

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique  
Université des Sciences et de la Technologie d'Oran – Mohamed Boudiaf  
**Faculté des Mathématiques et Informatique**  
*Département d'informatique*



## TP 2 Bio-informatique

# Alignement des séquences par FASTA et BLAST

Domaine : Mathématiques – Informatique  
Filière : Informatique  
Spécialité : M2 IAA

Présenté par :

- KEBIRI Issam Dine

Supervisé par :

- Mme Drioua.w, department d'informatique - USTO.

### Questions :

Ce TP a pour but de vous présenter les outils BLAST et FASTA et l'alignement des séquences en utilisant ces outils.

1. Sur le serveur NCBI, identifier toutes les séquences de la protéine de la pénicilline.
2. Choisir un type de pénicilline et récupérer la séquence au format FASTA.
3. En utilisant la séquence récupérée au format FASTA, faites un BLAST au NCBI.
4. A l'aide de la même séquence au format FASTA, interroger la banque de données SwissProt en recherchant les séquences voisines.
5. Est-ce que les séquences trouvées sont les mêmes que celle trouvées par BLAST.

### Figures :

Figure 1 : Les séquences de la protéine de la pénicilline

Figure 2 : Séquence FASTA choisi

Figure 3 : Résultat du BLAST au NCBI

Figure 4 : Recherche des séquences voisines sur SwissProt

## Réponse Q1 :

Après avoir cherché sur le serveur NCBI<sup>1</sup> toutes les séquences de la protéine de la pénicilline, j'ai trouvé plus de 1,618,232 de résultats comme indiqué sur la figure 1.

Literature		Genes		Proteins	
Bookshelf	647	Gene	8,487	Conserved Domains	63
MeSH	11	GEO DataSets	134,281	Identical Protein Groups	55,058
NLM Catalog	21	GEO Profiles	120	Protein	1,618,232
PubMed	17,985	HomoloGene	1	Protein Family Models	588
PubMed Central	396,123	PopSet	413	Structure	1,923

Figure 1: Les séquences de la protéine de la pénicilline

## Réponse Q2 :

J'ai choisi le type suivant : pénicilline protéine [Solibacillus silvestris StLB046]

La séquence en fasta<sup>2</sup> :

```
MAQQAKKSKGRKFINIVIWAVGILIIVGGAAVAGLHLYVSNKPVIEGELAVSILDEDVEVTRDGV
GVPHIVAKSDADLYRAQGYYVQAQDRLFQMDLARRQASGTLAEVVGESAVNTDKFFRTFSLRHA
AEQSWDDYDAQSKQILEWYTEGVNAFIDEVKDTSKLSYEFKLLGYKPEAWTPIDSLTIGKYMAYDL
GGNWNQQVFRSWAMQNYTQEQAEEFVEYPENASQSIIEANLNLETDIASAIPDYLPSHNGSN
NWVVSGLTKSGKPLLADDPHLLGLSTPSVWYQMHLSPEQNVSGVIFAGIPGILGHNEQIAWG
VTNVNPDVQDIYIETPNPENPYQFEYDGEWEEATVRQEPIKVKGGETVDFEVVETRHGPIISNIMLK
DTKMKEQFSMQWTALQSTQELKAVLGFNKATDWEFEVALEDFKAPAQNFVFASTDGTIAYKAN
GLVPIRKQGTGALPVPNGNSSEYGWESYIPFDELPTVINPEEGFIATANNQVVGEEYPYHISNLWAQ
PYRYERIVEMIKTPMYDEETGELLKLSVAEMKGMQMDKKNLHAQEFLPGLLETIKAQDKDGKYKN
VITLLEWEYYDDKDQGAPLVYHFLIENMEKALFQDAMPEDVYKLMPAKSQITDQMLRDAYAGN
PGVWITEEGGLDKFVYAFEDAVTKIKGKFGTSVSKWKWGSYNQLTFDHLASASDFLADYVNPE
```

<sup>1</sup> NCBI : Le National Center for Biotechnology Information (NCBI, « Centre américain pour les informations biotechnologiques ») est un institut national américain pour l'information biologique moléculaire.

<sup>2</sup> Le format FASTA (ou format Pearson) est un format de fichier texte utilisé pour stocker des séquences biologiques de nature nucléique ou protéique. Ces séquences sont représentées par une suite de lettres codant pour des acides nucléiques ou des acides aminés selon la nomenclature IUPAC.

KLPVGGSSVTVQAA SEDGSGNVNHGASWRFVADLDNLKSAYHIVGPGQSGHVKSKWYDNQT  
SNWVYGRYHKTKINGDIDHGYELILKAE

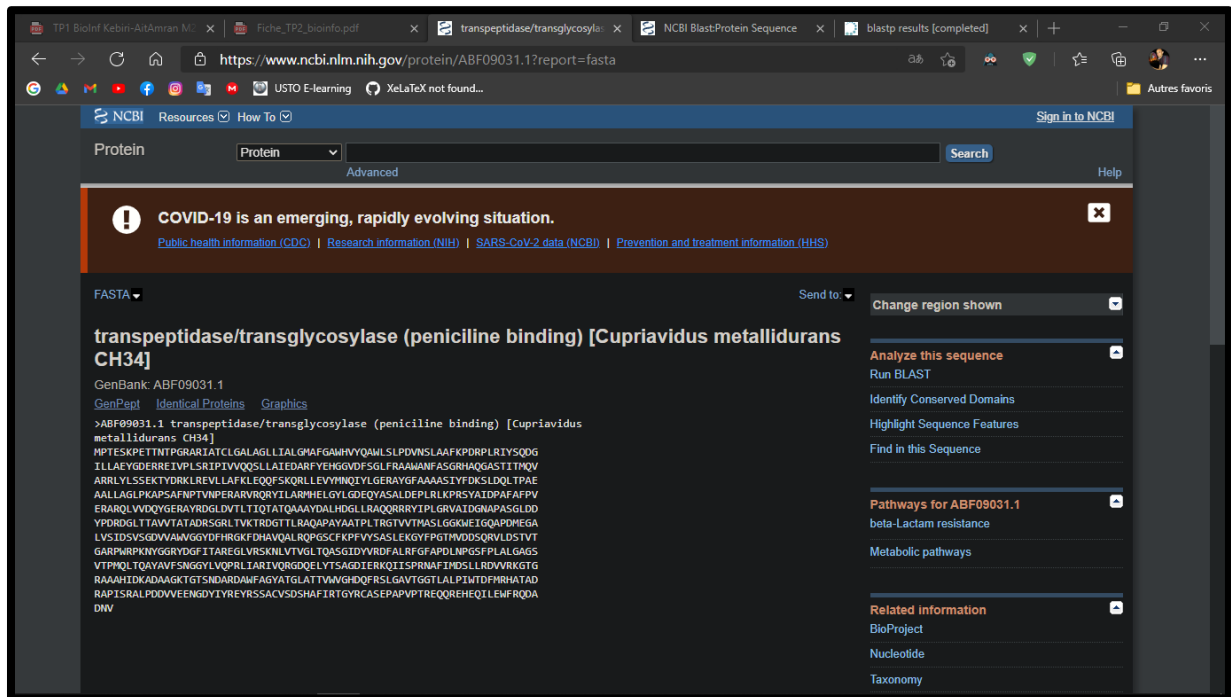


Figure 2: Séquence FASTA choisi

### Réponse Q3 :

Après application de BLAST<sup>3</sup> au NCBI les résultats sont montrés dans la figure 3 :

<sup>3</sup> BLAST est une méthode de recherche heuristique utilisée en bio-informatique. Il permet de trouver les régions similaires entre deux ou plusieurs séquences de nucléotides ou d'acides aminés, et de réaliser un alignement de ces régions homologues.

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Crystal structure of a double mutant of a class III engineered cephalosporin acylase [Pseudomonas]	Pseudomonas	186	186	64%	3e-50	27.12%	543	4HST_B
Crystal Structure of a class III engineered cephalosporin acylase [Pseudomonas]	Pseudomonas	186	186	64%	3e-50	27.12%	543	4HSR_B
Chain B. Penicillin G acylase [Priestia megaterium]	Priestia megaterium	137	137	64%	1e-33	26.37%	537	6NVW_B
Crystal structure of penicillin G acylase from Bacillus thermotolerans [Quasibacillus thermotolerans]	Quasibacillus ther...	136	136	41%	4e-33	30.12%	538	6NVY_B
Crystal structure of penicillin G acylase from Bacillus sp. FJAT-27231 [Bacillus sp. FJAT-27231]	Bacillus sp. FJAT-...	132	132	53%	8e-32	27.11%	538	6NVX_B
A slow processing precursor penicillin acylase from Escherichia coli [Escherichia coli]	Escherichia coli	129	129	63%	1e-30	24.82%	560	1E3A_B
Penicillin Acylase Mutant With Phenyl Propionic Acid [Escherichia coli]	Escherichia coli	128	128	63%	1e-30	24.86%	557	1KEC_B
Penicillin Acylase, mutant [Escherichia coli]	Escherichia coli	127	127	63%	2e-30	24.86%	557	1JX9_B
Crystal structures of penicillin acylase enzyme-substrate complexes: Structural insights into the catalytic mechanism ...	Escherichia coli	127	127	63%	4e-30	24.68%	557	1GK9_B
Altered substrate specificity mutant of penicillin acylase [Escherichia coli]	Escherichia coli	127	127	63%	5e-30	24.68%	557	1H2G_B
Penicillin Acylase Complexed With 3,4-Dihydroxyphenylacetic Acid [Escherichia coli]	Escherichia coli	126	126	63%	8e-30	24.49%	557	1AI4_B
Mutant Of Penicillin Acylase Impaired In Catalysis With Phenylacetic Acid In The Active Site [Escherichia coli]	Escherichia coli	125	125	63%	1e-29	24.68%	557	1FXH_B
S1C mutant of Penicillin G acylase from Kluyvera citrophila [Kluyvera cryocrescens]	Kluyvera cryocres...	125	125	41%	1e-29	26.96%	565	4PEI_B
Crystal Structure of S1G mutant of Penicillin G Acylase from Kluyvera citrophila [Kluyvera cryocrescens]	Kluyvera cryocres...	125	125	41%	1e-29	27.03%	568	4PEM_B
Crystal Structure of a class III engineered cephalosporin acylase [Pseudomonas]	Pseudomonas	117	117	25%	3e-29	33.65%	229	4HSR_A
Crystal structures of penicillin acylase enzyme-substrate complexes: Structural insights into the catalytic mechanism ...	Escherichia coli	124	124	63%	3e-29	24.49%	55	
Crystal Structure Of Penicillin G Acylase From The Bro1 Mutant Strain Of Providencia rettgeri [Providencia rettgeri]	Providencia rettgeri	123	123	41%	7e-29	27.62%	55	

Figure 3: Résultat du BLAST au NCBI

## Réponse Q4 :

Le résultat de la recherche sur la banque de données SwissProt<sup>4</sup> des séquences voisines est montré dans la figure 4 :

<sup>4</sup> Swissprot est une banque de données biologiques de séquences protéiques où les informations sont saisies et vérifiées manuellement.

Entry	Alignment overview	Info	Status
All 250 result(s) selected. (or select only the results on this page)			
Query: B20210216A94466D2655679D1FD8953E075198DA8001466Q			
<input checked="" type="checkbox"/> Q1LLE5	Q1LLE5_CUPMC - Penicillin-insensitive transglycosylase - Cupriavidus meta... - View alignment	E-value: 0.0 Score: 3,993 Ident.: 100.0%	
<input checked="" type="checkbox"/> C5T5P2	C5T5P2_ACIDE - DD-transpeptidase - Acidovorax dela... - View alignment	E-value: 0.0 Score: 2,046 Ident.: 55.2%	
<input checked="" type="checkbox"/> A0A410UR72	A0A410UR72_9BURK - DD-transpeptidase - Janthinobacteriu... - View alignment	E-value: 0.0 Score: 2,055 Ident.: 54.4%	
<input checked="" type="checkbox"/> A0A6P1IWQ7	A0A6P1IWQ7_9BURK - DD-transpeptidase - Hydrogenophaga s... - View alignment	E-value: 0.0 Score: 2,004 Ident.: 53.2%	
<input checked="" type="checkbox"/> A0A2S9H1M0	A0A2S9H1M0_9BURK - DD-transpeptidase - Solimicrobium si... - View alignment	E-value: 0.0 Score: 1,958 Ident.: 52.6%	
<input checked="" type="checkbox"/> A0A0M2WPK4	A0A0M2WPK4_9BURK - DD-transpeptidase - Janthinobacteriu... - View alignment	E-value: 0.0 Score: 1,936 Ident.: 50.9%	

**Figure 4: Recherche des séquences voisines sur SwissProt**

## Réponse Q5 :

Analyse des deux résultats : il existe une grande similitude entre les résultats trouvés avec le BLAST au NCBI et sur la banque de données SwissProt mais les résultats ne sont pas exacts à 100% car se sont des bases de données différents.

A retenir : il existe deux moyens (ou méthode) afin de trouver des séquences voisines.