

Université Mohamed KHIDER Biskra
Faculté des Sciences Exactes et Sciences de la Nature et de le Vie
Département des Sciences de la Nature et de la Vie

1^{ère} Année mastère : Biochimie, Microbiologie et Biotechnologie végétale
Module : Bioinformatique

TP 03 : Interrogation des banques nucléiques et protéiques

1. Familiarisation avec le portail NCBI

Nous tapons l'adresse suivante : <https://www.ncbi.nlm.nih.gov/>, pour explorer les différents volets de ce portail (Tools DNA and RNA, Banques de données bibliographiques).

- Dans le menu déroulant nous choisissons la banque « **Nucleotide** ».
- Sur l'interface de cette banque et dans la case de recherche nous tapons la molécule d'intérêt (Par exemple **Insulin** : toujours SVP en Anglais).
- Nous répétons la même requête mais cette fois on utilise l'un des **opérateurs Booléens (AND, NOT ou OR)**. Par exemple (Insulin AND homo sapiens). Que vous remarquer ?
- Nous choisissons une proposition (résultat numéro d'accession : **AH002844**) pour accéder à l'entrée correspondante.
- Définir les différentes parties de cette entrée (les champs).
- Nous voudrions accéder à la publication à partir de laquelle la séquence a été extraite en cliquant sur le lien vers **PUBMED**.
- Nous voudrions récupérer la séquence dans fichier Doc, sous format « texte » pour cela nous cliquons sur **FASTA**.
- Copier la séquence pour une ultérieure utilisation.

2. Banque de données EMBL de portail EBI.

- Nous tapons l'adresse suivante : <https://www.ebi.ac.uk/>, Pour accéder à la banque EMBL.

- Dans la case de recherche, nous tapons le nom de la molécule d'intérêt « toujours **Insulin** », nous cliquons dans le volet gauche sur « Nucleotide sequences), ensuite choisissons le premier résultat en explorant les différentes parties de l'entrée EMBL.

- Nous prendrons maintenant le numéro d'accession de l'entrée de GenBank «AH002844 » pour le coller dans la case de recherche de la banque EMBL.
Que vous remarquez ?

3. Banques de données protéiques

-UniProt

- Nous tapons l'adresse suivante : <https://www.uniprot.org/>, pour accéder à la banque « UniProtKB ».
- Nous choisissons « UniProtKB » ensuite nous tapons « Gliadin ».
- Nous cliquons sur un résultat (Format texte) pour découvrir l'entrée de cette banque.

(N.B. cette entrée comporte des liens vers des BD structurales comme PDB : Protein data base).

-Pour récupérer la séquence sous format « texte », toujours nous cliquons sur l'option **FASTA**.

4. Bases de données des Motifs et des domaines protéiques

4.1. Base de données pour les motifs protéiques « Prosite ».

-Pour y accéder nous utilisons l'adresse suivante : <https://prosite.expasy.org/>

- Pour l'obtention de motifs à partir de la protéine concernée, nous entrons le code ID de UniProtKB ou la séquence protéique (format FASTA), et nous cliquons sur le bouton "Scan".

- Finalement nous obtenons une entrée descriptive des résultats de la recherche de motifs protéiques.

4.2. Base de données pour les domaines protéiques « Pfam »

- Pour y accéder nous allons à l'adresse suivante : <https://pfam.xfam.org/>.

- Pour l'obtention de domaines à partir de la protéine concernée, nous entrons le code ID de UniProtKB ou la séquence protéique (format FASTA), et nous cliquons sur le bouton "GO". Nous le (les) domaine (s) de la protéine en questions.