

# Chapitre 2

## Les nouvelles techniques de la recherche scientifique.

**Introduction** : L'utilisation de l'Internet pour la recherche de l'information biologique est d'actualité. Si la méthode n'est pas structurée, le chercheur de l'information aura le sentiment d'être perdu au sein de cette gigantesque toile d'araignée qui est le web. C'est pour cela qu'une structuration et une modélisation de la méthode de recherche s'imposent. Cela permet, en effet de gagner énormément de temps et d'effectuer des recherches plus spécifiques.

ils y a des techniques nouvellement incluse dans les systèmes d'enseignement biologiques (elle émerge dans les années 1980). ce sont des technique qui permet l'analyse et l'interprétation des informations biologiques contenues On peut également la définir comme des technique de l'analyse" *in silico* " de l'information biologique .

La révolution extraordinaire que connaît la biologie ces dernières années, est en grande partie associée au développement spectaculaire de ces nouveaux outils. Celle-ci, grâce à ses méthodes qui se basent essentiellement sur les connaissances approfondies de la biologie moléculaire, des mathématiques et de l'informatique (car interdisciplinaire), vient contribuer spectaculairement à l'avancement des connaissances biologiques aussi bien au niveau génomique que protéomique.

. Elles sont devenue l'outil par excellence pour :

- interpréter les données biomoléculaires,
- analyser la structure des molécules,
- prédire le rôle et la fonction de cette structure, ...

Elle s'intéresse aux données du :

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

Le développement rapide et l'utilisation facilitée de Internet, grâce à ses interfaces graphiques conviviales, ont permis de gérer des ressources bibliographiques et de les partager ; on parlera alors de banques de données biologiques, de délocalisation des ressources.

## Les bases et les banques de données

1. qu'est ce qu'une base de données (BD) ?
2. différence entre banque de données et base de données ?

Une base de données, qu'est ce que ?

- Ensemble de données relatives à un domaine, organisées par traitement informatique, accessibles en ligne et à distance
- Souvent, les données sont stockées sous la forme d'un fichier texte formaté (respectant une disposition particulière)

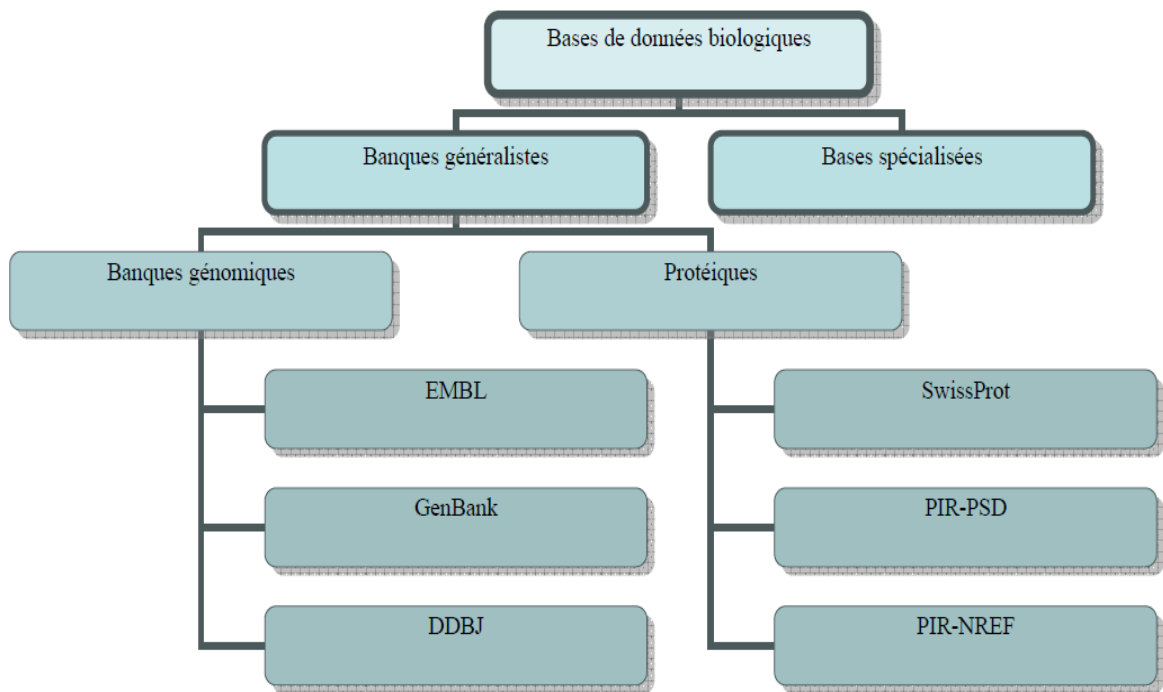
Besoin de développer des logiciels spécifiques pour interroger les données contenues dans ces banque Une **base de données**, usuellement abrégée en *BD* ou *BDD*, est un ensemble structuré et organisé permettant le stockage de grandes quantités d'informations afin d'en faciliter l'exploitation (ajout, mise à jour, recherche de données). Une **base de données** se traduit physiquement par un ensemble de fichiers présent sur une mémoire de masse (bien souvent un disque). Certaines peuvent être accessibles via les réseaux, on parle alors de base de données en ligne

## Les banques de données généralistes

- Ces banques contiennent des données hétérogènes
  - Collecte la plus exhaustive possible
  - Banques de séquences nucléiques
  - Banques de séquences protéiques
  - Banques de structure 3D de macromolécules
  - Banques d'articles scientifiques

## Les banques de données spécialisées

- Ces banques contiennent des données homogènes
  - Collecte établie autour d'une thématique particulière



### Quelques exemples de BD sur Internet:

1. **BD bibliographique Medline** : (de l'anglais : *Medical Literature Analysis and Retrieval System Online*) est une base de données bibliographiques regroupant la

littérature relative aux sciences biologiques et biomédicales. La base est gérée et mise à jour par la Bibliothèque nationale de médecine des États-Unis d'Amérique (NLM). L'interface PubMed permet de consulter gratuitement la base de données à partir d'un navigateur World Wide Web.

Bases de données biologiques

Banques généralistes Bases spécialisées

Banques génomiques Protéiques

EMBL

GenBank

DDBJ

SwissProt

PIR-PSD

PIR-NREF

La base de données MEDLINE possède en avril 2007 plus de 15 millions de citations publiées depuis 1950 dans environ 5000 revues biomédicales (revues en biologie et en médecine) distinctes. Elle est la base de données de référence pour les sciences biomédicales.

**2. BD bibliographique WorldCat :** WorldCat est le catalogue en ligne de l'OCLC (Online Computer Library Center), réputé le plus grand catalogue OPAC (*Online public access catalog*) du monde. Son nom est la contraction de l'anglais *World Catalog* (catalogue mondial).

**Exemples :**

- Base de données sur les maladies : GeneCards, OMIM, OMIA, ...
- Base de données taxonomique
- Bases de structure : PDB (Protein Data Bank)
- Bases d'expression : YPD (Yeast Proteome Database), MGED (Microarray and Gene Expression Data Home)
- Bases de voies métaboliques : KEGG
- Bases de cartographie
- Base de phénotypie : MIM, MIA

Une banque de données biologique, est une base de données généralisée. On y

trouve des informations sur nombreuses espèces, nombreuses molécules, en même

temps. Son utilisation concerne plusieurs domaines à la fois. **Elle est donc généraliste et correspond à un ensemble de données exhaustif avec des informations hétérogènes.**

Il faut savoir qu'il existe des banques de données génomiques et des banques de données protéiques.

Les trois principales banques de données nucléiques sont :

**1. GenBank** de NCBI (National Center for Biotechnology Information) :

<http://www.ncbi.nlm.nih.gov/>.

**2. EMBL** de EMBO (European Molecular Biology Organization):

<http://www.ebi.ac.uk/embl/> .

**3. DDBJ** : Dna Data Base of Japan : <http://www.ddbj.nig.ac.jp/searches-e.html>

Les deux principales banques de données protéiques sont :

**1. PIR-NBRF** : D'abord, elle fut créée par la NBRF (National Biomedical Research Foundation)

**2. SwissProt** : Créée par le biochimiste Amos BAIROCH en 1986 à l'Université de Genève. actuellement développée en collaboration entre l'Institut Suisse de BioInformatique (ISB-SIB [<http://www.isb-sib.ch/> .])

### **Interrogation des bases de données**

On peut interroger une BD pour plusieurs raisons :

- Pour connaître la séquence d'un gène ou d'une portion de ce gène
- Pour connaître la structure primaire d'une protéine
- Pour comparer deux séquences, ...

Le résultat de l'interrogation des BD est une fiche descriptive de la molécule. On parlera alors d'une **entrée** (ou fiche descriptive de la séquence recherchée). La structure d'une entrée est presque la même quelque soit la BD interrogée.