

# Cours 1 : Introduction à la bio-informatique

## 1. Définition de la bioinformatique

Selon le centre national de la biotechnologie informatique " National Center of Biotechnology Informatic NCBI: science in wich biology, computer sciences and information technology merge into a single discipline.

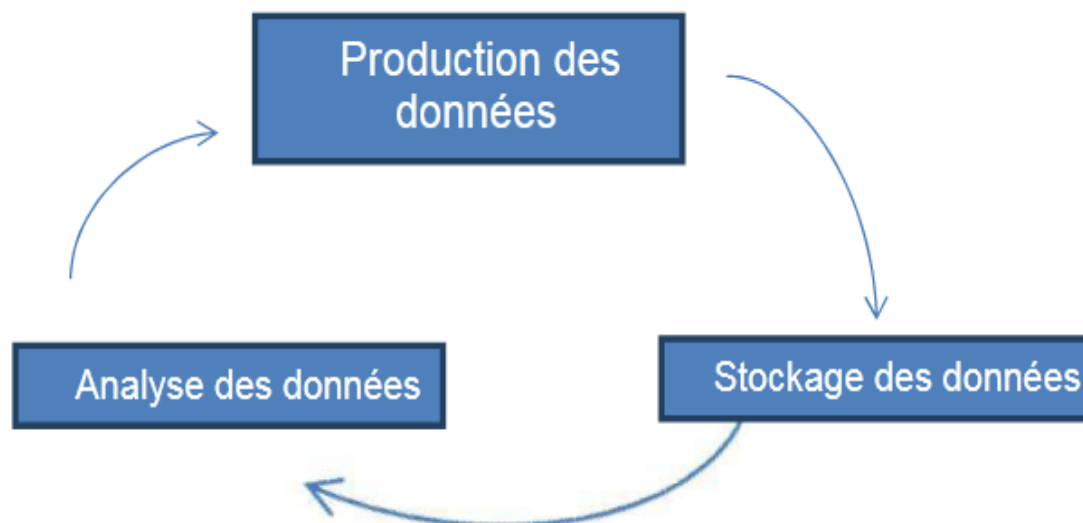
"la bioinformatique est la science dans laquelle la biologie, l'informatique et la technologie fusionnent en une seule science".

La bioinformatique, nouvellement incluse dans les systèmes d'enseignement biologiques (elle émerge dans les années 1980). C'est une discipline récente et un champ de recherche multidisciplinaire où travaillent en concert biologistes, médecins, informaticiens mathématiciens et des physiciens dans le but de résoudre un problème scientifique posé par la biologie.

On peut également la définir comme étant la discipline de l'analyse" in silico " de l'information biologique contenue dans les séquences nucléiques et protéiques.

Trois activités principales :

- Acquisition et organisation des données biologiques
- Conception de logiciels pour l'analyse, la comparaison et la modélisation des données
- Analyse des résultats produits par les logiciels.



## **2. Objectifs :**

Elle est devenue l'outil par excellence pour :

- interpréter les données biomoléculaires,
- analyser la structure des molécules,
- confronter cette structure au reste des molécules existantes dans des bases de données biologiques,
- confronter cette structure au reste des molécules existantes dans des bases de données biologiques,
- prédire le rôle et la fonction de cette structure, ...

## **3. Domaine et utilisation :**

Elle s'intéresse aux données du :

- génome (totalité du matériel génétique de la cellule),
- transcriptome (ARNm transcrits),- transcriptome (ARNm transcrits),
- protéome (l'ensemble des protéines bio synthétisées),
- métabolome (molécules organiques telles que lipides, glucides, faisant partie des activités métaboliques de la cellule vivante).

## Rappel en biologie moléculaire

### 1. L'acide désoxyribonucléique (ADN)

- L'ADN est le support de l'information génétique.
- L'ADN est une longue molécule, faite de deux brins s'enroulant en une double hélice.
- Les deux brins de la double hélice suggèrent un mécanisme de réplication de l'ADN
- Chaque brin est le support d'une succession de nucléotides
- Quatre types de nucléotides : (Adénine, Cytosine, Guanine, Thymine).
- Le texte génomique est écrit dans un alphabet de 4 lettres : A, C, G, T.

Ces 2 chaînes ont 3 propriétés:

#### • Antiparallèles:

Signifie que les deux brins de nucléotides sont parallèles mais dans des directions opposées.

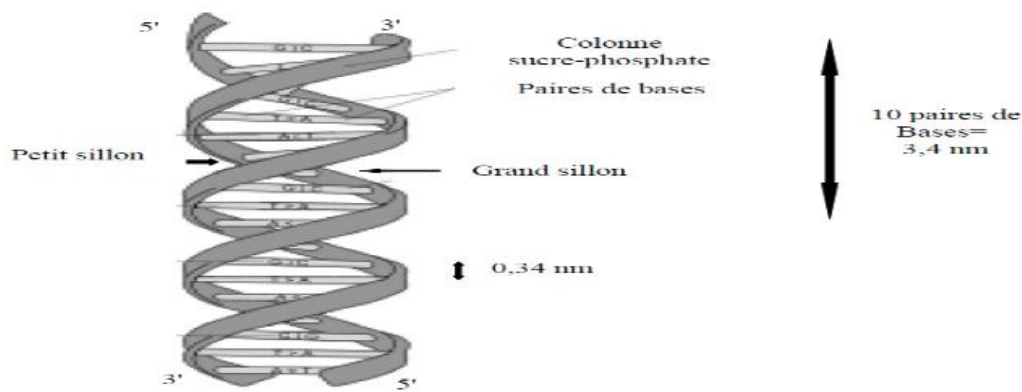
Pour un brin, la direction 5' vers 3' se trouve être, par exemple de haut en bas et pour le 2ème brin, la direction 5' vers 3' sera à l'inverse de bas en haut.

#### • Complémentaires

La règle de complémentarité est la suivante: en face de A on a T et en face de C on a G. En effet, la distribution des bases azotées de l'ADN n'est pas quelconque, le nombre de liaisons hydrogène pour le couple A - T est de deux et pour le couple C - G de trois.

#### • Hélicoïdales

Cette distribution égale de A et T d'une part, de G et C d'autre part implique donc une structure particulière de l'ADN, structure qui a beaucoup intrigué les biologistes.

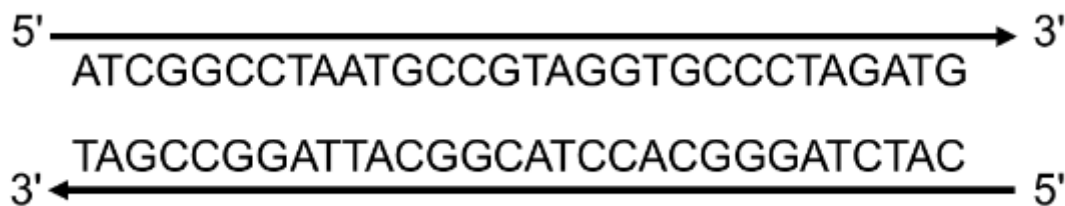


Représentation de la double hélice d'ADN

La structure de l'ADN représenté par la figure, les deux brins sont orientés, on parle de l'orientation 5' → 3' par des considérations biochimiques.

- L'extrémité 5' se termine par un groupement phosphate.
- L'extrémité 3' se termine par un groupement hydroxyle.

Il faut retenir qu'il y a un brin sens orienté 5' → 3' et un brin anti-sens orienté 3' → 5'.

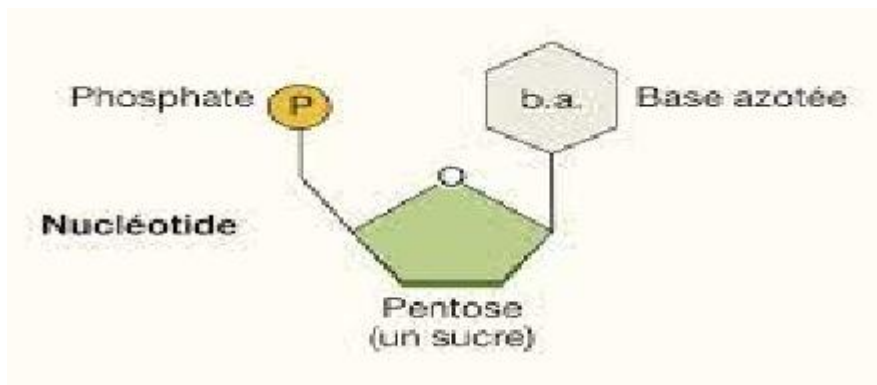


**Figure : Représentation schématique d'une succession de nucléotides d'un fragment d'ADN double brins.**

L'ADN est capable de se répliquer, c'est à dire de se recopier fidèlement. L'ADN se scinde en deux brins, chaque brin séparé servant de modèle pour fabriquer un brin complémentaire. Résultat : deux nouvelles molécules d'ADN, chacune avec un ancien et un nouveau brin.

L'ADN sert trois fonctions principales : **la formation de protéines et de l'ARN, l'échange de matériel génétique lors de la division cellulaire pendant la méiose et la facilitation de mutations génétiques au sein d'une population.**

Les **nucléotides** constituent les éléments de base des acides nucléiques, l'ADN et l'ARN. Ils **sont** composés de trois éléments principaux : une base azotée, liée à une molécule de sucre et à un groupement phosphate.



**Figure : les composants d'un nucléotide**

## **2. L'acide ribonucléique (ARN) :**

Les ARN sont des polymères de nucléotides, leur structure générale est très proche de celle des ADN, mais il faut noter 3 différences essentielles :

L'ARN est monocaténaire, l'ose est le ribose et il n'y a pas de thymine dans la structure de l'ARN, elle est remplacée par l'uracile.

Des molécules d'ARN sont constituées d'une seule chaîne polynucléotidique. Sauf chez certains virus, où des appariements par des doubles hélices peuvent s'établir entre A et U d'une part et entre G et C d'autre part, mais ces appariements se forment entre 2 bases de la même chaîne et n'ont aucun caractère de régularité contrairement à l'ADN.

### **La cellule comporte essentiellement 4 types d'ARN :**

ARNr (ribosomique), ARNt (de transfert), ARNm (messager) et ARNsn (smallnuclear).

Ce sont les ARNr qui sont les plus abondants (82%), suivis par les ARNt (16%) et les ARNm (~2%).

#### **2.1. Les ARN ribosomiques (ARNr) :**

Les ribosomes sont les particules nécessaires à la synthèse des protéines. Situés au niveau du cytosol, les ribosomes sont de véritables usines à protéines dans la cellule. Les ribosomes sont également trouvés au niveau des mitochondries. Les ARNr sont les constituants majeurs des ribosomes environ 65%.

Les ARNr entrent dans la composition du ribosome (nécessaire à la synthèse des protéines). Un ribosome fonctionnel est lui-même formé de deux sous unités, chacune est constituée d'un mélange de protéines (r- protéines) et d'ARN (ARNr). Les ribosomes sont situés dans le

cytoplasme et sont nécessaires à la synthèse des protéines. Ce sont de véritables "usines à protéines".

## **2.2. Les ARN de transfert (ARNt) :**

Ils sont appelés ainsi car ils vont transférer, véhiculer les acides aminés qui se trouvent dans le cytoplasme jusqu'au ribosome, lieu de synthèse protéique. Un ARNt possède la structure générale des ARN. La chaîne d'ARNt se replie pour donner un aspect général en forme de trèfle. Deux sites sont importants dans un ARNt :

- L'extrémité 3'OH ou sera fixé l'acide aminé à transporter
- L'anticodon (triplet) situé sur une boucle de l'ARNt qui va jouer un rôle très important car il reconnaîtra le codon de l'ARNm. Cet appariement anticodon- codon se fait de manière antiparallèle et complémentaire entre les bases du codon et de l'anticodon.

## **2.3. Les ARN messagers**

- il est formé d'une seule chaîne de nucléotides comprenant les mêmes sortes de bases AUCG. On l'appelle messenger car il porte l'information génétique contenue au niveau de l'ADN jusqu'au ribosome où s'effectuera la synthèse protéique. La taille de la molécule d'ARNm dépend de la longueur de la ou les chaînes polypeptidiques pour laquelle il code. Les ARNm se renouvellent très vite, ils sont rapidement produits et rapidement dégradés. Ils ne durent que le temps d'un message. Un ARNm pourra cependant être lu plusieurs fois au niveau du ribosome.

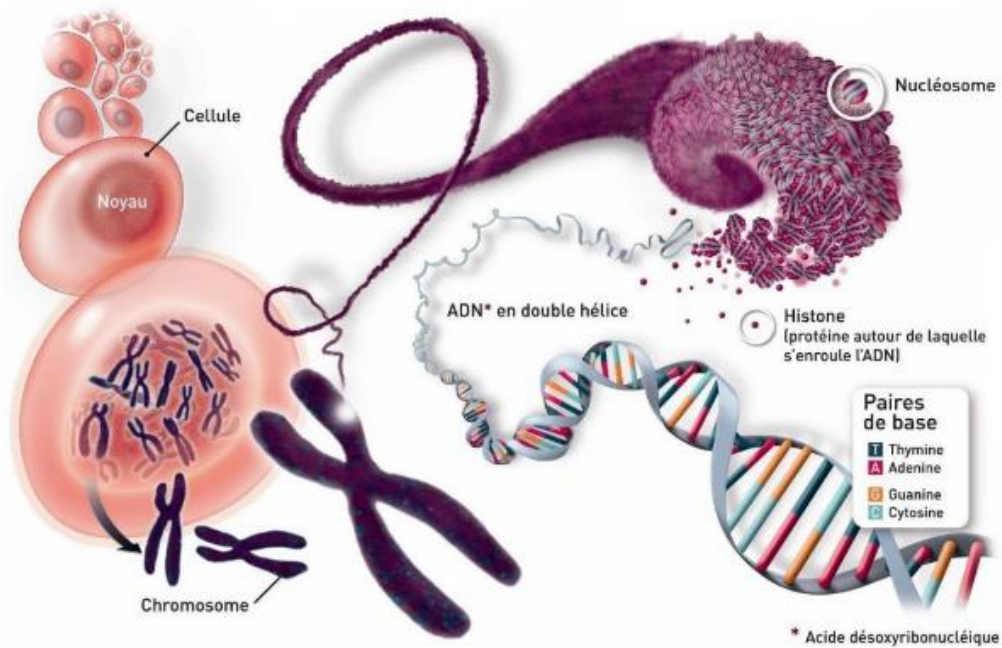
## **2.4. Les ARN nucléaires de petite taille (snRNA)**

- SnRNA (small nuclear) les plus petites molécules d'ARN, nous allons voir plus tard que ces ARN jouent un rôle important dans la maturation des Pré ARNm.

## **3. Le génome**

Le génome, c'est-à-dire l'ensemble des gènes d'un organisme, est structurellement défini par les chromosomes au sein de chaque cellule, selon la théorie chromosomique de l'hérédité.

Chaque chromosome est constitué d'une molécule d'ADN (Figure 5), support physique des gènes, et de protéines associées.



**Figure : Schéma du génome humain, avec la structure en double hélice de l'ADN.**

(Source : Premamag.wordpress.com).

### 3.1. Le génome des procaryotes

Le matériel génétique des bactéries n'est pas organisé de la même façon que chez les eucaryotes. Toutefois, les gènes bactériens sont disposés linéairement sur le chromosome. Les procaryotes possèdent en général un seul chromosome circulaire, en exemplaire souvent unique. Néanmoins, avec l'utilisation de l'électrophorèse en champ pulsé permettant de séparer les molécules d'ADN de grande taille, il a été montré que certaines bactéries ont un chromosome linéaire ou plusieurs chromosomes circulaires et linéaires. En plus les procaryotes possèdent de l'ADN sous forme extra-génomique. Il s'agit d'une petite molécule circulaire d'ADN, appelée plasmide, capable de se répliquer indépendamment du chromosome.

### 3.2. Le génome des eucaryotes

Chez les eucaryotes, les génomes sont en fait visualisés comme des structures filamenteuses, non circulaires, situées majoritairement dans le noyau, et qui peuvent présenter des configurations variables suivant le cycle cellulaire (cf. mitose). Il existe également des chromosomes mitochondriaux et chloroplastiques qui sont pour la plupart circulaires. Ceux-ci sont plus petits que les chromosomes nucléaires et ne présentent pas d'aspect filamenteux.

Les gènes présents sur ces chromosomes extra-nucléaires ne suivent pas les lois de la transmission mendélienne.

**4. La génomique** est une discipline de la biologie moderne. Elle étudie le fonctionnement d'un organisme, d'un organe, d'un cancer, etc. à l'échelle du génome, au lieu de se limiter à l'échelle d'un seul gène.

La génomique se divise en deux branches :

- **La génomique structurale** : qui se charge du séquençage du génome entier ;
  - **La génomique fonctionnelle** : qui vise à déterminer la fonction et l'expression des gènes séquencés en caractérisant le transcriptome et le protéome.
- Le **transcriptome** est l'ensemble des ARN issus de la transcription du génome. L'analyse transcriptomique peut caractériser le transcriptome d'un tissu particulier, d'un type cellulaire, ou comparer les transcriptomes entre différentes conditions expérimentales.
- Le **protéome** est l'ensemble des protéines exprimées dans une cellule, une partie d'une cellule (membranes, organites) ou un groupe de cellules (organe, organisme, groupe d'organismes) dans des conditions données et à un moment donné.