

# Chapter 1

## La Bioinformatique



## 1.1 Introduction

La bioinformatique est une "interdiscipline" à la frontière de la biologie, de l'informatique et des mathématiques. Les systèmes biologiques sont très complexes et les techniques modernes d'investigation du monde biologique fournissent une vaste quantité de données expérimentales. Le but ultime de la bioinformatique est d'intégrer ces données d'origines très diverses pour modéliser les systèmes vivants afin de comprendre et prédire leurs comportements (biologie systémique ou biologie des systèmes) dans des conditions de fonctionnement normales ou pathologiques. La bioinformatique est donc étroitement couplée à ses applications. Bon nombre de bioinformaticiens ne travaillent pas dans des laboratoires formellement estampillés [?] .

Aujourd'hui , tout projet de biologie comporte une étape d'analyse bioinformatique des données. Par conséquent, un biologiste passe environ 20-30/100 de son temps à utiliser des outils bioinformatiques.

Ce chapitre décrit de manière simple les tâches courantes de la bioinformatique qu'un biologiste/biochimiste doit savoir traiter par lui-même, à partir des progrès de l'intelligence informatique pour la bioinformatique qui nous avons utiliser pour notre contexte d'étude qui représente l'application de l'apprentissage profond pour la prédiction de l'activité biologie [?] .

## 1.2 Histoire du terme bioinformatique

Le terme de bioinformatique date du début des années 80. Cependant, le concept sous-jacent de traitement de l'information biologique est bien plus vieux. Durant les années 60, la biologie moléculaire a eu besoin de modélisation formelle, ce qui a mené à la création des bioinformatique [?].

## 1.3 Pour quoi la bioinformatique

La bioinformatique est l'étude de l'information biologique. Ce n'est pas simplement l'application à la biologie de l'informatique, c'est une branche à part entière de la biologie. La bioinformatique actuelle se concentre surtout sur l'étude des séquences d'ADN et sur le repliement des protéines, donc travaille surtout au niveau moléculaire [?] .

## 1.4 Qu'est-ce que la bioinformatique

### 1.4.1 Définition

La bioinformatique est l'étude de l'information biologique de façon automatique, alors elle inclut la création et le développement de technologies informatiques qui permettant le traitement et l'analyse de ces information. l'intégration des méthodes mathématiques ,statistiques ,informatiques pour analysées des données biologiques ,biochimiques et biophysiques [?] .



Figure 1.1: premier image

La bioinformatique se réfère spécifiquement à la recherche et à l'utilisation

de patterns et structures dans les données biologiques et au développement de nouvelles méthodes pour accéder aux bases de données [?] .

## 1.5 Apports à la biologie

L'informatique est devenue un apport fondamentale à la biologie moléculaire. Les moyens informatiques sont naturellement utilisés pour le stockage ou la gestion des données mais également pour l'interprétation de ces données. Le traitement informatique des séquences peut par exemple déterminer la fonction biologique d'un gène. Cet apport informatique concerne principalement quatre aspects :

1. Le premier est l'organisation des données avec essentiellement la création de bases de données afin de réunir le plus d'information possible sur les séquences.
2. Le deuxième aspect concerne les traitements que l'on peut effectuer sur les séquences afin de repérer un élément biologique intéressant. Ces programmes représentent les traitements couramment utilisés dans l'analyse des séquences comme la recherche des similitudes d'une séquence avec l'ensemble d'une base de données.
3. Le troisième aspect est celui qui permet d'élaborer des stratégies pour apporter des connaissances biologiques supplémentaires que l'on pourra ensuite intégrer dans des traitements standards. Par exemple la mise au point de nouvelles matrices de substitution des acides aminés,...etc
4. Enfin, le quatrième aspect est celui de l'évaluation des différentes approches citées précédemment dans le but de valider.

## 1.6 Domaine d'application

De plus en plus, la bioinformatique est développée dans un but d'application à l'agriculture, la pharmacologie, la médecine [?].

### 1.6.1 Structures moléculaires

Visualisation, analyse, classification, prédiction.

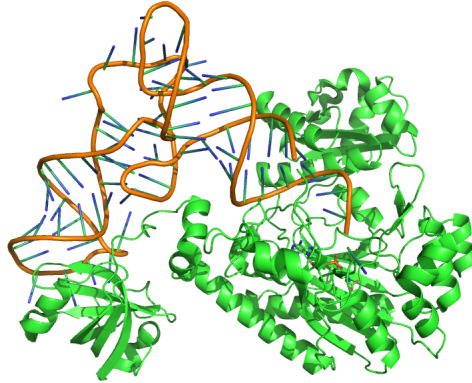


Figure 1.2: La structure moléculaire

### 1.6.2 Analyse de séquences

Alignements, recherches de similarités, détection de motifs.

### 1.6.3 La recherche pharmaceutique

1. Mécanismes d'action des médicaments.
2. Identification de cibles pharmaceutiques.

### 1.6.4 Génomique

Annotation des génomes, génomique comparative.

### 1.6.5 Analyse des réseaux biomoléculaires

Réseaux métaboliques, d'interactions protéiques, de régulation génétique.

## 1.7 Quantitative et le système Pharmacologie (QSP)

La pharmacologie quantitative (QSP) est une nouvelle discipline sur l'identification et la validation de cibles de médicaments, comprendre les thérapeutiques existantes. Le but de QSP est de comprendre la manière précise et prédictive aussi comment les médicaments modulent les réseaux cellulaires dans l'espace et dans le temps et leur impact sur la physiopathologie humaine [?].

Au cours des trois dernières décennies, le paradigme dominant dans la découverte de médicaments a été la conception de ligands sélectifs pour une cible spécifique afin d'éviter les effets indésirables. Cependant, dans l'ère postgénomique actuelle, l'objectif est de concevoir des médicaments à partir des modèles de relations quantitatives structure-activité **QSAR** pour prédire la réponse biologique quantifiée d'une molécule à partir de ses descripteurs [?] .

## 1.8 Progrès de l'intelligence informatique pour la bioinformatique

La bioinformatique est en train de devenir de plus en plus une science basée sur les données, une évolution tirée par les avancées technologiques en acquisition de données . Dans le domaine biomédical la grande quantité de données générées des scénarios difficiles pour les chercheurs, Les nouvelles exigences en matière de données requièrent de nouvelles approches en matière d'analyse des données. Certaines des plus intéressantes proviennent actuellement des domaines de l'intelligence computationnelle (IC) et de l'apprentissage automatique (ML) [?].

Cette session s'intéresse particulièrement à la proposition d'approches de l'apprentissage automatique ML pour résoudre les problèmes dans les domaines biomédical et bioinformatique. Les sujets qui intéressent cette session :

1. **Application de l'apprentissage profond pour la prédiction de l'activité biologique.**
2. Nouvelles applications des méthodes existantes d'apprentissage automatique ML à la bioinformatique.

3. Nouvelles techniques de d'apprentissage automatique ML pour la biomédecine et la bioinformatique.

## 1.9 l'apprentissage profond et la modélisation Qsar

### 1.9.1 QSAR

Relations entre l'industrie pharmaceutique et les modèles de relations quantitatives structure-activité (QSAR) pour prédire la réponse biologique quantifiée d'une molécule à partir de ses descripteurs, qui sont essentiellement des études étudiées de la molécule.

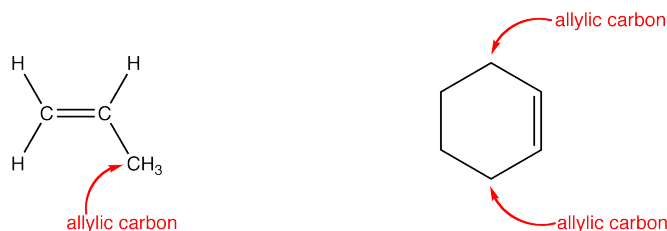


Figure 1.3: Allylic carbon

La complexité de ces descripteurs varie et peut aller de simples mesures de poids moléculaire à des caractéristiques géométriques complexes. La découverte de médicaments est un processus long et coûteux pour le secteur pharmaceutique.

### 1.9.2 L'apprentissage en profondeur

L'apprentissage en profondeur (deep learning) est un ensemble de méthodes d'apprentissage automatique et Les algorithmes d'apprentissage automatique, permettent aux ordinateurs de s'entraîner sur les entrées de données et utilisent l'analyse statistique .

fondées sur l'apprentissage de modèles de données, Une observation (une image, p. ex.) peut être représentée de différentes façons par un vecteur de données. Certaines représentations et une bonne capacité d'analyse automatique des différenciations rendent la tâche d'apprentissage plus efficace [?].



### 1.9.3 La prédiction de l'activité biologique

Le principe de la modélisation **Qsar**, consiste à trouver une relation reliant quantitativement une **activité biologique** mesurée expérimentalement. Ceci, est réalisé pour une série de composés similaire avec des descripteurs moléculaires à l'aide des méthodes computationnelles.

Les données des molécules biologiques disponibles dans les benchmarks sont tellement volumineuses (Big Data), qu'ils en deviennent difficiles à l'analyser avec des techniques de traitement de données classiques. Afin d'arriver à analyser ce type de données, nous faisons appel à des algorithmes de **Deep Learning**, qui sont en constante augmentation dans plusieurs domaines [?].

## 1.10 Utilisation de l'amarrage (Docking)

Dans le domaine de la modélisation moléculaire, l'amarrage est une méthode qui calcule l'orientation préférée d'une molécule vers une seconde lorsqu'elles sont liées pour former un complexe stable. Savoir l'orientation préférée sert à prévoir la solidité de l'union entre deux molécules [?] .

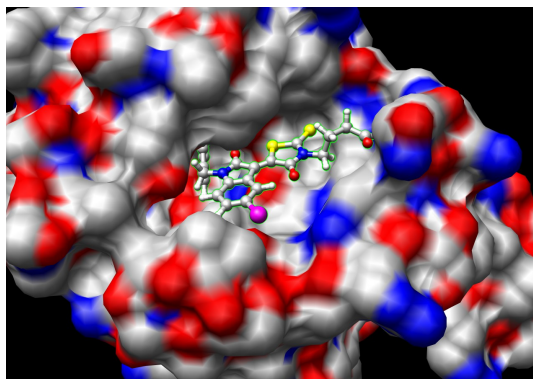


Figure 1.4: Docking

Les associations entre des molécules d'importance biologique, telles que les protéines, les acides nucléiques, les glucides et matières grasses jouent un rôle essentiel dans la transduction de signal. D'ailleurs, l'orientation relative des deux molécules associées peuvent avoir un effet sur le genre du signal produit (ex. antagoniste contre l'agoniste). Par conséquent, des études

d'amarrage sont utiles à calculer la force et le genre du signal produit [?] . Dans la mise au point de nouveaux médicaments, l'amarrage sert souvent à déterminer l'orientation de petites molécules liées à leurs protéines ciblées afin de calculer leurs affinité et niveau d'activité.

## 1.11 Conclusion

Dans ce chapitre nous avons défini les notions et l'évolution de la **bioinformatique** d'après le concept traitement de l'information biologique. les caractéristiques aussi les domaine d'application de la bioinformatique ,vers le problème biologique a traité sous domaine de la pharmacologie, d'objectif la prédiction de l'activité biologique pour concevoir des médicament à partir le modèle quantitative QSAR avec les moyennes informatique qui représente technique de l'apprentissage automatique (ML) qui fais appel à des algorithmes de Deep learning.

Dans le prochain chapitre nous allons exprimé notre démarche d'étude, aussi les outils de développement utilisés tel que le langage de programmation, sous chapitre de contribution.

# Bibliography

- [1] Anouar Boucheham, Mohamed Batouche, and Souham Meshoul. An ensemble of cooperative parallel metaheuristics for gene selection in cancer classification. In *International Conference on Bioinformatics and Biomedical Engineering*, pages 301–312. Springer, 2015.
- [2] Gilbert and Manolo Gouy. *Bioinformatique-2e édition: Cours et applications*. Dunod, 2015.
- [3] EL-MABROUK Nadia, Pr Ricardo Baeza-Yates, Pr Maxime Crochemore, and Fr ed erique Lisacek. Th ese de doctorat de l’universit e de paris vii. 2010.
- [4] Francisco Ortuño and Ignacio Rojas. *Bioinformatics and Biomedical Engineering: Third International Conference, IWBBIO 2015, Granada, Spain, April 15-17, 2015. Proceedings*, volume 9044. Springer, 2015.
- [5] Jacques van Helden. *Bioinformatique appliquée*. 2010.
- [6] Lu Zhang, Jianjun Tan, Dan Han, and Hao Zhu. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug discovery today*, 22(11):1680–1685, 2017.