

Université Mohamed Khider, Biskra
Faculté des Sciences Exactes et de la Nature et de la Vie
Département SNV

Master 1
Cours de Bioinformatique

Année universitaire: 2019-2020

Programme du module

Chapitre 1: Initiation à la conception des bases de données.

Chapitre 2: Banques de données biologiques (génomiques, protéiques et bibliographiques).

Chapitre 3: Alignement des séquences biologiques (séquences génomiques et protéique).

Chapitre 4: Motifs nucléiques (Matrice de poids).

Chapitre 5: Prédiction de structure tridimensionnelle des protéines (Algorithmes de première et deuxième génération).

Chapitre 6: Initiation à la phylogénie moléculaire.

❖ Introduction:

La bioinformatique, nouvellement incluse dans les systèmes d'enseignement biologiques (elle émerge dans les années 1980). C'est une discipline qui permet l'analyse et l'interprétation des informations biologiques contenues soit dans génome (séquences ADN, ARN) soit dans le protéome. 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.

❖ 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,
- prédire le rôle et la fonction de cette structure, ...

Domaine et utilisation:

Elle s'intéresse aux données du :

- génome (totalité du matériel génétique de la cellule),
- 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).

Chapitre 2 : Les Banques Et Bases De Données Biologiques

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

Il nous faut distinguer deux choses :

qu'est ce qu'une base de données (BD) ?

différence entre banque de données et base de données ?

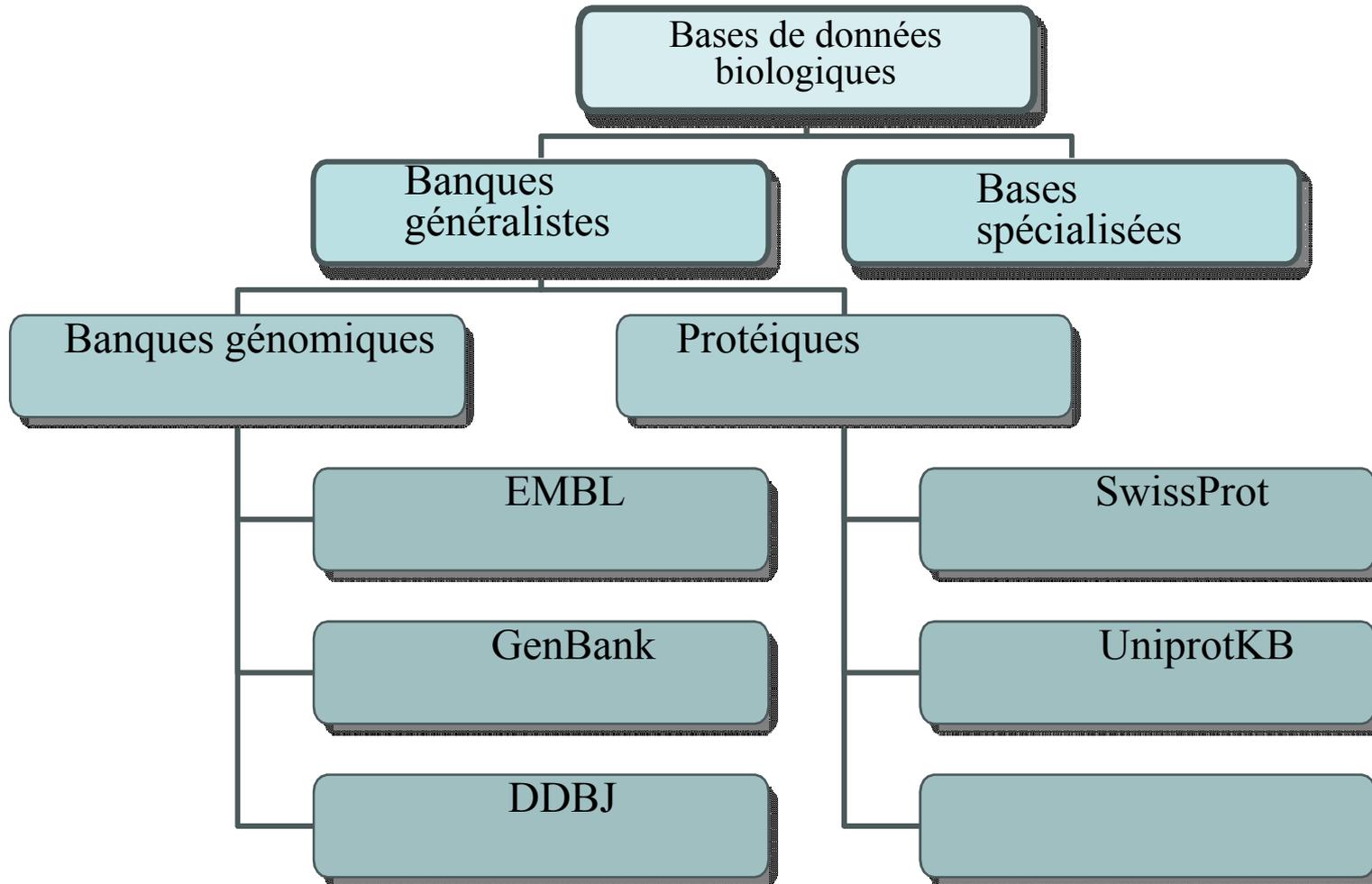
Une **base de données**, usuellement abrégée en *BD* ou *BDD* ,
Une base de données est un fichier ou un ensemble de fichiers permettant le stockage permanent ou temporaire des informations ainsi que l'accès à ces informations devenues structurées (<http://www.webadev.com/lexique-b-base-de-donnees.php>).

C'est un tableau dans lequel on intègre des informations de manière logique et structurée comme la liste d'un groupe d'étudiants :

Différence entre bases de données et banques de données

Il convient de dire qu'une banque de données est une base de données (car tableau structuré) mais qui contient des informations biologiques hétérogènes (virus, bactéries, champignons, végétaux, animaux) alors qu'une base de données est plus spécialisée (base spécifique à *E. coli*, à *Bacillus*, etc.).

Les Banques Et Bases De Données Biologiques



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

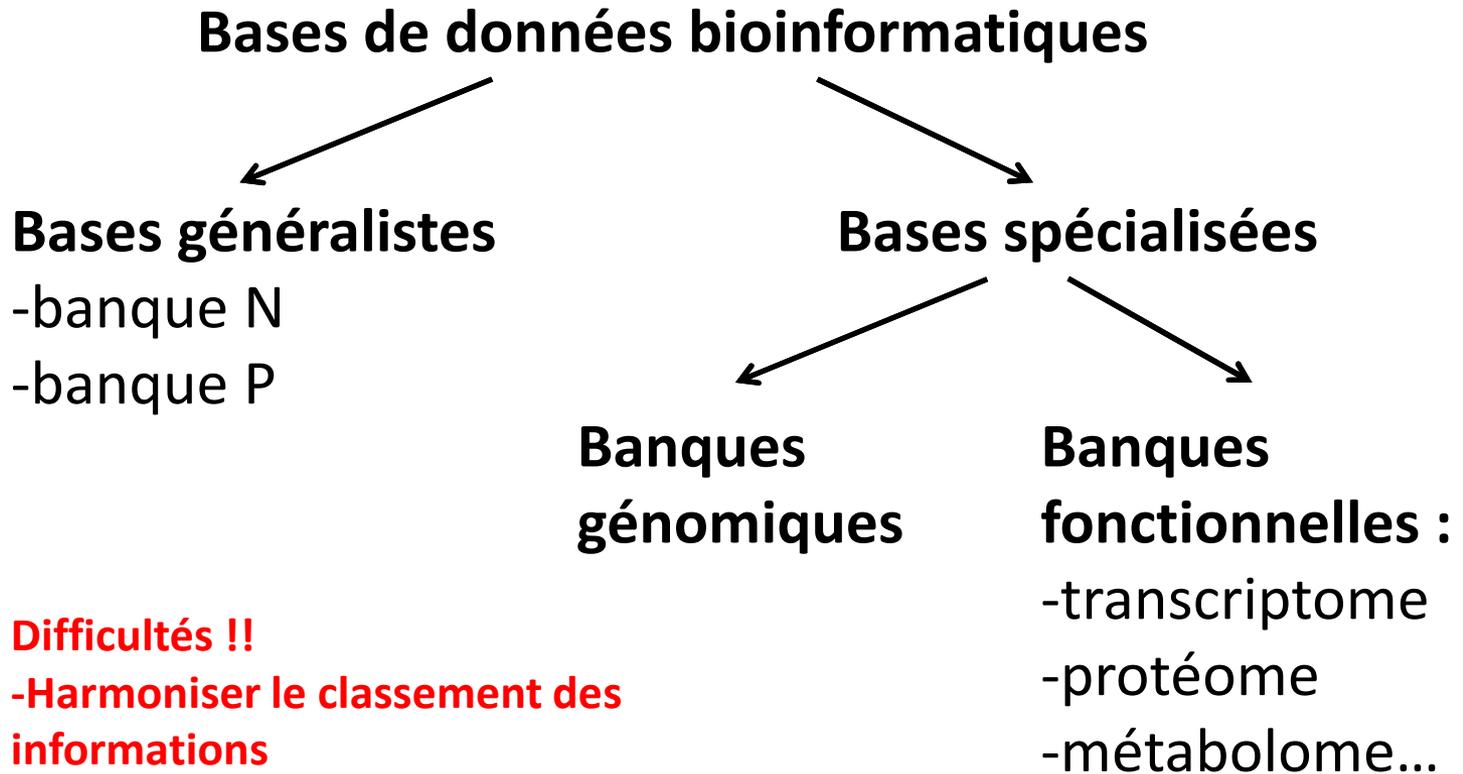
<http://www.ncbi.nlm.nih.gov/>. Créée par IntelliGenetics en 1982. jusqu'en octobre 2004 elle contenait **38 941 263 entrées** (ou séquences par auteur)

• **EMBL** de EMBO (European Molecular Biology Organization): <http://www.ebi.ac.uk/embl/> . La banque EMBL contient **44 538 943 entrées** jusqu'en octobre 2004.

• **DDBJ** : Dna Data Base of Japan : <http://www.ddbj.nig.ac.jp/searches-e.html> Crééé en 1986 et diffusée par NIG (National Institute of Genetics, Japan). En octobre 2004, elle contenait **37 926 117 entrées**.

Les bases de données

- Différentes catégories de bases de données :



Difficultés !!

- Harmoniser le classement des informations
- Utiliser un langage commun pour échanger des informations entre toutes ces bases

Les bases de données

➤ Harmonisation des fiches de données

Exemple de la fiche GENBANK d'un plasmide d'*E.faecalis*

```
LOCUS          KC297657                673 bp    DNA     linear   BCT 18-MAR-2013
DEFINITION    Enterococcus faecalis strain 493/96 plasmid OrfC gene, partial cds;
              RNAI (rnaI) and RNAII (rnaII) genes, complete sequence; and OrfD
              gene, partial cds.
ACCESSION     KC297657
VERSION       KC297657.1  GI:460758767
KEYWORDS      .
SOURCE        Enterococcus faecalis
  ORGANISM    Enterococcus faecalis
              Bacteria; Firmicutes; Bacilli; Lactobacillales; Enterococcaceae;
              Enterococcus.
REFERENCE     1 (bases 1 to 673)
  AUTHORS     Wardal,E., Sadowy,E. and Hryniewicz,W.
  TITLE       Diversity of plasmid-associated genes among Enterococcus faecalis
              clinical isolates
  JOURNAL     Unpublished
REFERENCE     2 (bases 1 to 673)
  AUTHORS     Wardal,E. and Sadowy,E.
  TITLE       Direct Submission
  JOURNAL     Submitted (10-DEC-2012) Molecular Microbiology, National Medicines
              Institute, Chelmska 30/34, Warsaw 00-725, Poland
COMMENT       ##Assembly-Data-START##
              Sequencing Technology :: Sanger dideoxy sequencing
              ##Assembly-Data-END##
```

Les bases de données

➤ Harmonisation des fiches de données

Suite de la fiche GENBANK d'un plasmide d'*E.faecalis*

```
FEATURES             Location/Qualifiers
    source             1..673
                       /organism="Enterococcus faecalis"
                       /mol_type="genomic DNA"
                       /strain="493/96"
                       /isolation_source="hospitalized patient"
                       /host="Homo sapiens"
                       /db_xref="taxon:1351"
                       /plasmid="unnamed"
                       /country="Poland"
                       /PCR_primers="fwd_name: par-F, fwd_seq:
ccatgcactactaggaacc, rev_name: par-R, rev_seq:
ctgtctagcaagcagagttacg"
    CDS              <1..65
                       /codon_start=3
                       /transl_table=11
                       /product="OrfC"
                       /protein_id="AGH14172.1"
                       /db_xref="GI:460758768"
                       /translation="YKCSWCKRVYTLRKDHKTAR"
    gene              90..334
                       /gene="rnaI"
    misc RNA         90..334
                       /gene="rnaI"
                       /product="RNAI"
                       /note="par toxin"
```

Les bases de données

➤ Harmonisation des fiches de données

Fin de la fiche GENBANK d'un plasmide d'*E.faecalis*

ORIGIN

```
1 agtataaatg ttcttggtgt aaacgagttt acacgcttag aaaagatcat aaacagcta
61 gataaattgt tgaaggtttt attattgaat tggcagaatt tcaatctatg ctataattaa
121 tacggcagct cgctcgatt ggaggtgtgt ttttgtgaa agatttaatg tcgttggtta
181 tcgcaccaat ctttgttaga ttggttctgg aatgatttc tcgtgtgttg gacgaggaag
241 acgatagcgc aaagtaagct gctatcaaca cacacgctag aagtcgcaac tagtgtaaaa
301 aaaagcaatc ctattcgccg taggattgct ttttgtgta tctgtacgat ttaatgtcgt
361 ttcgcacttt tagtatagca ttttttatt ttgggtcaag ttttgtgact atgcaggaat
421 tggtaaagaa tacagtggta gcaattttca tcgatgctat tttattaata aaatagtagt
481 agaaaaatat atttattgat aaacttatag ttatgaatct gtatagttag ttataataat
541 tggatTTTTT ttaggaaaat ttgagctttt gaattgaata agaaggagtg attttatgga
601 tttaaagtac aatgtttttg gtaattcaat gtattctttg aaagaaatgg agctaattca
661 actagcttca caa
```

//

Les bases de données

➤ Harmonisation des fiches de données

En résumé, une fiche comporte de nombreuses informations :

Locus	Identificateur (nom et taille de la séquence)
Definition	Description de la séquence
Accession / version	Numéro d'accès dans la base
Keyword / Source / Organism / Reference / Authors / Title / Journal	Informations diverses (taxonomie, publications...)
Features	Caractéristiques de la séquence / produits d'expression
Origin	Séquence (par blocs de caractères / par lignes)
//	Fin de l'entrée dans la base

Les bases de données

- Format commun de manipulation des données : le format FASTA (Fast – alignment)

Objectif : **manipuler facilement** des séquences dans les bases de données, à l'aide d'un **format universel**, compatibles avec les traitements de texte (sous forme de fichier texte), ou par copier – coller.

Exemple de la fiche précédente du plasmide d'*E.faecalis* en format FASTA :

```
>gi|460758767|gb|KC297657.1| Enterococcus faecalis strain 493/96 plasmid OrfC
gene, partial cds; RNAI (rnaI) and RNAII (rnaII) genes, complete sequence; and
OrfD gene, partial cds
AGTATAAATGTTCCCTGGTGTAAACGAGTTTACACGCTTAGAAAAGATCATAAAACAGCTAGATAAATTGT
TGAAGGTTTTATTATTGAATTGGCAGAATTTCAATCTATGCTATAATTAATACGGCAGCTCGCCTCGATT
GGAGGTGTGTTATTTGTGAAAGATTTAATGTCGTTGGTTATCGCACCAATCTTTGTAGGATTGGTTCCTGG
AAATGATTTCTCGTGTGTTGGACGAGGAAGACGATAGCCGAAAGTAAGCTGCTATCAACACACACGCTAG
AAGTCGCAACTAGTGTAAAAAAAAGCAATCCTATTTCGCCGTAGGATTGCTTTTTGTGTTATCTGTACGAT
TTAATGTCGTTTTCGCACTTTTAGTATAGCATATTTTTATTTTGGGTCAAGTTTTGTGACTATGCAGGAAT
TGGTAAAGAATACAGTGGTAGCAATTTTCATCGATGCTATTTTATTAATAAAATAGTAGAGAAAAATAT
ATTTATTGATAAACTTATAGTTATGAATCTGTATAGTTAGTTATAATAAATTGGTATTTTTTTAGGAAAAT
TTGAGCTTTTGAATTGAATAAGAAGGAGTGATTTTATGGATTTAAAGTACAATGTTTTTGGTAATTC AAT
GTATTCCTTTGAAAGAAATGGAGCTAATTC AACTAGCTTCACAA
```

Les bases de données

- Format commun de manipulation des données :
le format FASTA (Fast – alignment)

Remarques :

-Les bases nucléotidiques ne référencient que des **monobrans d'ADN** (même si la séquence soumise est de l'ADN bicaténaire ou de l'ARN)

→ **la séquence est toujours dans le sens 5'P – 3'OH**

-Les séquences nucléotidiques selon le degré de précision de l'enregistrement seront écrites le plus souvent avec **A, T, C et G** et/ou avec **R, Y** (base pu**R**ique A et G / base p**Y**rimidique C et T) et/ou **K, M** (base **K**eto G et T / base a**M**ino A et C).

-Les bases protéiques sont référencées :

→ **avec la séquence dans le sens N vers C terminal**

→ **avec le symboles d'acides aminés à 1 lettre**

Uniprot

- **Uniprot**
 - <http://www.uniprot.org/>
- Les données proviennent de deux sources
 - La base de données SwissProt, remplie manuellement à partir de publications
 - La traduction automatique des séquences d'ADN issues de la base EMBL
 - Il est plus facile de séquencer un gène qu'une protéine !
- Recherche principalement par nom de gène ou de protéine, et ou nom d'espèce

UniProtKB



UniProtKB

atpase human

x Advanced 1

BLAST Align Retrieve/ID Mapping

Help Contact

Show help for UniProtKB

b Basket

Results

Filter byⁱ

S Reviewed
(1,873)
Swiss-Prot

t Unreviewed
(40,176)
TrEMBL

Popular
organisms

Human (3,063)

Zebrafish (569)

Mouse (219)

S. cerevisiae
(42)

Rat (16)

Other organisms

Go

Search
terms

e Columns t BLAST i Align = Download b Add to basket 1 to 25 of 42,049 Show 25

<input type="checkbox"/>	Entry	Entry name		Protein names	Gene names	Organism	Length	e
<input type="checkbox"/>	P00846	ATP6_HUMAN	S	ATP synthase subunit a	MT-ATP6 , ATP6, ATPASE6, MTATP6	Homo sapiens (Human)	226	
<input type="checkbox"/>	P03928	ATP8_HUMAN	S	ATP synthase protein 8	MT-ATP8 , ATP8, ATPASE8, MTATP8	Homo sapiens (Human)	68	
<input type="checkbox"/>	P25685	DNJB1_HUMAN	S	DnaJ homolog subfamily B member 1	DNAJB1 , DNAJ1, HDJ1, HSPF1	Homo sapiens (Human)	340	
<input type="checkbox"/>	Q9UBS4	DJB11_HUMAN	S	DnaJ homolog subfamily B member 11	DNAJB11 , EDJ, ERJ3, HDJ9, PSEC0121, UNQ537/PRO1080	Homo sapiens (Human)	358	
<input type="checkbox"/>	Q15645	PCH2_HUMAN	S	Pachytene checkpoint protein 2 homo...	TRIP13 , PCH2	Homo sapiens (Human)	432	
<input type="checkbox"/>	Q9Y2G3	AT11B_HUMAN	S	Probable phospholipid-transporting ...	ATP11B , ATPIF, ATPIR, KIAA0956	Homo sapiens	1,177	

PDB

- PDB (BrookHaven Protein DataBank)
 - <http://www.rcsb.org>
 - **Séquences et structures des protéines**
 - Visualisation en 3D
 - Les données proviennent de cristallographie, de RMN,...
 - Pour certaines protéines, plusieurs structures sont disponibles
 - Structure de la protéine seule ou avec ligand
 - Structure de la protéine dans différents milieux
 - Structure obtenue avec des méthodes expérimentales différentes

WHAT'S NEW | [HELP](#) | [PRINT](#)

PDB ID or keyword

map kinase

Search

Advanced Search

[Home](#) Hide

[News & Publications](#)
[Policies](#)
[FAQ](#)
[Contact](#)
[Feedback](#)
[About Us](#)

[Deposition](#) Hide

[All Deposit Services](#)
[Electron Microscopy](#)
[NMR](#)
[Validation Server](#)
[BioSync Beamline](#)
[Related Tools](#)

[Search](#) Hide

[Advanced Search](#)
[Latest Release](#)
[Latest Publications](#)
[Sequence Search](#)
[Ligand Search](#)
[Unreleased Entries](#)
[Browse Database](#)
[Histograms](#)

Explorer:

315 Structure Hits

9 Unreleased Structures

158 Citations

231 Ligand Hits

107 Web Page Hits

[GO Hits](#)

[SCOP Hits](#)

[CATH Hits](#)

Advanced Keyword Query for: **MAP KINASE**

Query Options:

Display/Download:

Generate Reports:

Sort by:

Results per Page:

Displaying results 1 - 10 of 315 total | Page 1 of 32

3FXW



High resolution crystal structure of mitogen-activated protein kinase-activated protein kinase 3/inhibitor 2 complex

Characteristics

Release Date: 26-Jan-2010 Exp. Method: X-RAY DIFFRACTION
Resolution: 2.00 Å

Classification

Transferase

Compound

Molecule: MAP kinase-activated protein kinase 3

Polymer: 1 **Type:** polypeptide(L)

Length: 336

Chains: A

EC#: [2.7.11.1](#)

Fragment: Kinase domain, UNP residues 33-349

Authors

[Cheng, R.K.Y.](#) , [Barker, J.](#) , [Palan, S.](#) , [Felicetti, B.](#) ,
[Whittaker, M.](#) , [Hesterkamp, T.](#)



3KGA



Crystal structure of MAPKAP kinase 2 (MK2) complexed with a potent 3-aminopyrazole ATP site inhibitor

Crystal structure of MAPKAP kinase 2 (MK2) complexed with a potent 3-aminopyrazole ATP site inhibitor

3KGA

- Display Files
- Download Files

Share this Page

Sequence Display

The sequence display provides a graphical representation of the UniProtKB, PDB - ATOM and PDB - SEQRES sequences. Different 3rd party annotations can be graphically mapped on the sequence and displayed in the Jmol viewer.

The structure 3KGA has in total 1 chains.

Currently viewing **unique chains** only. [show all chains](#)

Sequence & Structure Relationships

Display Jmol

Enable Jmol to view annotations in 3D.

Chain A : MAP kinase-activated protein kinase 2

FASTA | [Sequence & DSSP](#) | [Image](#)

Polymer 1

Length: 299 residues

Chain Type: polypeptide(L)

Reference: [UniProtKB P49137](#)

Display Parameters

Currently displayed: **SEQRES**

sequence.

[Display external \(UniProtKB\) sequence](#)

Mouse over an annotation to see more details. Click annotation to enable Jmol.

Annotations

Add Annotations

Secondary Structure:DSSP
[hide] [reference]

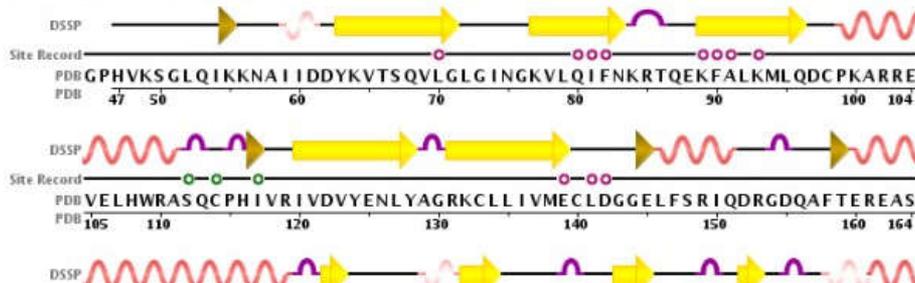
38% helical (15 helices; 116 residues)

19% beta sheet (14 strands; 57 residues)

Structural Feature:Site Record
[hide] [reference]

3KGA_A_AC2_16 BINDING SITE FOR RESIDUE LX9 A 365 (SOFTWARE)

3KGA_A_AC1_4 BINDING SITE FOR RESIDUE MG A 1 (SOFTWARE)



**PDB :
structure
secondaires**

PDB : séquence des protéines

Summary **Sequence** Annotations Seq. Similarity 3D Similarity Literature Biol. & Chem. Methods Geometry Links

Crystal structure of MAPKAP kinase 2 (MK2) complexed with a potent 3-aminopyrazole ATP site inhibitor

3KGA

Display Files
Download Files

Share this Page

Sequence Display

The sequence display provides a graphical representation of the UniProtKB, PDB - ATOM and PDB - SEQRES sequences. Different 3rd party annotations can be graphically mapped on the sequence and displayed in the Jmol viewer.

The structure 3KGA has in total 1 chains.

Currently viewing **unique chains** only. [show all chains](#)

Chain A : MAP kinase-activated protein kinase 2

FASTA Sequence & DSSP | Image

Polymer 1

Length: 299 residues

Chain Type: polypeptide(L)

Reference: [UniProtKB P49137](#)

Sequence & Structure Relationships

Display Jmol

Enable Jmol to view annotations in 3D.

Display Parameters

Currently displayed: **SEQRES** sequence.

[Display external \(UniProtKB\) sequence](#)

3KGA_A.fasta.txt [Lecture seule] (/tmp) - gedit

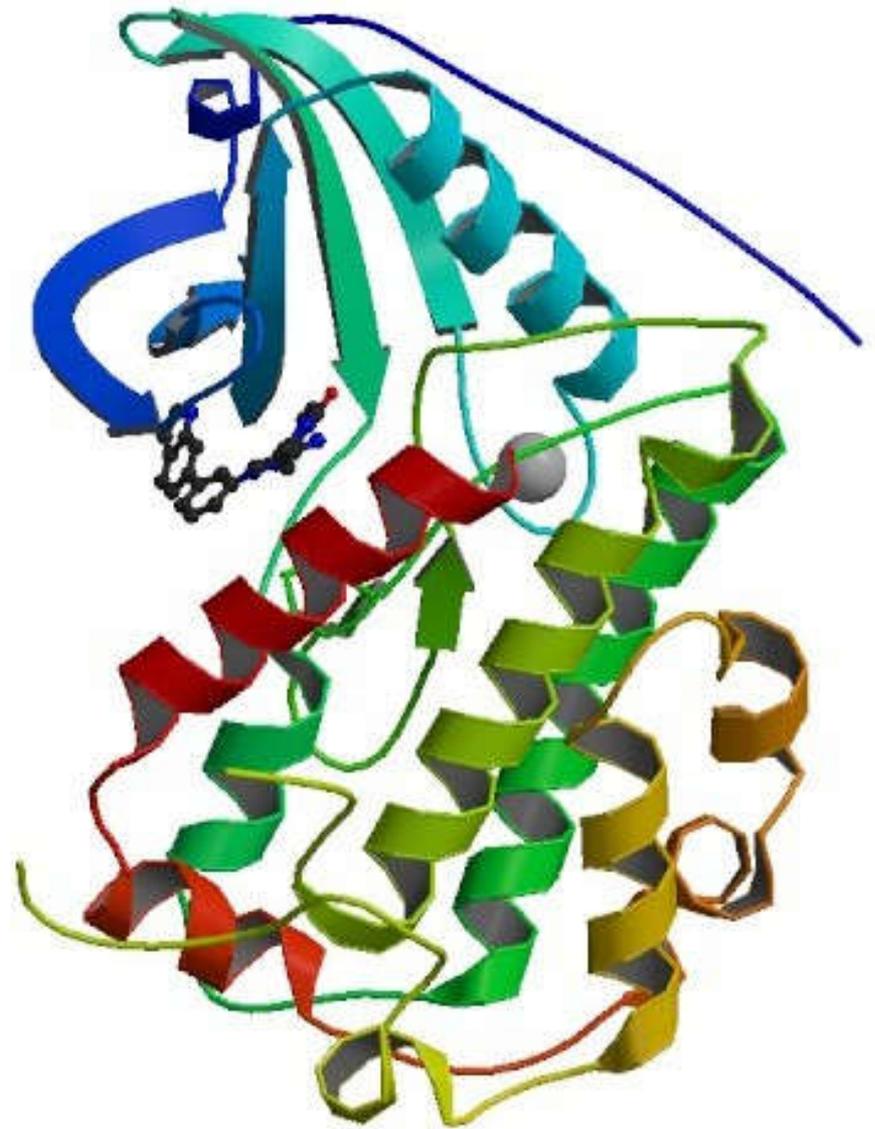
Fichier Édition Affichage Recherche Outils Documents Aide

Ouvrir Enregistrer Annuler

3KGA_A.fasta.txt

```
>3KGA: A | PDBID | CHAIN | SEQUENCE
GPHVKSGLQIKKNAIIDDYKVTSQLVGLGINGKVLQIFNKRTQEKFALKMLQDCPKARREVELHWRASQCPHIVRIVDVY
ENLYAGRKCLLIVMECLDGGELFSRIQDRGDQAFTEREASEIMKSIGEAIQYLHSINIAHRDVKPENLLYTSKRPNAILK
LTDGFGAKETTGEKYDKSCDMWSLGVIMYILLCGYPPFYSNHGLAISPGMKTRIRMGQYEFNPEWSEVSEEKMLIRNL
LKTEPTQRMTITEFMNHPWIMQSTKVPQTPLHTSRVLKEDKERWEDVKEEMTSALATMR
```

PDB : structure tertiaires



Biological Assembly Image for 3KGA

Crystal structure of MAPKAP kinase 2 (MK2) complexed with a potent 3-aminopyrazole ATP site inhibitor

Domaine protéique:

Un **domaine protéique** est une partie d'une protéine capable d'adopter une structure de manière autonome ou partiellement autonome du reste de la molécule. C'est un élément modulaire de la structure des protéines qui peuvent ainsi être composées de l'assemblage de plusieurs de ces domaines

PROSITE

Permet d'accéder au motif du domaine

Technical section

PROSITE methods (with tools and information) covered by this documentation:

ABC_TM1, PS50928; ABC transporter integral membrane type-1 domain profile (MATRIX)

- Sequences in UniProtKB/Swiss-Prot known to belong to this class: 490
 - detected by PS50928: 490 (true positives)
 - undetected by PS50928: 0 (false negative or 'partial')
- Other sequence(s) in UniProtKB/Swiss-Prot detected by PS50928: NONE.
- Domain architecture view of Swiss-Prot proteins matching PS50928



- Retrieve an alignment of UniProtKB/Swiss-Prot true positive hits:
Clustal format, color, condensed view / Clustal format, color / Clustal format, plain text / Fasta format
- Retrieve the sequence logo from the alignment
- Taxonomic distribution of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS50928
- Retrieve a list of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS50928
- Scan UniProtKB (Swiss-Prot and/or TrEMBL) entries against PS50928
- View ligand binding statistics of PS50928
- Matching PDB structures: 2ONK 2R6G 3DHW 3FH6 ... [ALL]

Recherche des « architectures »
(= suite de domaines)
dans les protéines de la base
Swiss Prot

Recherche toutes les protéines
Ayant ce domaine dans Uniprot

PROSITE

Permet d'accéder au motif du domaine

Technical section

PROSITE methods (with tools and information) covered by this documentation:

ABC_TM1, PS50928; ABC transporter integral membrane type-1 domain profile (MATRIX)

- Sequences in UniProtKB/Swiss-Prot known to belong to this class: 490
 - detected by PS50928: 490 (true positives)
 - undetected by PS50928: 0 (false negative or 'partial')
- Other sequence(s) in UniProtKB/Swiss-Prot detected by PS50928: NONE.
- Domain architecture view of Swiss-Prot proteins matching PS50928

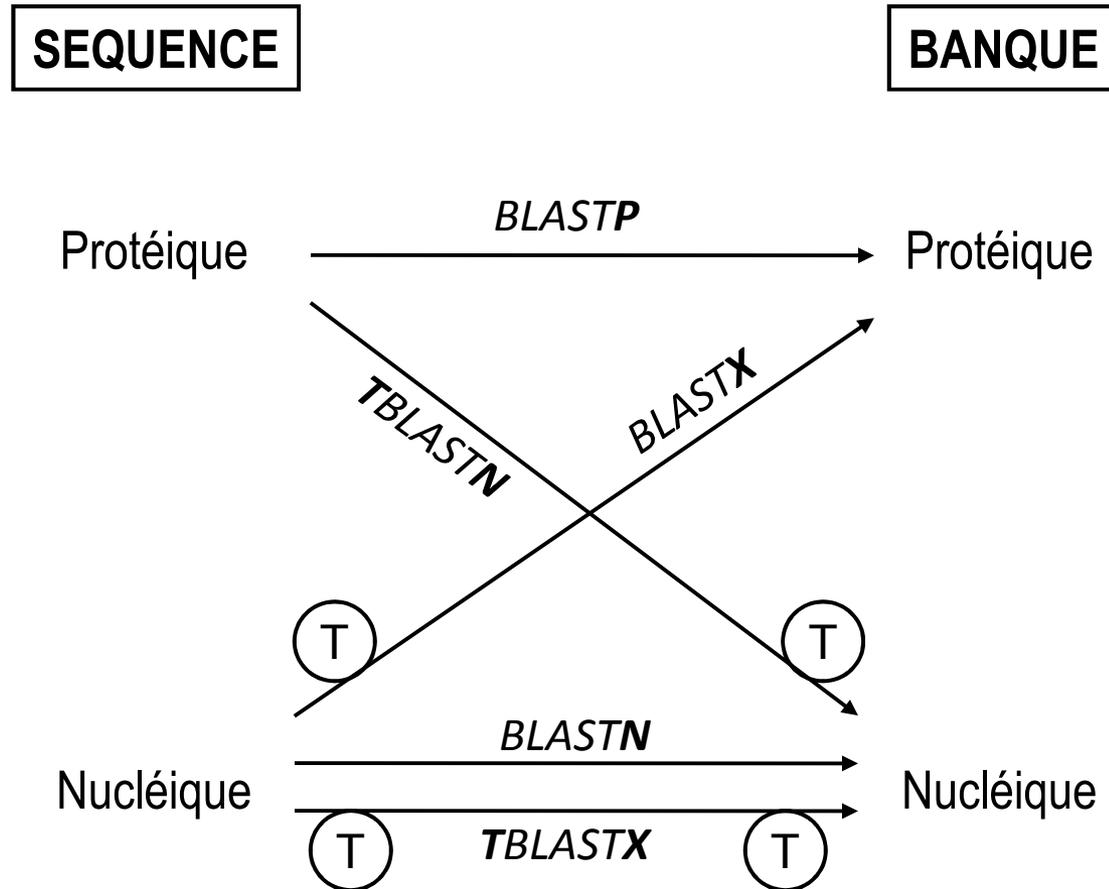


- Retrieve an alignment of UniProtKB/Swiss-Prot true positive hits:
Clustal format, color, condensed view / Clustal format, color / Clustal format, plain text / Fasta format
- Retrieve the sequence logo from the alignment
- Taxonomic distribution of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS50928
- Retrieve a list of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS50928
- Scan UniProtKB (Swiss-Prot and/or TrEMBL) entries against PS50928
- View ligand binding statistics of PS50928
- Matching PDB structures: 2ONK 2R6G 3DHW 3FH6 ... [ALL]

Recherche des « architectures »
(= suite de domaines)
dans les protéines de la base
Swiss Prot

Recherche toutes les protéines
Ayant ce domaine dans Uniprot

BLAST: Choix du programme



- ❖ blastn, de nucléotides, séquence nucléotidique contre une base de données de séquences nucléotidiques.
- ❖ blastp, de protéines, séquence de protéine contre une base de données de séquences de protéines.
- ❖ blastx, séquence nucléotidique traduite en séquence de protéine contre une base de données de séquences de protéines.
- ❖ tblastn, séquence de protéine contre une base de données de séquences nucléotidiques traduites en séquences de protéines.
- ❖ tblastx, séquence nucléotidique traduite en séquence de protéine contre une base de données de séquences nucléotidiques traduites en séquences de protéines.

Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

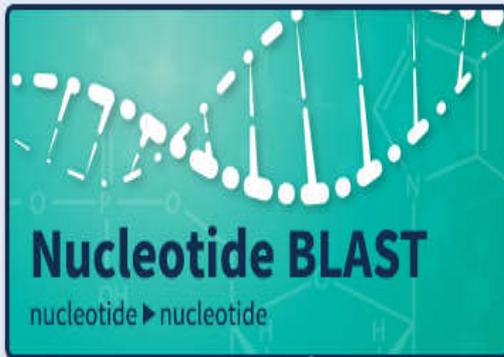
NEWS

[Search Betacoronavirus Database](#)

We have created a new BLAST database focused on the SARS-CoV-2 (Severe acute respiratory syndrome coronavirus 2) Sequences. For further detail please visit [NCBI GenBank](#).

Mon, 03 Feb 2020 10:00:00 EST [More BLAST news...](#)

Web BLAST



Nucleotide BLAST
nucleotide ► nucleotide



blastx
translated nucleotide ► protein



tblastn
protein ► translated nucleotide



Protein BLAST
protein ► protein



Cliquer ici pour choisir le Blast nucléique

Cliquer ici pour choisir le Blast protéique

On entre la séquence à chercher

Standard Nucleotide BLAST

[blastn](#) [blastp](#) [blastx](#) [tblastn](#) [tblastx](#)

Enter Query Sequence

BLASTN programs search nucleotide databases using a nucleotide query. [more...](#)

[Reset page](#) [Bookmark](#)

Enter accession number(s), gi(s), or FASTA sequence(s)

[Clear](#)

Query subrange

From

To

```
GGCCATCCACAATGTTGTTTCATGCTATTATTCTGCATCAACAACAAAAACCAACAACAACCATCGAGC
CAGGTCCTCTTCCAACAGCCTCTGCAACAATATCCATTAGGCCAGGGCTCCTCCGGCCATCTCAGCAA
ACCCACAGGCCCGGGCTCTGTCCAGCCTCAACAACCTGCCAGTTCGAGGAAATAAGGAACCTAGCGCT
ACAGACGCTACCCGCAATGTGCAATGTCTACATCCCTCCATATTGCACCATCGGCCATTGGCATCTTC
GGTACTAACTG
```

BLAST results will be displayed in a new format by default

You can always switch back to the Traditional Results page.



Or, upload file

Choisir un fichier Aucun fichier choisi

Job Title

Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database Standard databases (nr etc.) rRNA/ITS databases Genomic + transcript databases Betacoronavirus

Nucleotide collection (nr/nt)

Limit by

Organism BioProjectID WGS Project

Cliquer sur "Back to traditional page"

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: <https://www.coronavirus.gov>.
Get the latest research from NIH: <https://www.nih.gov/coronavirus>.

BLAST® » blastn suite » results for RID-8P3NW11G01R

[Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#)

[← Edit Search](#)

[Save Search](#)

[Search Summary](#) ▼

[? How to read this report?](#)

[▶ BLAST Help Video](#)

[↶ Back to Traditional Results Page](#)

Job Title	KC660359.1 Triticum aestivum clone pGli70...
RID	8P3NW11G01R <i>Search expires on 04-07 19:13 pm</i> Download All ▼
Program	BLASTN ? Citation ▼
Database	nt See details ▼
Query ID	lcl Query_64535
Description	KC660359.1 Triticum aestivum clone pGli70 gliadin (gli) gene ...
Molecule type	dna
Query Length	852

Filter Results

Organism *only top 20 will appear*

exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity

E value

Query Coverage

to

to

to



[Resubmit](#) [Save Search Strategies](#) [Formatting options](#) [Download](#) [YouTube](#) [How to read this page](#) [Blast report description](#) **NEW** [Click here to use the new BLAST results page](#)

KC660359.1 *Triticum aestivum* clone pGli70...

RID [8P3NW11G01R](#) (Expires on 04-07 19:13 pm)

Query ID [Id|Query_64535](#)

Description [KC660359.1 Triticum aestivum clone pGli70 gliadin \(gli\) gene, complete cds](#)

Sequence type [dna](#)

Sequence Length [852](#)

Database Name [nt](#)

Description [Nucleotide collection \(nt\)](#)

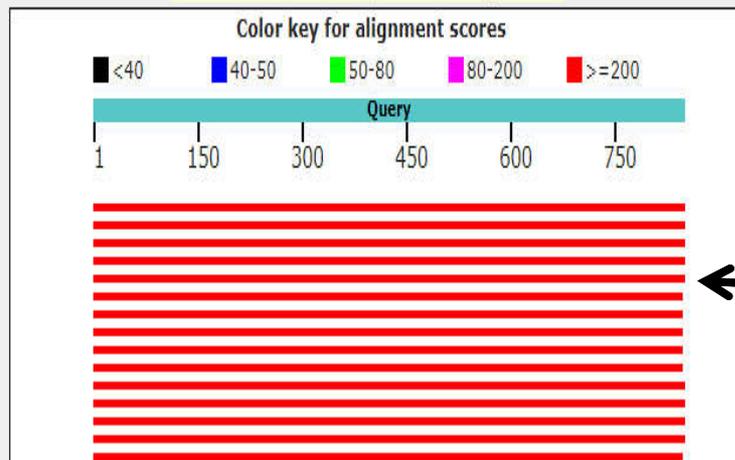
Program [BLASTN 2.10.0+](#) [Citation](#)

Reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#) [MSA viewer](#)

[Search Summary](#)

Distribution of the top 100 Blast Hits on 100 subject sequences

Mouse over to see the title, click to show alignments



Nombres de hits
(nombre de
comparaison est =
100)

Répartition des hits
(comparaisons) en
fonction du score

Job title: KC660359.1 Triticum aestivum clone pGli70...

RID [BP3NW11G01R](#) (Expires on 04-07 19:13 pm)

Query ID [ld|Query_64535](#)

Description [KC660359.1 Triticum aestivum clone pGli70 gliadin \(gli\) gene, complete cds](#)

Molecule type [dna](#)

Query Length [852](#)

Database Name [nt](#)

Description [Nucleotide collection \(nt\)](#)

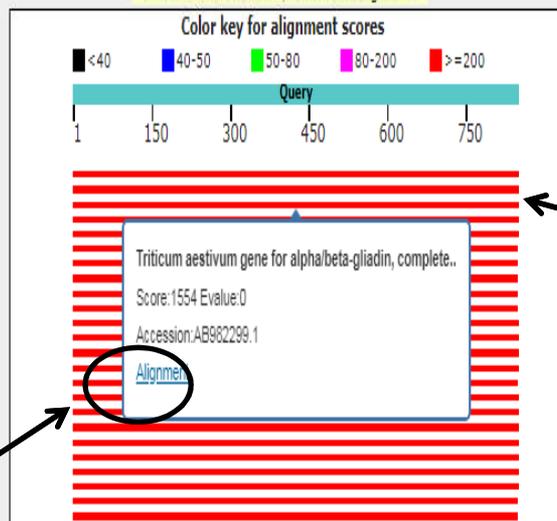
Program [BLASTN 2.10.0+](#) [Citation](#)

Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#) [MSA viewer](#)

Graphic Summary

Distribution of the top 100 Blast Hits on 100 subject sequences

Mouse over to see the title, click to show alignments



1-Cliquez 2 fois sur la 3^{ème} ligne rouge (chaque ligne correspond à une comparaison entre ma séquence et l'autre de la banque

2-Cliquez sur Alignment

E-value

% d'identité entre les 2 séquences
(Requête et sujet)

Triticum aestivum gene for alpha/beta-gliadin, complete cds, cultivar: Chinese Spring, clone: TAC2
Sequence ID: [AB982299.1](#) Length: 2161 Number of Matches: 1

Range 1: 596 to 1450 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
1554 bits(841)	0.0	851/855(99%)	3/855(0%)	Plus/Plus

Query	1	ATGAAGACCTTTCTCATCCTTGCCTCCTTGTCTATCGTGCGACCACCGCCACAAC	60
Sbjct	596	ATGAAGACCTTTCTCATCCTTGCCTCCTTGTCTATCGTGCGACCACCGCCACAAC	655
Query	61	GTTAGAGTTCAGTGCCACAATTGCAGCCACAACATCCATCTCAGCAACAGCCACAAGAG	120
Sbjct	656	GTTAGAGTTCAGTGCCACAATTGCAGCCACAACATCCATCTCAGCAACAGCCACAAGAG	715
Query	121	CAAGTTCATTGGTACAACAACAACAATTTCTAGGGCAGCAACAACCATTTCCACCACAA	180
Sbjct	716	CAAGTTCATTGGTACAACAACAACAATTTCTAGGGCAGCAACAACCATTTCCACCACAA	775
Query	181	CAACCATATCCACAGCCGCAACCATTTCCATCACAACAACCATATCTGCAACTACAACCA	240
Sbjct	776	CAACCATATCCACAGCCGCAACCATTTCCATCACAACAACCATATCTGCAACTACAACCA	835
Query	241	TTTCCGCAGCCGCAACTACCATATTCGAGCCACAACCATTTGACCCACAACAACCATAT	300
Sbjct	836	TTTCCGCAGCCGCAACTACCATATTCGAGCCACAACCATTTGACCCACAACAACCATAT	895
Query	301	CCACAACCGCAACCACAGTATTGcaaccacaacaaccaatttcacagcggcagcagcag	360
Sbjct	896	CCACAACCGCAACCACAGTATTGCAACCACAACAACCAATTTACAGCAGCAGCAGCAG	955
Query	361	cag---caacaacaacaacaacaacaacaacaataatccttcaacaattttgcaacaaca	417
Sbjct	956	CAGCAACAACAACAACAACAACAACAACAACAATCCTTCAACAATTTGCAACAACAA	1015
Query	418	cTGATTCCATGCATGGATGTTGTATTGCAGCAACACAACATAGCGCATGGAAGATCACAA	477
Sbjct	1016	CTGATTCCATGCATGGATGTTGTATTGCAGCAACACAACATAGCGCATGGAAGATCACAA	1075

Gaps: nombre de
délétion dans la
séquence requête

Suite d'alignement (comparaison)

```
Query 418 cTGATTCCATGCATGGATGTTGTATTGCAGCAACACAACATAGCGCATGGAAGATCACAA 477
          |||
Sbjct 1016 CTGATTCCATGCATGGATGTTGTATTGCAGCAACACAACATAGCGCATGGAAGATCACAA 1075

Query 478 GTTTTGCAACAAAGTACTTACCAGCTGTTACAAGAATTGTGTTGTCAGCACCTATGGCAG 537
          |||
Sbjct 1076 GTTTTGCAACAAAGTACTTACCAGCTGTTACAAGAATTGTGTTGTCAGCACCTATGGCAG 1135

Query 538 ATCCCTGAGCAGTCGCAGTGCCAGGCCATCCACAATGTTGTTTCATGCTATTATTCTGCAT 597
          |||
Sbjct 1136 ATCCCTGAGCAGTCGCAGTGCCAGGCCATCCACAATGTTGTTTCATGCTATTATTCTGCAT 1195

Query 598 CAACAACAAAAACCACAACAACAACCATCGAGCCAGGTCTCCTTCCAACAGCCTCTGCAA 657
          |||
Sbjct 1196 CAACAACAAAAACCACAACAACAACCATCGAGCCAGGTCTCCTTCCAACAGCCTCTGCAA 1255

Query 658 CAATATCCATTAGGCCAGGGCTCCTTCCGGCCATCTCAGCAAAACCCACAGGCCCGGGGC 717
          |||
Sbjct 1256 CAATATCCATTAGGCCAGGGCTCCTTCCGGCCATCTCAGCAAAACCCACAGGCCCGGGGC 1315

Query 718 TCTGTCCAGCCTCAACAACCTGCCCCAGTTCGAGGAAATAAGGAACCTAGCGCTACAGACG 777
          |||
Sbjct 1316 TCTGTCCAGCCTCAACAACCTGCCCCAGTTCGAGGAAATAAGGAACCTAGCGCTACAGACG 1375

Query 778 CTACCCGCAATGTGCAATGTCTACATCCCTCCATATTGCACCATCGCGCCATTTGGCATC 837
          |||
Sbjct 1376 CTACCCGCAATGTGCAATGTCTACATCCCTCCATATTGCACCATCGCGCCATTTGGCATC 1435

Query 838 TTCGGTACTAACTGA 852
          |||
Sbjct 1436 TTCGGTACTAACTGA 1450
```

La taille de notre séquence (Query: 852) et la taille de la séquence de la banque (Sbjet: 1450).

-Pour le Blast protein, on va suivre les même étapes précédentes que pour le Blast Nucleotide.

-Le résultat apparait est presque identique à celui du Blast nucleotide mais à la place des nucleotides on utilise des acides aminés (AKWYTR.....) car les séquences à comparer sont des séquences protéiques.

alpha-gliadin protein [Triticum monococcum]

Sequence ID: [AGI15866.1](#) Length: 285 Number of Matches: 1

Range 1: 1 to 285 [GenPept](#) [Graphics](#)

▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
499 bits(1285)	3e-177	Compositional matrix adjust.	281/285(99%)	282/285(98%)	2/285(0%)

Query	1	MKTFLILALLAIVATTATTAVRVPVPLQVPLVQQQFLGQQPFPPQ	60
Sbjct	1	MKTFLILALLAIVATTATTAVRVPVPLQVPLVQQQFLGQQPFPPQ	60
Query	61	QYPYQPQPFPSQQPYLQLQPFQPQLPYSQPQFRPQQYPYQPQPYSPQQPISQRQQQ	120
Sbjct	61	QYPYQPQPFPSQQPYLQLQPFQPQLPYSQPQFRPQQYPYQPQPYSPQQPISQ+QQQ	120
Query	121	QQQQQQQQQQ - - ILQQILQQQLIPCMDVVLQQHNI AHGRSQVLQQSTYQLLQELCCQHLW	178
Sbjct	121	QQQQQQQQQQ ILQQILQQQLIPCMDVVLQQHNI AHGRSQVLQQSTYQLLQELCCQHLW	180
Query	179	QIPEQSQCQAIHNVVHAIILHQQQKQQPSSQVSFQQPLQQYPLGQGSFRPSQQNPQAR	238
Sbjct	181	QIPEQSQCQAIHNVVHAIIPHQQQKQQPSSQVSFQQPLQQYPLGQGSFRPSQQNPQAR	240
Query	239	GSVQPQQLPQFEEIRNLALQTL PAMCNVYIPPYCTIAPFGIFGTN	283
Sbjct	241	GSVQPQQLPQFEEIRNLALQTL PAMCNVYIPPYCTIAPFGIFGTN	285

← C'est une confirmation des acides aminés et non une 3^{ème} séquence

Interprétation des résultats

Les résultats d'un Blast sont en 4 parties :

1. Un schéma graphique : les meilleurs résultats sont en rouge, suivis par les verts, les plus mauvais en bleu et noir)

2. La liste des meilleurs hits (liste de numéro d'accession cliquable), avec le E-value et le Bit score

- **Bit score** : mesure statistique de la validité de l'alignement : plus la valeur est élevée, plus les 2 séquences sont similaires. En-dessous de 50, le résultat n'est pas fiable.

- **E-value** : ?Expectation value? : estime la chance que vous auriez de trouver le même résultat par hasard. Plus la valeur est faible (proche de 0), meilleure est votre résultat. Au-dessus de 0.001, vos résultats ne sont pas bons.

3. **Alignements** : comparaison de votre séquence appelée (requête en français, query en anglais) avec la séquence la plus proche que Blast a trouvé appelée (Sujet, subject).