Анализ белковых последовательностей: Swiss-Prot

Swiss-Prot – одна из первых баз данных белковых последовательностей, “gold standard” белковой аннотации.

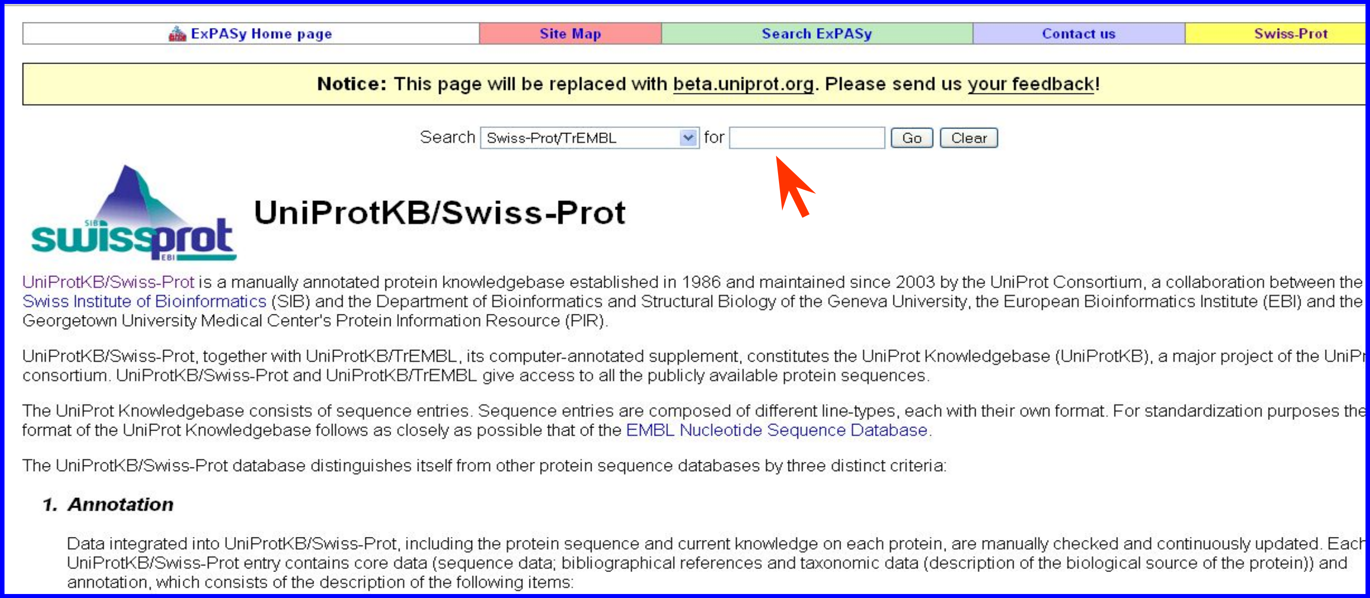
Аннотация выполнена вручную группой профессиональных экспертов на основе экспериментальной информации, описанной в научных статьях.

Организована в 1986 году – SIB+EBI+PIR+GU = prof. Amos Bairoch

На сегодняшний день – Release 56.2 - **398181** последовательностей

UniProt DB

- UniProt = Swiss-Prot + TrEMBL (Translated EMBL sequence database)
- TrEMBL – Release 39.2 - **6534543** sequences



Notice: This page will be replaced with beta.uniprot.org. Please send us your feedback!

UniProtKB/Swiss-Prot

UniProtKB/Swiss-Prot is a manually annotated protein knowledgebase established in 1986 and maintained since 2003 by the UniProt Consortium, a collaboration between the Swiss Institute of Bioinformatics (SIB) and the Department of Bioinformatics and Structural Biology of the Geneva University, the European Bioinformatics Institute (EBI) and the Georgetown University Medical Center's Protein Information Resource (PIR).

UniProtKB/Swiss-Prot, together with UniProtKB/TrEMBL, its computer-annotated supplement, constitutes the UniProt Knowledgebase (UniProtKB), a major project of the UniPr consortium. UniProtKB/Swiss-Prot and UniProtKB/TrEMBL give access to all the publicly available protein sequences.

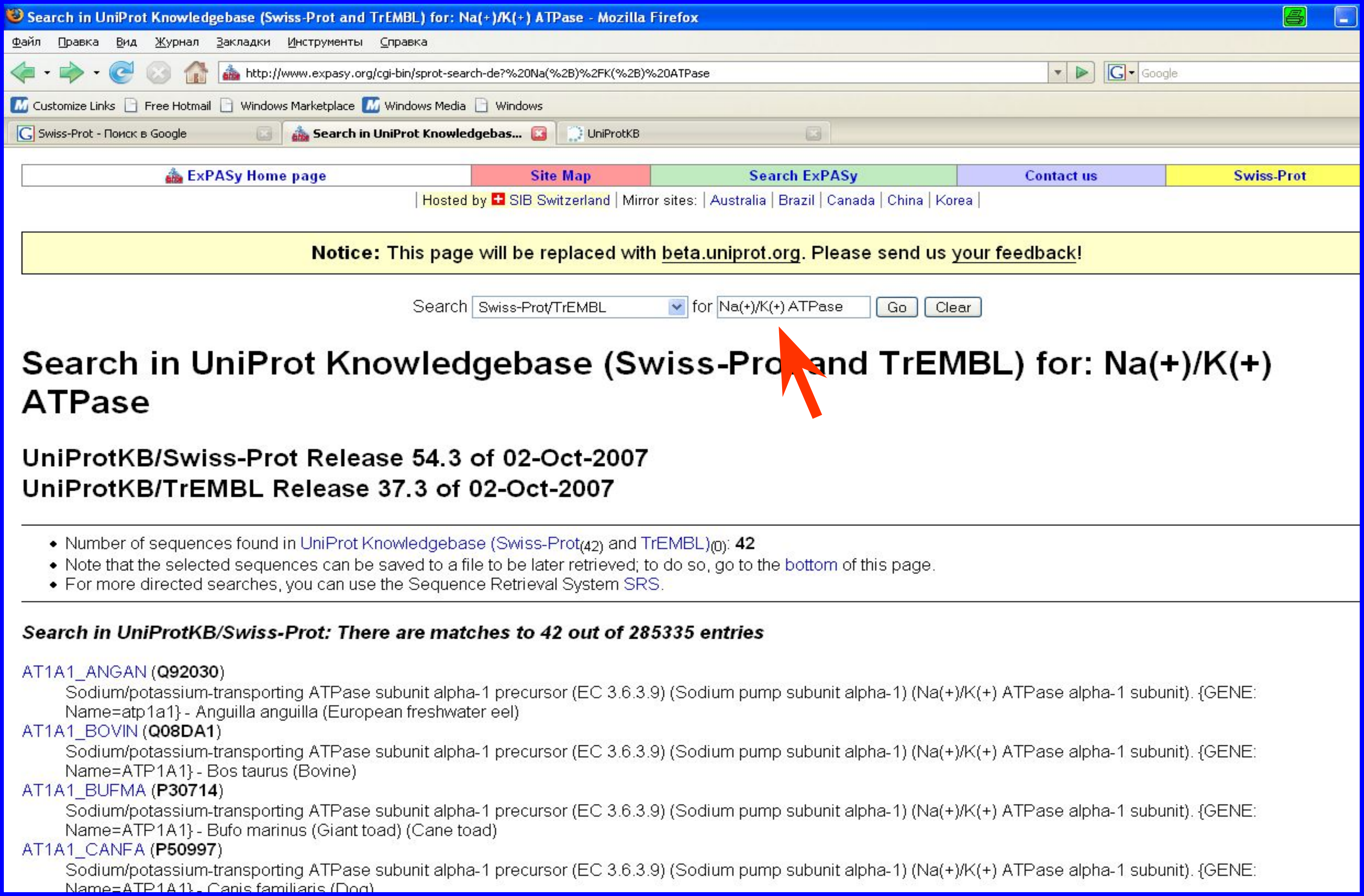
The UniProt Knowledgebase consists of sequence entries. Sequence entries are composed of different line-types, each with their own format. For standardization purposes the format of the UniProt Knowledgebase follows as closely as possible that of the EMBL Nucleotide Sequence Database.

The UniProtKB/Swiss-Prot database distinguishes itself from other protein sequence databases by three distinct criteria:

1. ***Annotation***

   Data integrated into UniProtKB/Swiss-Prot, including the protein sequence and current knowledge on each protein, are manually checked and continuously updated. Each UniProtKB/Swiss-Prot entry contains core data (sequence data; bibliographical references and taxonomic data (description of the biological source of the protein)) and annotation, which consists of the description of the following items:

Поиск белка в Swiss-Prot (по названию)



| Hosted by SIB Switzerland | Mirror sites: | Australia | Brazil | Canada | China | Korea |

Notice: This page will be replaced with beta.uniprot.org. Please send us your feedback!

Search Swiss-Prot/TrEMBL for Na(+)/K(+) ATPase

Search in UniProt Knowledgebase (Swiss-Prot and TrEMBL) for: Na(+)/K(+) ATPase

UniProtKB/Swiss-Prot Release 54.3 of 02-Oct-2007
UniProtKB/TrEMBL Release 37.3 of 02-Oct-2007

- Number of sequences found in UniProt Knowledgebase (Swiss-Prot(42) and TrEMBL)(0): **42**
- Note that the selected sequences can be saved to a file to be later retrieved; to do so, go to the bottom of this page.
- For more directed searches, you can use the Sequence Retrieval System SRS.

Search in UniProtKB/Swiss-Prot: There are matches to 42 out of 285335 entries

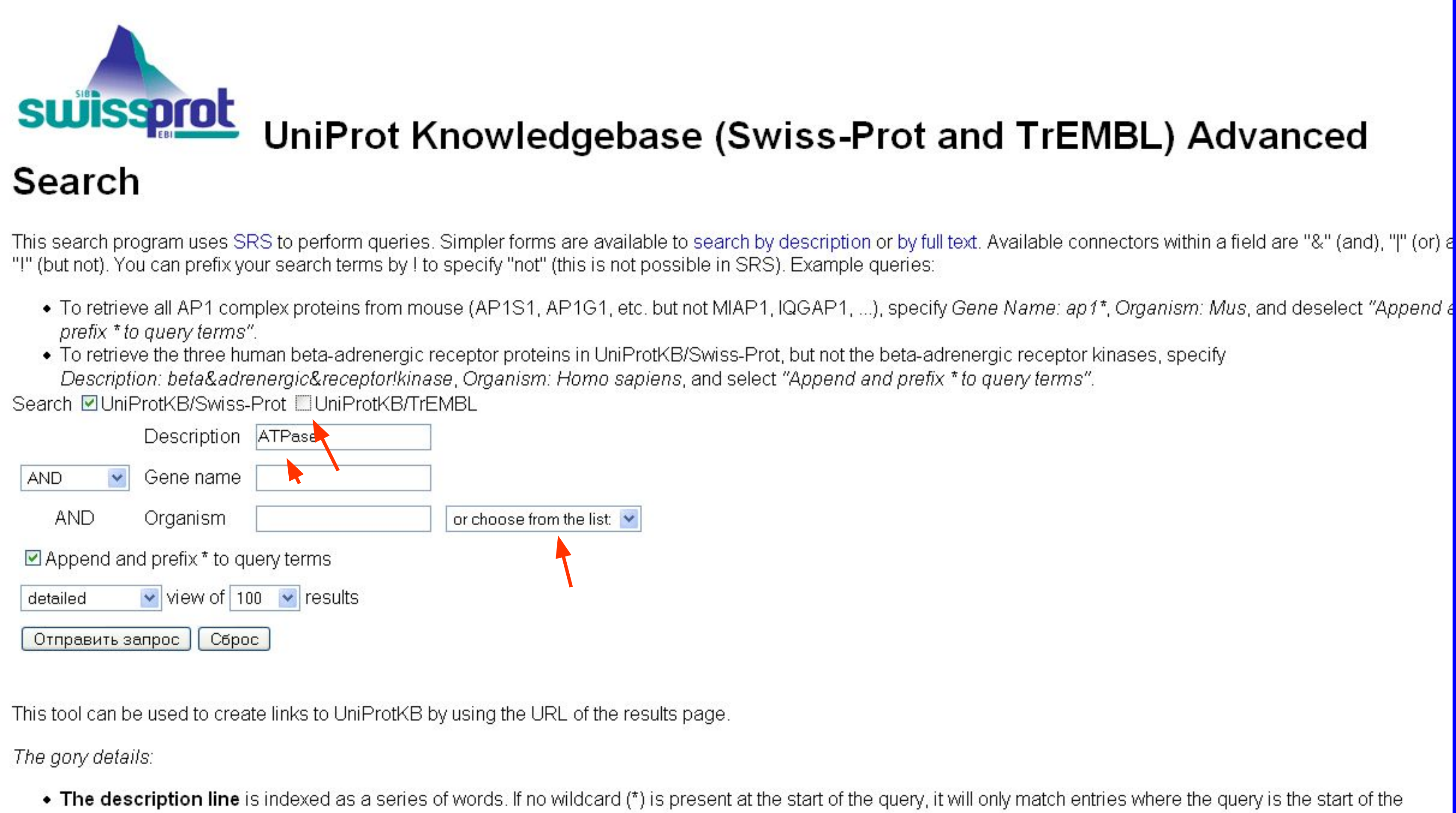
AT1A1_ANGAN (**Q92030**)
Sodium/potassium-transporting ATPase subunit alpha-1 precursor (EC 3.6.3.9) (Sodium pump subunit alpha-1) (Na(+)/K(+) ATPase alpha-1 subunit). {GENE: Name=atp1a1} - Anguilla anguilla (European freshwater eel)

AT1A1_BOVIN (**Q08DA1**)
Sodium/potassium-transporting ATPase subunit alpha-1 precursor (EC 3.6.3.9) (Sodium pump subunit alpha-1) (Na(+)/K(+) ATPase alpha-1 subunit). {GENE: Name=ATP1A1} - Bos taurus (Bovine)

AT1A1_BUFMA (**P30714**)
Sodium/potassium-transporting ATPase subunit alpha-1 precursor (EC 3.6.3.9) (Sodium pump subunit alpha-1) (Na(+)/K(+) ATPase alpha-1 subunit). {GENE: Name=ATP1A1} - Bufo marinus (Giant toad) (Cane toad)

AT1A1_CANFA (**P50997**)
Sodium/potassium-transporting ATPase subunit alpha-1 precursor (EC 3.6.3.9) (Sodium pump subunit alpha-1) (Na(+)/K(+) ATPase alpha-1 subunit). {GENE: Name=ATP1A1} - Canis familiaris (Dog)

Advances search



UniProt Knowledgebase (Swiss-Prot and TrEMBL) Advanced Search

This search program uses SRS to perform queries. Simpler forms are available to search by description or by full text. Available connectors within a field are "&" (and), "|" (or) a
"!" (but not). You can prefix your search terms by ! to specify "not" (this is not possible in SRS). Example queries:

- To retrieve all AP1 complex proteins from mouse (AP1S1, AP1G1, etc. but not MIAP1, IQGAP1, ...), specify *Gene Name: ap1**, *Organism: Mus*, and deselect *"Append a prefix * to query terms"*.
- To retrieve the three human beta-adrenergic receptor proteins in UniProtKB/Swiss-Prot, but not the beta-adrenergic receptor kinases, specify *Description: beta&adrenergic&receptor!kinase*, *Organism: Homo sapiens*, and select *"Append and prefix * to query terms"*.

Search ☑ UniProtKB/Swiss-Prot ☐ UniProtKB/TrEMBL

Description: ATPase

AND Gene name

AND Organism — or choose from the list:

☑ Append and prefix * to query terms

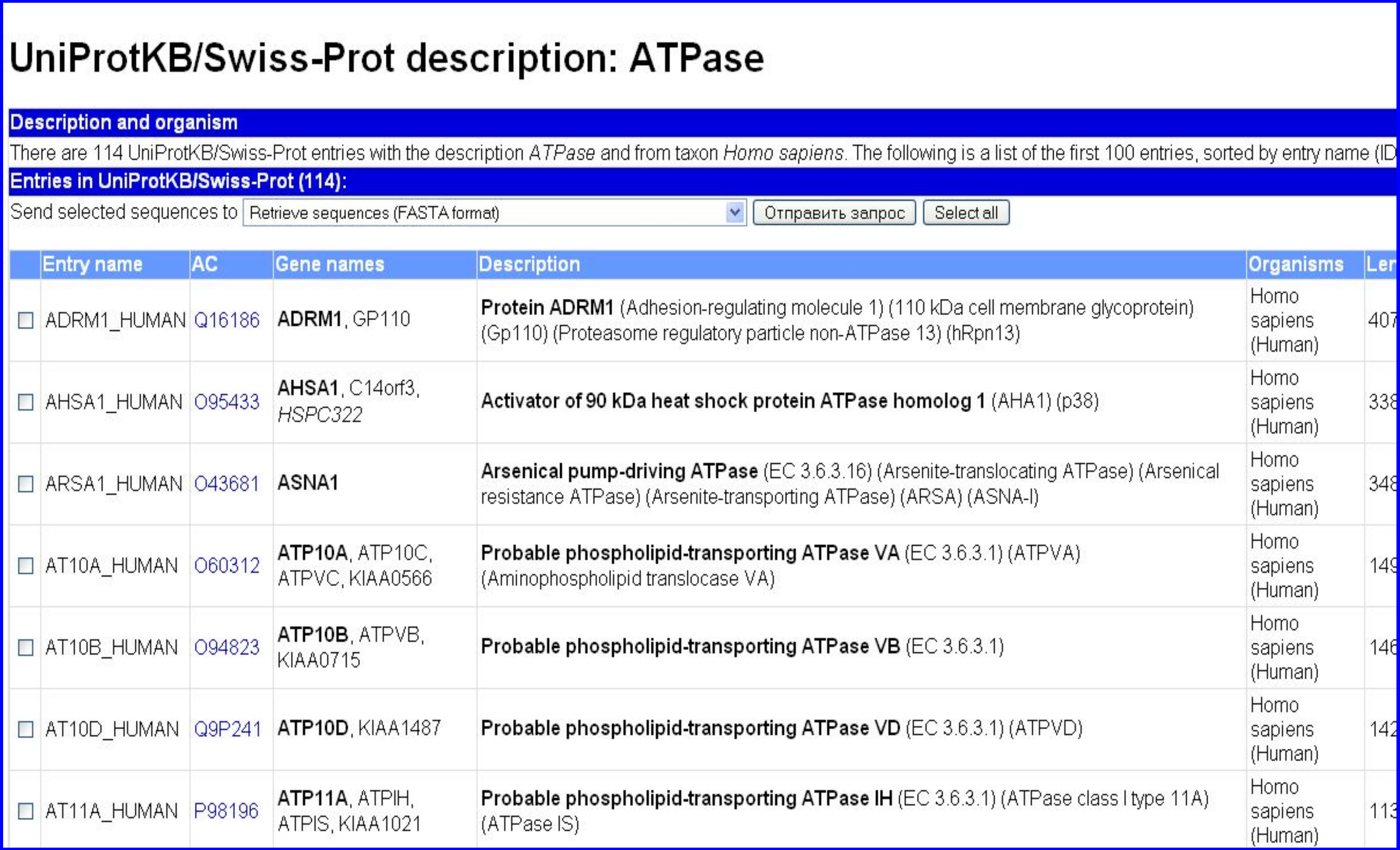
detailed view of 100 results

This tool can be used to create links to UniProtKB by using the URL of the results page.

The gory details:

- **The description line** is indexed as a series of words. If no wildcard (*) is present at the start of the query, it will only match entries where the query is the start of the

Результаты



UniProtKB/Swiss-Prot description: ATPase

Description and organism

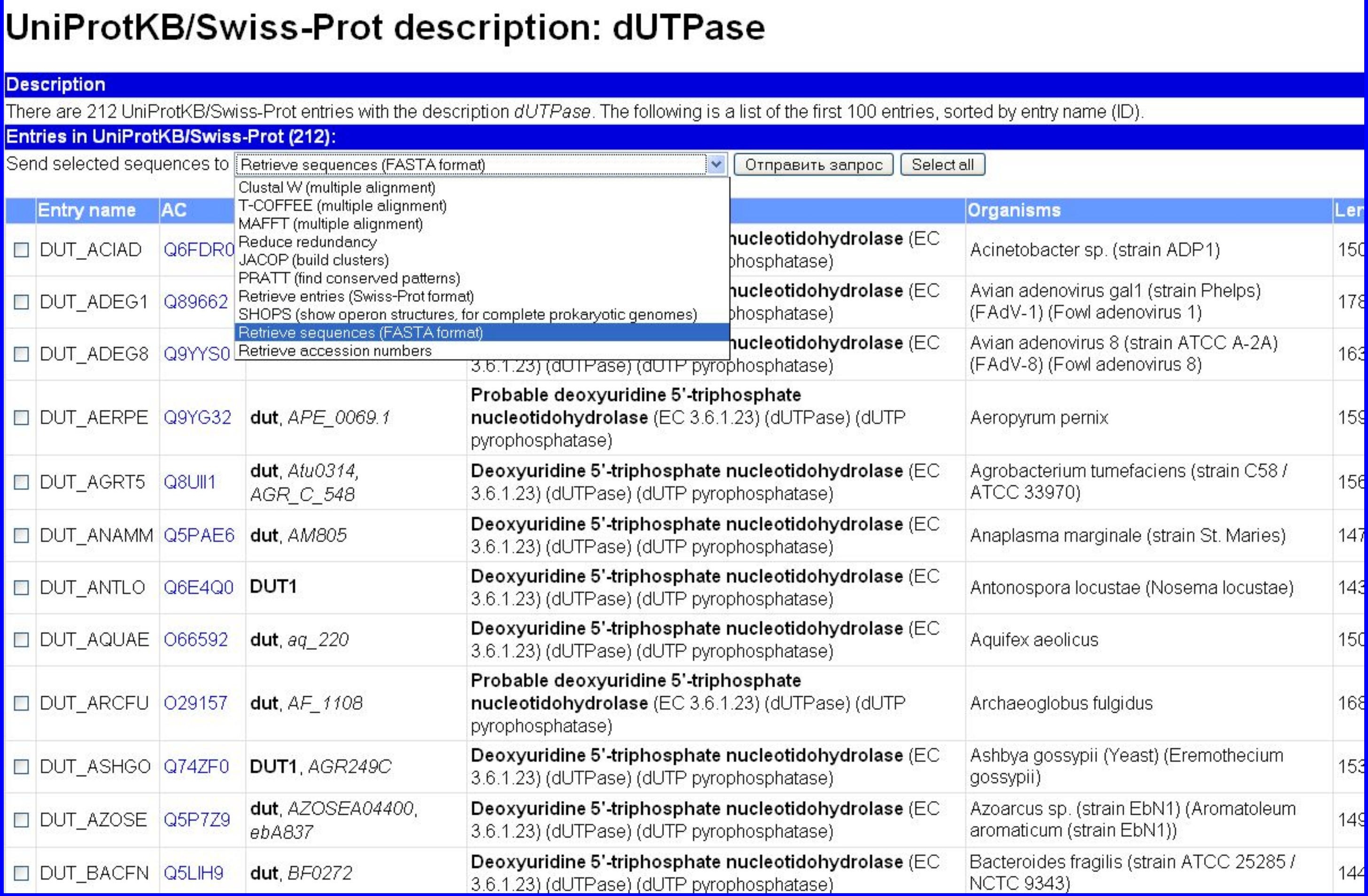
There are 114 UniProtKB/Swiss-Prot entries with the description *ATPase* and from taxon *Homo sapiens*. The following is a list of the first 100 entries, sorted by entry name (ID

Entries in UniProtKB/Swiss-Prot (114):

Send selected sequences to Retrieve sequences (FASTA format)

	Entry name	AC	Gene names	Description	Organisms	Len
☐	ADRM1_HUMAN	Q16186	**ADRM1**, GP110	**Protein ADRM1** (Adhesion-regulating molecule 1) (110 kDa cell membrane glycoprotein) (Gp110) (Proteasome regulatory particle non-ATPase 13) (hRpn13)	Homo sapiens (Human)	407
☐	AHSA1_HUMAN	O95433	**AHSA1**, C14orf3, *HSPC322*	**Activator of 90 kDa heat shock protein ATPase homolog 1** (AHA1) (p38)	Homo sapiens (Human)	338
☐	ARSA1_HUMAN	O43681	**ASNA1**	**Arsenical pump-driving ATPase** (EC 3.6.3.16) (Arsenite-translocating ATPase) (Arsenical resistance ATPase) (Arsenite-transporting ATPase) (ARSA) (ASNA-I)	Homo sapiens (Human)	348
☐	AT10A_HUMAN	O60312	**ATP10A**, ATP10C, ATPVC, KIAA0566	**Probable phospholipid-transporting ATPase VA** (EC 3.6.3.1) (ATPVA) (Aminophospholipid translocase VA)	Homo sapiens (Human)	149
☐	AT10B_HUMAN	O94823	**ATP10B**, ATPVB, KIAA0715	**Probable phospholipid-transporting ATPase VB** (EC 3.6.3.1)	Homo sapiens (Human)	146
☐	AT10D_HUMAN	Q9P241	**ATP10D**, KIAA1487	**Probable phospholipid-transporting ATPase VD** (EC 3.6.3.1) (ATPVD)	Homo sapiens (Human)	142
☐	AT11A_HUMAN	P98196	**ATP11A**, ATPIH, ATPIS, KIAA1021	**Probable phospholipid-transporting ATPase IH** (EC 3.6.3.1) (ATPase class I type 11A) (ATPase IS)	Homo sapiens (Human)	113

Выборка гомологичных белков



UniProtKB/Swiss-Prot description: dUTPase

Description

There are 212 UniProtKB/Swiss-Prot entries with the description *dUTPase*. The following is a list of the first 100 entries, sorted by entry name (ID).

Entries in UniProtKB/Swiss-Prot (212):

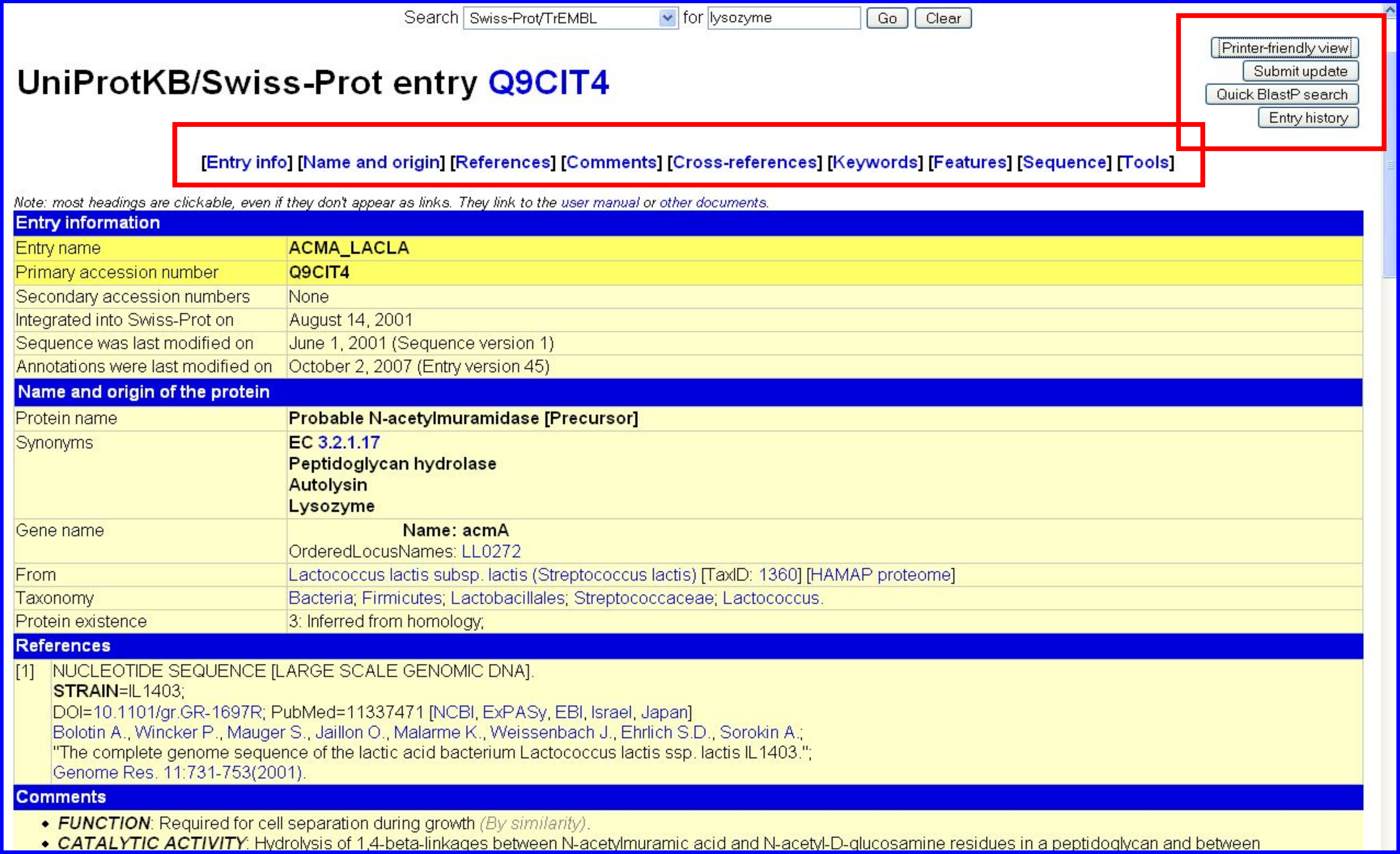
Send selected sequences to: Retrieve sequences (FASTA format)

	Entry name	AC			Organisms	Ler
☐	DUT_ACIAD	Q6FDR0		ucleotidohydrolase (EC phosphatase)	Acinetobacter sp. (strain ADP1)	150
☐	DUT_ADEG1	Q89662		ucleotidohydrolase (EC phosphatase)	Avian adenovirus gal1 (strain Phelps) (FAdV-1) (Fowl adenovirus 1)	178
☐	DUT_ADEG8	Q9YYS0		ucleotidohydrolase (EC 3.6.1.23) (dUTPase) (dUTP pyrophosphatase)	Avian adenovirus 8 (strain ATCC A-2A) (FAdV-8) (Fowl adenovirus 8)	163
☐	DUT_AERPE	Q9YG32	**dut**, *APE_0069.1*	**Probable deoxyuridine 5'-triphosphate nucleotidohydrolase** (EC 3.6.1.23) (dUTPase) (dUTP pyrophosphatase)	Aeropyrum pernix	159
☐	DUT_AGRT5	Q8UII1	**dut**, *Atu0314*, *AGR_C_548*	**Deoxyuridine 5'-triphosphate nucleotidohydrolase** (EC 3.6.1.23) (dUTPase) (dUTP pyrophosphatase)	Agrobacterium tumefaciens (strain C58 / ATCC 33970)	156
☐	DUT_ANAMM	Q5PAE6	**dut**, *AM805*	**Deoxyuridine 5'-triphosphate nucleotidohydrolase** (EC 3.6.1.23) (dUTPase) (dUTP pyrophosphatase)	Anaplasma marginale (strain St. Maries)	147
☐	DUT_ANTLO	Q6E4Q0	**DUT1**	**Deoxyuridine 5'-triphosphate nucleotidohydrolase** (EC 3.6.1.23) (dUTPase) (dUTP pyrophosphatase)	Antonospora locustae (Nosema locustae)	143
☐	DUT_AQUAE	O66592	**dut**, *aq_220*	**Deoxyuridine 5'-triphosphate nucleotidohydrolase** (EC 3.6.1.23) (dUTPase) (dUTP pyrophosphatase)	Aquifex aeolicus	150
☐	DUT_ARCFU	O29157	**dut**, *AF_1108*	**Probable deoxyuridine 5'-triphosphate nucleotidohydrolase** (EC 3.6.1.23) (dUTPase) (dUTP pyrophosphatase)	Archaeoglobus fulgidus	168
☐	DUT_ASHGO	Q74ZF0	**DUT1**, *AGR249C*	**Deoxyuridine 5'-triphosphate nucleotidohydrolase** (EC 3.6.1.23) (dUTPase) (dUTP pyrophosphatase)	Ashbya gossypii (Yeast) (Eremothecium gossypii)	153
☐	DUT_AZOSE	Q5P7Z9	**dut**, *AZOSEA04400*, *ebA837*	**Deoxyuridine 5'-triphosphate nucleotidohydrolase** (EC 3.6.1.23) (dUTPase) (dUTP pyrophosphatase)	Azoarcus sp. (strain EbN1) (Aromatoleum aromaticum (strain EbN1))	149
☐	DUT_BACFN	Q5LIH9	**dut**, *BF0272*	**Deoxyuridine 5'-triphosphate nucleotidohydrolase** (EC 3.6.1.23) (dUTPase) (dUTP pyrophosphatase)	Bacteroides fragilis (strain ATCC 25285 / NCTC 9343)	144

Сохранить в FASTA формате

```
>DUT_ACIAD
MKVQVKILDQRLGQEWPLPSYATTGSAGLDLRACLDEAIQIEPGQTVLIKTGMAIYIHDT
NFAGLILPRSGLGHKHGIVLGNLVGLIDSDYQGELMISVWNRGQNTFTLEPGERLAQYVL
VPVIQAEFEQVEEFVATDRGAGGFGHTGQK
>DUT_ADEG1
MDPFGSSSVPPCSTSDLPEPKLYFVRLSPHAVPPVRATHGAAGYDLFSAYDIKVPARGRA
LVPTDLVFQFPPGCYGRIAPRSGLAAKFFIDVGAGVIDPDYRGNVSVVLFNFSESSFNIR
RGDRVAQLILERIMVPELSELTQLGETDRGASGFGSTGMGAVDRNQRSVLEWLTPGSR
>DUT_ADEG8
MSFDSGCPPTPPVKLLFKKHSPFAVTPQRATSGAAGYDLCSSADVVVPPKSRSLIPTDLS
FQFPRGVYGRIAPRSGLAVKFFIDVGAGVIDSDYRGIVSVLLFNFSDHNFNVRRGDRIAQ
LILERHLTPDLEERSGLDETARGAAGFGSTGGFDTGVCPSSFS
>DUT_AERPE
MFLSGRDLVLLGVVKGHSNGAIQPAGVDLSVGEIESLADAGFLGEEDKIMPKGDRIQCEY
GVCELEPGAYRLRFNEVVSIPPGHVGFCFPRSSLLRMGCYLGCAVWDPGYTGRGQAMLLV
ANPHGLRLEMGSRIAQLVVARVEGPLTSLYKGDYQGEGL
>DUT_AGRT5
MTVQNDNRPLLRLVRLANGADLELPSYETRGAAGMDLRAAVPADEPLNLQPGERALVPTG
FIFEVPQGYEAQIRPRSGLAIKNGITCLNSPGTVDSDYRGEVKVILANLGQDDFTIERGM
RIAQMVIAPVTQVTVSEVTETSETARGAGGFGSTGV
>DUT_ANAMM
MLKVKILRLASGYGLPLPSYATPKSAGLDLYAAVDSKLVVHPGGRCAVKTGVALELPDGY
EAQIRSRSGLAANFGICVLNAPGTIDSDYRGEITVVLSNFGSEDYVISRGDRVAQMVIAP
VERVEWEEVNSITATSRGEGGFGSTGT
>DUT_ANTLO
MSEIITVKRLFSDAKIPVRHSEGAAAYDLYAYEDTVVAPNERKVIATGVRITVPLSCQGT
IYSRSGLALKYCIEIFGVNIGPGETKDIVVDIYNHGKMPFNVAKGDRIAQIVFIKLFGGD
LHEVSELSDTKRGSCGWGSTGIS
>DUT_AQUAE
MSKVILKIKRLPHAQDLPLPSYATPHSSGLDLRAAIEKPLKIKPFERVLIPTGLILEIPE
GYEGQVRPRSGLAWKKGLTVLNAPGTIDADYRGEVKVILVNLGNEEVVIERGERIAQLVI
APVQRVEVVEVEEVSQTQRGEGGFGSTGTK
>DUT_ARCFU
MAVLSGDEIRKLIQKEGLIRDYVDLETQIQPNGFDCTLRSVYRLRGCGRVDFDNSRRELP
ELEEVEFRDWVYLPKGVYRAKLNEVVRLGNDIMAIARPRSTLIRCGANVLTAVWDAGYEG
RSEVSIVVHNDYGIWLSRNARIIQLVFIRLSSPTKGYEGVYKGENIDS
>DUT_ASHGO
MTDQPAKKVHSAPTLKVQLRSENAIAPTKGSAAAAGYDIYASQDCVIPGRGQGLVATDVS
FTVPVGTYGRIAPRSGLAVKHGIQTGAGVVDRDYTGEVKIVLFNHSDRDYAVKRGDRVAQ
```

Стандартная запись Swiss-Prot



Search Swiss-Prot/TrEMBL for lysozyme

UniProtKB/Swiss-Prot entry Q9CIT4

[Entry info] [Name and origin] [References] [Comments] [Cross-references] [Keywords] [Features] [Sequence] [Tools]

Note: most headings are clickable, even if they don't appear as links. They link to the user manual or other documents.

Entry information

Entry name	**ACMA_LACLA**
Primary accession number	**Q9CIT4**
Secondary accession numbers	None
Integrated into Swiss-Prot on	August 14, 2001
Sequence was last modified on	June 1, 2001 (Sequence version 1)
Annotations were last modified on	October 2, 2007 (Entry version 45)

Name and origin of the protein

Protein name	**Probable N-acetylmuramidase [Precursor]**
Synonyms	**EC 3.2.1.17** **Peptidoglycan hydrolase** **Autolysin** **Lysozyme**
Gene name	**Name: acmA** OrderedLocusNames: LL0272
From	Lactococcus lactis subsp. lactis (Streptococcus lactis) [TaxID: 1360] [HAMAP proteome]
Taxonomy	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
Protein existence	3: Inferred from homology;

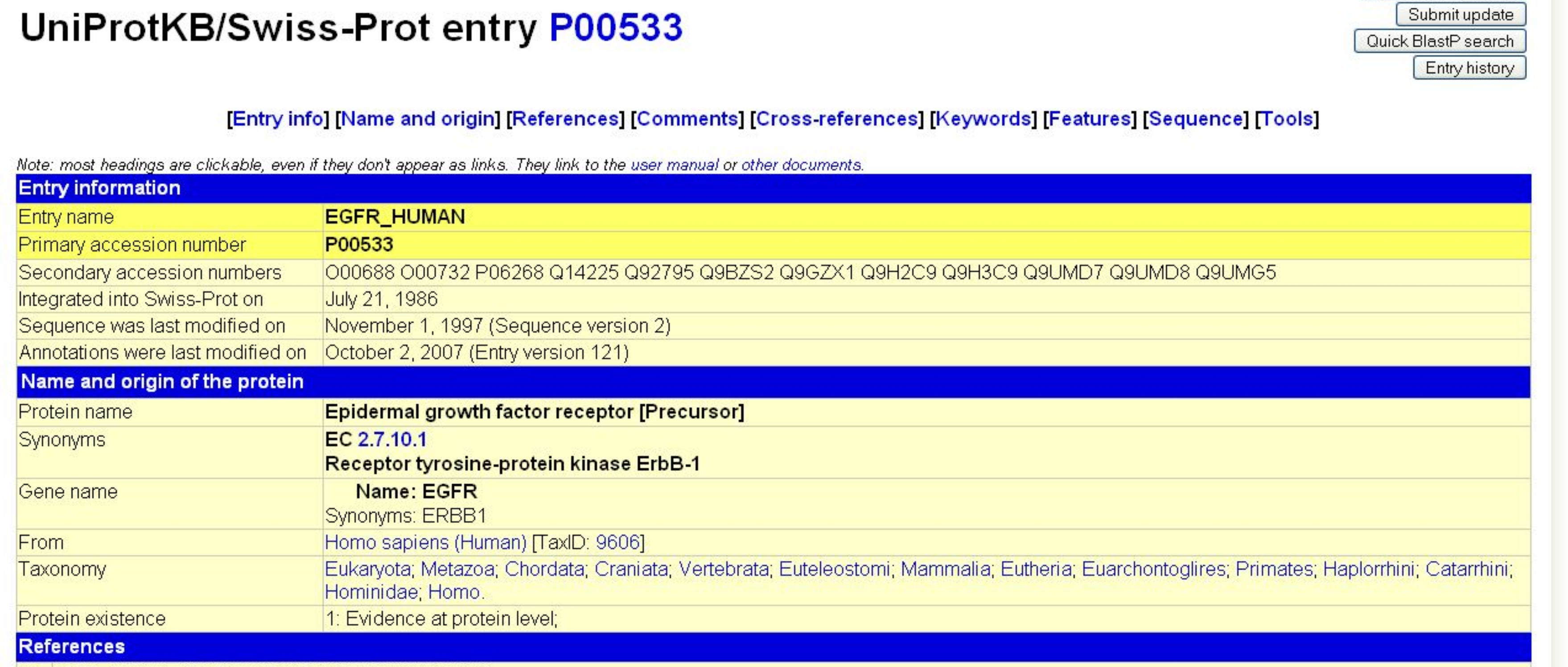
References

[1] NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
STRAIN=IL1403;
DOI=10.1101/gr.GR-1697R; PubMed=11337471 [NCBI, ExPASy, EBI, Israel, Japan]
Bolotin A., Wincker P., Mauger S., Jaillon O., Malarme K., Weissenbach J., Ehrlich S.D., Sorokin A.;
"The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. lactis IL1403.";
Genome Res. 11:731-753(2001).

Comments

- ***FUNCTION***: Required for cell separation during growth *(By similarity)*.
- ***CATALYTIC ACTIVITY***: Hydrolysis of 1,4-beta-linkages between N-acetylmuramic acid and N-acetyl-D-glucosamine residues in a peptidoglycan and between

Стандартные поля: entry, name, origin



UniProtKB/Swiss-Prot entry P00533

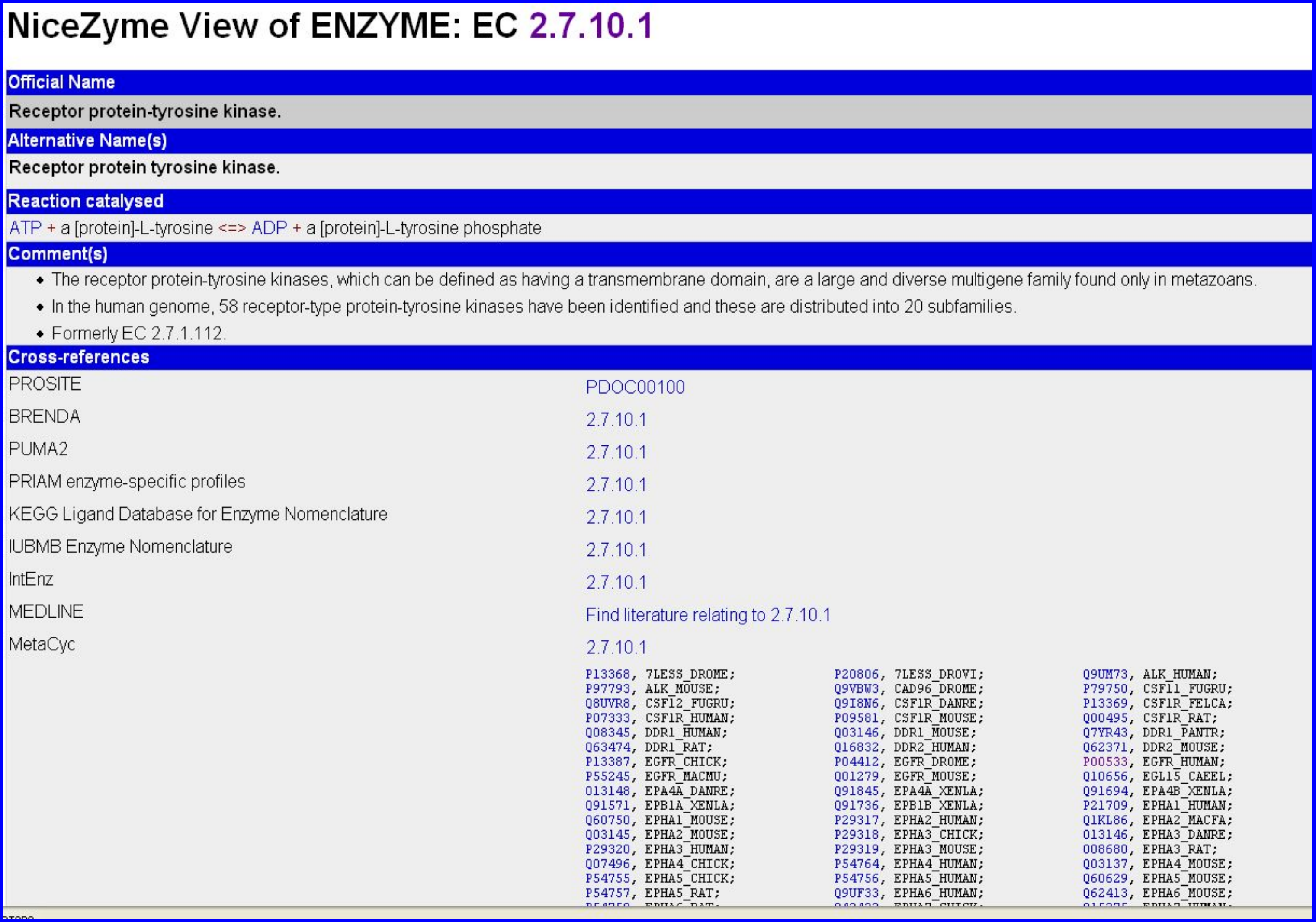
[Entry info] [Name and origin] [References] [Comments] [Cross-references] [Keywords] [Features] [Sequence] [Tools]

Note: most headings are clickable, even if they don't appear as links. They link to the user manual or other documents.

Entry information	
Entry name	**EGFR_HUMAN**
Primary accession number	**P00533**
Secondary accession numbers	O00688 O00732 P06268 Q14225 Q92795 Q9BZS2 Q9GZX1 Q9H2C9 Q9H3C9 Q9UMD7 Q9UMD8 Q9UMG5
Integrated into Swiss-Prot on	July 21, 1986
Sequence was last modified on	November 1, 1997 (Sequence version 2)
Annotations were last modified on	October 2, 2007 (Entry version 121)
Name and origin of the protein	
Protein name	**Epidermal growth factor receptor [Precursor]**
Synonyms	**EC** 2.7.10.1 **Receptor tyrosine-protein kinase ErbB-1**
Gene name	**Name: EGFR** Synonyms: ERBB1
From	Homo sapiens (Human) [TaxID: 9606]
Taxonomy	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.
Protein existence	1: Evidence at protein level;
References	

Название записи, уникальный идентификатор (ID), предыдущие идентификаторы соответствующей записи, даты первой и последней модификаций, распространенное название белка и его синонимы (EC номер для ферментов), название гена, организм и его таксономия, уровень подтверждения

NiceZyme (ферменты)



NiceZyme View of ENZYME: EC 2.7.10.1

Official Name

Receptor protein-tyrosine kinase.

Alternative Name(s)

Receptor protein tyrosine kinase.

Reaction catalysed

ATP + a [protein]-L-tyrosine <=> ADP + a [protein]-L-tyrosine phosphate

Comment(s)

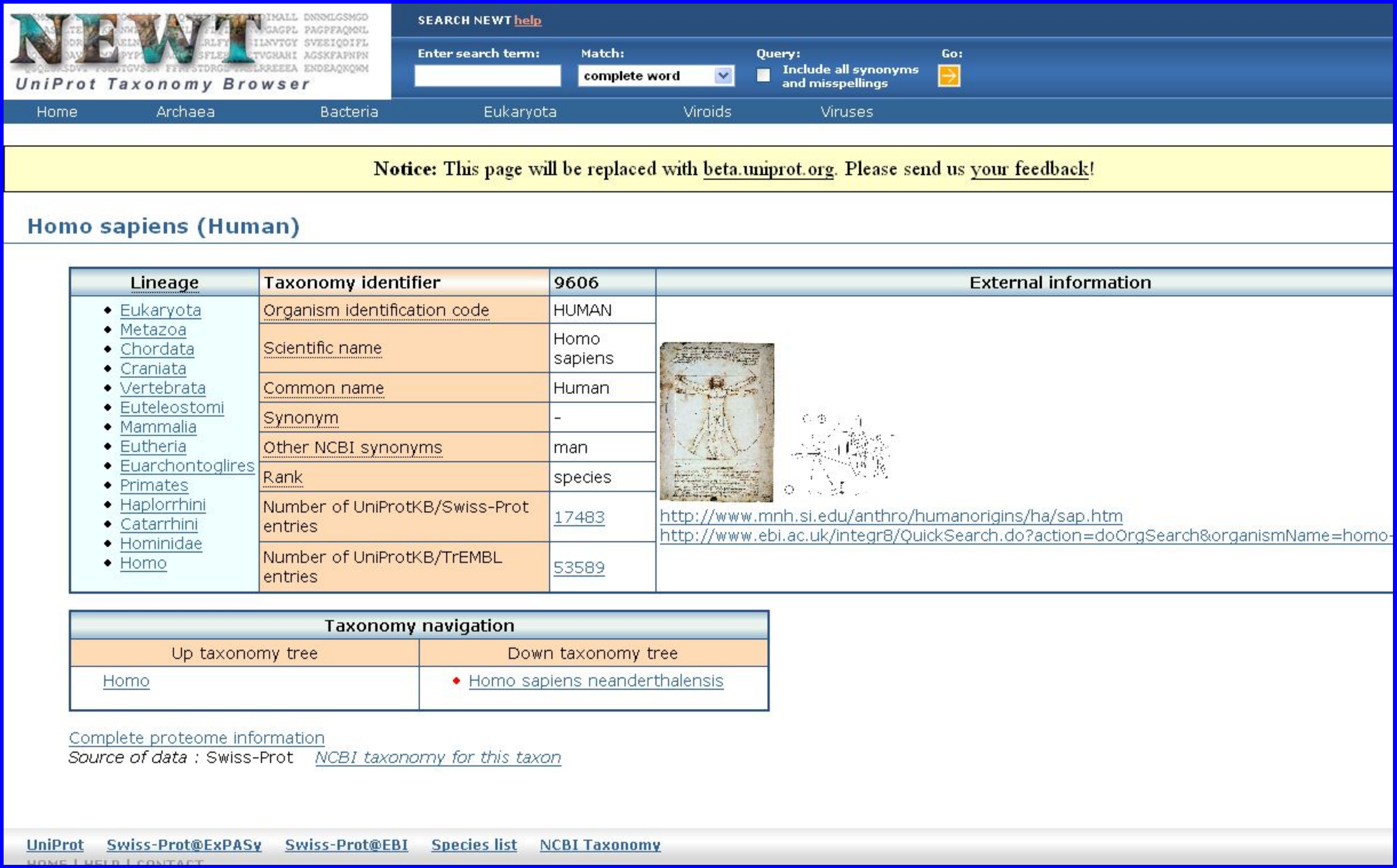
- The receptor protein-tyrosine kinases, which can be defined as having a transmembrane domain, are a large and diverse multigene family found only in metazoans.
- In the human genome, 58 receptor-type protein-tyrosine kinases have been identified and these are distributed into 20 subfamilies.
- Formerly EC 2.7.1.112.

Cross-references

PROSITE	PDOC00100
BRENDA	2.7.10.1
PUMA2	2.7.10.1
PRIAM enzyme-specific profiles	2.7.10.1
KEGG Ligand Database for Enzyme Nomenclature	2.7.10.1
IUBMB Enzyme Nomenclature	2.7.10.1
IntEnz	2.7.10.1
MEDLINE	Find literature relating to 2.7.10.1
MetaCyc	2.7.10.1

```
P13368, 7LESS_DROME;        P20806, 7LESS_DROVI;        Q9UM73, ALK_HUMAN;
P97793, ALK_MOUSE;          Q9VBW3, CAD96_DROME;        P79750, CSF11_FUGRU;
Q8UVR8, CSF12_FUGRU;        Q9I8N6, CSF1R_DANRE;        P13369, CSF1R_FELCA;
P07333, CSF1R_HUMAN;        P09581, CSF1R_MOUSE;        Q00495, CSF1R_RAT;
Q08345, DDR1_HUMAN;         Q03146, DDR1_MOUSE;         Q7YR43, DDR1_PANTR;
Q63474, DDR1_RAT;           Q16832, DDR2_HUMAN;         Q62371, DDR2_MOUSE;
P13387, EGFR_CHICK;         P04412, EGFR_DROME;         P00533, EGFR_HUMAN;
P55245, EGFR_MACMU;         Q01279, EGFR_MOUSE;         Q10656, EGL15_CAEEL;
O13148, EPA4A_DANRE;        Q91845, EPA4A_XENLA;        Q91694, EPA4B_XENLA;
Q91571, EPB1A_XENLA;        Q91736, EPB1B_XENLA;        P21709, EPHA1_HUMAN;
Q60750, EPHA1_MOUSE;        P29317, EPHA2_HUMAN;        Q1KL86, EPHA2_MACFA;
Q03145, EPHA2_MOUSE;        P29318, EPHA3_CHICK;        O13146, EPHA3_DANRE;
P29320, EPHA3_HUMAN;        P29319, EPHA3_MOUSE;        O08680, EPHA3_RAT;
Q07496, EPHA4_CHICK;        P54764, EPHA4_HUMAN;        Q03137, EPHA4_MOUSE;
P54755, EPHA5_CHICK;        P54756, EPHA5_HUMAN;        Q60629, EPHA5_MOUSE;
P54757, EPHA5_RAT;          Q9UF33, EPHA6_HUMAN;        Q62413, EPHA6_MOUSE;
```

Taxonomy Browser



Notice: This page will be replaced with beta.uniprot.org. Please send us your feedback!

Homo sapiens (Human)

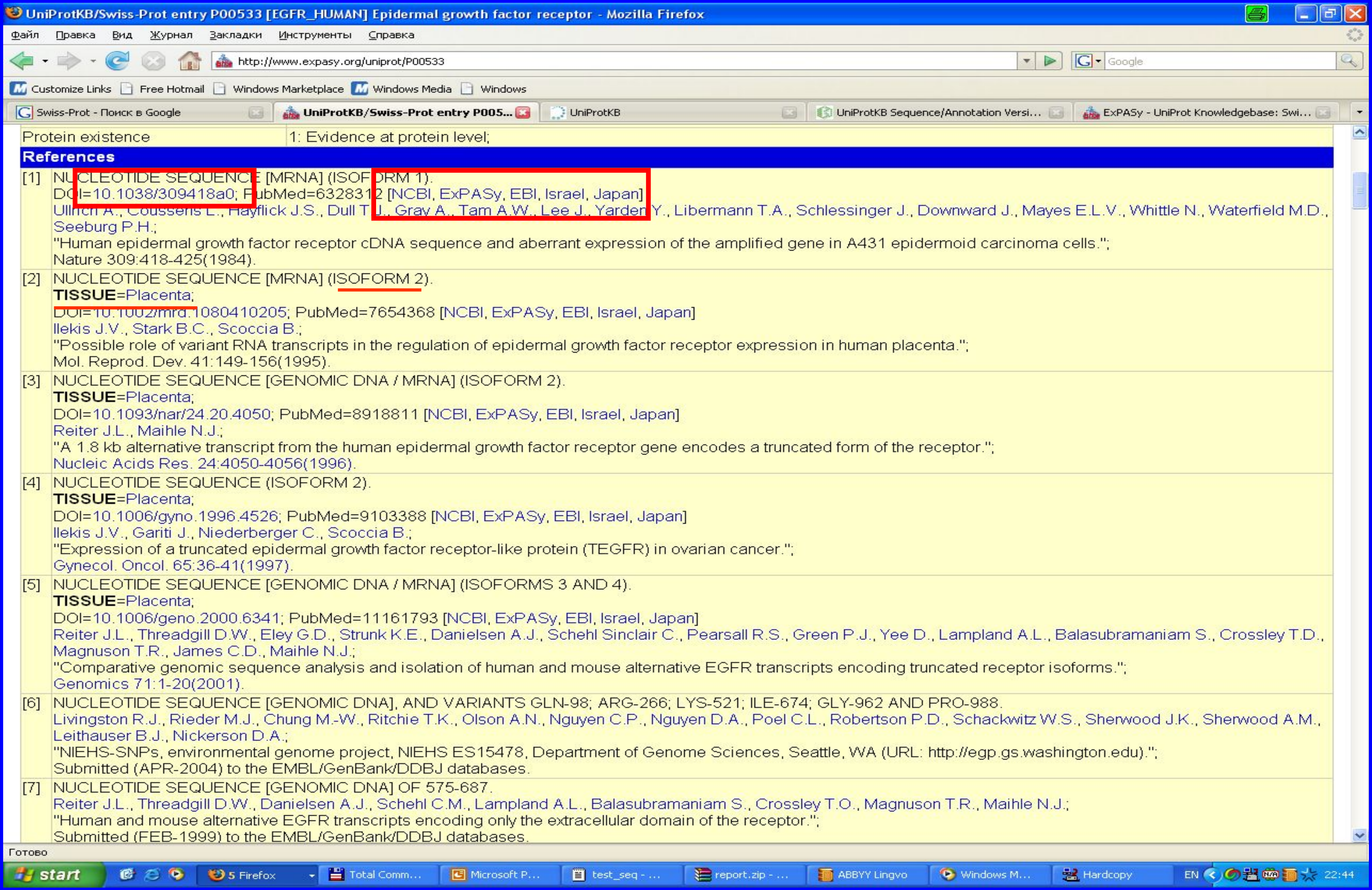
Lineage	Taxonomy identifier	9606	External information
Eukaryota Metazoa Chordata Craniata Vertebrata Euteleostomi Mammalia Eutheria Euarchontoglires Primates Haplorrhini Catarrhini Hominidae Homo	Organism identification code	HUMAN	http://www.mnh.si.edu/anthro/humanorigins/ha/sap.htm http://www.ebi.ac.uk/integr8/QuickSearch.do?action=doOrgSearch&organismName=homo-
	Scientific name	Homo sapiens	
	Common name	Human	
	Synonym	-	
	Other NCBI synonyms	man	
	Rank	species	
	Number of UniProtKB/Swiss-Prot entries	17483	
	Number of UniProtKB/TrEMBL entries	53589	

Taxonomy navigation	
Up taxonomy tree	Down taxonomy tree
Homo	Homo sapiens neanderthalensis

Complete proteome information
Source of data : Swiss-Prot *NCBI taxonomy for this taxon*

UniProt Swiss-Prot@ExPASy Swiss-Prot@EBI Species list NCBI Taxonomy

Ссылки на статьи, использованные для аннотации



Protein existence 1: Evidence at protein level;

References

[1] NUCLEOTIDE SEQUENCE [MRNA] (ISOFORM 1).
DOI=10.1038/309418a0; PubMed=6328312 [NCBI, ExPASy, EBI, Israel, Japan]
Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W., Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J., Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
"Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells.";
Nature 309:418-425(1984).

[2] NUCLEOTIDE SEQUENCE [MRNA] (ISOFORM 2).
TISSUE=Placenta;
DOI=10.1002/mrd.1080410205; PubMed=7654368 [NCBI, ExPASy, EBI, Israel, Japan]
Ilekis J.V., Stark B.C., Scoccia B.;
"Possible role of variant RNA transcripts in the regulation of epidermal growth factor receptor expression in human placenta.";
Mol. Reprod. Dev. 41:149-156(1995).

[3] NUCLEOTIDE SEQUENCE [GENOMIC DNA / MRNA] (ISOFORM 2).
TISSUE=Placenta;
DOI=10.1093/nar/24.20.4050; PubMed=8918811 [NCBI, ExPASy, EBI, Israel, Japan]
Reiter J.L., Maihle N.J.;
"A 1.8 kb alternative transcript from the human epidermal growth factor receptor gene encodes a truncated form of the receptor.";
Nucleic Acids Res. 24:4050-4056(1996).

[4] NUCLEOTIDE SEQUENCE (ISOFORM 2).
TISSUE=Placenta;
DOI=10.1006/gyno.1996.4526; PubMed=9103388 [NCBI, ExPASy, EBI, Israel, Japan]
Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;
"Expression of a truncated epidermal growth factor receptor-like protein (TEGFR) in ovarian cancer.";
Gynecol. Oncol. 65:36-41(1997).

[5] NUCLEOTIDE SEQUENCE [GENOMIC DNA / MRNA] (ISOFORMS 3 AND 4).
TISSUE=Placenta;
DOI=10.1006/geno.2000.6341; PubMed=11161793 [NCBI, ExPASy, EBI, Israel, Japan]
Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J., Schehl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L., Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D., Maihle N.J.;
"Comparative genomic sequence analysis and isolation of human and mouse alternative EGFR transcripts encoding truncated receptor isoforms.";
Genomics 71:1-20(2001).

[6] NUCLEOTIDE SEQUENCE [GENOMIC DNA], AND VARIANTS GLN-98; ARG-266; LYS-521; ILE-674; GLY-962 AND PRO-988.
Livingston R.J., Rieder M.J., Chung M.-W., Ritchie T.K., Olson A.N., Nguyen C.P., Nguyen D.A., Poel C.L., Robertson P.D., Schackwitz W.S., Sherwood J.K., Sherwood A.M., Leithauser B.J., Nickerson D.A.;
"NIEHS-SNPs, environmental genome project, NIEHS ES15478, Department of Genome Sciences, Seattle, WA (URL: http://egp.gs.washington.edu).";
Submitted (APR-2004) to the EMBL/GenBank/DDBJ databases.

[7] NUCLEOTIDE SEQUENCE [GENOMIC DNA] OF 575-687.
Reiter J.L., Threadgill D.W., Danielsen A.J., Schehl C.M., Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., Maihle N.J.;
"Human and mouse alternative EGFR transcripts encoding only the extracellular domain of the receptor.";
Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.

Комментарии

Comments

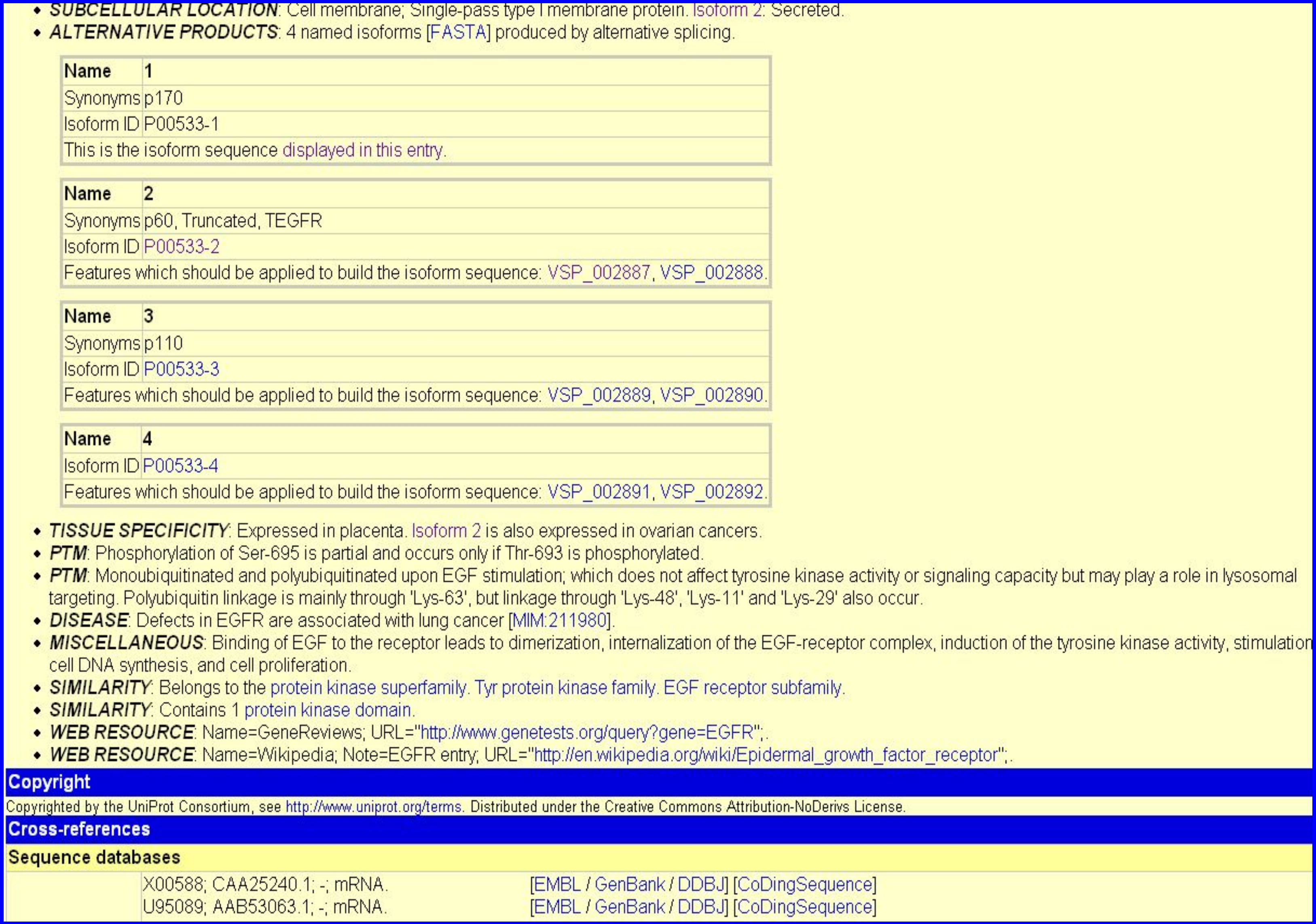
- ***FUNCTION***: Receptor for EGF, but also for other members of the EGF family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding EGF-like growth factor, GP30 an vaccinia virus growth factor. Is involved in the control of cell growth and differentiation.
- ***FUNCTION***: Isoform 2/truncated isoform may act as an antagonist.
- ***CATALYTIC ACTIVITY***: ATP + a [protein]-L-tyrosine = ADP + a [protein]-L-tyrosine phosphate.
- ***SUBUNIT***: Binds RIPK1. CBL interacts with the autophosphorylated C-terminal tail of the EGF receptor. Part of a complex with ERBB2 and either PIK3C2A or PIK3C2B. The autophosphorylated form interacts with PIK3C2B, maybe indirectly. Interacts with PELP1.
- ***INTERACTION***:
  Self; NbExp=1; IntAct=EBI-297353, EBI-297353;
  Q29376:- (xeno); NbExp=1; IntAct=EBI-297353, EBI-1256881;
  Q53FG0:-; NbExp=2; IntAct=EBI-297353, EBI-913954;
  P62157:CALM (xeno); NbExp=1; IntAct=EBI-297353, EBI-397403;
  P62158:CALM1; NbExp=1; IntAct=EBI-297353, EBI-397435;
  P62161:Calm1 (xeno); NbExp=2; IntAct=EBI-297353, EBI-397530;
  P62204:Calm1 (xeno); NbExp=1; IntAct=EBI-297353, EBI-397460;
  P22681:CBL; NbExp=1; IntAct=EBI-297353, EBI-518228;
  P22682:Cbl (xeno); NbExp=1; IntAct=EBI-297353, EBI-640919;
  P13987:CD59; NbExp=1; IntAct=EBI-297353, EBI-297972;
  P01133:EGF; NbExp=2; IntAct=EBI-297353, EBI-640857;
  P04626:ERBB2; NbExp=2; IntAct=EBI-297353, EBI-641062;
  P21860:ERBB3; NbExp=2; IntAct=EBI-297353, EBI-720706;
  Q15303:ERBB4; NbExp=2; IntAct=EBI-297353, EBI-80371;
  P62993:GRB2; NbExp=2; IntAct=EBI-297353, EBI-401755;
  O00750:PIK3C2B; NbExp=4; IntAct=EBI-297353, EBI-641107;
  P98083:Shc1 (xeno); NbExp=1; IntAct=EBI-297353, EBI-300201;
  P63104:YWHAZ; NbExp=1; IntAct=EBI-297353, EBI-347088;
- ***SUBCELLULAR LOCATION***: Cell membrane; Single-pass type I membrane protein. Isoform 2: Secreted.
- ***ALTERNATIVE PRODUCTS***: 4 named isoforms [FASTA] produced by alternative splicing.

Name	**1**
Synonyms	p170
Isoform ID	P00533-1
This is the isoform sequence displayed in this entry.	

Name	**2**
Synonyms	p60, Truncated, TEGFR
Isoform ID	P00533-2
Features which should be applied to build the isoform sequence: VSP_002887, VSP_002888.	

Name	**3**
Synonyms	p110
Isoform ID	P00533-3

Продолжение



- ***SUBCELLULAR LOCATION***: Cell membrane; Single-pass type I membrane protein. Isoform 2: Secreted.
- ***ALTERNATIVE PRODUCTS***: 4 named isoforms [FASTA] produced by alternative splicing.

Name	**1**
Synonyms	p170
Isoform ID	P00533-1
This is the isoform sequence displayed in this entry.	

Name	**2**
Synonyms	p60, Truncated, TEGFR
Isoform ID	P00533-2
Features which should be applied to build the isoform sequence: VSP_002887, VSP_002888.	

Name	**3**
Synonyms	p110
Isoform ID	P00533-3
Features which should be applied to build the isoform sequence: VSP_002889, VSP_002890.	

Name	**4**
Isoform ID	P00533-4
Features which should be applied to build the isoform sequence: VSP_002891, VSP_002892.	

- ***TISSUE SPECIFICITY***: Expressed in placenta. Isoform 2 is also expressed in ovarian cancers.
- ***PTM***: Phosphorylation of Ser-695 is partial and occurs only if Thr-693 is phosphorylated.
- ***PTM***: Monoubiquitinated and polyubiquitinated upon EGF stimulation; which does not affect tyrosine kinase activity or signaling capacity but may play a role in lysosomal targeting. Polyubiquitin linkage is mainly through 'Lys-63', but linkage through 'Lys-48', 'Lys-11' and 'Lys-29' also occur.
- ***DISEASE***: Defects in EGFR are associated with lung cancer [MIM:211980].
- ***MISCELLANEOUS***: Binding of EGF to the receptor leads to dimerization, internalization of the EGF-receptor complex, induction of the tyrosine kinase activity, stimulation cell DNA synthesis, and cell proliferation.
- ***SIMILARITY***: Belongs to the protein kinase superfamily. Tyr protein kinase family. EGF receptor subfamily.
- ***SIMILARITY***: Contains 1 protein kinase domain.
- ***WEB RESOURCE***: Name=GeneReviews; URL="http://www.genetests.org/query?gene=EGFR";.
- ***WEB RESOURCE***: Name=Wikipedia; Note=EGFR entry; URL="http://en.wikipedia.org/wiki/Epidermal_growth_factor_receptor";.

Copyright

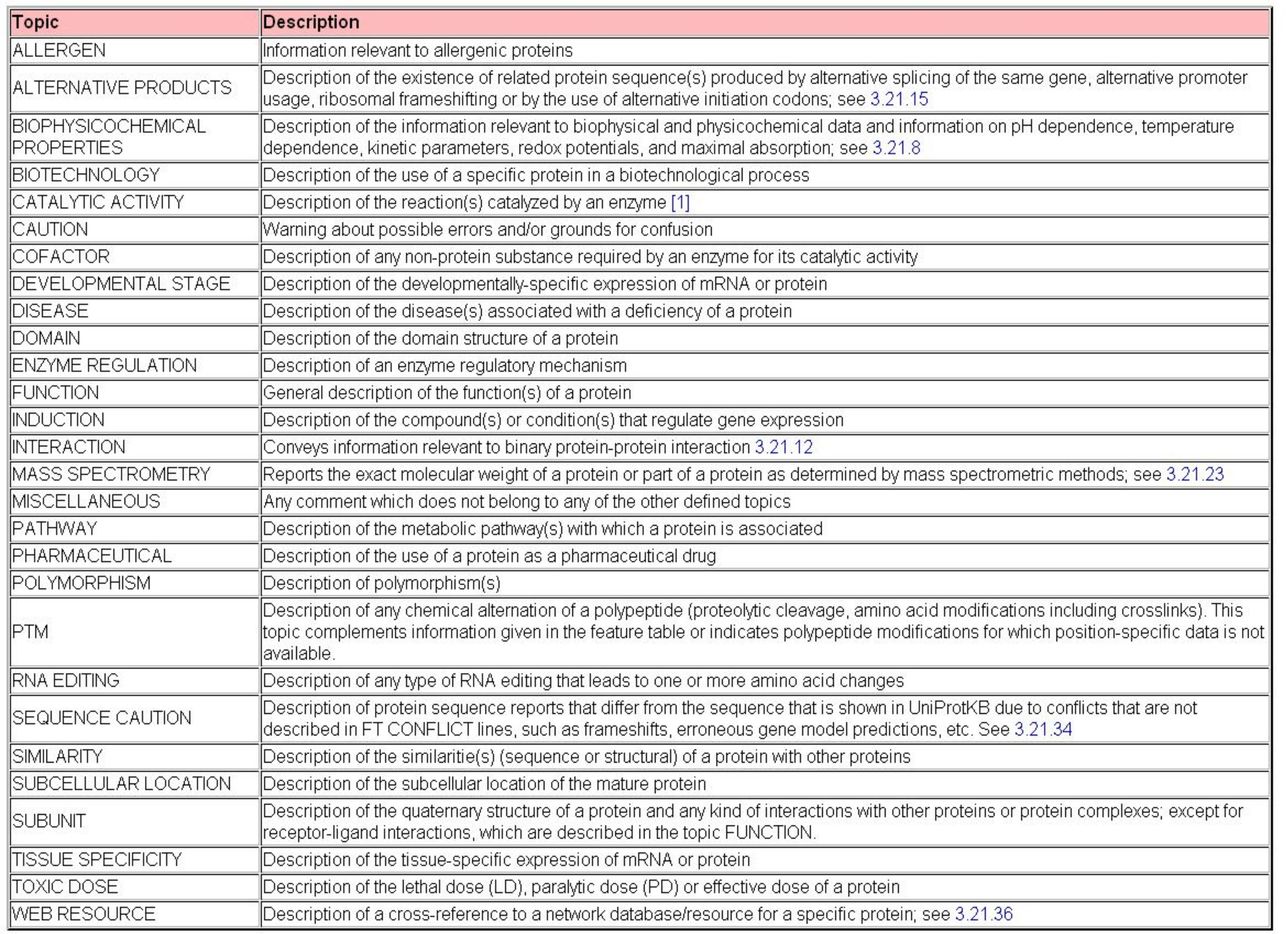
Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms. Distributed under the Creative Commons Attribution-NoDerivs License.

Cross-references

Sequence databases

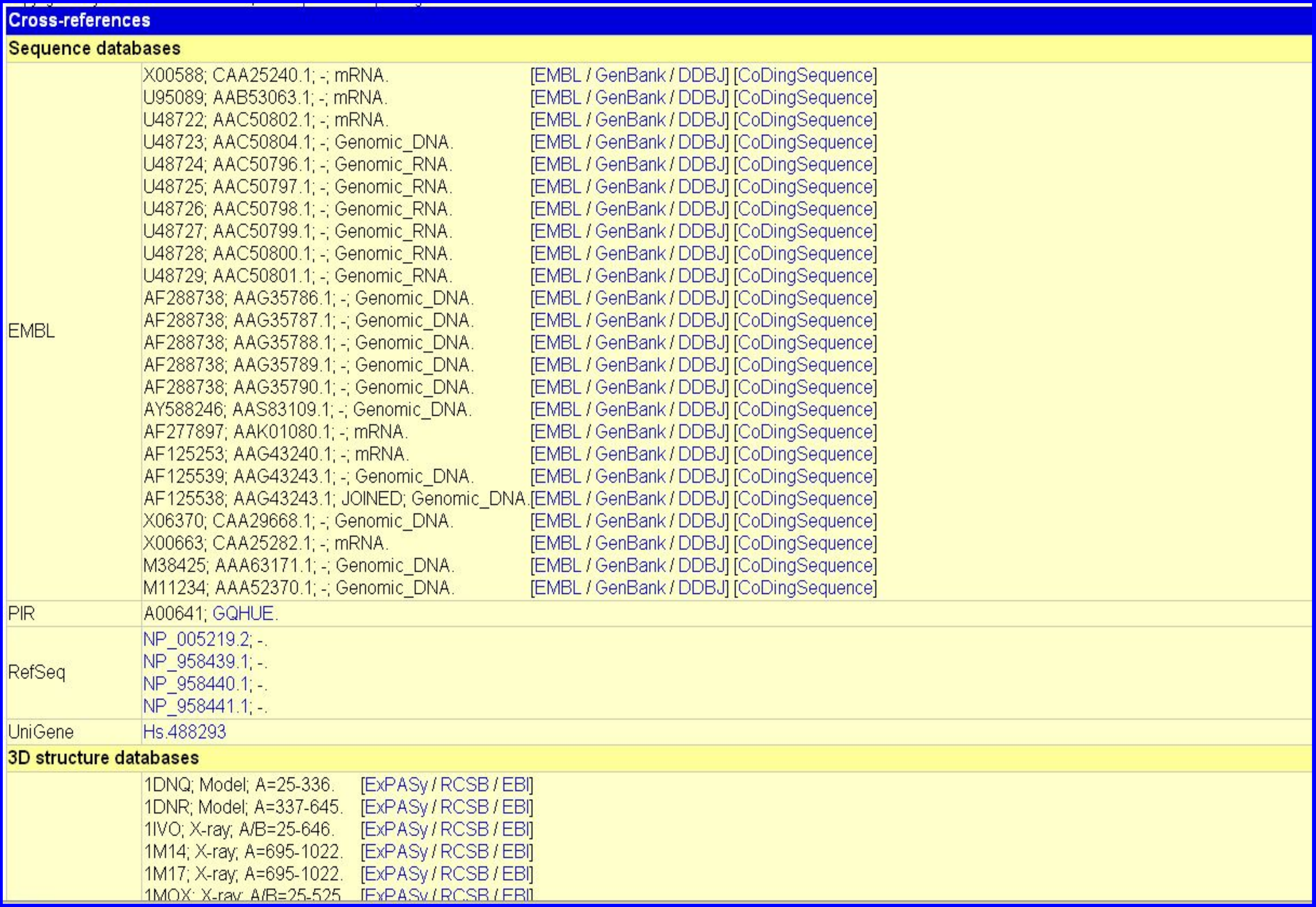
	X00588; CAA25240.1; -; mRNA.	[EMBL / GenBank / DDBJ] [CoDingSequence]
	U95089; AAB53063.1; -; mRNA.	[EMBL / GenBank / DDBJ] [CoDingSequence]

Возможные разделы Комментариев



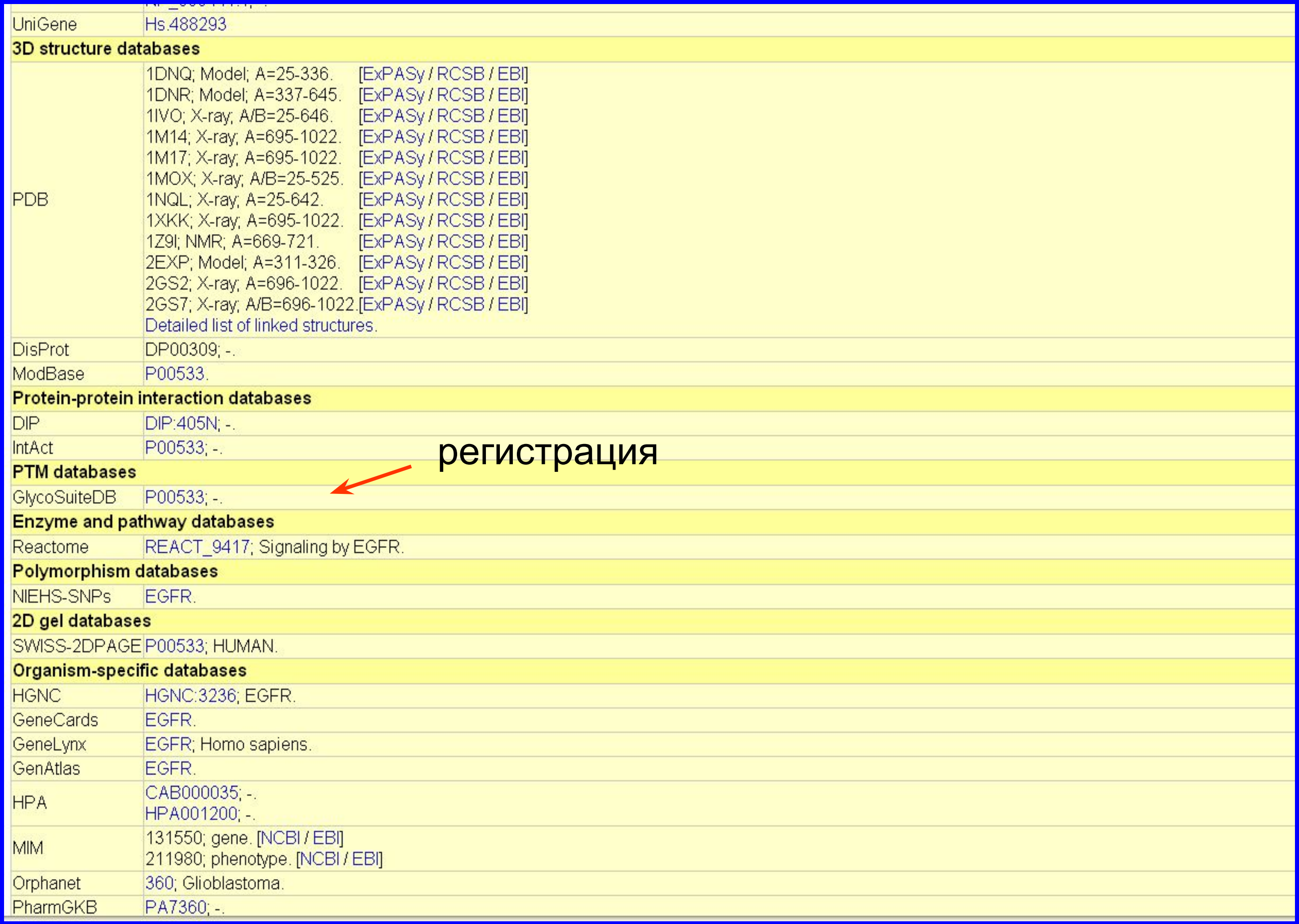
Topic	Description
ALLERGEN	Information relevant to allergenic proteins
ALTERNATIVE PRODUCTS	Description of the existence of related protein sequence(s) produced by alternative splicing of the same gene, alternative promoter usage, ribosomal frameshifting or by the use of alternative initiation codons; see 3.21.15
BIOPHYSICOCHEMICAL PROPERTIES	Description of the information relevant to biophysical and physicochemical data and information on pH dependence, temperature dependence, kinetic parameters, redox potentials, and maximal absorption; see 3.21.8
BIOTECHNOLOGY	Description of the use of a specific protein in a biotechnological process
CATALYTIC ACTIVITY	Description of the reaction(s) catalyzed by an enzyme [1]
CAUTION	Warning about possible errors and/or grounds for confusion
COFACTOR	Description of any non-protein substance required by an enzyme for its catalytic activity
DEVELOPMENTAL STAGE	Description of the developmentally-specific expression of mRNA or protein
DISEASE	Description of the disease(s) associated with a deficiency of a protein
DOMAIN	Description of the domain structure of a protein
ENZYME REGULATION	Description of an enzyme regulatory mechanism
FUNCTION	General description of the function(s) of a protein
INDUCTION	Description of the compound(s) or condition(s) that regulate gene expression
INTERACTION	Conveys information relevant to binary protein-protein interaction 3.21.12
MASS SPECTROMETRY	Reports the exact molecular weight of a protein or part of a protein as determined by mass spectrometric methods; see 3.21.23
MISCELLANEOUS	Any comment which does not belong to any of the other defined topics
PATHWAY	Description of the metabolic pathway(s) with which a protein is associated
PHARMACEUTICAL	Description of the use of a protein as a pharmaceutical drug
POLYMORPHISM	Description of polymorphism(s)
PTM	Description of any chemical alternation of a polypeptide (proteolytic cleavage, amino acid modifications including crosslinks). This topic complements information given in the feature table or indicates polypeptide modifications for which position-specific data is not available.
RNA EDITING	Description of any type of RNA editing that leads to one or more amino acid changes
SEQUENCE CAUTION	Description of protein sequence reports that differ from the sequence that is shown in UniProtKB due to conflicts that are not described in FT CONFLICT lines, such as frameshifts, erroneous gene model predictions, etc. See 3.21.34
SIMILARITY	Description of the similaritie(s) (sequence or structural) of a protein with other proteins
SUBCELLULAR LOCATION	Description of the subcellular location of the mature protein
SUBUNIT	Description of the quaternary structure of a protein and any kind of interactions with other proteins or protein complexes; except for receptor-ligand interactions, which are described in the topic FUNCTION.
TISSUE SPECIFICITY	Description of the tissue-specific expression of mRNA or protein
TOXIC DOSE	Description of the lethal dose (LD), paralytic dose (PD) or effective dose of a protein
WEB RESOURCE	Description of a cross-reference to a network database/resource for a specific protein; see 3.21.36

Cross-References



Cross-references		
Sequence databases		
EMBL	X00588; CAA25240.1; -; mRNA.	[EMBL / GenBank / DDBJ] [CoDingSequence]
	U95089; AAB53063.1; -; mRNA.	[EMBL / GenBank / DDBJ] [CoDingSequence]
	U48722; AAC50802.1; -; mRNA.	[EMBL / GenBank / DDBJ] [CoDingSequence]
	U48723; AAC50804.1; -; Genomic_DNA.	[EMBL / GenBank / DDBJ] [CoDingSequence]
	U48724; AAC50796.1; -; Genomic_RNA.	[EMBL / GenBank / DDBJ] [CoDingSequence]
	U48725; AAC50797.1; -; Genomic_RNA.	[EMBL / GenBank / DDBJ] [CoDingSequence]
	U48726; AAC50798.1; -; Genomic_RNA.	[EMBL / GenBank / DDBJ] [CoDingSequence]
	U48727; AAC50799.1; -; Genomic_RNA.	[EMBL / GenBank / DDBJ] [CoDingSequence]
	U48728; AAC50800.1; -; Genomic_RNA.	[EMBL / GenBank / DDBJ] [CoDingSequence]
	U48729; AAC50801.1; -; Genomic_RNA.	[EMBL / GenBank / DDBJ] [CoDingSequence]
	AF288738; AAG35786.1; -; Genomic_DNA.	[EMBL / GenBank / DDBJ] [CoDingSequence]
	AF288738; AAG35787.1; -; Genomic_DNA.	[EMBL / GenBank / DDBJ] [CoDingSequence]
	AF288738; AAG35788.1; -; Genomic_DNA.	[EMBL / GenBank / DDBJ] [CoDingSequence]
	AF288738; AAG35789.1; -; Genomic_DNA.	[EMBL / GenBank / DDBJ] [CoDingSequence]
	AF288738; AAG35790.1; -; Genomic_DNA.	[EMBL / GenBank / DDBJ] [CoDingSequence]
	AY588246; AAS83109.1; -; Genomic_DNA.	[EMBL / GenBank / DDBJ] [CoDingSequence]
	AF277897; AAK01080.1; -; mRNA.	[EMBL / GenBank / DDBJ] [CoDingSequence]
	AF125253; AAG43240.1; -; mRNA.	[EMBL / GenBank / DDBJ] [CoDingSequence]
	AF125539; AAG43243.1; -; Genomic_DNA.	[EMBL / GenBank / DDBJ] [CoDingSequence]
	AF125538; AAG43243.1; JOINED; Genomic_DNA.	[EMBL / GenBank / DDBJ] [CoDingSequence]
	X06370; CAA29668.1; -; Genomic_DNA.	[EMBL / GenBank / DDBJ] [CoDingSequence]
	X00663; CAA25282.1; -; mRNA.	[EMBL / GenBank / DDBJ] [CoDingSequence]
	M38425; AAA63171.1; -; Genomic_DNA.	[EMBL / GenBank / DDBJ] [CoDingSequence]
	M11234; AAA52370.1; -; Genomic_DNA.	[EMBL / GenBank / DDBJ] [CoDingSequence]
PIR	A00641; GQHUE.	
RefSeq	NP_005219.2; -. NP_958439.1; -. NP_958440.1; -. NP_958441.1; -.	
UniGene	Hs.488293	
3D structure databases		
	1DNQ; Model; A=25-336.	[ExPASy / RCSB / EBI]
	1DNR; Model; A=337-645.	[ExPASy / RCSB / EBI]
	1IVO; X-ray; A/B=25-646.	[ExPASy / RCSB / EBI]
	1M14; X-ray; A=695-1022.	[ExPASy / RCSB / EBI]
	1M17; X-ray; A=695-1022.	[ExPASy / RCSB / EBI]
	1MOX; X-ray; A/B=25-525.	[ExPASy / RCSB / EBI]

Cross-References



UniGene	Hs.488293
3D structure databases	
PDB	1DNQ; Model; A=25-336. [ExPASy / RCSB / EBI] 1DNR; Model; A=337-645. [ExPASy / RCSB / EBI] 1IVO; X-ray; A/B=25-646. [ExPASy / RCSB / EBI] 1M14; X-ray; A=695-1022. [ExPASy / RCSB / EBI] 1M17; X-ray; A=695-1022. [ExPASy / RCSB / EBI] 1MOX; X-ray; A/B=25-525. [ExPASy / RCSB / EBI] 1NQL; X-ray; A=25-642. [ExPASy / RCSB / EBI] 1XKK; X-ray; A=695-1022. [ExPASy / RCSB / EBI] 1Z9I; NMR; A=669-721. [ExPASy / RCSB / EBI] 2EXP; Model; A=311-326. [ExPASy / RCSB / EBI] 2GS2; X-ray; A=696-1022. [ExPASy / RCSB / EBI] 2GS7; X-ray; A/B=696-1022.[ExPASy / RCSB / EBI] Detailed list of linked structures.
DisProt	DP00309; -.
ModBase	P00533.
Protein-protein interaction databases	
DIP	DIP:405N; -.
IntAct	P00533; -.
PTM databases	
GlycoSuiteDB	P00533; -.
Enzyme and pathway databases	
Reactome	REACT_9417; Signaling by EGFR.
Polymorphism databases	
NIEHS-SNPs	EGFR.
2D gel databases	
SWISS-2DPAGE	P00533; HUMAN.
Organism-specific databases	
HGNC	HGNC:3236; EGFR.
GeneCards	EGFR.
GeneLynx	EGFR; Homo sapiens.
GenAtlas	EGFR.
HPA	CAB000035; -. HPA001200; -.
MIM	131550; gene. [NCBI / EBI] 211980; phenotype. [NCBI / EBI]
Orphanet	360; Glioblastoma.
PharmGKB	PA7360; -.

регистрация

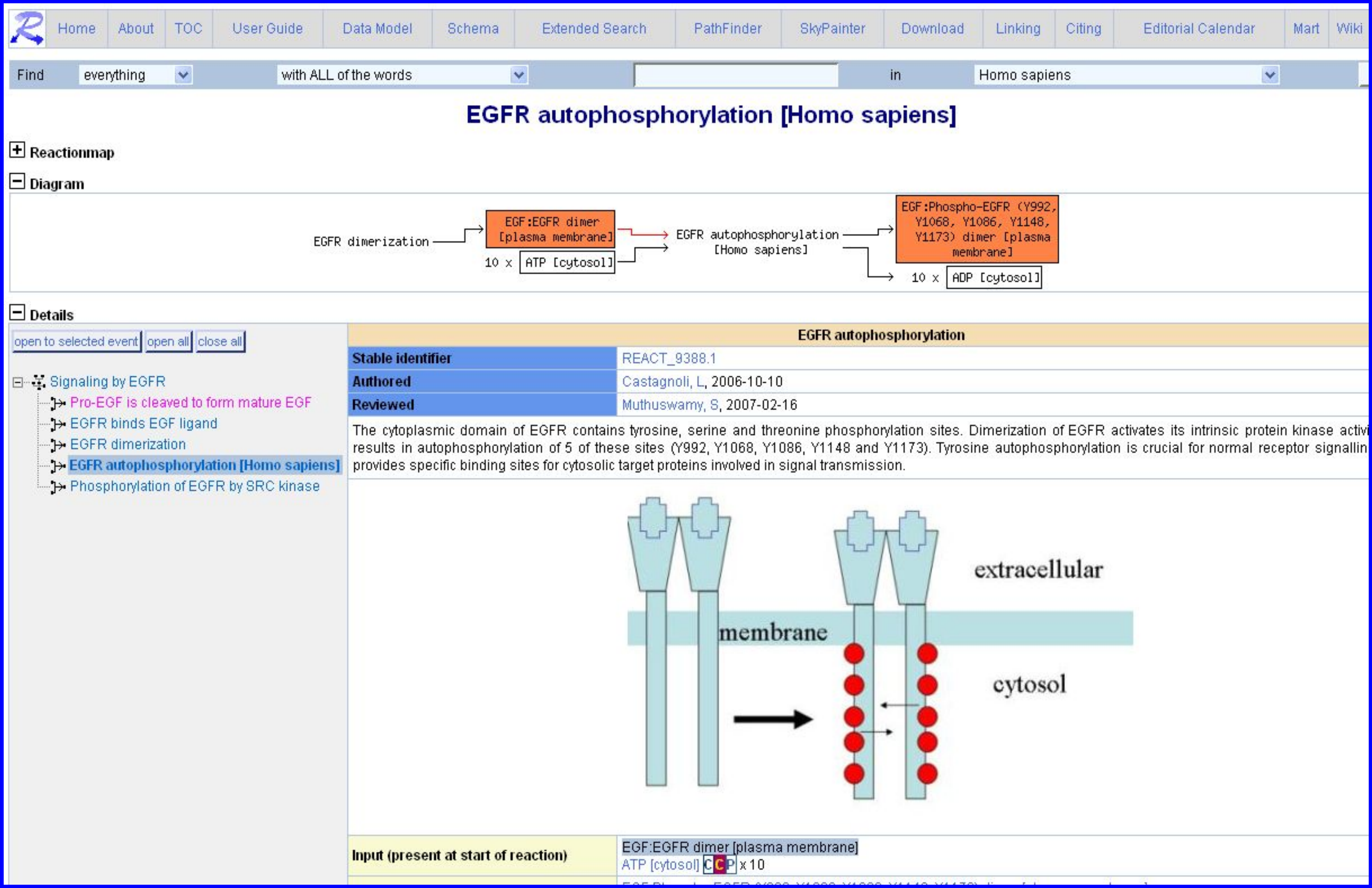
3D-Structure (список структур)

PDB structures for UniProtKB entry P00533 (EGFR_HUMAN)

Epidermal growth factor receptor precursor (EC 2.7.10.1) (Receptor tyrosine-protein kinase ErbB-1).

PDB	Swiss-Prot entries	Compound name *(Submission date)*
1DNQ	EGFR_HUMAN (P00533)	MOL_ID: 1; MOLECULE: EPIDERMAL GROWTH FACTOR RECEPTOR; CHAIN: A; FRAGMENT: L1 AND S1 DOMAINS; EC: 2.7.1.1 *(16-DEC-99)*
1DNR	EGFR_HUMAN (P00533)	MOL_ID: 1; MOLECULE: EPIDERMAL GROWTH FACTOR RECEPTOR; CHAIN: A; FRAGMENT: L2 AND S2 DOMAINS; EC: 2.7.1.1 *(16-DEC-99)*
1IVO	EGFR_HUMAN (P00533) EGF_HUMAN (P01133)	MOL_ID: 1; MOLECULE: EPIDERMAL GROWTH FACTOR RECEPTOR; CHAIN: A, B; FRAGMENT: EXTRACELLULAR DOMAINS I, II AND IV; EC: 2.7.1.112; ENGINEERED: YES; MOL_ID: 2; MOLECULE: EPIDERMAL GROWTH FACTOR; CHAIN: C, D; FRAGMENT RESIDUES 1-53; SYNONYM: EGF; ENGINEERED: YES [3.30 Å] *(28-MAR-02)*
1M14	EGFR_HUMAN (P00533)	MOL_ID: 1; MOLECULE: EPIDERMAL GROWTH FACTOR RECEPTOR; CHAIN: A; FRAGMENT: TYROSINE KINASE DOMAIN (RESIDUES 671-998); SYNONYM: RECEPTOR PROTEIN-TYROSINE KINASE ERBB-1; EC: 2.7.1.112; ENGINEERED: YES; MUTATION: YES [2.60 Å] *(17-JUN-02)*
1M17	EGFR_HUMAN (P00533)	MOL_ID: 1; MOLECULE: EPIDERMAL GROWTH FACTOR RECEPTOR; CHAIN: A; FRAGMENT: TYROSINE KINASE DOMAIN (RESIDUES 671-998); SYNONYM: RECEPTOR PROTEIN-TYROSINE KINASE ERBB-1; EC: 2.7.1.112; ENGINEERED: YES [2.60 Å] *(17-JUN-02)*
1MOX	EGFR_HUMAN (P00533) TGFA_HUMAN (P01135)	MOL_ID: 1; MOLECULE: EPIDERMAL GROWTH FACTOR RECEPTOR; CHAIN: A, B; FRAGMENT: EXTRACELLULAR FRAGMENT EC: 2.7.1.112; ENGINEERED: YES; MOL_ID: 2; MOLECULE: TRANSFORMING GROWTH FACTOR ALPHA; CHAIN: C, D; ENGINEERED: YES [2.50 Å] *(10-SEP-02)*
1NQL	EGFR_HUMAN (P00533) EGF_HUMAN (P01133)	MOL_ID: 1; MOLECULE: EPIDERMAL GROWTH FACTOR RECEPTOR; CHAIN: A; FRAGMENT: EXTRACELLULAR DOMAIN; ENGINEERED: YES; MOL_ID: 2; MOLECULE: EPIDERMAL GROWTH FACTOR; CHAIN: B; SYNONYM: EGF, [CONTAINS: EPIDERMAL GROWTH FACTOR UROGASTRONE,]; ENGINEERED: YES [2.80 Å] *(21-JAN-03)*
1XKK	EGFR_HUMAN (P00533)	MOL_ID: 1; MOLECULE: EPIDERMAL GROWTH FACTOR RECEPTOR; CHAIN: A; FRAGMENT: EGFR KINASE DOMAIN; SYNONY RECEPTOR PROTEIN-TYROSINE KINASE ERBB-1; EC: 2.7.1.112; ENGINEERED: YES [2.40 Å] *(29-SEP-04)*
1Z9I	EGFR_HUMAN (P00533)	MOL_ID: 1; MOLECULE: EPIDERMAL GROWTH FACTOR RECEPTOR; CHAIN: A; FRAGMENT: SEQUENCE DATABASE RESIDUES 669-721; SYNONYM: RECEPTOR TYROSINE-PROTEIN KINASE ERBB-1; EC: 2.7.1.112; ENGINEERED: YES [NMR, 25 STRUCTURES] *(02-APR-05)*
2EXP	EGFR_HUMAN (P00533) HVM61_MOUSE (P18532)	MOL_ID: 1; MOLECULE: LIGHT CHAIN FROM FV FRAGMENT OF MONOCLONAL ANTIBODY 806; CHAIN: L; MOL_ID: 2; MOLECULE: HEAVY CHAIN FROM FV FRAGMENT OF MONOCLONAL ANTIBODY 806; CHAIN: H; MOL_ID: 3; MOLECULE: EPIDERMAL GROWTH FACTOR RECEPTOR; CHAIN: A; SYNONYM: RECEPTOR TYROSINE-PROTEIN KINASE ERBB-1; EC: 2.7.1.112 *(08-NOV-05)*
2GS2	EGFR_HUMAN (P00533)	MOL_ID: 1; MOLECULE: EPIDERMAL GROWTH FACTOR RECEPTOR; CHAIN: A; FRAGMENT: KINASE DOMAIN; SYNONYM:

Reactome



EGFR autophosphorylation [Homo sapiens]

⊞ **Reactionmap**

⊟ **Diagram**

EGFR dimerization → EGF:EGFR dimer [plasma membrane] → EGFR autophosphorylation [Homo sapiens] → EGF:Phospho-EGFR (Y992, Y1068, Y1086, Y1148, Y1173) dimer [plasma membrane]

10 x ATP [cytosol] → EGFR autophosphorylation [Homo sapiens] → 10 x ADP [cytosol]

⊟ **Details**

- Signaling by EGFR
  - Pro-EGF is cleaved to form mature EGF
  - EGFR binds EGF ligand
  - EGFR dimerization
  - **EGFR autophosphorylation [Homo sapiens]**
  - Phosphorylation of EGFR by SRC kinase

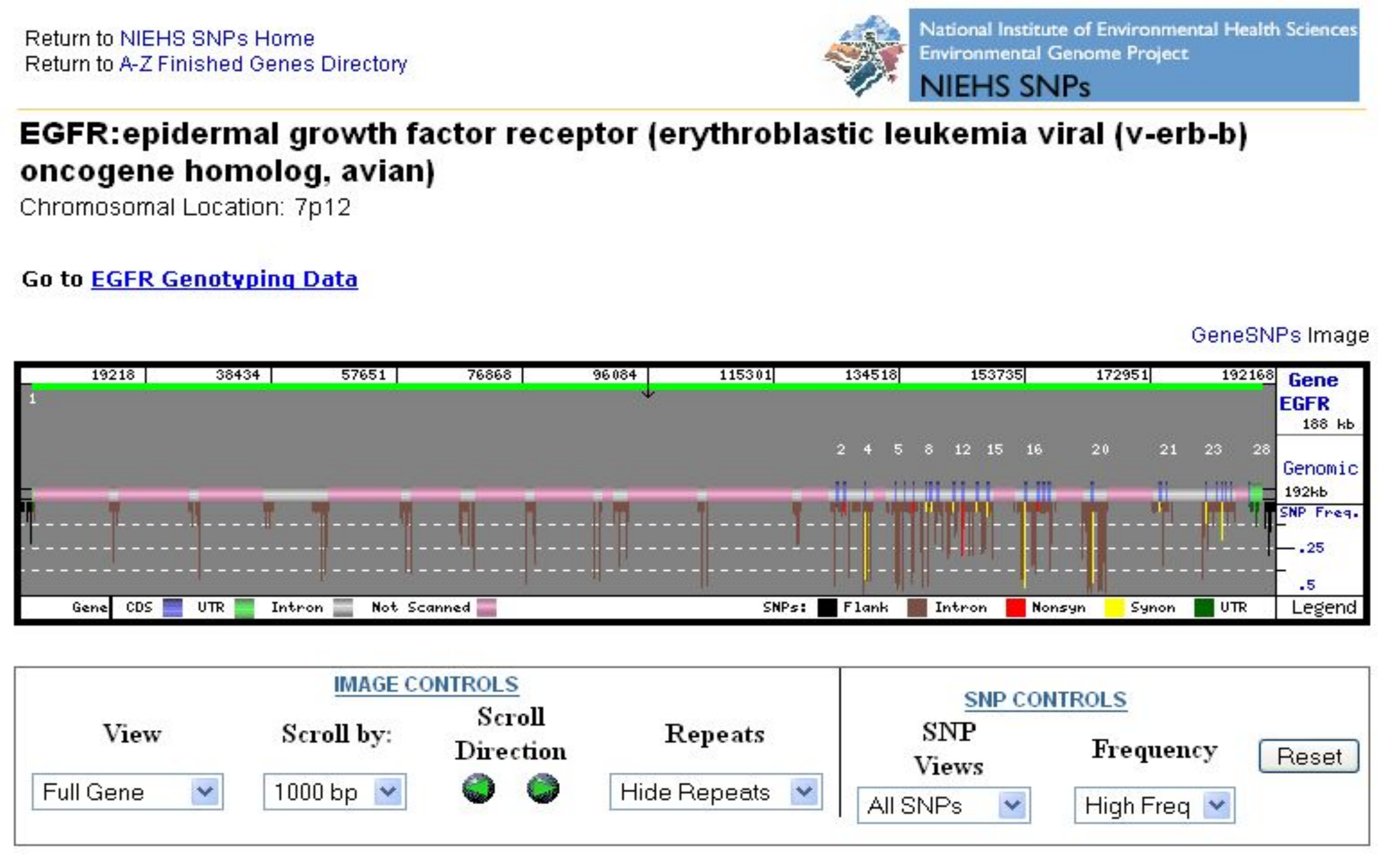
EGFR autophosphorylation

Stable identifier	REACT_9388.1
Authored	Castagnoli, L, 2006-10-10
Reviewed	Muthuswamy, S, 2007-02-16

The cytoplasmic domain of EGFR contains tyrosine, serine and threonine phosphorylation sites. Dimerization of EGFR activates its intrinsic protein kinase activi results in autophosphorylation of 5 of these sites (Y992, Y1068, Y1086, Y1148 and Y1173). Tyrosine autophosphorylation is crucial for normal receptor signallin provides specific binding sites for cytosolic target proteins involved in signal transmission.

Input (present at start of reaction)	EGF:EGFR dimer [plasma membrane] ATP [cytosol] C C P x 10

Полиморфизмы



Return to NIEHS SNPs Home
Return to A-Z Finished Genes Directory

National Institute of Environmental Health Sciences
Environmental Genome Project
NIEHS SNPs

EGFR:epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)

Chromosomal Location: 7p12

Go to EGFR Genotyping Data

GeneSNPs Image

IMAGE CONTROLS

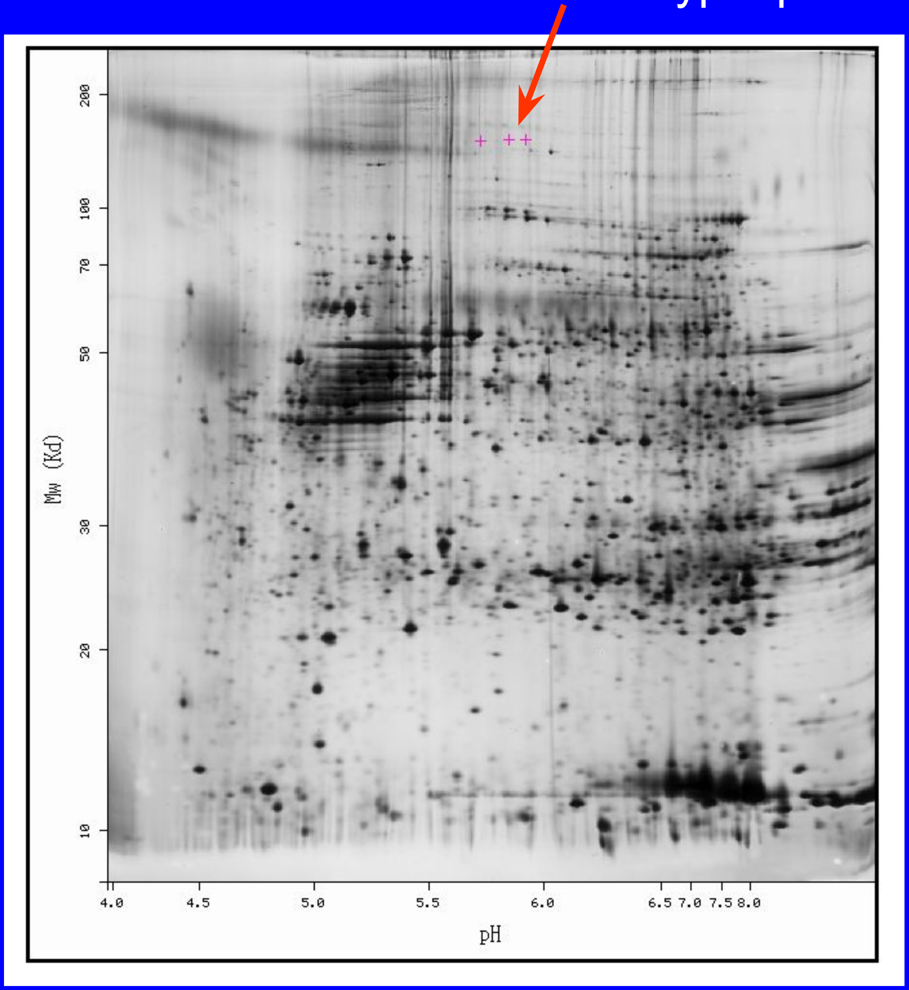
View: Full Gene
Scroll by: 1000 bp
Scroll Direction
Repeats: Hide Repeats

SNP CONTROLS

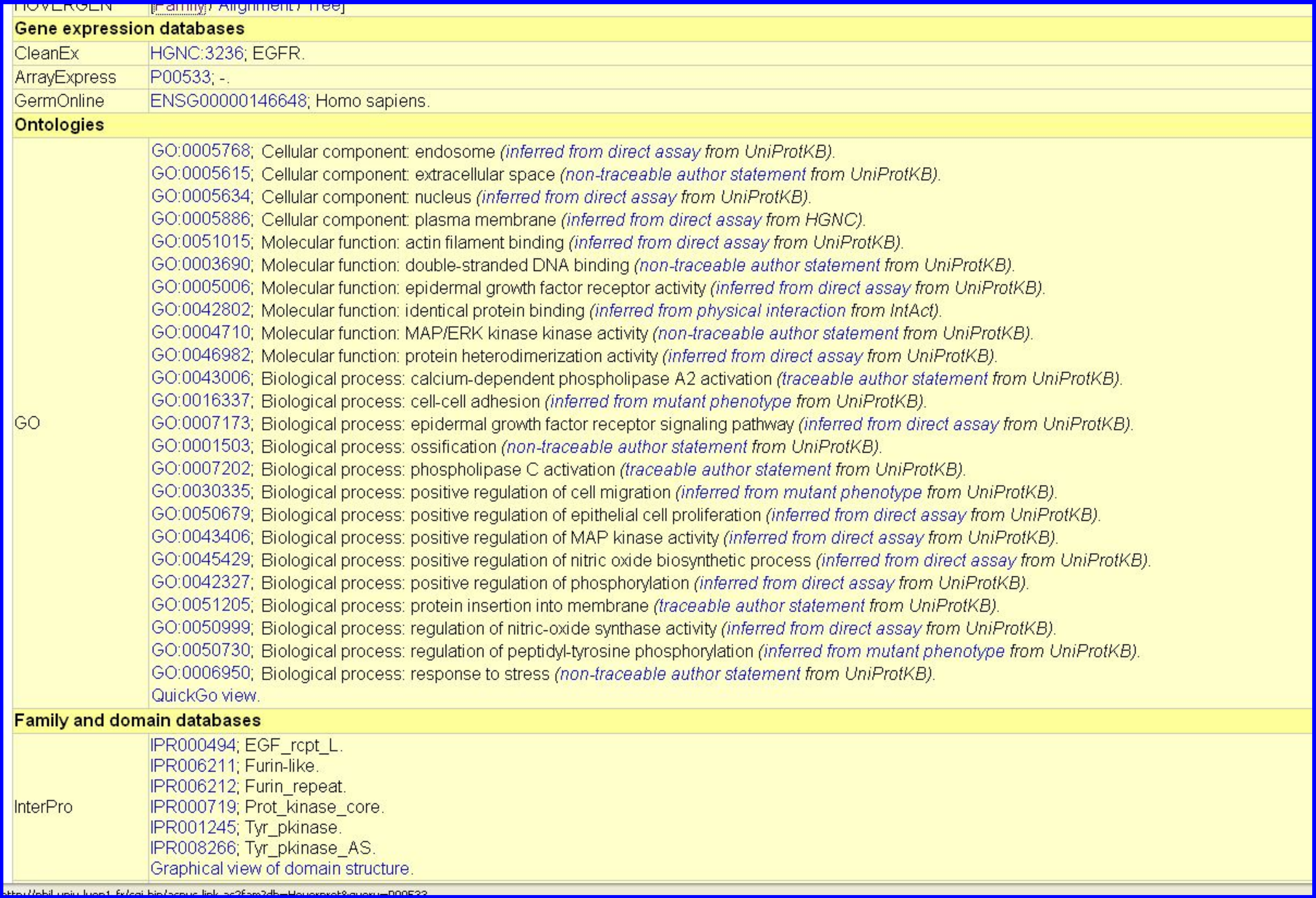
SNP Views: All SNPs
Frequency: High Freq

2D GEL

+ курсор



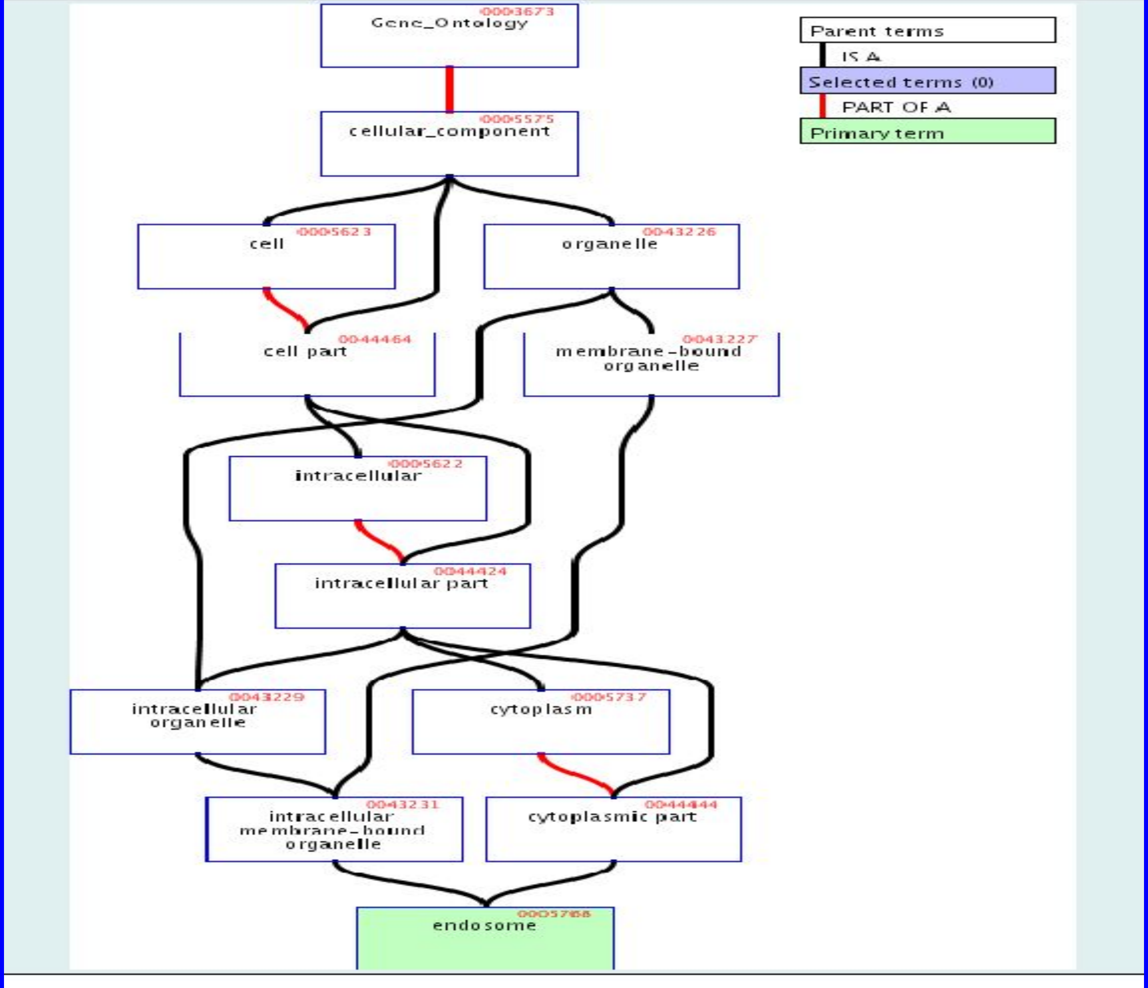
Cross-References



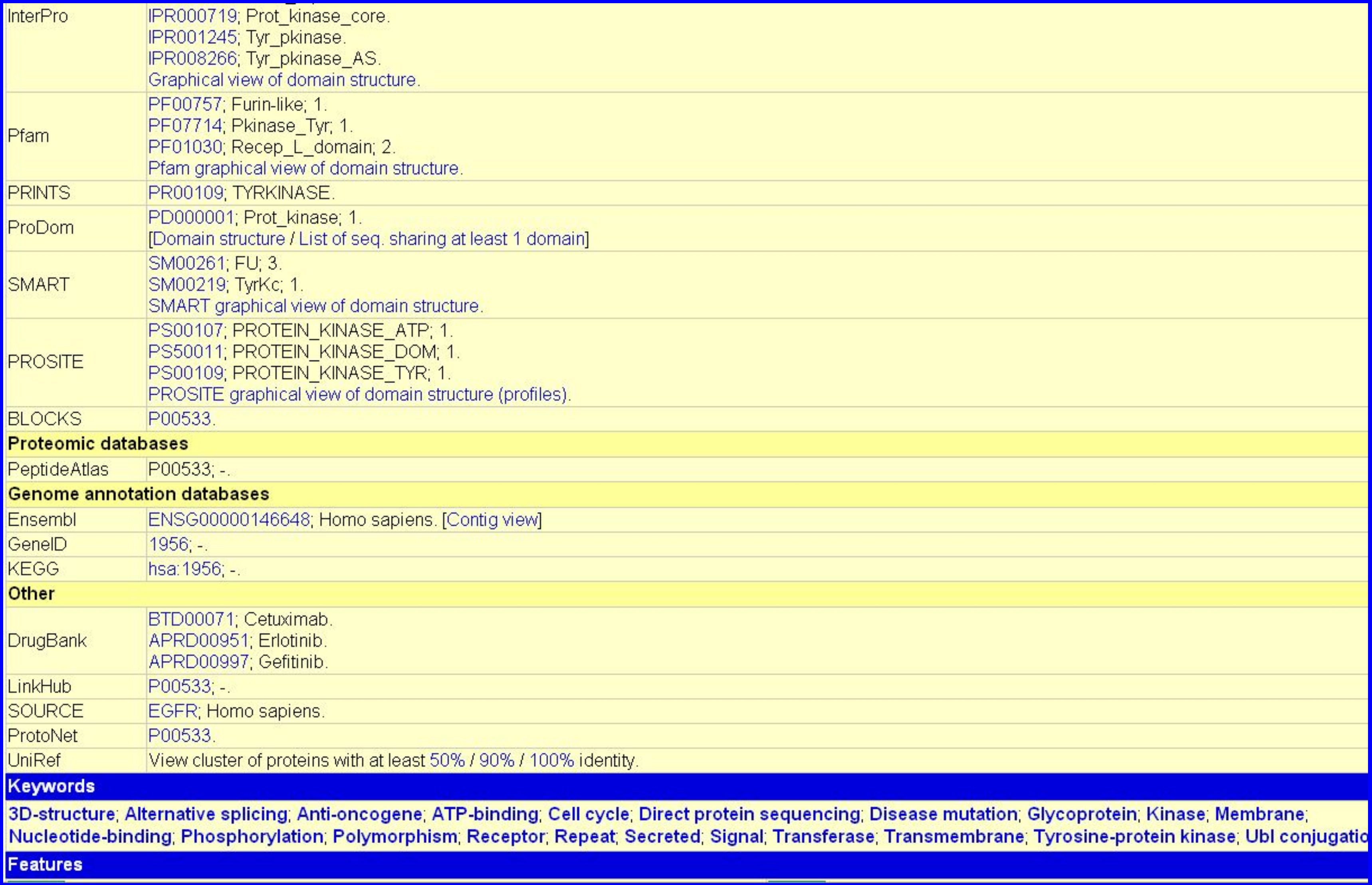
HOVERGEN	[Family / Alignment / Tree]
Gene expression databases	
CleanEx	HGNC:3236; EGFR.
ArrayExpress	P00533; -.
GermOnline	ENSG00000146648; Homo sapiens.
Ontologies	
GO	GO:0005768; Cellular component: endosome *(inferred from direct assay from UniProtKB).* GO:0005615; Cellular component: extracellular space *(non-traceable author statement from UniProtKB).* GO:0005634; Cellular component: nucleus *(inferred from direct assay from UniProtKB).* GO:0005886; Cellular component: plasma membrane *(inferred from direct assay from HGNC).* GO:0051015; Molecular function: actin filament binding *(inferred from direct assay from UniProtKB).* GO:0003690; Molecular function: double-stranded DNA binding *(non-traceable author statement from UniProtKB).* GO:0005006; Molecular function: epidermal growth factor receptor activity *(inferred from direct assay from UniProtKB).* GO:0042802; Molecular function: identical protein binding *(inferred from physical interaction from IntAct).* GO:0004710; Molecular function: MAP/ERK kinase kinase activity *(non-traceable author statement from UniProtKB).* GO:0046982; Molecular function: protein heterodimerization activity *(inferred from direct assay from UniProtKB).* GO:0043006; Biological process: calcium-dependent phospholipase A2 activation *(traceable author statement from UniProtKB).* GO:0016337; Biological process: cell-cell adhesion *(inferred from mutant phenotype from UniProtKB).* GO:0007173; Biological process: epidermal growth factor receptor signaling pathway *(inferred from direct assay from UniProtKB).* GO:0001503; Biological process: ossification *(non-traceable author statement from UniProtKB).* GO:0007202; Biological process: phospholipase C activation *(traceable author statement from UniProtKB).* GO:0030335; Biological process: positive regulation of cell migration *(inferred from mutant phenotype from UniProtKB).* GO:0050679; Biological process: positive regulation of epithelial cell proliferation *(inferred from direct assay from UniProtKB).* GO:0043406; Biological process: positive regulation of MAP kinase activity *(inferred from direct assay from UniProtKB).* GO:0045429; Biological process: positive regulation of nitric oxide biosynthetic process *(inferred from direct assay from UniProtKB).* GO:0042327; Biological process: positive regulation of phosphorylation *(inferred from direct assay from UniProtKB).* GO:0051205; Biological process: protein insertion into membrane *(traceable author statement from UniProtKB).* GO:0050999; Biological process: regulation of nitric-oxide synthase activity *(inferred from direct assay from UniProtKB).* GO:0050730; Biological process: regulation of peptidyl-tyrosine phosphorylation *(inferred from mutant phenotype from UniProtKB).* GO:0006950; Biological process: response to stress *(non-traceable author statement from UniProtKB).* QuickGo view.
Family and domain databases	
InterPro	IPR000494; EGF_rcpt_L. IPR006211; Furin-like. IPR006212; Furin_repeat. IPR000719; Prot_kinase_core. IPR001245; Tyr_pkinase. IPR008266; Tyr_pkinase_AS. Graphical view of domain structure.

GO terms

Определение термина, синонимы, родительские (Hierarchy) и дочерние термины, ключевые слова, дата последней модификации



Cross-References, Keywords



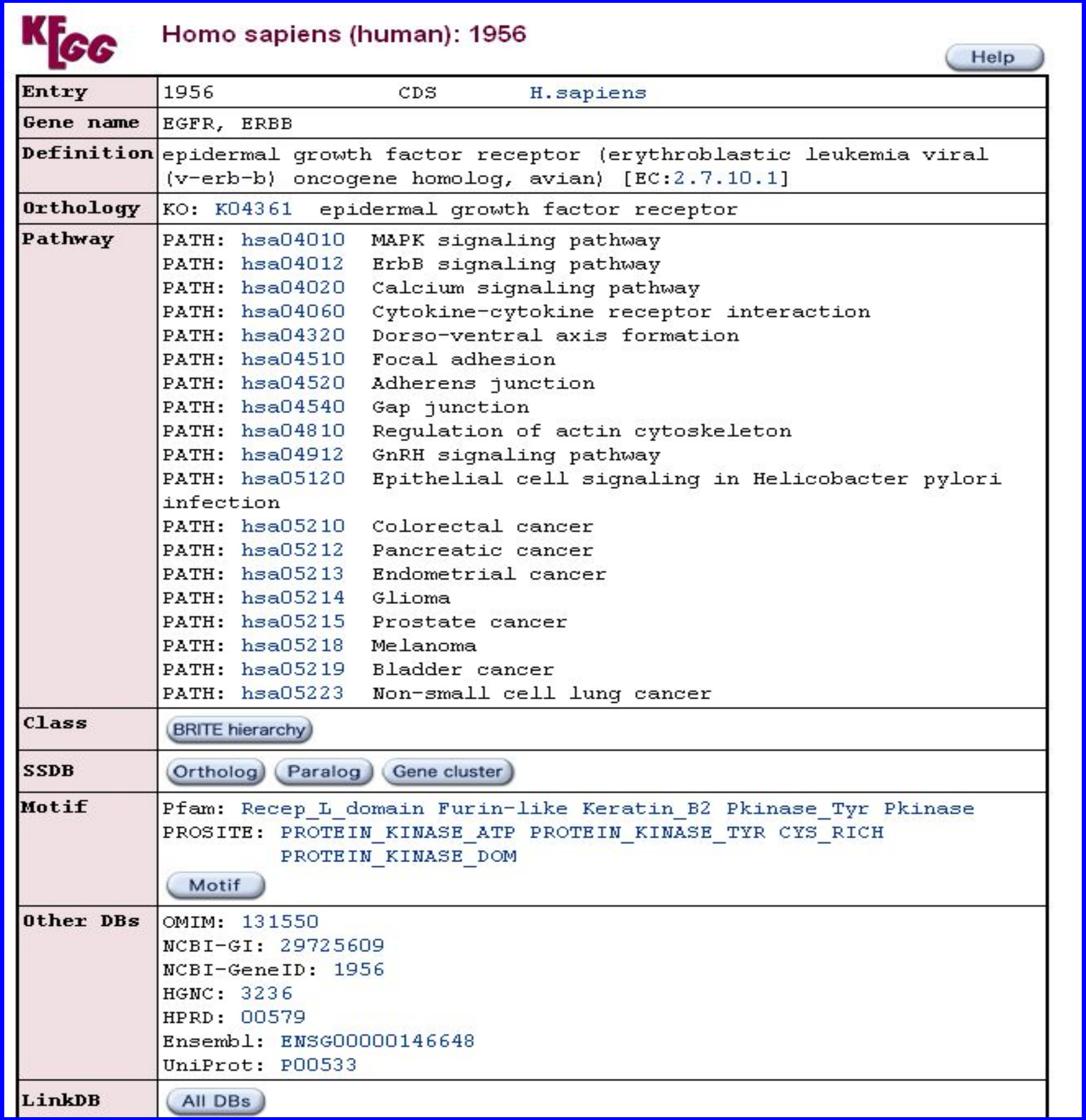
InterPro	IPR000719; Prot_kinase_core. IPR001245; Tyr_pkinase. IPR008266; Tyr_pkinase_AS. Graphical view of domain structure.
Pfam	PF00757; Furin-like; 1. PF07714; Pkinase_Tyr; 1. PF01030; Recep_L_domain; 2. Pfam graphical view of domain structure.
PRINTS	PR00109; TYRKINASE.
ProDom	PD000001; Prot_kinase; 1. [Domain structure / List of seq. sharing at least 1 domain]
SMART	SM00261; FU; 3. SM00219; TyrKc; 1. SMART graphical view of domain structure.
PROSITE	PS00107; PROTEIN_KINASE_ATP; 1. PS50011; PROTEIN_KINASE_DOM; 1. PS00109; PROTEIN_KINASE_TYR; 1. PROSITE graphical view of domain structure (profiles).
BLOCKS	P00533.
Proteomic databases	
PeptideAtlas	P00533; -.
Genome annotation databases	
Ensembl	ENSG00000146648; Homo sapiens. [Contig view]
GeneID	1956; -.
KEGG	hsa:1956; -.
Other	
DrugBank	BTD00071; Cetuximab. APRD00951; Erlotinib. APRD00997; Gefitinib.
LinkHub	P00533; -.
SOURCE	EGFR; Homo sapiens.
ProtoNet	P00533.
UniRef	View cluster of proteins with at least 50% / 90% / 100% identity.

Keywords

3D-structure; Alternative splicing; Anti-oncogene; ATP-binding; Cell cycle; Direct protein sequencing; Disease mutation; Glycoprotein; Kinase; Membrane; Nucleotide-binding; Phosphorylation; Polymorphism; Receptor; Repeat; Secreted; Signal; Transferase; Transmembrane; Tyrosine-protein kinase; Ubl conjugatio

Features

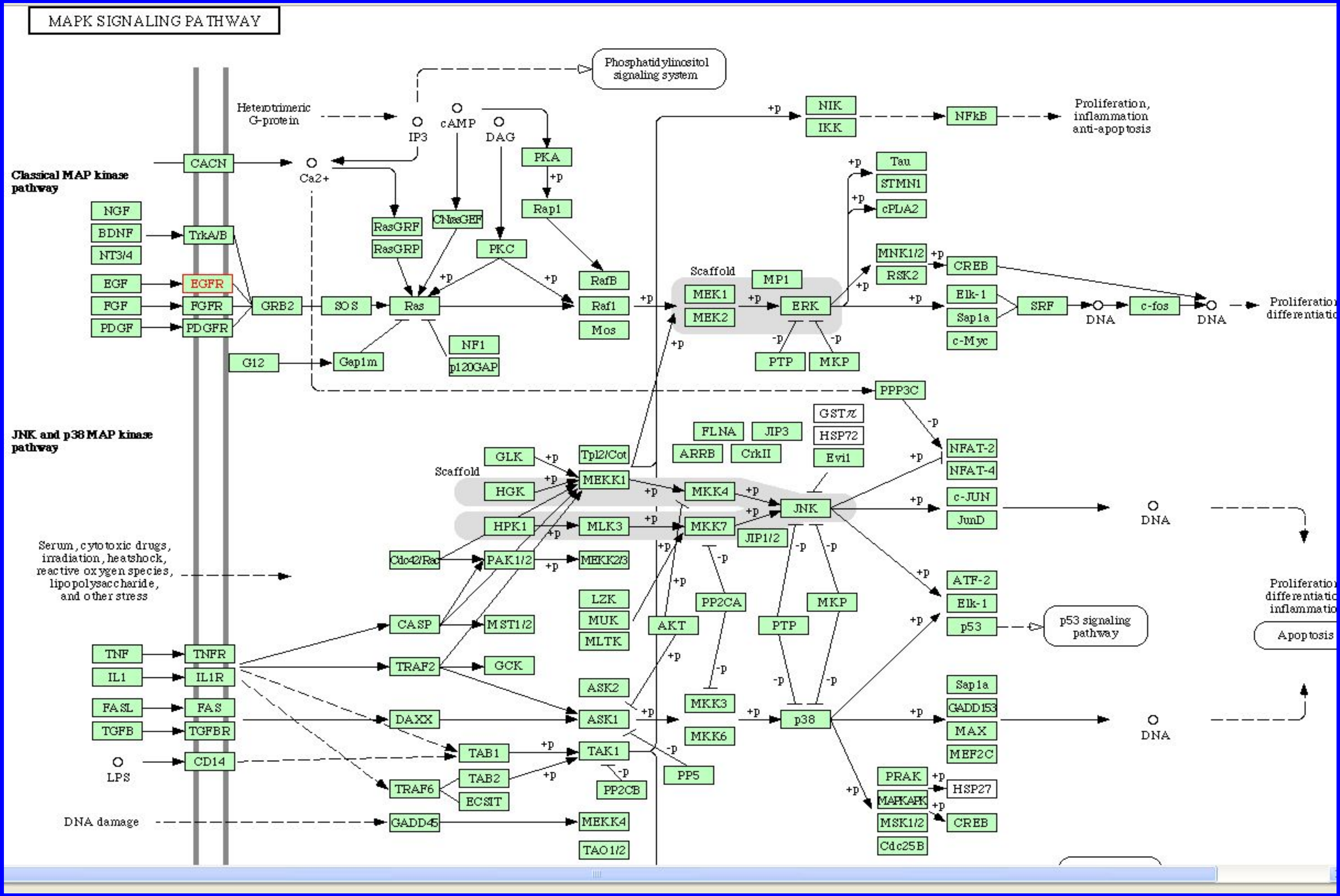
KEGG



Homo sapiens (human): 1956

Entry	1956 CDS H.sapiens
Gene name	EGFR, ERBB
Definition	epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian) [EC:2.7.10.1]
Orthology	KO: K04361 epidermal growth factor receptor
Pathway	PATH: hsa04010 MAPK signaling pathway PATH: hsa04012 ErbB signaling pathway PATH: hsa04020 Calcium signaling pathway PATH: hsa04060 Cytokine-cytokine receptor interaction PATH: hsa04320 Dorso-ventral axis formation PATH: hsa04510 Focal adhesion PATH: hsa04520 Adherens junction PATH: hsa04540 Gap junction PATH: hsa04810 Regulation of actin cytoskeleton PATH: hsa04912 GnRH signaling pathway PATH: hsa05120 Epithelial cell signaling in Helicobacter pylori infection PATH: hsa05210 Colorectal cancer PATH: hsa05212 Pancreatic cancer PATH: hsa05213 Endometrial cancer PATH: hsa05214 Glioma PATH: hsa05215 Prostate cancer PATH: hsa05218 Melanoma PATH: hsa05219 Bladder cancer PATH: hsa05223 Non-small cell lung cancer
Class	BRITE hierarchy
SSDB	Ortholog Paralog Gene cluster
Motif	Pfam: Recep_L_domain Furin-like Keratin_B2 Pkinase_Tyr Pkinase PROSITE: PROTEIN_KINASE_ATP PROTEIN_KINASE_TYR CYS_RICH PROTEIN_KINASE_DOM Motif
Other DBs	OMIM: 131550 NCBI-GI: 29725609 NCBI-GeneID: 1956 HGNC: 3236 HPRD: 00579 Ensembl: ENSG00000146648 UniProt: P00533
LinkDB	All DBs

KEGG (pathway viewer)

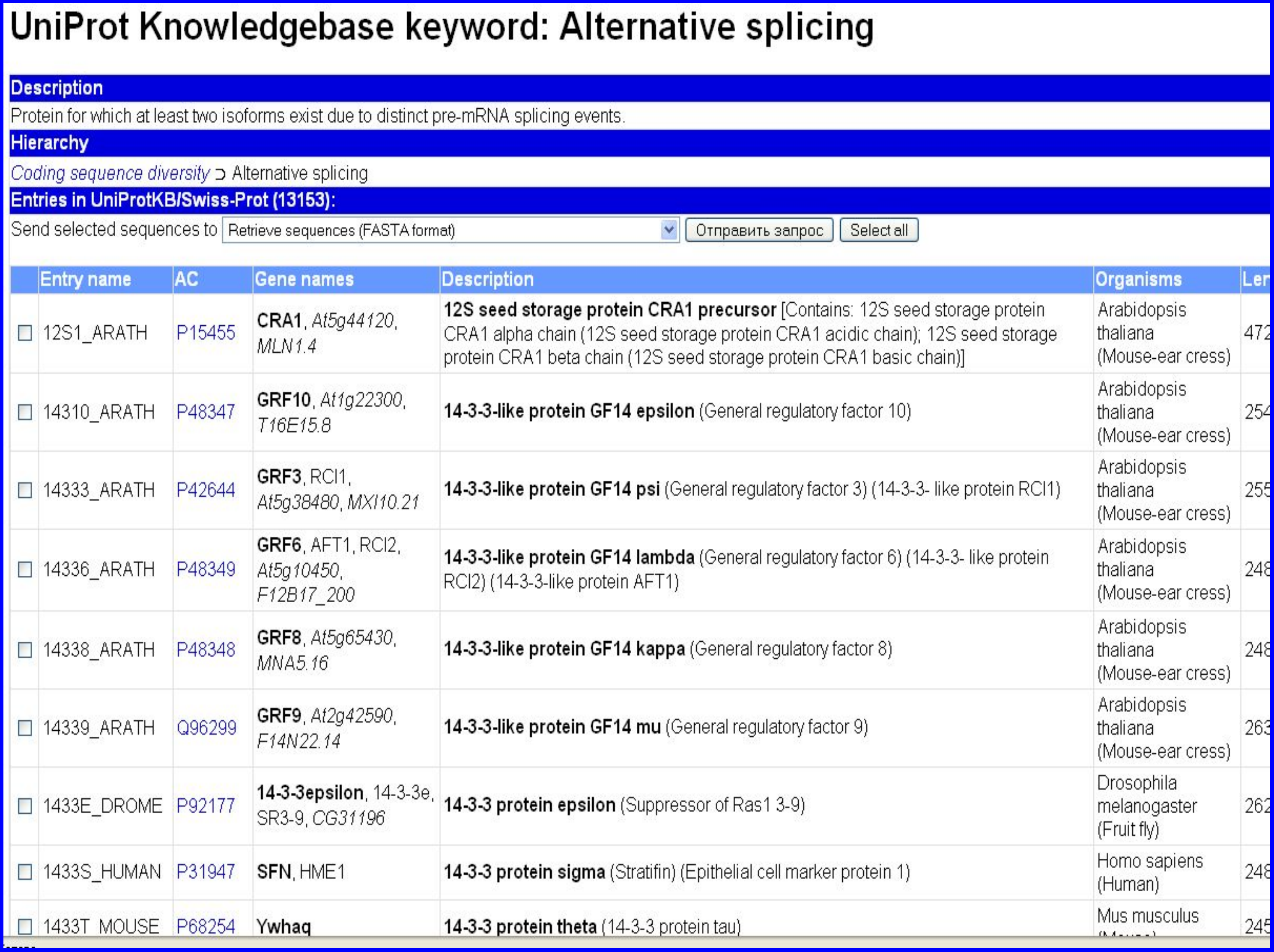


DrugBank

DrugBank
Cetuximab
(BTD00071)

Creation Date	2005/6/28 0:11:12 GMT
Last Update	Feb 01, 2007
Accession Number	BTD00071
Generic Name	Cetuximab
Synonyms	1. Anti EGFR 2. Ig gamma-1 chain C region
Brand Names	Erbitux (ImClone Systems Inc)
Description	Epidermal growth factor receptor binding FAB. Cetuximab is composed of the Fv (variable; antigen-binding) regions of the 225 murine EGFr monoclonal antibody specific for the N-terminal portion of human EGFr with human IgG1 heavy and kappa light chain constant (framework) regions.
Chemical IUPAC Name	Humanized anti-EGF receptor (EGFr) antibody
Chemical Formula	$C_{6484}H_{10042}N_{1732}O_{2023}S_{36}$
	> Cetuximab > Anti-EGFR heavy chain 1 QVQLKQSGPGLVQPSQSLSITCTVSGFSLTNYGVHWVRQSPGKGLEWLGVIWSGGNTDYN TPFTSRLSINKDNSKSQVFFKMNSLQSNDTAIYYCARALTYYDYEFAYWGQGTLVTVSAA STKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSG LYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKRVEPKSPKSCDKTHTCPPCPAPELL GGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQ YNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSR

Keywords



UniProt Knowledgebase keyword: Alternative splicing

Description

Protein for which at least two isoforms exist due to distinct pre-mRNA splicing events.

Hierarchy

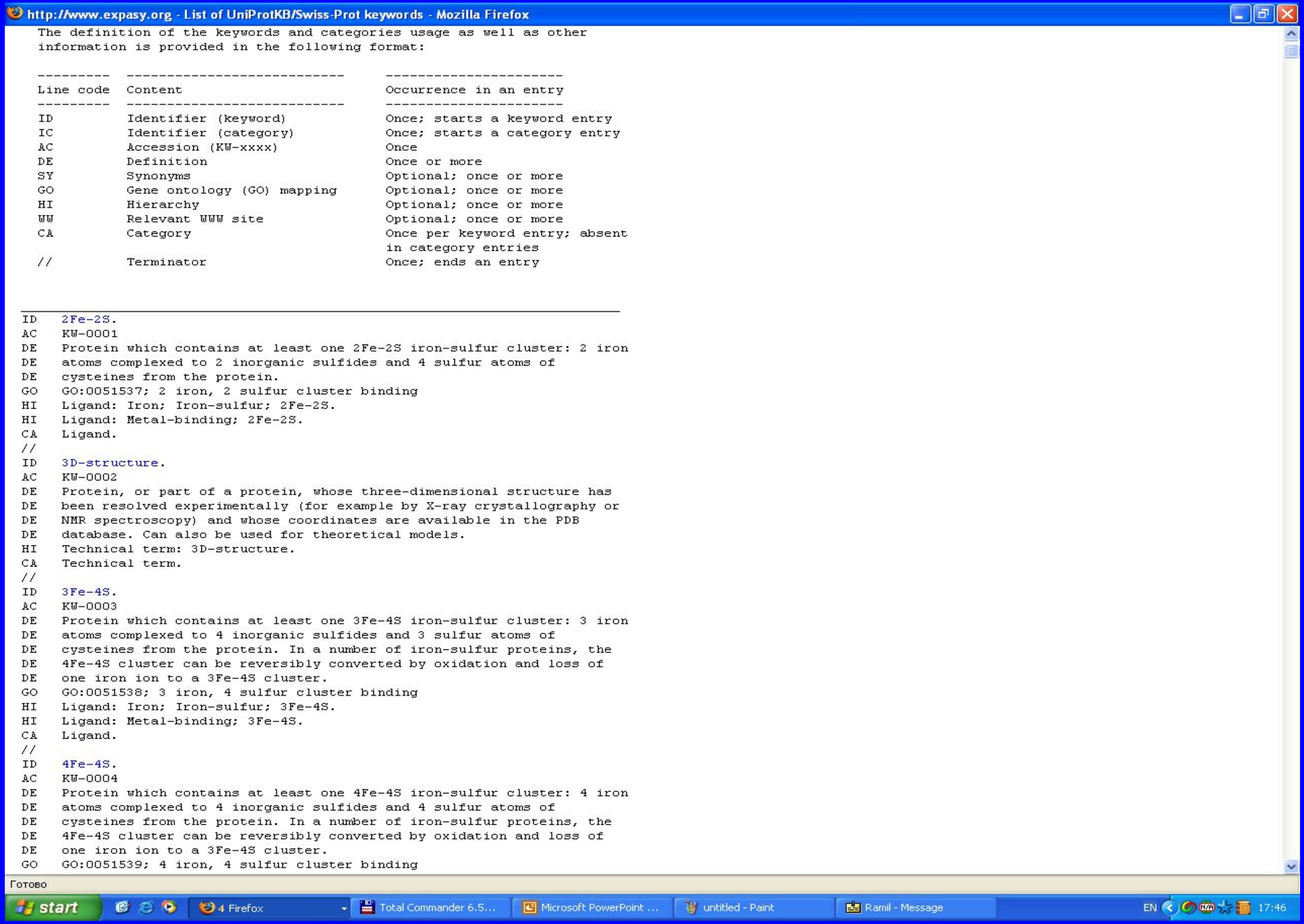
Coding sequence diversity ⊃ Alternative splicing

Entries in UniProtKB/Swiss-Prot (13153):

Send selected sequences to Retrieve sequences (FASTA format)

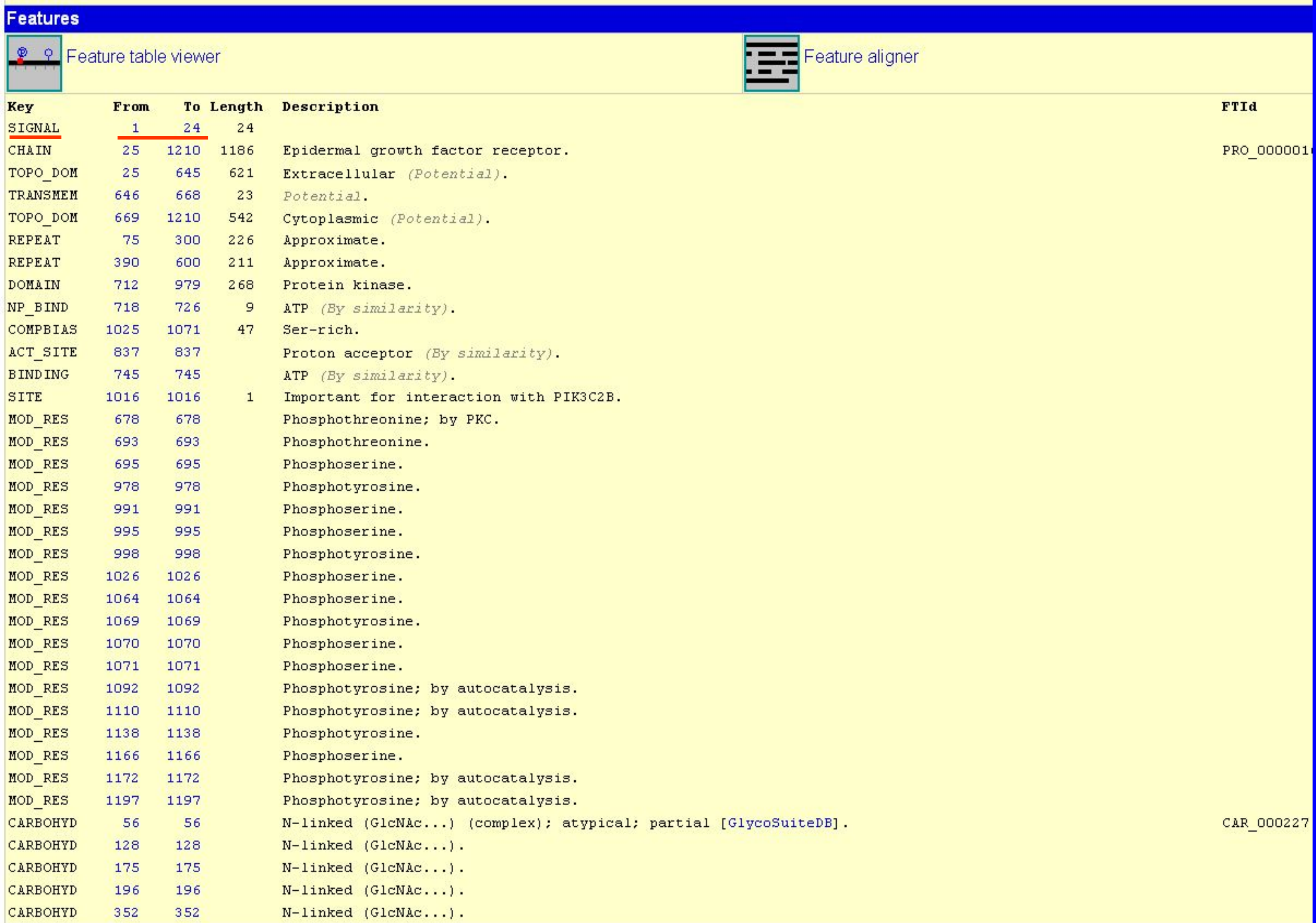
	Entry name	AC	Gene names	Description	Organisms	Len
☐	12S1_ARATH	P15455	**CRA1**, *At5g44120*, *MLN1.4*	**12S seed storage protein CRA1 precursor** [Contains: 12S seed storage protein CRA1 alpha chain (12S seed storage protein CRA1 acidic chain); 12S seed storage protein CRA1 beta chain (12S seed storage protein CRA1 basic chain)]	Arabidopsis thaliana (Mouse-ear cress)	472
☐	14310_ARATH	P48347	**GRF10**, *At1g22300*, *T16E15.8*	**14-3-3-like protein GF14 epsilon** (General regulatory factor 10)	Arabidopsis thaliana (Mouse-ear cress)	254
☐	14333_ARATH	P42644	**GRF3**, RCI1, *At5g38480*, *MXI10.21*	**14-3-3-like protein GF14 psi** (General regulatory factor 3) (14-3-3- like protein RCI1)	Arabidopsis thaliana (Mouse-ear cress)	255
☐	14336_ARATH	P48349	**GRF6**, AFT1, RCI2, *At5g10450*, *F12B17_200*	**14-3-3-like protein GF14 lambda** (General regulatory factor 6) (14-3-3- like protein RCI2) (14-3-3-like protein AFT1)	Arabidopsis thaliana (Mouse-ear cress)	248
☐	14338_ARATH	P48348	**GRF8**, *At5g65430*, *MNA5.16*	**14-3-3-like protein GF14 kappa** (General regulatory factor 8)	Arabidopsis thaliana (Mouse-ear cress)	248
☐	14339_ARATH	Q96299	**GRF9**, *At2g42590*, *F14N22.14*	**14-3-3-like protein GF14 mu** (General regulatory factor 9)	Arabidopsis thaliana (Mouse-ear cress)	263
☐	1433E_DROME	P92177	**14-3-3epsilon**, 14-3-3e, SR3-9, *CG31196*	**14-3-3 protein epsilon** (Suppressor of Ras1 3-9)	Drosophila melanogaster (Fruit fly)	262
☐	1433S_HUMAN	P31947	**SFN**, HME1	**14-3-3 protein sigma** (Stratifin) (Epithelial cell marker protein 1)	Homo sapiens (Human)	248
☐	1433T_MOUSE	P68254	**Ywhaq**	**14-3-3 protein theta** (14-3-3 protein tau)	Mus musculus	245

Словарь ключевых слов



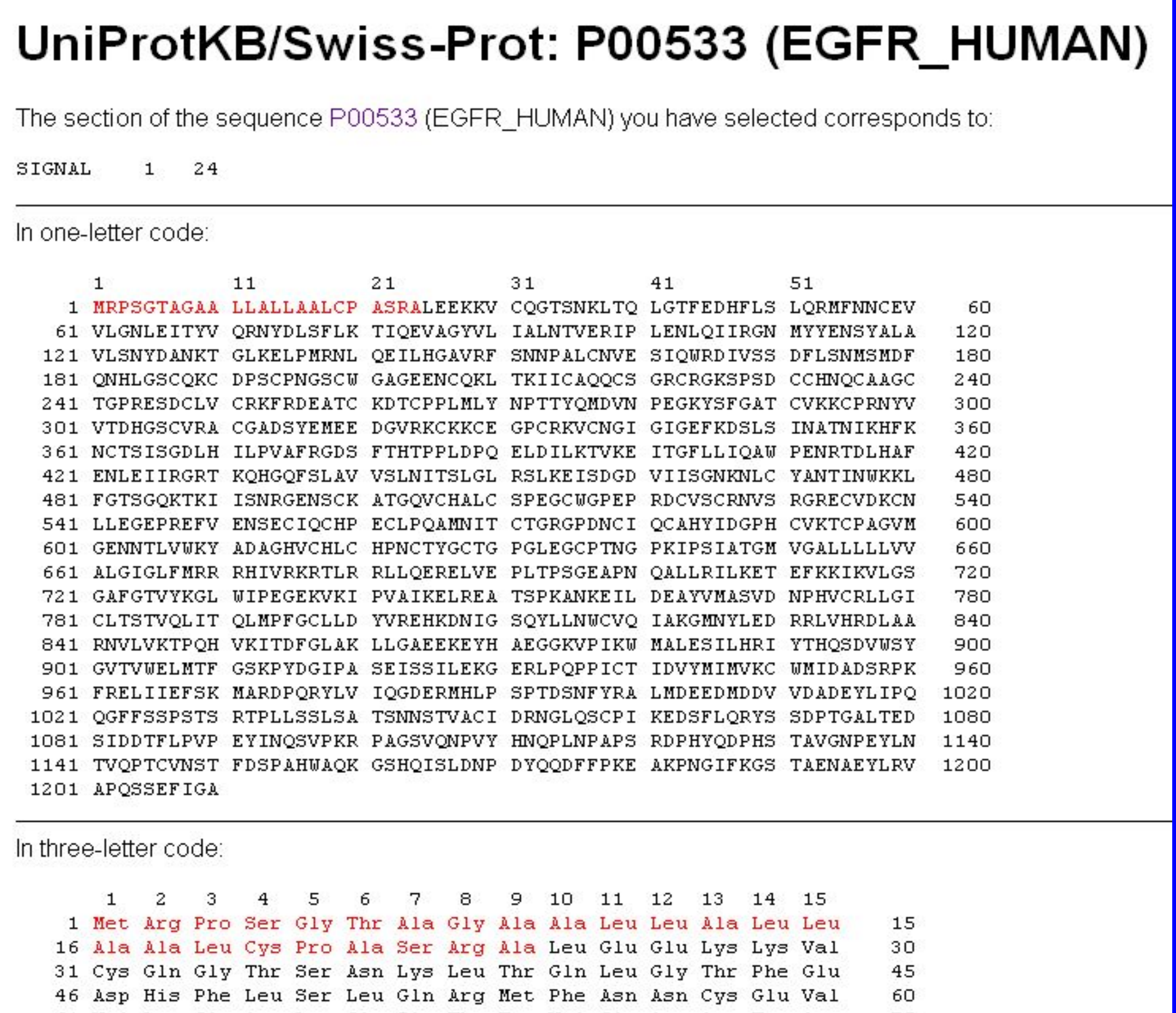
```
  The definition of the keywords and categories usage as well as other
  information is provided in the following format:

  ---------  ---------------------------     ----------------------
  Line code  Content                         Occurrence in an entry
  ---------  ---------------------------     ----------------------
  ID         Identifier (keyword)            Once; starts a keyword entry
  IC         Identifier (category)           Once; starts a category entry
  AC         Accession (KW-xxxx)             Once
  DE         Definition                      Once or more
  SY         Synonyms                        Optional; once or more
  GO         Gene ontology (GO) mapping      Optional; once or more
  HI         Hierarchy                       Optional; once or more
  WW         Relevant WWW site               Optional; once or more
  CA         Category                        Once per keyword entry; absent
                                             in category entries
  //         Terminator                      Once; ends an entry


ID   2Fe-2S.
AC   KW-0001
DE   Protein which contains at least one 2Fe-2S iron-sulfur cluster: 2 iron
DE   atoms complexed to 2 inorganic sulfides and 4 sulfur atoms of
DE   cysteines from the protein.
GO   GO:0051537; 2 iron, 2 sulfur cluster binding
HI   Ligand: Iron; Iron-sulfur; 2Fe-2S.
HI   Ligand: Metal-binding; 2Fe-2S.
CA   Ligand.
//
ID   3D-structure.
AC   KW-0002
DE   Protein, or part of a protein, whose three-dimensional structure has
DE   been resolved experimentally (for example by X-ray crystallography or
DE   NMR spectroscopy) and whose coordinates are available in the PDB
DE   database. Can also be used for theoretical models.
HI   Technical term: 3D-structure.
CA   Technical term.
//
ID   3Fe-4S.
AC   KW-0003
DE   Protein which contains at least one 3Fe-4S iron-sulfur cluster: 3 iron
DE   atoms complexed to 4 inorganic sulfides and 3 sulfur atoms of
DE   cysteines from the protein. In a number of iron-sulfur proteins, the
DE   4Fe-4S cluster can be reversibly converted by oxidation and loss of
DE   one iron ion to a 3Fe-4S cluster.
GO   GO:0051538; 3 iron, 4 sulfur cluster binding
HI   Ligand: Iron; Iron-sulfur; 3Fe-4S.
HI   Ligand: Metal-binding; 3Fe-4S.
CA   Ligand.
//
ID   4Fe-4S.
AC   KW-0004
DE   Protein which contains at least one 4Fe-4S iron-sulfur cluster: 4 iron
DE   atoms complexed to 4 inorganic sulfides and 4 sulfur atoms of
DE   cysteines from the protein. In a number of iron-sulfur proteins, the
DE   4Fe-4S cluster can be reversibly converted by oxidation and loss of
DE   one iron ion to a 3Fe-4S cluster.
GO   GO:0051539; 4 iron, 4 sulfur cluster binding
```

Feature Table



Key	From	To	Length	Description	FTId
SIGNAL	1	24	24		
CHAIN	25	1210	1186	Epidermal growth factor receptor.	PRO_000001
TOPO_DOM	25	645	621	Extracellular *(Potential)*.	
TRANSMEM	646	668	23	*Potential*.	
TOPO_DOM	669	1210	542	Cytoplasmic *(Potential)*.	
REPEAT	75	300	226	Approximate.	
REPEAT	390	600	211	Approximate.	
DOMAIN	712	979	268	Protein kinase.	
NP_BIND	718	726	9	ATP *(By similarity)*.	
COMPBIAS	1025	1071	47	Ser-rich.	
ACT_SITE	837	837		Proton acceptor *(By similarity)*.	
BINDING	745	745		ATP *(By similarity)*.	
SITE	1016	1016	1	Important for interaction with PIK3C2B.	
MOD_RES	678	678		Phosphothreonine; by PKC.	
MOD_RES	693	693		Phosphothreonine.	
MOD_RES	695	695		Phosphoserine.	
MOD_RES	978	978		Phosphotyrosine.	
MOD_RES	991	991		Phosphoserine.	
MOD_RES	995	995		Phosphoserine.	
MOD_RES	998	998		Phosphotyrosine.	
MOD_RES	1026	1026		Phosphoserine.	
MOD_RES	1064	1064		Phosphoserine.	
MOD_RES	1069	1069		Phosphotyrosine.	
MOD_RES	1070	1070		Phosphoserine.	
MOD_RES	1071	1071		Phosphoserine.	
MOD_RES	1092	1092		Phosphotyrosine; by autocatalysis.	
MOD_RES	1110	1110		Phosphotyrosine; by autocatalysis.	
MOD_RES	1138	1138		Phosphotyrosine.	
MOD_RES	1166	1166		Phosphoserine.	
MOD_RES	1172	1172		Phosphotyrosine; by autocatalysis.	
MOD_RES	1197	1197		Phosphotyrosine; by autocatalysis.	
CARBOHYD	56	56		N-linked (GlcNAc...) (complex); atypical; partial [GlycoSuiteDB].	CAR_000227
CARBOHYD	128	128		N-linked (GlcNAc...).	
CARBOHYD	175	175		N-linked (GlcNAc...).	
CARBOHYD	196	196		N-linked (GlcNAc...).	
CARBOHYD	352	352		N-linked (GlcNAc...).	

Координаты в Feature table



UniProtKB/Swiss-Prot: P00533 (EGFR_HUMAN)

The section of the sequence P00533 (EGFR_HUMAN) you have selected corresponds to:

```
SIGNAL     1    24
```

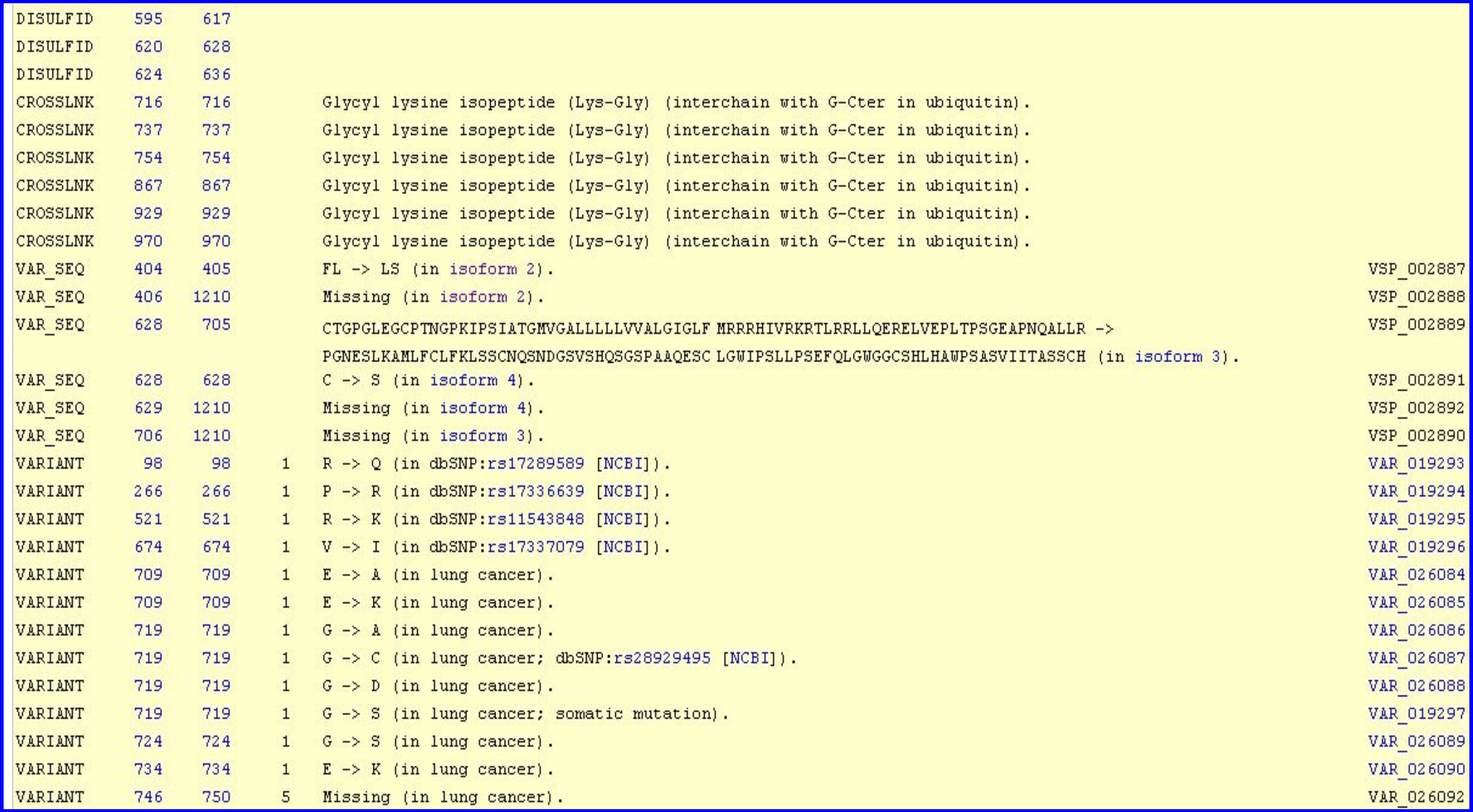
In one-letter code:

```
         1          11         21         31         41         51
     1 MRPSGTAGAA LLALLAALCP ASRALEEKKV CQGTSNKLTQ LGTFEDHFLS LQRMFNNCEV    60
    61 VLGNLEITYV QRNYDLSFLK TIQEVAGYVL IALNTVERIP LENLQIIRGN MYYENSYALA   120
   121 VLSNYDANKT GLKELPMRNL QEILHGAVRF SNNPALCNVE SIQWRDIVSS DFLSNMSMDF   180
   181 QNHLGSCQKC DPSCPNGSCW GAGEENCQKL TKIICAQQCS GRCRGKSPSD CCHNQCAAGC   240
   241 TGPRESDCLV CRKFRDEATC KDTCPPLMLY NPTTYQMDVN PEGKYSFGAT CVKKCPRNYV   300
   301 VTDHGSCVRA CGADSYEMEE DGVRKCKKCE GPCRKVCNGI GIGEFKDSLS INATNIKHFK   360
   361 NCTSISGDLH ILPVAFRGDS FTHTPPLDPQ ELDILKTVKE ITGFLLIQAW PENRTDLHAF   420
   421 ENLEIIRGRT KQHGQFSLAV VSLNITSLGL RSLKEISDGD VIISGNKNLC YANTINWKKL   480
   481 FGTSGQKTKI ISNRGENSCK ATGQVCHALC SPEGCWGPEP RDCVSCRNVS RGRECVDKCN   540
   541 LLEGEPREFV ENSECIQCHP ECLPQAMNIT CTGRGPDNCI QCAHYIDGPH CVKTCPAGVM   600
   601 GENNTLVWKY ADAGHVCHLC HPNCTYGCTG PGLEGCPTNG PKIPSIATGM VGALLLLLVV   660
   661 ALGIGLFMRR RHIVRKRTLR RLLQERELVE PLTPSGEAPN QALLRILKET EFKKIKVLGS   720
   721 GAFGTVYKGL WIPEGEKVKI PVAIKELREA TSPKANKEIL DEAYVMASVD NPHVCRLLGI   780
   781 CLTSTVQLIT QLMPFGCLLD YVREHKDNIG SQYLLNWCVQ IAKGMNYLED RRLVHRDLAA   840
   841 RNVLVKTPQH VKITDFGLAK LLGAEEKEYH AEGGKVPIKW MALESILHRI YTHQSDVWSY   900
   901 GVTVWELMTF GSKPYDGIPA SEISSILEKG ERLPQPPICT IDVYMIMVKC WMIDADSRPK   960
   961 FRELIIEFSK MARDPQRYLV IQGDERMHLP SPTDSNFYRA LMDEEDMDDV VDADEYLIPQ  1020
  1021 QGFFSSPSTS RTPLLSSLSA TSNNSTVACI DRNGLQSCPI KEDSFLQRYS SDPTGALTED  1080
  1081 SIDDTFLPVP EYINQSVPKR PAGSVQNPVY HNQPLNPAPS RDPHYQDPHS TAVGNPEYLN  1140
  1141 TVQPTCVNST FDSPAHWAQK GSHQISLDNP DYQQDFFPKE AKPNGIFKGS TAENAEYLRV  1200
  1201 APQSSEFIGA
```

In three-letter code:

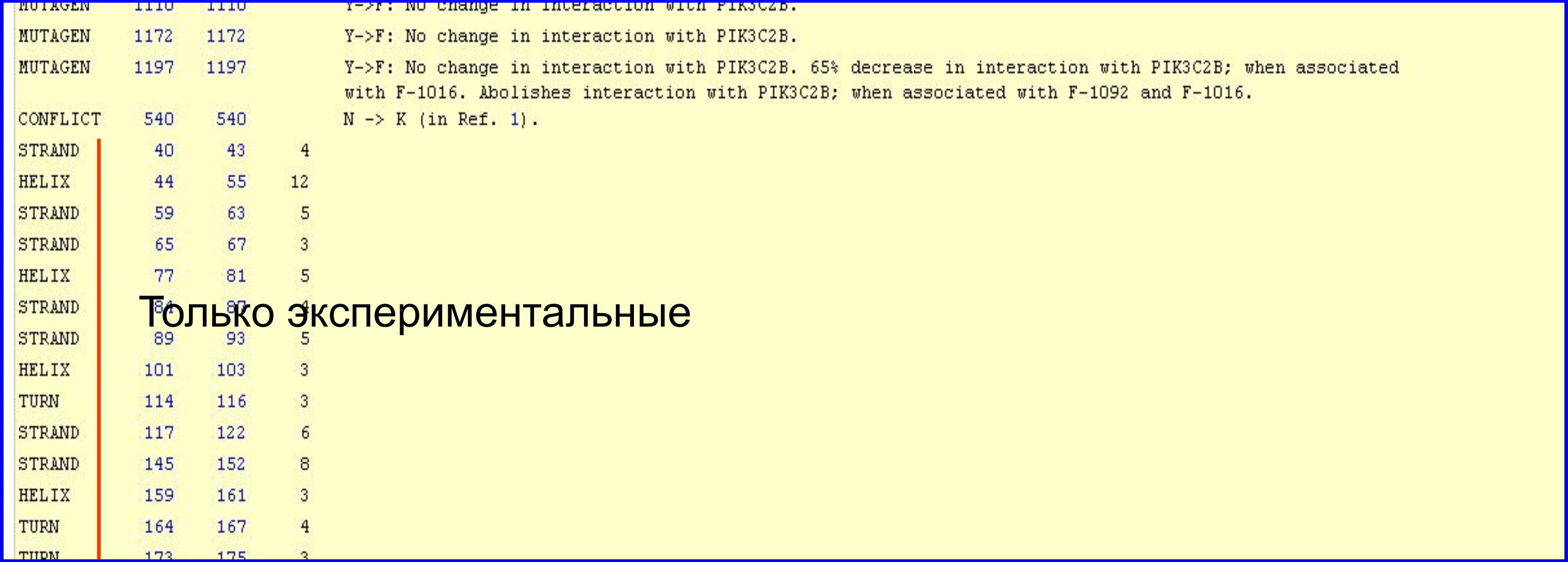
```
      1   2   3   4   5   6   7   8   9  10  11  12  13  14  15
   1 Met Arg Pro Ser Gly Thr Ala Gly Ala Ala Leu Leu Ala Leu Leu    15
  16 Ala Ala Leu Cys Pro Ala Ser Arg Ala Leu Glu Glu Lys Lys Val    30
  31 Cys Gln Gly Thr Ser Asn Lys Leu Thr Gln Leu Gly Thr Phe Glu    45
  46 Asp His Phe Leu Ser Leu Gln Arg Met Phe Asn Asn Cys Glu Val    60
```

Feature table, продолжение



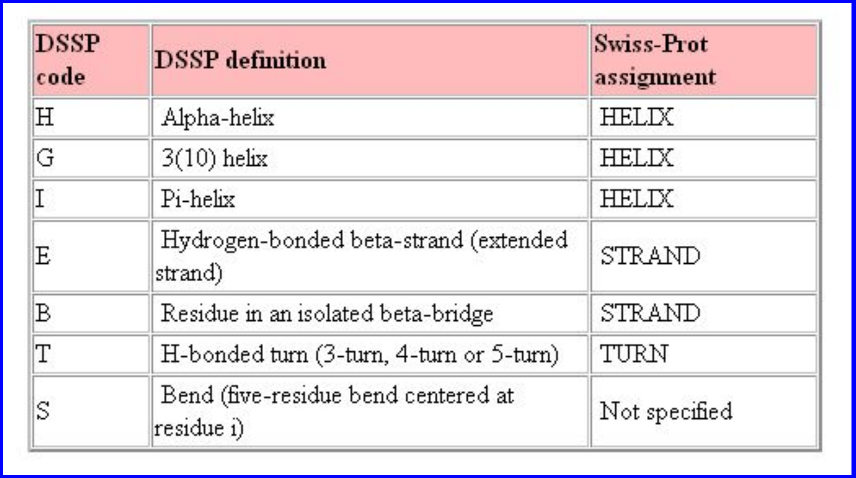
```
DISULFID   595   617
DISULFID   620   628
DISULFID   624   636
CROSSLNK   716   716        Glycyl lysine isopeptide (Lys-Gly) (interchain with G-Cter in ubiquitin).
CROSSLNK   737   737        Glycyl lysine isopeptide (Lys-Gly) (interchain with G-Cter in ubiquitin).
CROSSLNK   754   754        Glycyl lysine isopeptide (Lys-Gly) (interchain with G-Cter in ubiquitin).
CROSSLNK   867   867        Glycyl lysine isopeptide (Lys-Gly) (interchain with G-Cter in ubiquitin).
CROSSLNK   929   929        Glycyl lysine isopeptide (Lys-Gly) (interchain with G-Cter in ubiquitin).
CROSSLNK   970   970        Glycyl lysine isopeptide (Lys-Gly) (interchain with G-Cter in ubiquitin).
VAR_SEQ    404   405        FL -> LS (in isoform 2).                                                      VSP_002887
VAR_SEQ    406  1210        Missing (in isoform 2).                                                       VSP_002888
VAR_SEQ    628   705        CTGPGLEGCPTNGPKIPSIATGMVGALLLLLVVALGIGLF MRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLR ->   VSP_002889
                            PGNESLKAMLFCLFKLSSCNQSNDGSVSHQSGSPAAQESC LGWIPSLLPSEFQLGWGGCSHLHAWPSASVIITASSCH (in isoform 3).
VAR_SEQ    628   628        C -> S (in isoform 4).                                                        VSP_002891
VAR_SEQ    629  1210        Missing (in isoform 4).                                                       VSP_002892
VAR_SEQ    706  1210        Missing (in isoform 3).                                                       VSP_002890
VARIANT     98    98     1  R -> Q (in dbSNP:rs17289589 [NCBI]).                                          VAR_019293
VARIANT    266   266     1  P -> R (in dbSNP:rs17336639 [NCBI]).                                          VAR_019294
VARIANT    521   521     1  R -> K (in dbSNP:rs11543848 [NCBI]).                                          VAR_019295
VARIANT    674   674     1  V -> I (in dbSNP:rs17337079 [NCBI]).                                          VAR_019296
VARIANT    709   709     1  E -> A (in lung cancer).                                                      VAR_026084
VARIANT    709   709     1  E -> K (in lung cancer).                                                      VAR_026085
VARIANT    719   719     1  G -> A (in lung cancer).                                                      VAR_026086
VARIANT    719   719     1  G -> C (in lung cancer; dbSNP:rs28929495 [NCBI]).                             VAR_026087
VARIANT    719   719     1  G -> D (in lung cancer).                                                      VAR_026088
VARIANT    719   719     1  G -> S (in lung cancer; somatic mutation).                                    VAR_019297
VARIANT    724   724     1  G -> S (in lung cancer).                                                      VAR_026089
VARIANT    734   734     1  E -> K (in lung cancer).                                                      VAR_026090
VARIANT    746   750     5  Missing (in lung cancer).                                                     VAR_026092
```

Feature Table, продолжение



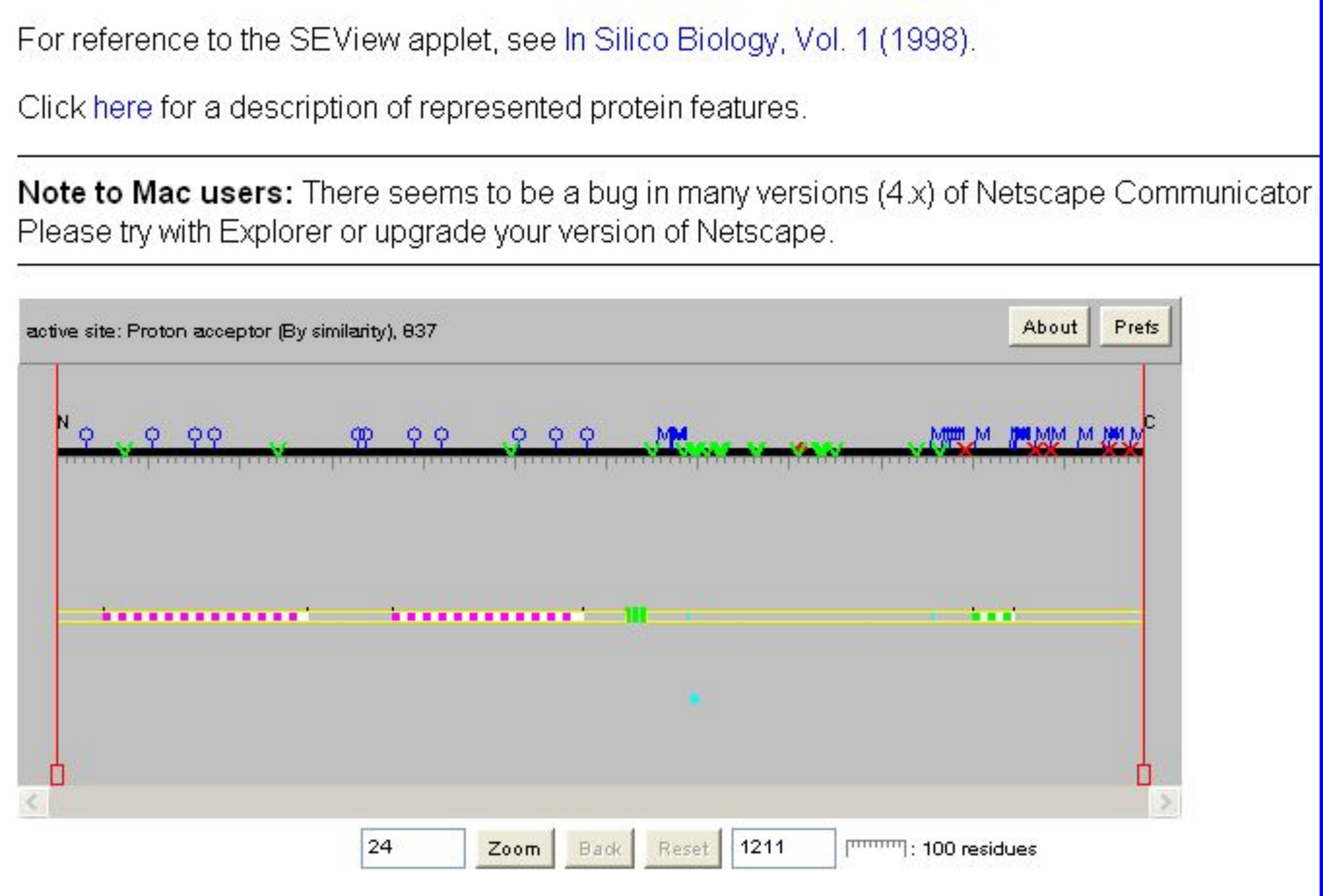
```
MUTAGEN     1172    1172          Y->F: No change in interaction with PIK3C2B.
MUTAGEN     1197    1197          Y->F: No change in interaction with PIK3C2B. 65% decrease in interaction with PIK3C2B; when associated
                                  with F-1016. Abolishes interaction with PIK3C2B; when associated with F-1092 and F-1016.
CONFLICT     540     540          N -> K (in Ref. 1).
STRAND        40      43     4
HELIX         44      55    12
STRAND        59      63     5
STRAND        65      67     3
HELIX         77      81     5
STRAND        84      87     4
STRAND        89      93     5
HELIX        101     103     3
TURN         114     116     3
STRAND       117     122     6
STRAND       145     152     8
HELIX        159     161     3
TURN         164     167     4
```

Только экспериментальные



DSSP code	DSSP definition	Swiss-Prot assignment
H	Alpha-helix	HELIX
G	3(10) helix	HELIX
I	Pi-helix	HELIX
E	Hydrogen-bonded beta-strand (extended strand)	STRAND
B	Residue in an isolated beta-bridge	STRAND
T	H-bonded turn (3-turn, 4-turn or 5-turn)	TURN
S	Bend (five-residue bend centered at residue i)	Not specified

Feature Table viewer
(Sequence Element viewer)



For reference to the SEView applet, see In Silico Biology, Vol. 1 (1998).

Click here for a description of represented protein features.

Note to Mac users: There seems to be a bug in many versions (4.x) of Netscape Communicator
Please try with Explorer or upgrade your version of Netscape.

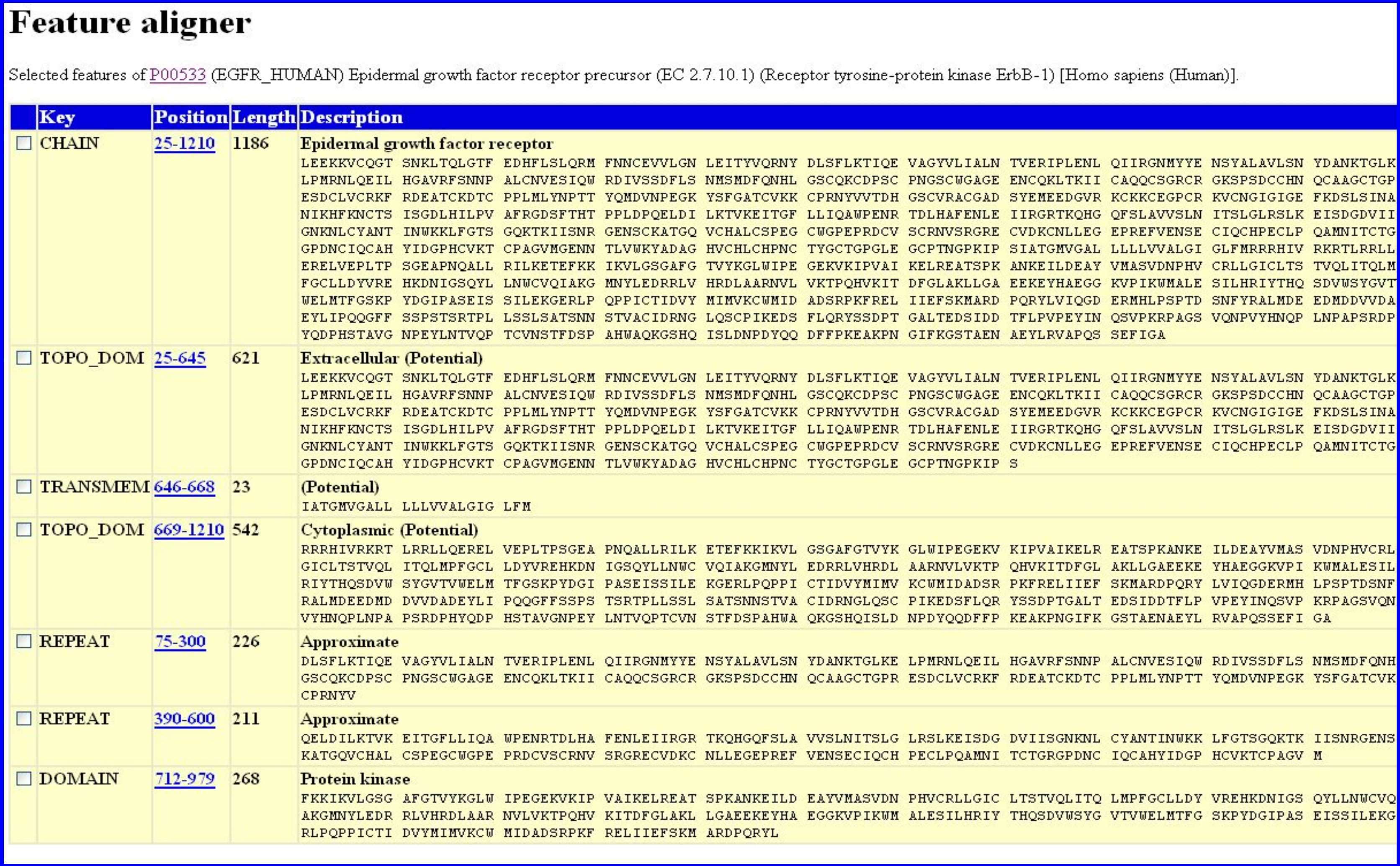
---

active site: Proton acceptor (By similarity), 837

24

1211

Feature aligner



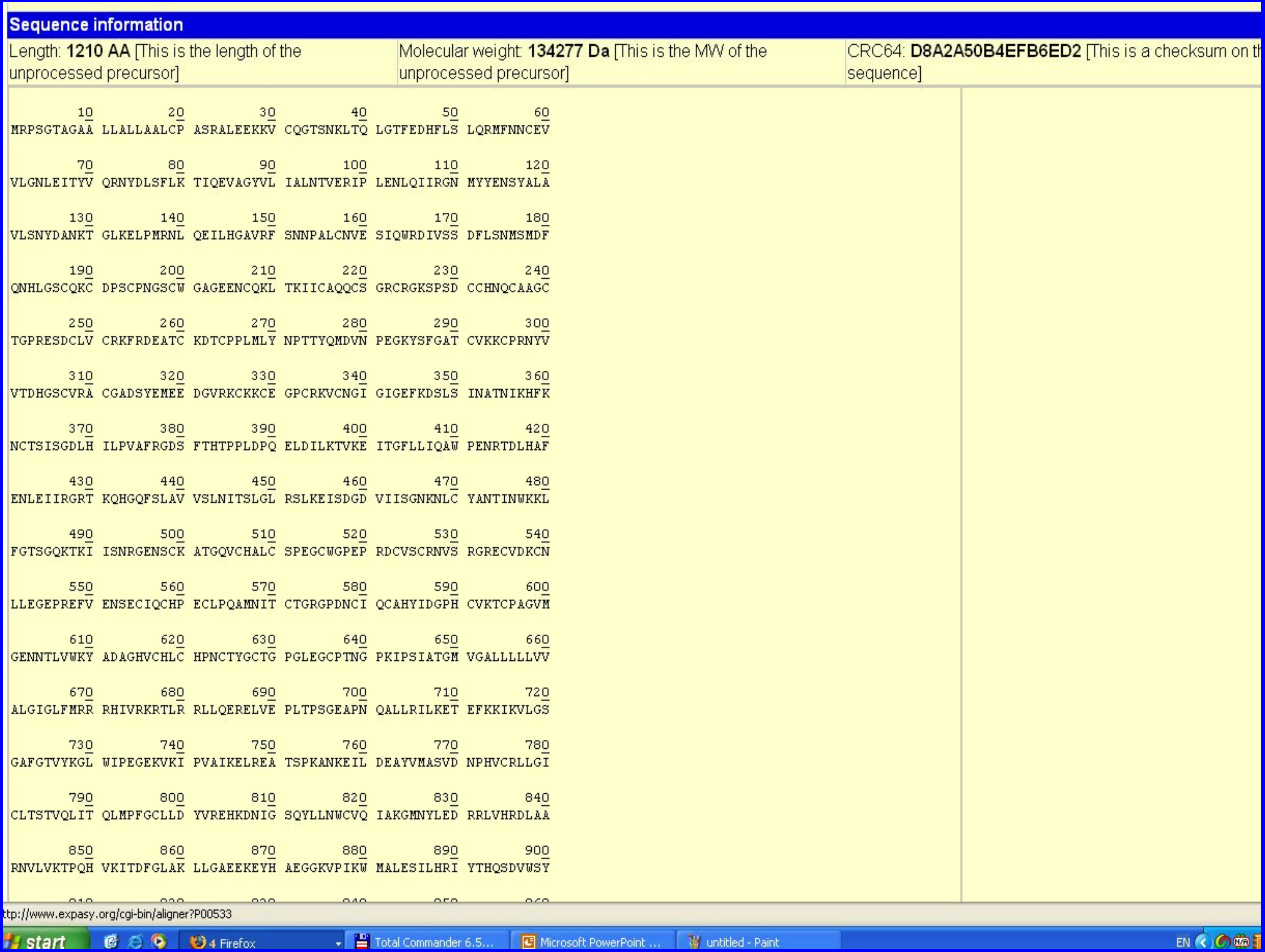
Feature aligner

Selected features of P00533 (EGFR_HUMAN) Epidermal growth factor receptor precursor (EC 2.7.10.1) (Receptor tyrosine-protein kinase ErbB-1) [Homo sapiens (Human)].

	Key	Position	Length	Description
☐	CHAIN	25-1210	1186	**Epidermal growth factor receptor** LEEKKVCQGT SNKLTQLGTF EDHFLSLQRM FNNCEVVLGN LEITYVQRNY DLSFLKTIQE VAGYVLIALN TVERIPLENL QIIRGNMYYE NSYALAVLSN YDANKTGLK LPMRNLQEIL HGAVRFSNNP ALCNVESIQW RDIVSSDFLS NMSMDFQNHL GSCQKCDPSC PNGSCWGAGE ENCQKLTKII CAQQCSGRCR GKSPSDCCHN QCAAGCTGP ESDCLVCRKF RDEATCKDTC PPLMLYNPTT YQMDVNPEGK YSFGATCVKK CPRNYVVTDH GSCVRACGAD SYEMEEDGVR KCKKCEGPCR KVCNGIGIGE FKDSLSINA NIKHFKNCTS ISGDLHILPV AFRGDSFTHT PPLDPQELDI LKTVKEITGF LLIQAWPENR TDLHAFENLE IIRGRTKQHG QFSLAVVSLN ITSLGLRSLK EISDGDVII GNKNLCYANT INWKKLFGTS GQKTKIISNR GENSCKATGQ VCHALCSPEG CWGPEPRDCV SCRNVSRGRE CVDKCNLLEG EPREFVENSE CIQCHPECLP QAMNITCTG GPDNCIQCAH YIDGPHCVKT CPAGVMGENN TLVWKYADAG HVCHLCHPNC TYGCTGPGLE GCPTNGPKIP SIATGMVGAL LLLLVVALGI GLFMRRRHIV RKRTLRRLL ERELVEPLTP SGEAPNQALL RILKETEFKK IKVLGSGAFG TVYKGLWIPE GEKVKIPVAI KELREATSPK ANKEILDEAY VMASVDNPHV CRLLGICLTS TVQLITQLM FGCLLDYVRE HKDNIGSQYL LNWCVQIAKG MNYLEDRRLV HRDLAARNVL VKTPQHVKIT DFGLAKLLGA EEKEYHAEGG KVPIKWMALE SILHRIYTHQ SDVWSYGVT WELMTFGSKP YDGIPASEIS SILEKGERLP QPPICTIDVY MIMVKCWMID ADSRPKFREL IIEFSKMARD PQRYLVIQGD ERMHLPSPTD SNFYRALMDE EDMDDVVDA EYLIPQQGFF SSPSTSRTPL LSSLSATSNN STVACIDRNG LQSCPIKEDS FLQRYSSDPT GALTEDSIDD TFLPVPEYIN QSVPKRPAGS VQNPVYHNQP LNPAPSRDP YQDPHSTAVG NPEYLNTVQP TCVNSTFDSP AHWAQKGSHQ ISLDNPDYQQ DFFPKEAKPN GIFKGSTAEN AEYLRVAPQS SEFIGA
☐	TOPO_DOM	25-645	621	**Extracellular (Potential)** LEEKKVCQGT SNKLTQLGTF EDHFLSLQRM FNNCEVVLGN LEITYVQRNY DLSFLKTIQE VAGYVLIALN TVERIPLENL QIIRGNMYYE NSYALAVLSN YDANKTGLK LPMRNLQEIL HGAVRFSNNP ALCNVESIQW RDIVSSDFLS NMSMDFQNHL GSCQKCDPSC PNGSCWGAGE ENCQKLTKII CAQQCSGRCR GKSPSDCCHN QCAAGCTGP ESDCLVCRKF RDEATCKDTC PPLMLYNPTT YQMDVNPEGK YSFGATCVKK CPRNYVVTDH GSCVRACGAD SYEMEEDGVR KCKKCEGPCR KVCNGIGIGE FKDSLSINA NIKHFKNCTS ISGDLHILPV AFRGDSFTHT PPLDPQELDI LKTVKEITGF LLIQAWPENR TDLHAFENLE IIRGRTKQHG QFSLAVVSLN ITSLGLRSLK EISDGDVII GNKNLCYANT INWKKLFGTS GQKTKIISNR GENSCKATGQ VCHALCSPEG CWGPEPRDCV SCRNVSRGRE CVDKCNLLEG EPREFVENSE CIQCHPECLP QAMNITCTG GPDNCIQCAH YIDGPHCVKT CPAGVMGENN TLVWKYADAG HVCHLCHPNC TYGCTGPGLE GCPTNGPKIP S
☐	TRANSMEM	646-668	23	**(Potential)** IATGMVGALL LLLVVALGIG LFM
☐	TOPO_DOM	669-1210	542	**Cytoplasmic (Potential)** RRRHIVRKRT LRRLLQEREL VEPLTPSGEA PNQALLRILK ETEFKKIKVL GSGAFGTVYK GLWIPEGEKV KIPVAIKELR EATSPKANKE ILDEAYVMAS VDNPHVCRL GICLTSTVQL ITQLMPFGCL LDYVREHKDN IGSQYLLNWC VQIAKGMNYL EDRRLVHRDL AARNVLVKTP QHVKITDFGL AKLLGAEEKE YHAEGGKVPI KWMALESIL RIYTHQSDVW SYGVTVWELM TFGSKPYDGI PASEISSILE KGERLPQPPI CTIDVYMIMV KCWMIDADSR PKFRELIIEF SKMARDPQRY LVIQGDERMH LPSPTDSNF RALMDEEDMD DVVDADEYLI PQQGFFSSPS TSRTPLLSSL SATSNNSTVA CIDRNGLQSC PIKEDSFLQR YSSDPTGALT EDSIDDTFLP VPEYINQSVP KRPAGSVQN VYHNQPLNPA PSRDPHYQDP HSTAVGNPEY LNTVQPTCVN STFDSPAHWA QKGSHQISLD NPDYQQDFFP KEAKPNGIFK GSTAENAEYL RVAPQSSEFI GA
☐	REPEAT	75-300	226	**Approximate** DLSFLKTIQE VAGYVLIALN TVERIPLENL QIIRGNMYYE NSYALAVLSN YDANKTGLKE LPMRNLQEIL HGAVRFSNNP ALCNVESIQW RDIVSSDFLS NMSMDFQNH GSCQKCDPSC PNGSCWGAGE ENCQKLTKII CAQQCSGRCR GKSPSDCCHN QCAAGCTGPR ESDCLVCRKF RDEATCKDTC PPLMLYNPTT YQMDVNPEGK YSFGATCVK CPRNYV
☐	REPEAT	390-600	211	**Approximate** QELDILKTVK EITGFLLIQA WPENRTDLHA FENLEIIRGR TKQHGQFSLA VVSLNITSLG LRSLKEISDG DVIISGNKNL CYANTINWKK LFGTSGQKTK IISNRGENS KATGQVCHAL CSPEGCWGPE PRDCVSCRNV SRGRECVDKC NLLEGEPREF VENSECIQCH PECLPQAMNI TCTGRGPDNC IQCAHYIDGP HCVKTCPAGV M
☐	DOMAIN	712-979	268	**Protein kinase** FKKIKVLGSG AFGTVYKGLW IPEGEKVKIP VAIKELREAT SPKANKEILD EAYVMASVDN PHVCRLLGIC LTSTVQLITQ LMPFGCLLDY VREHKDNIGS QYLLNWCVQ AKGMNYLEDR RLVHRDLAAR NVLVKTPQHV KITDFGLAKL LGAEEKEYHA EGGKVPIKWM ALESILHRIY THQSDVWSYG VTVWELMTFG SKPYDGIPAS EISSILEKG RLPQPPICTI DVYMIMVKCW MIDADSRPKF RELIIEFSKM ARDPQRYL

Можно построить множественное выравнивание подмножества этих элементов (ClustalW) или скопировать их в FASTA формате

Sequence



Sequence information

Length: **1210 AA** [This is the length of the unprocessed precursor]

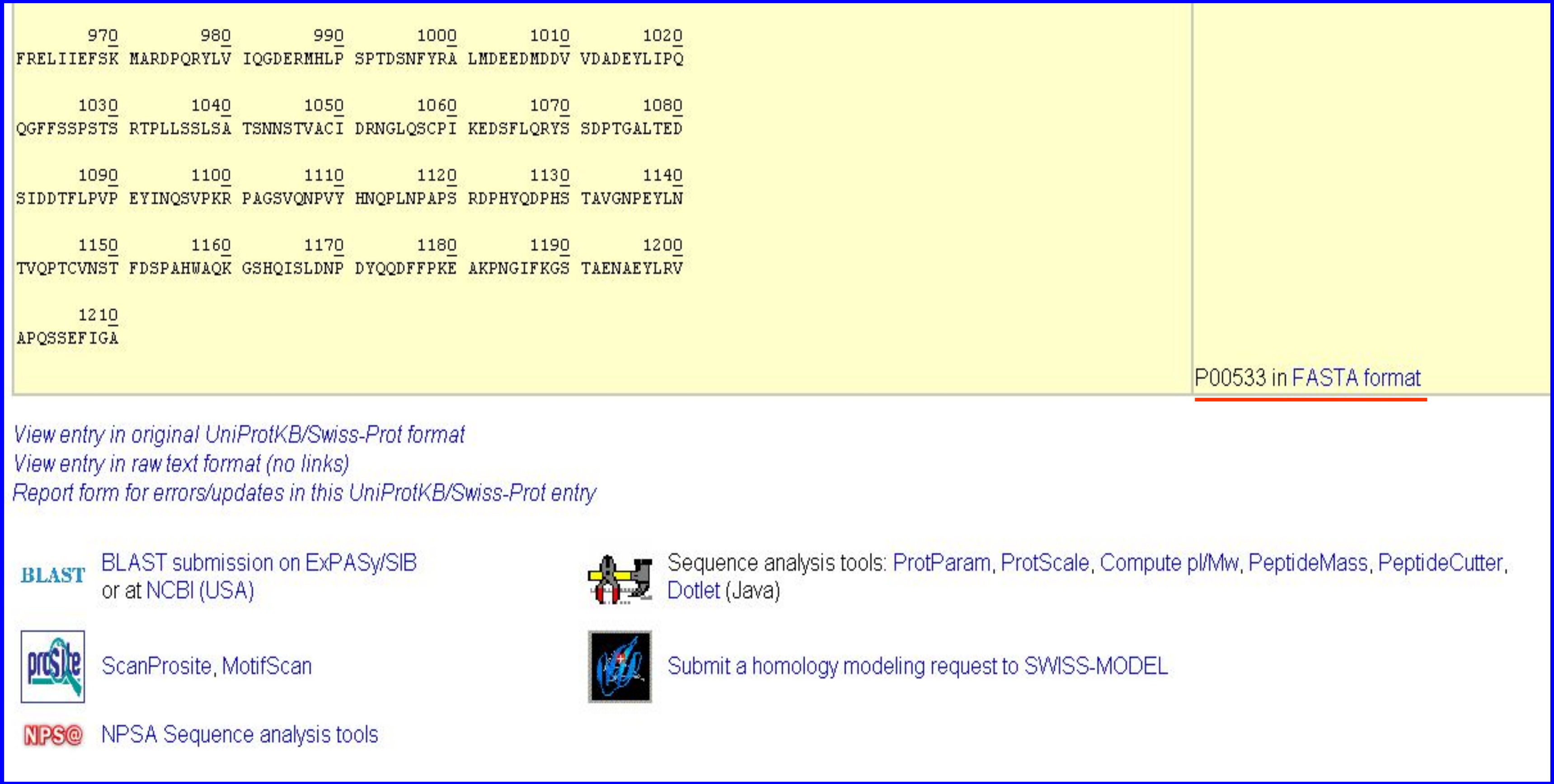
Molecular weight: **134277 Da** [This is the MW of the unprocessed precursor]

CRC64: **D8A2A50B4EFB6ED2** [This is a checksum on th sequence]

```
        10         20         30         40         50         60
MRPSGTAGAA LLALLAALCP ASRALEEKKV CQGTSNKLTQ LGTFEDHFLS LQRMFNNCEV

        70         80         90        100        110        120
VLGNLEITYV QRNYDLSFLK TIQEVAGYVL IALNTVERIP LENLQIIRGN MYYENSYALA

       130        140        150        160        170        180
VLSNYDANKT GLKELPMRNL QEILHGAVRF SNNPALCNVE SIQWRDIVSS DFLSNMSMDF

       190        200        210        220        230        240
QNHLGSCQKC DPSCPNGSCW GAGEENCQKL TKIICAQQCS GRCRGKSPSD CCHNQCAAGC

       250        260        270        280        290        300
TGPRESDCLV CRKFRDEATC KDTCPPLMLY NPTTYQMDVN PEGKYSFGAT CVKKCPRNYV

       310        320        330        340        350        360
VTDHGSCVRA CGADSYEMEE DGVRKCKKCE GPCRKVCNGI GIGEFKDSLS INATNIKHFK

       370        380        390        400        410        420
NCTSISGDLH ILPVAFRGDS FTHTPPLDPQ ELDILKTVKE ITGFLLIQAW PENRTDLHAF

       430        440        450        460        470        480
ENLEIIRGRT KQHGQFSLAV VSLNITSLGL RSLKEISDGD VIISGNKNLC YANTINWKKL

       490        500        510        520        530        540
FGTSGQKTKI ISNRGENSCK ATGQVCHALC SPEGCWGPEP RDCVSCRNVS RGRECVDKCN

       550        560        570        580        590        600
LLEGEPREFV ENSECIQCHP ECLPQAMNIT CTGRGPDNCI QCAHYIDGPH CVKTCPAGVM

       610        620        630        640        650        660
GENNTLVWKY ADAGHVCHLC HPNCTYGCTG PGLEGCPTNG PKIPSIATGM VGALLLLLVV

       670        680        690        700        710        720
ALGIGLFMRR RHIVRKRTLR RLLQERELVE PLTPSGEAPN QALLRILKET EFKKIKVLGS

       730        740        750        760        770        780
GAFGTVYKGL WIPEGEKVKI PVAIKELREA TSPKANKEIL DEAYVMASVD NPHVCRLLGI

       790        800        810        820        830        840
CLTSTVQLIT QLMPFGCLLD YVREHKDNIG SQYLLNWCVQ IAKGMNYLED RRLVHRDLAA

       850        860        870        880        890        900
RNVLVKTPQH VKITDFGLAK LLGAEEKEYH AEGGKVPIKW MALESILHRI YTHQSDVWSY
```

ttp://www.expasy.org/cgi-bin/aligner?P00533

Sequence, продолжение



```
       970        980        990       1000       1010       1020
FRELIIEFSK MARDPQRYLV IQGDERMHLP SPTDSNFYRA LMDEEDMDDV VDADEYLIPQ

      1030       1040       1050       1060       1070       1080
QGFFSSPSTS RTPLLSSLSA TSNNSTVACI DRNGLQSCPI KEDSFLQRYS SDPTGALTED

      1090       1100       1110       1120       1130       1140
SIDDTFLPVP EYINQSVPKR PAGSVQNPVY HNQPLNPAPS RDPHYQDPHS TAVGNPEYLN

      1150       1160       1170       1180       1190       1200
TVQPTCVNST FDSPAHWAQK GSHQISLDNP DYQQDFFPKE AKPNGIFKGS TAENAEYLRV

      1210
APQSSEFIGA
```

P00533 in FASTA format

View entry in original UniProtKB/Swiss-Prot format
View entry in raw text format (no links)
Report form for errors/updates in this UniProtKB/Swiss-Prot entry

BLAST submission on ExPASy/SIB
or at NCBI (USA)

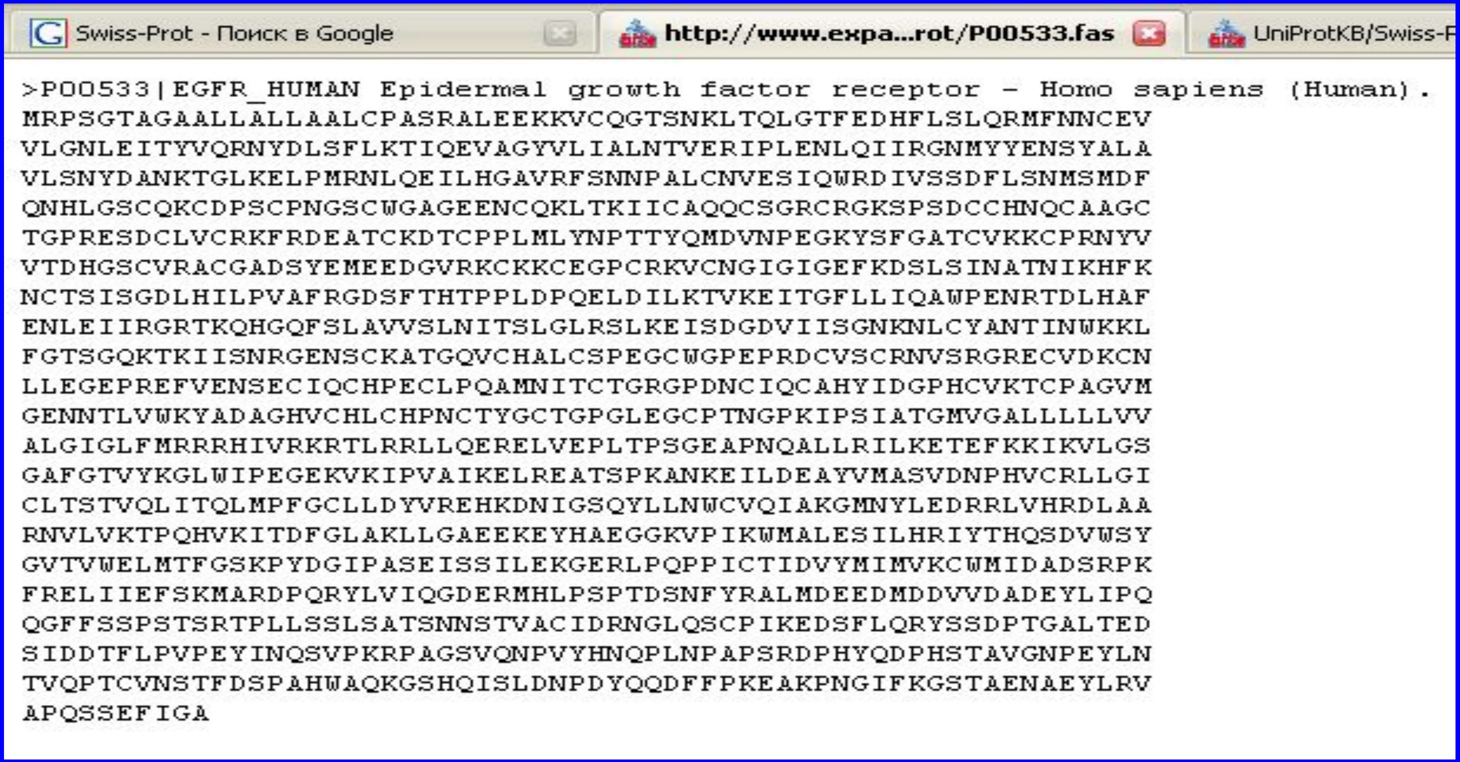
Sequence analysis tools: ProtParam, ProtScale, Compute pI/Mw, PeptideMass, PeptideCutter, Dotlet (Java)

ScanProsite, MotifScan

Submit a homology modeling request to SWISS-MODEL

NPSA Sequence analysis tools

FASTA format

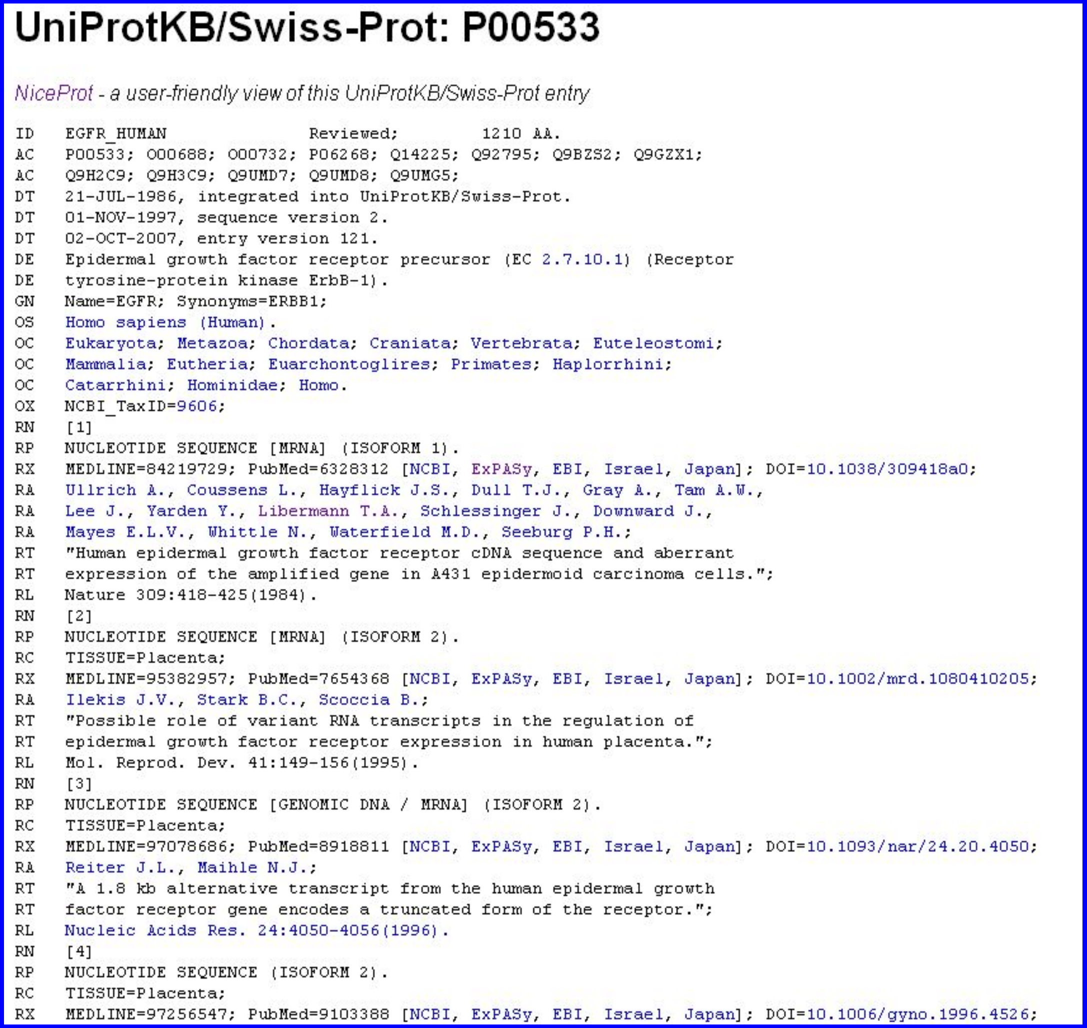


```
>P00533|EGFR_HUMAN Epidermal growth factor receptor - Homo sapiens (Human).
MRPSGTAGAALLALLAALCPASRALEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNNCEV
VLGNLEITYVQRNYDLSFLKTIQEVAGYVLIALNTVERIPLENLQIIRGNMYYENSYALA
VLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSMDF
QNHLGSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRGKSPSDCCHNQCAAGC
TGPRESDCLVCRKFRDEATCKDTCPPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYV
VTDHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFK
NCTSISGDLHILPVAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAF
ENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISGNKNLCYANTINWKKL
FGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCN
LLEGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPAGVM
GENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPTNGPKIPSIATGMVGALLLLLVV
ALGIGLFMRRRHIVRKRTLRRLLQERELVEPLTPSGEAPNQALLRILKETEFKKIKVLGS
GAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGI
CLTSTVQLITQLMPFGCLLDYVREHKDNIGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAA
RNVLVKTPQHVKITDFGLAKLLGAEEKEYHAEGGKVPIKWMALESILHRIYTHQSDVWSY
GVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVKCWMIDADSRPK
FRELIIEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDMDDVVDADEYLIPQ
QGFFSSPSTSRTPLLSSLSATSNNSTVACIDRNGLQSCPIKEDSFLQRYSSDPTGALTED
SIDDTFLPVPEYINQSVPKRPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLN
TVQPTCVNSTFDSPAHWAQKGSHQISLDNPDYQQDFFPKEAKPNGIFKGSTAENAEYLRV
APQSSEFIGA
```

Программа FASTA (1988, WR Pearson & DJ Lipman):
>(the definition line)_уникальный_ID + короткое описание
ПОСЛЕДОВАТЕЛЬНОСТЬ БЕЛКА (ИЛИ ДНК) В ОДНОБУКВЕННОМ КОДЕ

RAW format – без definition line

NiceProt view



UniProtKB/Swiss-Prot: P00533

NiceProt - a user-friendly view of this UniProtKB/Swiss-Prot entry

```
ID   EGFR_HUMAN              Reviewed;        1210 AA.
AC   P00533; O00688; O00732; P06268; Q14225; Q92795; Q9BZS2; Q9GZX1;
AC   Q9H2C9; Q9H3C9; Q9UMD7; Q9UMD8; Q9UMG5;
DT   21-JUL-1986, integrated into UniProtKB/Swiss-Prot.
DT   01-NOV-1997, sequence version 2.
DT   02-OCT-2007, entry version 121.
DE   Epidermal growth factor receptor precursor (EC 2.7.10.1) (Receptor
DE   tyrosine-protein kinase ErbB-1).
GN   Name=EGFR; Synonyms=ERBB1;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
OC   Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE [MRNA] (ISOFORM 1).
RX   MEDLINE=84219729; PubMed=6328312 [NCBI, ExPASy, EBI, Israel, Japan]; DOI=10.1038/309418a0;
RA   Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
RA   Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
RA   Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
RT   "Human epidermal growth factor receptor cDNA sequence and aberrant
RT   expression of the amplified gene in A431 epidermoid carcinoma cells.";
RL   Nature 309:418-425(1984).
RN   [2]
RP   NUCLEOTIDE SEQUENCE [MRNA] (ISOFORM 2).
RC   TISSUE=Placenta;
RX   MEDLINE=95382957; PubMed=7654368 [NCBI, ExPASy, EBI, Israel, Japan]; DOI=10.1002/mrd.1080410205;
RA   Ilekis J.V., Stark B.C., Scoccia B.;
RT   "Possible role of variant RNA transcripts in the regulation of
RT   epidermal growth factor receptor expression in human placenta.";
RL   Mol. Reprod. Dev. 41:149-156(1995).
RN   [3]
RP   NUCLEOTIDE SEQUENCE [GENOMIC DNA / MRNA] (ISOFORM 2).
RC   TISSUE=Placenta;
RX   MEDLINE=97078686; PubMed=8918811 [NCBI, ExPASy, EBI, Israel, Japan]; DOI=10.1093/nar/24.20.4050;
RA   Reiter J.L., Maihle N.J.;
RT   "A 1.8 kb alternative transcript from the human epidermal growth
RT   factor receptor gene encodes a truncated form of the receptor.";
RL   Nucleic Acids Res. 24:4050-4056(1996).
RN   [4]
RP   NUCLEOTIDE SEQUENCE (ISOFORM 2).
RC   TISSUE=Placenta;
RX   MEDLINE=97256547; PubMed=9103388 [NCBI, ExPASy, EBI, Israel, Japan]; DOI=10.1006/gyno.1996.4526;
```