

TP Bioinformatique

**TP 2 : Recherche d'une séquence nucléique associée à une pathologie (Genbank)**

**Les étapes :**

Lancer une recherche de Genbank (base de données nucléotidiques) sur le moteur de recherche Google

Environ 10 800 000 résultats (0,40 secondes)

[www.ncbi.nlm.nih.gov/genbank](http://www.ncbi.nlm.nih.gov/genbank) Traduire cette page

**GenBank Overview - NCBI - NIH**

20 oct. 2020 — What is GenBank? GenBank® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences (Nucleic ...

Vous avez consulté cette page de nombreuses fois. Date de la dernière visite : 10/11/20.

**Sample GenBank Record**  
This page presents an annotated sample GenBank record ...

**Submission Types**  
Standard. GenBank accepts mRNA or genomic sequence ...

**Statistics**  
Notes on GenBank statistics. The following table lists the number ...

**International Nucleotide**  
International Nucleotide Sequence Database Collaboration.

**GenBank**

La GenBank est une base de données de la séquences d'ADN, comprenant toutes les séquences de nucléotides publiquement disponibles et leur traduction en protéines. Wikinédia

Cliquer sur le premier résultat : Genbank Overview-NCBI- NIH

Entrer le mot clé convenable à votre séquence dans la zone de recherche, puis cliquer sur «Search».

NCBI Resources How To

Nucleotide Nucleotide glucose-6-phosphate dehydrogenase deficiency |  Create alert Advanced

**COVID-19 is an emerging, rapidly evolving situation.**  
Get the latest public health information from CDC: <https://www.coronavirus.gov>  
Get the latest research information from NIH: <https://www.nih.gov/coronavirus>  
Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>

Species Animals (26)  
Plants (8)  
Fungi (26)  
Bacteria (61)  
Archaea (1)  
Customize ...

Molecule types genomic DNA/RNA (114)  
mRNA (9)  
Customize ...

Source databases INSDC (GenBank) (103)  
RefSeq (19)  
Customize ...

Sequence Type Nucleotide (123)

Summary 20 per page Sort by Default order ▾

Items: 1 to 20 of 123

<< First < Prev Page 1 of 7 Next > Last >>

**Homo sapiens isolate E0068V glucose-6-phosphate dehydrogenase (G6PD).gene..partial cds**

1. 340 bp linear DNA  
Accession: MF796526.1 GI: 1279397414  
Protein PubMed Taxonomy  
GenBank FASTA Graphics PopSet

**Homo sapiens isolate E0119V glucose-6-phosphate dehydrogenase (G6PD).gene..partial cds**

2. 340 bp linear DNA  
Accession: MF796525.1 GI: 1279397412  
Protein PubMed Taxonomy  
GenBank FASTA Graphics PopSet

Send to: Filters: Manage Filters

Results by taxon Top Organisms Tree  
Homo sapiens (21)  
Saccharomyces cerevisiae (13)  
Bacillus subtilis (9)  
Pseudomonas aeruginosa (8)  
Arabidopsis thaliana (6)  
All other taxa (66)  
More...

Find related data Database: Select

Les résultats de l'information sont organisés sous forme de fichier, chaque fichier est appelé « une entrée »

## Homo sapiens isolate E0068V glucose-6-phosphate dehydrogenase (G6PD) gene, partial cds

GenBank: MF796526.1

[FASTA](#) [Graphics](#) [PopSet](#)

Go to:

Locus: MF796526      Definition: Homo sapiens isolate E0068V glucose-6-phosphate dehydrogenase (G6PD) gene, partial cds.

Accession: MF796526      Version: MF796526.1

Keywords:

Source: Homo sapiens (human)  
Organism: [Homo sapiens](#)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.  
Reference: 1 (bases 1 to 340)  
Authors: Carter,T.E., Mekonnen,S.K., Lopez,K., Bonnell,V., Damodaran,L., Aseffa,A. and Janies,D.A.  
Title: Glucose-6-Phosphate Dehydrogenase Deficiency Genetic Variants in Malaria Patients in Southwestern Ethiopia  
Journal: Am. J. Trop. Med. Hyg. (2017) In press  
Pubmed: [29141760](#)  
Remark:  
Publication Status: Available-Online prior to print  
2 (bases 1 to 340)  
Authors: Carter,T.E., Mekonnen,S.K., Lopez,K., Bonnell,V., Damodaran,L., Aseffa,A. and Janies,D.A.  
Title: Direct Submission  
Journal: Submitted (31-AUG-2017) Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, 9201 University City Blvd, Charlotte, NC 28223, USA  
Comment: ##Assembly-Data-START##  
Sequencing Technology :: Sanger dideoxy sequencing

Origin:

```
1 cacacatgg ctcaaagaga ggggctgaca tctgtctgtg tgtctgtctg tccgtgtctc
61 ccaggcccacc ccagaggaga agctcaagct ggaggacttc ttggcccgca actcctatgt
121 ggctggccag tacgatgtac cagcctcta ccagcgcctc aacagccaca tgaatgccct
181 ccacctgggg tcacaggcca accgcctctt ctacctggcc ttggcccccga ccgtctacga
241 ggccgtcacc aagaacattc acgagtccctg catgagccag atgtaaggct tgccgttgcc
301 ctcccttccc gcctgccagg ctggccagg cagtgtcccc
```

Séquence nucléotidique

Taille et type de molécule, division de l'organisme, date de dépôt de séquence

Organisme +taxonomie

Références bibliographiques et nom des auteurs

## Base de données OMIM

OMIM est une base de données bibliographique qui donne des informations sur les maladies héréditaires chez l'être humain.

About Statistics Downloads Contact Us MIMmatch Donate Help ?

**50 YEARS**  
**OMIM**  
Human Genetics Knowledge for the World

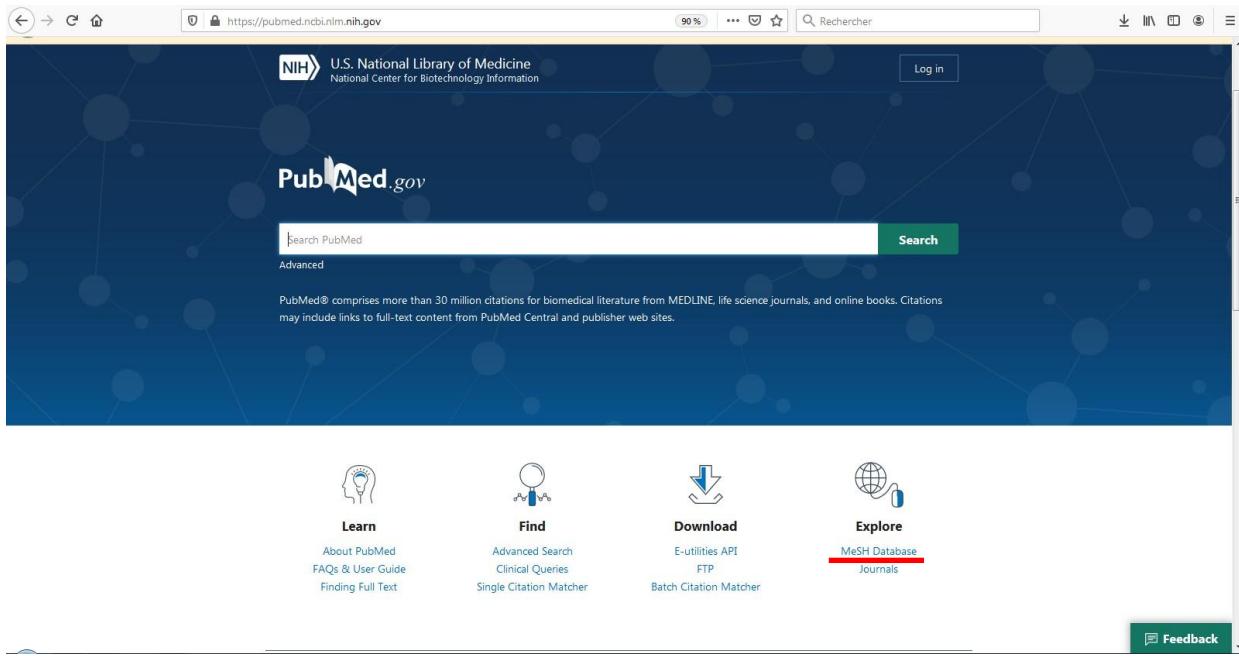
**OMIM®**  
**Online Mendelian Inheritance in Man®**  
**An Online Catalog of Human Genes and Genetic Disorders**  
Updated December 29, 2020

Search OMIM for clinical features, phenotypes, genes, and more...

Advanced Search : OMIM, Clinical Synopses, Gene Map  
Need help? : Example Searches, OMIM Search Help, OMIM Video Tutorials  
Mirror site : <https://mirror.omim.org>

OMIM is supported by a grant from NHGRI, licensing fees, and generous contributions from people like you.

## **TP3 : Bases de données bibliographiques (Pubmed) (TP réalisé)**



## **TP 4 : Comparaison des séquences (Alignement)**

### **1- Comparaison d'une séquence nucléique avec une base de données nucléique**

- Le programme utilisé est Blast de Genbank sur le site du NCBI

- Il existe 5 types de BLAST (blastn, blastp, blastx, tblastn, tblastx)

Pour comparer la séquence nucléotidique avec les séquences nucléotidiques de la base de données genbank on utilise le Blastn (Nucleotide BLAST).

Lancer une recherche BLAST pour la séquence suivante :

ggggggggtt	tccggattta	ttggcgtaa	gcgagcgcag	gcggttttt	aagtctgtatg	tgaaagccct
cggcttaacc	gaggaagcgc	atcgaaaaact	ggaaacttg	agtgcagaag	aggacagtgg	aactccatgt
gtagcggtaga	aatgcgtaga	tatatggaag	aacaccagtg	gcgaaggcg	ctgtctggtc	tgtaactgac
gctgaggctc	gaaagcatgg	gtagcgaaca	ggattagata	ccctggtagt	ccatgccgt	aacgatgaat
gctagggttt	ggagggtttc	cgccttcag	tgcccgagct	aacgcattaa	gcattccgccc	tggggagtac
gaccgcaagg	ttgaaactca	aaggaattga	cggggcccg	cacaagcggt	ggagcatgtg	gtttaattcg
aagcaacgcg	aagaacctta	ccaggtcttg	acatcttttg	atcacctgag	agatcagggtt	tccccttcgg
ggcggaaaatg	acaggtggtg	catgggtgtc	gtcagctcgt	gtcgtgagat	gttgggttaa	gtccccgcaac
gagcgcaacc	cttatgacta	gttgcgcagca	tttagttggg	cactctagta	agactgcccgg	tgaccaaaccg
gaggaagggtg	gggatgacgt	caaatcatca	tgcccttat	gacctgggt	acacacgtgc	tacaatggat
ggtacaacga	gttgcgcagac	cgcgagggtca	agctaatactc	ttaaagccat	tctcagttcg	gactgttaggc
tgcaactcgc	ctacacgaag	tcggaatcgc	tagtaatcgc	ggatcagcac	ggccgcgggtga	atacgttccc
ggcccttgta	cacccssca					

## Les étapes :

### Lancer une recherche de blast sur le moteur de recherche Google

The screenshot shows a Google search results page for the query "blast". A red arrow points to the search bar at the top. Another red arrow points to the first search result, which is a link to the NCBI Basic Local Alignment Search Tool (BLAST) website.

Environ 432 000 000 résultats (0,39 secondes)

blast.ncbi.nlm.nih.gov Traduire cette page

**BLAST: Basic Local Alignment Search Tool** ⓘ

The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to ...

**BLASTn** ⓘ BLASTN programs search nucleotide databases using a ...

**BLASTp** ⓘ BLASTP programs search protein databases using a protein query ...

**Nucleotide BLAST** ⓘ Search database Nucleotide collection (nr/nr) using Blastn ...

**Blastx** ⓘ Start typing in the text box, then select your taxid. Use the "plus" ...

Autres résultats pour nih.gov >

Videos

Basic Local Alignment Search Tool

BLAST est une méthode de recherche heuristique utilisée en bioinformatique. 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. Wikipedia

Système d'exploitation : Type Unix, Linux, macOS et Microsoft Windows

Type : Outil bioinformatique

Dernière version : 2.9.0+ (1er avril 2019)

Licence : Domaine public

Développé par : Altschul S.F., Gish W., Miller E.W., Lipman D.J., NCBI

SafeSearch activé

Cliquer sur le premier résultat : Blast : Basic local alignment Search Tool  
<https://blast.ncbi.nlm.nih.gov/Blast.cgi>

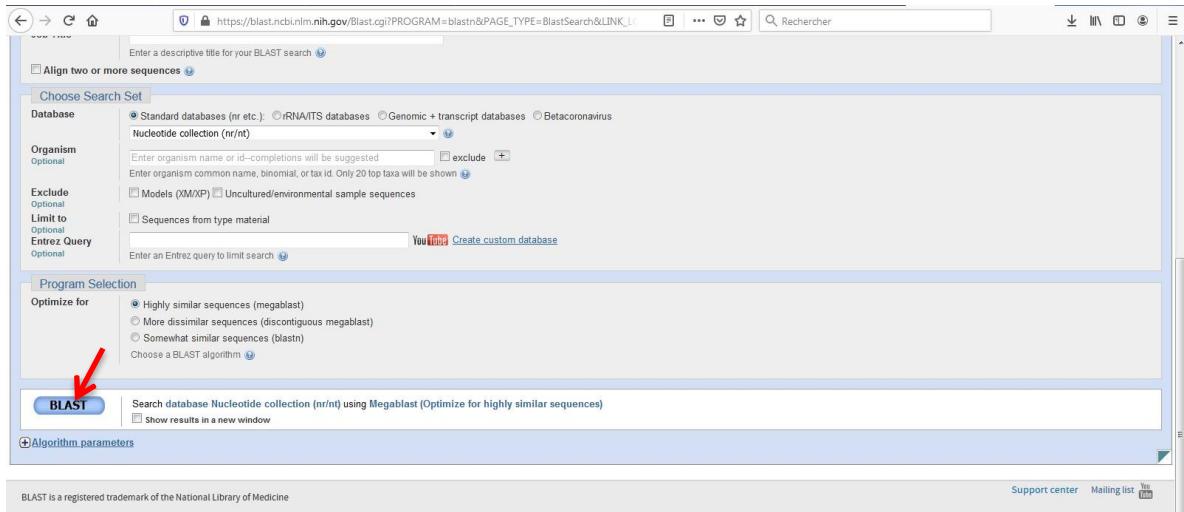
The screenshot shows the NCBI BLAST homepage. A red arrow points to the "Web BLAST" section, which features a large green button labeled "Nucleotide BLAST" with the text "nucleotide ▶ nucleotide". To the right, there are links for "blastx" (translated nucleotide ▶ protein) and "tblastn" (protein ▶ translated nucleotide). Further down, there is a section for "BLAST Genomes" with a search bar.

Choisir le nucleotide BLAST

The screenshot shows the "Standard Nucleotide BLAST" search interface. A red arrow points to the "Enter Query Sequence" field, which contains the placeholder text "Enter accession number(s), g(i)s, or FASTA sequence(s)". Below this field are options for "Or, upload file" and "Job Title". Further down, there is a checkbox for "Align two or more sequences". On the right side, a message states "BLAST results will be displayed in a new format by default". At the bottom, there are sections for "Choose Search Set", "Organism", "Exclude", "Limit to", "Entrez Query", and "Program Selection".

## Entrer la séquence dans la fenêtre

## Cliquer sur Blast



Le programme blast va comparer la séquence (la séquence requête ou query) avec les séquences de la base.

The screenshot shows the NCBI BLAST search results page. The search parameters are: Job Title (RID-7PWP4M50016), Program (BLASTN), Database (nt), Query ID (lcl|Query\_62969). The results table has columns: Description, Max Score, Total Score, Query Cover, E value, Per Ident, and Accession. A red box highlights the 'Feedback' button in the top right corner of the results table. To the right, a red box contains the text: 'Plus le score est grand, plus la qualité est bonne et plus le pourcentage d'identité est élevé'.

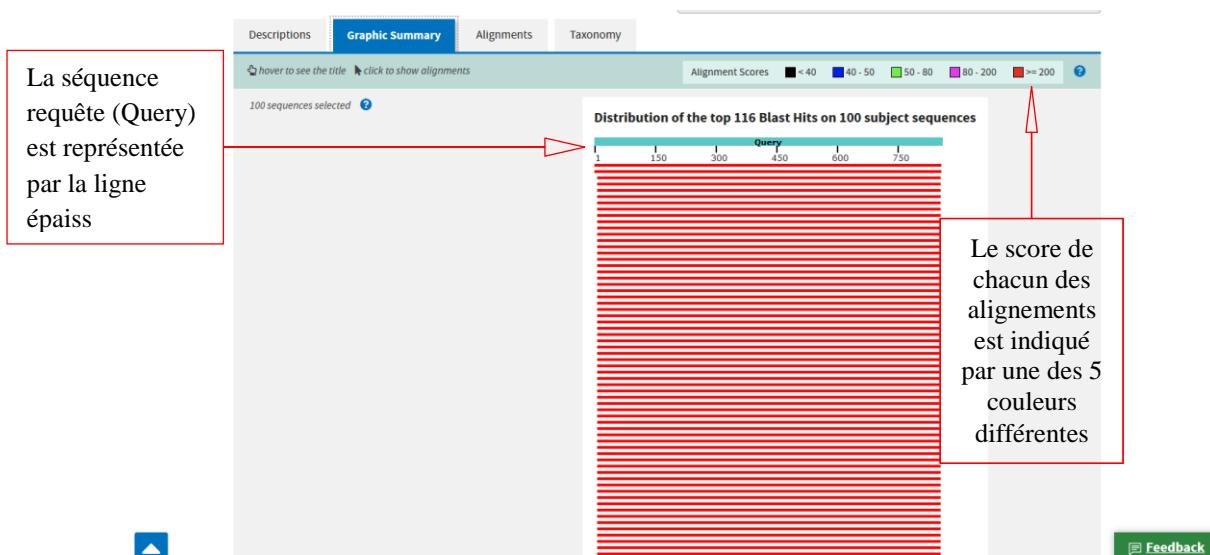
Description	Max Score	Total Score	Query Cover	E value	Per Ident	Accession
Lactobacillus casei strain B1 16S ribosomal RNA gene _partial sequence	1570	1570	99%	0.0	99.77%	KY764324_1
Lactobacillus casei strain BM10 16S ribosomal RNA gene _partial sequence	1561	1561	99%	0.0	99.53%	KY764329_1
Lactobacillus casei strain 2SA3 16S ribosomal RNA gene _partial sequence	1557	1557	98%	0.0	99.88%	KC967210_1
Lactobacillus fermentum strain MBO003 16S ribosomal RNA gene _partial sequence	1554	1554	98%	0.0	99.76%	MT044253_1
Lactobacillus paracasei strain ES81 16S ribosomal RNA gene _partial sequence	1554	1554	98%	0.0	99.76%	MT044251_1
Lactobacillus paracasei strain NS16 16S ribosomal RNA gene _partial sequence	1554	1554	98%	0.0	99.76%	MT044244_1
Lactobacillus paracasei strain NS12 16S ribosomal RNA gene _partial sequence	1554	1554	98%	0.0	99.76%	MT044244_1
Lactobacillus paracasei strain ABK 16S ribosomal RNA gene _partial sequence	1554	1554	98%	0.0	99.76%	MN994625_1
Lactobacillus paracasei strain CACC 556 chromosome _complete genome	1554	7750	98%	0.0	99.76%	CP048003_1
Lactobacillus paracasei strain TD 062 chromosome _complete genome	1554	7759	98%	0.0	99.76%	CP044361_1
Lactobacillus paracasei strain 1490 16S ribosomal RNA gene _partial sequence	1554	1554	98%	0.0	99.76%	MN750022_1
Lactobacillus paracasei strain MG5191 16S ribosomal RNA gene _partial sequence	1554	1554	98%	0.0	99.76%	MN749945_1
Lactobacillus paracasei strain MG5190 16S ribosomal RNA gene _partial sequence	1554	1554	98%	0.0	99.76%	MN749944_1
Lactobacillus paracasei strain MG5189 16S ribosomal RNA gene _partial sequence	1554	1554	98%	0.0	99.76%	MN749943_1
Lactobacillus paracasei strain MG5179 16S ribosomal RNA gene _partial sequence	1554	1554	98%	0.0	99.76%	MN749939_1
Lactobacillus paracasei strain f3 16S ribosomal RNA gene _partial sequence	1554	1554	98%	0.0	99.76%	MN744244_1
Lactobacillus paracasei strain f2 16S ribosomal RNA gene _partial sequence	1554	1554	98%	0.0	99.76%	MT051212_1
Lactobacillus paracasei strain MG5310 16S ribosomal RNA gene _partial sequence	1554	1554	98%	0.0	99.76%	MN704668_1
Lactobacillus paracasei strain MG5245 16S ribosomal RNA gene _partial sequence	1554	1554	98%	0.0	99.76%	MN688621_1
Lactobacillus paracasei strain LMT13-95 16S ribosomal RNA gene _partial sequence	1554	1554	98%	0.0	99.76%	LR650498_1
uncultured bacterium partial 16S rRNA gene	1554	1554	98%	0.0	99.76%	MN59715_1
Lactobacillus paracasei strain SHTR-1 16S ribosomal RNA gene _partial sequence	1554	1554	98%	0.0	99.76%	MN59715_1

## Les résultats de BLAST

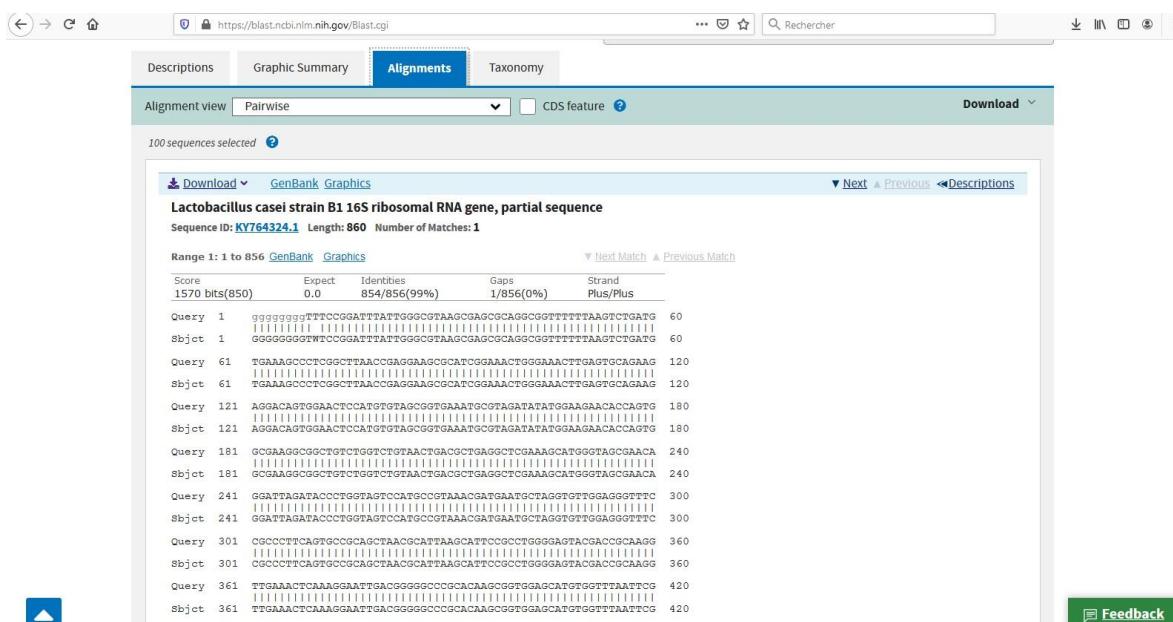
La page résultat est divisée en 3 parties :

- La liste des séquences avec leur score et leur E-value

- Une vue graphique générale (Graphic summary) des séquences résultats avec différentes couleurs



- Résultat détaillé, fournissant pour chaque séquence résultat (subject), l'alignement avec la séquence requête (query).

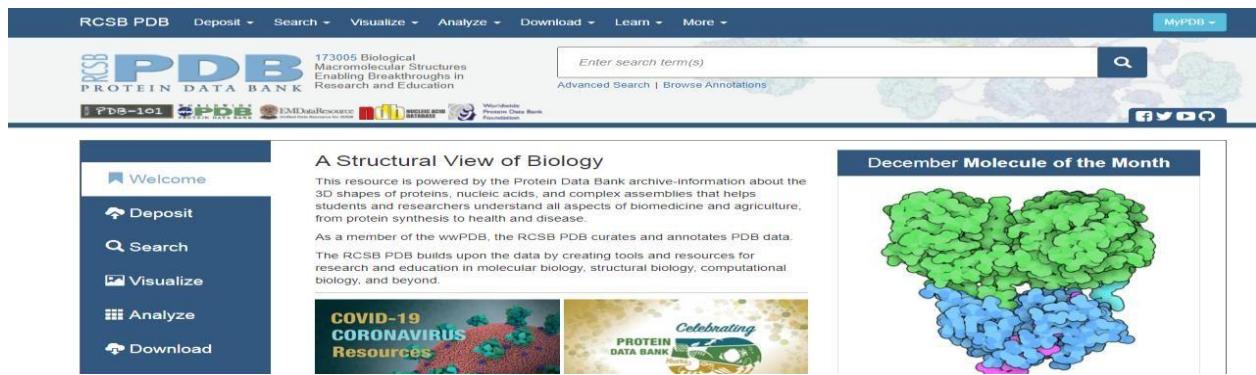


## **TP5. Analyse structurale et modélisation moléculaire**

La modélisation moléculaire permet la prédition de la structure 3D d'une protéine à partir de sa structure primaire (l'enchaînement des acides aminés qui la composent), en prenant en compte les différentes propriétés physico-chimiques des acides aminés.

**La Protein Data Bank (PDB)** est la base de données dédiée au stockage de la structure des protéines obtenues expérimentalement (par RMN, cristallographie et diffraction de RX, et par microscopie électronique).

**Base de données PDB** : conçues pour la structure tridimensionnelle des protéines



## **Applications**

- 1- Trouver dans la base Genbank la séquence dont le numéro d'accession est le suivant : MW494397.  
Quelles sont les informations accompagnant la séquence.
- 2- Trouver la structure tridimensionnelle de cette protéine dont le numéro identifiant est le suivant : 2GTL