

TP 4 : Banques de données et alignement

1- Banques de données :

NCBI (The *National Center for Biotechnology Information*) :

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

<ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/>

EMBL-EBI (European Molecular Biology Laboratory - The European Bioinformatics Institute) :

<https://www.ebi.ac.uk/ena>

UniProt (Universal Protein resource) :

<https://www.uniprot.org/>

ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/reference_proteomes/

2- Alignements :

Logiciels d'alignements EMBOS : <https://www.ebi.ac.uk/Tools/emboss/>

Local Alignment

Target Sequence
5' ACTACTAGATTACTTTACGGATCAGGTACTTTAGAGGCTTGCAACCA 3'

Query Sequence 5' TACTCACGGATGAGGTACTTTAGAGGC 3'

Global Alignment

Target Sequence

5' ACTACTAGATTACTTACGGATCAGGTACTTTAGAGGCTTGCAACCA 3'

|||||

5' ACTACTAGATT---ACGGATC--GTACTTTAGAGGCTAGCAACCA 3'

Query Sequence

Alignement global vs alignement local

© Major differences

Exercice :

Bases de données :

Sur le site NCBI -

1. Section Nucleotide, cherchez une séquence appartenant à l'espèce Homo sapiens et à la maladie diabète. Combien d'entrées d'ADN et ARN trouvez-vous?
 - Sélectionnez la première entrée d'ARN, à quel date a-t-elle été ajoutée à la base de donnée?
2. Section Pubmed, cherchez un article scientifique qui dont le titre contient **lipid homeostasis** d'auteurs **Baek SH** et **Kim K**, l'année d'apparition est **2017**.

Alignement :

Prenez les deux exemples, Homo sapiens et Mus musculus en format fasta (<https://www.ebi.ac.uk/seqdb/confluence/display/JDSAT/Pairwise+Sequence+Alignment+Tool+Input+Examples>) et faite un alignement global et local.

Quelles sont les différences entre des deux alignements?

Programmation :

Matrice d'alignement

Cette semaine vous devriez faire un programme qui ouvre un fichier en format FASTA extrait les deux séquences, aligne et affiche l'alignement en forme d'une matrice.

L'alignement de deux séquences se fait par un match et mismatch. Si c'est un match le programme affiche le symbole « - » si c'est un mismatch il affiche « * ».

Exemple :

l'alignement de deux séquences AAT TAT

AAT	AAT	AAT	AAT	AAT	AAT	AAT	AAT	AAT
TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT
*	-	*	*	-	*	-	*	-

Matrice d'alignement

_

_

_

PS : "\n" permet le passage de la ligne suivante.