

# TUTORIELS

Recherche de séquences dans le NCBI

ORFinder

Recherche de séquences dans UniProt

Recherche de séquences homologues par BLAST

Alignement de séquences par Clustal Omega

Choix de sites: Gblocks

Construction d'arbres phylogénétiques avec IQ-Tree

Edition d'arbres avec iTOL

AlphaFold Protein Structure Database

ColabFold – utilisation d'AlphaFold 2

Foldseek – recherche des structures similaires dans le AFDB

http://www.ncbi.nlm.nih.gov/

The screenshot shows the NCBI homepage with a dark blue header containing the logo and navigation links for 'Resources' and 'How To'. Below the header is a search bar with a dropdown menu set to 'All Databases' and a 'Search' button. A left-hand navigation menu lists various categories such as 'NCBI Home', 'Resource List (A-Z)', 'All Resources', 'Chemicals & Bioassays', 'Data & Software', 'DNA & RNA', 'Domains & Structures', 'Genes & Expression', 'Genetics & Medicine', 'Genomes & Maps', 'Homology', 'Literature', 'Proteins', 'Sequence Analysis', 'Taxonomy', 'Training & Tutorials', and 'Variation'. The main content area features a 'Welcome to NCBI' message and a grid of six service categories: 'Submit' (Deposit data or manuscripts into NCBI databases), 'Download' (Transfer NCBI data to your computer), 'Learn' (Find help documents, attend a class or watch a tutorial), 'Develop' (Use NCBI APIs and code libraries to build applications), 'Analyze' (Identify an NCBI tool for your data analysis task), and 'Research' (Explore NCBI research and collaborative projects). On the right side, there are sections for 'Popular Resources' (PubMed, Booksshelf, PubMed Central, PubMed Health, BLAST, Nucleotide, Genome, SNP, Gene, Protein, PubChem) and 'NCBI Announcements' (December 17th webinar: 1000 Genomes Project I, On Thursday December staff will demonstrate how, Registration open for Dec NCBI Minute: "New Face Search in dbGaP Provide

## Search NCBI databases

[Help](#)

### Results found in 27 databases for "YrdC"

#### Literature

<a href="#">Books</a>	1	books and reports
<a href="#">MeSH</a>	2	ontology used for PubMed indexing
<a href="#">NLM Catalog</a>	0	books, journals and more in the NLM Collections
<a href="#">PubMed</a>	29	scientific & medical abstracts/citations
<a href="#">PubMed Central</a>	137	full-text journal articles

#### Health

<a href="#">ClinVar</a>	1	human variations of clinical significance
<a href="#">dbGaP</a>	0	genotype/phenotype interaction studies
<a href="#">GTR</a>	1	genetic testing registry
<a href="#">MedGen</a>	0	medical genetics literature and links
<a href="#">OMIM</a>	1	online mendelian inheritance in man
<a href="#">PubMed Health</a>	0	clinical effectiveness, disease and drug reports

#### Genomes

<a href="#">Assembly</a>	0	genome assembly information
<a href="#">BioProject</a>	0	biological projects providing data to NCBI
<a href="#">BioSample</a>	0	descriptions of biological source materials
<a href="#">Clone</a>	1,195	genomic and cDNA clones
<a href="#">dbVar</a>	60	genome structural variation studies
<a href="#">Genome</a>	25	genome sequencing projects by organism

#### Genes

<a href="#">EST</a>	13	expressed sequence tag sequences
<a href="#">Gene</a>	4,764	collected information about gene loci
<a href="#">GEO DataSets</a>	0	functional genomics studies
<a href="#">GEO Profiles</a>	3,822	gene expression and molecular abundance profiles
<a href="#">HomoloGene</a>	1	homologous gene sets for selected organisms
<a href="#">PopSet</a>	4	sequence sets from phylogenetic and population studies
<a href="#">UniGene</a>	29	clusters of expressed transcripts

#### Proteins

<a href="#">Conserved Domains</a>	4	conserved protein domains
<a href="#">Protein</a>	104,816	protein sequences
<a href="#">Protein Clusters</a>	87	sequence similarity-based protein clusters
<a href="#">Structure</a>	34	experimentally-determined biomolecular structures

#### Chemicals

<a href="#">BioSystems</a>	151	molecular pathways with links to genes, proteins and chemicals
<a href="#">PubChem BioAssay</a>	1	bioactivity screening studies
<a href="#">PubChem Compound</a>	0	chemical information with structures, information and links
<a href="#">PubChem Substance</a>	80	deposited substance and chemical information

NCBI Resources How To Sign in to NCBI

Gene    [Create RSS](#) [Create alert](#) [Advanced](#) [Help](#)

Gene sources: Genomic, Organelles, Plasmids

Categories: Alternatively spliced, Annotated genes, Non-coding, Protein-coding, Pseudogene

Sequence content: CCDS, Ensembl, RefSeq, RefSeqGene

Status:  Current

Chromosome locations: more...

[Clear all](#)

[Show additional filters](#)

Tabular 20 per page Sort by Relevance Send to: [Hide sidebar >>](#)

Did you mean YrdC as a gene symbol?  
Search Gene for [YrdC](#) as a symbol.

**Search results**

Items: 1 to 20 of 2370 << First < Prev Page 1 of 119 Next > Last >>

[See also 2394 discontinued or replaced items.](#)

Name/Gene ID	Description	Location	Aliases	MIM
<input type="checkbox"/> <a href="#">YRDC</a> ID: 79693	YrdC N6-threonylcarbamoyltransferase domain containing [ <i>Homo sapiens</i> (human)]	Chromosome 1, NC_000001.11 (37802942..37808193, complement)	DRIP3, IRIP, SUA5	612276
<input type="checkbox"/> <a href="#">Yrdc</a> ID: 230734	YrdC domain containing (E.coli) [ <i>Mus musculus</i> (house mouse)]	Chromosome 4, NC_000070.6 (124850759..124855242)	AV303379, BC023823, IRIP, ITIP	
<input type="checkbox"/> <a href="#">Yrdc</a> ID: 319113	YrdC N(6)-threonylcarbamoyltransferase domain containing [ <i>Rattus norvegicus</i> (Norway rat)]	Chromosome 5, NC_005104.4 (142845265..142850110)	Isrip	
<input type="checkbox"/> <a href="#">yrdc</a> ID: 797343	YrdC N(6)-threonylcarbamoyltransferase domain containing [ <i>Danio rerio</i> (zebrafish)]	Chromosome 16, NC_007127.6 (4353606..4361437)	zgc:162301	
<input type="checkbox"/> <a href="#">yrdc</a> ID: 548554	YrdC N(6)-threonylcarbamoyltransferase domain containing [ <i>Xenopus tropicalis</i> (tropical clawed frog)]	Chromosome 2, NC_030678.1 (57893501..57903444)	drip3, irip, sua5	
<input type="checkbox"/> <a href="#">YRDC</a> ID: 456764	YrdC N6-threonylcarbamoyltransferase domain containing [ <i>Pan troglodytes</i> (chimpanzee)]	Chromosome 1, NC_006468.4 (37172228..37177593, complement)		
<input type="checkbox"/> <a href="#">YRDC</a> ID: 419610	YrdC N6-threonylcarbamoyltransferase domain containing [ <i>Gallus gallus</i> (chicken)]	Chromosome 23, NC_006110.4 (3571061..3573778)	null	

**Filters:** [Manage Filters](#)

**Results by taxon**

Top Organisms [\[Tree\]](#)

- Metallosphaera sedula (14)
- Homo sapiens (12)
- Nicotiana tabacum (9)
- Nicotiana glauca (8)
- Nicotiana glauca (7)
- Streptomyces virginiae (7)
- All other taxa (2320)

[More...](#)

**Find related data**

Database:

**Search details**

YrdC[All Fields] AND alive[prop]

[See more...](#)

**Recent activity**

Numéro d'accèsion, ID

Descriptif

Name/Gene ID	Description	Location	Aliases	MIM
<input type="checkbox"/> <a href="#">YRDC</a> ID: 79693	yrDC N6-threonylcarbamoyltransferase domain containing [ <i>Homo sapiens</i> (human)]	Chromosome 1, NC_000001.11 (37802942..37808193, complement)	DRIP3, IRIP, SUA5	612276
<input type="checkbox"/> <a href="#">Yrdc</a> ID: 230734	yrDC domain containing ( <i>E.coli</i> ) [ <i>Mus musculus</i> (house mouse)]	Chromosome 4, NC_000070.6 (124850759..124855242)	AV303379, BC023823, IRIP, ITIP	
<input type="checkbox"/> <a href="#">Yrdc</a> ID: 319113	yrDC N(6)-threonylcarbamoyltransferase domain containing [ <i>Rattus norvegicus</i> (Norway rat)]	Chromosome 5, NC_005104.4 (142845265..142850110)	Isrip	
<input type="checkbox"/> <a href="#">yrdc</a> ID: 797343	yrDC N(6)-threonylcarbamoyltransferase domain containing [ <i>Danio rerio</i> (zebrafish)]	Chromosome 16, NC_007127.6 (4353606..4361437)	zgc:162301	
<input type="checkbox"/> <a href="#">yrdc</a> ID: 548554	yrDC N(6)-threonylcarbamoyltransferase domain containing [ <i>Xenopus tropicalis</i> (tropical clawed frog)]	Chromosome 2, NC_030678.1 (57893501..57903444)	drip3, irip, sua5	
<input type="checkbox"/> <a href="#">YRDC</a> ID: 456764	yrDC N6-threonylcarbamoyltransferase domain containing [ <i>Pan troglodytes</i> (chimpanzee)]	Chromosome 1, NC_006468.4 (37172228..37177593, complement)		
<input type="checkbox"/> <a href="#">YRDC</a> ID: 419610	yrDC N6-threonylcarbamoyltransferase domain containing [ <i>Gallus gallus</i> (chicken)]	Chromosome 23, NC_006110.4 (3571061..3573778)	null	

NCBI Resources How To Sign in to NCBI

Gene  Search [Advanced](#) [Help](#)

Full Report Send to: [Hide sidebar >>](#)

## YRDC yrdC N6-threonylcarbamoyltransferase domain containing [ *Homo sapiens* (human) ]

Gene ID: 79693, updated on 6-Dec-2016

**Summary**

**Official Symbol** YRDC provided by HGNC

**Official Full Name** yrdC N6-threonylcarbamoyltransferase domain containing provided by HGNC

**Primary source** [HGNC:HGNC:28905](#)

**See related** [Ensembl:ENSG00000196449](#) [MIM:612276](#) [Vega:OTTHUMG00000004318](#)

**Gene type** protein coding

**RefSeq status** VALIDATED

**Organism** [Homo sapiens](#)

**Lineage** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo

**Also known as** IRIP; SUA5; DRIP3

**Orthologs** [mouse](#) [all](#)

**Genomic context**

**Location:** 1p34.3 [See YRDC in Genome Data Viewer](#) [Map Viewer](#)

**Exon count:** 5

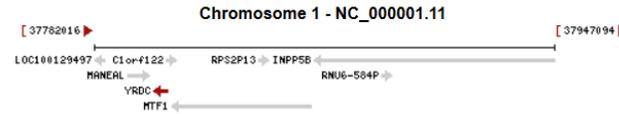
Annotation release	Status	Assembly	Chr	Location
<a href="#">108</a>	current	GRCh38.p7 ( <a href="#">GCF_000001405.33</a> )	1	NC_000001.11 (37802942..37808193, complement)
<a href="#">105</a>	previous assembly	GRCh37.p13 ( <a href="#">GCF_000001405.25</a> )	1	NC_000001.10 (38268614..38273865, complement)

**Table of contents**

- Summary
- Genomic context
- Genomic regions, transcripts, and products
- Bibliography
- Phenotypes
- Variation
- Interactions
- General gene information
  - Markers, Related pseudogene(s), Clone Names, Homology, Gene Ontology
- General protein information
- NCBI Reference Sequences (RefSeq)
- Related sequences
- Additional links

**Genome Browsers**

- Genome Data Viewer
- Map Viewer
- Variation Viewer (GRCh37.p13)
- Variation Viewer (GRCh38)

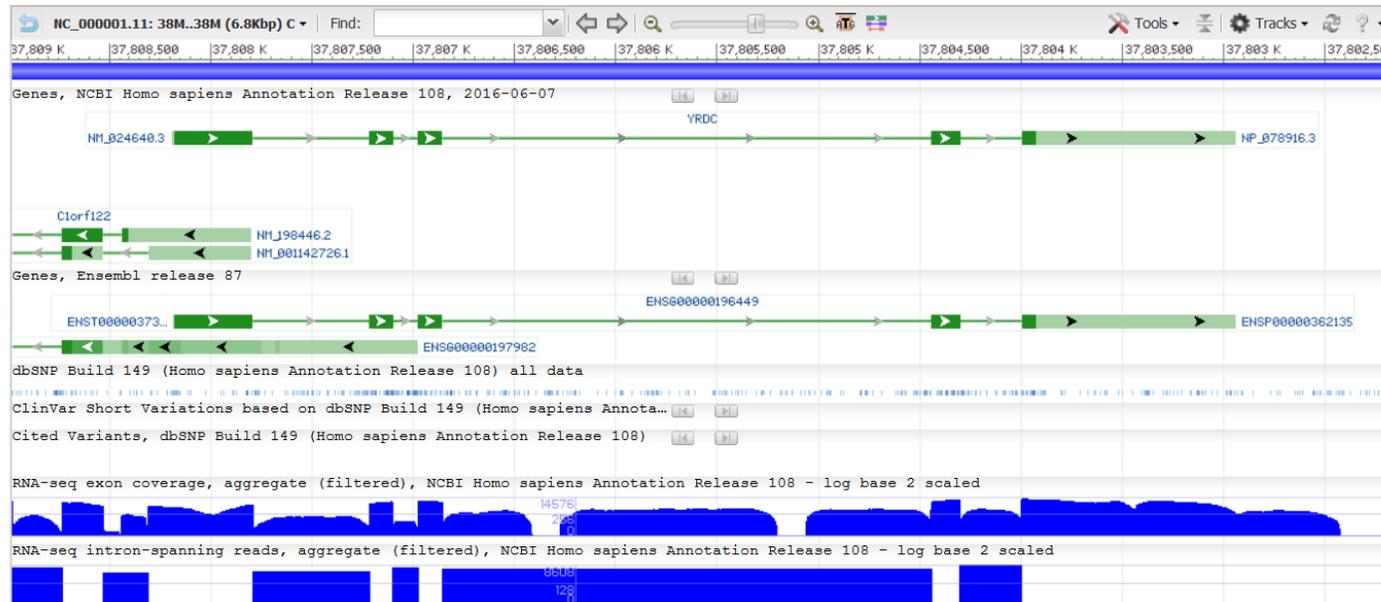


Genomic regions, transcripts, and products

Go to [reference sequence details](#)

Genomic Sequence: NC\_000001.11 Chromosome 1 Reference GRCh38.p7 Primary Assembly

Go to nucleotide: [Graphics](#) [FASTA](#) [GenBank](#)



1000 Genomes Browser (GRCh37.p13)

Ensembl

UCSC

Related information

Order cDNA clone

BioAssay by Target (List)

BioAssay by Target (Summary)

BioAssay, by Gene target

BioAssays, RNAi Target, Active

BioAssays, RNAi Target, Tested

BioProjects

CCDS

ClinVar

Conserved Domains

dbVar

EST

Full text in PMC

Full text in PMC\_nucleotide

Gene neighbors

Genes with a similar H3K4me3 profile

Genome

GEO Profiles

GTR

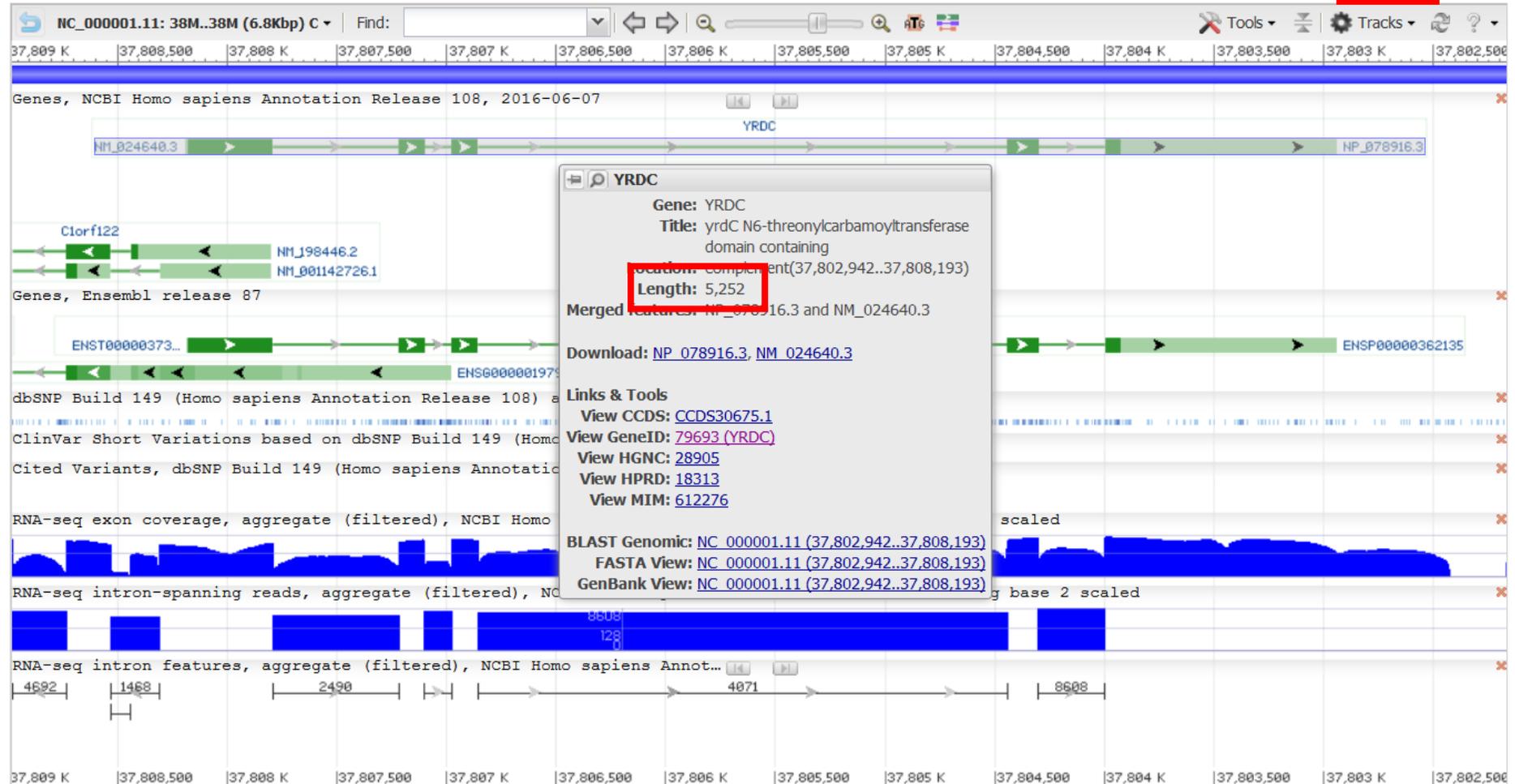
HomoloGene

Map Viewer

Go to [reference sequence details](#)

Genomic Sequence:

Go to nucleotide: [Graphics](#) **FASTA** [GenBank](#)



[NCBI](#)
[Resources](#)
[How To](#)
[Sign in to NCBI](#)

Nucleotide Nucleotide

Advanced Help

FASTA

Send:

## Homo sapiens chromosome 1, GRCh38.p7 Primary Assembly

NCBI Reference Sequence: NC\_000001.11

[GenBank](#) [Graphics](#)

>NC\_000001.11:c37808193-37802942 Homo sapiens chromosome 1, GRCh38.p7 Primary Assembly

```

CGGGCCTGGGCGGATGCTCCGGCGCGTCCGGTGCAGGGGGATGAGGGCCGCGGTGGCTGCCAGCGTGGGG
TTGAGCGAGGGGCTGCTGGCTCCCGGAGCGGTCCGCTCTCCGCCCGAGTCCCGCTCCGGCGGGCC
CCGGCGCCCGGCTGTTGCGGCTCCCGGGAGCGGGCCGTGCAGGCCCGAGCCCGGAGCGCCCGGCTG
GACCGAGGGCGTGCGGGGCCGCGTGGCCGAGCTGCGCGCCGGCCCGTGGTGGCCGCTCCCCACCGATA
CTGTACGGCCTGGCCTGCGCGGCGAGCTGCTCCGGCGCTGTCGCGCTGTGTACCGCCTCAAGGGTCGCA
GCGAGGCCAAGCCTCTGGCCGTATGCCTCGGCCGCGTGGCCGACGTCTACAGGTAAGGCCGCCCGACCC
GGCCCCGCTAGGGGCGGCACCCGCGGAGGGGTTCCGGTGACAGGCTTGGGAGACTGAGGCGCAGAGAGC
GGGAGGGGCTTGTCCAGGCTCACCAAGGAATCCGGAACCTGGAAATATGTACCAGGACTTAACGTGCCCC
GGCCCTGAACACTACATGCTGCTGGGAAGCAGCAACACACGAGAAAGGCAGGGACCCTGTTCCCTTGGCGT
ATTTGTGCTAATTGGGGAGGCCCGGGGAAAAAGCAAACAACACAGTAGCTTAATTTAATAGCGAATG
CCTTGAAGAAAATAAAACGTGGAGCTTGATCCCACTATCATGATATCACAGCGGTGCAGCTGTTGTAGTC
CCTGAGACCAGAGGCACATGGCGTCAAGCCAGGGGTGCGCCTCTGGAGCCCTGCTTTCTGCTCAAGCTC
TTAGGGAGGTGAAGCAGGGAGACAGGCGGCTCCCATTTGTAACCTCGAAGGATTCTTTCTACGTCTGCC
TAGAGTCTAGGAAAGGACCACAGTGGCCCTCAATTTGAAATGGATTATTCTTCTGTGTTTCCAGAT
ACTGCCGTGTGAGAGTACCTGAGGGGCTCCTGAAAGACCTACTGCCAGGACCAGTGACCCTGGTGATGGA
ACGCTCGGAGGAGCTCAACAAGGACCTAAACCCTTTTACGCCTGTGAGTCAAGTTTTCTTGTGTCCCC
AGACTCCAGGCCACTTGGCTTTTACTAGGATCCAGATAAGACTTACCCTCTTATCCCTTCCAACCTTTCT
CATGATCCCTTTCCCTTGCCTTACAGCTTGTAGGCATTGCGATTCTGATCATGCTTTTATGCAAGACT
TGGCTCAGATGTTTGGGGTCCGCTTGCCTCACTAGTGCCAACCTCAGCTCCAGGCCAGTTCTCTGAA
TGTCGAGGTGAGTTTACCAGTCTTCCCTGGAAATGTACCAGCGCACATACTGTTTCTGTGAATCA
GTGGCATTGATAGTACACTGGGATGGCCAGGCATGAGAGTTCTTGGGCCAACAGTGGGAAGGTGCCCCAG
CTGGCAGGGAGGTGGTGGGGGCAATATAAATTAATAAATTTGAAAGATTTTATCTCAAAGTCTGAAA
GCCAGGGGAGAAATGGTCTTTCTTCTCAGTCTTTATAATAATAGGAAAGATTATAAGAACAATGTCTCC
TGTTTTTCTGGAGATAACAGTTTCTGGAGAACACAGTTAGGCTATAAATCTTACGACAGTGTGTAAG
AAAGATGAAGAAGGCCAAGGGCAGTGGCTCATGCTGTGACCCAGCACTTTGGGAGGCCAAGCAGGTG
GATCATGTGAGGTGAGGTTTGGAGACCAGCCTGGCTAACATGGCAAACTCCATCTAATAAAAAATA
CAAAGATTAGCCAGACATGGTGGCGCATGCTGTAATCCAGCTACTTGGGAGGCTGAGGCAGGAGAATC
    
```

5252 bp – séquence non-épisé

**Change region shown**

Whole sequence  
 Selected region  
 from:  to:

**Customize view**

**Display options**

Show reverse complement

- Analyze this sequence**
- Run BLAST
  - Pick Primers
  - Highlight Sequence Features
  - Find in this Sequence

- Related information**
- Assembly
  - BioAssay by RNA target
  - BioProject
  - Components (Core)
  - Full text in PMC

→ Comment trouver la séquence codante (CDS)?

Go to [reference sequence details](#)

Genomic Sequence:

Go to nucleotide: [Graphics](#) [FASTA](#) [GenBank](#)

The screenshot displays the NCBI genome browser interface for the YRDC gene on Chromosome 1. The main view shows the gene structure with exons and introns, along with various annotations including dbSNP, ClinVar, and RNA-seq coverage. A pop-up window for the YRDC gene is open, providing detailed information:

- Gene:** YRDC
- Title:** yrdC N6-threonylcarbamoyltransferase domain containing
- Location:** complement(37,802,942..37,808,193)
- Length:** 5,252 (highlighted in red)
- Merged features:** NP\_078916.3 and NM\_024640.3
- Download:** [NP\\_078916.3](#), [NM\\_024640.3](#)
- Links & Tools:**
  - [View CCDS: CCDS30675.1](#)
  - [View GeneID: 79693 \(YRDC\)](#)
  - [View HGNC: 28905](#)
  - [View HPRD: 18313](#)
  - [View MIM: 612276](#)
- BLAST Genomic:** [NC\\_000001.11 \(37,802,942..37,808,193\)](#)
- FASTA View:** [NC\\_000001.11 \(37,802,942..37,808,193\)](#)
- GenBank View:** [NC\\_000001.11 \(37,802,942..37,808,193\)](#)

Nucleotide

Nucleotide

[Advanced](#)

[Help](#)

GenBank

Send:

## Homo sapiens chromosome 1, GRCh38.p7 Primary Assembly

NCBI Reference Sequence: NC\_000001.11

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

LOCUS NC\_000001 5252 bp DNA linear CON 06-JUN-2016

DEFINITION Homo sapiens chromosome 1, GRCh38.p7 Primary Assembly.

ACCESSION [NC\\_000001](#) REGION: complement(37802942..37808193) GPC\_000001293

VERSION NC\_000001.11

DBLINK BioProject: [PRJNA168](#)

Assembly: [GCF\\_000001405.33](#)

KEYWORDS RefSeq.

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 5252)

AUTHORS Gregory,S.G., Barlow,K.F., McLay,K.E., Kaul,R., Swarbreck,D., Dunham,A., Scott,C.E., Howe,K.L., Woodfine,K., Spencer,C.C., Jones,M.C., Gillson,C., Searle,S., Zhou,Y., Kokocinski,F., McDonald,L., Evans,R., Phillips,K., Atkinson,A., Cooper,R., Jones,C., Hall,R.E., Andrews,T.D., Lloyd,C., Ainscough,R., Almeida,J.P., Ambrose,K.D., Anderson,F., Andrew,R.W., Ashwell,R.I., Aubin,K., Babbage,A.K., Bagguley,C.L., Bailey,J., Beasley,H., Bethel,G., Bird,C.P., Bray-Allen,S., Brown,J.Y., Brown,A.J., Buckley,D., Burton,J., Bye,J., Carder,C., Chapman,J.C., Clark,S.Y., Clarke,G., Clee,C., Cobley,V., Collier,R.E., Corby,N., Coville,G.J., Davies,J., Deadman,R., Dunn,M., Earthrowl,M., Ellington,A.G., Errington,H., Frankish,A., Frankland,J., French,L., Garner,P., Garnett,J., Gay,L., Ghorri,M.R., Gibson,R., Gilby,L.M., Gillett,W., Glithero,R.J., Grafham,D.V., Griffiths,C., Griffiths-Jones,S., Grocock,R., Hammond,S., Harrison,E.S., Hart,E., Haugen,E., Heath,P.D., Holmes,S., Holt,K., Howden,P.J., Hunt,A.R.,

### Change region shown

Whole sequence (abbreviated view)

Selected region

from:  to:

### Customize view

Abbreviated view

Customize

#### Basic Features

Default features

Gene, RNA, and CDS features only

#### Display options

Show sequence

Show reverse complement

### Analyze this sequence

[Run BLAST](#)

[Pick Primers](#)

[Highlight Sequence Features](#)

### Related information

[Assembly](#)

[BioAssembly DNA Assets](#)

FEATURES	Location/Qualifiers
source	1..5252 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="1"
gene	complement(<1..393) /gene="Clorf122" /gene_synonym="ALAESM" /note="chromosome 1 open reading frame 122; Derived by automated computational analysis using gene prediction method: BestRefSeq." /db_xref="GeneID:127687" /db_xref="HGNC:HGNC:24789"
mRNA	complement(1..393) /gene="Clorf122" /gene_synonym="ALAESM" /product="chromosome 1 open reading frame 122, transcript variant 1" /note="Derived by automated computational analysis using gene prediction method: BestRefSeq." /transcript_id="NM_198446.2" /db_xref="GeneID:127687" /db_xref="HGNC:HGNC:24789"
mRNA	complement(1..393) /gene="Clorf122" /gene_synonym="ALAESM" /product="chromosome 1 open reading frame 122, transcript variant 2" /note="Derived by automated computational analysis using gene prediction method: BestRefSeq." /transcript_id="NM_001142726.1" /db_xref="GeneID:127687" /db_xref="HGNC:HGNC:24789"
gene	1..5252 /gene="YRDC" /gene_synonym="DRIP3; IRIP; SUA5" /note="yrdC N6-threonylcarbamoyltransferase domain containing; Derived by automated computational analysis using gene prediction method: BestRefSeq." /db_xref="GeneID:79693" /db_xref="HGNC:HGNC:28905" /db_xref="HPRD:18313"

Région 1-393 correspond à un cadre de lecture ouvert (ORF) nommé Clorf122 qui est présent sous deux formes (variant 1 et 2)

Cet ORF a été prédit par une analyse automatisé de séquences

```

gene      1..5252 ←
          /gene="YRDC"
          /gene_synonym="DRIP3; IRIP; SUA5"
          /note="yrdC N6-threonylcarbamoyltransferase domain
          containing; Derived by automated computational analysis
          using gene prediction method: BestRefSeq."
          /db_xref="GeneID:79693"
          /db_xref="HGNC:HGNC:28905"
          /db_xref="HPRD:18313"
          /db_xref="MIM:612276"
mRNA     join(1..402,979..1093,1218..1337,3750..3892,4197..5252)
          /gene="YRDC"
          /gene_synonym="DRIP3; IRIP; SUA5"
          /product="yrdC N6-threonylcarbamoyltransferase domain
          containing"
          /note="Derived by automated computational analysis using
          gene prediction method: BestRefSeq."
          /transcript_id="NM_024640.3"
          /db_xref="GeneID:79693"
          /db_xref="HGNC:HGNC:28905"
          /db_xref="HPRD:18313"
          /db_xref="MIM:612276"
CDS      join(14..402,979..1093,1218..1337,3750..3892,4197..4269)
          /gene="YRDC"
          /gene_synonym="DRIP3; IRIP; SUA5"
          /note="Derived by automated computational analysis using
          gene prediction method: BestRefSeq."
          /codon_start=1
          /product="yrdC domain-containing protein, mitochondrial
          precursor"
          /protein_id="NP_078916.3"
          /db_xref="CCDS:CCDS30675.1"
          /db_xref="GeneID:79693"
          /db_xref="HGNC:HGNC:28905"
          /db_xref="HPRD:18313"
          /db_xref="MIM:612276"
          /translation="MSPARRCRGMRAAVAASVGLSEGPAGSRSGRLFRPPSPAPAAPG
          ARLLRLPGSGAVQAASPERAGWTEALRAAVAELRAGAVVAVPTDTLYGLACAASCSAA
          LRAVYRLKGRSEAKPLAVCLGRVADVRYRCRVRVPEGLLKDLLPGPVTLMERSEELN
          KDLNPFPTPLVGIRIPDHAFMQDLAQMFEGPLALTSANLSSQASSLNVEEFQDLWPQLS
          LVIDGGQIGDGQSPECLRGSTVVDLSVPGKFGIIRPGCALESTTAILQKYGLLPSHA
          SYL"

```

Gène YrdC: transcrit non-épissé contient 5252 nt  
Ce gène a été prédit par une analyse automatisé de séquences

mRNA: transcrit épissé est composé de 5 exons

CDS: la séquence traduite correspond à:

nr. d'accésion correspondant à la séquence CDS



Consensus CDS protein set

## CCDS Database

EBI • HGNC • MGI • NCBI • UCSC • WTSI



PubMed
Entrez
Gene
BLAST
OMIM

Search CCDS ID for CCDS30675.1 in All Organisms and Current Releases

**CCDS**

Home  
FTP  
Process  
Releases & Statistics  
Curation Guidelines

**Collaborators**

EBI  
HGNC  
MGI  
NCBI  
UCSC  
WTSI

**Contact Us**

email CCDS

**Genome Displays**

E Ensembl  
U Genome Browser  
N Map Viewer  
V VEGA

**Related Resources**

Gene  
HomoloGene  
RefSeq  
UniGene

### Report for CCDS30675.1 (current version)

CCDS	Status	Species	Chrom.	Gene	CCDS Release	NCBI Annotation Release	Ensembl Annotation Release	Links
30675.1	Public	<i>Homo sapiens</i>	1	YRDC	20	108	85	<a href="#">H</a> <a href="#">G</a> <a href="#">C</a> <a href="#">G</a> <a href="#">S</a>

**Public since:** CCDS release 3, NCBI annotation release 36.2, Ensembl annotation release 41

**Review status:** Reviewed (by RefSeq and Havana)

### Sequence IDs included in CCDS 30675.1

Original	Current	Source	Nucleotide ID	Protein ID	Status in CCDS	Seq. Status	Links
✓	✓	EBI,WTSI	ENST00000373044	ENSP00000362135	Accepted	alive	<a href="#">N</a> <a href="#">P</a> <a href="#">N</a> <a href="#">P</a>
✓	✓	EBI,WTSI	OTTHUMT00000012470	OTTHUMP00000004511	Accepted	alive	<a href="#">N</a> <a href="#">P</a> <a href="#">N</a> <a href="#">P</a>
✓	✓	NCBI	NM_024640.3	NP_078916.3	Accepted	alive	<a href="#">N</a> <a href="#">P</a> <a href="#">N</a> <a href="#">P</a> <a href="#">B</a>

RefSeq	Length	Related UniProtKB/SwissProt	Length	Identity	Gaps	Mismatches
<a href="#">NP_078916.3</a>	279	<a href="#">Q86U90</a>	279	100%	0	0

### Chromosomal Locations for CCDS 30675.1

Assembly GRCh38.p7 ([GCF\\_000001405.33](#))

**On '-' strand of Chromosome 1 (NC\_000001.11)**

Genome Browser links: [N](#)[N](#)[U](#)[E](#)[V](#)[S](#)

## CCDS Sequence Data

Blue highlighting indicates alternating exons.

Red highlighting indicates amino acids encoded across a splice junction.

Mouse over the nucleotide or protein sequence below and click on the highlighted codon or residue to select the pair.

### Nucleotide Sequence (840 nt):

```

ATGTCTCCGGCGCGTCCGGTGCAGGGGGATGAGGGCCGCGGTGGCTGCCAGCGTGGGGTTGAGCGAGGGGC
CTGCTGGCTCCCGGAGCGGTGCCTCTTCCGCCCCGAGTCCCCTCCGGCGGCCCCCGGCGCCCGGCT
GTTGCGGCTCCCGGGGAGCGGGGCCGTGCAGGCCGCGAGCCCCGAGCGCGCCGGCTGGACCGAGGCGCTG
CGGGCCCGCGTGGCCGAGCTGCGCGCCGGCGCCGTGGTGGCCGTCCCCACCGATACGCTGTACGGCCTGG
CCTGCGCGGCGAGCTGCTCGGCGGCTCTGCGCGCTGTGTACCGCCTCAAGGGTCGCAGCGAGGCCAAGCC
TCTGGCCGTATGCCTCGGCCGCGTGGCCGACGTCTACAGATACTGCCGTGTGAGAGTACCTGAGGGGCTC
CTGAAAGACCTACTGCCAGGACCAGTGACCCTGGTGTGGAACGCTCGGAGGAGCTCAACAAGGACCTAA
ACCCTTTTACGCCTCTTGTAGGCATTCGGATTCTGATCATGCTTTTATGCAAGACTTGGCTCAGATGTT
TGAGGGTCCGCTTGCTCTCACTAGTGCCAACCTCAGCTCCCAGGCCAGTTCTCTGAATGTCGAGGAGTTC
CAGGATCTCTGGCCTCAGTTGTCCTTGGTTATTGATGGGGGACAAATTGGGGATGGCCAGAGCCCCGAGT
GTCGCCTTGGCTCAACTGTGGTTGATTTGTCTGTGCCCGGAAAGTTGGCATCATTCGTCAGGCTGTGC
CCTGGAAAGTACTACAGCCATCCTCCAACAGAAGTACGGACTGCTCCCCTCACATGCGTCCTACCTGTGA
  
```

### Translation (279 aa):

```

MSPARRCRGMRAAVAASVGLSEGPAGSRSGRLFRPPSPAPAAPGARLLRLPGSGAVQAASPERAGWTEAL
RAAVAE LRAGAVVAVPTDTLYGLACAASCSAALRAVYRLKGRSEAKPLAVCLGRVADVRYCVRVPEGL
LKDLLPGPVTLVMERSEELNKDLNPFPLVGI R I PDHAFMQDLAQMFEGPLALTSANLSSQASSLNVEEF
QDLWPQLSLVIDGGQIGDQSPECLRGSTVVVDLSVPGKFGIIRPGCALESTTAILQOKYGLLPSHASYL
  
```

→ Comment extraire une séquence nucléotidique fasta à partir d'une séquence génomique?

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide Search Advanced Help

GenBank Send to

## Thermococcus kodakarensis KOD1 DNA, complete genome

GenBank: AP006878.1  
[FASTA](#) [Graphics](#)

Go to

LOCUS AP006878 2088737 bp DNA circular BCT  
 DEFINITION Thermococcus kodakarensis KOD1 DNA, complete genome.  
 ACCESSION AP006878  
 VERSION AP006878.1  
 DBLINK BioProject: [PRJNA13213](#)  
 BioSample: [SAM00061071](#)  
 KEYWORDS .  
 SOURCE Thermococcus kodakarensis KOD1  
 ORGANISM [Thermococcus kodakarensis KOD1](#)  
 Archaea; Euryarchaeota; Thermococci; Thermococcales;  
 Thermococcaceae; Thermococcus.  
 REFERENCE 1  
 AUTHORS Fukui,T., Atomi,H., Kanai,T., Matsumi,R., Fujiwara,S. and  
 Imanaka,T.  
 TITLE Complete genome sequence of the hyperthermophilic archaeon  
 Thermococcus kodakaraensis KOD1 and comparison with Pyrococcus  
 genomes  
 JOURNAL Genome Res. 15 (3), 352-363 (2005)  
 PUBMED [15710748](#)

Dans la fiche GenBank cliquer sur change region shown, puis selected region, puis donner les coordonnées de la séquence souhaitée

**Change region shown**

Whole sequence  
 Selected region  
 from:  to:

**Customize view**

**Basic Features**  
 Default features  
 Gene, RNA, and CDS features only

**Features added by NCBI**  
 3421 conserved domains

**Display options**  
 Show sequence  
 Show reverse complement

**Analyze this sequence**

[Run BLAST](#)  
[Pick Primers](#)

# ORFinder

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

« ORF finder » cherche les ORF (entre 2 STOP) et donne la séquence codante trouvée à l'intérieur et la signale en trait plein rouge.

NCBI Resources How To Sign in to NCBI

ORFfinder PubMed Search

### Open Reading Frame Finder

ORF finder searches for open reading frames (ORFs) in the DNA sequence you enter. The program returns the range of each ORF, along with its protein translation. Use ORF finder to search newly sequenced DNA for potential protein encoding segments, verify predicted protein using newly developed SMART BLAST or regular BLASTP.

This web version of the ORF finder is limited to the subrange of the query sequence up to 50 kb long. Stand-alone version, which doesn't have query sequence length limitation, is available for [Linux x64](#).

**Examples** (click to set values, then click Submit button) :

- NC\_011604 Salmonella enterica plasmid pWES-1; genetic code: 11; 'ATG' and alternative initiation codons; minimal ORF length: 300 nt
- NM\_000059; genetic code: 1; start codon: 'ATG only'; minimal ORF length: 150 nt

### Enter Query Sequence

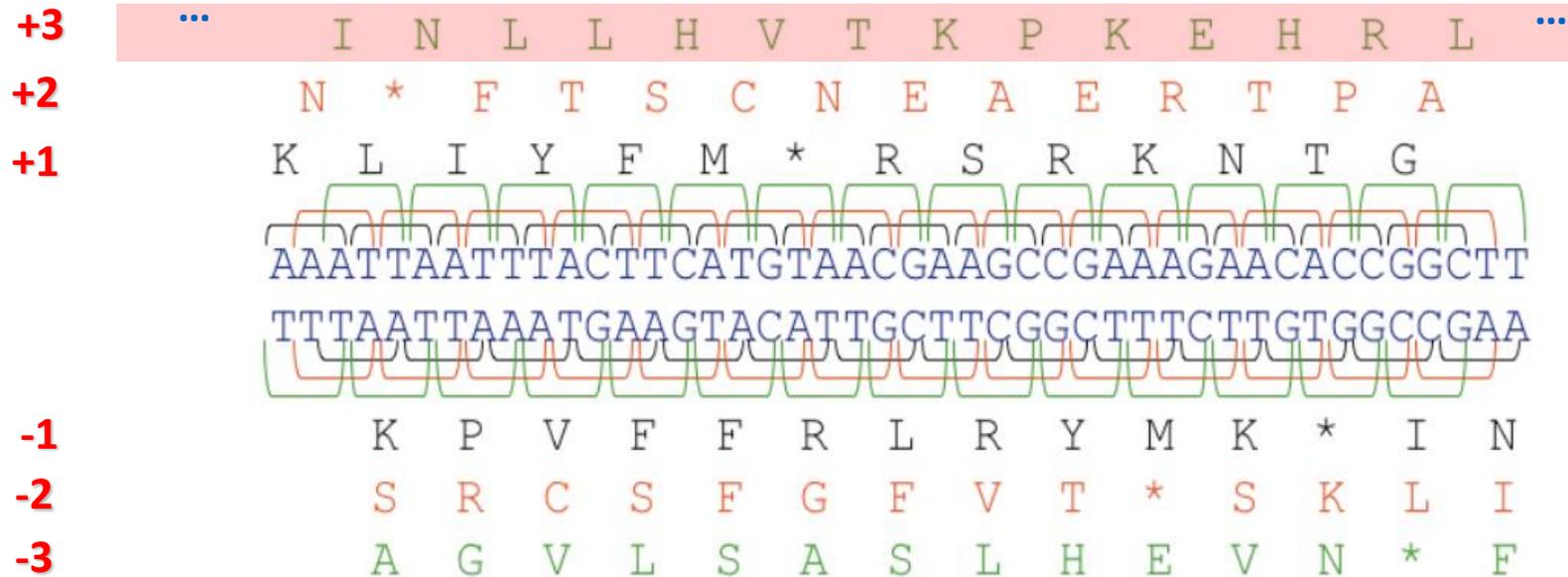
Enter accession number, gi, or nucleotide sequence in FASTA format:

From: To:

# ORF finder

- Traduction à l'aveugle

6 phases de lecture = 6 séquences protéiques possibles



### Open Reading Frame Finder

ORF finder searches for open reading frames (ORFs) in the DNA sequence you enter. The program returns the location of each ORF, along with its protein translation. Use ORF finder to search newly sequenced DNA for potential protein encoding segments, verify predicted protein using newly developed SMART BLAST or regular BLASTP.

This web version of the ORF finder is limited to the subrange of the query sequence up to 50 kb long. Stand-alone version, which doesn't have query sequence length limitation, is available for [Linux x64](#).

**Examples** (click to set values, then click Submit button) :

- [NC\\_011604 Salmonella enterica plasmid pWES-1](#); genetic code: 11; 'ATG' and alternative initiation codons; minimal ORF length: 300 nt
- [NM\\_000059](#); genetic code: 1; start codon: 'ATG' only; minimal ORF length: 150 nt



**Enter Query Sequence**

Enter accession number, gi, or sequence in FASTA format:

From:  To:

Copier-coller votre séquence

**Choose Search Parameters**

Minimal ORF length (nt):

Genetic code:

ORF start codon to use:

"ATG" only

"ATG" and alternative initiation codons

Any sense codon

Ignore nested ORFs:

Choisir 11. Bacterial, Archaeal, plastids Code

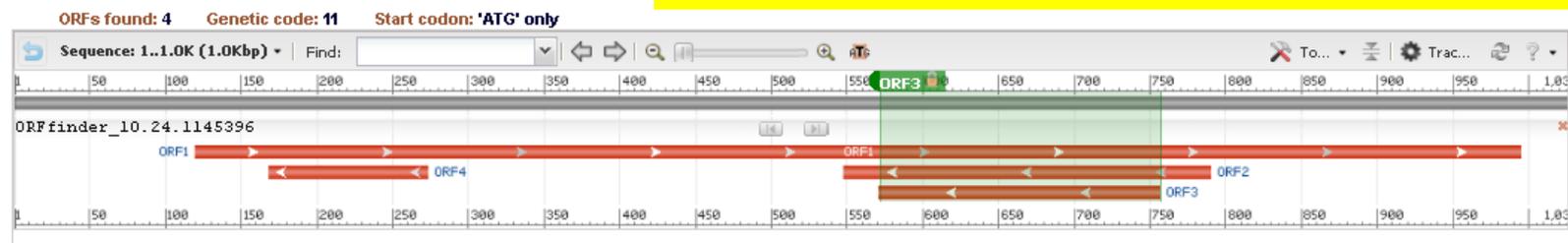
Puis Cliquer sur

**Start Search / Clear**

### Open Reading Frame Viewer

Sequence inconnue au format Fasta

Représentation schématique des CDS sur la séquence ADN



### Traduction de la CDS sélectionnée

ORF3 (61 aa)

```
>1c1|ORF3
MSPAAPVWGCRPHLYPASDSAEFAPLPEVKW
YHADRPEWRCLTSARYGRDGTFRLLPAVGRT
R
```

BLAST Database:

### Comparaison aux banques de données - BLAST

Mark subset Marked: 0 Download marked set as FASTA

Label	Strand	Frame	Start	Stop	Length (bp   aa)
ORF1	+	3	120	995	876   291
ORF2	-	2	790	548	243   80
<b>ORF3</b>	<b>-</b>	<b>3</b>	<b>756</b>	<b>571</b>	<b>186   61</b>
ORF4	-	3	273	169	105   34



## UniProt – base de données dédié aux séquences protéiques

Requête: YrdC homo sapiens



The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

**UniProtKB**  
UniProt Knowledgebase

**Swiss-Prot (553,231)**  
Manually annotated and reviewed.  
Records with information extracted from literature and curator-evaluated computational analysis.

**TrEMBL (71,002,161)**  
Automatically annotated and not reviewed.  
Records that await full manual annotation.

**UniRef**  
The UniProt Reference Clusters (UniRef) provide clustered sets of sequences from the UniProt Knowledgebase (including isoforms) and selected UniParc records.

**UniParc**  
UniParc is a comprehensive and non-redundant database that contains most of the publicly available protein sequences in the world.

**Proteomes**  
A proteome is the set of proteins thought to be expressed by an organism. UniProt provides proteomes for species with completely sequenced genomes.

**Supporting data**

- Literature citations
- Taxonomy
- Subcellular locations
- Cross-ref. databases
- Diseases
- Keywords

**News**

- Forthcoming changes  
Planned changes for UniProt
- UniProt release 2016\_11  
From mouth to gut, a new mechanism for fimbriae assembly | Change of URIs for Ensembl and Ensembl Genomes
- UniProt release 2016\_10  
N-acyl amino acids: a new treatment for obesity? | Cross-references to DisGeNET and OpenTargets | Change of the cross-references to Phos...

[News archive](#)

### Getting started

- [Text search](#)  
Our basic text search allows you to search all the resources available
- [BLAST](#)



### UniProt data

- [Download latest release](#)  
Get the UniProt data
- [Statistics](#)  
View Swiss-Prot and TrEMBL statistics

### Protein spotlight



#### Tenacious Memory

November 2016

We need to remember. Yet there are things we would like to forget. They just hang on in there regardless of feeling and time. What is it that keeps

UniProtKB  Advanced

[BLAST](#) [Align](#) [Retrieve/ID mapping](#) [Peptide search](#) [Help](#) [Contact](#)

## UniProtKB results

[About UniProtKB](#) [Basket](#)

[BLAST](#) [Align](#) [Download](#) [Add to basket](#) [Columns](#) [Share](#) 1 to 2 of 2 Show 25

**Quote terms: "homo sapiens"**

<input type="checkbox"/>	Entry	Entry name	<input type="checkbox"/>	Protein names	<input type="checkbox"/>	Gene names	Organism	Length	<input type="button" value="Edit"/>
<input type="checkbox"/>	Q86U90	YRDC_HUMAN	<input type="checkbox"/>	YrdC domain-containing protein, mit...	<input type="checkbox"/>	YRDC DRIP3, IRIP	Homo sapiens (Human)	279	<input type="button" value="Edit"/>
<input type="checkbox"/>	Q92681	RSCA1_HUMAN	<input type="checkbox"/>	Regulatory solute carrier protein f...	<input type="checkbox"/>	RSC1A1	Homo sapiens (Human)	617	<input type="button" value="Edit"/>

1 to 2 of 2 Show 25

**Filter by**<sup>i</sup>

- [Reviewed \(2\)](#)  
Swiss-Prot
- Popular organisms**  
[Human \(2\)](#)
- Search terms**  
Filter "yrdc" as:  
[domain \(1\)](#)  
[gene name \(1\)](#)  
[protein name \(1\)](#)
- View by**  
[Results table](#)  
[Taxonomy](#)  
[Keywords](#)  
[Gene Ontology](#)  
[Enzyme class](#)  
[Pathway](#)
- UniRef**

nr. d'accession

## UniProtKB - Q86U90 (YRDC\_HUMAN)

### Display

- Entry
- Publications
- Feature viewer
- Feature table

- None
- Function
  - Names & Taxonomy
  - Subcellular location
  - Pathology & Biotech
  - PTM / Processing
  - Expression
  - Interaction
  - Structure
  - Family & Domains
  - Sequence
  - Cross-references
  - Entry information
  - Miscellaneous
  - Similar proteins

- BLAST
- Align
- Format
- Add to basket
- History

**Protein** | YrdC domain-containing protein, mitochondrial

**Gene** | YRDC

**Organism** | *Homo sapiens (Human)*

**Status** |  Reviewed - Annotation score: ●●●●○ - Experimental evidence at protein level<sup>i</sup>

### Function<sup>i</sup>

May regulate the activity of some transporters.  By similarity

#### GO - Molecular function<sup>i</sup>

- double-stranded RNA binding  Source: InterPro
- nucleotidyltransferase activity  Source: GO\_Central
- tRNA binding  Source: GO\_Central

#### GO - Biological process<sup>i</sup>

- negative regulation of transport  Source: MGI
- regulation of translational fidelity  Source: GO\_Central
- tRNA threonylcarbamoyladenosine modification  Source: GO\_Central

← complete GO annotation...

#### Enzyme and pathway databases

BioCyc<sup>i</sup> | ZFISH:G66-32489-MONOMER.

### Names & Taxonomy<sup>i</sup>

Protein name: | Recommended name:

## Display

Entry

Publications

Feature viewer

Feature table

None

Function

Names & Taxonomy

Subcellular location

Pathology & Biotech

PTM / Processing

Expression

Interaction

Structure

Family & Domains

Sequence

Cross-references

Entry information

Miscellaneous

Similar proteins

[▲ Top](#)

## Sequence<sup>1</sup>

Sequence status<sup>1</sup>: Complete.

Sequence processing<sup>1</sup>: The displayed sequence is further processed into a mature form.

Q86U90-1 [UniParc]

[FASTA](#)

[Add to basket](#)

[« Hide](#)

10	20	30	40	50
MSPARRCRGM	RAAVAASVGL	SEGFAGSRSG	RLFRPPSPAP	AAPGARLLRL
60	70	80	90	100
PGSGAVQAAS	PERAGWTEAL	RAAVAELRAG	AVVAVPTDTL	YGLACAASCS
110	120	130	140	150
AALRAVYRLK	GRSEAKPLAV	CLGRVADVYR	YCRVRVPEGL	LKDLLPGPVT
160	170	180	190	200
LVMERSEELN	KDLNPFPTPLV	GIRIPDHAFM	QDLAQMFEQP	LALTSANLSS
210	220	230	240	250
QASSLNVEEF	QDLWPQLSLV	IDGGQIGDQQ	SPECRLGSTV	VDLSVPGKFG
260	270			
IIRPGCALES	TTAILQQKYG	LLPSHASYL		

**Length:** 279

**Mass (Da):** 29,328

**Last modified:** June 1, 2003 - v1

**Checksum:**<sup>1</sup> F8F5C30274E2BB12

BLAST

[GO](#)

## Sequence caution<sup>1</sup>

The sequence [AAH08984](#) differs from that shown. Reason: Erroneous initiation. [Curated](#)

The sequence [AAP37053](#) differs from that shown. Reason: Erroneous initiation. [Curated](#)

The sequence [AAP37054](#) differs from that shown. Reason: Erroneous initiation. [Curated](#)

The sequence [BAB15668](#) differs from that shown. Reason: Erroneous initiation. [Curated](#)

## Sequence databases

Select the link [AY172561](#) mRNA. Translation: [AAO41711.1](#).

destinations: [AL929472](#) Genomic DNA. Translation: [CAH70083.1](#).

[EMBL](#)<sup>1</sup> [CH471059](#) Genomic DNA. Translation: [EAX07317.1](#).

# Comment faire un BLAST?

c'est très simple:

NCBI/BLAST/blastp suite Standard Protein BLAST

blastn **blastp** blastx tblastn tblastx

BLASTP programs search protein databases using a protein query. [more...](#)

**Enter Query Sequence**

Enter accession number(s), gi(s), or FASTA sequence(s)  [Clear](#) Query subrange  From  To

Or, upload file  Aucun fichier sélectionné.

Job Title

Enter a descriptive title for your BLAST search

Align two or more sequences

**Choose Search Set**

Database

Organism   Exclude  +

Exclude  Models (XM/XP)  Uncultured/environmental sample sequences

Entrez Query  [YouTube](#) [Create custom database](#)

Enter an Entrez query to limit search

**Program Selection**

Algorithm  blastp (protein-protein BLAST)  
 PSI-BLAST (Position-Specific Iterated BLAST)  
 PHI-BLAST (Pattern Hit Initiated BLAST)  
 DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)  
Choose a BLAST algorithm

Search database Non-redundant protein sequences (nr) using Blastp (protein-protein BLAST)  
 Show results in a new window

1. requête (votre séquence)

2. portée

(à qui vous voulez la comparer, base de données, organisme, groupe d'organismes...)

3. et hop !

# Les paramètres cachés de BLAST

**Algorithm parameters**

**General Parameters**

**Max target sequences**  Select the maximum number of aligned sequences to display

**Short queries**  Automatically adjust parameters for short input sequences

**Expect threshold**  seuil sur l'E-value

**Word size**  taille de l'amorce (mot, k)

**Max matches in a query range**

**Scoring Parameters**

**Matrix**  choix de la matrice de substitution

**Gap Costs**  cout de gaps

**Compositional adjustments**

**Filters and Masking**

**Filter**  Low complexity regions

**Mask**  Mask for lookup table only  
 Mask lower case letters

nombre max. de séquences alignées affichées

seuil sur l'E-value

taille de l'amorce (mot, k)

choix de la matrice de substitution

cout de gaps

# Page de résultats

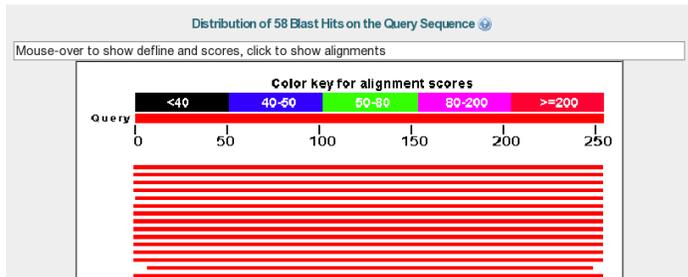
sp|P04156|PRIO\_HUMAN Major prion protein OS=Homo...

Query ID: IJ46010  
Description: sp|P04156|PRIO\_HUMAN Major prion protein OS=Homo sapiens GN=PRNP PE=1 SV=1  
Molecule type: amino acid  
Query Length: 253

Database Name: swissprot  
Description: Non-redundant SwissProt sequences  
Program: BLAST 2.2.21+ >Citation

Other reports: >Search Summary [Taxonomy reports] [Distance tree of results] [Multiple alignment]

1. récapitulatif de la requête



2. représentation graphique des résultats

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value
P36914.2	glucoamylase [Aspergillus oryzae RIB40] >gi 94730359 sp P36914.2 AMY	1245	1245	100%	0.0
P22832.1	RecName: Full=Glucosylase; AltName: Full=1,4-alpha-D-glucan glucohy	845	845	99%	0.0
P69327.1	RecName: Full=Glucosylase; AltName: Full=1,4-alpha-D-glucan glucohy	843	843	99%	0.0
P23176.1	RecName: Full=Glucosylase I; AltName: Full=1,4-alpha-D-glucan glucohy	842	842	99%	0.0
P14804.3	RecName: Full=Glucosylase; AltName: Full=1,4-alpha-D-glucan glucohy	631	631	95%	0.0
Q03045.1	RecName: Full=Glucosylase P; AltName: Full=1,4-alpha-D-glucan glucohy	560	560	99%	0.0
Q60087.1	RecName: Full=Probable glucosylase; AltName: Full=1,4-alpha-D-glucan glucohy	294	294	73%	4e-92
P07683.2	RecName: Full=Glucosylase 1; Short=Gluc 1; AltName: Full=1,4-alpha-D-glucan glucohy	267	267	68%	3e-80
P42042.1	RecName: Full=Glucosylase; AltName: Full=1,4-alpha-D-glucan glucohy	245	245	67%	1e-71
P26989.2	RecName: Full=Glucosylase GLA1; AltName: Full=1,4-alpha-D-glucan glucohy	224	224	70%	6e-65
P08017.1	RecName: Full=Glucosylase GLU1; AltName: Full=1,4-alpha-D-glucan glucohy	220	220	70%	2e-63
P08019.2	RecName: Full=Glucosylase, intracellular sporulation-specific; AltName: Full=1,4-alpha-D-glucan glucohy	203	203	70%	8e-57
P04065.2	RecName: Full=Glucosylase S1; AltName: Full=1,4-alpha-D-glucan glucohy	183	183	54%	3e-48
P29780.1	RecName: Full=Glucosylase S2; AltName: Full=1,4-alpha-D-glucan glucohy	182	182	54%	4e-48
P22988.1	RecName: Full=Alpha-amylase; AltName: Full=1,4-alpha-D-glucan glucohy	80.1	80.1	14%	2e-14
Q03065.1	RecName: Full=Cyclomaltoextrin glucoamylase; AltName: Full=Cyclomaltoextrin glucoamylase	74.7	74.7	14%	9e-13
P29750.1	RecName: Full=Alpha-amylase; AltName: Full=1,4-alpha-D-glucan glucohy	71.2	71.2	12%	1e-11
P05618.1	RecName: Full=Cyclomaltoextrin glucoamylase; AltName: Full=Cyclomaltoextrin glucoamylase	71.2	71.2	14%	1e-11

3. résumé des résultats

GENE ID: 396452 PRNP | prion protein (p27-30) (Creutzfeldt-Jakob disease, Gerstmann-Sträussler-Scheinker syndrome, fatal familial insomnia) [Gallus gallus] (10 or fewer PubMed Links)

Score = 80.9 bits (198), Expect = 6e-15, Method: Compositional matrix adjust. Identities = 94/230 (40%), Positives = 120/230 (52%), Gaps = 33/230 (14%)

Query 18 DLGLC...KRRPKPGWNTGGSRYPGGSPGGRYPPGGGGGGOPHGGGGOPHGGGG 74  
Sbjct 20 DVALSKKGRGKPSGGGGAGSHRQPSYPROPQYPHNPGYPHNPGY - PHNPGY - 76

Query 75 OPHGGGGW...PHG...GGWG...GGGTHSOWNKPSK -PKTNMKHMAGAAAAA 120  
Sbjct 77 -PHNPGYPPNPGYPHNPGYPPNPGYPPNPGYPPNPGYPPNPGYPPNPGYPPNPGY 133

Query 121 VVGGGGYMLGASMSRPIHFGSDYEDRYRENHRYVNOVYRPMDEYSNONNFVHDCV 180  
Sbjct 134 VVGGGGYAMGRVMSGNNYHFDSPDYRWSENSARYPNRYVRYDYSSPVPQDVFVADCF 193

Query 181 NI TIKOHTVYTTTK - - - - - GENFTE TDV - KMMERVVEORCI TOYER 220  
Sbjct 194 NI TVYESTIGPAAKNTSEAVAAAANTEVEHENKVVTKVIREKCVQYRE 243

sp|A415X2\_1|IF2\_DESRM RecName: Full=Translation initiation factor IF-2 Length=985

GENE ID: 4955422 Pfad\_1857 | translation initiation factor IF-2 (Drosophila melanogaster) (10 or fewer PubMed Links)

Score = 46.6 bits (109), Expect = 2e-04, Method: Composition-based stats. Identities = 39/103 (37%), Positives = 51/103 (49%), Gaps = 32/103 (31%)

Query 29 GQWNTGGSRYP...GGSPGGRYPPGG...GGGGOPHG...GGGG 67  
Sbjct 105 GGRPGRGSRPYGDRPQGGGRPYGDR - PGGGGRPYGDRPQGGGRPYGDRPQGGGR 223

Query 181 NI TIKOHTVYTTTK - - - - - GENFTE TDV - KMMERVVEORCI TOYER 220  
Sbjct 223 PYGDRPQGGGRPYGDRPQGGGRPYGDR - PGGGGRPYGDRPQGGGRPYGDRPQGGGR 264

4. les alignements

# Page de résultats

## 1. récapitulatif de la requête

quelle séquence a été soumise ("query") ;  
identifiant, descriptif, longueur, type

[Edit and Resubmit](#) [Save Search Strategies](#) [▶Formatting options](#) [▶Download](#)

sp|P04156|PRIO\_HUMAN Major prion protein OS=Homo...

**Query ID** Id|46010

**Description** sp|P04156|PRIO\_HUMAN Major prion protein OS=Homo sapiens GN=PRNP  
PE=1 SV=1

**Molecule type** amino acid

**Query Length** 253

**Database Name** swissprot

**Description** Non-redundant SwissProt sequences

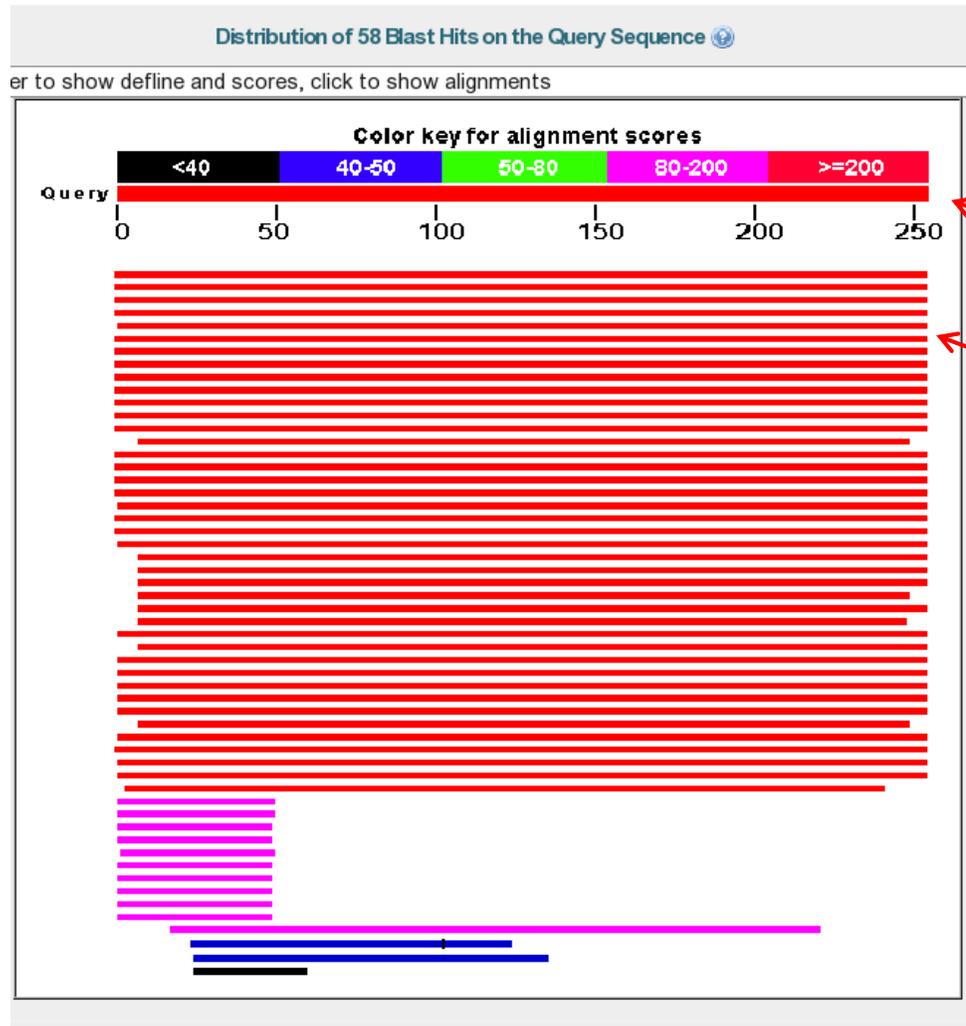
**Program** BLASTP 2.2.21+ [▶Citation](#)

Other reports: [▶Search Summary](#) [\[Taxonomy reports\]](#) [\[Distance tree of results\]](#) [\[Multiple alignment\]](#)

quelle banque de données est interrogée ?

quel programme est utilisé ?

# Page de résultats



## 2. représentation graphique des résultats

ce trait représente  
la séquence soumise (long. 253 AA)

chaque trait de couleur  
représente un alignement  
entre la séquence de départ  
et une séquence de la banque  
de donnée sélectionnée  
couleur → score  
longueur → taille de l'alignement

# Page de résultats

## 3. résumé des résultats

identifiant                      descriptif                      score                      couverture                      E-value

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value
<a href="#">P36914.2</a>	glucoamylase [Aspergillus oryzae RIB40] >gi 94730359 sp P36914.2 AMY	<a href="#">1245</a>	1245	100%	0.0
<a href="#">P22832.1</a>	RecName: Full=Glucoamylase; AltName: Full=1,4-alpha-D-glucan glucohy	<a href="#">845</a>	845	99%	0.0
<a href="#">P69327.1</a>	RecName: Full=Glucoamylase; AltName: Full=1,4-alpha-D-glucan glucohy	<a href="#">843</a>	843	99%	0.0
<a href="#">P23176.1</a>	RecName: Full=Glucoamylase I; AltName: Full=1,4-alpha-D-glucan glucof	<a href="#">842</a>	842	99%	0.0
<a href="#">P14804.3</a>	RecName: Full=Glucoamylase; AltName: Full=1,4-alpha-D-glucan glucohy	<a href="#">631</a>	631	95%	0.0
<a href="#">Q03045.1</a>	RecName: Full=Glucoamylase P; AltName: Full=1,4-alpha-D-glucan glucc	<a href="#">560</a>	560	99%	0.0
<a href="#">O60087.1</a>	RecName: Full=Probable glucoamylase; AltName: Full=1,4-alpha-D-gluca	<a href="#">294</a>	294	73%	4e-92
<a href="#">P07683.2</a>	RecName: Full=Glucoamylase 1; Short=Gluc 1; AltName: Full=1,4-alpha-D	<a href="#">267</a>	267	68%	3e-80
<a href="#">P42042.1</a>	RecName: Full=Glucoamylase; AltName: Full=1,4-alpha-D-glucan glucohy	<a href="#">245</a>	245	67%	1e-71
<a href="#">P26989.2</a>	RecName: Full=Glucoamylase GLA1; AltName: Full=1,4-alpha-D-glucan g	<a href="#">224</a>	224	70%	6e-65
<a href="#">P08017.1</a>	RecName: Full=Glucoamylase GLU1; AltName: Full=1,4-alpha-D-glucan g	<a href="#">220</a>	220	70%	2e-63
<a href="#">P08019.2</a>	RecName: Full=Glucoamylase, intracellular sporulation-specific; AltName:	<a href="#">203</a>	203	70%	8e-57
<a href="#">P04065.2</a>	RecName: Full=Glucoamylase S1; AltName: Full=1,4-alpha-D-glucan glu	<a href="#">183</a>	183	54%	3e-48
<a href="#">P29760.1</a>	RecName: Full=Glucoamylase S2; AltName: Full=1,4-alpha-D-glucan glu	<a href="#">182</a>	182	54%	4e-48
<a href="#">P22998.1</a>	RecName: Full=Alpha-amylase; AltName: Full=1,4-alpha-D-glucan glucan	<a href="#">80.1</a>	80.1	14%	2e-14
<a href="#">Q30565.1</a>	RecName: Full=Cyclomaltodextrin glucanotransferase; AltName: Full=Cyc	<a href="#">74.7</a>	74.7	14%	9e-13
<a href="#">P29750.1</a>	RecName: Full=Alpha-amylase; AltName: Full=1,4-alpha-D-glucan glucan	<a href="#">71.2</a>	71.2	12%	1e-11
<a href="#">P05618.1</a>	RecName: Full=Cyclomaltodextrin glucanotransferase; AltName: Full=Cyc	<a href="#">71.2</a>	71.2	14%	1e-11

# Page de résultats

## 4. les alignements

```
> sp|P27177.2|PRIO\_CHICK  RecName: Full=Major prion protein homolog; AltName: Full=PR-LP;  
AltName: Full=Acetylcholine receptor-inducing activity; Short=ARIA;  
AltName: Full=65-21 protein; Flags: Precursor  
Length=273
```

```
GENE ID: 396452 PRNP | prion protein (p27-30) (Creutzfeldt-Jakob disease,  
Gerstmann-Strausler-Scheinker syndrome, fatal familial insomnia) [Gallus gallus]  
(10 or fewer PubMed links)
```

```
Score = 80.9 bits (198), Expect = 6e-15, Method: Compositional matrix adjust.  
Identities = 94/230 (40%), Positives = 120/230 (52%), Gaps = 33/230 (14%)
```

```
Query 18 DLGLC---KKRKP+GGWNTGGSKRYPGQ+SPGGNRYPPQGGGGWQPHGGGGWQPHGGGGW 74  
D+ L K +P GGW G R P G + P G PH G+ PH G+  
Sbjct 20 DVALSKKGGKPKSGGGWAGSHRQPSYPRQPGYPHNP+GYPHNP+GYPHNP+GY--PHNP+GY- 76  
Query 75 QPHGGGGWQ----PHG---GGWQ-----GGGTHSQWNKPSK-PKTNMKHMAGAAAAGA 120  
PH G+ Q PH GWGQ GG H+Q KP K PKTN KH+AGAAAAGA  
Sbjct 77 -PHNPGYPQNP+GYPHNP+GYPGWQGYNPSSGGSYHNQ--KPWKPKTNFKHVAVAGAAAAGA 133  
  
Query 121 VVGGLGGYMLGSAMSRPIIHFGSDYEDRYRENMMHRYPNQVYYRPMDEYSNQNNFVHDCV 180  
VVGGLGGY +G MS HF S E R++ EN RYPN+VYYR Q+ FV DC  
Sbjct 134 VVGGLGGYAMGRVMSGMNYHFDSPDEYRWWSENSARYPNRVYYR+DYSSPVPQDV+FVADCF 193  
  
Query 181 NITIKQHTVTTTTK-----GENFTETDV--KMMERVVEQMCITQYER 220  
NIT+ +++++ K N TE ++ K++ +V+ +MC+ QY  
Sbjct 194 NITVTEYSIGPAKKNTSEAVAAANQTEVEMENKVVTKVIREMCVQYRE 243
```

```
> sp|A4J5X2.1|IF2\_DESRM  RecName: Full=Translation initiation factor IF-2  
Length=985
```

```
GENE ID: 4955422 Dred 1957 | translation initiation factor IF-2  
(Desulfotomaculum reducens MT-1)
```

```
Score = 46.6 bits (109), Expect = 2e-04, Method: Composition-based stats.  
Identities = 39/103 (37%), Positives = 51/103 (49%), Gaps = 32/103 (31%)
```

```
Query 29 GGWNTGGSRYP-----GQGGSPGGNRYPPQGG-----GGWQPHG---GGWQ 67  
GG GG P GQG P G+R PQQG GG G+P+G GG G+  
Sbjct 165 GGRPQGGQSRPYGDRPQGGQGRPYGDR--PQGGQGRPYGDRPQGGQGRPYGDRPQGGQGR 222  
  
Query 68 PHG---GGWQPHG---GGWQPHGGGGWQGGGTHSQWNK 102  
P+G GG G+P+G GG G+P+G QGG + ++P  
Sbjct 223 PYGDRPQGGQGRPYGDRPQGGQGRPYGDR-PQGGQSRPYGDR 264
```

query, la séquence soumise  
subject, la séquence trouvée  
dans la bdd

les valeurs quantitatives  
pour chaque alignement:  
score  
E-value  
% identité  
% similarité  
% gaps (indéls)

# Page de résultats - score

score est normalisé (bits), il permet de comparer 2 alignements, et de dire lequel est le meilleur

```
> sp|Q92125.1|ANXA7\_XENLA  RecName: Full=Annexin A7; AltName: Full=Annexin-7; AltName: Full=Annexin VII; AltName: Full=Synexin  
Length=512
```

```
GENE ID: 397854 LOC397854 | annexin VII [Xenopus laevis]  
(10 or fewer PubMed links)
```

```
Score = 40.0 bits (92), Expect = 0.011, Method: Compositional matrix adjust.  
Identities = 45/128 (35%), Positives = 57/128 (44%), Gaps = 34/128 (26%)
```

```
Query 26 PKPGGWNTGGSRYPGQGSPGGNRYPPQGGGGWGQPHG-----GGWGQ 67  
P PGG+ G YPG +PG P GG G+G P G GG+G  
Sbjct 67 PAPGGYPGGMPSYPG--APGFGA--PAGGQGYGAPPGAPAYGVPGYGGPGFNAPAGGYGA 122  
  
Query 68 PHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPSKPKTNMKHMAGAAAAGAVVGGGLGG 127  
P+ GG+G P GG+G P GG G GG +++PS GA G + G + G  
Sbjct 123 PNAGGFGVPPAGGYGSP--GGAPGYGG-----FSQPS-----SQSYGAGGPGQMPGQMPG 170  
  
Query 128 YMLGSAMS 135  
M G A S  
Sbjct 171 QMPGQAPS 178
```

longueur: 128  
identité: 35%  
gap: 26%

```
> sp|Q92125.1|ANXA7\_XENLA  RecName: Full=Annexin A7; AltName: Full=Annexin-7; AltName: Full=Annexin VII; AltName: Full=Synexin  
Length=512
```

```
GENE ID: 397854 LOC397854 | annexin VII [Xenopus laevis]  
(10 or fewer PubMed links)
```

```
Score = 34.7 bits (76), Expect = 0.62, Method: Compositional matrix adjust.  
Identities = 34/86 (39%), Positives = 41/86 (47%), Gaps = 23/86 (26%)
```

```
Query 26 PKPGGWNTGGSRYPGQGSPGGNRYPPQGGGGWGQPHG-----GGWGQ 67  
P PGG+ G YPG +PG P GG G+G P G GG+G  
Sbjct 67 PAPGGYPGGMPSYPG--APGFGA--PAGGQGYGAPPGAPAYGVPGYGGPGFNAPAGGYGA 122  
  
Query 68 PHGGGWGQPHGGGWGQPHGGGWGQGG 93  
P+ GG+G P GG+G P GG G GG  
Sbjct 123 PNAGGFGVPPAGGYGSP--GGAPGYGG 147
```

longueur: 86  
identité: 39%  
gap: 26%

## Multiple Sequence Alignment

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between **three or more** sequences. For the alignment of two sequences please instead use our [pairwise sequence alignment tools](#).

**Important note:** This tool can align up to 4000 sequences or a maximum file size of 4 MB.

### STEP 1 - Enter your input sequences

Enter or paste a set of

PROTEIN

sequences in any supported [format](#):

Or, [upload a file](#):  Aucun fichier sélectionné.

[Use a example sequence](#) | [Clear sequence](#) | [See more example inputs](#)

Insérer les  
séquences sous  
format FASTA

### STEP 2 - Set your parameters

OUTPUT FORMAT

ClustalW with character counts

*The default settings will fulfill the needs of most users.*

*(Click here, if you want to view or change the default settings.)*

GO!

### STEP 3 - Submit your job

Be notified by email *(Tick this box if you want to be notified by email when the results are available)*

Submit

# Clustal Omega

## Results for job clustalo-l20181126-213954-0578-88728344-p2m

Alignments

Result Summary

Phylogenetic Tree

Submission Details

Download Alignment File

Show Colors

View result with Jalview

Send to Simple Phylogeny

Send to MView

CLUSTAL O(1.2.4) multiple sequence alignment

```
WP_042702101.1  --MKIVQMDQNNPDLDLIDEAIEVLTSGGVVLYPDDTVYGLGANVFNEKAVEKVYNIKNR      58
RLI59347.1      -MVRIFR--WPLRESE-REMLSAILMNGGVVVYPTDTIYGISSGIYDRRVVDKIYKIKER      56
RLG13225.1      MEVEIIVKVEKVLADNE-LIQAIRYLEQEKTVVFPDPTVYGIADLFFKEKAVKNLFEIKKR      59
AIF13504.1      -----MKISCNDVD-IQIATKAINDGAIVVFPDPTVYGLGCNPYNHDAVLSLFEIKKR      52
                : :           : . *::*****::*   .. .*  ::*:~*~*

WP_042702101.1  DYFKPLSVCVSSIDEILLIADVGNKTHQILKNNLPGPFTFIFYKKEPIHNYA-TKNHKV      117
RLI59347.1      LRDKGLPILFGSIEKIEEIA-ILNPLAKKLAQRFWPGMLTLVVPLKRRELRFVTGLEDKI      115
RLG13225.1      PLDKPINALVSSMKQAEMIVENIPDSAKKLIKEFWPGGLTIILKKKDIVPDIVTAGRETI      119
AIF13504.1      KTKKPPFVVLGYSKKELEKIA-EFNSLEEKIAEKFWPGPITLILKVKDKIEIQKSLDLEGKI      111
                * :      * :.  *.      .:: :.  ** :~*~*  *      . . :

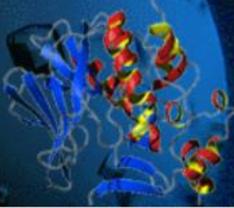
WP_042702101.1  GIRIPENIISRKLTQS--FPITTTSANLSGKKTLLNPNDEIINQLNEGIDFVIDVGRKQS      175
RLI59347.1      AVRIPSDEIVLEVIKLAGGAIIGTSANLSGKLPCTNIKCVISQLGERMDAIVDGGERTG      175
RLG13225.1      GVRKPNNEIILQILEKFGKPLAATSANISGLTSPITAKEVLQQLGEKSYLLIDGGKTIEK      179
AIF13504.1      AVRVPNNQCILALLKEC-KLLVGTSANISGTAPFNDPKECGENLS-GYDLLIDGGVISSQ      169
                .:* *.:      : :      : *****~*      .  .:~*      :~* *

WP_042702101.1  EPSTIVDLTK-KEPKILRKGSGILKSF-----      201
RLI59347.1      FPSTIVEIQDDKKIKIIREGSIEVEKIKEAIGLEFS      211
RLG13225.1      QPSTIIDLSS-KKPLILRKGSVSVEKIKEILPSIIE      214
AIF13504.1      GESTIVEIEN-NDVKILRKGSVSEMIKELT-----      199
                ***::~ . . . *::~*~*      : :
```

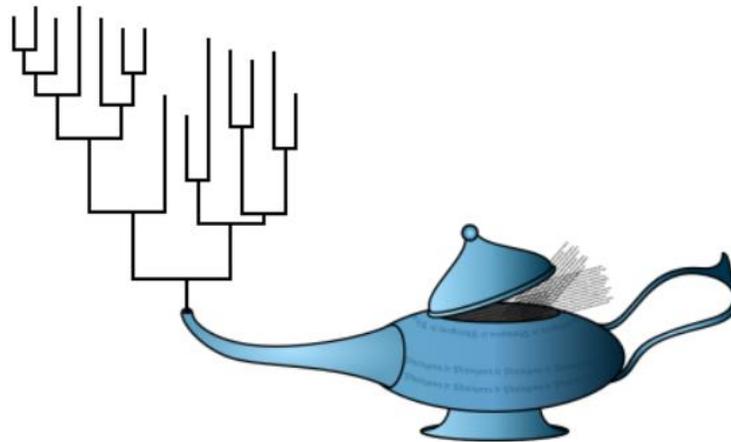
Télécharger le  
fichier  
d'alignement sous  
forme clustal

# Choix des sites: Gblocks

[http://www.phylogeny.fr/one\\_task.cgi?task\\_type=gblocks](http://www.phylogeny.fr/one_task.cgi?task_type=gblocks)



## Phylogeny.fr Robust Phylogenetic Analysis For The Non-Specialist



Phylogeny.fr is a free, simple to use web service dedicated to reconstructing and analysing phylogenetic relationships between molecular sequences.

Phylogeny.fr runs and connects various bioinformatics programs to reconstruct a **robust phylogenetic tree from a set of sequences**.

If you use this site, as I am managing it alone since years, could you please add me in the acknowledgments (and let me know):

**Sebastien Santini - CNRS/AMU IGS UMR7256**

Moreover, please cite:

Dereeper A \* Guignon V\*, Blanc C, Audic S, Buffet S, Chevillon E, Duforet JE, Guindon S, Lefort V, Leclerc M, Claverie JM

# Choix des sites: Gblocks

Phylogeny.fr  
Robust Phylogenetic Analysis For The Non-Specialist

Phylogeny.fr is a free, simple to use web service dedicated to reconstructing and analysing phylogenetic relationships between molecular sequences.

Phylogeny.fr runs and connects various bioinformatics programs to reconstruct a **robust phylogenetic tree from a set of sequences**.

If you use this site, as I am managing it alone since years, could you please add me in the acknowledgments (and let me know):

**Sebastien Santini - CNRS/AMU IGS UMR7256**

# Gblocks 0.91b (doc)

1. Overview

2. Data & Settings

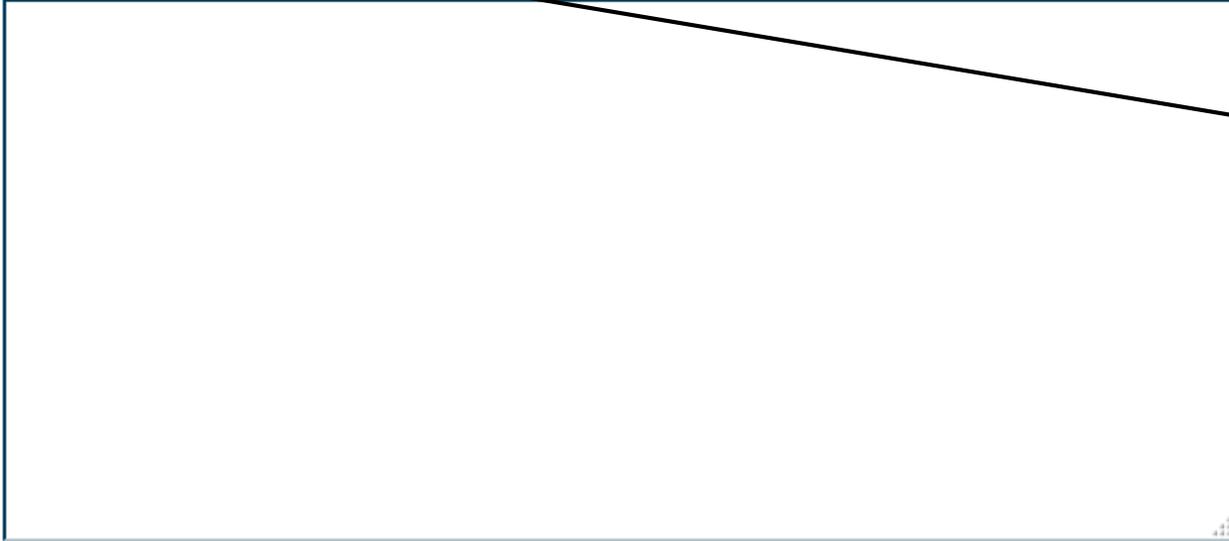
3. Results

Upload your alignment (FASTA, Phylip, Clustal, EMBL or NEXUS format) from a file:

Parcourir...

Aucun fichier sélectionné.

Or paste it here [\(load an example of alignment\)](#)



Clear

Insérer votre alignement (FASTA, Phylip, Clustal format)

Copier - coller ou charger un fichier

## Settings

For a less stringent selection:

- Allow smaller final blocks
- Allow gap positions within the final blocks
- Allow less strict flanking positions

For a more stringent selection:

- Do not allow many contiguous nonconserved positions

Les paramètres par défaut sont conservatifs (stringents)  
Vous pouvez moduler le niveau de stringence:

Submit

Lancer Gblocks

# Choix des sites: Gblocks

## Gblocks 0.91b (doc)

1. Overview

2. Data & Settings

3. Results

## Curation: Gblocks

Processed file: **input.fasta**

Number of sequences: **12**

Alignment assumed to be: **DNA**

New number of positions: **886** (selected positions are underlined in blue)

	10	20	30	40	50	60
Tarsius_syrichth	AAGTTCATTGGAGCCACCACTCTTATAATTGCCCATGGCCTCACCTCCTCCCTATTATT					
Lemur_catta	AAGCTTCATAGGAGCAACCATTCCTAATAATCGCACATGGCCTTACATCATCCATATTATT					
Homo_sapiens	AAGCTTCACCGGCGCAGTCACTTCTCATAATCGCCCACGGGCTTACATCCTCATTCCTATT					
Pan	AAGCTTCACCGGCGCAATTATCCTCATAATCGCCCACGGACTTACATCCTCATATTATTATT					
Gorilla	AAGCTTCACCGGCGCAGTTGTTTCTTATAATTGCCCAACGGACTTACATCATCATATTATT					
Pongo	AAGCTTCACCGGCGCAACCACTTCATGATGGCCATGGACTCACATCCTCCCTACTGTT					
Hylobates	AAGCTTTACAGGTGCAACCGTCCCTCATAATCGCCCAACGGACTAACCTCTCCCTGCTATT					
Macaca_fuscata	AAGCTTTTCCGGCGCAACCATCCTTATGATCGCTCACGGACTCACCTCTTCCATATATTATT					
M_mulatta	AAGCTTTTCTGGCGCAACCATCCTCATGATGGCTCACGGACTCACCTCTTCCATATATTATT					
M_fascicularis	AAGCTTCTCCGGCGCAACCACTTATAATCGCCCAACGGCTCACCTCTTCCATATTATT					
M_sylvanus	AAGCTTCTCCGGTGCAACTATCCTTATAGATGGCCATGGACTCACCTCTTCCATATACTT					
Saimiri_sciureu	AAGCTTCACCGGCGCAATGATCCTAATAATCGCTCACGGTCTTACTTCCTATAGCTATT					

Tirer vers le bas pour visualiser l'ensemble de l'alignement

Les blocks retenus sont soulignés en bleu

Figure 1: Selected alignment blocks (underlined in blue).

# Choix des sites: Gblocks

Alignment assumed to be: DNA

New number of positions: 886 (selected positions are underlined in blue)

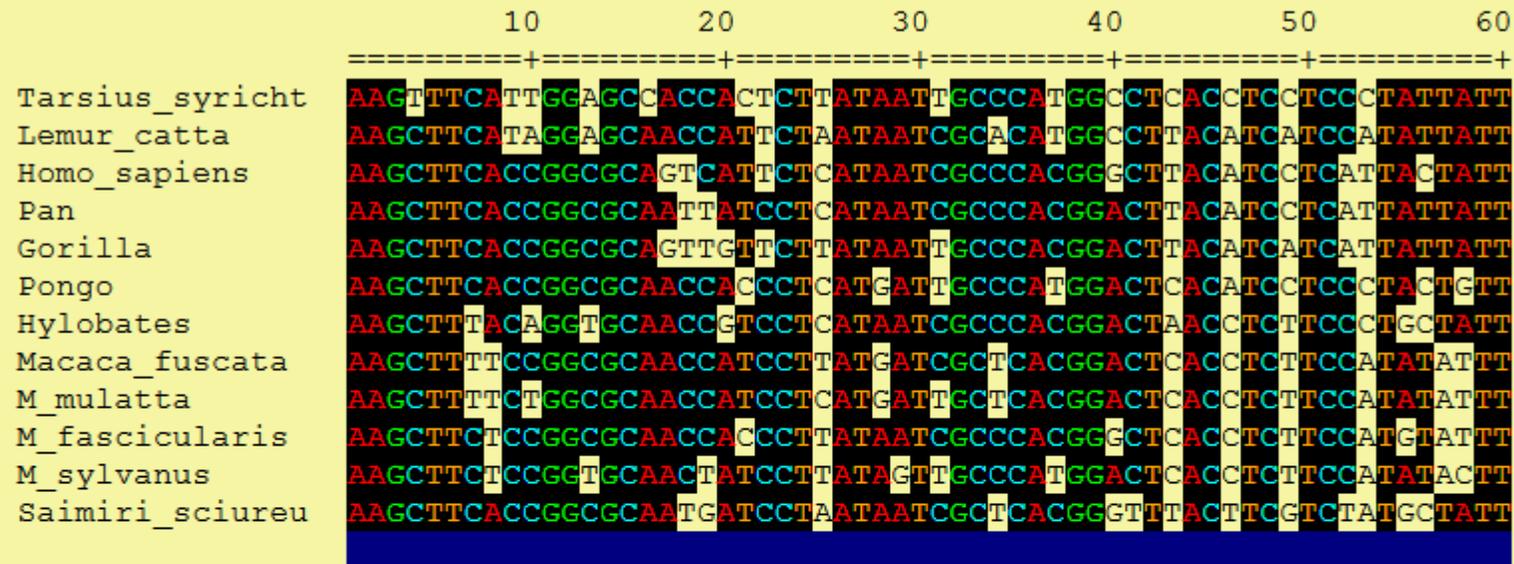


Figure 1: Selected

Input:

Input raw alignment

Outputs:

◆ Cured alignment in FASTA Format

◆ Cured alignment in PHYLIP Format

Conserved positions: 98%

Clique droit pour enregistrer le fichier d'alignement

# Construction d'arbre en utilisant IQ-Tree

<https://www.hiv.lanl.gov/content/sequence/IQTREE/iqtree.html>

**IQ-TREE web server: fast and accurate phylogenetic trees under maximum likelihood**

Server load: 31%

Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ (2016) *Nucl. Acids Res.* 44 (W1): W232-W235. doi: [10.1093/nar/gkw256](https://doi.org/10.1093/nar/gkw256)

Tree Inference

Model Selection

Analysis Results

For a quick start, take a look at the [tutorial](#) for the IQ-TREE web server.

Please visit the [IQ-TREE homepage](#) for more information or if you want to download the main software.

Data Privacy Statement: All your personal data are strictly confidential and will not be shared with any third parties. Your data will be automatically deleted after 180 days.

## Input Data

Alignment file :

Browse...

Show example >

Use example alignment:

Yes

?

Sequence type:

Auto-detect

DNA

Protein

Codon

?

DNA->AA

Binary

Morphology

Partition file:

This field is optional.

Browse...

Show example >

Partition type:

Edge-linked

Edge-unlinked

?

## Substitution Model Options

Substitution model:

Auto

?

FreeRate heterogeneity:

Yes [+R]

Rate heterogeneity:

Gamma [+G]

Invar. sites [+I]

?

#rate categories:

4

Choisir le fichier d'alignement sous format FASTA ou PHYLIP

Le logiciel teste une multitude des modèles d'évolution de séquences et choisi celui qui donne le meilleur arbre au vu des séquences

# Construction d'arbre en utilisant IQ-Tree

## Branch Support Analysis

**Bootstrap analysis:**  None  Ultrafast  Standard

**Number of bootstrap alignments:**

**Create .ufboot file:**  Yes (write bootstrap trees to .ufboot file)

**Maximum iterations:**

**Minimum correlation coefficient:**

**Single branch tests:**

**SH-aLRT branch test:**  No  Yes #replicates:

**Approximate Bayes test:**  Yes

Pour petit nombre de séquences (<20), analyse bootstrap classique, changer « number of bootstrap alignments » à 100

Pour grand nombre de séquences (>20), analyse Ultrafast bootstrap, 1000 alignments

## IQ-TREE Search Parameters

**Perturbation strength:**

**IQ-TREE stopping rule:**

**Email (optional, to retrieve results):**

**SUBMIT JOB**

GO!

# Construction d'arbre en utilisant IQ-Tree

IQ-TREE web server: fast and accurate phylogenetic trees under maximum likelihood

Server load: 28%

Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ (2016) *Nucl. Acids Res.* 44 (W1): W232-W235. doi: [10.1093/nar/gkw256](https://doi.org/10.1093/nar/gkw256)

Tree Inference

Model Selection

Analysis Results

User name or Email:

guest

QUERY STATUS

<input checked="" type="checkbox"/>	No.	Submission Time ▾	Status
<input checked="" type="checkbox"/>	1	2018-11-27 13:00	Running

Summary

Run Log

Full Result

Please bookmark the following link to later monitor/retrieve results:

<http://iqtree.cibiv.univie.ac.at/?user=guest&jobid=181127130014>

Cliquer sur le lien pour rafraichir la page

If you want to monitor the progress, click on Run Log above. If you hit QUERY STATUS, the page is reloaded. You can [download IQ-TREE](#) and run it locally with the command-line:

```
path_to_iqtree -s GBlock_Sua5_Archaea_default.fasta -m TEST -bb 1000 -alrt 1000
```

**Note:** The CPU time limit is 24 hours and RAM limit is 1GB. Your job will be stopped if it exceeds these limits. (In that case, please download the stopped job and use the above command-line to resume the run from the last checkpoint on your local PC as described [here](#).)

# Construction d'arbre en utilisant IQ-Tree

IQ-TREE web server: fast and accurate phylogenetic trees under maximum likelihood

Server load: 28%

Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ (2016) *Nucl. Acids Res.* 44 (W1): W232-W235. doi: [10.1093/nar/gkw256](https://doi.org/10.1093/nar/gkw256)

Tree Inference Model Selection Analysis Results

Page de résumé

User name or Email: guest

QUERY STATUS

<input checked="" type="checkbox"/>	No.	Submission Time	Status
<input checked="" type="checkbox"/>	1	2018-11-27 13:00	Success

Summary Run Log Full Result

Please bookmark the following link to later monitor/retrieve results:

<http://iqtree.cbiv.univie.ac.at/?user=guest&jobid=184127130014>

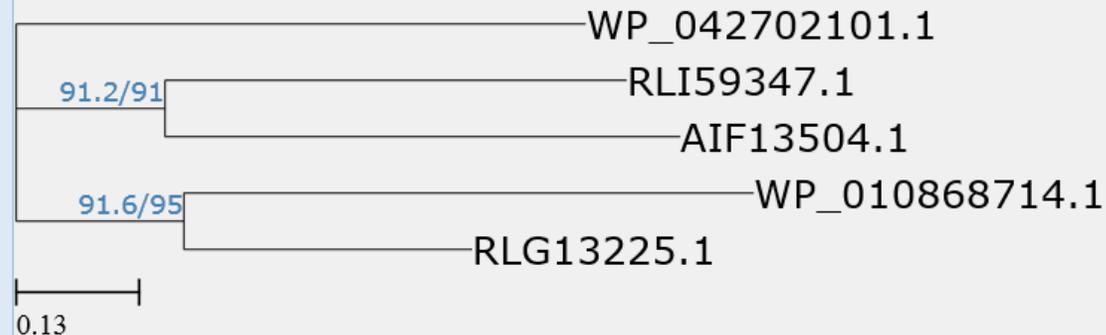
If you want to monitor the progress, click on Run Log above. If you hit QUERY STATUS, the page is reloaded. You can [download IQ-TREE](#) and run it locally with the command-line:

```
path_to_iqtree -s GBlock_Sua5_Archaea_default.fasta -m TEST -bb 1000 -alrt 1000
```

Note: The CPU time limit is 24 hours and RAM limit is 1GB. Your job will be stopped if it exceeds these limits. (In that case, please download the stopped job and use the above command-line to resume the run from the last checkpoint on your local PC as described [here](#).)

**Job has FINISHED!** Detailed analysis can be found in the Full Result tab. You can click on [DOWNLOAD SELECTED JOBS](#) to retrieve all result files.

The tree below was visualized with [ete-view](#). Support values written on the branches: SH-aLRT support (%) / ultrafast bootstrap support (%)



Cliquer pour obtenir le résultat complet

# Construction d'arbre en utilisant IQ-Tree

## SEQUENCE ALIGNMENT

```
-----  
Input data: 5 sequences with 169 amino-acid sites  
Number of constant sites: 27 (= 15.9763% of all sites)  
Number of invariant (constant or ambiguous constant) sites: 27 (= 15.9763% of all sites)  
Number of parsimony informative sites: 35  
Number of distinct site patterns: 151
```

Les informations  
sur l'alignement

## ModelFinder

```
-----  
Best-fit model according to BIC: Blos62+I
```

Le modèle d'évolution de séquences  
donnant le meilleur arbre

# Construction d'arbre en utilisant IQ-Tree

L'arbre de maximum vraisemblance

MAXIMUM LIKELIHOOD TREE

Log-likelihood of the tree: -1827.6917 (s.e. 47.6123)  
Unconstrained log-likelihood (without tree): -834.3614  
Number of free parameters (#branches + #model parameters): 8  
Akaike information criterion (AIC) score: 3671.3835  
Corrected Akaike information criterion (AICc) score: 3672.2835  
Bayesian information criterion (BIC) score: 3696.4227

Total tree length (sum of branch lengths): 2.8223  
Sum of internal branch lengths: 0.3282 (11.6304% of tree length)

NOTE: Tree is UNROOTED although outgroup taxon 'WP\_042702101.1' is drawn at root  
Numbers in parentheses are SH-aLRT support (%) / standard bootstrap support (%)

```
-----WP_042702101.1
|
|
|
+-----RLI59347.1
| (91/92)
|
+-----AIF13504.1
|
|
+-----WP_010868714.1
| (92.2/98)
|
+-----RLG13225.1
```

Tree in newick format:

```
(WP_042702101.1:0.5907845559,(RLI59347.1:0.4788295087,AIF13504.1:0.5343783778)91/92:0.1543572719,(WP_010868714.1:0.5910644208,RLG13225.1:0.2989554890)92.2/98:0.1738816251);
```

Soutient pour la branche: on considère que la branche est robuste quand SH-aLRT > 80 % et le bootstrap > 90%

L'arbre sous format Newick

# Construction d'arbre en utilisant IQ-Tree

L'arbre consensus

```
CONSENSUS TREE
-----
Consensus tree is constructed from 100bootstrap trees
Log-likelihood of consensus tree: -1827.691779
Robinson-Foulds distance between ML tree and consensus tree: 0

Branches with support >0.000000% are kept (extended consensus)
Branch lengths are optimized by maximum likelihood on original alignment
Numbers in parentheses are bootstrap supports (%)

+-----WP_042702101.1
|
+-----RLI59347.1
| (92)
+-----AIF13504.1
|
+-----WP_010868714.1
| (98)
+-----RLG13225.1

Consensus tree in newick format:
(WP_042702101.1:0.5907348262, (RLI59347.1:0.4787256895,AIF13504.1:0.5342567529) 92:0.1543493891, (WP_010868714.1:0.5910255765,RLG13225.1:0.2983703428) 98:0.1739553161);
```

Soutient pour la branche: on considère que la branche est robuste quand bootstrap > 90%

L'arbre sous format Newick

# Edition d'arbres avec iTOL

<https://itol.embl.de/>

ITOL INTERACTIVE TREE OF LIFE

Tree of Life Upload Sharing data Help

Login

Permet de charger un arbre sous différents formats

**Create high quality figures**  
iTOL provides WYSIWYG high quality export into vector graphics formats. Interactively adjust the tree as you export. The exported figure will mirror what is shown on your screen.

Vous pouvez créer un compte si vous souhaitez sauvegarder vos arbres

Current changelog: [version 4.1](#)

# Upload a new tree

Use this page to upload and visualize a new phylogenetic tree. It should be in a plain text file and in a supported format (Newick, Nexus or PhyloXML). You can also use *.jplace* files generated by RaxML or pplacer, or *.qza* trees generated by QIIME 2. Please check the [help pages](#) for detailed instructions.

Trees uploaded anonymously will be stored for 30 days, and are not protected from modifications by other users. If you want to keep them private and protected, or have multiple trees to visualize, we recommend creating [an iTOL personal account](#). If you already have an account, please [login first](#).

Datasets and other annotation should be dragged and dropped directly onto the interactive tree. Please check the [help pages](#) for detailed instructions and dataset template files.

## Upload a new tree

Tree name:

optional

Vous pouvez donner un nom à votre arbre ici

Paste your tree into the box below, or select a file using the **Tree file selector**. You can also simply drag and drop the tree file onto the page (only a regular plain text file, not QIIME QZA files).

Tree text:

Inserez votre arbre ici sous forme Newick

Tree file:

Parcourir...

Aucun fichier sélectionné.

Ou chargez un fichier en format adéquat (Newick, Nexus ou PhyloXML)

Upload

GO!

# Edition d'arbres avec iTOL

ITOL INTERACTIVE TREE OF LIFE Tree of Life Upload Sharing Help

Tree scale: 0.1

EZG56502.1  
WP 010868714.1  
RLG13225.1  
WP 042702101.1  
RLI59347.1  
AIE13504.1

Afficher les valeurs de bootstrap sur les branches: Parametres avancés

Controls

Basic **Advanced** Datasets Export

Display mode Circular **Normal** Unrooted

Parameters 0 ° rotation

Invert Yes No

Slanted Yes No

Branch lengths Use Ignore

Labels Aligned At tips Off

Label rotation On Off

Label shift 0 Dashes On Off

Label font Arial Add

Font style 20 px B I

Branch lines 1 px

Save/restore view Reset all

# Edition d'ar

**Controls**

Basic Advanced Datasets Export

**Scaling factors** Hor: 1 Vert: 1

**Leaf sorting** Default None

**Branch lengths** Display Hide

**Bootstraps / metadata** Display Hide

**Data source:** bootstrap

**Display range:** 91 - 98

Symbol Text Color Width

Circle Triangle Square Star

Min size: 5 px Fill: 

Max size: 15 px

Position on branch: 50 %

**Internal node symbols** All One child Off

**Internal tree scale** Display Hide

**Auto collapse clades** BRL <  Collapse

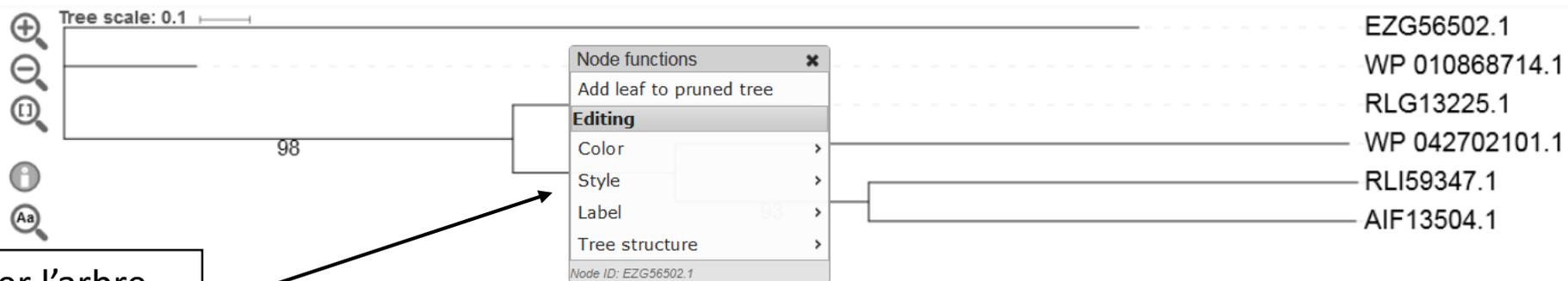
**Delete branches** Bootstrap <  Delete

**Auto assign taxonomy** Assign

Save/restore view Reset all

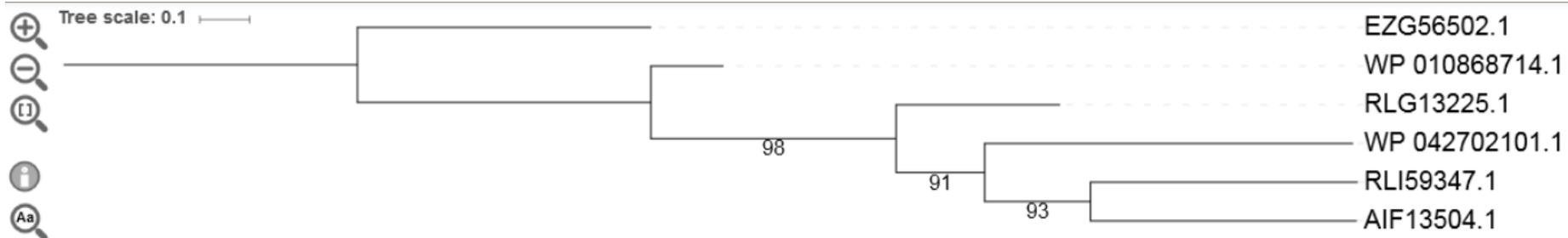
Afficher les valeurs de bootstrap sur les branches: Display puis text, choisissez la taille de police 16

# Edition d'arbres avec iTOL



Pour enracer l'arbre choisir la branche de l'outgroup, puis clique gauche sur cette branche ouvre le menu. Choisir « Tree structure » puis « Reroot the tree here »

Arbre enraciné entre l'outgroup et le groupe étudié:



# Edition d'arbres avec iTOL

ITOL INTERACTIVE TREE OF LIFE Tree of Life Upload Sharing Help Login

Tree scale: 0.1

EZG56502.1

- Node functions
- Add leaf to pruned tree
- Editing
  - Color
  - Style
  - Label
  - Tree structure
- Node ID: EZG56502.1

Controls

Basic Advanced Dataset

Scaling factors Hor:

Leaf sorting

Branch lengths

Set label color

Set label background

Edit: EZG56502.1

Style: **B** *I*

Size: 1 x

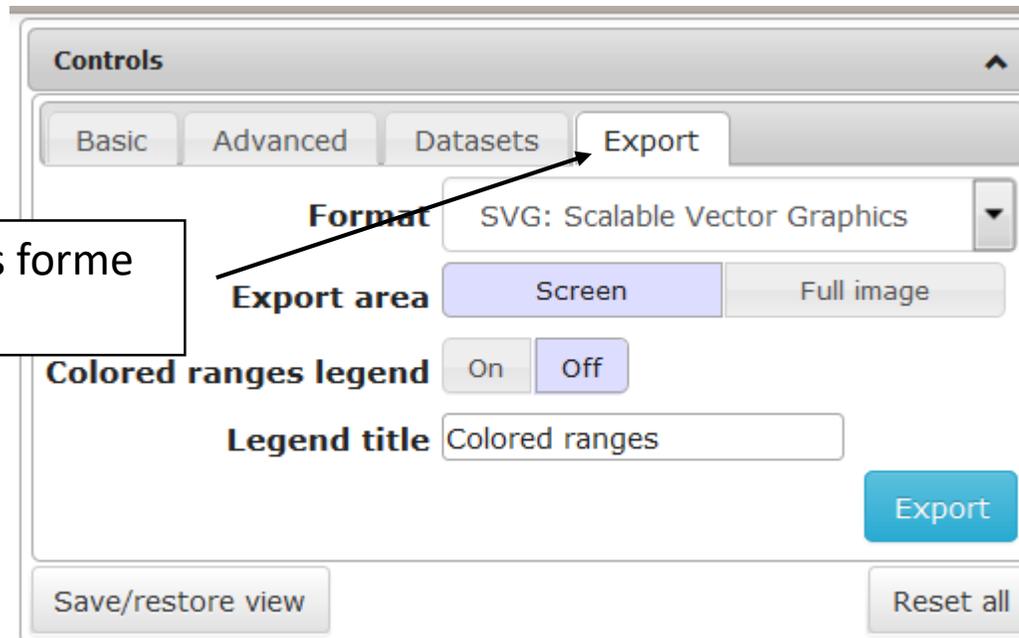
Pour renommer les séquences, click gauche sur le nom, puis « Label » puis « Edit »

Vous devez sauvegarder les changements à chaque fois.

Vous pouvez également choisir la couleur et la taille de police

# Edition d'arbres avec iTOL

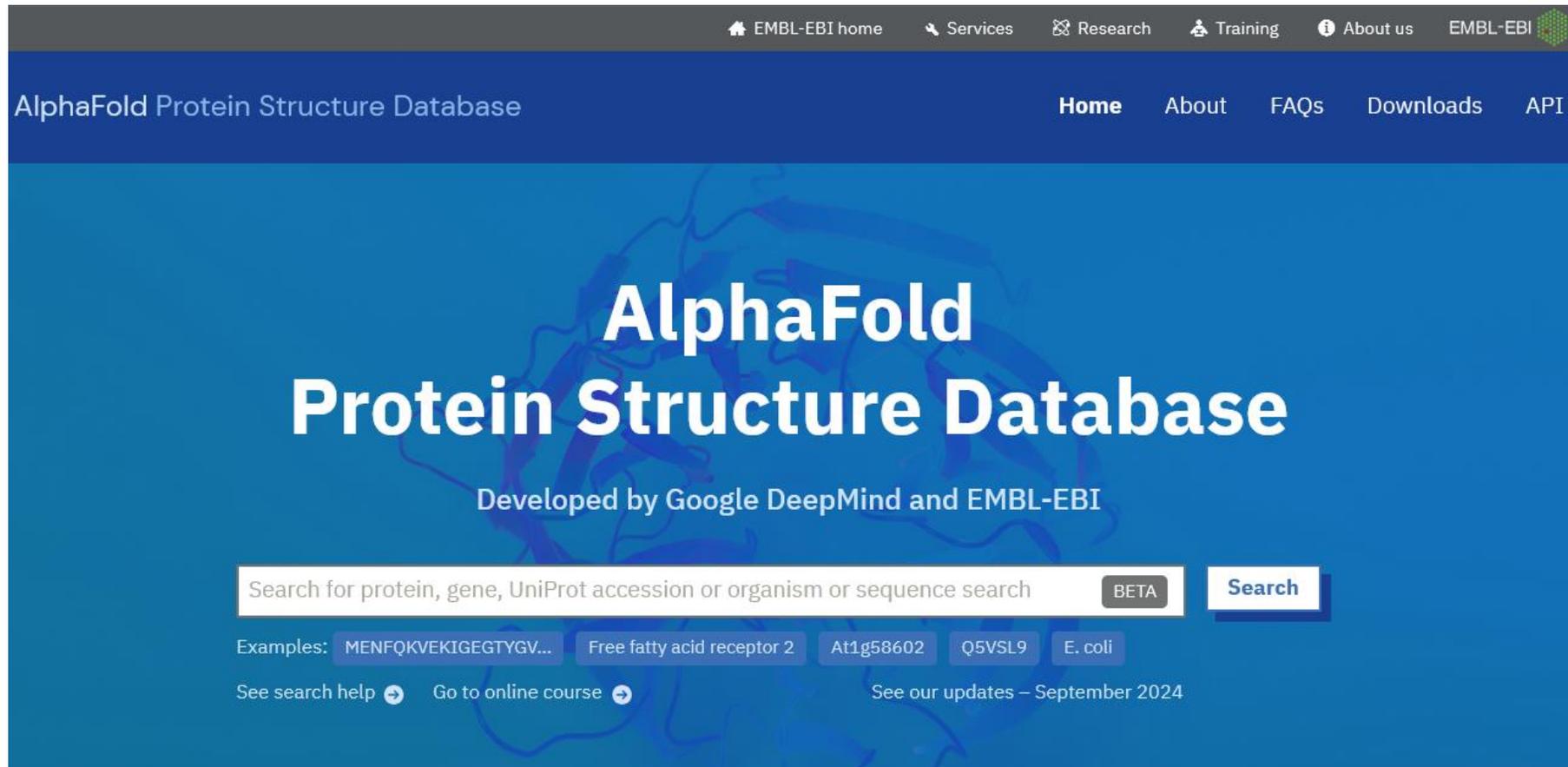
Pour récupérer votre arbre sous forme d'une image vectorielle



# AlphaFold Protein Structure Database

<https://alphafold.ebi.ac.uk/>

**Répertorie 200 millions de structures prédites**



The screenshot shows the homepage of the AlphaFold Protein Structure Database. At the top, there is a navigation bar with links for EMBL-EBI home, Services, Research, Training, About us, and EMBL-EBI logo. Below this, the site title "AlphaFold Protein Structure Database" is displayed on the left, and navigation links for Home, About, FAQs, Downloads, and API are on the right. The main content area features a large blue background with a faint protein structure. The title "AlphaFold Protein Structure Database" is prominently displayed in white, followed by the text "Developed by Google DeepMind and EMBL-EBI". A search bar is located below the title, with the placeholder text "Search for protein, gene, UniProt accession or organism or sequence search" and a "BETA" label. A "Search" button is positioned to the right of the search bar. Below the search bar, there are examples of search terms: "MENFQKVEKIGEGTYGV...", "Free fatty acid receptor 2", "At1g58602", "Q5VSL9", and "E. coli". At the bottom, there are links for "See search help", "Go to online course", and "See our updates – September 2024".

EMBL-EBI home Services Research Training About us EMBL-EBI

AlphaFold Protein Structure Database Home About FAQs Downloads API

# AlphaFold Protein Structure Database

Developed by Google DeepMind and EMBL-EBI

Search for protein, gene, UniProt accession or organism or sequence search **BETA** **Search**

Examples: MENFQKVEKIGEGTYGV... Free fatty acid receptor 2 At1g58602 Q5VSL9 E. coli

See search help Go to online course See our updates – September 2024

# AlphaFold Protein Structure Database

EMBL-EBI home Services Research Training About us EMBL-EBI

AlphaFold Protein Structure Database Home About FAQs Downloads API

# AlphaFold Protein Structure Database

Developed by Google DeepMind and EMBL-EBI

Search for protein, gene, UniProt accession or organism or sequence search **BETA** **Search**

Examples: MENFQKVEKIGEGTYGV... Free fatty acid receptor 2 At1g58602 Q5VSL9 E. coli

See search help → Go to online course → See our updates – September 2024

Vous pouvez chercher avec le Nr. d'accèsion UniProt, nom d'un organisme, une séquence protéique etc.

# AlphaFold Protein Structure Database

Recherche avec UniProt Q5JH81:

Showing all search results for Q5JH81

1 - 1 of 1 results

Filter by:

Status

Review

Reviewed (Swiss-Prot) (1)

Reference proteome

Show predictions for sequences found only in UniProt reference proteomes (1)

Average pLDDT score



Show results with the score between

and  [Apply score range](#)

Other organisms

[Thermococcus kodakarensis \(strain ATCC BAA-918 / JCM 12380 / KOD1\)](#) (1)

Type 2 DNA topoisomerase 6 subunit B

AF-Q5JH81-F1-v4

Protein Type 2 DNA topoisomerase 6 subunit B

Gene top6B

Source Organism [Thermococcus kodakarensis \(strain ATCC BAA-918 / JCM 12380 / KOD1\)](#) [search this organism](#)

UniProt [Q5JH81](#) [go to UniProt](#)

Average pLDDT 90.4 (Very High)

Sequence length 569

Nom de la protéine

L'organisme

Score de confiance en la prédiction

“pLDDT is a per-residue measure of local confidence, scaled from 0 to 100. Higher scores indicate higher confidence and usually a more accurate prediction.”

Cliquer sur le nom de la protéine pour visualiser la structure

# AlphaFold Protein Structure Database

## Type 2 DNA topoisomerase 6 subunit B

AF-Q5JH81-F1-v4

Download

PDB file

mmCIF file

Predicted aligned error

Share your feedback on structure with Google DeepMind

Looks great

Could be improved

### Information

Protein	Type 2 DNA topoisomerase 6 subunit B
Gene	top6B
Source organism	Thermococcus kodakarensis (strain ATCC BAA-918 / JCM 12380 / KOD1) (Pyrococcus kodakaraensis (strain KOD1)) <a href="#">go to search</a> 
UniProt	Q5JH81 <a href="#">go to UniProt</a> 
Experimental structures	None available in the PDB
Biological function	Relaxes both positive and negative superturns and exhibits a strong decatenase activity. <a href="#">go to UniProt</a> 

Vous pouvez télécharger le fichier PDB – vous pouvez utiliser ce fichier pour chercher des structures similaires dans Foldseek (voir diapos plus loin)

# AlphaFold Protein Structure Database

## Structure viewer

Sequence of AF-Q5JH81-F1 Chain 1: Type 2 DNA ... A

MAEANQLFKE FKI QSVSEFF RRNAAMLGYT GKVRSLT TLV HEAVINSLDACEEAGIPPVVRVEI  
EELGNEHYKVVVEDNGPGIPEKYITHVFGKMLAGTKAHRNIQSRGQ QGIGISGAVMFAQITSGK  
ATRVIITSGNDE IIEAWVKIDV DKNNEGKIVKKEKHPNPKGWR GTRIELEVKIVKYMRSKQGV FW

Structure Tools

Structure

AF-Q5JH81-F1

Type Model

Nothing Focused

Quick Styles

Default Illustrative Stylize Current

Components AF-Q5JH81-F1

Preset + Add

Polymer Cartoon

Measurements

+ Add

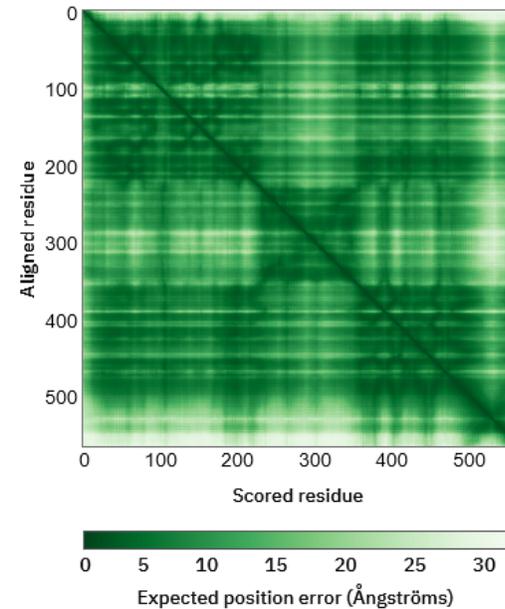
### Model Confidence

Very high (pLDDT > 90) High (90 > pLDDT > 70) Low (70 > pLDDT > 50) Very low (pLDDT < 50)

AlphaFold produces a per-residue model confidence score (pLDDT) between 0 and 100. Some regions below 50 pLDDT may be unstructured in isolation.

Vous pouvez manipuler la structure

Cliquer sur cette icone pour récupérer une image de la structure



### Predicted aligned error (PAE)

Click and drag a box on the PAE viewer to select regions of the structure and highlight them on the 3D viewer.

PAE data is useful for assessing inter-domain accuracy – go to [Help section below](#) for more information.

PAE permet d'évaluer l'erreur (exprimé en Angströms) dans le positionnement des résidus les uns aux autres. Cela vous permet de voir quelle est la confiance du programme dans la prédiction de l'interaction entre différents domaines du polypeptide. En effet, il est possible que la prédiction de structure de domaines individuels au sein d'un polypeptide est très robuste alors que leur positionnement mutuel est prédit avec une grande marge d'erreur (PAE)

## ColabFold - utilisation d'AlphaFold 2

Si votre protéine n'est pas répertoriée dans AlphaFold Protein Structure Database vous pouvez faire vous-même une prédiction en utilisant ColabFold:

# ColabFold - utilisation d'AlphaFold 2

**Vous avez besoin d'une adresse gmail pour pouvoir utiliser ColabFold**

AlphaFold2.ipynb  
Fichier Modifier Affichage Insérer Exécution Outils Aide

+ Code + Texte Copier sur Drive

RAM Disque Modification

## ColabFold: AlphaFold2 w/ MMseqs2

Easy to use AlphaFold2 protein structure ([Jumper et al. 2021](#)) and complex ([Evans et al. 2021](#)) prediction using multiple sequence alignments generated through MMseqs2. For details, refer to our manuscript: [Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold - Making protein folding accessible to all. bioRxiv, 2021](#)

- This notebook provides basic functionality for **protein structure (AlphaFold2)** and **complex prediction (AlphaFold2-multimer)**. Advanced features such as recycles, sampling, ... -> [advanced notebook](#).
- This notebook replaces the homology detection and MSA pairing of AlphaFold2 with MMseqs2. For a comparison against the [Deepmind Colab](#) and the full [AlphaFold2](#) system read our [preprint](#).

Old versions: [v1.0-alpha](#), [v1.1-permultimer](#)

For more details, see [bottom](#) of the notebook and checkout the [ColabFold GitHub](#).

[1] Input protein sequence, then hit Runtime -> Run all

```
query_sequence: "PIAQIHILEGRSDEQKETLIREVSEAI SRSLDAPLTSVRVIITEMAKGHFGIGGELASK"
```

• Use : to specify inter-protein chainbreaks for **modeling complexes** (supports homo- and hetro-oligomers). For example **PI...SK:PI...SK** for a mono-dimer

```
jobname: "test"
```

Insérer la séquence sous format brut

Nom de la séquence, du projet...

# ColabFold - utilisation d'AlphaFold 2



AlphaFold2.ipynb

Fichier Modifier Affichage Insérer Exécution Outils Aide

Pour démarrer le calcul, cliquer sur « exécution » puis « tout exécuter »

+ Code + Texte Copier sur Drive

## ColabFold: AlphaFold2 w/ MMseqs2

Easy to use AlphaFold2 protein structure ([Jumper et al. 2021](#)) and complex ([Evans et al. 2021](#)) prediction using multiple sequence alignments generated through MMseqs2. For details, refer to our manuscript:

[Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold - Making protein folding accessible to all. bioRxiv, 2021](#)

- This notebook provides basic functionality for **protein structure (Alphafold2)** and **complex prediction (Alphafold2-multimer)**. Advanced features such as recycles, sampling, ... ->[advanced notebook](#).
- This notebook replaces the homology detection and MSA pairing of AlphaFold2 with MMseqs2. For a comparison against the [Deepmind Colab](#) and the full [AlphaFold2](#) system read our [preprint](#).

Old versions: [v1.0-alpha](#), [v1.1-permultimer](#)

For more details, see [bottom](#) of the notebook and checkout the [ColabFold GitHub](#).



[1] Input protein sequence, then hit `Runtime -> Run all`

```
query_sequence: "PIAQIHILEGRSDEQKETLIREVSEAISRSLDAPLTSVRVIITEMAKGHFGIGGELASK"
```

## ColabFold - utilisation d'AlphaFold 2

→ À la fin du calcul un fichier .zip peut être téléchargé contenant tous les résultats (y compris les fichiers pdb des modèles, sinon vous pouvez visualiser directement dans le navigateur les résultats pour le meilleur modèle :

# ColabFold - utilisation d'AlphaFold 2

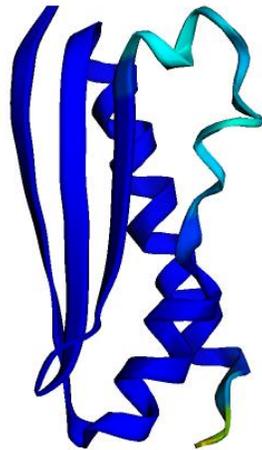
Résultats :

color: IDDT

show\_sidechains:

show\_mainchains:

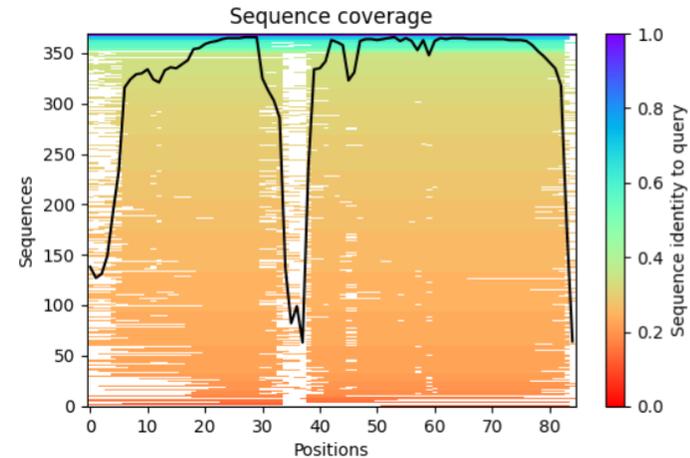
[Afficher le code](#)



Structure prédite : la couleur indique la robustesse de la prédiction

Rouge : la prédiction est très peu fiable

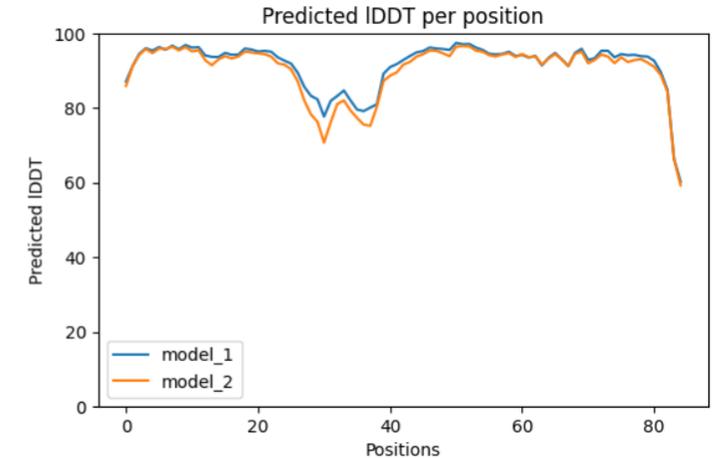
Bleu : la prédiction est très fiable



Ordonnée : nombre des séquences qui ont été alignées avec votre séquence pour créer le modèle

Abscisse : la position dans la séquence (ici votre séquence contient environ 85 résidus)

Chaque ligne correspond à une séquence, code couleur indique le % d'identité de séquence par rapport à votre protéine



Ordonnée : un indicateur de la qualité de prédiction

Abscisse : la position dans la séquence (ici votre séquence contient environ 85 résidus)

# ColabFold - utilisation d'AlphaFold 2

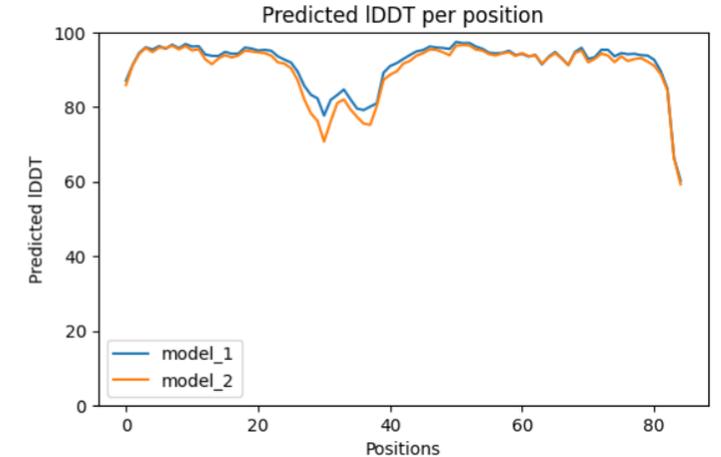
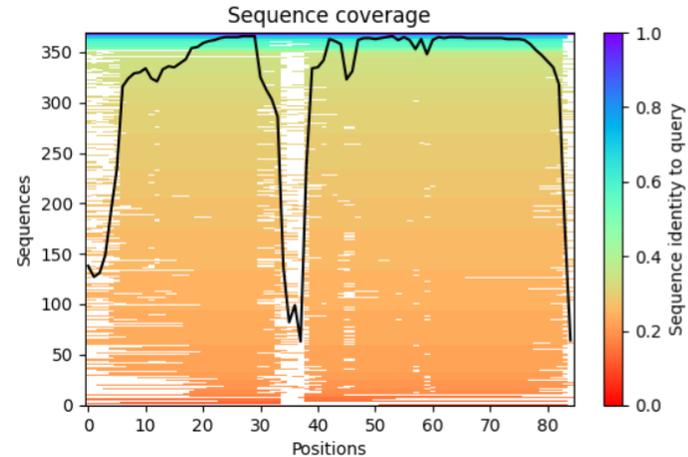
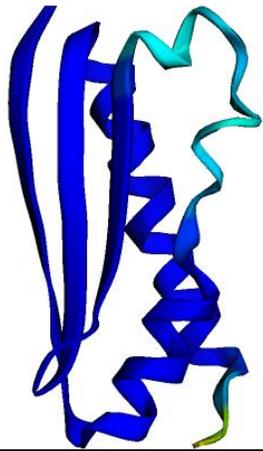
Résultats :

color: IDDT

show\_sidechains:

show\_mainchains:

[Afficher le code](#)



Cette image montre qu'environ 350 séquences ont été alignées pour calculer le modèle de la structure. On note que les résidus 30-40 sont présents chez une minorité des séquences y compris la séquence qui a été analysée.

On note également faible conservation de séquence en particulier en N-ter et C-ter

Cet indicateur montre le niveau de confiance (allant de 0 à 100) de prédiction. On voit que ce niveau est globalement très élevé sauf pour la région 30-40 et les deux extrémités de la séquence.

Structure prédite : couleur bleu indique que cette prédiction est globalement très fiable mis à part une boucle qui est affichée en vert. Cette boucle correspond aux résidus 30-40 AA qui sont sous-représentés dans l'alignement des séquences (voir le graphe à droite)

## Foldseek: chercher avec la structure 3D dans AFDB

Avec le fichier PDB que vous avez obtenu avec ColabFold ou en cherchant dans AlphaFold Protein Structure Database vous pouvez utiliser FoldSeek pour chercher les structures similaires:

**<https://search.foldseek.com/search>**

# Foldseek: chercher avec la structure 3D dans AFDB

Thermococcus kodakarensis KC x | ORFfinder Viewer - NCBI x | Taxonomy browser (Thermococ x | NCBI Blast:Protein Sequence x | Transcription factor Pcc1 (PF09: x | taxonomy results | HMMER x | Foldseek Search Server x

search.foldseek.com/search

Foldseek Search  GITHUB SÖDING LAB STEINEGGER LAB

Input protein structure (PDB) or sequence (FASTA) API ?

ATOM	866	N	PHE	A 111	11.187	-12.768	-6.000
ATOM	867	CA	PHE	A 111	11.895	-11.516	-5.804
ATOM	868	C	PHE	A 111	13.203	-11.457	-6.592
ATOM	870	CB	PHE	A 111	12.169	-11.360	-4.310
ATOM	877	N	GLY	A 112	13.543	-10.277	-7.094
ATOM	878	CA	GLY	A 112	14.800	-10.107	-7.788
ATOM	879	C	GLY	A 112	14.816	-9.982	-9.286
ATOM	881	N	TYR	A 113	13.670	-10.112	-9.938
ATOM	882	CA	TYR	A 113	13.648	-10.024	-11.397
ATOM	883	C	TYR	A 113	12.764	-8.904	-11.929
ATOM	885	CB	TYR	A 113	13.182	-11.355	-11.997

LOAD ACCESSION **UPLOAD PDB** PREDICT UPLOAD PREVIOUS RESULTS

**Databases & search settings**

**SEARCH** **Summary**  
Search **all available** databases with Foldseek in **3Di/AA** mode.

**Reference**  
van Kempen M, Kim S, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, Söding J, and Steinegger M. [Fast and accurate protein structure search with Foldseek](#). Nature Biotechnology, 2023.

**Charger le fichier PDB et lancer la recherche**



# Foldseek: chercher avec la structure 3D dans AFDB

Nr. de résultats dans les différentes bases de données (PDB100 contient des structures réelles (déterminées expérimentalement))

Results: job.pdb

[ALL DATABASES](#)
 AFDB-PROTEOME (22)
 AFDB-SWISSPROT (19)
 AFDB50 (1000)
 CATH50 (9)
 GMGCL\_ID (5)
 MGNIFY\_ESM30 (1000)
 PDB100 (13)

AFDB-PROTEOME 22 hits

GRAPHICAL NUMERIC

Target	Description	Scientific Name	Prob.	Seq. Id.	E-Value	Position in query	Alignment
<a href="#">AF-P0A1G5-F1-model_v4</a>	RNA polymerase-binding transcription...	<a href="#">Salmonella enterica subsp. enterica s...</a>	1.00	100	6.18e-5	1-41	≡
<a href="#">AF-A0A0H3GIE5-F1-model_v4</a>	RNA polymerase-binding transcription...	<a href="#">Klebsiella pneumoniae subsp. pneum...</a>	1.00	100	6.18e-5	1-41	≡
<a href="#">AF-Q32JW2-F1-model_v4</a>	RNA polymerase-binding transcription...	<a href="#">Shigella dysenteriae Sd197</a>	1.00	100	7.18e-5	1-41	≡
<a href="#">AF-P0ABS1-F1-model_v4</a>	RNA polymerase-binding transcription...	<a href="#">Escherichia coli K-12</a>	1.00	100	7.18e-5	1-41	≡
<a href="#">AF-G3XD14-F1-model_v4</a>	RNA polymerase-binding transcription...	<a href="#">Pseudomonas aeruginosa PAO1</a>	1.00	82.9	2.06e-4	1-41	≡
<a href="#">AF-P43758-F1-model_v4</a>	RNA polymerase-binding transcription...	<a href="#">Haemophilus influenzae Rd KW20</a>	1.00	90.2	3.00e-4	1-41	≡
<a href="#">AF-Q9HT38-F1-model_v4</a>	RNA polymerase-binding transcription...	<a href="#">Pseudomonas aeruginosa PAO1</a>	1.00	48.7	1.75e-2	1-39	≡
<a href="#">AF-Q0PC12-F1-model_v4</a>	DksA C4-type domain-containing protein	<a href="#">Campylobacter jejuni subsp. jejuni NC...</a>	1.00	45.9	5.82e-2	1-37	≡
<a href="#">AF-Q5F5L9-F1-model_v4</a>	RNA polymerase-binding transcription...	<a href="#">Neisseria gonorrhoeae FA 1090</a>	1.00	41.4	6.76e-2	1-41	≡
<a href="#">AF-Q93GN5-F1-model_v4</a>	Conjugative transfer	<a href="#">Salmonella enterica subsp. enterica s...</a>	0.99	30	3.54e-1	1-40	≡
<a href="#">AF-Q9HVK7-F1-model_v4</a>	DksA C4-type domain-containing protein	<a href="#">Pseudomonas aeruginosa PAO1</a>	0.98	42.4	4.44e-1	1-33	≡
<a href="#">AF-A0A0H3GKN8-F1-model_v4</a>	Phage/conjugal plasmid C-4 type zinc...	<a href="#">Klebsiella pneumoniae subsp. pneum...</a>	0.94	35.8	9.41e-1	1-39	≡
<a href="#">AF-P41039-F1-model_v4</a>	Uncharacterized protein Ybil	<a href="#">Escherichia coli K-12</a>	0.93	37.1	7.51e-1	1-35	≡
<a href="#">AF-Q8ZQN5-F1-model_v4</a>	Putative DnaK suppressor protein	<a href="#">Salmonella enterica subsp. enterica s...</a>	0.91	34.1	7.51e-1	1-41	≡
<a href="#">AF-P44221-F1-model_v4</a>	Uncharacterized protein HI_1497	<a href="#">Haemophilus influenzae Rd KW20</a>	0.87	37.5	1.72e+0	2-33	≡
<a href="#">AF-Q32I79-F1-model_v4</a>	DksA C4-type domain-containing protein	<a href="#">Shigella dysenteriae Sd197</a>	0.85	37.1	1.18e+0	1-35	≡
<a href="#">AF-Q9HUU3-F1-model_v4</a>	DksA C4-type domain-containing protein	<a href="#">Pseudomonas aeruginosa PAO1</a>	0.84	29.2	1.27e+0	1-41	≡
<a href="#">AF-G3XDA6-F1-model_v4</a>	Repressor, PtrB	<a href="#">Pseudomonas aeruginosa PAO1</a>	0.72	38.3	2.15e+0	1-41	≡

# Foldseek: chercher avec la structure 3D dans AFDB

Probabilité qu'il s'agit de la même structure

Identité de séquence entre votre protéine et le résultat de recherche

E-value, le résultat est-il statistiquement significatif

Alignement de séquence entre votre protéine et le résultat

Results: job.pdb

ALL DATABASES    AFDB-PROTEOME (22)    AFDB-SWISSPROT (19)    AFDB50 (1000)    CATH50 (9)    GMGCL\_ID (5)    MGNIFY\_ESM30 (1000)    PDB100 (13)

AFDB-PROTEOME 22 hits

GRAPHICAL    NUMERIC

Target	Description	Scientific Name	Prob.	Seq. Id.	E-Value	Position in query	Alignment
<a href="#">AF-P0A1G5-F1-model_v4</a>	RNA polymerase-binding transcription...	<a href="#">Salmonella enterica subsp. enterica s...</a>	1.00	100	6.18e-5	1-41	≡
<a href="#">AF-A0A0H3GIE5-F1-model_v4</a>	RNA polymerase-binding transcription...	<a href="#">Klebsiella pneumoniae subsp. pneum...</a>	1.00	100	6.18e-5	1-41	≡
<a href="#">AF-Q32JW2-F1-model_v4</a>	RNA polymerase-binding transcription...	<a href="#">Shigella dysenteriae Sd197</a>	1.00	100	7.18e-5	1-41	≡
<a href="#">AF-P0ABS1-F1-model_v4</a>	RNA polymerase-binding transcription...	<a href="#">Escherichia coli K-12</a>	1.00	100	7.18e-5	1-41	≡
<a href="#">AF-G3XD14-F1-model_v4</a>	RNA polymerase-binding transcription...	<a href="#">Pseudomonas aeruginosa PAO1</a>	1.00	82.9	2.06e-4	1-41	≡
<a href="#">AF-P43758-F1-model_v4</a>	RNA polymerase-binding transcription...	<a href="#">Haemophilus influenzae Rd KW20</a>	1.00	90.2	3.00e-4	1-41	≡
<a href="#">AF-Q9HT38-F1-model_v4</a>	RNA polymerase-binding transcription...	<a href="#">Pseudomonas aeruginosa PAO1</a>	1.00	48.7	1.75e-2	1-39	≡
<a href="#">AF-Q0PC12-F1-model_v4</a>	DksA C4-type domain-containing protein	<a href="#">Campylobacter jejuni subsp. jejuni NC...</a>	1.00	45.9	5.82e-2	1-37	≡
<a href="#">AF-Q5F5L9-F1-model_v4</a>	RNA polymerase-binding transcription...	<a href="#">Neisseria gonorrhoeae FA 1090</a>	1.00	41.4	6.76e-2	1-41	≡
<a href="#">AF-Q93GN5-F1-model_v4</a>	Conjugative transfer	<a href="#">Salmonella enterica subsp. enterica s...</a>	0.99	30	3.54e-1	1-40	≡
<a href="#">AF-Q9HVK7-F1-model_v4</a>	DksA C4-type domain-containing protein	<a href="#">Pseudomonas aeruginosa PAO1</a>	0.98	42.4	4.44e-1	1-33	≡
<a href="#">AF-A0A0H3GKN8-F1-model_v4</a>	Phage/conjugal plasmid C-4 type zinc...	<a href="#">Klebsiella pneumoniae subsp. pneum...</a>	0.94	35.8	9.41e-1	1-39	≡
<a href="#">AF-P41039-F1-model_v4</a>	Uncharacterized protein Ybil	<a href="#">Escherichia coli K-12</a>	0.93	37.1	7.51e-1	1-35	≡
<a href="#">AF-Q8ZQN5-F1-model_v4</a>	Putative DnaK suppressor protein	<a href="#">Salmonella enterica subsp. enterica s...</a>	0.91	34.1	7.51e-1	1-41	≡
<a href="#">AF-P44221-F1-model_v4</a>	Uncharacterized protein HI_1497	<a href="#">Haemophilus influenzae Rd KW20</a>	0.87	37.5	1.72e+0	2-33	≡
<a href="#">AF-Q32I79-F1-model_v4</a>	DksA C4-type domain-containing protein	<a href="#">Shigella dysenteriae Sd197</a>	0.85	37.1	1.18e+0	1-35	≡
<a href="#">AF-Q9HUU3-F1-model_v4</a>	DksA C4-type domain-containing protein	<a href="#">Pseudomonas aeruginosa PAO1</a>	0.84	29.2	1.27e+0	1-41	≡
<a href="#">AF-G3XDA6-F1-model_v4</a>	Repressor, PtrB	<a href="#">Pseudomonas aeruginosa PAO1</a>	0.72	38.3	2.15e+0	1-41	≡

# Foldseek: chercher avec la structure 3D dans AFDB

Results: job.pdb

ALL DATABASES AFDB-PROTEOME (22) AFDB-SWISSPROT (19) AFDB50 (1000) CATH50 (9) GMECL\_ID (5) MGNIFY\_ESM30 (1000) PDB100 (13)

AFDB-PROTEOME 22 hits

Target	Description	Scientific Name	Prob.	Seq. Id.	E-Value	Position in query	Alignment
<a href="#">AF-P0A1G5-F1-model_v4</a>	RNA polymerase-binding transcription...	<a href="#">Salmonella enterica subsp. enterica s...</a>	1.00	100	6.18e-5	1-41	
<a href="#">AF-A0A0H3GKN8-F1-model_v4</a>	RNA polymerase-binding transcription...	<a href="#">Klebsiella pneumoniae subsp. pneum...</a>	1.00	100	6.18e-5	1-41	
<a href="#">AF-Q32VQ2-F1-model_v4</a>	RNA polymerase-binding transcription...	<a href="#">Shigella dysenteriae Sd197</a>	1.00	100	7.18e-5	1-41	
<a href="#">AF-Q9M8L1-F1-model_v4</a>	RNA polymerase-binding transcription...	<a href="#">Escherichia coli K.12</a>	1.00	100	7.18e-5	1-41	
<a href="#">AF-Q30V14-F1-model_v4</a>	RNA polymerase-binding transcription...	<a href="#">Pseudomonas aeruginosa PAO1</a>	1.00	82.9	2.05e-4	1-41	
<a href="#">AF-P43758-F1-model_v4</a>	RNA polymerase-binding transcription...	<a href="#">Haemophilus influenzae Rd KW20</a>	1.00	90.2	3.00e-4	1-41	
<a href="#">AF-Q9H178-F1-model_v4</a>	RNA polymerase-binding transcription...	<a href="#">Pseudomonas aeruginosa PAO1</a>	1.00	48.7	1.75e-2	1-41	
<a href="#">AF-Q9PC12-F1-model_v4</a>	DksA C4-type domain-containing protein	<a href="#">Carnobacterium subsp. jeikei NC...</a>	1.00	45.9	5.02e-2	1-41	
<a href="#">AF-Q9F5L9-F1-model_v4</a>	RNA polymerase-binding transcription...	<a href="#">Neisseria gonorrhoeae FA1090</a>	1.00	41.4	6.75e-2	1-41	
<a href="#">AF-Q93QNS-F1-model_v4</a>	Conjugative transfer	<a href="#">Salmonella enterica subsp. enterica s...</a>	0.99	30	2.54e-1	1-41	
<a href="#">AF-Q9HVK7-F1-model_v4</a>	DksA C4-type domain-containing protein	<a href="#">Pseudomonas aeruginosa PAO1</a>	0.98	42.4	4.44e-1	1-41	
<a href="#">AF-A0A0H3GKN8-F1-model_v4</a>	Phage/conjugal plasmid C-4 type zinc...	<a href="#">Klebsiella pneumoniae subsp. pneum...</a>	0.94	35.8	9.41e-1	1-41	
<a href="#">AF-P41009-F1-model_v4</a>	Uncharacterized protein Y81	<a href="#">Escherichia coli K.12</a>	0.93	37.1	7.51e-1	1-41	
<a href="#">AF-Q8ZQNS-F1-model_v4</a>	Putative DnaK suppressor protein	<a href="#">Salmonella enterica subsp. enterica s...</a>	0.91	34.1	7.51e-1	1-41	
<a href="#">AF-P44921-F1-model_v4</a>	Uncharacterized protein HL_1497	<a href="#">Haemophilus influenzae Rd KW20</a>	0.87	37.5	1.72e+0	1-41	
<a href="#">AF-Q30V79-F1-model_v4</a>	DksA C4-type domain-containing protein	<a href="#">Shigella dysenteriae Sd197</a>	0.85	37.1	1.18e+0	1-41	
<a href="#">AF-Q9HVK7-F1-model_v4</a>	DksA C4-type domain-containing protein	<a href="#">Pseudomonas aeruginosa PAO1</a>	0.84	29.2	1.27e+0	1-41	
<a href="#">AF-P41009-F1-model_v4</a>	Uncharacterized protein Y81	<a href="#">Escherichia coli K.12</a>	0.79	36.9	9.44e+0	1-41	

Cliquer sur camembert bleu pour voir l'intégralité de votre structure.

Cliquer sur camembert jaune pour voir l'intégralité de la structure trouvée dans la base de données

Cliquer sur « Alignment » pour voir la superposition de deux structures (votre et structure de la base de données)

AFDB-PROTEOME 22 hits

Target	Description	Scientific Name	Prob.	Seq. Id.	E-Value	Position in query	Alignment
<a href="#">AF-P0A1G5-F1-model_v4</a>	RNA polymerase-binding transcription...	<a href="#">Salmonella enterica subsp. enterica s...</a>	1.00	100	6.18e-5	1-41	
<a href="#">AF-Q9HVK7-F1-model_v4</a>	DksA C4-type domain-containing protein	<a href="#">Pseudomonas aeruginosa PAO1</a>	0.98	42.4	4.44e-1	1-33	
<a href="#">AF-A0A0H3GKN8-F1-model_v4</a>	Phage/conjugal plasmid C-4 type zinc...	<a href="#">Klebsiella pneumoniae subsp. pneum...</a>	0.94	35.8	9.41e-1	1-39	
<a href="#">AF-P41009-F1-model_v4</a>	Uncharacterized protein Y81	<a href="#">Escherichia coli K.12</a>	0.93	37.1	7.51e-1	1-41	

Q 1 FGYCESGVEIGIRRLRLEARPTADLCIDCKTLAEIREKQ MAG  
 FGYCESGVEIGIRRLRLEARPTADLCIDCKTLAEIREKQ MAG  
 T 111 FGYCESGVEIGIRRLRLEARPTADLCIDCKTLAEIREKQ MAG

Select target residues to highlight their structure

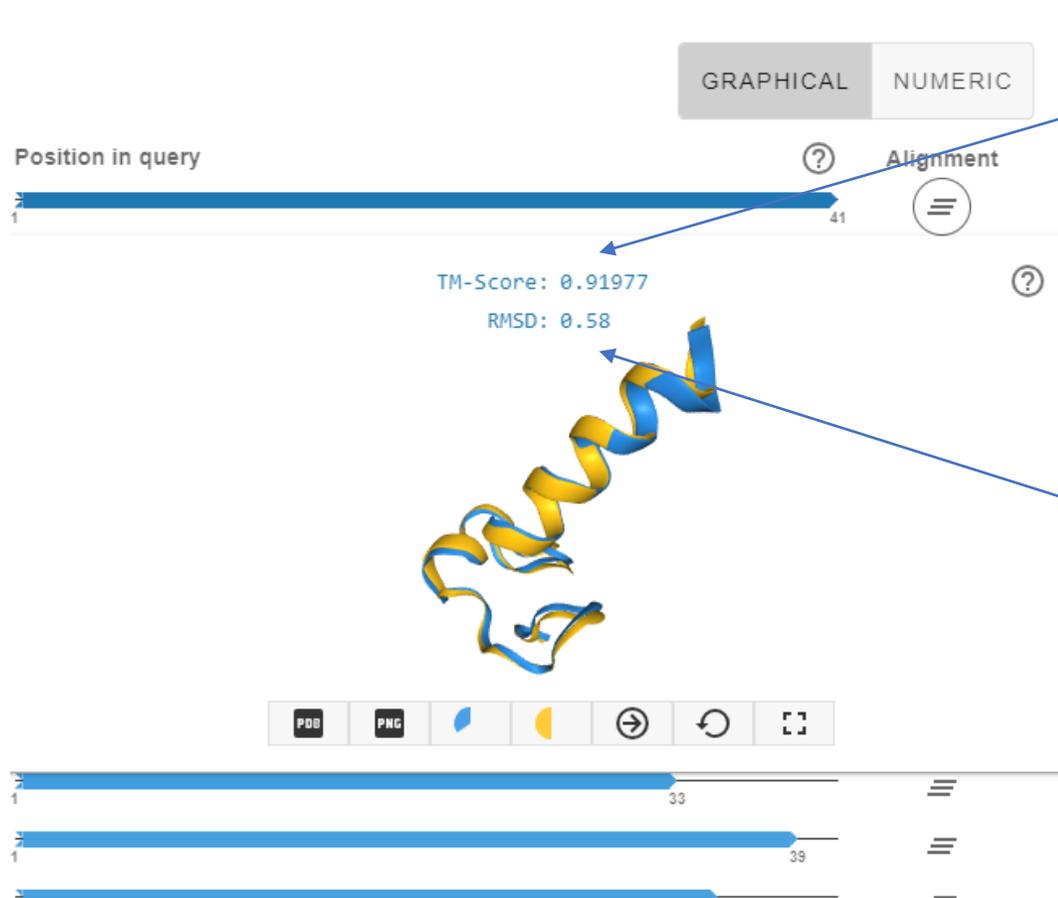
TM-Score: 0.91977  
RMSD: 0.58

Buttons: PDB, PNC, [Color selection], [Rotation], [Refresh], [Zoom]

Alignement de séquence

# Foldseek: chercher avec la structure 3D dans AFDB

Indicateurs de la qualité d'alignement de structures:



TM-score (0 – 1); 1 = les deux structures sont identiques, <0.17 les structures ne se ressemblent pas du tout; > 0.5 les structures ont un repliement global commun

La valeur RMSD donne l'écart moyen de distance entre les atomes correspondants de deux structures protéiques : plus la valeur rmsd est petite, plus les deux structures sont similaires. Les valeurs de RMSD < 2 Angström sont considérées comme très fiables.