**CHAPITRE I : LA BIOINFORMATIQUE : DEFINITION, DESCRIPTION, DEMARCHE ET PRINCIPALES ETAPES**

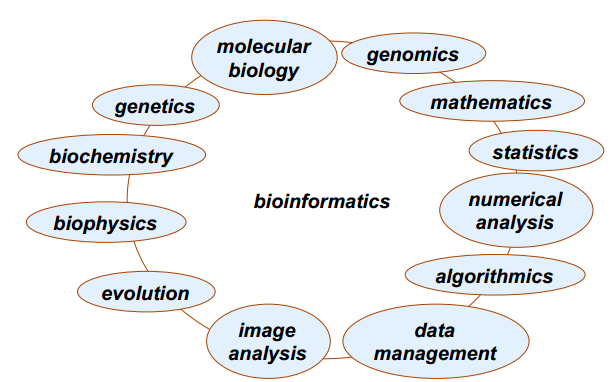
### Introduction

Au cours de ces trente dernières années, la récolte de données en biologie a connu un boom quantitatif grâce notamment au développement de nouveaux moyens techniques servant à comprendre l'ADN et d’autres composants d’organismes vivants. Pour analyser ces données, plus nombreuses et plus complexes aussi, les scientifiques se sont tournés vers les nouvelles technologies de l’information.

L'immense capacité de stockage et d’analyse des données qu'offre l'informatique leur a permis de gagner en puissance pour leurs recherches. Et la rencontre entre la biologie et l'informatique, c'est ce qu'on appelle la bioinformatique. Celle-ci couvre des disciplines des sciences de la vie telles que la génomique, la protéomique et la biologie des systèmes.

### Définition de la Bioinformatique

* La bioinformation est l'information liée aux molécules biologiques : leur séquence, leur nombre, leur(s) structure(s), leur(s) fonction(s), leurs liens de "parenté", leurs interactions et leur intégration dans la cellule ...
* Cette bioinformation est issue de diverses disciplines : la biochimie, la génétique, la génomique structurale, la [génomique fonctionnelle](http://biochimej.univ-angers.fr/Page2/COURS/9ModulGenFoncVeg/1PageDepModGFV.htm), la transcriptomique, la protéomique, la [biologie structurale](http://biochimej.univ-angers.fr/Page2/COURS/7RelStructFonction/3Structure/3ProteinStructure/1ProteinStructure.htm) (structure spatiale des molécules biologiques, [modélisation moléculaire](http://biochimej.univ-angers.fr/Page2/TexteTD/9BioinfoSTRUCTURALE/1MecaniqueMoleculaire/1MecaniqueMoleculaire.htm) ... ), ...
* Une définition de la bioinformatique : analyse de la bioinformation par des moyens informatiques.



**Figure 1 : Champs d’application de la bioinformatique**

### Déscription de la Bioinformatique

### La bioinformatique est une science née de la rencontre entre la biologie et l'informatique.

Selon une définition de l'Institut Suisse de Bioinformatique : La bioinformatique fournit des bases de données centrales, accessibles mondialement, qui permettent aux scientifiques de présenter, rechercher et analyser de l'information.

Elle propose des logiciels d'analyse de données pour les études de données et les comparaisons et fournit des outils pour la modélisation, la visualisation, l'exploration et l'interprétation des données.

**IV. Démarche de la Bioinformatique**

**1.** Compilation et organisation des données biologiques dans des bases de données :

* bases de données généralistes (elles contiennent le plus d'information possible sans expertise très poussée de l'information déposée).
* bases de données spécialisées autour de thèmes précis.

**2.** Traitements systématiques des données : l'un des objectifs est de repérer et de caractériser une fonction et/ou une structure biologique importante. Les résultats de ces traitements constituent de nouvelles données biologiques obtenues "*in silico*".

**3.** Elaboration de stratégies :

* apporter des connaissances biologiques supplémentaires en combinant les données biologiques initiales et les données biologiques obtenues "*in silico*".
* ces connaissances permettent, à leur tour, de développer de nouveaux concepts en biologie.
* concepts qui, pour être validés, peuvent nécessiter le développement de nouvelles théories et outils en mathématiques et en informatique.

### De l’ADN à la fonction cellulaire

### 

### Information manipulée

### 1. ADN (Génome) 2. ARN (Transcriptome) 3. Protéines (Protéome) - Séquences de nucléotides - Séquence - Séquence - Séquence de gènes - Structure - Structure - Banques de données - Réseaux d’intéraction

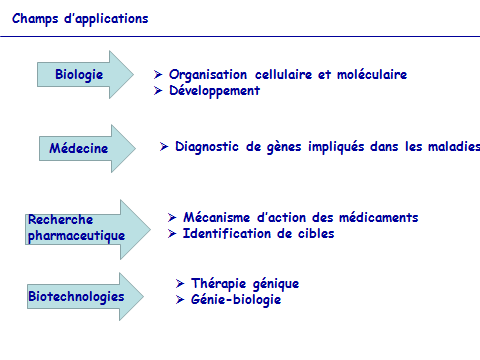
### A quelles questions répond la bioinformatique?

La bioinformatique nous aide à visualiser les structures invisibles tels que les protéines et d'en apprendre davantage sur leur travail et leur fonction. Cela conduit à comprendre les questions essentielles de la vie :

* Comment les organismes fonctionnent-ils ?
* Comment la vie s'est-elle développée ?
* Comment peuvent se développer de nouveaux traitements contre des maladies telles que le cancer ?

1. **Domaines d’applications**

* Gestion des données
* Structures moléculaires " Visualisation, analyse, classification, prédiction
* Analyse de séquences " Alignements, recherches de similarités, détection de motifs
* Génomique " Annotation des génomes, génomique comparative
* Phylogénie " Relations évolutives entre gènes, entre génomes, entre organismes " Inférence de scénarios évolutifs
* Génomique fonctionnelle " Transcriptome, protéome, interactome
* Analyse des réseaux biomoléculaires " Réseaux métaboliques, d’interactions protéiques, de régulation génétique, …
* Biologie des systèmes " Modélisation et simulation des propriétés dynamiques des systèmes biologiques



1. **Principales étapes**

* **1965 :** Première compilation de protéines ("Atlas of Protein Sequences").
* **1967 :** Article : "Construction of Phylogenetic Trees"
* **1970 :** Algorithme pour l'alignement global de séquences
* **1974 :** Programme de prédiction de [structures secondaires](http://biochimej.univ-angers.fr/Page2/COURS/7RelStructFonction/3Structure/1StructPrimQuat/4StructuresII/1StructSecond.htm) des protéines : "Prediction of Protein Conformation
* **1977 :** Techniques de [séquençage d'ADN](http://biochimej.univ-angers.fr/Page2/COURS/9ModulGenFoncVeg/5MethEtudGenFonc/1MethodeSEQUENGAGE/1SEQUENGAGE.htm)
* **1978 – 1980 :** - Séquençage du 1er génome, le bactériophage phiX174 : [Frederick Sanger](http://en.wikipedia.org/wiki/Frederick_Sanger)

- Premières bases de données : EMBL, GenBank, PIR

* **1981 :** Programme d'alignement local de séquences
* **1984 :** Amplification de l'ADN : réaction de polymérisation en chaîne (PCR)
* **1985 :** "[FASTA](http://www.ebi.ac.uk/Tools/sss/fasta/)" : Programme d'alignement local de séquences
* **1987 :** Nouveau vecteur permettant de cloner des fragments d'ADN 20 fois plus grands : le YAC (Yeast Artificial Chromosome) qui rend possible le [séquençage de grands génomes](http://biochimej.univ-angers.fr/Page2/COURS/9ModulGenFoncVeg/5MethEtudGenFonc/1MethodeSEQUENGAGE/1SEQUENGAGE.htm).
* **1988 :** Création du "National Centre for Biotechnology Information" ([NCBI](http://www.ncbi.nlm.nih.gov/)).
* **1993 :** "European Bioinformatics Institute" ([EMBL](http://www.embl.org/)). Création à terme du "European Bioinformatics Institute" ([EMBL - EBI](http://www.ebi.ac.uk/)).
* **1995 :** Analyse du transcriptome : début des [puces à ADN](http://biochimej.univ-angers.fr/Page2/COURS/9ModulGenFoncVeg/5MethEtudGenFonc/3PucesADN/1PucesADN.htm)
* **1996 :** Séquençage complet de la levure (consortium européen).
* **1997 :** -11 génomes bactériens séquencés

- Evolutions de BLAST : "Gapped BLAST" et "[PSI-BLAST](http://www.ncbi.nlm.nih.gov/books/NBK2590/)

* **2000 :** Séquençage du 1er génome de plante : [*Arabidopsis thaliana*](http://biochimej.univ-angers.fr/Page2/COURS/9ModulGenFoncVeg/4Arabidopsis/1Arabidopsis.htm)
* **2001 :** Séquence "premier jet" complète du génome humain
* **2007 - 2008 :** Avènement des nouvelles [technologies de séquençage](http://biochimej.univ-angers.fr/Page2/COURS/9ModulGenFoncVeg/5MethEtudGenFonc/1MethodeSEQUENGAGE/1SEQUENGAGE.htm) à très haut débit, dites de seconde génération et maintenant de 3è génération.
* **Mars 2019 :** Plus de 18.900 génomes eucaryotes et procaryotes séquencés et des milliers en cours de séquençage.