

# Βιοπληροφορική

Βάσεις Δεδομένων  
1ο εργαστήριο

Γρηγόρης Αμούτζιας

# Βάσεις Δεδομένων: Εισαγωγή

Χρησιμοποιούνται για:

- Οργάνωση
- Αποθήκευση
- Επεξεργασία
- Αναζήτηση/επαναπόκτηση της βιολογικής πληροφορίας

Κύρια είδη:

Επίπεδης οργάνωσης (Flat-files:) Το ποιο απλό είδος. Ουσιαστικά είναι κατάλογοι

Σχεσιακές βάσεις. Πιο περίπλοκες και πλέον πολύ διαδεδομένες . Π.χ., SQL. Η πληροφορία οργανώνεται σε πίνακες που σχετίζονται μεταξύ τους. Έτσι αποφεύγεται η επανάληψη και συσσώρευση δεδομένων

Αντικειμενοστρεφείς βάσεις κ.α.

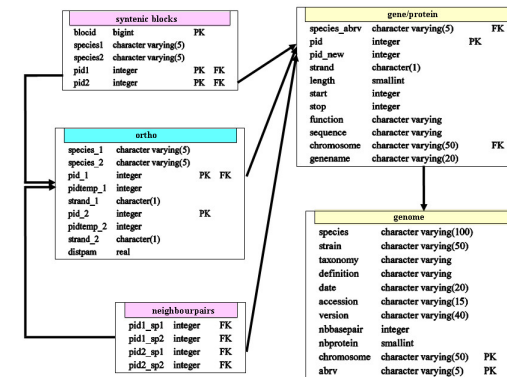
Διακρίνονται κυρίως σε αρχειακές/πρωτεύοντες και δευτερεύοντες

Στις αρχειακές γίνεται κατάθεση δεδομένων ενώ στις δευτερεύοντες τα δεδομένα είναι περαιτέρω επεξεργασμένα/σχολιασμένα/αλληλοσυνδεδεμένα

```

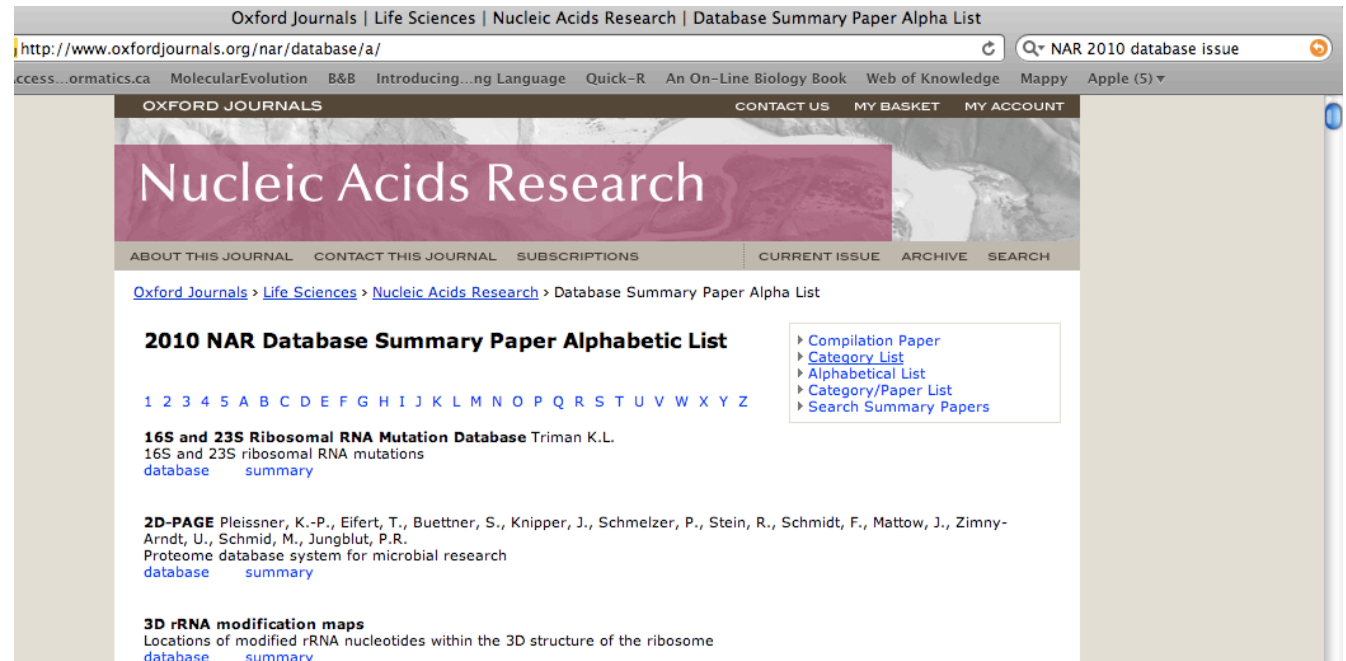
LOCUS       name of locus, length and type of sequence,
            classification of organism, data of entry
DEFINITION  description of entry
ACCESSION   accession numbers of original source
KEYWORDS    key words for cross referencing this entry
SOURCE      source organism of DNA
ORGANISM    description of organism
REFERENCE
COMMENT     biological function or database information
FEATURES   information about sequence by base position or range of positions
            source          range of sequence, source organism
            misc_signal     range of sequence, type of function or signal
            mRNA           range of sequence, mRNA
            CDS            range of sequence, protein coding region
            intron         range of sequence, position of intron
            mutation       sequence position, change in sequence for mutation
BASE COUNT  count of A, C, G, T and other symbols
ORIGIN      text indicating start of sequence
            1 gaattcgata aatctctggt ttattgtgca gtttatggtt ccaaaatcgc
            51 atatactcac agcataaactg tatatacaacc cagggggcgg aatgaaagcg
            //
    
```

Figure 2.5. GenBank DNA sequence entry.



# Ετήσιος κατάλογος Β.Δ.

- Κάθε Ιανουάριο στο Nucleic Acids Research (Special database issue)
- 2010: 58 νέες και 73 ανανεωμένες
- Σύνολο: 1230
- 5% ετήσια ανάπτυξη
- Επίσης υπάρχει το περιοδικό Database: the journal of biological databases and curation



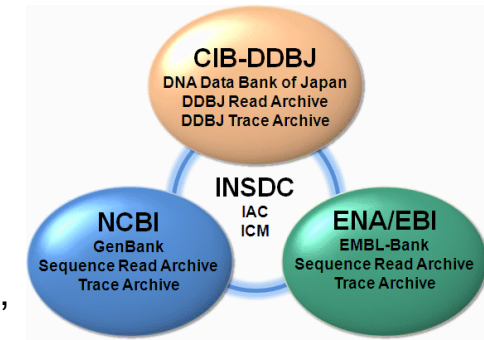
The screenshot displays the Oxford Journals website interface. At the top, the navigation bar includes 'Oxford Journals | Life Sciences | Nucleic Acids Research | Database Summary Paper Alpha List'. The browser address bar shows 'http://www.oxfordjournals.org/nar/database/a/'. The main header features the 'Nucleic Acids Research' logo. Below the header, there are navigation links for 'ABOUT THIS JOURNAL', 'CONTACT THIS JOURNAL', 'SUBSCRIPTIONS', 'CURRENT ISSUE', 'ARCHIVE', and 'SEARCH'. The main content area is titled '2010 NAR Database Summary Paper Alphabetic List' and includes a list of databases with their respective authors and links to 'database' and 'summary' pages. The databases listed are: '16S and 23S Ribosomal RNA Mutation Database' by Triman K.L., '2D-PAGE' by Pleissner, K.-P., Eifert, T., Buettner, S., Knipper, J., Schmelzer, P., Stein, R., Schmidt, F., Mattow, J., Zimny-Arndt, U., Schmid, M., Jungblut, P.R., and '3D rRNA modification maps'.

<http://www.oxfordjournals.org/nar/database/a/>

<http://www.oxfordjournals.org/nar/database/c/>

# Βάσεις νουκλεοτιδικών δεδομένων (I)

- Αρχειακές ΒΔ για νουκλεοτιδικές αλληλουχίες:
  - EMBL-BANK. European Nucleotide Archive (ENA), EBI. Hinxton, UK.
  - GENBANK. NCBI, NIH. Bethesda, USA
  - DNA databank of Japan (DDBJ). National institute of Genetics, Mishima, JP
- Η ακολουθία κατατίθεται σε μία απο τις ΒΔ, η οποία έχει και την δυνατότητα να την αναθεωρήσει (μόνο αυτή, για αποτροπή 'συγκρούσεων')
- Και οι 3 ΒΔ ανήκουν στο International nucleotide sequence database collection (INSDC). Κάθε μέρα ανταλλάσσουν δεδομένα. Η ίδια ακολουθία Χ3. Νέα έκδοση ανά δίμηνο.
- Από το 2009, το INSDC ξεκίνησε να καταχωρεί και αμορφοποίητα δεδομένα από μεγάλης κλίμακας αλληλουχίσεις (Sequencing projects), είτε αυτά προέρχονται από κλασσικές μεθόδους αλληλούχισης (Trace archive) (capillary sequencing), είτε από μεθόδους αλληλούχισης 2ης γενιάς (Read Archive) (454, Solexa, Solid, Helicos)



# EMBL bank help page

[http://www.ebi.ac.uk/embl/Documentation/User\\_manual/usrman.html](http://www.ebi.ac.uk/embl/Documentation/User_manual/usrman.html)

Note that each line begins with a two-character line code, which indicates the type of information contained in the line. The currently used line types, along with their respective line codes, are listed below:

ID - identification	(begins each entry; 1 per entry)
AC - accession number	(>=1 per entry)
PR - project identifier	(0 or 1 per entry)
DT - date	(2 per entry)
DE - description	(>=1 per entry)
KW - keyword	(>=1 per entry)
OS - organism species	(>=1 per entry)
OC - organism classification	(>=1 per entry)
OG - organelle	(0 or 1 per entry)
RN - reference number	(>=1 per entry)
RC - reference comment	(>=0 per entry)
RP - reference positions	(>=1 per entry)
RX - reference cross-reference	(>=0 per entry)
RG - reference group	(>=0 per entry)
RA - reference author(s)	(>=0 per entry)
RT - reference title	(>=1 per entry)
RL - reference location	(>=1 per entry)
DR - database cross-reference	(>=0 per entry)
CC - comments or notes	(>=0 per entry)
AH - assembly header	(0 or 1 per entry)
AS - assembly information	(0 or >=1 per entry)
FH - feature table header	(2 per entry)
FT - feature table data	(>=2 per entry)
XX - spacer line	(many per entry)
SQ - sequence header	(1 per entry)
CO - contig/construct line	(0 or >=1 per entry)
bb - (blanks) sequence data	(>=1 per entry)
// - termination line	(ends each entry; 1 per entry)

Note that some entries will not contain all of the line types, and some line types occur many times in a single entry. As indicated, each entry begins with an identification line (ID) and ends with a terminator line (//). The various line types appear in entries in the order in which they are listed above (except for XX lines which may appear anywhere between the ID and SQ lines). A detailed description of each line type is given in the following sections.

# Βάσεις νουκλεοτιδικών δεδομένων. EMBL format (i)

```
ID X56734; SV 1; linear; mRNA; STD; PLN; 1859 BP.
XX
AC X56734; S46826;
XX
DT 12-SEP-1991 (Rel. 29, Created)
DT 25-NOV-2005 (Rel. 85, Last updated, Version 11)
XX
DE Trifolium repens mRNA for non-cyanogenic beta-glucosidase
XX
KW beta-glucosidase.
XX
OS Trifolium repens (white clover)
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Trifolium.
XX
RN [ 5]
RP 1-1859
RX PUBMED; 1907511.
RA Oxtoby E., Dunn M.A., Pancoro A., Hughes M.A.;
RT "Nucleotide and derived amino acid sequence of the cyanogenic
RT beta-glucosidase (linamarase) from white clover (Trifolium repens L.)";
RL Plant Mol. Biol. 17(2):209-219(1991).
XX
RN [ 6]
RP 1-1859
RA Hughes M.A.;
RT ;
RL Submitted (19-NOV-1990) to the EMBL/GenBank/DDBJ databases.
RL Hughes M.A., University of Newcastle Upon Tyne, Medical School, Newcastle
RL Upon Tyne, NE2 4HH, UK
XX
```

---

# Βάσεις νουκλεοτιδικών δεδομένων.

## EMBL format (ii)

```
FH   Key                Location/Qualifiers
FH
FT   source             1..1859
FT                       /organism="Trifolium repens"
FT                       /mol_type="mRNA"
FT                       /clone_lib="lambda gt10"
FT                       /clone="TRE361"
FT                       /tissue_type="leaves"
FT                       /db_xref="taxon:3899"
FT   CDS                14..1495
FT                       /product="beta-glucosidase"
FT                       /EC_number="3.2.1.21"
FT                       /note="non-cyanogenic"
FT                       /db_xref="GOA:P26204"
FT                       /db_xref="HSSP:P26205"
FT                       /db_xref="InterPro:IPR001360"
FT                       /db_xref="UniProtKB/Swiss-Prot:P26204"
FT                       /protein_id="CAA40058.1"
FT                       /translation="MDFIVAIFALFVISSFTITSTNAVEASTLLDIGNLSRSSFPARGFI
FT                       FGAGSSAYQFEGAVNEGGRGPSIWDTFTHKYPEKIRDGNSADITVDQYHRYKEDVGIMK
FT                       DQNMDSYRFSISWPRILPKGKLSGGINHEGIKYYNNLINELLANGIQPFVTLFHWDLPO
FT                       VLEDEYGGFLNSGVINDFRDYTDLCFKEFGDRVRYWSTLNEPWVFSNSGYALGTNAPGR
FT                       CSASNVAKPGDSGTGPYIVTHNQILAHAEAVHVYKTKYQAYQKGIKIGITLVSNWLMPLD
FT                       DNSIPDIKAAERSLDFQFGLFMEQLTTGDYSKSMRRIVKNRLLPKFSKFESSLVNGSDFD
FT                       IGINYSSSYISNAPSHGNAKPSYSTNPMTNISFEKHGIPLGPRASIWIYVYPYMFQ
FT                       EDFEIFCYILKINITILQFSITENGMNEFNATLPVEEALLNTYRIDYVYRHLYYIRSA
FT                       IRAGSNVKGIFYAWSFLDCNEWFAGFTVRFGLNFVD"
FT   mRNA              1..1859
FT                       /experiment="experimental evidence, no additional details
FT                       recorded"
XX
```

# Βάσεις νουκλεοτιδικών δεδομένων. EMBL format (ii)

```
SQ   Sequence 1859 BP; 609 A; 314 C; 355 G; 581 T; 0 other;
aaacaaacca aatatggatt ttattgtagc catatattgct ctgtttgta ttagctcatt      60
cacaattact tccacaaatg cagttgaagc ttctactcct cttgacatag gtaacctgag      120
tcggagcagt tttcctcgtg gcttcatcct tgggtcgtga tcttcagcat accaatttga      180
aggtgcagta aacgaaggcg gtagaggacc aagtatttgg gataccttca cccataaata      240
tccagaaaaa ataagggatg gaagcaatgc agacatcacg gttgaccaat atcaccgcta      300
caaggaagat gttgggatta tgaaggatca aaatatggat tcgtatagat totcaatctc      360
ttggccaaga atactcccaa agggaaagt ttagcggaggc ataaatcacg aaggaatcaa      420
atattacaac aaccttatca acgaactatt ggctaacggg atacaaccat ttgtaactct      480
ttttcattgg gatcttcccc aagtcttaga agatgagtat ggtggtttct taaactcggg      540
tgtaataaat gattttcgag actatacggg tctttgcttc aaggaatttg gagatagagt      600
gaggtattgg agtactctaa atgagccatg ggtgtttagc aattctggat atgcactagg      660
aacaaatgca ccaggctgat gttcggcctc caacgtggcc aagcctgggtg attctggaac      720
aggaccttat atagttacac acaatcaaat tcttgctcat gcagaagctg tacatgtgta      780
taagactaaa taccaggcat atcaaaaggg aaagataggc ataacgttgg tatctaactg      840
gttaatgcca cttgatgata atagcatacc agatataaag gctgccgaga gatcacttga      900
cttocaattt ggattgttta tggacaacatt aacaacagga gattattcta agagcatgog      960
gogtatagtt aaaaaccgat tacctaagtt ctcaaaattc gaatcaagcc tagtgaatgg     1020
ttcatttgat tttattggta taaactatta ctcttctagt tatattagca atgccccttc     1080
acatggcaat gccaaaccca gttactcaac aaatcctatg accaatattt catttgaaaa     1140
acatgggata cccttaggtc caagggctgc tcoaatttgg atatatgttt atccatatat     1200
gtttatocaa gaggacttgc agatcttttg ttacatatta aaaataaata taacaatcct     1260
gcaattttca atcactgaaa atggtatgaa tgaattcaac gatgcaacac ttccagtaga     1320
agaagctcct ttgaataact acagaattga ttactattac cgtcacttat actacattcg     1380
ttctgcaatc agggctggct caaatgtgaa gggtttttac gcatggctat ttttggactg     1440
taatgaatgg tttgcaggct ttactgttgc ttttggatta aactttgtag attagaaaaga     1500
tggattaaaa aggtacccta agctttctgc ccaatggtag aagaactttc tcaaaagaaa     1560
ctagctagta ttattaaag aactttgtag tagattacag tacatcgttt gaagttgagt     1620
tgggtcacct aattaaataa aagaggttac tottaacata tttttaggcc attcgttgtg     1680
aagttgttag gctgttatct ctattatact atgttgtagt aataagtgca ttgttgtacc     1740
agaagctatg atcataacta taggttgatc cttcatgtat cagtttgatg ttgagaatac     1800
tttgaattaa aagtcttttt ttattttttt aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa     1859
```

//





# Βάσεις νουκλεοτιδικών δεδομένων. FASTA format

```
>ENA|X56734|X56734.1 Trifolium repens mRNA for non-cyanogenic beta-glucosidase
aaacaaaccaaataatggatTTTTATTGTAGCCATATTTGCTCTGTTTGTTATTAGCTCATT
cacaattacttccacaaatgcagttgaagcttctactcttcttgacataggtaacctgag
tcgggagcagtttccctcgtggcttcatctttgggtgctggatcttcagcataccaatttga
aggtgcagtaaacgaaggcggtagaggaccaagtatttgggataccttcaccataaata
tccagaaaaaataagggatggaagcaatgcagacatcacggttgaccaatatcacgcta
caaggaagatgttgggattatgaaggatcaaaatatggattcgtatagattctcaatctc
ttggccaagaatactcccaaagggaaagttgagcggaggcataaatcacgaaggaatcaa
atattacaacaaccttatcaacgaactattggctaacggtatacaaccatttgaactct
ttttcattgggatcttccccaaagtcttagaagatgagtatggtggtttcttaaactccgg
tgtaataaatgattttcagagactatacggatctttgcttcaaggaatttggagatagagt
gaggtattggagtactctaaatgagccatgggtggttagcaattctggatagcactagg
aacaatgcaccaggtcgatgctcggcctccaacgctggccaagcctggtgattctggaac
aggaccttatatagttacacacaatcaaattcttgctcatgcagaagctgtacatgtgta
taagactaaataccaggcatatcaaaagggaaagataggcataacgttggtatctaactg
gttaatgccacttgatgataatagcataaccagatataaaggctgccgagagatcacttga
cttccaatttggattgtttatggaacaattaacaacaggagattattctaagagcatgcg
gcgtatagttaaaaaccgattacctaagttctcaaaattcgaatcaagcctagtgaatgg
ttcatttgattttattggtataaactattactcttctagttatattagcaatgcccttc
acatggcaatgccaaaccagttactcaacaaatcctatgaccaatatttcatttgaaa
acatgggatacccttaggtccaagggctgcttcaatttggatataatgtttatccatata
gtttatccaagaggacttcgagatcttttggtacatataaaaaataaatataacaatcct
gcaattttcaatcactgaaaatggtatgaatgaattcaacgatgcaacacttccagtaga
agaagctctttgaatacttacagaattgattaactattaccgtcacttatactacattcg
ttctgcaatcagggtggctcaaatgtgaagggttttacgcatggtcatttttgactg
taatgaatggtttgcaggcttactgttctgttttgattaaactttgtagattagaaaga
tggattaaaaaggtaccctaagcttctgccaatggtacaagaactttctcaaaagaaa
ctagctagtattattaaagaactttgtagtagattacagtacatcgtttgaagttgagt
tgggtgcacctaatattaaataaaagaggttactcttaacatatttttaggccattcgttgtg
aagttgttaggctgttatttctattatactatggttagtaataagtgcatgttgtgacc
agaagctatgatcataactatagggtgatccttcatgtatcagtttgatgttgagaatac
tttgaattaaaagtctttttttatttttttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa
```

---

# Παράδειγμα εγγραφής στην EMBL bank

- Πηγαίνετε στην διεύθυνση:  
<http://www.ebi.ac.uk/embl/>
- X03635 : Estrogen receptor alpha, Human
- Αναζητήστε την ακολουθία του παραδείγματος χρησιμοποιώντας το accession number του (X03635).
- Δείτε το Nucleotide Sequence του mRNA σε μορφή ENA και σε μορφή EMBL format.
- Στην μορφή ENA, δείτε την ακολουθία ως FASTA format.

**All results (12)**[Genomes \(2\)](#)[Nucleotide Sequences \(7\)](#)[Protein Sequences \(2\)](#)[Gene Expression \(1\)](#) **ADVANCED SEARCH** **QUERY SUGGESTIONS**EBI > Search for **X03635** in *All results***Nucleotide Sequences / EMBL Release (Normal Divisions)****X03635**

Homo sapiens mRNA for oestrogen receptor

View: [in ENA](#) [in EMBL format](#) [in SRS](#) [in EMBL-SVA](#) [Launch NCBI BLAST](#) [Launch FASTA](#)References: [Taxonomy](#) [InterPro](#) [Ensembl Gene](#) [UniProtKB](#) [EMBL-Bank \(Coding Sequence\)](#) [PDBe](#) [HGNC](#) [Medline](#)[▶ View all 4 results...](#)**Genomes / HGNC****HGNC:3467**Approved Symbol: ESR1 [▶ Discover more about this gene...](#)

Approved Name: estrogen receptor 1

Status: Approved

Aliases: NR3A1 Era

Locus Type: gene with protein product

Chromosome: 6q24-q27

References: [Medline](#) [UniProtKB](#) [Ensembl Gene](#) [EMBL-Bank](#)**Genomes / Ensembl Gene****ENSG00000091831** [▶ Discover more about this gene...](#)

estrogen receptor 1 [Source:HGNC Symbol;Acc:3467]

Species: Homo sapiens

References: [Taxonomy](#) [Ensembl Genomes Gene](#) [UniProtKB](#) [Ensembl](#) [OMIM](#) [GO](#) [PDBe](#) [HGNC](#) [EMBL-Bank](#)**Nucleotide Sequences / ASTD****TRAN00000039246**

Homo sapiens transcript, product of gene ENSG00000091831 with CDS and translation.

References: [Ensembl](#) [Taxonomy](#) [EMBL-Bank](#)**Nucleotide Sequences / EMBL-Bank (Coding Sequence)****CAD97416**

Homo sapiens (human) hypothetical protein

View: [in ENA](#) [in EMBL format](#) [in SRS](#) [Launch NCBI BLAST](#) [Launch FASTA](#)References: [EMBL-Bank \(Coding Sequence\)](#) [UniProtKB](#) [EMBL-Bank](#)



- ENA Home
- Search & Browse
- Submit & Update
- About ENA
- Contact

Text search

Sequence search

Enter or paste text or ENA accession number:

Search

Upload file of accessions:

Choose File no file selected

Search

EMBL-Bank: X03635.1 : Homo sapiens mRNA for oestrogen receptor

View: [TEXT](#) [FASTA](#) [XML](#)Download: [TEXT](#) [FASTA](#) [XML](#)[Overview](#) [Source Feature\(s\)](#) [Other Features](#) [Assembly](#) [References](#) [Comments](#) [Sequence](#)[Send Feedback](#)**Organism**

Homo sapiens

**Molecule type**

mRNA

**Topology**

linear

**Data class**

STD

**Taxonomic Division**

HUM

**Sequence length**

6,450

**Sequence Version**

1

**First public**

18-NOV-1986

**Last updated**

07-OCT-2008

**Keywords**

estrogen receptor, receptor, steroid hormone receptor.

**Secondary Accession(s)**

M11457.

**Lineage**[Eukaryota](#), [Metazoa](#), [Chordata](#), [Craniata](#), [Vertebrata](#), [Euteleostomi](#), [Mammalia](#), [Eutheria](#), [Euarchontoglires](#), [Primates](#), [Haplorrhini](#), [Catarrhini](#), [Hominidae](#), [Homo](#)**Navigation**

# Βάσεις πρωτεϊνικών δεδομένων

- Swissprot. 1987, Uni Geneva + SIB. Σχολιασμός των εγγραφών/ πρωτεϊνών από επιστήμονες.
- TrEMBL. 1996. SIB + EBI. Αυτόματη μετάφραση των ακολουθιών που βρίσκονται στην EMBL. Δεδομένα στην ίδια μορφή με την Swissprot. Μπορεί να είναι υποθετικές ή ο σχολιασμός να μην είναι εκτενής, όπως στην Swissprot.
- PIR. 1984, USA
- UniProt. 2002. Ενώθηκαν οι παραπάνω βάσεις.
- UniMes: για μεταγενωμικά δεδομένα, όπου δεν γνωρίζουμε από ποιά είδη προέρχονται οι ακολουθίες.

# Swissprot (I)

- Από την εγγραφή του προηγούμενου παραδείγματος, ακολουθήστε τον σύνδεσμο (link) προς την Β.Δ. UniprotKB/Swissprot, με κωδικό εγγραφής P03372

<b>CDS</b>	<b>361..2148</b>
<b>product</b>	oestrogen receptor
<b>translation</b>	MTMTLHTKASGMALLHQIQGNELEPLNRPQLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYGQTGLPY GPGSEAAAFSGNGLGGFPPLNSVSPSPMLMLLHPPQLSPFLOPHGQQVPPYLENEPSGYTVREAGPPAFYRPNSDNRRQG GRERLASTNDKGSMAESAKETRYCAVCNDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQAC RLRKCCEVGMKGGIRKDRRGGRMLKHKRQRDDGEGRGEVGSAGDMRAANLWPSPLMIKRSKKNLALSITADQMVSA DAEPPILYSEYDPTRFSEASMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQVHLLCWLLEILMIGLVWRSMEHPV KLLFAPNLLDRNQKCEGMVEIFDMLLATSSRFMMNLQGEFVCLKSIILLNSGVYTFLLSSTLKSLEEKDHIHRVLD KITDTLIHLMKAGLTLOQQHQRLAQLLLILSHIRHMSNKGMEHLYSMCKKNVPLYDLLEMLDAHRLHAPTSRGGASV EETDQSHLATAGSTSSHSLQKYYITGEAEGFPATV
↓ EMBL-Bank CDS:	<a href="#">CAA27284</a>
→ GOA	<a href="#">P03372</a>
→ HGNC	<a href="#">3467</a>
→ InterPro	<a href="#">IPR000536</a> , <a href="#">IPR001292</a> , <a href="#">IPR001628</a> , <a href="#">IPR001723</a> , <a href="#">IPR008946</a> , <a href="#">IPR013088</a> , <a href="#">IPR024178</a>
→ PDB	<a href="#">1A52</a> , <a href="#">1AKE</a> , <a href="#">1ERE</a> , <a href="#">1ERR</a> , <a href="#">1G50</a> , <a href="#">1GWQ</a> , <a href="#">1GWR</a> , <a href="#">1HCP</a> , <a href="#">1HCQ</a> , <a href="#">1L2I</a> , <a href="#">1PCG</a> , <a href="#">1QKT</a> , <a href="#">1QKU</a> , <a href="#">1R5K</a> , <a href="#">1SJ0</a> , <a href="#">1UOM</a> , <a href="#">1X7E</a> , <a href="#">1X7R</a> , <a href="#">1XP1</a> , <a href="#">1XP6</a> , <a href="#">1XP9</a> , <a href="#">1XPC</a> , <a href="#">1XQC</a> , <a href="#">1YIM</a> , <a href="#">1YIN</a> , <a href="#">1ZKY</a> , <a href="#">2AYR</a> , <a href="#">2B1V</a> , <a href="#">2B1Z</a> , <a href="#">2B23</a> , <a href="#">2BJ4</a> , <a href="#">2FAI</a> , <a href="#">2G44</a> , <a href="#">2G50</a> , <a href="#">2I0J</a> , <a href="#">2IOG</a> , <a href="#">2I0K</a> , <a href="#">2JF9</a> , <a href="#">2JFA</a> , <a href="#">2OCF</a> , <a href="#">2OUZ</a> , <a href="#">2P15</a> , <a href="#">2POG</a> , <a href="#">2Q6J</a> , <a href="#">2Q70</a> , <a href="#">2QA6</a> , <a href="#">2QA8</a> , <a href="#">2QAB</a> , <a href="#">2QE4</a> , <a href="#">2QGT</a> , <a href="#">2QGW</a> , <a href="#">2QH6</a> , <a href="#">2QR9</a> , <a href="#">2QSE</a> , <a href="#">2QXM</a> , <a href="#">2QXS</a> , <a href="#">2QZO</a> , <a href="#">2R6W</a> , <a href="#">2R6Y</a> , <a href="#">2YAT</a> , <a href="#">3CBM</a> , <a href="#">3CBO</a> , <a href="#">3CBP</a> , <a href="#">3DT3</a> , <a href="#">3ERD</a> , <a href="#">3ERT</a> , <a href="#">3HLV</a> , <a href="#">3HM1</a> , <a href="#">3L03</a> , <a href="#">3OS8</a> , <a href="#">3OS9</a> , <a href="#">3OSA</a>
→ UniProtKB/Swiss-Prot	<b>P03372</b>

# Swissprot (I)

– Δείτε

- το όνομα και τα συνώνυμα της ακολουθίας
- Την ταξινόμηση του οργανισμού. Η ταξινόμηση μπορεί επίσης να βρεθεί και στην ιστοσελίδα του NCBI taxonomy <http://www.ncbi.nlm.nih.gov/>
- Λειτουργίες της πρωτεΐνης (και στο τμήμα των Ontologies)
- Την ακολουθία σε FASTA format
- Ακολουθείστε το σύνδεσμο (Hs.208124) προς την Β.Δ. Unigene και από εκεί δείτε το προφίλ γονιδιακής έκφρασης μέσω του link 'EST profile'
- Από την προηγούμενη ιστοσελίδα του Uniprot, ακολουθείστε το σύνδεσμο P03372 προς την Β.Δ. Intact (στο τμήμα protein-protein interaction databases) για να δείτε πόσες πρωτεϊνικές αλληλεπιδράσεις έχει το estrogen receptor alpha.

Search

Blast \*

Align \*

Retrieve

ID Mapping \*

Search in


Query

Protein Knowledgebase (UniProtKB) ▾


Search

Advanced Search »

Clear

**P03372 (ESR1\_HUMAN)** ★ Reviewed, UniProtKB/Swiss-ProtLast modified September 21, 2011. Version 176.  [This entry in the past...](#)

Contribute

 [Send feedback](#) [Read comments \(0\) or add your own](#) Clusters with 100%, 90%, 50% identity |  Documents (6) |  Third-party data[text](#) [xml](#) [rdf/xml](#) [gff](#) [fasta](#)[Names](#) · [Attributes](#) · [General annotation](#) · [Ontologies](#) · [Interactions](#) · [Alt products](#) · [Sequence annotation](#) · [Sequences](#) · [References](#) · [Web links](#) · [Cross-refs](#) · [Entry info](#) · [Documents](#) [Customize order](#)

## Names and origin

Protein names	<i>Recommended name:</i> <b>Estrogen receptor</b> Short name=ER <i>Alternative name(s):</i> ER-alpha Estradiol receptor Nuclear receptor subfamily 3 group A member 1
Gene names	Name: <b>ESR1</b> Synonyms:ESR, NR3A1
Organism	<a href="#">Homo sapiens (Human)</a>
Taxonomic identifier	<a href="#">9606 [NCBI]</a>
Taxonomic lineage	<a href="#">Eukaryota</a> › <a href="#">Metazoa</a> › <a href="#">Chordata</a> › <a href="#">Craniata</a> › <a href="#">Vertebrata</a> › <a href="#">Euteleostomi</a> › <a href="#">Mammalia</a> › <a href="#">Eutheria</a> › <a href="#">Euarchontoglires</a> › <a href="#">Primates</a> › <a href="#">Haplorrhini</a> › <a href="#">Catarrhini</a> › <a href="#">Hominidae</a> › <a href="#">Homo</a>

## Protein attributes

Sequence length	595 AA.
Sequence status	Complete.
Protein existence	<a href="#">Evidence at protein level</a>



# ΒΔ πρωτεϊνικών επικρατειών

- Πρωτεϊνική επικράτεια: Μια περιοχή της πρωτεΐνης με συγκεκριμένη λειτουργία/δομή και καλά συντηρημένη.
- Διάφορες βάσεις δεδομένων, όπως:
  - PROSITE
  - Pfam
  - PRINTS
  - ProDom
  - SMART
  - TIGRFAMs
  - PIR superfamily
  - Superfamily
- Έχουν ενσωματωθεί στο INTERPRO.
- Το INTERPRO περιέχει πρωτεϊνικές επικράτειες. Το πρόγραμμα INTERPROscan ανιχνεύει αυτές τις επικράτειες στις πρωτεΐνες.

# Pfam

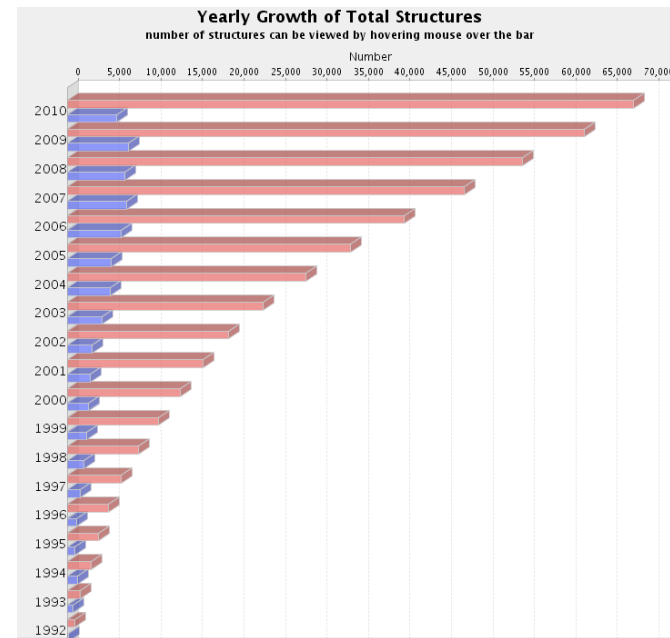
- Για την ακολουθία του Estrogen receptor alpha, από τη Uniprot ακολουθείστε τη σύνδεση για την Β.Δ. πρωτεϊνικών επικρατειών (domains) Pfam (graphical view).
- Δείτε την αρχιτεκτονική της πρωτεΐνης.
- Ποιά είναι τα βασικά domains;
- Δείτε λεπτομερέστερα την εγγραφή για το Hormone receptor / ligand binding domain.
- Δείτε σε ποιά είδη έχει βρεθεί αυτή η επικράτεια (σύνδεσμος 'species' στα αριστερά της ιστοσελίδας) (Tree).

**Family and domain databases**

InterPro	<a href="#">IPR008946</a> . Nucl_hormone_rcpt_ligand-bd. <a href="#">IPR000536</a> . Nucl_hrmn_rcpt_lig-bd_core. <a href="#">IPR001292</a> . Oestr_rcpt. <a href="#">IPR024178</a> . Oestrogen_rcpt-rel. <a href="#">IPR001723</a> . Str_hrmn_rcpt. <a href="#">IPR001628</a> . Znf_hrmn_rcpt. <a href="#">IPR013088</a> . Znf_NHR/GATA. <a href="#">[Graphical view]</a>
Gene3D	<a href="#">G3DSA:1.10.565.10</a> . Nucl_hrmn_rcpt_lig_bd. 1 hit. <a href="#">G3DSA:3.30.50.10</a> . Znf_NHR/GATA. 1 hit.
Pfam	<a href="#">PF00104</a> . Hormone_recep. 1 hit. <a href="#">PF02159</a> . Oest_recep. 1 hit. <a href="#">PF00105</a> . zf-C4. 1 hit. <a href="#">[Graphical view]</a>
PIRSF	<a href="#">PIRSF002527</a> . ER-like_NR. 1 hit.
PRINTS	<a href="#">PR00543</a> . OESTROGENR. <a href="#">PR00398</a> . STRDHORMONER. <a href="#">PR00047</a> . STROIDFINGER.
SMART	<a href="#">SM00430</a> . HOLI. 1 hit. <a href="#">SM00399</a> . ZnF_C4. 1 hit. <a href="#">[Graphical view]</a>
SUPFAM	<a href="#">SSF48508</a> . Str_ncl_receptor. 1 hit.
PROSITE	<a href="#">PS00031</a> . NUCLEAR_REC_DBD_1. 1 hit. <a href="#">PS51030</a> . NUCLEAR_REC_DBD_2. 1 hit. <a href="#">[Graphical view]</a>
ProtoNet	<a href="#">Search...</a>

# ΒΔ τρισδιάστατων δομών PDB

- Protein Data Bank (PDB)
  - Πρωτεΐνες
  - Νουκλεϊκά οξέα
  - Σύμπλοκα των παραπάνω
- Μέθοδοι
  - X-ray (~59000)
  - NMR (~8500)
  - Κρύο-ηλεκτρονική μικροσκοπία (~300)
- Οι παραπάνω μέθοδοι βρίσκουν τις συντεταγμένες (3D) των ατόμων του βιολογικού μορίου.
- Τα αρχεία με τις συντεταγμένες διαβάζονται από ειδικά προγράμματα (π.χ Rasmol) που απεικονίζουν τη δομή στο χώρο

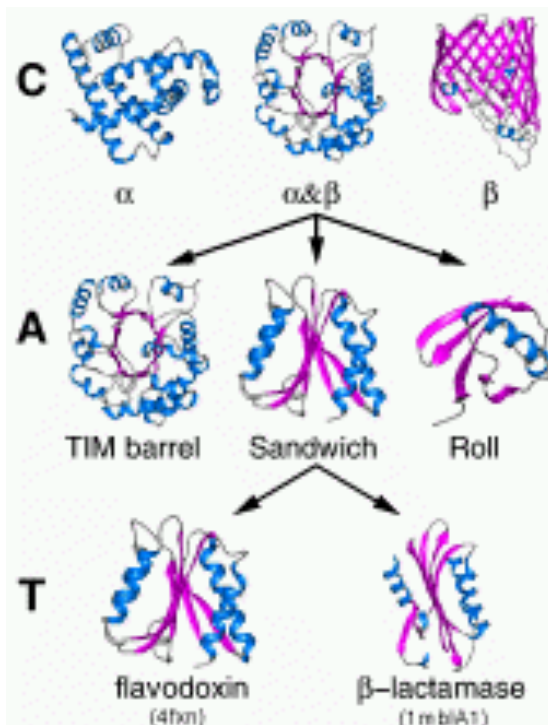


# PDB

- Από την προηγούμενη ιστοσελίδα του Uniprot για την εγγραφή estrogen receptor alpha, στο τμήμα 3D structure databases, επιλέξτε RCSB PDB και ακολουθείστε το σύνδεσμο για την 1A52 (είναι ο κωδικός εγγραφής στην PDB). Είναι η κρυσταλλική δομή της επικράτειας σε σύμπλεγμα με την οιστραδιόλη.
- Στην δεξιά πλευρά της ιστοσελίδας μπορείτε να δείτε την τρισδιάστατη δομή μέσω του συνδέσμου 'view in Jmol'.

# Β.Δ. τρισδιάστατων δομών

- CATH: κατηγοριοποιεί τις τρισδιάστατες δομές των πρωτεϊνικών επικρατειών ιεραρχικά, σε 4 βασικά επίπεδα.
- Η κατηγοριοποίηση γίνεται με ένα συνδυασμό αυτόματων μεθόδων και ανθρώπινης κρίσης.



## What do the letters "C.A.T.H.S.O.L.I.D" mean?

CATH is a tree-like, hierarchical classification that starts off at the tree "trunk" by clustering protein domains into broad categories (e.g. C, or class, where domains are clustered solely based on their general secondary structure content). As the hierarchy moves away from the "trunk" to the "branches", more stringent clustering criteria are applied to provide clusters of domains with finer granularity of similarity.

Depth	Letter	Name	Clustering criteria
1	C	Class	Secondary structure content
2	A	Architecture	General spatial arrangement of secondary structures
3	T	Topology	Spatial arrangement and connectivity of secondary structures (fold)
4	H	Homologous Superfamily	Manual curation of evidence of evolutionary relationship (at least two criteria from sequence/structure/function must be observed)
5	S	Sequence Family (S35)	>= 35% sequence similarity
6	O	Orthologous Family (S60)	>= 60% sequence similarity
7	L	"Like" domain (S95) *	>= 95% sequence similarity
8	I	Identical domain (S100)	100% sequence similarity
9	D	Domain counter	Unique domains

# Βάσεις τρισδιάστατων δομών

CATH Domain: [1cukA01](#) [XML](#)

PDB [1cuk](#), Chain A, Domain 1

CATH Code	Level Description	Links
<a href="#">2</a>	<a href="#">Mainly Beta</a>	
<a href="#">A 2.40</a>	<a href="#">Beta Barrel</a>	
<a href="#">T 2.40.50</a>	<a href="#">OB fold (Dihydrolipoamide Acetyltransferase, E2P)</a>	
<a href="#">H 2.40.50.140</a>	<a href="#">Nucleic acid-binding proteins</a>	<a href="#">Gene3D</a>
<a href="#">S 2.40.50.140.47</a>		
<a href="#">O 2.40.50.140.47.1</a>		
<a href="#">L 2.40.50.140.47.1.1</a>		
<a href="#">I 2.40.50.140.47.1.1.1</a>		
<a href="#">D 2.40.50.140.47.1.1.1.1</a>		



## What do the letters "C.A.T.H.S.O.L.I.D" mean?

CATH is a tree-like, hierarchical classification that starts off at the tree "trunk" by clustering protein domains into broad categories (e.g. C, or class, where domains are clustered solely based on their general secondary structure content). As the hierarchy moves away from the "trunk" to the "branches", more stringent clustering criteria are applied to provide clusters of domains with finer granularity of similarity.

Depth	Letter	Name	Clustering criteria
1	<a href="#">C</a>	Class	Secondary structure content
2	<a href="#">A</a>	Architecture	General spatial arrangement of secondary structures
3	<a href="#">T</a>	Topology	Spatial arrangement and connectivity of secondary structures (fold)
4	<a href="#">H</a>	Homologous Superfamily	Manual curation of evidence of evolutionary relationship (at least two criteria from sequence/structure/function must be observed)
5	<a href="#">S</a>	Sequence Family (S35)	>= 35% sequence similarity
6	<a href="#">O</a>	Orthologous Family (S60)	>= 60% sequence similarity
7	<a href="#">L</a>	"Like" domain (S95) *	>= 95% sequence similarity
8	<a href="#">I</a>	Identical domain (S100)	100% sequence similarity
9	<a href="#">D</a>	Domain counter	Unique domains

# Μεταβολικά μονοπάτια

## Metabolic and Signaling Pathways

### Enzymes and enzyme nomenclature

### Metabolic pathways

BioCarta

BioCyc

Bionemo

BioSilico

BRITE - Biomolecular Relations in Information Transmission and Expression

BSD - Biodegradative Strain Database

HMDB

HMDB - The Human Metabolome Database

KEGG - Kyoto Encyclopedia of Genes and Genomes

Klotho

LIGAND

MedicCyc

MetaCrop

MetaCyc

Metagrowth

MMCD

MODOMICS

NMPDR - National Microbial Pathogen Data Resource

Pathguide

PMAP

PUMA2

SYSTEMONAS

UM-BBD

### Protein-protein interactions

### Signalling pathways

---



# KEGG pathways

- Kyoto encyclopedia of genes and genomes.
- 2010: 374 μεταβολικά μονοπάτια.



## KEGG PATHWAY Database

Wiring diagrams of molecular interactions, reactions, and relations

KEGG2 PATHWAY BRITE DISEASE DRUG KO GENES GENOME LIGAND DBGET

Select prefix

map

Organism

Enter keywords

Go

Help

### Pathway Maps

**KEGG PATHWAY** is a collection of manually drawn pathway maps (see [new maps](#), [change history](#), and [last updates](#)) representing our knowledge on the molecular interaction and reaction networks for:

#### 0. Global Map

#### 1. Metabolism

Carbohydrate Energy Lipid Nucleotide Amino acid Other amino acid Glycan  
Cofactor/vitamin Terpenoid/PK Other secondary metabolite Xenobiotics Overview

#### 2. Genetic Information Processing

#### 3. Environmental Information Processing

#### 4. Cellular Processes

#### 5. Organismal Systems

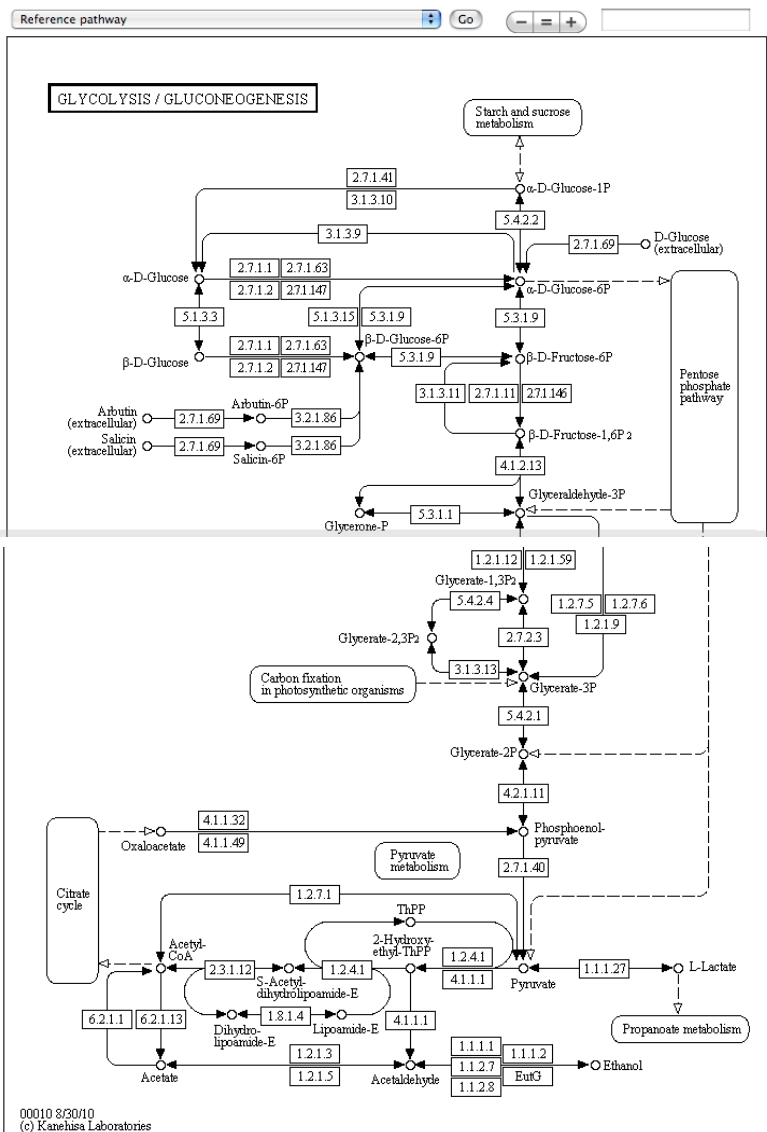
#### 6. Human Diseases

and also on the structure relationships (KEGG drug structure maps) in:

#### 7. Drug Development

**KEGG Atlas** may now be used to examine any of the KEGG pathway maps.

Glycolysis is the process of converting glucose into pyruvate and generating small amounts of ATP (energy) and NADH (reducing power). It is a central pathway that produces important precursor metabolites: six-carbon compounds of glucose-6P and fructose-6P and three-carbon compounds of glyceraldehyde-3P, phosphoenolpyruvate, and pyruvate [MD:M00001]. Acetyl-CoA, another important precursor metabolite, is produced by oxidative decarboxylation of pyruvate [MD:M00679]. When the enzyme genes of this pathway are examined in completely sequenced genomes, the reaction steps of three-carbon compounds from glyceraldehyde-3P to pyruvate form a conserved core module [MD:M00002], which is found in almost all organisms and which often corresponds to operon structures in bacterial genomes. Gluconeogenesis is a synthesis pathway of glucose from noncarbohydrate precursors. It is essentially a reversal of glycolysis with minor variations of alternative paths [MD:M00003].



# KEGG pathways

<b>Entry</b>	EC 3.1.3.9	Enzyme
<b>Name</b>	glucose-6-phosphatase; glucose 6-phosphate phosphatase	
<b>Class</b>	Hydrolases; Acting on ester bonds; Phosphoric-monoester hydrolases (BRITE hierarchy)	
<b>Synname</b>	D-glucose-6-phosphate phosphohydrolase	
<b>Reaction (IUBMB)</b>	D-glucose 6-phosphate + H <sub>2</sub> O = D-glucose + phosphate [RN:R00303]	
<b>Reaction (KEGG)</b>	R00303 > R01788 <a href="#">Show all</a>	
<b>Substrate</b>	D-glucose 6-phosphate [CPD:C00092]; H <sub>2</sub> O [CPD:C00001]	
<b>Product</b>	D-glucose [CPD:C00031]; phosphate [CPD:C00009]	
<b>Comment</b>	Wide distribution in animal tissues. Also catalyses potent transphosphorylations from carbamoyl phosphate, hexose phosphates, diphosphate, phosphoenolpyruvate and nucleoside di- and triphosphates, to D-glucose, D-mannose, 3-methyl-D-glucose or 2-deoxy-D-glucose [cf. EC 2.7.1.62 (phosphoramidate---hexose phosphotransferase), EC 2.7.1.79 (diphosphate---glycerol phosphotransferase) and EC 3.9.1.1 (phosphoamidase)].	
<b>Pathway</b>	ec00010 Glycolysis / Gluconeogenesis ec00052 Galactose metabolism ec00500 Starch and sucrose metabolism ec01100 Metabolic pathways	
<b>Orthology</b>	K01084 glucose-6-phosphatase	
<b>Genes</b>	HSA: 2538(G6PC) 57818(G6PC2) PTR: 741431(G6PC2) MCC: 712053 MMU: 14377(G6pc) RNO: 25634(G6pc) CFA: 403492(G6PC) BTA: 538710(G6PC) SSC: 100134959(G6PC)	

All links	
Ontology (5)	
KEGG BRTE (5)	
Pathway (89)	
KEGG PATHWAY (89)	
Disease (1)	
OMIM (1)	
Chemical substance (6)	
KEGG COMPOUND (6)	
Chemical reaction (12)	
KEGG ENZYME (4)	
KEGG REACTION (2)	
KEGG RPAIR (5)	
KEGG RCLASS (1)	
Genome (2)	
KEGG GENOME (2)	
Gene (58)	
KEGG ORTHOLOGY (1)	
KEGG GENES (20)	
KEGG DGENES (8)	
KEGG EGENES (29)	
Protein sequence (64)	
UniProt (25)	
PRF (3)	
RefSeq(pep) (25)	
RefSeq(nc) (1)	
DNA sequence (50)	
RefSeq(nuc) (30)	
GenBank (10)	
EMBL (10)	
Protein domain (2)	
InterPro (1)	
Pfam (1)	
Literature (3)	
PubMed (3)	
Enzyme (4)	
BRENDA (1)	
EXPASY-ENZYME (1)	
EXPLORENZ (1)	
IUBMB (1)	
All databases (296)	

<b>Entry</b>	R01788	Reaction
<b>Name</b>	alpha-D-Glucose 6-phosphate phosphohydrolase	
<b>Definition</b>	alpha-D-Glucose 6-phosphate + H <sub>2</sub> O <=> alpha-D-Glucose + Orthophosphate	
<b>Equation</b>	C00668 + C00001 <=> C00267 + C00009	
<b>RPair</b>	RP00216 C00267_C00668 main RP05676 C00001_C00009 leave RP06709 C00009_C00668 leave	
<b>Enzyme</b>	3.1.3.9	
<b>Pathway</b>	rn00010 Glycolysis / Gluconeogenesis rn00052 Galactose metabolism rn00500 Starch and sucrose metabolism rn01100 Metabolic pathways	
<b>Orthology</b>	K01084 glucose-6-phosphatase [EC:3.1.3.9]	

All links	
Ontology (2)	
KEGG BRTE (2)	
Pathway (8)	
KEGG PATHWAY (8)	
Chemical substance (4)	
KEGG COMPOUND (4)	
Chemical reaction (5)	
KEGG ENZYME (1)	
KEGG RPAIR (3)	
KEGG RCLASS (1)	
Gene (1)	
KEGG ORTHOLOGY (1)	
All databases (20)	

# KEGG

- Από την ιστοσελίδα του Uniprot για το Estrogen receptor alpha, ακολουθείστε το σύνδεσμο hsa:2099 προς τη Β.Δ. KEGG.

Genome annotation databases	
Ensembl	<a href="#">ENST00000206249</a> ; <a href="#">ENSP00000206249</a> ; <a href="#">ENSG00000091831</a> . <a href="#">ENST00000338799</a> ; <a href="#">ENSP00000342630</a> ; <a href="#">ENSG00000091831</a> . <a href="#">ENST00000440973</a> ; <a href="#">ENSP00000405330</a> ; <a href="#">ENSG00000091831</a> . <a href="#">ENST00000443427</a> ; <a href="#">ENSP00000387500</a> ; <a href="#">ENSG00000091831</a> .
GeneID	<a href="#">2099</a> .
KEGG	<a href="#">hsa:2099</a> .
UCSC	<a href="#">uc003qom.2</a> . human.

Organism-specific databases	
-----------------------------	--


- Δεξιά της νέας ιστοσελίδας (στο KEGG), ακολουθείστε το σύνδεσμο KEGG disease και στη συνέχεια το σύνδεσμο H00026 για endometrial cancer.
- Στη νέα ιστοσελίδα, στο τμήμα 'markers' δείτε ποιά γονίδια χρησιμοποιούνται ως μοριακοί δείκτες της ασθένειας.
- Στα δεξιά της ιστοσελίδας ακολουθείστε το σύνδεσμο KEGG pathways, για να δείτε το μοριακό μονοπάτι του καρκίνου του ενδομητρίου (link: hsa05213).


# Pubmed

- ΒΔ του NCBI. Ξεκίνησε τον Ιανουάριο του 1996.
- Καταχωρεί όλες τις δημοσιευμένες εργασίες που προέρχονται από τον ευρύτερο χώρο της βιοϊατρικής
- ~20 εκατομύρια εργασίες καταχωρημένες (Ιούλιος 2010)
- Όταν μια εργασία γίνεται δεκτή από το περιοδικό, κατατίθεται και στην Pubmed
- Η Pubmed δίνει ένα μοναδικό κωδικό εγγραφής (PMID) και λέξεις κλειδιά που χαρακτηρίζουν το περιεχόμενο της εργασίας (MeSH terms).
- Από το 2007, το NIH απαιτεί όποιες ερευνητικές εργασίες έχουν χρηματοδοτηθεί από αυτό, τα αποτελέσματά τους να γίνονται προσβάσιμα σε όλους, μέσω του Pubmed Central (εντός 12 μηνών από την ημερομηνία δημοσίευσης). (~ 1 εκατομύριο εργασίες)



# Pubmed


[Resources](#)  [How To](#) 
[My NCBI](#) [Sign In](#)


 U.S. National Library of Medicine  
 National Institutes of Health

Search: 
[Limits](#) [Advanced search](#) [Help](#)

[Display Settings](#):  Abstract

[Send to:](#)



[Science](#). 1996 Oct 25;274(5287):546, 563-7.

## Life with 6000 genes.

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Gallibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG.

Université Catholique de Louvain, Unité de Biochimie Physiologique, Place Croix du Sud, 2/20, 1348 Louvain-la-Neuve, Belgium.

Comment in:

[Science](#). 1997 Feb 21;275(5303):1051-2.

### Abstract

The genome of the yeast *Saccharomyces cerevisiae* has been completely sequenced through a worldwide collaboration. The sequence of 12,068 kilobases defines 5885 potential protein-encoding genes, approximately 140 genes specifying ribosomal RNA, 40 genes for small nuclear RNA molecules, and 275 transfer RNA genes. In addition, the complete sequence provides information about the higher order organization of yeast's 16 chromosomes and allows some insight into their evolutionary history. The genome shows a considerable amount of apparent genetic redundancy, and one of the major problems to be tackled during the next stage of the yeast genome project is to elucidate the biological functions of all of these genes.

PMID: 8849441 [PubMed - indexed for MEDLINE]

[+](#) Publication Types, MeSH Terms, Substances, Grant Support

[+](#) LinkOut - more resources

### Related citations

[Sequence analysis of a near-subtelomeric 35.4 kb DNA segment on the right arm o \[Yeast. 1997\]](#)

[Complete nucleotide sequence of \*Saccharomyces cerevisiae\* chror \[Science. 1994\]](#)

[The sequence of a 36 kb segment on the left arm of yeast chromosome X identifies 2 \[Yeast. 1994\]](#)

**Review** [Sequencing the yeast genome: an international achievement. \[Yeast. 1994\]](#)

**Review** [Complete nucleotide sequence of \*Saccharomyces cerevisiae\* chror \[EMBO J. 1996\]](#)

[See reviews...](#)

[See all...](#)

### Cited by over 100 PubMed Central articles

[Species concepts in \*Calonectria\* \(\*Cylindrocladium\*\). \[Stud Mycol. 2010\]](#)

[Genome sequence of the necrotrophic plant pathogen \*Pythium ultimum\* I \[Genome Biol. 2010\]](#)

[Reconstruction and validation of RefRec: a global model for the yeast molI \[PLoS One. 2010\]](#)

[See all...](#)

# Pubmed

PMID- 8849441  
 OWN - NLM  
 STAT- MEDLINE  
 DA - 19961122  
 DCOM- 19961122  
 LR - 20090929  
 IS - 0036-8075 (Print)  
 IS - 0036-8075 (Linking)  
 VI - 274  
 IP - 5287  
 DP - 1996 Oct 25  
 TI - Life with 6000 genes.  
 PG - 546, 563-7  
 AB - The genome of the yeast *Saccharomyces cerevisiae* has been completely sequenced through a worldwide collaboration. The sequence of 12,068 kilobases defines 5885 potential protein-encoding genes, approximately 140 genes specifying ribosomal RNA, 40 genes for small nuclear RNA molecules, and 275 transfer RNA genes. In addition, the complete sequence provides information about the higher order organization of yeast's 16 chromosomes and allows some insight into their evolutionary history. The genome shows a considerable amount of apparent genetic redundancy, and one of the major problems to be tackled during the next stage of the yeast genome project is to elucidate the biological functions of all of these genes.  
 AD - Universite Catholique de Louvain, Unite de Biochimie Physiologique, Place Croix du Sud, 2/20, 1348 Louvain-la-Neuve, Belgium.  
 FAU - Goffeau, A  
 AU - Goffeau A  
 FAU - Oliver, S G  
 AU - Oliver SG  
 LA - eng  
 GR - Wellcome Trust/United Kingdom  
 PT - Journal Article  
 PT - Review  
 PL - UNITED STATES  
 TA - Science  
 JT - Science (New York, N.Y.)  
 JID - 0404511  
 RN - 0 (DNA, Fungal)  
 RN - 0 (Fungal Proteins)  
 RN - 0 (RNA, Fungal)  
 SB - IM  
 CIN - Science. 1997 Feb 21;275(5303):1051-2. PMID: 9054002  
 MH - Amino Acid Sequence  
 MH - Base Sequence  
 MH - \*Chromosome Mapping  
 MH - Chromosomes, Fungal/genetics  
 MH - Computer Communication Networks  
 MH - DNA, Fungal/genetics  
 MH - Evolution, Molecular  
 MH - Fungal Proteins/chemistry/genetics/physiology  
 MH - Gene Library  
 MH - \*Genes, Fungal  
 MH - \*Genome, Fungal  
 MH - International Cooperation  
 MH - Multigene Family  
 MH - Open Reading Frames  
 MH - RNA, Fungal/genetics  
 MH - *Saccharomyces cerevisiae*/\*genetics  
 MH - Sequence Analysis, DNA  
 RF - 86  
 EDAT- 1996/10/25  
 MHDA- 1996/10/25 00:01  
 CRDT- 1996/10/25 00:00  
 PST - ppublish  
 SO - Science. 1996 Oct 25;274(5287):546, 563-7.

# Pubmed

- Πόσες εργασίες υπάρχουν για το estrogen receptor alpha;

# Κατάλογος με ΒΔ: Pathguide

- <http://www.pathguide.org/>

Pathguide the pathway resource list

Home BioPAX cBio MSKCC

**Navigation**

- Protein-Protein Interactions
- Metabolic Pathways
- Signaling Pathways
- Pathway Diagrams
- Transcription Factors / Gene Regulatory Networks
- Protein-Compound Interactions
- Genetic Interaction Networks
- Protein Sequence Focused
- Other

**Search**

Organisms: All

Availability: All

Standards: All

Reset Search

**Analysis**

- Statistics
- Database Interactions

**Contact**

Comments, Questions, Suggestions are Always Welcome!

**Database Interactions**

Network: All (Pathways) Databases

This network shows the links among many databases in Pathguide.

Selecting node(s) shows a summary of database information below the network, with linkouts to database details from Pathguide, and to the database itself.

Reset Layout | Show Pan-Zoom Control

**Resources**

Database Name	Categories	Full Record	Availability	Standards
---------------	------------	-------------	--------------	-----------

[Back to the Top](#)

**Legends**

**Resource Type**

- Interactions
- Pathways
- Predictive Interactions
- Metamining
- Exchange format language
- Unifying efforts
- Not categorized

**Interaction Type**

- Source → Mining source data
- Source ◁ Maps to source
- ↔ Bidirectional exchange agreement



# Bionumbers

<http://www.bionumbers.hms.harvard.edu/>

Browse -> genome

## B10NUMB3R5

THE DATABASE OF USEFUL BIOLOGICAL NUMBERS

Home \ Search | Browse | Resources | BioNumber of The Month | About Us | Login \ Submit

Popular BioNumbers | Recent BioNumbers | Key BioNumbers | Amazing BioNumbers

Find Terms   
e.g., [ribosome](#), [p53](#), [glucose](#), [CO2](#)

Organism

Did you ever need to look up a number like the volume of a cell or the cellular concentration of ATP, only to find yourself spending much more time than you wanted on the Internet or flipping through textbooks - all without much success?

Well, it didn't happen only to you. It is often surprising how difficult it can be to find concrete biological numbers, even for properties that have been measured numerous times. To help solve this for one and all, BioNumbers (**the database of key numbers in molecular biology**) was created. Along with the numbers, you'll find the relevant **references to the original literature**, useful comments, and related numbers.

Though we have made an honest first try at simplifying the process of finding useful biological numbers, there is still much work to be done. **A key challenge is filling in the large number of missing items. Another challenge involves setting up a reliable and discriminating search engine** which on a first try yields the numbers a user is actually interested in finding.


**FEEDBACK**

Didn't find what you looked for?  
Let us know and we will try to help! (include email for an answer)

submit

BioNumber of the Month

**JUNE** >>



Length 2-4 mm

# Textpresso

- Μηχανή αναζήτησης που ελέγχει ολόκληρο το κείμενο μιας εργασίας (full text).
- <http://www.textpresso.org/>

## Textpresso

---

[Home](#) : [Sites](#) : [Downloads](#) : [Linking](#) : [Publications](#) : [About](#) : [Contact](#)

### Textpresso Sites

[C. elegans](#)  
[Mouse](#)  
[D. melanogaster](#)  
[Neuroscience](#)  
[Arabidopsis](#)  
[Dicty](#)  
[Mycoplasma](#)  
[Rat](#)  
[Zebrafish](#)  
[Nematode](#)  
[Alzheimer's](#)  
-----  
[S. cerevisiae](#)  
[Candida](#)  
[Textpresso for Site-Specific Recombinases](#)  
[Regulon DB](#)  
[Ecoliwiki Textpresso](#)  
[Ecocyc Textpresso](#)  
[Vaxpresso](#)  
[Pharmspresso](#)  
[O. sativa](#)  
[Textpresso for sea urchin](#)

**NEW! Textpresso for Mouse and Textpresso for Mycoplasma now available.** Click on *Mouse* or *Mycoplasma* on the left side menu.

Textpresso is a text-mining system for scientific literature. Textpresso's two major elements are (1) access to full text, so that entire articles can be searched, and (2) introduction of categories of biological concepts and classes that relate two objects (e.g., association, regulation, etc.) or describe one (e.g., methods, etc). A search engine enables the user to search for one or a combination of these categories and/or keywords within an entire literature.

Textpresso is useful as a search engine for researchers as well as a curation tool. It was developed as a part of [WormBase](#) and is used extensively by *C. elegans* curators. Textpresso has currently been implemented for 17 different literatures, and can readily be extended to other corpora of text.

# Clinical tests webpages

- <http://labtestsonline.org/>
- Ποιά test για ποιές ασθένειες
  
- <http://informedna.com/index.php/>
- Informed Medical Decisions, Inc. is the only nationwide network of independent genetic counselors.

# Εργασία

Να ετοιμάσετε μια αναφορά για τον ανθρώπινο υποδοχέα οιστρογόνου άλφα (estrogen receptor alpha) που να περιγράφει τα παρακάτω:

- Τον κωδικό εγγραφής στην Β.Δ. Uniprot.
- Την ταξινόμηση του οργανισμού από τον οποίο προέρχεται η πρωτεΐνη.
- Τις λειτουργίες της πρωτεΐνης.
- Τα domains της πρωτεΐνης και την λειτουργία του καθενός από αυτά.
- Να δείξετε από μια εικόνα της τρισδιάστατης δομής για τα domains της πρωτεΐνης (εφόσον υπάρχουν).
- Την έκφραση του γονιδίου στους ανθρώπινους ιστούς.
- Σε ποιές ασθένειες εμπλέκεται η πρωτεΐνη αυτή.
- Ποιά άλλα γονίδια είναι μοριακοί δείκτες στις ασθένειες που εμπλέκεται αυτός ο υποδοχέας οιστρογόνου.

Για κάθε στάδιο της εργασίας να αναφέρετε την πηγή (Β.Δ.) στο διαδίκτυο μαζί με το URL καθώς επίσης και μια σύντομη περιγραφή της Β.Δ.