

Faculté des Sciences de la Nature et de la Vie  
Département BMC, Université de Jijel

# Cours de Bioinformatique

Licence Biochimie/Sciences pharmacologiques

Dr. BOULHISSA Ilham

**Année universitaire: 2023-2024**

# Chapitre 2 : Les Banques et les Bases de Données Biologiques

# Base de données Vs banques de données

Une **base de données**, usuellement abrégée en **BD** ou **BDD**, Une base de données est un fichier ou un ensemble de fichiers permettant le stockage permanent ou temporaire des informations ainsi que l'accès à ces informations devenues.

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

•Exemple: Organisation des données

Notes de la matière : Bioinformatique (M1-Biochimie Appliquée) S1 (2018-2019)

N°	Matricule	Nom	prénoms	Exam	TD	TP	Moyenne
1	1534048283	ABDELOUAHAB	Nihed	14.5		16	15.1
2	1534049328	ABDENNOUR	Fatima	14		12	13.2
3	1534046258	ABDERREZZAK	Oualid	13		15	13.8
4	1534065132	AGGOUNE	Kaouther	15.5		16	15.7
5	1534044511	AISSANI	Zeyneb	14.75		15	14.85
6	1534046107	AKLOUCHE	Nour El Houda	12		14	12.8
7	1534063927	ALIOUCHE	Amira	13.75		15	14.25
8	1534044540	ALLAG	Sara	12.25		11	11.75

Grâce à cette organisation, on peut retirer facilement des informations (nombre d'étudiants, nombre de garçons, notes de TP supérieures à 15...etc.)

# Base de données Vs banques de données

## Banque de données

Ensemble de données relatif à un domaine ; Ensemble de fichiers manuels ou informatiques sans relation entre eux (*fichier plat*)



Des données exhaustives, donc offre un ensemble plutôt hétérogène d'informations



Bases de données **généralistes**

## Base de données

Ensemble de relations entre les données, gérées à l'aide d'un système de gestion de base de données



Des données plus homogènes établies autour d'une thématique



Bases de données **spécialisées**

# **Les Bases de données biologiques**

Elles sont divisées en 4 catégories :

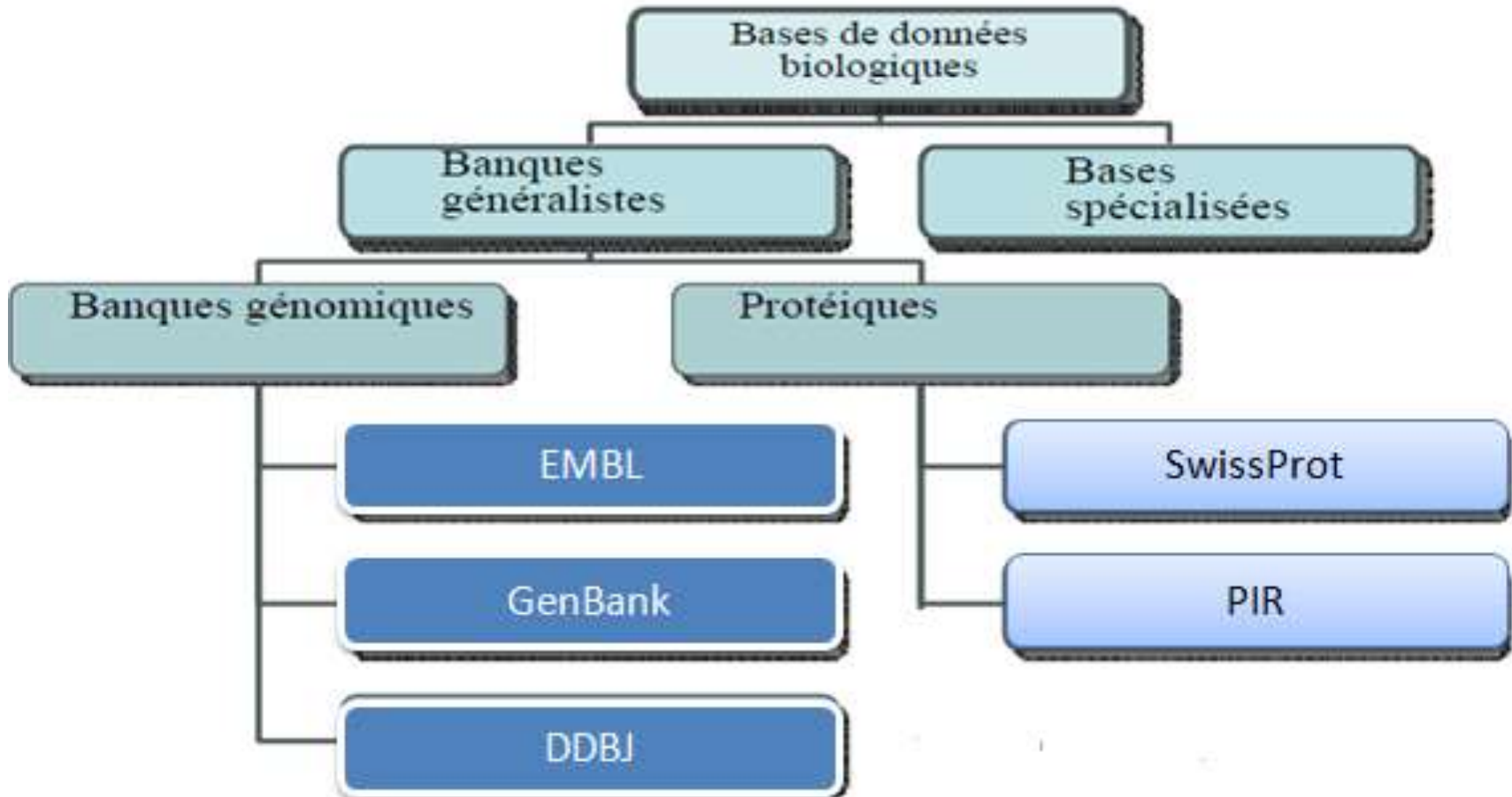
1. **Bases de données généralistes**
2. **Bases de données spécialisées**
3. **Bases de données de structures**

**Exemple: PDB (est une base généraliste aussi).**

4. **Bases de données bibliographiques**

**Exemple: Pubmed.**

# Les banques et les Bases de données biologiques



# 1- Banques de données généralistes

## Banques de données de séquences nucléiques

- ❖ **EMBL** (European Molecular Biology Laboratory ou Laboratoire Européen de Biologie Moléculaire) : banque européenne créée en 1980 et financée par l'EMBO (European Molecular Biology Organization), elle est aujourd'hui diffusée par l'EBI (European Bioinformatics Institute, Cambridge, UK) ;
- ❖ **GenBank** : créée en 1982 par la société IntelliGenetics et diffusée maintenant par le NCBI (National Center for Biotechnology Information, Los Alamos, US) ;
- ❖ **DDBJ (DNA Data Bank of Japan)** : créée en 1986 et diffusée par le NIG (National Institute of Genetics, Japon) ;

# 1- Banques de données généralistes

## Banques de données de séquences nucléiques

- EMBL (Europe) devenue aujourd'hui l'ENA= **E**uropean **N**ucleotide **A**rchive  
→ <http://www.ebi.ac.uk/ena>
- GenBank (USA) → <http://www.ncbi.nlm.nih.gov/genbank/>
- DDBJ (Japon) **D**NA **D**ATA **B**ANQUE OF **J**APAN → <http://www.ddbj.nig.ac.jp>



Ces trois principales bases sont interconnectées dans le cadre d'un consortium appelé **International Nucleotide Sequences databases Collaboration (INSDC)**

# 1- Banques de données généralistes

## Banques de données de séquences protéiques

- ❖ **PIR-NBRF (Protein Information Resource-National Biomedical Research Foundation)** : créée en 1984 par la NBRF (National Biomedical Research Foundation). Elle est maintenant un ensemble de données issues du MIPS (Martinsried Institute for Protein Sequences, Munich, Allemagne) et de la banque japonaise JIPID (Japan International Protein Information Database) ;
- ❖ **SwissProt** : créée en 1986 à l'Université de Genève et maintenue depuis 1987 dans le cadre d'une collaboration, entre cette université (via ExPASy, Expert Protein Analysis System) et l'EBI. Celle-ci regroupe aussi des séquences annotées de la banque PIR-NBRF ainsi que des séquences codantes, traduites de l'EMBL.

Elles contiennent la protéine obtenue de plusieurs manières différentes :

- *in silico* : déduite à partir de la séquence nucléique, par simple traduction du ou des exons la codant
- isolée à partir de la cellule
- ou encore par génie génétique



Ces deux bases de données sont regroupées dans **un consortium**  
***Universal Protein Resource «UniProt »***

## 2- Bases de données spécialisées

Dédiées à un organisme :



- Flybase : Drosophile <http://flybase.org>
- HIV database : [www.hiv.lanl.gov/](http://www.hiv.lanl.gov/)
- Porteco: Escherichia coli <http://www.porteco.org>
- *Arabidopsis thaliana*: TAIR <https://www.arabidopsis.org>

## 2- Bases de données spécialisées

Dédiées à un type de séquences particulier :



**OMIM**® Online Mendelian Inheritance in Man®  
An Online Catalog of Human Genes and Genetic Disorders  
Updated 9 October 2015

- IMGT : données d'immunologie <http://www.imgt.org>
- EPD : Eukaryotic Promoter Database <http://epd.vital-it.ch>
- The European ribosomal RNA database  
<http://bioinformatics.psb.ugent.be/webtools/rRNA>
- Online Mendelian Inheritance in Man <http://www.omim.org>
- GOLD: Genome Online Database <https://gold.jgi.doe.gov>

## **2- Bases de données spécialisées**

**Ensembl** : Base intégrative génomique (annotation du génome)

**Prosite**: Recense les motifs protéiques ayant une signification biologique.

**Reactome**: Base intégrative métabolique.

**Pathway database**: Visualisation, interprétation, et analyse des connaissances des voies.

**Kegg pathway**: Interactions moléculaires et réaction.

**PFAM**: base de domaines protéiques

# Avantage et inconvénient des bases généralistes et spécialisées

## Généralistes

**Avantage:** Tout est consultable en une fois.

**Inconvénient:** Difficile à maintenir, difficile à interroger.

## Spécialisées

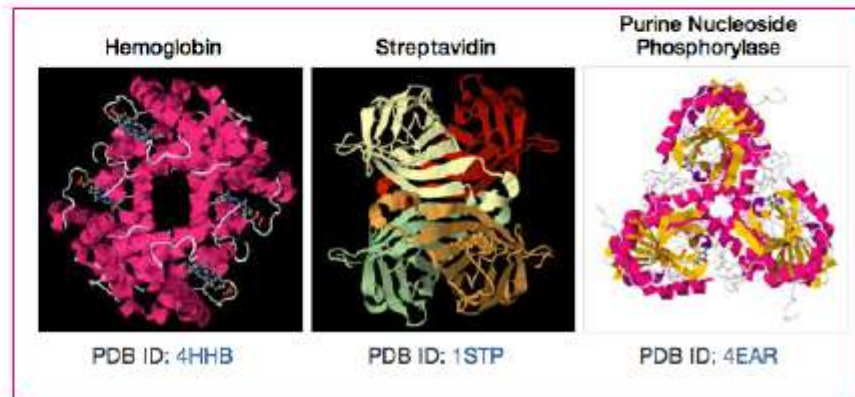
**Avantage:** Facile à mettre à jour les données, vérifier leur intégrité, offrir une interface adaptée.

**Inconvénient:** Ne cible pas toujours ce que l'on veut, toutes les banques possibles n'existent pas.

### 3- Bases de données de structure

- La Protein Database (PDB) stockent les structures protéiques obtenues par RMN ou cristallographie
- Une entrée contient donc les coordonnées de tous les atomes de la structure

<http://mm.rcsb.org>



## 4- Bases de données bibliographiques



Relative aux sciences biologiques et médicales

Appartient à la bibliothèque nationale de médecine des États-Unis d'Amérique (NLM).

Accessible en ligne via l'interface *PubMed* (<https://www.ncbi.nlm.nih.gov/pubmed/>)

**Donne accès à:** plus de 26 millions d'articles scientifiques provenant d'environ 6000 revues

Mot Clé

https://www.ncbi.nlm.nih.gov/pubmed/?term=acetylcholinesterase

NCBI Resources How To Sign in to NCBI

PubMed acetylcholinesterase Search

US National Library of Medicine National Institutes of Health

Article types: Clinical Trial, Review, Customize ...

Text availability: Abstract, Free full text, Full text

Publication dates: 5 years, 10 years, Custom range...

Species: Humans, Other Animals

Format: Summary Sort by: Best Match Per page: 20

Send to Filters: Manage Filters

Sort by: Best match Most recent

Search results

Items: 1 to 20 of 31887

<< First < Prev Page 1 of 1595 Next > Last >>

[Acetylcholinesterase: A Primary Target for Drugs and Insecticides.](#)

1. Thapa S, Lv M, Xu H. Mini Rev Med Chem. 2017;17(17):1665-1676. doi: 10.2174/1389557517666170120153930. Review. PMID: 28117022 [Similar articles](#)

[Immunopurification of Acetylcholinesterase from Red Blood Cells for Detection of Nerve Agent Exposure.](#)

2. Daffemer AJ, Schopfer LM, Xiao G, Cashman JR, Yerramalla U, Johnson RC, Blake TA, Lockridge O. Chem Res Toxicol. 2017 Oct 16;30(10):1897-1910. doi: 10.1021/acs.chemrestox.7b00209. Epub 2017 Sep 25. PMID: 28892361 [Free PMC Article](#) [Similar articles](#)

[Recombinant Acetylcholinesterase purification and its interaction with silver nanoparticle.](#)

3. Mirzajani F, Motevalli SM, Jabbari S, Ranaei Siadat SO, Sefidbakht Y. Protein Expr Purif. 2017 Aug;136:58-65. doi: 10.1016/j.pep.2017.05.007. Epub 2017 May 26. PMID: 28554568 [Similar articles](#)

[Acetylcholinesterase Inhibitory Alkaloids from the Whole Plants of Zephyranthes carinata.](#)

Results by year

Affichage des résultats

Download CSV

Related searches

acetylcholinesterase inhibitor

Titles with your search terms

Caregivers' perspectives and experiences of withdrawing acetylcholi [BMC Palliat Care. 2019]

Observation of Acetylcholinesterase in Stress-induced Depression Phei [J Am Chem Soc. 2019]

Large-scale separation of acetylcholinesterase

Format: Abstract ▾

[Mini Rev Med Chem](#). 2017;17(17):1665-1676. doi: 10.2174/1389557517666170120153930.

## Acetylcholinesterase: A Primary Target for Drugs and Insecticides.

[Thapa S](#)<sup>1</sup>, [Lv M](#)<sup>1</sup>, [Xu H](#)<sup>1</sup>.

### ⊕ Author information

#### Abstract

**BACKGROUND:** Acetylcholinesterase is a serine hydrolase that terminates the action of the neurotransmitter acetylcholine by hydrolyzing it into acetic acid and choline.

**OBJECTIVE:** The enzyme, containing an ellipsoidal structure, possesses three binding sites such as active site (with catalytic anionic and esteratic subsites), aromatic gorge and peripheral anionic site, where the inhibiting compounds interact. The acetylcholinesterase inhibitors bind to the enzyme and interfere with the breakdown of acetylcholine, leading to the deposition of acetylcholine in the nerve synapses and causing disrupted neurotransmission. Based on this principle of action, many therapeutic drugs for the treatment of different diseases, pesticides and chemical warfare agents have been synthesized targeting the acetylcholinesterase.

**CONCLUSION:** The present review summarizes the current knowledge about acetylcholinesterase, its structure, function and biosynthesis, its inhibitors, and mode of action of inhibitors on it. Besides, the review also presents an overview about the resistance mechanism that the organisms develop due to the over-application of acetylcholinesterase inhibitors.

Copyright© Bentham Science Publishers; For any queries, please email at [epub@benthamscience.org](mailto:epub@benthamscience.org).

**KEYWORDS:** Acetylcholine; acetylcholinesterase; acetylcholinesterase inhibitors; active sites; insecticide resistance; neurotransmission

PMID: 28117022 DOI: [10.2174/1389557517666170120153930](https://doi.org/10.2174/1389557517666170120153930)

[Indexed for MEDLINE]

# Recherche avancée

(((Acetylcholinesterase) AND butyrylcholinesterase) OR alzheimer) NOT cancer

[Edit](#)

[Clear](#)

## Builder

	All Fields	Acetylcholinesterase	-	<a href="#">Show index list</a>
AND	All Fields	butyrylcholinesterase	-	<a href="#">Show index list</a>
OR	All Fields	alzheimer	-	<a href="#">Show index list</a>
NOT	All Fields	cancer	-	<a href="#">Show index list</a>
AND	All Fields		- +	<a href="#">Show index list</a>

[Search](#) or [Add to history](#)

## Article types

Clinical Trial  
Review  
Customize ...

## Text availability

Abstract  
Free full text  
Full text

## Publication dates

5 years  
10 years  
Custom range...

## Species

Humans  
Other Animals

[Clear all](#)

[Show additional filters](#)

Format: Summary ▾ Sort by: Best Match ▾ Per page: 20 ▾

Send to ▾ Filters: [Manage Filters](#)

## Search results

Items: 1 to 20 of 110895

<< First < Prev Page 1 of 5545 Next > Last >>

Sort by:

Best match

[Most recent](#)

Results by year

Titles with your search terms

Oxidative stress, dysfunctional glucose metabolism and **Alzhei** [Nat Rev Neurosci. 2019]

Inhibition of **acetylcholinesterase** and **butyrylcholin** [J Enzyme Inhib Med Chem. 2019]

Dementia in Down syndrome: unique insights for **Alzheimer** disease reseal [Nat Rev Neurol. 2019]

[See more...](#)

Find related data

Database:

Find items

- [Rivastigmine: the advantages of dual inhibition of \*\*acetylcholinesterase\*\* and \*\*butyrylcholinesterase\*\* and its role in subcortical vascular dementia and Parkinson's disease dementia.](#)  
Kandiah N, Pai MC, Senanarong V, Looi I, Ampil E, Park KW, Karanam AK, Christopher S.  
Clin Interv Aging. 2017 Apr 18;12:697-707. doi: 10.2147/CIA.S129145. eCollection 2017. Review.  
PMID: 28458525 [Free PMC Article](#)  
[Similar articles](#)
- [Status of \*\*acetylcholinesterase\*\* and \*\*butyrylcholinesterase\*\* in \*\*Alzheimer's disease\*\* and type 2 diabetes mellitus.](#)  
Mushtaq G, Greig NH, Khan JA, Kamal MA.  
CNS Neurol Disord Drug Targets. 2014;13(8):1432-9. Review.  
PMID: 25345511 [Free PMC Article](#)  
[Similar articles](#)
- [In silico repurposing of antipsychotic drugs for \*\*Alzheimer's disease\*\*.](#)  
Kumar S, Chowdhury S, Kumar S.  
BMC Neurosci. 2017 Oct 27;18(1):76. doi: 10.1186/s12868-017-0394-8.  
PMID: 29078760 [Free PMC Article](#)  
[Similar articles](#)



Banque de donnée des séquences génétiques (ADN, ARN et protéines)

Banque Américaine appartenant au NCBI (centre national américain pour l'information en biotechnologie)

Accessibles gratuitement en ligne via son site web:

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

**Contenu:** 215 millions de séquences (390 milliards de bases) provenant de 100000 organismes.

Cette banque se construit en recevant des séquences provenant de différents laboratoires à travers le monde.

**Formats disponibles:** Text (Genbank) et Fasta

# Format GenBank

Fichier stockant une séquence nucléique avec une organisation bien précise

Compte courant de nucléotides

ORIGIN

```
1 ctctccctc atctttgcca acctgcccc cctcctctgc agctgagcga taacccttgg
61 gccgacagt ccctaattc ctccctcctg gcttctcgac cgacccttca ccttttccct
121 ttctttctcc cagcagacgc cgctgacct gcagccatga ggcccccgca gtgtctgctg
181 cacacgcctt ccctggcttc cccactcctt ctctcctcc tctggctcct ggggtggagga
241 gtgggggctg agggccggga ggatgcagag ctgctgggta cggtgctggt gggccggctg
301 cggggcattc gcctgaagac ccccgggggc cctgtctctg ctttctctgg catccccctt
361 gcggagccac ccatgggacc ccgtcgcttt ctgccaccgg agcccaagca gccttgggtca
421 ggggtggtag acgctacaac cttccagagt gtctgtctacc aatatgtgga caccctatac
481 ccaggttttg agggcaccga gatgtggaac cccaaccgtg agctgagcga ggactgcctg
541 tacctcaacg tgtggacacc atacccccgg cctacatccc ccaccctgt cctcgtctgg
601 atctatgggg gtggcttcta cagtggggcc tcctccttgg acgtgtacga tggccgcttc
661 ttgttacagg ccgagaggac tgtgctggtg tccatgaact accgggtggg agcctttggc
721 ttcttgccc tgccggggag ccgagaggcc ccgggcaatg tgggtctcct ggatcagagg
781 ctggccctgc agtgggtgca ggagaactg gcagccttc ggggtgacc gacatcagt
841 acgctgtttg gggagagcgc gggagccgcc tcggtgggca tgcacctgct gtccccgcc
901 agccggggcc tgttcacag ggcctgctg cagagcggtg ccccaatgg acctggggc
961 acggtgggca tgggagaggc ccgtcgagg gccacgcagc tggccacct tgtgggctgt
1021 cctccaggcg gcactggtgg gaatgacaca gagctggtag cctgccttcg gacacgacca
1081 gcgcaggtcc tggatgaacca cgaatggcac gtgctgcctc aagaaagcgt cttccggttc
1141 tccttcgtgc ctgtggtaga tggagacttc ctcagtgaca cccagaggc cctcatcaac
1201 gcgggagact tccacggcct gcaggtgctg gtgggtggtg tgaaggatga gggctcgtat
1261 tttctggttt acggggcccc aggcttcagc aaagacaacg agtctctcat cagccgggcc
1321 gaggctcctg ccgggggtgc ggtcgggggt ccccaggtaa gtgacctggc agccgaggct
1381 gtggtcctgc attacacaga ctggctgcat cccgaggacc cggcacgcct gagggaggcc
1441 ctgagcgatg tgggtgggca ccacaatgtc gtgtgccccg tggcccagct ggctgggcca
1501 ctggtgccc aggggtgccc ggtctacgcc tacgtctttg aacaccgtgc ttccacgctc
1561 tcctggcccc tgtggatggg ggtgccccac ggctacgaga tcgagttcat ctttgggatc
1621 cccctggacc cctctcgaaa ctacacggca gaggagaaaa tcttcgcca gcgactgatg
1681 cgatactggg ccaacttgc ccgcacaggg gatccaatg agccccgaga cccaaggcc
1741 ccacaatggc ccccgtacac ggcgggggct cagcagtacg ttagtctgga cctgcggccg
1801 ctggaggtgc ggcgggggct gcgcgccag gcctgcgct tctggaaccg cttcctccc
1861 aaattgctca gcgccaccga cacgctcgac gaggcggagc gccagtggaa ggccgagttc
1921 caccgctgga gtcctacat ggtgcaactg aagaaccagt tcgaccacta cagcaagcag
1981 gatcgtgctc cagacctgtg acccggcgg gacccccatg tcttcgctc cgcccggccc
2041 cctagctgta tatactattt atttcagggc tgggctataa cacagacgag cccagactc
2101 tgcccatccc caccacccc cgagctccc cggggctccc ggtcctctgg catgtcttca
2161 ggctgagctc ctccccgct gccttcgcc tctggctgca aataaactgt tacaggcc
```

Paquet de 10 nucléotides

séquence nucléique (format GenBank)

60 nucléotides par ligne (6 paquets)

Descripteur de fin de fichier

//

# Human acetylcholinesterase (ACHE) mRNA, complete cds

GenBank: M55040.1

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

Go to:

**Date de soumission**

**Définition et numéro d'accèsion**

LOCUS HUMACHE 2218 bp mRNA **linear** PRI 09-JUN-1995  
DEFINITION Human acetylcholinesterase (ACHE) mRNA, complete cds.  
ACCESSION M55040  
VERSION M55040.1  
KEYWORDS acetylcholinesterase.

**Type de molécule: linéaire ou circulaire**

**Organisme et taxonomie**

SOURCE Homo sapiens (human)  
ORGANISM [Homo sapiens](#)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.

**Références bibliographiques**

REFERENCE 1 (bases 1 to 2218)  
AUTHORS Soreq,H.E., Ben-Aziz,R., Prody,C.A., Seidman,S., Gnatt,A., Neville,L., Lieman-Hurwitz,J., Lev-Lehman,E., Ginzberg,D., Lapidot-Lifson,Y. and Zakut,H.  
TITLE Molecular cloning and construction of the coding region for human acetylcholinesterase reveals a G + C-rich attenuating structure  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 87 (24), 9688-9692 (1990)  
PUBMED [2263619](#)  
COMMENT Original source text: Human 21-week old fetus DNA, and cDNA to mRNA.

**Caractéristiques De la séquence: gène, région...etc.**

FEATURES Location/Qualifiers  
source 1..2218  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:[9606](#)"  
/map="7q22"  
/dev\_stage="21-week-old fetus"  
/tissue\_lib="lambda-gt10"  
gene 1..2218  
/gene="ACHE"

Ce format contient également une foule de métadonnées relative à la séquence en question

# Suite

Caractéristiques  
De la séquence:  
gène, région...etc.

gene

mRNA

CDS

```
/mol_type="mRNA"  
/db_xref="taxon:9606"  
/map="7q22"  
/dev_stage="21-week-old fetus"  
/tissue_lib="lambda-gt10"  
1..2218  
/gene="ACHE"  
1..2218  
/gene="ACHE"  
/product="acetylcholinesterase"  
/note="G00-118-746"  
157..2001  
/gene="ACHE"  
/EC_number="3.1.1.7"  
/codon_start=1  
/product="acetylcholinesterase"  
/protein_id="AAA68151.1"  
/db_xref="GDB:G00-118-746"
```

Propriétés et  
séquence  
peptidique

sig\_peptide

mat\_peptide

```
/translation="MRPPQCLLHTPSLASPLLLLLLWLLGGGVGAEGREDAELLVTVR  
GGRLRGIRLKTGGPVSAFLGIPFAEPPMGPRRFLPPEPKQPWSGVVDATTFQSVCYQ  
YVDTLYPGFEGTEMWNPRELSEDCLYLNVWTPYPRPTSPTPVLVWIYGGGFYSGASS  
LDVYDGRFLVQAERTVLVSMNYRVGAFGLALPGSREAPGNVGLLDQRLALQWVQENV  
AAFGGDPTSVTLFGESAGAASVGMHLLSPPSRGLFHRAVLQSGAPNGPWATVGMGEAR  
RRATQLAHLVGCPPGGTGGNDTELVACLRTRPAQVLVNHWHVLPQESVFRFSFVPVW  
DGDFLSDTPEALINAGDFHGLQVLVGVVKDEGSYFLVYGAPGFSKDNESSLISRAEFLA  
GVRVGVQVSDLAAEAVVLHYTDNLHPEDPARLREALSDVVGDHNVVCPVAQLAGRLA  
AQGARVYAYVFEHRASLWPLWVGVPHYEIEFIFGIPLDPSRNYTAEKIFAQRLM  
RYWANFARTGDPNEPRDPKAPQWPPYTAGAQQYVSLDLRPLEVRRGLRAQACAFWNR  
LPKLLSATDLDLDEAERQKAEFHRWSSYVHMKNQFDHYSKQDRCSDL"  
229..291  
/gene="ACHE"  
/note="G00-118-746; does not fit consensus"  
310..1998  
/gene="ACHE"  
/product="acetylcholinesterase"  
/EC_number="3.1.1.7"  
/note="G00-118-746"
```



**Créée en:**1980 en Allemagne

**Financée par :**l'EMBO (European Molecular Biology Organisation)

**Hiérarchisée par :** 80 groupes de recherche provenant de 24 pays européens

**Contenu:** 248 millions de séquences nucléiques

**Accessible gratuitement en ligne :** <https://www.ebi.ac.uk>

**Formats disponibles:** Text (Genbank), Fasta et XML

## Format Fasta

Une seule ligne d'en-tete

Le premier caractère est obligatoirement >

La séquence commence en deuxième ligne et se continue sur autant de lignes que nécessaire

Pas d'espace entre les nucléotides

>ENA|M55040|M55040.1 Human acetylcholinesterase (ACHE) mRNA, complete cds.

CTCTCCCCTCATCTTTGCCAACCTGCCCCACCTCCTCTGCAGCTGAGCGATAACCCTTGG  
GCCGACAGTGCCCTAATCTCCTCCCTCCTGGCTTCTCGACCGACCCTTCACCCTTTCCT  
TTCTTTCTCCCAGCAGACGCCGCCTGCCCTGCAGCCATGAGGCCCCCGCAGTGTCTGCTG  
CACACGCCTTCCCTGGCTTCCCCACTCCTTCTCCTCCTCCTCTGGCTCCTGGGTGGAGGA  
GTGGGGGCTGAGGGCCGGGAGGATGCAGAGCTGCTGGTGACGGTGCCTGGGGGCCGGCTG  
CGGGGCATTCGCCTGAAGACCCCGGGGGCCCTGTCTCTGCTTTCCTGGGCATCCCCTTT  
GCGGAGCCACCCATGGGACCCCGTTCGCTTTCTGCCACCGGAGCCCAAGCAGCCTTGGTCA  
GGGGTGGTAGACGCTACAACCTTCCAGAGTGTCTGCTACCAATATGTGGACACCCTATAC  
CCAGGTTTTGAGGGCACCGAGATGTGGAACCCCAACCCTGAGCTGAGCGAGGACTGCCTG  
TACCTCAACGTGTGGACACCATAACCCCGGCCTACATCCCCACCCCTGTCCTCGTCTGG  
ATCTATGGGGGTGGCTTCTACAGTGGGGCCTCCTCCTTGGACGTGTACGATGGCCGCTTC  
TTGGTACAGGCCGAGAGGACTGTGCTGGTGTCCATGAACTACCGGGTGGGAGCCTTTGGC  
TTCCTGGCCCTGCCGGGGAGCCGAGAGGCCCGGGCAATGTGGGTCTCCTGGATCAGAGG  
CTGGCCCTGCAGTGGGTGCAGGAGAACGTGGCAGCCTTCGGGGGTGACCCGACATCAGTG  
ACGCTGTTTGGGGAGAGCGCGGGAGCCGCCCTCGGTGGGCATGCACCTGCTGTCCCCGCC  
AGCCGGGGCCTGTTCCACAGGGCCGTGCTGCAGAGCGGTGCCCCCAATGGACCCTGGGCC  
ACGGTGGGCATGGGAGAGGCCCGTTCGACAGGGCCACGCAGCTGGCCACCTTGTGGGCTGT  
CCTCCAGGCGGCACCTGGTGGGAATGACACAGAGCTGGTAGCCTGCCTTCGGACACGACCA  
GCGCAGGTCCTGGTGAACCACGAATGGCACGTGCTGCCTCAAGAAAGCGTCTTCCGGTTC  
TCCTTCGTGCCTGTGGTAGATGGAGACTTCCTCAGTGACACCCAGAGGCCCTCATCAAC  
GCGGGAGACTTCCACGGCCTGCAGGTGCTGGTGGGTGTGGTGAAGGATGAGGGCTCGTAT  
TTTCTGGTTTACGGGGCCCCAGGCTTCAGCAAAGACAACGAGTCTCTCATCAGCCGGGCC  
GAGTTCCTGGCCGGGGTGCGGGTCCGGGTTCCCCAGGTAAGTGACCTGGCAGCCGAGGCT  
GTGGTCTTGCATTACACAGACTGGCTGCATCCCGAGGACCCGGCACGCCTGAGGGAGGCC  
CTGAGCGATGTGGTGGGCAGCCACAATGTCGTGTGCCCCGTGGCCAGCTGGCTGGGCGA  
CTGGCTGCCCAGGGTGCCCGGGTCTACGCCCTACGTCTTTGAACACCGTGCTTCCACGCTC  
TCCTGGCCCCTGTGGATGGGGGTGCCCCACGGCTACGAGATCGAGTTCATCTTTGGGATC  
CCCCTGGACCCCTCTCGAACTACACGGCAGAGGAGAAAATCTTCGCCCAGCGACTGATG  
CGATACTGGGCCAACTTTGCCCGCACAGGGGATCCCAATGAGCCCCGAGACCCCAAGGCC  
CCACAATGGCCCCCGTACACGGCGGGGGCTCAGCAGTACGTTAGTCTGGACCTGCGGGCCG  
CTGGAGGTGCGGCGGGGGCTGCGCGCCCAGGCCTGCGCCTTCTGGAACCGCTTTCCTCCCC  
AAATTGCTCAGCGCCACCAGACAGCTCGACGAGGCGGAGCGCCAGTGGAAAGGCCGAGTTC  
CACCGCTGGAGCTCCTACATGGTGCACCTGGAAGAACCAGTTCGACCACTACAGCAAGCAG  
GATCGCTGCTCAGACCTGTGACCCCGGCGGGACCCCATGTCCTCCGCTCCGCCCGGCC  
CCTAGCTGTATACTATTTATTTACAGGGCTGGGCTATAACACAGACGAGCCCCAGACTC  
TGCCCATCCCCACCCACCCCGACGTCCCCCGGGGCTCCCGGTCTCTGGCATGTCTTCA  
GGCTGAGCTCCTCCCCGCGTGCCTTCGCCCTCTGGCTGCAAATAAACTGTTACAGGCC

# Dna Data Base of Japan

**Créée en:**1984 au Japon

**Contenu:** 240 millions de séquences nucléiques, 6.56 milliards de bases

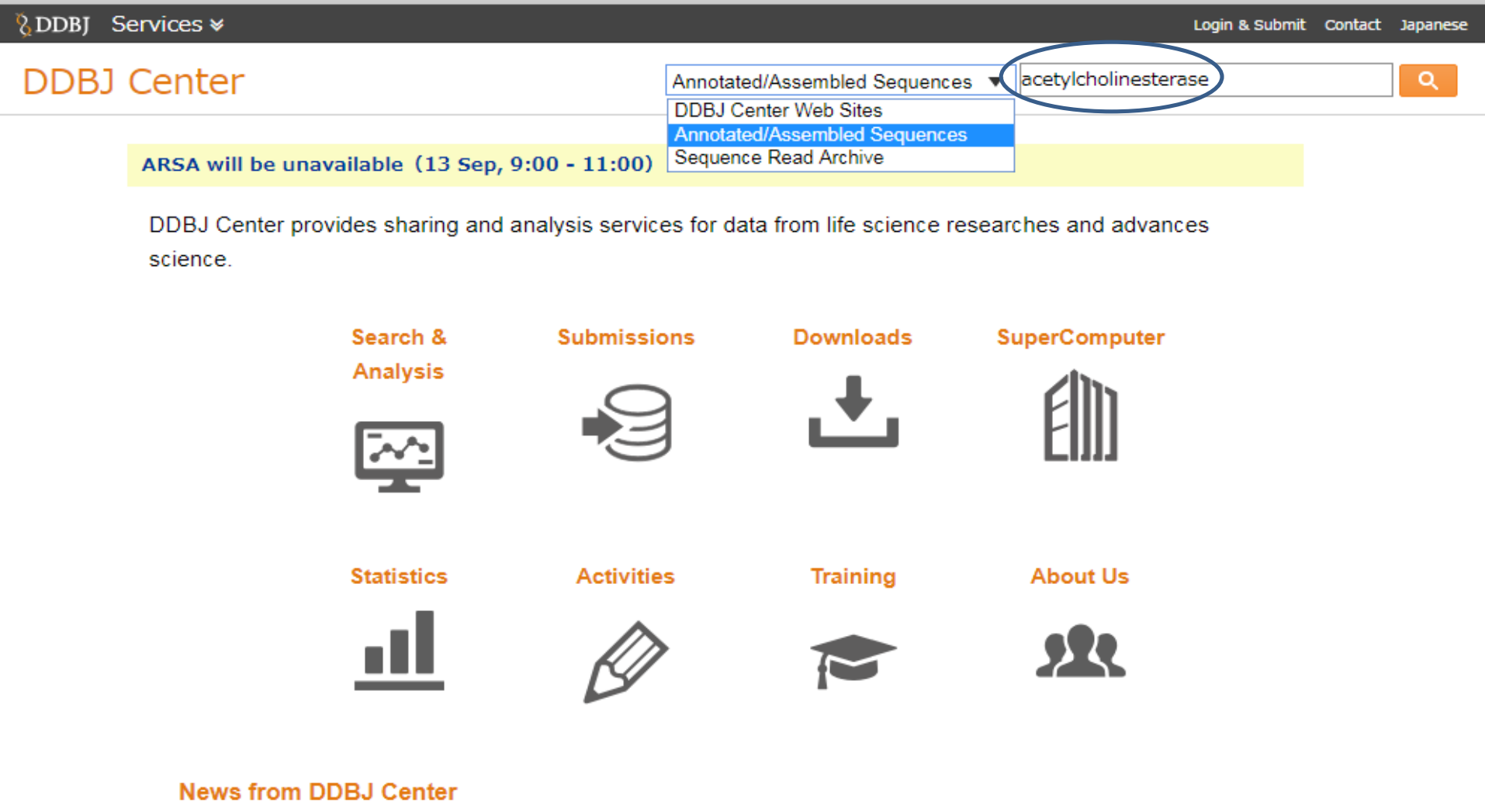
**Accessible gratuitement via son site web:** <https://www.ddbj.nig.ac.jp>

**Formats disponibles:** Fasta, XML

# Recherche des séquences nucléiques via DDBJ

1. Sur une page web, tapez l'adresse de DDBJ: <https://www.ddbj.nig.ac.jp/>
2. Sélectionner « Annotated/Assembled Sequences » et écrire « Acetylcholinesterase »

risé | <https://www.ddbj.nig.ac.jp/index-e.html>



The screenshot shows the DDBJ Center website interface. At the top, there is a navigation bar with the DDBJ logo, 'Services', and links for 'Login & Submit', 'Contact', and 'Japanese'. Below this, the 'DDBJ Center' logo is on the left, and a search bar is on the right. The search bar contains the text 'acetylcholinesterase' and has a magnifying glass icon. A dropdown menu is open below the search bar, showing options: 'Annotated/Assembled Sequences' (selected), 'DDBJ Center Web Sites', 'Annotated/Assembled Sequences', and 'Sequence Read Archive'. A yellow banner below the search bar reads 'ARSA will be unavailable (13 Sep, 9:00 - 11:00)'. Below the banner, a paragraph states: 'DDBJ Center provides sharing and analysis services for data from life science researches and advances science.' The main content area features eight service tiles arranged in a 2x4 grid, each with an icon and a title: 'Search & Analysis' (monitor with graph), 'Submissions' (database cylinder), 'Downloads' (download arrow), 'SuperComputer' (server rack), 'Statistics' (bar chart), 'Activities' (pencil), 'Training' (graduation cap), and 'About Us' (three people silhouettes). At the bottom, there is a section titled 'News from DDBJ Center'.

### 3. Raffiner la recherche en cliquant sur « condition search ». Choisir « AND » et ajouter « homo sapiens » pour rechercher l'acétylcholinesterase humaine.

#### Search Condition

##### Quick Search

acetylcholinesterase homo sapiens

Search AND  
OR  
AND

##### Available Fields

#### Search Result

##### Facet

##### List of Entries

1 - 30 entries / Number of finds: 271  FlatFile  XML  Fasta

PrimaryAccessionNumber	Definition	SequenceLength	MolecularType	Organism
<input type="checkbox"/> <a href="#">S71129</a>	Definition:acetylcholinesterase {I4-E5 doman} [human, tumor cell lines, Genomic, 847 nt].	SequenceLength:847	MolecularType:DNA	Organism:Homo sapiens
<input type="checkbox"/> <a href="#">AY225516</a>	Definition:Homo sapiens acetylcholinesterase membrane anchor precursor PRiMA variant I mRNA, complete cds. Organism:Homo sapiens	SequenceLength:910	MolecularType:mRNA	
<input type="checkbox"/> <a href="#">AY225517</a>	Definition:Homo sapiens acetylcholinesterase membrane anchor precursor PRiMA variant II mRNA, complete cds. Organism:Homo sapiens	SequenceLength:1081	MolecularType:mRNA	
<input type="checkbox"/> <a href="#">AH005291</a>	Definition:Homo sapiens Human acetylcholinesterase (ACHE) gene.	SequenceLength:3419	MolecularType:DNA	Organism:Homo sapiens
<input type="checkbox"/> <a href="#">AY389977</a>	Definition:Homo sapiens N-terminal extended acetylcholinesterase (ACHE) mRNA, exon E1d and partial cds.	SequenceLength:470	MolecularType:mRNA	

#### 4. Sélectionner le code « [M55040](#) » correspondant à la séquence de mRNA (séquences codantes) de l'Acétylcholinestérase humaine.

##### List of Entries

1 - 60 entries / Number of founds: 271  FlatFile  XML  Fasta

PrimaryAccessionNumber	Definition	SequenceLength	MolecularType	Organism
<input type="checkbox"/> <a href="#">S71129</a>	Definition:acetylcholinesterase {I4-E5 domain} [human, tumor cell lines, Genomic, 847 nt].	SequenceLength:847	MolecularType:DNA	Organism:Homo sapiens
<input type="checkbox"/> <a href="#">AY225516</a>	Definition:Homo sapiens acetylcholinesterase membrane anchor precursor PRiMA variant I mRNA, complete cds.	SequenceLength:910	MolecularType:mRNA	Organism:Homo sapiens
<input type="checkbox"/> <a href="#">AY225517</a>	Definition:Homo sapiens acetylcholinesterase membrane anchor precursor PRiMA variant II mRNA, complete cds.	SequenceLength:1081	MolecularType:mRNA	Organism:Homo sapiens
<input type="checkbox"/> <a href="#">AH005291</a>	Definition:Homo sapiens Human acetylcholinesterase (ACHE) gene.	SequenceLength:3419	MolecularType:DNA	Organism:Homo sapiens
<input type="checkbox"/> <a href="#">AY389977</a>	Definition:Homo sapiens N-terminal extended acetylcholinesterase (ACHE) mRNA, exon E1d and partial cds.	SequenceLength:470	MolecularType:mRNA	Organism:Homo sapiens
<input type="checkbox"/> <a href="#">DQ140346</a>	Definition:Synthetic construct acetylcholinesterase erythrocytic isoform (oAChE-E) mRNA, complete cds.	SequenceLength:1857	MolecularType:mRNA	Organism:synthetic construct
<input type="checkbox"/> <a href="#">DQ140347</a>	Definition:Synthetic construct acetylcholinesterase readthrough isoform (oAChE-R) RNA, complete cds.	SequenceLength:1806	MolecularType:RNA	Organism:synthetic construct
<input checked="" type="checkbox"/> <a href="#">M55040</a>	Definition:Human acetylcholinesterase (ACHE) mRNA, complete cds.	SequenceLength:2218	MolecularType:mRNA	Organism:Homo sapiens
<input type="checkbox"/> <a href="#">M76539</a>	Definition:Human acetylcholinesterase (ACHE) gene, exons 2 and 3H, partial cds.	SequenceLength:535	MolecularType:DNA	Organism:Homo sapiens
<input type="checkbox"/> <a href="#">DQ140345</a>	Definition:Synthetic construct acetylcholinesterase synaptic isoform (oAChE-S) mRNA, complete cds.	SequenceLength:1848	MolecularType:mRNA	Organism:synthetic construct
<input type="checkbox"/> <a href="#">AJ225895</a>	Definition:Homo sapiens mRNA for acetylcholinesterase-associated collagen COLQ.	SequenceLength:1867	MolecularType:mRNA	Organism:Homo sapiens
<input type="checkbox"/> <a href="#">GM880660</a>	Definition:Sequence 5 from Patent WO2008135790.	SequenceLength:20	MolecularType:DNA	Organism:Homo sapiens
<input type="checkbox"/> <a href="#">KY684069</a>	Definition:Homo sapiens monoclonal_1G anti-human acetylcholinesterase immunoglobulin heavy chain variable region (IGK) mRNA, partial cds.	SequenceLength:1338	MolecularType:mRNA	Organism:Homo sapiens
<input type="checkbox"/> <a href="#">KY684074</a>	Definition:Homo sapiens monoclonal_AE-1 anti-human acetylcholinesterase immunoglobulin light chain variable region (IGK) mRNA, partial cds.	SequenceLength:321	MolecularType:mRNA	Organism:Homo sapiens
<input type="checkbox"/> <a href="#">AF230924</a>	Definition:Homo sapiens brain acetylcholinesterase putative membrane anchor mRNA, complete cds.	SequenceLength:624	MolecularType:mRNA	Organism:Homo sapiens
<input type="checkbox"/> <a href="#">GM880661</a>	Definition:Sequence 6 from Patent WO2008135790.	SequenceLength:19	MolecularType:DNA	Organism:Homo sapiens
<input type="checkbox"/> <a href="#">KY684068</a>	Definition:Homo sapiens monoclonal_10D anti-human acetylcholinesterase immunoglobulin light chain variable region (IGK) mRNA, partial cds.	SequenceLength:645	MolecularType:mRNA	Organism:Homo sapiens
<input type="checkbox"/> <a href="#">KY684073</a>	Definition:Homo sapiens monoclonal_AE-1 anti-human acetylcholinesterase immunoglobulin heavy chain variable region (IGK) mRNA, partial cds.	SequenceLength:360	MolecularType:mRNA	Organism:Homo sapiens
<input type="checkbox"/> <a href="#">KU048096</a>	Definition:Homo sapiens acetylcholinesterase (ACHE) gene, ACHE-YT*A 01 variant allele, partial cds.	SequenceLength:470	MolecularType:DNA	Organism:Homo sapiens

**Créée en:**1984 aux états unis d'amérique

**Contenu:** Séquences protéiques et outils d'analyses

**Accessible gratuitement via le site web:** <https://pir.georgetown.edu/>

**Formats disponibles:** Format PIR

Chaque ligne comporte 30 acides aminés

Compte courant des acides aminés au début de chaque ligne

Pas d'espaces entre les acides aminés

En-tête contenant des métadonnées relatives à la protéine

ENTRY           A39256 #type complete                    [iProClass](#) View  
 TITLE           acetylcholinesterase (EC 3.1.1.7) precursor, brain splice  
                   form - human  
 ORGANISM        #formal\_name Homo sapiens #common\_name man  
 DATE            18-Oct-1991 #sequence\_revision 18-Oct-1991 #text\_change  
                   09-Jul-2004  
 ACCESSIONS     A39256; S03959  
 REFERENCE       A39256  
           #authors   Soreq, H.; Ben-Aziz, R.; Prody, C.A.; Seidman, S.; Gnatt,  
                       A.; Neville, L.; Lieman-Hurwitz, J.; Lev-Lehman, E.;  
                       Ginzberg, D.; Lapidot-Lifson, Y.; Zakut, H.  
           #journal   Proc. Natl. Acad. Sci. U.S.A. (1990) 87:9688-9692  
           #title     Molecular cloning and construction of the coding region for  
                       human acetylcholinesterase reveals a G+C-rich attenuating  
                       structure.  
           #cross-references MUID:91088577; PMID:2263619  
           #accession  A39256  
           ##molecule\_type mRNA; DNA  
           ##residues  1-614 ##label SOR  
           ##cross-references UNIPROT:P22303; UNIPARC:UPI0000125216; GB:M55040;  
                               NID:g177974; PIDN:AAA68151.1; PID:g177975  
           ##note this sequence represents composite of clones including clone  
                       ABGACHE from adult brain basal nuclei cDNA library; no  
                       alternative splicing should occur in the region 5' to  
                       ABGACHE so this sequence should represent an authentic brain  
                       splice form  
 REFERENCE       S03959  
           #authors   Chhajlani, V.; Derr, D.; Earles, B.; Schmell, E.; August, T.  
           #journal   FEBS Lett. (1989) 247:279-282  
           #title     Purification and partial amino acid sequence analysis of  
                       human erythrocyte acetylcholinesterase.  
           #cross-references MUID:89232136; PMID:2714437  
           #accession  S03959  
           ##molecule\_type protein  
           ##residues  256-266, 'Y', 268-273; 306-308, 'X', 310-313, 'X', 315-316, 'D',

## SEQUENCE

                  5                  10                  15                  20                  25                  30  
1 M R P P Q C L L H T P S L A S P L L L L L L W L L G G G V G  
31 A E G R E D A E L L V T V R G G R L R G I R L K T P G G P V  
61 S A F L G I P F A E P P M G P R R F L P P E P K Q P W S G V  
91 V D A T T F Q S V C Y Q Y V D T L Y P G F E G T E M W N P N  
121 R E L S E D C L Y L N V W T P Y P R P T S P T P V L V W I Y  
151 G G G F Y S G A S S L D V Y D G R F L V Q A E R T V L V S M  
181 N Y R V G A F G F L A L P G S R E A P G N V G L L D Q R L A  
211 L Q W V Q E N V A A F G G D P T S V T L F G E S A G A A S V  
241 G M H L L S P P S R G L F H R A V L Q S G A P N G P W A T V  
271 G M G E A R R R A T Q L A H L V G C P P G G T G G N D T E L  
301 V A C L R T R P A Q V L V N H E W H V L P Q E S V F R F S F  
331 V P V V D G D F L S D T P E A L I N A G D F H G L Q V L V G  
361 V V K D E G S Y F L V Y G A P G F S K D N E S L I S R A E F  
391 L A G V R V G V P Q V S D L A A E A V V L H Y T D W L H P E  
421 D P A R L R E A L S D V V G D H N V V C P V A Q L A G R L A  
451 A Q G A R V Y A Y V F E H R A S T L S W P L W M G V P H G Y  
481 E I E F I F G I P L D P S R N Y T A E E K I F A Q R L M R Y  
511 W A N F A R T G D P N E P R D P K A P Q W P P Y T A G A Q Q  
541 Y V S L D L R P L E V R R G L R A Q A C A F W N R F L P K L  
571 L S A T D T L D E A E R Q W K A E F H R W S S Y M V H W K N  
601 Q F D H Y S K Q D R C S D L

## Banques de données protéiques



**Créée en:**1971

**Contenu:** plus grande banque de données protéiques avec plus de 160 000 séquences

**Contenu:** Structures 3D, séquences de protéines, ADN et ARN

**Accessible gratuitement via son site web:** <https://www.rcsb.org/>

**Formats disponibles:** Fasta (séquence), PDB (3D)

acetylcholinesterase

Go

Advanced Search | Browse by Annotations



Welcome

Deposit

Search

Visualize

Analyze

Download

Learn

### A Structural View of Biology

This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

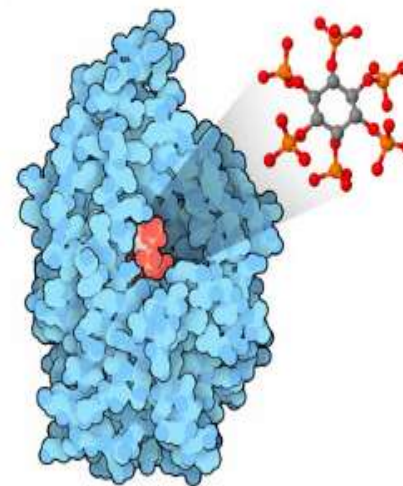
As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

#### Award-Winning Videos on Antibiotic Resistance



### September Molecule of the Month



Phytase

#### Latest Entries

As of Tuesday Sep 11 2018



#### Features & Highlights



New Architecture and Services Enable Faster Access to More Information

#### News

Publications



Meet PDB in VR with Nanome Explore PDB structures in

## 2. Télécharger la séquence au format « FASTA »

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB

RCSB PDB PROTEIN DATA BANK 144211 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

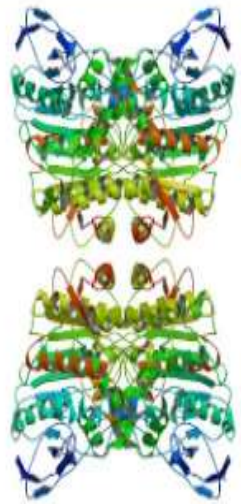
Search by PDB ID, author, macromolecule, sequence, or ligands Go

Advanced Search | Browse by Annotations

PDB-101 PDB EMDataBank PROTEIN DATA BANK

Structure Summary 3D View Annotations Sequence Sequence Similarity Structure Similarity Experiment

Biological Assembly 1



**1EEA**  
Acetylcholinesterase  
DOI: 10.2210/pdb1EEA/pdb  
Classification: [HYDROLASE](#)  
Organism(s): [Tetronarce californica](#)

Deposited: 1999-01-26 Released: 1999-02-01  
Deposition Author(s): [Raves, M.L.](#), [Giles, K.](#), [Schrag, J.D.](#), [Schmid, M.F.](#), [Silman, I.](#), [Sussman, J.L.](#)

Experimental Data Snapshot  
Method: X-RAY DIFFRACTION  
Resolution: 4.5 Å

wwPDB Validation

- Metric
- Clashscore
- Ramachandran outliers
- Sidechain outliers

Structure Factors (CIF)  
Structure Factors (CIF - gz)  
2fo-fc Map (DSN6)  
fo-fc Map (DSN6)  
Map Coefficients (MTZ format)

3D View: [Structure](#) | [Electron Density](#)

Standalone Viewers  
[Protein Workshop](#) | [Ligand Explorer](#)

Global Symmetry: Dihedral - D2 (3D View)

This is version 1.4 of the entry. See complete [history](#).

[Download Primary Citation](#)

[www.rcsb.org/pdb/download/downloadFastaFiles.do?structureIdList=1EEA&compressionType=uncompressed](http://www.rcsb.org/pdb/download/downloadFastaFiles.do?structureIdList=1EEA&compressionType=uncompressed)

>1EEA:A | PDBID | CHAIN | SEQUENCE

DDHSELLVNTKSGKVMGTRVPVLSSHISAF LGIPFAEPPVGNMRFRRPEPKKPWSGVWNA  
ST  
YPNNCQQYVDEQFPGFSG  
SEMWNPNREMSDCLYLNIWVPSRPRKSTTVMVWIYGGGFYSGSSTLDVYNGKYLAYTEEVV  
LVSLSYRVGAFGFLALHG  
SQEAPGNVGLLDQRMALQWVHDNIQFFGGDPKTVTIFGESAGGASVGMHILSPGSRDLFRA  
ILQSGSPNCPWASVSVAE  
GRRRAVELGRNLNCNLNSDEELIHCLREKKPQELIDVEWNVLPFDSIFRFSFVPVIDGEFFP  
TSLESMLNSGNFKKTQIL  
LGVNKDEGSFFLLYGAPGFSKDSISKISREDFMSGVKLSVPHANDLGLDAVTLQYTDWMDDN  
NGIKNRDGLDDIVGDHNV  
ICPLMHFVNKYTKFGNGTYLYFFNHRASNLVWPEWVGVIHG YEIEFVFGLPLVKELNYTAE  
EALSRRIMHYWATFAKTG  
NPNEPHSQESKWPLFTTKEQKFIDLNTEPMKVHQRLRVQMCVFWNQFLPKLLNA

# Systemes d'interrogation des banques de données biologiques

Facilitent l'exploitation et l'extraction des données

Permettent une interrogation multicritère et simple

Adoptés à plusieurs systèmes biologiques (EMBL, GenBank...etc.)

**Exemples:** ACNUC, SRS, ENTREZ

# Le système ACNUC

Développé à Lyon chez le Professeur Grantham

Il est principalement utilisé par les grandes banques de données biologiques (EMBL, GenBank, PIR-NBRF, PDB).

Permet une exploitation facile par **plusieurs critères de sélection** (mot-clé, auteurs, espèces, revues, type de molécule...)

Permet d'effectuer des liens entre critères en utilisant un langage basé sur les opérations logiques **ET, OU et NON**

Peut répondre à ce type de requêtes « je recherche dans l'EMBL les séquences codantes pour des protéines de levure ou de souris qui ne soient pas mitochondriales et ayant une longueur supérieure à 500 paires de bases

# Le système SRS

## Sequence Retrieval System

Créé en 1993 par Etzold et argos

Offre la possibilité d'exploiter en même temps et avec le même langage de requête plusieurs bases de données

Exemple: on peut rechercher dans PIR-NBRF les séquences issues d'une interrogation de la PDB

Le langage de requête est similaire à celui d'ACNUC (requête par plusieurs critères de sélection avec les opérateurs logiques ET, OU et NON

## Le système ENTREZ

Le système global de recherches inter-bases de données **Entrez**

permet d'accéder à des bases de données du NCBI.

Couvre actuellement plus de 20 bases de données génomiques et protéiques telles que : PIR, Swiss-Prot, PDB, GenBank, DDBJ et EMBL