'autres classes de technologies, s'intéressant beaucoup plus au stockage des grandes bases de données moléculaires qui ne cessent d'augmenter encore et encore. Ces systèmes, appelés aussi des systèmes de parallélisation, sont principalement le « Cloud computing » et le « Grid computing ».

Des améliorations aussi performantes et élargies, en sens de capacité de stockage et d'analyse des données, pourraient être envisagées par collaboration avec des efforts informatiques plus intenses durant les années à venir.

Références bibliographiques

1. Gini G. QSAR Methods. In: Benfenati E, éditeur. In Silico Methods for Predicting Drug Toxicity [Internet]. New York, NY: Springer New York; 2016 [cité 19 sept 2021]. p. 1- 20. (Methods in Molecular Biology; vol. 1425). Disponible sur: http://link.springer.com/10.1007/978-1-4939-3609-0_1
2. Hercend C. Contribution de la modélisation moléculaire à l'étude de pathologies humaines: Application au transporteur ATP7B et au récepteur 5HT2B [PhD Thesis]. Université René Descartes-Paris V; 2012.
3. Kukol A. Molecular modeling of proteins. Vol. 443. Springer; 2008.
4. Cohen NC, Blaney JM, Humblet C, Gund P, Barry DC. Molecular modeling software and methods for medicinal chemistry. Journal of medicinal chemistry. 1990;33(3):883- 94.
5. Schlick T. Molecular modeling and simulation: an interdisciplinary guide. Vol. 2. Springer; 2010.
6. Engel T. Basic overview of chemoinformatics. Journal of chemical information and modeling. 2006;46(6):2267- 77.
7. Ramachandran KI, Deepa G, Namboori K. Computational chemistry and molecular modeling: principles and applications. Springer Science & Business Media; 2008.
8. Éditions Le Robert : la référence en langues pour définir, traduire, corriger et certifier - Dictionnaire Le Robert [Internet]. Le Robert. [cité 18 sept 2021]. Disponible sur: <https://www.lerobert.com/>
9. Caplat G. Modèles et métamodèles. PPUR presses polytechniques; 2008. 205 p.
10. Andrew RL. Molecular modeling principles and applications. Prentice Hall, London. 2001;
11. Aminpour M, Montemagno C, Tuszynski JA. An overview of molecular modeling for drug discovery with specific illustrative examples of applications. Molecules. 2019;24(9):1693.
12. Somer Jr FL. Molecular Modelling for Beginners (Alan Hinchliffe). ACS Publications; 2004.
13. Asses Y. Conception par modélisation et criblage in silico d'inhibiteurs du récepteur c-Met [PhD Thesis]. Nancy 1; 2011.
14. Futura. Simulation informatique [Internet]. Futura. [cité 18 sept 2021]. Disponible sur: <https://www.futura-sciences.com/tech/definitions/informatique-simulation-informatique-11319/>

REFERENCES BIBLIOGRAPHIQUES

15. Analyse conformationnelle : signification et origine de l'expression [Internet]. [cité 18 sept 2021]. Disponible sur: <https://www.linternaute.fr/expression/langue-francaise/14952/analyse-conformationnelle/>
16. Cohen NC. Guidebook on molecular modeling in drug design. Gulf Professional Publishing; 1996.
17. ARNAUD T. Modélisation Moléculaire [Internet]. France; 2020 [cité 18 sept 2021]. Disponible sur: <https://hal.archives-ouvertes.fr/hal-03017942>
18. Holtje H-D, Folkers G, Luzar A. Molecular Modeling, Basic Principles and Applications. American Institute of Physics; 1998.
19. Ealy JB. A student evaluation of molecular modeling in first year college chemistry. *Journal of Science Education and Technology*. 1999;8(4):309- 21.
20. Goodfellow JM. Computer modelling in molecular biology. John Wiley & Sons; 2008.
21. Vayer P, Arrault A, Lesur B, Bertrand M, Walther B. Apports de la chémo-informatique dans la recherche et l'optimisation des molécules d'intérêt thérapeutique. *médecine/sciences*. 2009;25(10):871- 7.
22. Le Guilloux V. Développement de méthodes et d'outils chémo-informatiques pour l'analyse et la comparaison de chimiothèques [PhD Thesis]. Université d'Orléans; 2013.
23. Xu J, Hagler A. Chemoinformatics and drug discovery. *Molecules*. 2002;7(8):566- 600.
24. Duelen R, Corvelyn M, Tortorella I, Leonardi L, Chai Y, Sampaolesi M. Medicinal Biotechnology for Disease Modeling, Clinical Therapy, and Drug Discovery and Development. In 2019. p. 89- 128.
25. Chtita S. Modélisation de molécules organiques hétérocycliques biologiquement actives par des méthodes QSAR/QSPR. Recherche de nouveaux médicaments [PhD Thesis]. Université Moulay Ismail, Meknès; 2017.
26. Wieber F. Racines scientifiques de la conception dite rationnelle de médicaments sur ordinateur: pour une histoire de la modélisation (bio) moléculaire. *Philosophia Scientiæ Travaux d'histoire et de philosophie des sciences*. 2008;(12- 2):93- 109.
27. Marshall GR, Oprea TI, Leach MMHAR, Green DV, Matter MRCLH, Gozalbes DHBMR, et al. Chemioinformatics in drug discovery. Wiley Online Library; 2004.
28. Leroux V. Modélisation d'inhibiteurs du domaine SH2 de la protéine Grb2 par dynamique moléculaire, docking et criblage virtuel [PhD Thesis]. Université Henri Poincaré-Nancy 1; 2006.
29. Louyot A. La gestion de projet en développement pharmaceutique: application pratique au développement industriel de principes actifs. :146.
30. Futura CP. Le cycle du médicament [Internet]. Futura. [cité 22 sept 2021]. Disponible sur: <https://www.futura-sciences.com/sante/dossiers/medecine-cycle-medicament-1125/>

REFERENCES BIBLIOGRAPHIQUES

31. Branciard A. Des modèles de recherche-développement ouverts et collaboratifs dans le domaine pharmaceutique: vers des «communs»? 2012;
32. Weinmann N. LA GLOBALISATION DES LEADERS PHARMACEUTIQUES. :191.
33. Martin E. Recherche pharmaceutique, prix du médicament et assurance-maladie. Revue française d'économie. 1996;11(1):49- 86.
34. Cozza G. The development of CK2 inhibitors: From traditional pharmacology to in silico rational drug design. *Pharmaceuticals*. 2017;10(1):26.
35. Hassan Baig M, Ahmad K, Roy S, Mohammad Ashraf J, Adil M, Haris Siddiqui M, et al. Computer aided drug design: success and limitations. *Current pharmaceutical design*. 2016;22(5):572- 81.
36. Ferenczy GG. Application of Molecular Modeling in Analogue-Based Drug Discovery. *Analogue-Based Drug Discovery II*. 2010;61 - 82.
37. Schwede T. Protein modeling: what happened to the “protein structure gap”? *Structure*. 2013;21(9):1531- 40.
38. Salzberg SL, Searls DB, Kasif S. *Computational Methods in Molecular Biology*. Elsevier; 1998. 399 p.
39. Hardy S. Méthodes systémiques d'analyse des données de simulation de modèles de voies de signalisation cellulaire. ProQuest; 2007.
40. Kumar N, Hendriks BS, Janes KA, de Graaf D, Lauffenburger DA. Applying computational modeling to drug discovery and development. *Drug discovery today*. 2006;11(17- 18):806- 11.
41. Andrieux G. Modélisation dynamique de la signalisation cellulaire: aspects différentiels et discrets; application à la signalisation du facteur de croissance TGF-beta dans le cancer. :225.
42. Bezzo F, Galvanin F. On the identifiability of physiological models: optimal design of clinical tests. In: *Computer Aided Chemical Engineering*. Elsevier; 2018. p. 85- 110.
43. Pruet WA, Clemmer JS, Hester RL. Physiological Modeling and Simulation—Validation, Credibility, and Application. *Annual Review of Biomedical Engineering*. 2020;22:185- 206.
44. SALAH T. Etude par la chimie computationnelle des corrélations 2D-QSAR et 3D-QSAR de quelques composés bioactifs. [PhD Thesis]. Université Mohamed Khider-Biskra; 2017.
45. Merz Jr KM, Ringe D, Reynolds CH. *Drug design: structure-and ligand-based approaches*. Cambridge University Press; 2010.
46. Poso A. *Chemoinformatics for Drug Discovery*. Edited by Jürgen Bajorath. Wiley Online Library; 2014.

REFERENCES BIBLIOGRAPHIQUES

47. Kuntz ID. Structure-based strategies for drug design and discovery. *Science*. 1992;257(5073):1078- 82.
48. Kalyaanamoorthy S, Chen Y-PP. Structure-based drug design to augment hit discovery. *Drug discovery today*. 2011;16(17- 18):831- 9.
49. Amzel LM. Structure-based drug design. *Current opinion in biotechnology*. 1998;9(4):366- 9.
50. Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T. Protein structure homology modeling using SWISS-MODEL workspace. *Nature protocols*. 2009;4(1):1- 13.
51. Traore MDM. Synthèse et études de modélisation moléculaire dans l'optimisation de la sélectivité de nouveaux agents antiparasitaires inspirés de produits naturels [PhD Thesis]. Université Grenoble Alpes; 2016.
52. Klebe G. Recent developments in structure-based drug design. *Journal of Molecular Medicine*. 2000;78(5):269- 81.
53. Whittle PJ, Blundell TL. Protein structure-based drug design. *Annual review of biophysics and biomolecular structure*. 1994;23(1):349- 75.
54. Verlinde CL, Hol WG. Structure-based drug design: progress, results and challenges. *Structure*. 1994;2(7):577- 87.
55. Wlodawer A, Vondrasek J. Inhibitors of HIV-1 protease: A major success of structure-assisted drug design. *Annual review of biophysics and biomolecular structure*. 1 févr 1998;27:249- 84.
56. Anderson AC. The process of structure-based drug design. *Chemistry & biology*. 2003;10(9):787- 97.
57. Batool M, Ahmad B, Choi S. A structure-based drug discovery paradigm. *International journal of molecular sciences*. 2019;20(11):2783.
58. Muhammed MT, Aki-Yalcin E. Homology modeling in drug discovery: Overview, current applications, and future perspectives. *Chemical biology & drug design*. 2019;93(1):12- 20.
59. Floudas CA, Fung HK, McAllister SR, Mönnigmann M, Rajgaria R. Advances in protein structure prediction and de novo protein design: A review. *Chemical Engineering Science*. 2006;61(3):966- 88.
60. Webb B, Sali A. Comparative protein structure modeling using MODELLER. *Current protocols in bioinformatics*. 2016;54(1):5- 6.
61. Lee J, Freddolino PL, Zhang Y. Ab initio protein structure prediction. In: *From protein structure to function with bioinformatics*. Springer; 2017. p. 3- 35.

REFERENCES BIBLIOGRAPHIQUES

62. Bonneau R, Strauss CE, Rohl CA, Chivian D, Bradley P, Malmström L, et al. De novo prediction of three-dimensional structures for major protein families. *Journal of molecular biology*. 2002;322(1):65- 78.
63. Tomar V, Mazumder M, Chandra R, Yang J, Sakharkar MK. Small Molecule Drug Design. In: *Encyclopedia of Bioinformatics and Computational Biology* [Internet]. Elsevier; 2019 [cité 22 sept 2021]. p. 741- 60. Disponible sur: <https://linkinghub.elsevier.com/retrieve/pii/B978012809633820157X>
64. Fortuné A. Techniques de Modélisation Moléculaire appliquées à l'Etude et à l'Optimisation de Molécules Immunogènes et de Modulateurs de la Chimiorésistance. [PhD Thesis]. Université Joseph-Fourier-Grenoble I; 2006.
65. Joubert F, Tastan Bishop AO, De Beer TA. Protein homology modelling and its use in South Africa: research in action. *South African Journal of Science*. 2008;104(1):2- 6.
66. Lambert C. Développement d'une méthode automatique fiable de modélisation de la structure tridimensionnelle des protéines par homologie et application au protéome de *Brucella melitensis*. département de biologie. FUNDP, Namur. 2003;
67. Tramontano A. Homology modeling with low sequence identity. *Methods*. 1998;14(3):293- 300.
68. Krieger E, Nabuurs SB, Vriend G. Homology modeling. *Methods of biochemical analysis*. 2003;44:509- 24.
69. Eswar N, John B, Mirkovic N, Fiser A, Ilyin VA, Pieper U, et al. Tools for comparative protein structure modeling and analysis. *Nucleic acids research*. 2003;31(13):3375- 80.
70. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. *Nucleic acids research*. 2004;32(suppl_2):W526- 31.
71. Cavasotto CN, Phatak SS. Homology modeling in drug discovery: current trends and applications. *Drug discovery today*. 2009;14(13- 14):676- 83.
72. Chow KYC, Hon CC, Hui RKH, Wong RTY, Yip CW, Zeng F, et al. Molecular Advances in Severe Acute Respiratory Syndrome-associated Coronavirus (SARS-CoV). *Genomics, Proteomics & Bioinformatics*. nov 2003;1(4):247- 62.
73. Allain F. Calcul efficace de la structure des protéines à partir de contacts évolutifs [PhD Thesis]. Université Pierre et Marie Curie-Paris VI; 2017.
74. Bornot A. Analyse et prédiction de la relation séquence-structure locale et flexibilité au sein des protéines globulaires [PhD Thesis]. Université Paris-Diderot-Paris VII; 2009.
75. Dorn M, e Silva MB, Buriol LS, Lamb LC. Three-dimensional protein structure prediction: Methods and computational strategies. *Computational biology and chemistry*. 2014;53:251- 76.
76. biochemistry - Why is ab initio protein secondary structure prediction less reliable than alternatives? [Internet]. *Biology Stack Exchange*. [cité 22 sept 2021]. Disponible sur:

REFERENCES BIBLIOGRAPHIQUES

- <https://biology.stackexchange.com/questions/43315/why-is-ab-initio-protein-secondary-structure-prediction-less-reliable-than-alter>
77. Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and structure-based drug design strategies. *Molecules*. 22 juill 2015;20(7):13384- 421.
 78. Salmaso V, Moro S. Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: an overview. *Frontiers in pharmacology*. 2018;9:923.
 79. Fan J, Fu A, Zhang L. Progress in molecular docking. *Quantitative Biology*. 2019;1- 7.
 80. Eweas AF, Maghrabi IA, Namarneh AI. Advances in molecular modeling and docking as a tool for modern drug discovery. *Der Pharma Chemica*. 2014;6(6):211 - 28.
 81. Fan J, Fu A, Zhang L. Progress in molecular docking. *Quant Biol*. 1 juin 2019;7(2):83- 9.
 82. AYACHI H. Compréhension du mécanisme enzyme-substrat par modélisation moléculaire [PhD Thesis].2017 Université de Tlemcen-Abou Bekr Belkaid;
 83. Torres PHM, Sodero ACR, Jofily P, Silva-Jr FP. Key Topics in Molecular Docking for Drug Design. *Int J Mol Sci*. 15 sept 2019;20(18):E4574.
 84. Kroemer RT. Structure-based drug design: docking and scoring. *Current protein and peptide science*. 2007;8(4):312- 28.
 85. Wang ZJ. Application of molecular modeling to drug discovery and functional genomics [PhD Thesis]. Northeastern University; 2012.
 86. Leelananda SP, Lindert S. Computational methods in drug discovery. *Beilstein J Org Chem*. 12 déc 2016;12:2694- 718.
 87. Gschwend DA, Good AC, Kuntz ID. Molecular docking towards drug discovery. *Journal of Molecular Recognition: An Interdisciplinary Journal*. 1996;9(2):175- 86.
 88. Yuriev E, Ramsland PA. Latest developments in molecular docking: 2010–2011 in review. *Journal of Molecular Recognition*. 2013;26(5):215- 39.
 89. Lee A, Lee K, Kim D. Using reverse docking for target identification and its applications for drug discovery. *Expert Opinion on Drug Discovery*. 2 juill 2016;11(7):707- 15.
 90. Chen YZ, Zhi DG. Ligand–protein inverse docking and its potential use in the computer search of protein targets of a small molecule. *Proteins: Structure, Function, and Bioinformatics*. 2001;43(2):217- 26.
 91. Pinzi L, Rastelli G. Molecular Docking: Shifting Paradigms in Drug Discovery. *Int J Mol Sci*. 4 sept 2019;20(18):E4331.
 92. Computational methods for drug design and discovery: focus on China: Trends in Pharmacological Sciences [Internet]. [cité 3 oct 2021]. Disponible sur: [https://www.cell.com/trends/pharmacological-sciences/fulltext/S0165-6147\(13\)00153-3](https://www.cell.com/trends/pharmacological-sciences/fulltext/S0165-6147(13)00153-3)

REFERENCES BIBLIOGRAPHIQUES

93. Xu X, Huang M, Zou X. Docking-based inverse virtual screening: methods, applications, and challenges. *Biophys Rep.* févr 2018;4(1):1- 16.
94. Do P-C, Lee EH, Le L. Steered Molecular Dynamics Simulation in Rational Drug Design. *J Chem Inf Model.* 27 août 2018;58(8):1473- 82.
95. Naqvi AA, Mohammad T, Hasan GM, Hassan M. Advancements in docking and molecular dynamics simulations towards ligand-receptor interactions and structure-function relationships. *Current topics in medicinal chemistry.* 2018;18(20):1755- 68.
96. Hollingsworth SA, Dror RO. Molecular Dynamics Simulation for All. *Neuron.* 19 sept 2018;99(6):1129- 43.
97. Karplus M, Petsko GA. Molecular dynamics simulations in biology. *Nature.* 1990;347(6294):631- 9.
98. Liu X, Shi D, Zhou S, Liu H, Liu H, Yao X. Molecular dynamics simulations and novel drug discovery. *Expert Opin Drug Discov.* janv 2018;13(1):23- 37.
99. Alonso H, Bliznyuk AA, Gready JE. Combining docking and molecular dynamic simulations in drug design. *Medicinal research reviews.* 2006;26(5):531- 68.
100. Ghemtio L. Simulation numérique et approche orientée connaissance pour la découverte de nouvelles molécules thérapeutiques. :163.
101. Kalyaanamoorthy S, Chen Y-PP. Modelling and enhanced molecular dynamics to steer structure-based drug discovery. *Progress in biophysics and molecular biology.* 2014;114(3):123- 36.
102. Huang H-J, Yu HW, Chen C-Y, Hsu C-H, Chen H-Y, Lee K-J, et al. Current developments of computer-aided drug design. *Journal of the Taiwan Institute of Chemical Engineers.* 2010;41(6):623- 35.
103. Shim J, Mackerell AD. Computational ligand-based rational design: Role of conformational sampling and force fields in model development. *Medchemcomm.* mai 2011;2(5):356- 70.
104. Veselovsky A, Ivanov A. Strategy of Computer-Aided Drug Design. *Current drug targets Infectious disorders.* 1 avr 2003;3:33- 40.
105. Kutlushina A, Khakimova A, Madzhidov T, Polishchuk P. Ligand-Based Pharmacophore Modeling Using Novel 3D Pharmacophore Signatures. *Molecules.* 27 nov 2018;23(12):3094.
106. Usha T, Shanmugarajan D, Goyal AK, Kumar CS, Middha SK. Recent Updates on Computer-aided Drug Discovery: Time for a Paradigm Shift. *CTMC.* 9 févr 2018;17(30):3296- 307.
107. Roy K. *Advances in QSAR modeling. Applications in Pharmaceutical, Chemical, Food, Agricultural and Environmental Sciences*; Springer: Cham, Switzerland. 2017;555:39.

REFERENCES BIBLIOGRAPHIQUES

108. Lee C-H, Huang H-C, Juan H-F. Reviewing ligand-based rational drug design: the search for an ATP synthase inhibitor. *International journal of molecular sciences*. 2011;12(8):5304- 18.
109. Aparoy P, Kumar Reddy K, Reddanna P. Structure and ligand based drug design strategies in the development of novel 5-LOX inhibitors. *Current medicinal chemistry*. 2012;19(22):3763- 78.
110. Hmamouchi R, Bouachrine M, Lakhlifi T. Tentative Pratique du Relation Quantitatives Structure-Activité/Propriété (QSAR/QSPR). *Revue Interdisciplinaire*. 2016;1(1).
111. Gini G. QSAR: What Else? *Methods Mol Biol*. 2018;1800:79- 105.
112. Dudek AZ, Arodz T, Gálvez J. Computational methods in developing quantitative structure-activity relationships (QSAR): a review. *Combinatorial chemistry & high throughput screening*. 2006;9(3):213- 28.
113. Grisoni F, Ballabio D, Todeschini R, Consonni V. Molecular Descriptors for Structure–Activity Applications: A Hands-On Approach. In: Nicolotti O, éditeur. *Computational Toxicology [Internet]*. New York, NY: Springer New York; 2018 [cité 7 oct 2021]. p. 3- 53. (Methods in Molecular Biology; vol. 1800). Disponible sur: http://link.springer.com/10.1007/978-1-4939-7899-1_1
114. Mauri A, Consonni V, Todeschini R. Molecular Descriptors. In: Leszczynski J, Kaczmarek-Kedziera A, Puzyn T, G. Papadopoulos M, Reis H, K. Shukla M, éditeurs. *Handbook of Computational Chemistry [Internet]*. Cham: Springer International Publishing; 2017 [cité 8 oct 2021]. p. 2065- 93. Disponible sur: http://link.springer.com/10.1007/978-3-319-27282-5_51
115. Bacilieri M, Moro S. Ligand-based drug design methodologies in drug discovery process: an overview. *Current drug discovery technologies*. 2006;3(3):155- 65.
116. Danishuddin, Khan AU. Descriptors and their selection methods in QSAR analysis: paradigm for drug design. *Drug Discovery Today*. août 2016;21(8):1291- 302.
117. Acharya C, Coop A, E Polli J, D MacKerell A. Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. *Current computer-aided drug design*. 2011;7(1):10- 22.
118. Verma J, Khedkar VM, Coutinho EC. 3D-QSAR in drug design--a review. *Curr Top Med Chem*. 2010;10(1):95- 115.
119. Damale MG, Harke SN, Kalam Khan FA, Shinde DB, Sangshetti JN. Recent advances in multidimensional QSAR (4D-6D): a critical review. *Mini Rev Med Chem*. janv 2014;14(1):35- 55.
120. RAGHANI A. *Drug Design*. 2018;
121. Timmerman H. *QSAR and drug design: new developments and applications*. Elsevier; 1995.

REFERENCES BIBLIOGRAPHIQUES

122. Macalino SJY, Gosu V, Hong S, Choi S. Role of computer-aided drug design in modern drug discovery. *Archives of pharmacal research*. 2015;38(9):1686- 701.
123. Lu X, Yang H, Chen Y, Li Q, He S, Jiang X, et al. The development of pharmacophore modeling: Generation and recent applications in drug discovery. *Current pharmaceutical design*. 2018;24(29):3424- 39.
124. Güner OF. *Pharmacophore Perception. Development and Use in Drug Design*, International University Line, La Jolla, CA, USA. 2000;
125. Khedkar SA, Malde AK, Coutinho EC, Srivastava S. Pharmacophore modeling in drug discovery and development: an overview. *Medicinal Chemistry*. 2007;3(2):187- 97.
126. Schwab CH. Conformations and 3D pharmacophore searching. *Drug Discovery Today: Technologies*. 2010;7(4):e245- 53.
127. Yang S-Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug discovery today*. 2010;15(11- 12):444- 50.
128. Bielska E, Lucas X, Czerwoniec A, Kasprzak JM, Kaminska KH, Bujnicki JM. Virtual screening strategies in drug design—methods and applications. *BioTechnologia Journal of Biotechnology Computational Biology and Bionanotechnology*. 2011;92(3).
129. Qing X, Lee XY, De Raeymaecker J, Tame JR, Zhang KY, De Maeyer M, et al. Pharmacophore modeling: advances, limitations, and current utility in drug discovery. *Journal of Receptor, Ligand and Channel Research*. 2014;7:81- 92.
130. Kaserer T, Beck KR, Akram M, Odermatt A, Schuster D. Pharmacophore models and pharmacophore-based virtual screening: concepts and applications exemplified on hydroxysteroid dehydrogenases. *Molecules*. 2015;20(12):22799- 832.
131. Lima AN, Philot EA, Trossini GHG, Scott LPB, Maltarollo VG, Honorio KM. Use of machine learning approaches for novel drug discovery. *Expert opinion on drug discovery*. 2016;11(3):225- 39.
132. Jordan AM. *Artificial intelligence in drug design—the storm before the calm?* ACS Publications; 2018.
133. Bini SA. Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care? *The Journal of Arthroplasty*. août 2018;33(8):2358- 61.
134. Duch W, Swaminathan K, Meller J. Artificial intelligence approaches for rational drug design and discovery. *Current pharmaceutical design*. 2007;13(14):1497- 508.
135. Schneider P, Walters WP, Plowright AT, Sieroka N, Listgarten J, Goodnow RA, et al. Rethinking drug design in the artificial intelligence era. *Nature Reviews Drug Discovery*. 2020;19(5):353- 64.
136. Brown N, Ertl P, Lewis R, Luksch T, Reker D, Schneider N. *Artificial intelligence in chemistry and drug design*. Springer; 2020.

REFERENCES BIBLIOGRAPHIQUES

137. Lin X, Li X, Lin X. A Review on Applications of Computational Methods in Drug Screening and Design. *Molecules*. 18 mars 2020;25(6):E1375.
138. Wang L, Ding J, Pan L, Cao D, Jiang H, Ding X. Artificial intelligence facilitates drug design in the big data era. *Chemometrics and Intelligent Laboratory Systems*. 2019;194:103850.
139. Haghightlari M, Hachmann J. Advances of machine learning in molecular modeling and simulation. *Current Opinion in Chemical Engineering*. 2019;23:51- 7.
140. LeCun Y, Bengio Y, Hinton G. Deep learning. *nature*. 2015;521(7553):436- 44.
141. Ballester PJ. Machine Learning for Molecular Modelling in Drug Design. *Biomolecules*. 4 juin 2019;9(6):E216.
142. Stephenson N, Shane E, Chase J, Rowland J, Ries D, Justice N, et al. Survey of Machine Learning Techniques in Drug Discovery. *Curr Drug Metab*. 2019;20(3):185- 93.
143. Korb O, Finn PW, Jones G. The cloud and other new computational methods to improve molecular modelling. *Expert Opinion on Drug Discovery*. oct 2014;9(10):1121- 31.
144. Sobeslav V, Maresova P, Krejcar O, Franca TCC, Kuca K. Use of cloud computing in biomedicine. *Journal of Biomolecular Structure and Dynamics*. 17 févr 2016;1- 10.
145. Moghadam BT, Alvarsson J, Holm M, Eklund M, Carlsson L, Spjuth O. Scaling predictive modeling in drug development with cloud computing. *J Chem Inf Model*. 26 janv 2015;55(1):19- 25.
146. Ebejer J-P, Fulle S, Morris GM, Finn PW. The emerging role of cloud computing in molecular modelling. *Journal of Molecular Graphics and Modelling*. juill 2013;44:177- 87.
147. Koehler M, Ruckebauer M, Janciak I, Benkner S, Lischka H, Gansterer WN. A grid services cloud for molecular modelling workflows. *IJWGS*. 2010;6(2):176.
148. Claus BL, Johnson SR. Grid computing in large pharmaceutical molecular modeling. *Drug Discovery Today*. juill 2008;13(13- 14):578- 83.
149. Hassan Baig M, Ahmad K, Roy S, Mohammad Ashraf J, Adil M, Haris Siddiqui M, et al. Computer Aided Drug Design: Success and Limitations. *CPD*. 26 janv 2016;22(5):572- 81.

Résumé

Le monde de la chimie, aussi vaste et compliqué, nécessite des efforts immenses pour essayer de comprendre ses phénomènes. Parmi ces derniers, la modélisation moléculaire est élaborée depuis l'arrivée de l'ère de l'informatique.

Cette discipline, étant en pleine émergence, trouve sa place dans de nombreux domaines, plus particulièrement, le domaine pharmaceutique. Actuellement, elle est devenue assistée par ordinateur, c'est le « Computer Aided Drug Design » (CADD) donnant naissance à deux grandes catégories de stratégies, en fonction de la disponibilité de structure 3D de la protéine cible, celle basée sur la cible (SBDD) et l'autre basée sur le ligand (LBDD).

La SBDD nécessite l'existence préalable de la structure 3D de la cible élaborée avec modélisation par homologie ou bien prédiction « de novo ». Cette classe est caractérisée par le Docking précédé par la dynamique moléculaire (DM) du complexe protéine-ligand.

La LBDD, basant sur l'interaction entre le ligand et la cible et ne requière pas la structure 3D de la cible, les méthodes utilisées majoritairement le QSAR et les modèles pharmacophores.

Récemment, suite à une modernisation très intéressante en termes de puissance de logiciel de calcul et de stockage, l'intelligence artificielle (IA) s'intègre dans ce processus. Ce domaine connaît une évolution remarquable et se révolutionne encore et encore dans les années à venir.

Mots clés : Modélisation moléculaire, CADD, SBDD, LBDD, Docking, QSAR

Abstract

The world of chemistry, so vast and complicated, requires a great deal of effort to understand its phenomena. Among these efforts, Molecular modelling is developed since the emergence of the computer.

This discipline, being in full emergence, finds its place in a wide range of fields, most notably, the pharmaceutical field. Nowadays, it has become Computer Aided Drug Design (CADD) that is giving birth to two main categories of strategies, depending on the availability of the 3D structure of the target protein, target-based drug design (SBDD) and ligand-based drug design (LBDD).

The SBDD needs the prior existence of the 3D structure of the target developed with homology modelling or "de novo" prediction, and it is characterised by Docking preceded by the Molecular Dynamics (MD) of the protein-ligand complex.

The LBDD, based on the interaction between the ligand and the target without requiring the 3D structure of the target, the methods used are mainly QSAR and pharmacophore models.

Currently, due to a very interesting update in terms of computational and storage software power, the artificial intelligence (AI) fits into this process. This field is undergoing a remarkable revolution so that it would be much more improved in the future.

Key words: Molecular modelling, CADD, SBDD, LBDD, Docking, QSAR.

ملخص

عالم الكيمياء ، واسع ومعقد حيث يحتاج الى مجهودات عظيمة في المجال العلمي لفهم مجموع ظواهره ، من بين هذه الأخيرة ، النمذجة الجزيئية التي تم تطويرها منذ وصول عهد الكمبيوتر.

هذا المجال ، كونه في تطور مستمر ، نجح في إيجاد مكانة معتبرة بين مجالات عديدة ، من بينها المجال الصيدلاني ، الذي في الوقت الحالي ، أصبح مُدارًا بالكمبيوتر ، تحت مسمى "تصميم الأدوية المدار بالكمبيوتر" مما أدى إلى بروز فئتان أساسيتان من الإستراتيجيات ، وفقا لإتاحة البنية الثلاثية الأبعاد للبروتين المستهدف ، إحداهما تصميم الأدوية على أساس هيكل هذا البروتين ، والأخرى على أساس المكمّل.

إن تصميم الأدوية على أساس بنية البروتين المستهدف يتطلب وجود بنية ثلاثية الأبعاد له ، و التي يتم تكوينها مسبقا بواسطة النمذجة بالتمائل أو النمذجة من البداية. تتمثل هذه الفئة في التنسيق المسبوق بالديناميكيات الجزيئية.

في حين أن تصميم الأدوية على أساس المكمّل يتركز على التفاعل بين هذا الأخير و بين البروتين المستهدف حيث لا يحتاج إلى الهيكل الثلاثي الأبعاد للبروتين. أما عن الطرق المستعملة لتنفيذ هذه العملية ، فهي في أغلب الأحيان دراسة العلاقة الكمية بين الهيكل و النشاط بالإضافة إلى النموذج الفارماكوفوري.

في الأونة الأخيرة و بعد مجموعة من التحديثات الملحوظة فيما يخص قوة برمجيات الحساب و التخزين الآلي، شهد الحيز العلمي دخول عنصر هام ألا و هو مجال الذكاء الاصطناعي. إن مجال تصميم الأدوية يشهد تطورا استثنائيا مما يشير إلى بلوغه أفق التقدم في السنوات القليلة القادمة.

الكلمات المفتاحية : النمذجة الجزيئية ، تصميم الأدوية المدار بالكمبيوتر ، تصميم الأدوية على أساس هيكل البروتين المستهدف ، تصميم الأدوية على أساس المكمّل. العلاقة الكمية بين الهيكل و النشاط.