



BMG 744

Bioinformatics/Genomics/Proteomics

The “ics” revolution
March 4, 2003



Contact Information

Elliot Lefkowitz

✦ Email

– ElliotL@uab.edu

✦ Web Site

– <http://www.genome.uab.edu>

✦ Office

– BBRB 277A

✦ Phone

– 934-1946



UAB-BCRF Home - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://www.genome.uab.edu/> Go Links »

UAB

Molecular and Genetic Bioinformatics Facility

General Information

- [UAB Bioinformatics Resources](#)
- [Sequence Analysis at UAB](#)
- [MIC753 - "Practical Applications of Sequence Analysis"](#)
- [CIS 640 - Bioinformatics I: "Lectures on Practical Bioinformatics" pdf](#)

Genomic Sequencing

- [The Poxvirus Bioinformatics Resource](#)
- [The *Streptococcus pneumoniae* genome diversity project](#)
- [The *Streptococcus pneumoniae* strain SpR6 genome sequencing project](#)
- [The *Ureaplasma urealyticum* genomic sequencing project](#)

UAB Only (Password required. Call or Email Elliot for access)

- [GCG at UAB](#)
- [SeqWeb - Web interface to GCG](#)
- [GCG 10 Documentation](#)
- [GCG 10 Documentation - Downloadable pdf files](#)

For information contact:	
Elliot Lefkowitz	
Phone:	205-934-1946
Email:	ElliotL@uab.edu
Office:	BBRB 277A

Funding for the Molecular and Genetic Bioinformatics Facility has been provided in part by the UAB Health Services Foundation and the UAB Center for AIDS Research



Definitions

(From NIH)

✓ Bioinformatics

- Research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data.

✓ Computational Biology

- The development and application of data-analytical and theoretical methods, mathematical modeling and computational simulation techniques to the study of biological, behavioral, and social systems.



Elliot's Definition of Bioinformatics

- ✓ Computer-aided analysis of biological information

- ✓ Caveat:
 - In the end, bioinformatics (a.k.a. computers) can only help in making inferences concerning biological processes.
 - These inferences (or hypotheses) have to be tested in the laboratory



Why Bioinformatics?

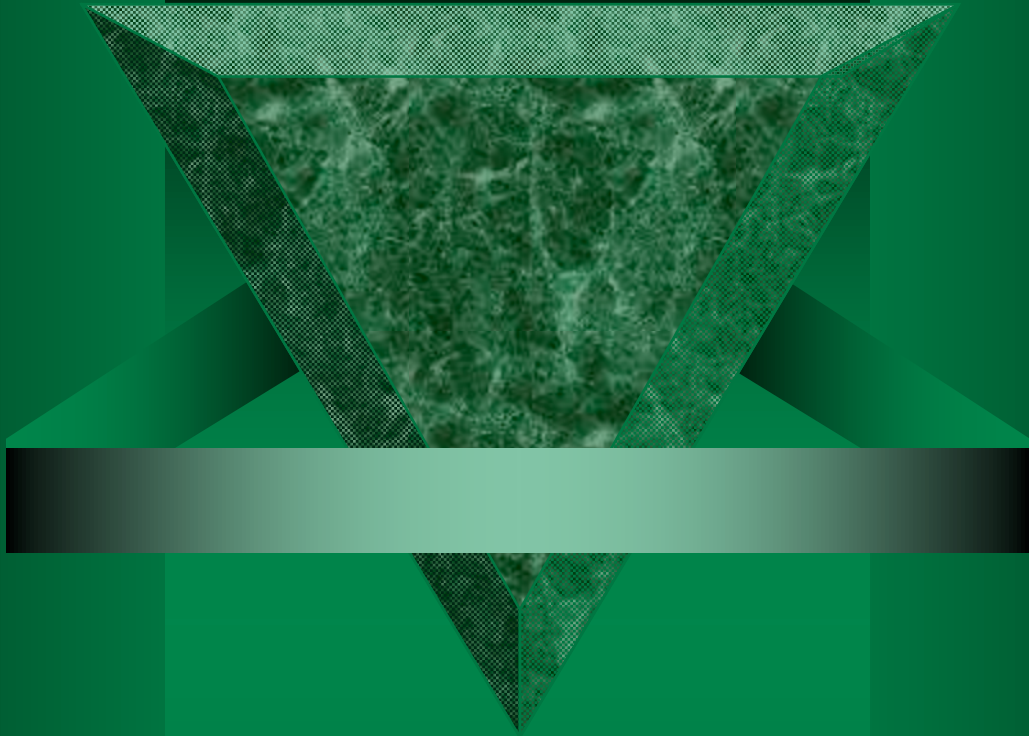
- ✓ Assists in the understanding of the basic genetic components of living systems.
- ✓ Understanding Genetics
 - Sequence
 - Genetic organization
 - Comparative analysis
 - Variation
- ✓ Structure Prediction
- ✓ Functional Inference
 - Enzymatic/Regulatory/Structural
 - Expression
 - Protein interactions



Why Bioinformatics Now?

✓ Data overload

- Genomics
- Gene expression arrays
- Proteomics
- Structural biology
- Future biological revolutions



Genomics



The Human Genome Project

- ✔ Mapping and Sequencing the Genomes of Model Organisms
- ✔ Data Collection and Distribution
- ✔ Ethical, Legal, and Social Considerations
- ✔ Research Training
- ✔ Technology Development
- ✔ Technology Transfer



Genomes of Humans and their “cousins”

- ✓ Eukaryotic
- ✓ Prokaryotic
- ✓ Archaea
- ✓ Viruses

Human Genome Sequencing as of 1/5/2003

95.8% finished; chromosomes 6, 7, 13, 20, 21, 22 and Y are considered complete

Chr.	Tiling Path	Sequenced Clones (%)	
	Clones	Draft	Finished
1	2202	71 (3.2)	2089 (94.9)
2	1966	14 (0.7)	1948 (99.1)
3	1740	82 (4.7)	1630 (93.7)
4	1629	47 (2.9)	1572 (96.5)
5	1775	54 (3.0)	1700 (95.8)
6	1795	2 (0.1)	1790 (99.7)
7	1526	1 (0.1)	1514 (99.2)
8	1239	79 (6.4)	1133 (91.4)
9	999	22 (2.2)	963 (96.4)
10	1132	6 (0.5)	1117 (98.7)
11	1147	43 (3.7)	1088 (94.9)
12	1140	74 (6.5)	1044 (91.6)
13	854	0 (0.0)	854 (100.0)
14	655	1 (0.2)	642 (98.0)
15	710	68 (9.6)	618 (87.0)
16	725	44 (6.1)	672 (92.7)
17	692	132 (19.1)	541 (78.2)
18	600	6 (1.0)	585 (97.5)
19	861	9 (1.0)	852 (99.0)
20	632	0 (0.0)	632 (100.0)
21	473	0 (0.0)	472 (99.8)
22	527	0 (0.0)	527 (100.0)
X	1588	37 (2.3)	1508 (95.0)
Y	200	0 (0.0)	200 (100.0)
total	26807	792 (3.0)	25691 (95.8)



Genome Project Organization

- ✓ Cloning
- ✓ Mapping
- ✓ Sequencing
- ✓ Annotation
- ✓ Analysis



Bioinformatic Information Flow

- ✓ “Raw” data generation
 - Sequence generation and assembly
- ✓ Analytical tools
 - Pattern matching
- ✓ Database generation
 - Construction and data import
- ✓ Visualization (publication) of results
 - Static: Table or graph
 - Dynamic: Web page/Java applet



Annotation and Analysis

- ✓ Gene prediction
 - Identify patterns characteristic of ORFs
- ✓ Functional assignment
 - Similarity searching
- ✓ Metabolic pathway modeling
- ✓ Comparative analysis
 - Identification and comparison with related genes



Comparative Genomics

✓ “Similar” sequences

- Sequences related by primary sequence similarity

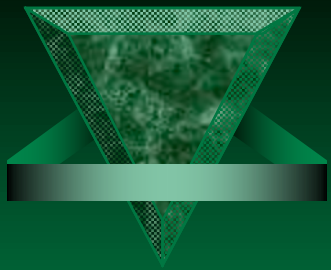
✓ Homologs

- Sequences related by evolution
- Orthologs
 - Related due to speciation
- Paralogs
 - Related due to gene duplication



Bioinformatic Databases

Something to compare against



Sequence Databases

✓ DNA

- Genbank (NCBI)
- EMBL
- DDBJ

✓ Protein

- PIR
- Swiss-Prot
- Swiss-Prot TrEMBL
- PDB



Other Databases

✓ Structural

- Protein Data Bank (PDB): <http://www.rcsb.org/pdb/>

✓ Expression

- Microarray Gene Expression Data Society (MGED): <http://www.mged.org/>
- Gene Expression Omnibus (GEO – NCBI)

✓ Proteomic

- Mascot: <http://www.matrixscience.com/>

✓ Metabolism

- BioCyc: <http://biocyc.org/>

✓ Ontology

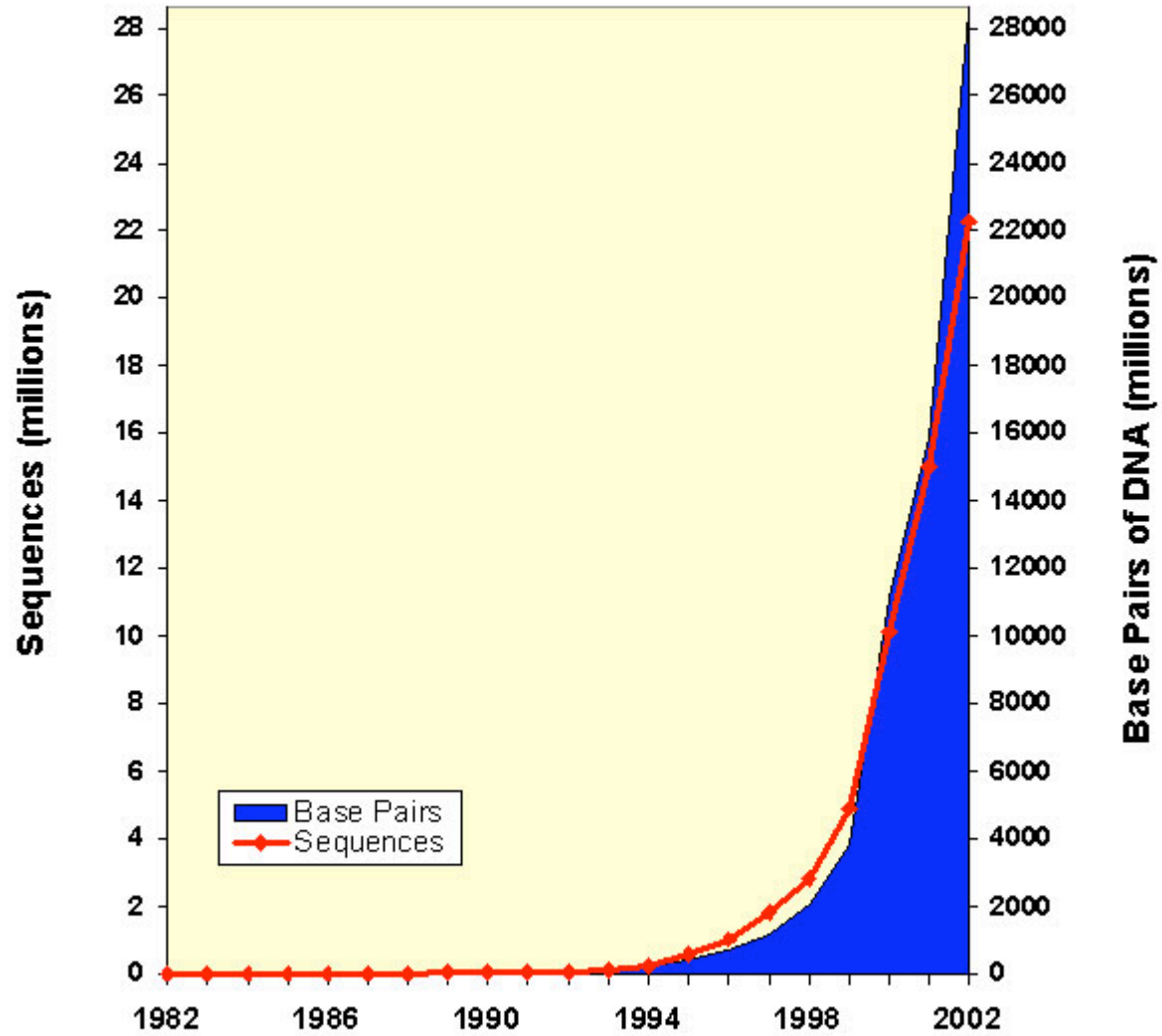
- Gene Ontology (GO) Consortium: <http://www.geneontology.org/>



Genbank

- ✓ Primary nucleic acid sequence database
- ✓ Maintained by NCBI
 - National Center for Biotechnology Information
 - <http://www.ncbi.nlm.nih.gov>
- ✓ February 15, 2003; Release 134
 - 29,358,082,791 bases
 - 23,035,823 sequences

Growth of GenBank





Genbank Divisions

1. PRI - primate sequences
2. ROD - rodent sequences
3. MAM - other mammalian sequences
4. VRT - other vertebrate sequences
5. INV - invertebrate sequences
6. PLN - plant, fungal, and algal sequences
7. BCT - bacterial sequences
8. VRL - viral sequences
9. PHG - bacteriophage sequences
10. SYN - synthetic sequences
11. UNA - unannotated sequences
12. EST - EST sequences (expressed sequence tags)
13. PAT - patent sequences
14. STS - STS sequences (sequence tagged sites)
15. GSS - GSS sequences (genome survey sequences)
16. HTG - HTGS sequences (high throughput genomic sequences)
17. HTC - unfinished high-throughput cDNA sequencing



HTGS

High Throughput Genomic Sequences

- ✓ 'Unfinished' DNA sequences generated by the high-throughput sequencing centers
- ✓ Phase 0
 - Single-few pass reads of a single clone (not contigs)
- ✓ Phase 1
 - Unfinished, may be unordered, unoriented contigs, with gaps
- ✓ Phase 2
 - Unfinished, ordered, oriented contigs, with or without gaps
- ✓ Phase 3
 - Primary division (Genbank)
 - Finished, no gaps (with or without annotations)



EST

- ✓ Expressed Sequence Tags
 - “Single-pass” cDNA sequences
 - Generally representative of the 3’ ends of cDNAs
 - More “full-length” ESTs now available
- ✓ NCBI also has a dbEST database
 - Same content
 - Different format



STS

- ✓ Sequence Tagged Sites
 - Sequence and mapping data
 - Short genomic landmark sequences
- ✓ NCBI also has a dbSTS database
 - Same content
 - Different format



GSS

- ✓ Genome Survey Sequences
- ✓ Similar to the EST division, except that its sequences are genomic in origin, rather than cDNA
 - Random “single pass read” genome survey sequences.
 - Cosmid/BAC/YAC end sequences
 - Exon trapped genomic sequences
 - alu PCR sequences
- ✓ NCBI also has a dbGSS database
 - Same content
 - Different format



Other NCBI Databases

- ▼ RefSeq
- ▼ Unigene
- ▼ HomoloGene
- ▼ Genomic
- ▼ dbSNP



RefSeq

- ✓ NCBI Reference Sequence project
- ✓ Provides reference sequence standards for the naturally occurring molecules from chromosomes to mRNAs to proteins
- ✓ Stable reference point for:
 - mutation analysis
 - gene expression studies
 - polymorphism discovery
- ✓ Accession numbers have two letters, an underscore, and six numbers
 - NM_123456



RefSeq...

- ✓ Curated RefSeq
 - transcripts and proteins
- ✓ Genome Annotation
 - contigs, transcripts, and proteins
- ✓ Complete Genomes
 - genomes, chromosomes, and proteins



Unigene

- ✔ GenBank sequences partitioned into a non-redundant set of gene-oriented clusters
 - Each UniGene cluster contains sequences that represent a unique gene, as well as related information such as the tissue types in which the gene has been expressed and map location.
- ✔ Includes EST and complete cDNA sequences
- ✔ Provides information on differentially-spliced transcripts

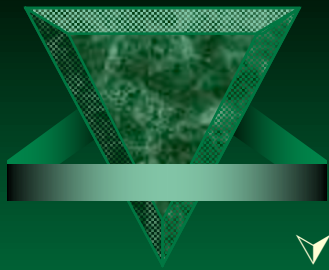
Unigene Organisms

	Vertebrata		
	Mammalia		
	Bos taurus (cow)	12,808 entries	
	Homo sapiens (human)	128,826 entries	
	Mus musculus (mouse)	90,444 entries	
	Rattus norvegicus (rat)	63,253 entries	
	Sus scrofa (pig)	14,344 entries	
	Aves		
	Gallus gallus (chicken)	5,068 entries	
	Amphibia		
	Xenopus laevis (frog)	19,512 entries	
	Actinopterygii		
	Danio rerio (zebrafish)	16,355 entries	
	Urochordata		
	Ascidiacea		
	Ciona intestinalis (sea squirt)	13,674 entries	
	Arthropoda		
	Insecta		
	Anopheles gambiae (malaria mosquito)	3,270 entries	
	Drosophila melanogaster (fruit fly)	14,779 entries	
	Nematoda		
	Chromadorea		
	Caenorhabditis elegans	20,137 entries	
	Embryophyta		
	Eudicotyledons		
	Arabidopsis thaliana (thale cress)	27,141 entries	
	Glycine max (soybean)	8,987 entries	
	Lycopersicon esculentum (tomato)	3,740 entries	
	Medicago truncatula (barrel medic)	5,729 entries	
	Liliopsida		
	Hordeum vulgare (barley)	7,944 entries	
	Oryza sativa (rice)	19,223 entries	
	Triticum aestivum (wheat)	20,454 entries	
	Zea mays (maize)	13,512 entries	
	Chlorophyta		
	Chlorophyceae		
	Chlamydomonas reinhardtii	6,448 entries	



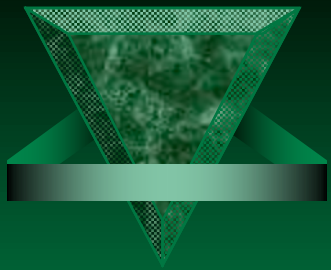
HomoloGene

- ✓ Curated and calculated orthologs and homologs for genes represented in UniGene and LocusLink. Organisms include:
 - *Arabidopsis thaliana*
 - *Bos taurus*
 - *Caenorhabditis elegans*
 - *Danio rerio*
 - *Drosophila melanogaster*
 - *Homo sapiens*
 - *Hordeum vulgare*
 - *Lycopersicon esculentum*
 - *Medicago truncatula*
 - *Mus musculus*
 - *Oryza sativa*
 - *Rattus norvegicus*
 - *Sus scrofa*
 - *Triticum aestivum*
 - *Xenopus laevis*
 - *Zea mays*



Genomic DBs

- ✓ Human
- ✓ Mouse
- ✓ Rat
- ✓ Zebrafish
- ✓ Drosophila
- ✓ Nematode
- ✓ Plant genomes
- ✓ Yeast
- ✓ Malaria
- ✓ Microbial genomes
- ✓ Viruses
- ✓ Viroids
- ✓ Plasmids
- ✓ Eukaryotic organelles



dbSNP

- ✓ Single Nucleotide Polymorphisms
 - Single base changes
 - Small-scale insertions/deletions
 - Polymorphic repetitive elements
 - Microsatellite variation



LocusLink

- ✓ Provides a single query interface to curated sequence and descriptive information about genetic loci
 - Nomenclature
 - Aliases
 - Sequence accessions
 - Phenotypes
 - EC numbers
 - MIM numbers
 - UniGene clusters
 - Homology
 - Map locations
 - Web sites



OMIM

- ✓ Online Mendelian Inheritance in Man
- ✓ Database of gene-linked genetic disorders
- ✓ Maintained at Johns Hopkins University
 - Dr. Victor A. McKusick



Sample OMIM Queries

(From the OMIM Help Docs)

- ✔ What human genes are related to hypertension? Which of those genes are on chromosome 17?
- ✔ List the OMIM entries that describe genes on chromosome 10.
- ✔ List the OMIM entries that contain information about allelic variants.
- ✔ Retrieve the OMIM record for the cystic fibrosis transmembrane conductance regulator (CFTR), and link to related protein sequence records via Entrez.
- ✔ Find the OMIM record for the p53 tumor protein, and link out to related information in LocusLink and the p53 Mutation Database.



EMBL and DDBJ

- ✓ European Molecular Biology Laboratory
 - Hinxton, UK
 - <http://www.ebi.ac.uk/>
- ✓ DNA Data Bank of Japan
 - Mishima, Japan
 - <http://www.ddbj.nig.ac.jp/>



Coordination with Genbank

- ✔ Prevents duplication
- ✔ Genbank enters sequences from U.S. journals and researchers
- ✔ EMBL handles European data
- ✔ DDBJ handles Asian data
- ✔ Data exchanged daily



Accession Numbers

- ✔ Each sequence submitted to a database is assigned a unique primary accession number
- ✔ Accession numbers do not change
- ✔ If a sequence is merged with another, a new accession number is assigned, and the original number becomes a secondary accession number
- ✔ Accession numbers may include version numbers
 - AO2428.2



The Sequence Record

- ✔ Different for each database
- ✔ Locus (Name)
- ✔ Accession Number
- ✔ Keywords
- ✔ Description
- ✔ Properties
- ✔ References
- ✔ The Sequence



GenBank Sample Record

✓ <http://www.ncbi.nlm.nih.gov/Sitemap/samplerecord.html>

analyze% typedata ge:humcftrm

!!NA_SEQUENCE 1.0

LOCUS HUMCFTRM 6129 bp mRNA PRI 15-DEC-1989

DEFINITION Human cystic fibrosis mRNA, encoding a presumed transmembrane conductance regulator (CFTR).

ACCESSION M28668

NID g180331

KEYWORDS cystic fibrosis; transmembrane conductance regulator.

SOURCE Human, cDNA to mRNA.

ORGANISM Homo sapiens

Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 6129)

AUTHORS Riordan, J.R., Rommens, J.M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J.-L., Drumm, M.L., Iannuzzi, M.C., Collins, F.S. and Tsui, L.-C.

TITLE Identification of the cystic fibrosis gene: Cloning and characterization of complementary DNA

JOURNAL Science 245, 1066-1073 (1989)

MEDLINE 89368940

COMMENT A three base-pair deletion spanning positions 1654-1656 is observed in cDNAs from cystic fibrosis patients.

FEATURES Location/Qualifiers
source 1. .6129
/organism="Homo sapiens"
/db_xref="taxon:9606"
CDS 133. .4575
/note="cystic fibrosis transmembrane conductance regulator"
/codon_start=1
/db_xref="PID:g180332"
/translation="MQRSPLEKASVVSKLFFSWTRPILRKGYRQRLELSDIYQIPSVDSADNLSEKLEREWDRELASKKNPKLINALRRCFFWRFMFYGIFLYLGEVTKAVQPLLLLNRFSDKDIAILDLLPLTIFDFIQLLLIVIGAIAVVAVLQPYIFVATVPVIVAFIMLRAYFLQTSQQLKQLESEGRSPIFTHLVTSLKGLWTLRAFGROPYFETLFHKALNLHTANWFLYLSTLRWFQMRIEMIFVIFFFIAVTFISILTTEGEGRVIILTLAMNIMSTLQWAVNSSIDVDSLMRSVSRVFKFIDMPTEGKPTKSTKPYKNGQLSKVMI IENSHVKKDDIWPSGGQMTVKDLTAKYTEGGNAILENISFSISPGQRVGLLGRTGSGKSTLLSAFLRLNTEGEIQIDGVSWDSITLQQWRKAFGVIPQKVFIFSGTFRKNLDPYEQWSDQEIWKVAD EVGLRSVIEQFPKGLDFVLVDGGCVLSHGKQLMCLARSVLSKAKILLLDEPSAHLDPVTYQIIRRTLKQAFADCTVILCEHRIEAMLECQQFLVIEENKVRQYDSIQKLLNERSLFRQAISPSDRVKLFPHRNSSKCKSKPQIAALKEETEEEVQDTRL"

BASE COUNT 1886 a 1181 c 1330 g 1732 t
ORIGIN

HUMCFTRM Length: 6129 April 13, 1998 13:00 Type: N Check: 6781 ..

```
1  AATTGGAAGC AAATGACATC ACAGCAGGTC AGAGAAAAAG GGTTGAGCGG
51  CAGGCACCCA GAGTAGTAGG TCTTTGGCAT TAGGAGCTTG AGCCCAGACG
101 GCCCTAGCAG GGACCCCAGC GCCCGAGAGA CCATGCAGAG GTCGCCTCTG
151 GAAAAGGCCA GCGTTGTCTC CAAACTTTTT TTCAGCTGGA CCAGACCAAT
201 TTTGAGGAAA GGATACAGAC AGCGCCTGGA ATTGTCAGAC ATATACCAA
251 TCCCTTCTGT TGATTCTGCT GACAATCTAT CTGAAAATT GGAAAGAGAA
301 TGGGATAGAG AGCTGGCTTC AAAGAAAAT CCTAAACTCA TTAATGCCCT
351 TCGGCGATGT TTTTTCTGGA GATTTATGTT CTATGGAATC TTTTATATT
401 TAGGGGAAGT CACCAAAGCA GTACAGCCTC TCTTACTGGG AAGAATCATA
451 GCTTCCTATG ACCCGGATAA CAAGGAGGAA CGCTCTATCG CGATTTATCT
```

analyze% typedata -ref GB_PR:HUMIFNRF1A

!!NA_SEQUENCE 1.0

LOCUS HUMIFNRF1A 7721 bp DNA PRI 10-NOV-1992

DEFINITION Homo sapiens interferon regulatory factor 1 gene, complete cds.

ACCESSION L05072

NID g184648

KEYWORDS interferon regulatory factor 1.

SOURCE Homo sapiens Placenta DNA.

ORGANISM Homo sapiens

Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 7721)

AUTHORS Cha,Y., Sims,S.H., Romine,M.F., Kaufmann,M. and Deisseroth,A.B.

TITLE Human interferon regulatory factor 1: intron/exon organization

JOURNAL DNA Cell Biol. 11, 605-611 (1992)

MEDLINE 93000481

```

FEATURES             Location/Qualifiers
     source           1. .7721
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="Placenta"
                     /map="5q23-q31"
     exon            1. .219
                     /gene="IRF1"
                     /note="putative"
                     /number=1
     5'UTR          join(1. .219,1279. .1287)
                     /gene="IRF1"
     gene            join(1. .219,1279. .1287)
                     /gene="IRF1"
     intron          220. .1278
                     /gene="IRF1"
                     /number=1
     exon            1279. .1374
                     /gene="IRF1"
                     /number=2
     CDS             join(1288. .1374,2738. .2837,3630. .3806,3916. .3965,
                     4073. .4202,4386. .4508,5040. .5089,6248. .6383,6670.
                     .6794)
                     /gene="IRF1"
                     /codon_start=1
                     /product="interferon regulatory factor 1"
                     /db_xref="PID:g184649"
                     /translation="MPITRMRMRPWLEMQINSNQIPGLIWINKEEMIFQIPWKHAAKH
                     GWDINKDACLFRSWAIHTGRYKAGEKEPDPKTWKANFRCAMNSLPDIEEVKDQSRNKG
                     SSAVRVYRMLPPLTKNQKQKSKSSRDAKSKAKRKSCGDSSPDTFSDGLSSSTLPDD
                     HSSYTVPGYMQDLEVEQALTPALSPCAVSSSTLPDWHIPVEVVPDSTSDLYNFQVSPMP
                     STCEATPTDEDEFECKI PEDI MKL I FQCEWQPTNYDCKCYI I NEDCGVQPTGCVYGCDFSCKE

```

```

intron      1375. .2737
             /gene="IRF1"
             /number=2
exon        2738. .2837
             /gene="IRF1"
             /number=3
intron      2838. .3629
             /gene="IRF1"
             /number=3
exon        3630. .3806
             /gene="IRF1"
             /number=4
intron      3807. .3915
             /gene="IRF1"
             /number=4
exon        3916. .3965
             /gene="IRF1"
             /number=5
intron      3966. .4072
             /gene="IRF1"
             /number=5
             ...
exon        5040. .5089
             /gene="IRF1"
             /number=8
intron      5090. .6247
             /gene="IRF1"
             /number=8
exon        6248. .6383
             /gene="IRF1"
             /number=9
intron      6384. .6669
             /gene="IRF1"
             /number=9
exon        6670. .7656
             /gene="IRF1"
             /number=10
3'UTR      6795. .7656
BASE COUNT 1750 a   1946 c   2253 g   1772 t
ORIGIN

```

```
analyze% typedata -ref est:hum091226f
```

```
!!NA_SEQUENCE 1.0
```

```
LOCUS          HUM091226F      152 bp      mRNA          EST          02-APR-1996
```

```
DEFINITION    Homo sapiens retinal fovea EST HFV091226 sequence.
```

```
ACCESSION     L48850
```

```
NID           g1254959
```

```
KEYWORDS      EST; expressed sequence tag.
```

```
SOURCE        Homo sapiens (clone: EST HFV091226) age normalized retinal foveae  
              cDNA to mRNA.
```

```
ORGANISM      Homo sapiens
```

```
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE     1 (sites)
```

```
AUTHORS       Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.
```

```
TITLE         3,400 new expressed sequence tags identify diversity of transcripts  
              in human brain
```

```
JOURNAL       Nature Genet. 4 (3), 256-267 (1993)
```

```
MEDLINE       93364420
```

```
REFERENCE     2 (sites)
```

```
AUTHORS       Liew,C.C., Hwang,D.M., Fung,Y.W., Laurensen,C., Cukerman,E.,  
              Tsui,S. and Lee,C.Y.
```

```
TITLE         A catalogue of genes in the cardiovascular system as identified by  
              expressed sequence tags
```

```
JOURNAL       Proc. Natl. Acad. Sci. U.S.A. 91 (22), 10645-10649 (1994)
```

```
MEDLINE       95024171
```

```
REFERENCE     3 (bases 1 to 152)
```

```
AUTHORS       Bernstein,S.L., Borst,D.E., Neuder,M.E. and Wong,P.
```

```
TITLE         Characterization of a human fovea cDNA library and regional  
              differential gene expression in the human retina
```

```
JOURNAL       Genomics 32 (3), 301-308 (1996)
```



```

FEATURES             Location/Qualifiers
    source           1. .152
                    /organism="Homo sapiens"
                    /note="Expressed sequence tags (first pass sequencing)
                    from randomly selected bacteriophage clones (mRNA-cDNA)
                    from human retinal fovea. The library is age normalized
                    from ten sets of donor foveae 2-79 years old.
                    /db_xref="taxon:9606"
                    /clone="EST HFV091226"
                    /dev_stage="age normalized"
                    /tissue_type="retinal foveae"
    mRNA             <1. .>152
                    /standard_name="EST HFV091226"
BASE COUNT          31 a      42 c      41 g      36 t      2 others
ORIGIN

```

analyze% typedata -ref sts:humswx153

!!NA_SEQUENCE 1.0

LOCUS HUMSWX153 192 bp DNA STS 24-MAY-1993

DEFINITION Human chromosome X STS sWXD153; single read.

ACCESSION L15212

NID g292645

KEYWORDS STS; primer; sequence tagged site.

SOURCE Homo sapiens DNA.

ORGANISM Homo sapiens

Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 192)

AUTHORS Kere,J., Nagaraja,R., Mumm,S.R., Ciccodicola,A., D'Urso,M.
and
Schlessinger,D.

TITLE Mapping human chromosomes by walking with sequence-tagged sites

from end fragments of yeast artificial chromosome inserts

JOURNAL Genomics 14, 241-248 (1992)

MEDLINE 93052321

COMMENT

Submitted by: David Schlessinger,
Center for Genetics in Medicine,
Washington University School of Medicine, Box 8232 4566 Scott
Avenue, St. Louis, MO 63110, USA
e-mail: davids@wugenmail.wustl.edu
Primer A: TAAAGGGATCGCCAAGGAC
Primer B: CTTACTCATTTGCTGGATTCTC
STS size: 85bp
Template: 600 ng/100ul
Primer: 40 pmoles/100ul
dNTPs: 100 uM
MgCl₂: 1.5 mM
KCl: 100 mM
TrisHCl: 10 mM
Taq Polymerase: 0.125 U
NH₄Cl: 5 mM
pH: 8.6
Total Vol: 5 ul
PCR Profile:
Denaturation: 94 degrees C for 1.00 minute(s)
Annealing: 55 degrees C for 2.00 minute(s)
Polymerization: 72 degrees C for 2.00 minute(s)
PCR Cycles: 35
Thermal Cycler: P-E.

```

FEATURES                               Location/Qualifiers
    source                               1. .192
                                         /organism="Homo sapiens"
                                         /db_xref="taxon:9606"
                                         /map="Xq13-q24"
    STS                                  60. .144
                                         /standard_name="sWXD153"
    primer_bind                          60. .78
    primer_bind                          complement(123. .144)
BASE COUNT                               72 a      26 c      60 g      29 t      5 others
ORIGIN
analyze%

```



Entrez Searching

- ✓ <http://www.ncbi.nlm.nih.gov/entrez/>
- ✓ Search via text patterns
- ✓ Access to all of NCBI's databases
 - Sequence
 - PubMed
 - OMIM
 - Linkage information



Swiss-Prot

- ✓ <http://www.expasy.ch/sprot/>
- ✓ Protein Database
- ✓ University of Geneva
- ✓ Arranged by protein function
- ✓ Release 40.44
- ✓ February 22, 2003
- ✓ 44,864,044 amino acids 122,214 entries
- ✓ Provides annotated protein records



Swiss-Prot TrEMBL

- ✔ Translation of all EMBL Nucleic Acid coding sequences not yet present in Swiss-Prot
- ✔ Allows rapid availability without immediate annotation
- ✔ Release 21.13
- ✔ February 14, 2003
- ✔ 725,373 entries



National Biomedical Research Foundation

- ✔ Database begun over twenty years ago by Margaret O. Dayhoff
- ✔ Originally published sequences in book form
- ✔ Started with sequences derived from direct amino acid sequencing



PIR

- ✓ <http://pir.georgetown.edu/>
- ✓ Protein Identification Resource
 - PIR-International Protein Sequence Database (PSD)
- ✓ National Biomedical Research Foundation
- ✓ Georgetown University
- ✓ Release 75.04, March 03, 2003
- ✓ 283,290 Entries



PIR-NREF

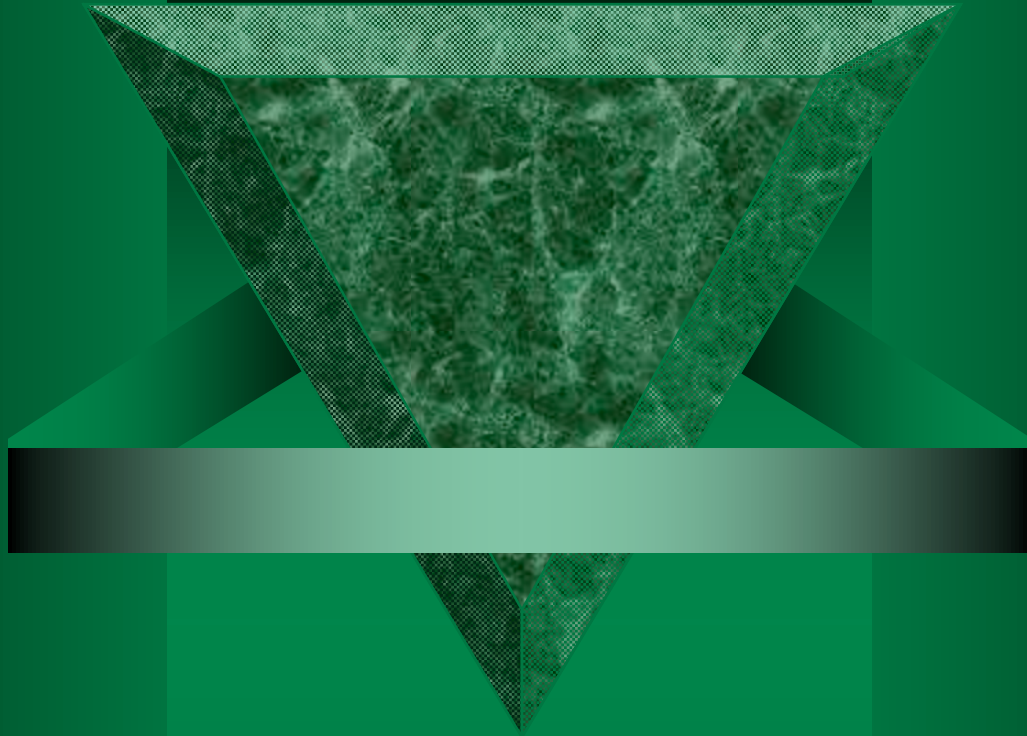
- ✔ Non-redundant REFerence protein database
- ✔ Current Release 1.17
- ✔ March 3, 2003
- ✔ 1,159,203 Entries



iProClass Database - PIR

<http://pir.georgetown.edu/iproclass/>

- ✓ Comprehensive family relationships and structural/functional classifications and features of proteins
 - Superfamilies
 - Families
 - Domains



So I have a protein. What next?



NCBI Sequence Services

- ✓ Obtain sequences directly from NCBI using Entrez
 - Sequence Searches
 - Sequence Retrieval
- ✓ Other services
 - BLAST Searches
 - Sequence Submission
 - PubMed Searches
- ✓ <http://www.ncbi.nlm.nih.gov>



Sequence Similarities

- ✔ What other sequences have some primary sequence similarity to my query sequence?
- ✔ Time and cost of the search is dependent on the size of the database
 - Restrict the size of the database



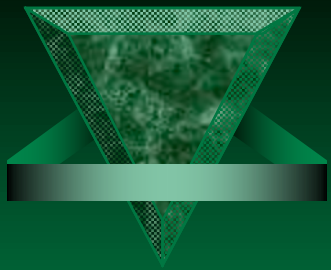
BLAST

- ✔ Search a sequence database for primary sequence similarities to some query sequence
- ✔ Provides a measure of the significance of the similarity
- ✔ Does not necessarily imply common evolutionary origin



BLAST

- ✓ All search combinations possible
- ✓ nt vs. nt database
 - blastn
- ✓ protein vs. protein database
 - blastp
- ✓ translated nt vs. protein database
 - blastx
- ✓ protein vs. translated nt database
 - tblastn
- ✓ translated nt vs. translated nt database
 - tblastx



A Few Genome Resources

✓ NCBI Genome Resources

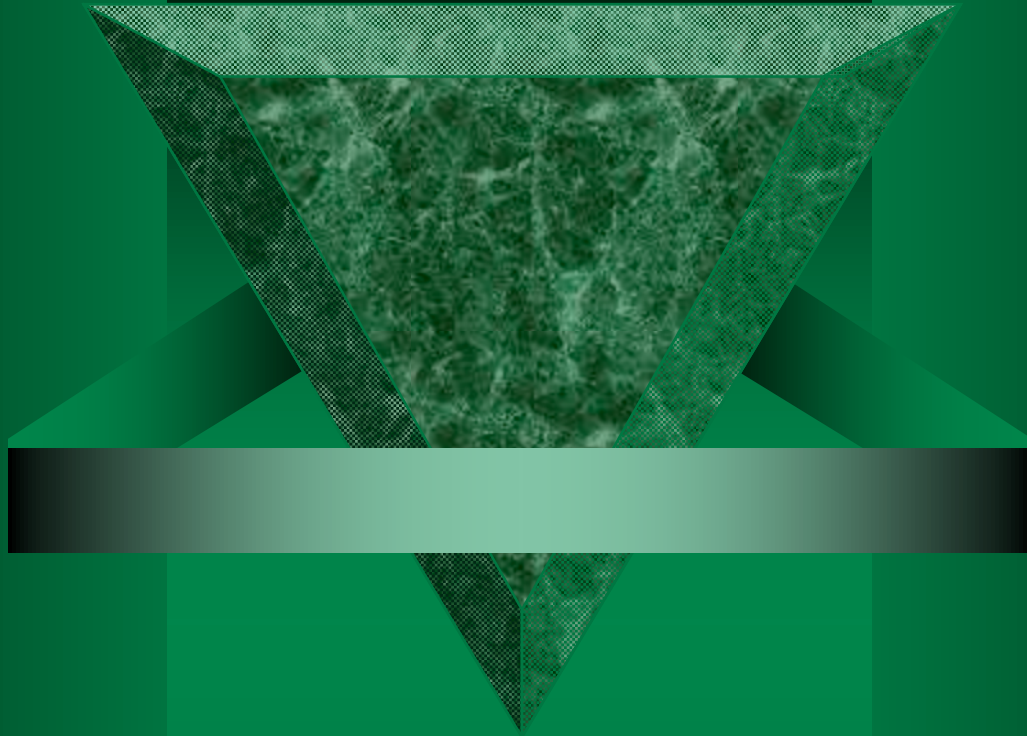
- National Center for Biotechnology Information
- www.ncbi.nih.gov

✓ Ensembl Human Genome Server

- www.ensembl.org

✓ UCSC Human Genome Browser

- genome.ucsc.edu



Analysis of PnP

Human Purine nucleoside phosphorylase

NCBI Sequence Viewer - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Copy Paste

Address http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=protein&list_uids=130377&dopt=GenPept Go Links »

NCBI Entrez Protein

PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books

Search Protein for Go Clear

Limits Preview/Index History Clipboard Details

Display default Show: 20 Send to File Get Subsequence

1: P00491. Purine nucleoside...[gi:130377] [BLink](#), [Domains](#), [Links](#)

LOCUS P00491 289 aa linear PRI 15-JUN-2002

DEFINITION Purine nucleoside phosphorylase (Inosine phosphorylase) (PNP).

ACCESSION P00491

VERSION P00491 GI:130377

DBSOURCE swissprot: locus PNP_HUMAN, accession P00491;
class: standard.
extra accessions:Q15160,created: Jul 21, 1986.
sequence updated: Jul 21, 1986.
annotation updated: Jun 15, 2002.
xrefs: gi: [35564](#), gi: [35565](#), gi: [190150](#), gi: [387033](#), gi: [190147](#),
gi: [190148](#), gi: [190149](#), gi: [66583](#), gi: [230387](#), gi: [230388](#)
xrefs (non-sequence databases): Aarhus/Ghent-2DPAGE2108, MIM
[164050](#), InterProIPR001369, PfamPF00896, PROSITEPS01240

KEYWORDS Transferase; Glycosyltransferase; Polymorphism; Disease mutation;
3D-structure.

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

REFERENCE 1 (residues 1 to 289)
AUTHORS Williams,S.R., Goddard,J.M. and Martin,D.W. Jr.
TITLE Human purine nucleoside phosphorylase cDNA sequence and genomic
clone characterization
JOURNAL Nucleic Acids Res. 12 (14), 5779-5787 (1984)
MEDLINE [84272252](#)
PUBMED [6087295](#)
REMARK SEQUENCE FROM N.A.

REFERENCE 2 (residues 1 to 289)
AUTHORS Williams,S.R., Gekeler,V., McIvor,R.S. and Martin,D.W. Jr.
TITLE A human purine nucleoside phosphorylase deficiency caused by a

Done Internet

NCBI Sequence Viewer - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=protein&list_uids Go Links »

FEATURES	Location/Qualifiers
source	1..289 /organism="Homo sapiens" /db_xref="taxon:9606"
gene	1..289 /gene="NP" /note="synonym: PNP"
Protein	1..289 /gene="NP" /product="Purine nucleoside phosphorylase" /EC_number="2.4.2.1"
Region	7 /gene="NP" /region_name="Hydrogen bonded turn"
Region	8..17 /gene="NP" /region_name="Helical region"
Region	18..20 /gene="NP" /region_name="Hydrogen bonded turn"
Region	27..31 /gene="NP" /region_name="Beta-strand region"
Region	36..41 /gene="NP" /region_name="Helical region"
Region	46..48 /gene="NP" /region_name="Beta-strand region"
Region	50..52 /gene="NP" /region_name="Helical region"
Region	51 /gene="NP" /region_name="Variant" /note="S -> G (IN DBSNP:1049564). /FTId=VAR_002243."
Region	54..55 /gene="NP"

Internet

NCBI Sequence Viewer - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=protein&list_uids= Go Links >>

```

    /gene="NP"
    /region_name="Helical region"
    Region 231..232
    /gene="NP"
    /region_name="Hydrogen bonded turn"
    Region 234..244
    /gene="NP"
    /region_name="Beta-strand region"
    Region 234
    /gene="NP"
    /region_name="Variant"
    /note="R -> P (IN NP DEFICIENCY). /FTId=VAR_002247."
    Region 263..278
    /gene="NP"
    /region_name="Helical region"
    Region 279..281
    /gene="NP"
    /region_name="Helical region"

ORIGIN
    1 mengytyedy kntaewllsh tkhrpqvaih cgsglggltd kltqaqifdy seipnfprst
    61 vpghagrllvf gflngracvm mqgrfihmyeg yplwkvtfpv rvfhllgvdt lvttnaaggl
    121 npkfvevgdim lirdhinlpg fsgqnlrpg nderfgdrfp amsdaydrtm rqralstwkq
    181 mgeqrelqeg tyvmvagpsf etvaecrvlq klgadavgms tvpevivarh cglrvfgfsl
    241 itnkvimdye slekanheev laagkqaqk leqfvsilma siplpdkas

//

Revised: August 5, 2002.
```

Internet

NCBI Sequence Viewer - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=protein&list_uids=130377&dopt=GenPept

NCBI Entrez Protein

PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books

Search Protein for Go Clear

Limits Preview/Index History Clipboard Details

Display FASTA Show: 20 Send to File Get Subsequence

1: P00491 Inoside...[gi:130377] [BLink](#), [Domains](#), [Links](#)

LOCUS 289 aa linear PRI 15-JUN-2002
 DEFINITION nucleoside phosphorylase (Inosine phosphorylase) (PNP).
 ACCESSION GenPept
 VERSION GBSeq XML
 DBSOURCE Graphics
 XML
 default

extra accessions: Q15160, created: Jul 21, 1986.
 sequence updated: Jul 21, 1986.
 annotation updated: Jun 15, 2002.
 xrefs: gi: [35564](#), gi: [35565](#), gi: [190150](#), gi: [387033](#), gi: [190147](#),
 gi: [190148](#), gi: [190149](#), gi: [66583](#), gi: [230387](#), gi: [230388](#)
 xrefs (non-sequence databases): Aarhus/Ghent-2DPAGE2108, MIM
[164050](#), InterProIPR001369, PfamPF00896, PROSITEPS01240

KEYWORDS Transferase; Glycosyltransferase; Polymorphism; Disease mutation;
 3D-structure.

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

REFERENCE 1 (residues 1 to 289)
 AUTHORS Williams, S.R., Goddard, J.M. and Martin, D.W. Jr.

Internet

NCBI Sequence Viewer - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?cmd=&bxt=&save=&cfm=&query_key=60&db=protein&extrafeat=-1&view=1 Go Links >>

NCBI Entrez Protein

PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books

Search Protein for Go Clear

Limits Preview/Index History Clipboard Details

Display FASTA Show: 20 Send to File Get Subsequence

1: P00491. Purine nucleoside...[gi:130377] [BLink](#), [Domains](#), [Links](#)

```
>gi|130377|sp|P00491|PNPH_HUMAN Purine nucleoside phosphorylase (Inosine phosphorylase) (PNP)
MENGTYEDYKNTAEWLLSHTKHRPQVAICGSGLGGLTDKLTQAQIFDYSEIPNFP RSTVPGHAGRLVF
GFLNGRACVMMQGRFHMYYEGYPLWKVTF PVRVFLHLLGVDTLVVVTNAAGGLNPKFEVGDIMLIRDHINLPG
FSGQNPLRGPNDERFGDRFPAMSDAYDRITMRQRALSTWKQMGEQRELQEGTYVMVAGPSFETVAECRVLQ
KLGADAVGMS TVPEVIVARHCGLRVFGFSLITNKVIMDYESLEKANHEEVLAAGKQAAQKLEQFVSILMA
SIPLPKAS
```

Revised: August 5, 2002.

[Disclaimer](#) | [Write to the Help Desk](#)
[NCBI](#) | [NLM](#) | [NIH](#)

Feb 19 2003 14:49:48

Done Internet



Searching Genomic Sequences

- ✔ Where is my sequence located in the human genome?
 - Chromosome; band; mapping data
 - Genetic linkage relationships
- ✔ What is the genomic context of my sequence?
 - Alternative splicing
 - Regulation
- ✔ Are there any paralogs?
- ✔ Are there any pseudogenes?
- ✔ Comparative analysis with the same gene in other genomes

BLAST the Human Genome - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://www.ncbi.nlm.nih.gov/genome/seq/page.cgi?F=HsBlast.html&&ORG=Hs>

NCBI Genomic Biology Human Genome Guide Human Sequence

Search LocusLink for Go

BLAST
overview
FAQs
news
manual
references

BLAST the Human genome

Compare your query sequence to the working draft sequence of the human genome or its mRNA and protein products.

Database: genome Program tblastn
 use MegaBLAST

Begin Search

Enter an accession, gi, or a sequence in FASTA format:

```
>PNP [Homo sapiens] gi|35565|emb|CAA25320.1|
MENGYTYEDYKNTAEWLLSHTKHRPQVAIICGSLGGLTDKLTQAQIFDYSEIPNFPRST
VPGHAGRLVF
GFLNGRACVMMQGRFHMIEGYPLWKVTFPVRVFHLLGVDTLVVTNAAGGLNPKFEVGDIM
LIRDHINLPG
FSGQNPLRGPNDERFGDRFPAMSDAYDRITMRQRALSTWKQMGEQRELQEGTYVMVAGPSF
```

Optional parameters
[Expect](#) [Filter](#) [Descriptions](#) [Alignments](#)

0.01 default 100 100

Advanced options:

Begin Search Clear Input

Internet

BLAST the Human Genome - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Refresh Stop

Address http://www.ncbi.nlm.nih.gov/genome/seq/page.cgi?F=HsBlast.html&&ORG=Hs Go Links »

NCBI Genomic Biology Human Genome Guide Human Sequence

Search LocusLink for Go

BLAST
 overview
 FAQs
 news
 manual
 references

BLAST the Human genome

Compare your query sequence to the working draft sequence of the human genome or its mRNA and protein products.

Database: genome Program tblastn

use MegaBLAST

Begin Search

Enter an accession in FASTA format:

>PNP [Homo sapiens] cDNA for PNPase, partial
 MENGYTYE
 VPGHAGRL
 GFLNGRAC
 LIRDHINL
 FSGQNPLRGPNDERFGDRFPAMSDAYDRITMRQRALSTWKQMGEQRELQEGTYVMVAGPSF

65 | emb | CAA25320.1 |
 VAIICGSGLGGLTDKLTQAQIFDYSEIPNFPRST
 TF PVRVFHLLGVDTLVVTNAAGGLNPKFEVGDIM

Optional parameters

[Expect](#) [Filter](#) [Descriptions](#) [Alignments](#)

0.01 default 100 100

Advanced options:

Begin Search Clear Input

Internet

BLAST the Human Genome - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Refresh Stop

Address http://www.ncbi.nlm.nih.gov/genome/seq/page.cgi?F=HsBlast.html&&ORG=Hs Go Links »

NCBI Genomic Biology Human Genome Guide Human Sequence

Search LocusLink for Go

BLAST
overview
FAQs
news
manual
references

BLAST the Human genome

Compare your query sequence to the working draft sequence of the human genome or its mRNA and protein products.

Database: genome Program tblastn
 use MegaBLAST
Begin Search

blastn
blastp
blastx
tblastn

Enter an accession, gi, or a sequence in FASTA format:

```
>PNP [Homo sapiens] gi|35565|emb|CAA25320.1|
MENGYTYEDYKNTAEWLLSHTKHRPQVAIICGSLGGLTDKLTQAQIFDYSEIPNFPRST
VPGHAGRLVF
GFLNGRACVMMQGRFHMIEGYPLWKVTFPVRVFHLLGVDTLVVTNAAGGLNPKFEVGDIM
LIRDHINLPG
FSGQNPLRGPNDERFGDRFPAMSDAYDRITMRQRALSTWKQMGEQRELQEGTYVMVAGPSF
```

Optional parameters
Expect Filter Descriptions Alignments
0.01 default 100 100

Advanced options:

Begin Search Clear Input


Internet

NCBI Blast - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Copy Paste Refresh

Address http://www.ncbi.nlm.nih.gov/blast/Blast.cgi Go Links »

 **NCBI** *formatting* **BLAST**

Nucleotide Protein Translations Retrieve results for an RID

Your request has been successfully submitted and put into the Blast Queue.

Query = PNP [Homo sapiens] gi|35565|emb|CAA25320.1| (289 letters)

The request ID is

or

The results are estimated to be ready in 6 minutes but may be done sooner.

Please press "FORMAT!" when you wish to check your results. You may change the formatting options for your result via the form below and press "FORMAT!" again. You may also request results of a different search by entering any other valid request ID to see other recent jobs.

Format

Show [Graphical Overview](#) [Linkout](#) [Sequence Retrieval](#) [NCBI-gi](#) in [format](#)

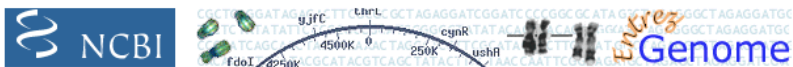
Number of: [Descriptions](#) [Alignments](#)

[Alignment view](#)

[Limit results by entrez query](#) or select from:

[Expect value range:](#)

Internet



TBLASTN 2.2.5 [Nov-16-2002]

Reference:

Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

RID: 1046634425-024719-9809

Query= PNP [Homo sapiens] gi|35565|emb|CAA25320.1|
(289 letters)

Database: Homo sapiens genomic contig sequences
953 sequences; 2,861,589,808 total letters

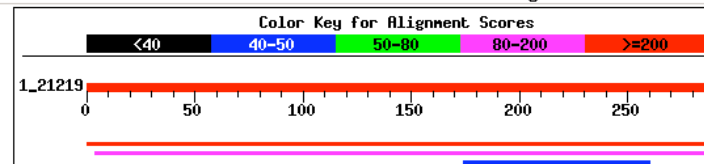
If you have any problems or questions with the results of this search please refer to the [BLAST FAQs](#)

Genome View

Show positions of the BLAST hits in the human genome using the Entrez Genomes MapViewer

Distribution of 7 Blast Hits on the Query Sequence

Mouse-over to show defline and scores. Click to show alignments



Sequences producing significant alignments:		Score	E
		(bits)	Value
ref NT_022184.10 Hs2_22340	Homo sapiens chromosome 2 genom...	327	5e-88
ref NT_037845.1 Hs14_37849	Homo sapiens chromosome 14 genom...	141	9e-62
ref NT_037734.1 Hs9_37738	Homo sapiens chromosome 9 genomic...	47	0.001

Alignments

>[ref|NT_022184.10|Hs2_22340](#) Homo sapiens chromosome 2 genomic contig
 Length = 13913408

Score = 327 bits (838), Expect = 5e-88
 Identities = 182/298 (61%), Positives = 211/298 (70%), Gaps = 9/298 (3%)
 Frame = -2

```

Query: 1      MENGYTYEDYKNTAEWLLSHTKHRPQVAIICGSGLGGLTDKLTQAQIFDYSEIPNFFRST 60
              MENGYTYEDY++TAEWLL HTKH  QV +ICGS LG LTDKL QAQIF+ SE+ NF +ST
Sbjct: 6971389 MENGYTYEDYQSTAEWLLFHTKH*TQVTVICGSELGDLTDKLIQAQIFNNSMLNFFQST 6971210

Query: 61      VPGHAGRLVFGFLNGRACVMMQGRFHMIEGYPLWKVTFPVRVFHLLGVDTLVVTNAAGGL 120
              VPGHA LVFGLNG CVMMQGRF++Y+GY LW + F VF LLG + LV T+AAGGL
Sbjct: 6971209 VPGHAV*LVFGLNGTVCVMMQGRFYLYDGYLLWNMIFLHEVFQLLGGNILVATDAAGGL 6971030

Query: 121     NPKFEVGDIMLIRDHINLPGFSGQNPLRGPNDERFGDRFPAMSDAYDRITMRQALSTWKQ 180
              NPK EVG IML+ DHI L GF QN +GPNDERFG FPA SDAY+ TM+Q+AL++ Q
Sbjct: 6971029 NPKSEVGRIMLLCDHIKLLGFCDQNSPKGPNDRFGVHFPATSDAYNWTMKQKALNS*NQ 6970850

Query: 181     MGEQRELQEGTYVMVAGPSFETVAECRVLQKLGADAVGMSTVPEVIV--ARHCGLRVFG- 237
              MG+Q+E+Q+ TYVM +FET G D+ + A+H
Sbjct: 6970849 MGKQQEVQKDTYVMAVNCNFET-----GRDSSDAEAGDGCCLA*AQHQS*SCMAL 6970700

Query: 238     -----FSLITNKVIMDYESLEKANHEEVLAAAGKQAAQKLEQFVSILMASIPLPKAS 289
              FSLITNKVIMDYESL+KANHE V A KQAAQKLEQFVSIL ASIPLD A+
Sbjct: 6970699 WTWSLCFSLITNKVIMDYESLKKANHE*V*EAVKQAAQKLEQFVSILKASIPLPDNAN 6970526
    
```

Entrez Genome view - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Copy Paste

Address http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?chr=hum_chr.inf&RID=1046634425-024719-9809&Cl Go Links

NCBI Entrez Genome

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM Help

Search for on chromosome(s) haplotype

Find

Show linked entries [Help](#) [FTP](#) Advanced search

[Homo sapiens genome view](#) [build 31](#) [BLAST search the human genome](#)

Hit GIs:
Hits: 1

Hit GIs:
Hits: 5

Color key for scores: < 40 40-50 50-80 80-200 ≥ 200 [Back to BLAST alignments page](#)

BLAST search results: 3 BLAST hits found

Query PNP [Homo sapiens] gi|35565|emb|CAA25320.1

[Sort results by score](#)

Chr	Hit GI	Hits	Score	E value	Map element
2	27481183	1	327	5e-88	NT_022184 Homo sapiens chromosome 2 genomic contig
9	27481904	1	47.4	0.001	NT_037734 Homo sapiens chromosome 9 genomic contig
14	27499846	5	141	4e-32	NT_037845 Homo sapiens chromosome 14 genomic contig

Internet



Ensembl Human Genome Browser (BlastView) - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Copy Paste

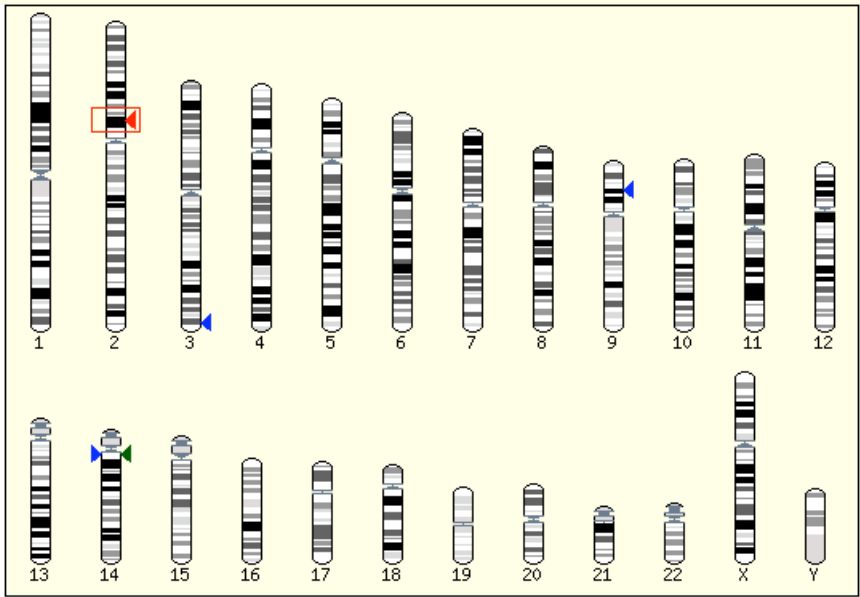
Address http://www.ensembl.org/Homo_sapiens/blastview?id=hs_s3a69Ev66Yg3&format=karyo_format Go Links

Google Search Web Search Site News PageRank Page Info Up Highlight

e! Ensembl Human BLASTView  

Home Human What's New BLAST SSAHA EnsMart Export Data Download Disease Browser Docs

Find All **Lookup** [e.g. AP000462, RH9632, cancer] **Help**



Blast score ranges for this search: [The highest scoring hit(s) are boxed]

42 - 308	309 - 575	576 - 842
----------	-----------	-----------

Location of Blast hits

Internet

Human BLAT Search

BLAT Search Genome

Genome: Assembly: Query type: Sort output: Output type:

Please paste in a query sequence to search in the genome. Multiple sequences can be searched at once if separated by > and the sequence name.

- BLAT's guess
- DNA
- protein
- translated RNA
- translated DNA

```
>PNP [Homo sapiens] gi|35565|emb|CAA25320.1|
MENGYTYEDYKNTAEWLLSHTKHRPQVAIICGSGLGGGLTDKLTQAQIFDYSEIPNFPRSTVPGHAGRLVF
GFLNGRACVMMQGRFHMVEGYPLWKVTFPVRVVFHLLGVDTLVVTNAAGGLNPKFEVGDIMLIRDHINLPG
FSGQNPLRGPNDERFGDRFPAMSDAYDRITMRQALSTWKQMGEORELQEGTYVMVAGPSFETVAECRVLQ
KLGADAVGMSTVPEVIVARHCGLRVFGFSLITNKVIMDYESLEKANHEEVLAAGKQAAQKLEQFVSILMA
SIPLPKAS
```

Rather than pasting a sequence, you can choose to upload a text file containing the sequence.

Upload sequence:

Only DNA sequences of 25,000 or less bases and protein or translated sequence of 5000 or less letters will

Human BLAT Results - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Copy Paste

Address http://genome.ucsc.edu/cgi-bin/hgBlat Go Links

Google Search Web Search Site News PageRank Page Info Up Highlight

Home - Genome Browser - Blat Search - Table Browser - FAQ - User Guide

Human BLAT Results

BLAT Search Results

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END
browser details	PNP	279	4	289	289	99.7%	14	++	14727930	14732220
browser details	PNP	101	0	288	289	70.0%	2	+ -	76697947	76698808

Done Internet

Entrez Map View - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://www.ncbi.nlm.nih.gov/mapview/maps.cgi?org=hum&MAPS=loc%2Ccntg-r&db=genome%2C> Go Links >>

Homo sapiens Map View build 31 BLAST
the Human Genome

Chromosome: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#) [12](#) [13](#) [[14](#)] [15](#)
[16](#) [17](#) [18](#) [19](#) [20](#) [21](#) [22](#) [X](#) [Y](#)

Query: [BLAST](#): PNP [Homo sapiens]
 gi|35565|emb|CAA25320.1| [\[clear\]](#)

Color Key for Alignment Scores: <40 40-50 50-80 80-200 >=200

Master Map: Contig **Maps & Options**

Total Contigs On Chromosome: 7 [[1 not localized](#)]
 Region Displayed: 14,727K-14,732K
 bp [Download/View Sequence/Evidence](#)
 Contigs Labeled: 6 Total Contigs in Region: 6

Gen... Contig accession orient

Contig	Accession	Orientation	Identity	Query
Blast hit			94%	5..62
Blast hit			97%	60..96
Blast hit			98%	95..154
Blast hit			100%	154..217
NT_037845.1				
Blast hit			100%	218..288

Region Shown:

out zoom in

ideogram master

Error on page. Internet



Ensembl Human Genome Browser (ContigView) - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Copy Paste

Address http://www.ensembl.org/Homo_sapiens/contigview?chr=14&vc_start=18265719&vc_end=18365719&f Go Links »

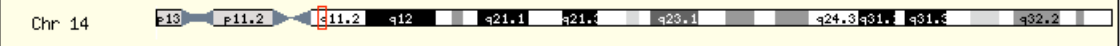
Google Search Web Search Site News PageRank Page Info Up Highlight

e! Ensembl Human ContigView  

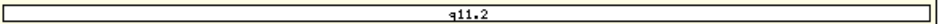
Home Human What's New BLAST SSAHA EnsMart Export Data Download Disease Browser Docs

Find Sequence 14 [e.g. AC067852, AP000869]

Chromosome 14

Chr 14 

Overview

Chromosome band 

17.82 Mb 18.82 Mb

DNA(contigs) < AL359218 AL163152 > AL356419 > < AL355075 > < AL163195 AL163636 > AL133371 > <

Markers

D14S964	D19667	D14S1338	D14S1291	D14S1192
AFM2384d6		D14S1294	D14S1332	D14S72
D14S261	D2S2636			D14S72
AFM2384d6				D14S72
D14S261				

Genes

LNOVEL	LQ96R33	LQ96HF9	LTEP1	LNOVEL	LQ96PS6	LQ8TR1	LRNASE4	LEP3A_HUMAN	LRNASE3	LRN
LQ96R71		LNM_021178	LADPRTL2	LOSSEP	LNOVEL	LNOVEL	LNOVEL	LRNASE6	LRNASE1	LQ12762
		LQ96HF9	LQ96HF9	LQ96HF9	LQ96HF9	LQ96HF9	LQ96HF9	LQ96HF9	LQ96HF9	LQ96HF9
		LQ96HF9	LQ96HF9	LQ96HF9	LQ96HF9	LQ96HF9	LQ96HF9	LQ96HF9	LQ96HF9	LQ96HF9

Gene legend ■ ENSEMBL PREDICTED GENES (KNOWN) ■ ENSEMBL PREDICTED GENES (NOVEL)

Internet

Ensembl Human Genome Browser (ContigView) - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Refresh Stop

Address http://www.ensembl.org/Homo_sapiens/contigview?chr=14&highlight=BLAST%3Ahs_s3a69Ev66Yg3&vc_start=1826 Go Links

Google Search Web Search Site News PageRank Page Info Up Highlight

Detailed View

Jump to Chromosome: bp to Refresh

Features ▾ DAS Sources ▾ Repeats ▾ Decorations ▾ Export ▾ Jump to ▾ Image size ▾ Help ▾

Right 2 Mb

Length 100.00 Kb

Rat matches
 Other mRNAs
 Human mRNAs
 Proteins
 Genscans
 Ensembl trans.
 BLAST hits
 DNA(contigs)
 BLAST hits
 Ensembl trans.
 Genscans
 Proteins
 Human mRNAs
 Other mRNAs
 SNPs

Tilepath

■ ENSEMBL PREDICTED GENES (KNOWN) ■ ENSEMBL PREDICTED GENES (NOVEL)
 ■ CODING SNPS ■ UTR SNPS ■ INTRONIC SNPS ■ FLANKING SNPS ■ OTHER SNPS

There are currently 39 tracks switched off; use the menus above the image to turn these on.

http://www.ensembl.org/Homo_sapiens/contigview?chr=14&vc_start=20265719&vc_end=20365719 Internet

Ensembl Human Genome Browser (ContigView) - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Copy Paste

Address http://www.ensembl.org/Homo_sapiens/contigview?chr=14&vc_start=18310719&vc_end=18320719&highlight Go Links

Google Search Web Search Site News PageRank Page Info Up Highlight

Detailed View

Jump to Chromosome: bp to Refresh

Features ▾ DAS Sources ▾ Repeats ▾ Decorations ▾ Export ▾ Jump to ▾ Image size ▾ Help ▾

Length: 18.311 Mb to 18.320 Mb (10.00 Kb scale)

Rat matches
 Other mRNAs
 Human mRNAs
 Proteins
 Genscans
 Ensembl trans.
 BLAST hits
 DNA(contigs) < AL355075
 BLAST hits
 Ensembl trans. Q96P56
 SNPs

Tilepath: RP11-203M5, RP11-14J7

Gene legend: ENSEMBL PREDICTED GENES (KNOWN), ENSEMBL PREDICTED GENES (NOVEL)
 SNP legend: CODING SNPS, INTRONIC SNPS, FLANKING SNPS

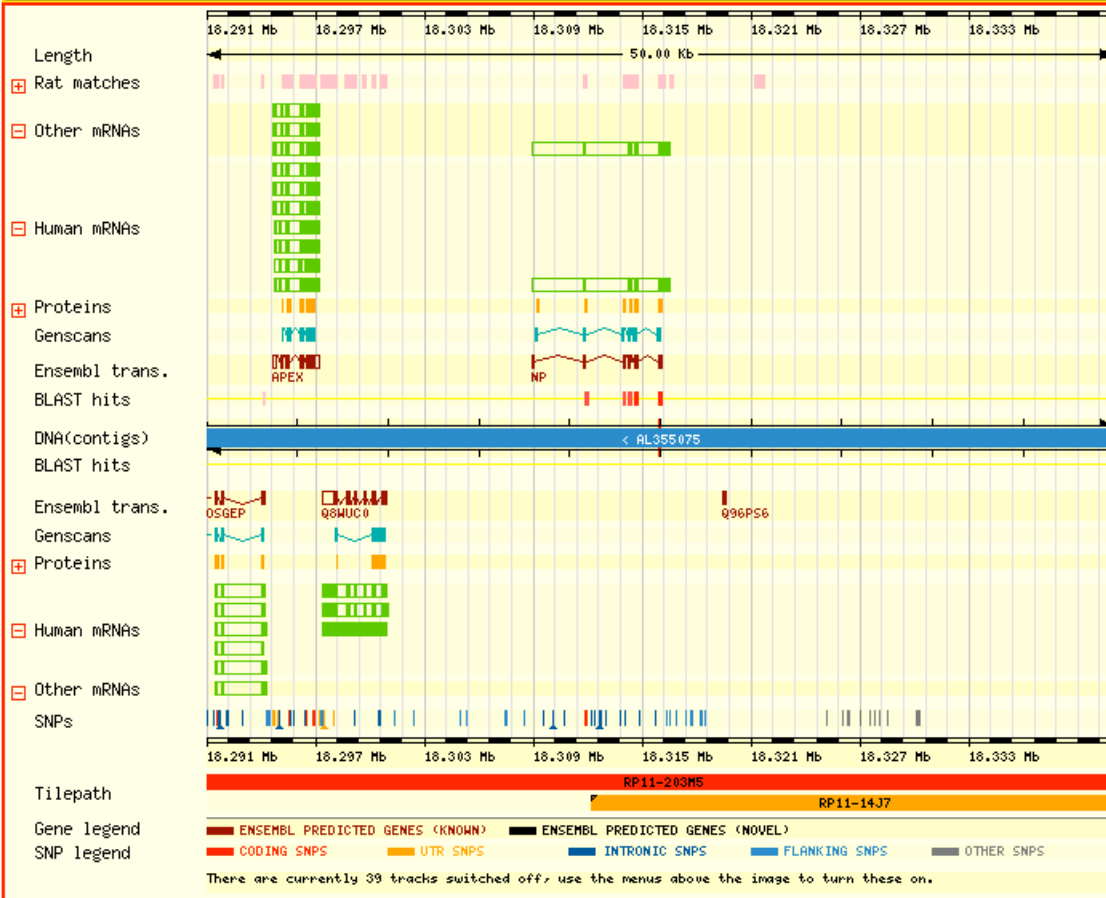
There are currently 39 tracks switched off, use the menus above the image to turn these on.

Internet

Detailed View

Jump to Chromosome: bp to

Features ▾ DAS Sources ▾ Repeats ▾ Decorations ▾ Export ▾ Jump to ▾ Image size ▾ Help ▾



Ensembl Human Genome Browser (SyntenyView) - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://www.ensembl.org/Homo_sapiens/syntenyview?species=Mus_musculus&chr=14&loc=18290719

Google Search Web Search Site News PageRank Page Info Up Highlight

Ensembl Human SyntenyView The Wellcome Trust Sanger Institute EBI

Home Human What's New BLAST SSAHA EnsMart Export Data Download Disease Browser Docs

Find All [e.g. AP000462, RH9632, cancer]

Homo sapiens chromosome 14

Chr 14 Mus musculus

Chr 12 Mus musculus

Human Chromosome 14

Jump to chromosome

[Jump to mapview](#) for chromosome statistics.

Homology Matches

Homo_sapiens Genes	Mus_musculus Homologues
OSGEP (18.29 Mb)	-> Q99LN8 (chr 14 : 42.78 Mb)
APEX (18.29 Mb)	-> Apex1 (chr 14 : 42.79 Mb)
Q8WUC0 (18.30 Mb)	-> ENSMUSG00000035953 (chr 14 : 42.79 Mb)
NP (18.31 Mb)	-> Pnp (chr 14 : 42.81 Mb)
Q96PS6 (18.32 Mb)	
ENSG00000165787 (18.35 Mb)	-> 4930474F22Rik (chr 14 : 42.87 Mb)
Q8TAA1 (18.42 Mb)	
ENSG00000169431 (18.48 Mb)	-> ENSMUSG00000035932 (chr 14 : 42.93 Mb)
RNASE4 (18.52 Mb)	-> Rnase4 (chr 14 : 42.96 Mb)
ANG (18.53 Mb)	-> Angrp (chr 14 : 43.06 Mb)
	-> Ang (chr 14 : 42.97 Mb)
EP3A_HUMAN (18.59 Mb)	-> ENSMUSG00000021878 (chr 14 : 42.98 Mb)

Internet

Ensembl Human Genome Browser (BlastView) - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Copy Paste

Address http://www.ensembl.org/Homo_sapiens/blastview?id=hs_s3a69983966YU38&format=karyo_format Go Links

Google Search Web Search Site News PageRank Page Info Up Highlight

e! Ensembl Human BLASTView The Wellcome Trust Sanger Institute EBI

Home Human What's New BLAST SSAHA EnsMart Export Data Download Disease Browser Docs

Find All **Lookup** [e.g. AP000462, RH9632, cancer] **Help**

1 2 3 4 5 6 7 8 9 10 11 12

13 14 15 16 17 18 19 20 21 22 X Y

Done Internet

Ensembl Human Genome Browser (ContigView) - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Copy Paste

Address http://www.ensembl.org/Homo_sapiens/contigview?chr=2&vc_start=76573222&vc_end=76673222&highlight=BL Go Links »

Google Search Web Search Site News PageRank Page Info Up Highlight

Detailed View

Jump to Chromosome: bp to Refresh

Features ▼ DAS Sources ▼ Repeats ▼ Decorations ▼ Export ▼ Jump to ▼ Image size ▼ Help ▼

Length: 100.00 Kb
 BLAST hits
 DNA(contigs): AC073091 > AC068597 >
 BLAST hits
 Genscans
 Proteins
 Rat matches
 SNPs
 Tilepath: RP11-335E6 RP11-320L2
 SNP legend: OTHER SNPs

There are currently 39 tracks switched off, use the menus above the image to turn these on.

Done Internet

Ensembl Human Genome Browser (ContigView) - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Copy Paste

Address http://www.ensembl.org/Homo_sapiens/contigview?chr=2&vc_start=76618222&vc_end=76628222&highlight=BL Go Links »

Google Search Web Search Site News PageRank Page Info Up Highlight

Detailed View

Jump to Chromosome: bp to Refresh

Features ▼ DAS Sources ▼ Repeats ▼ Decorations ▼ Export ▼ Jump to ▼ Image size ▼ Help ▼

Length: 10.00 Kb
 BLAST hits
 DNA(contigs): AC073091 >
 BLAST hits
 Genscans
 Proteins
 Mouse matches
 Rat matches
 SNPs
 Tilepath: RP11-335E8
 SNP legend: OTHER SNPS

There are currently 38 tracks switched off, use the menus above the image to turn these on.

Done Internet

Ensembl Human Genome Browser (BlastView) - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://www.ensembl.org/Homo_sapiens/blastview?format=hit_format&id=hs_s3a69983966YU3&hit=AC073091.5.1.185174

Google Search Web Search Site News PageRank Page Info Up Highlight

e! Ensembl Human BLASTView The Wellcome Trust Sanger Institute EBI

Home Human What's New BLAST SSAHA EnsMart Export Data Download Disease Browser Docs

Find All [e.g. AP000462, RH9632, cancer]

TBLASTN 2.0a13MP-WashU [10-Jun-1997] [Build 23:08:22 Jun 10 1997]
Query= PNP
(289 letters)
Database: ensembl/Homo_sapiens.latestgp.fa
44521 sequences; 3200338544 total letters
>AC073091.5.1.185174
Length: 185,174

Minus Strand HSPs:

Score = 842 (296.4 bits), Expect = 2.2e-81, P = 2.2e-81
Identities = 181/868 (63%), Positives = 213/868 (74%), Frame = -1

Query: 1 MENGYTYEDYKNTAEWLLSHTKHRPQVAIICGSLGGLTDKLTQAQIFDYSEIPNFPFRST 60
MENGYTYEDY++TAEWLL HTKH QV +ICGS LG LTKL QAQIF+ SE+ NF +ST
Sbjct: 166478 MENGYTYEDYQSTAEWLLPHTKH*TQVTVICGSELGDLTDKLIQAQIFNNSSEMLNFFQST 166299

Query: 61 VPGHAGRLVPGFLNGRACVMMQGRFMYEGYPLWKVTFPVRVPHLLGVDTLVVTNAAGGL 120
VPGHA LVPGLNG CVMMQGRF++Y+GY LW + F VF LLG + LV T+AAAGGL
Sbjct: 166298 VPGHAV*LVPGLNGTVCVMMQGRFYLYDGYLLWNMIFLHEVPQLLGGNIVATDAAGGL 166119

Query: 121 NPKFEVGDIMLIRDHINLPGFSGQNPLRGPNDERFGDRFPAMSDAYDRMTMRQALSTWKQ 180
NPK EVG IML+ DHI L GF QN +GPNDERFG FPA SDAY+ TM+Q+AL++ Q
Sbjct: 166118 NPKSEVGRIMLLCDHIKLLGFCDQNSPKGPNDERFGVHFPATSDAYNWTMKQKALNS*NQ 165939

Query: 181 MGEQRELQEGTYVMVAGPSFETVAECRVLQKLGADAVGMSTVPEVIVARHC -GLRVFG-- 237
MG+Q+E+Q+ TYVM +FET + + D ++ + C L +
Sbjct: 165938 MGKQQEVQKDTYVMVAVNCNFETGRDSSDAE--AGDGCLA*AQHQ--S*SCMALWTWSLC 165771

Query: 238 FSLITNKVIMDYESLEKANHEEVLAAQKAAQKLEQFVSILMASIPLPKAS 289
FSLITNKVIMDYESL+KANHE V A KQAAQKLEQFVSIL ASIPLPD A+
Sbjct: 165770 FSLITNKVIMDYESLKKANHE*V*EAVKQAAQKLEQFVSILKASIPLPDNAN 165615

Done Internet

UCSC Genome Browser v17 - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Home Search Favorites Media Print Copy Paste

Address http://genome.ucsc.edu/cgi-bin/hgTracks?position=chr14:14727931-14732220&db=hg138&ss=../trash/hgSs_genome Go Links

Google Search Web Search Site News PageRank Page Info Up Highlight

Home BLAT DNA Tables Convert Map View PDF/PS Guide

UCSC Genome Browser on Human Nov. 2002 Freeze

move <<< << < > >> >>> zoom in 1.5x 3x 10x zoom out 1.5x 3x 10x

position size 4,290 image width jump

move start > Click on a feature for details. Click on base position to zoom in around cursor. Click on left mini-buttons for track-specific options

move end < >

Internet

UCSC Genome Browser v17 - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://genome.ucsc.edu/cgi-bin/hgTracks?position=chr14:14727931-14732220&db=hg13&ss=../trash/hgSs_genomr Go Links

Google Search Web Search Site News PageRank Page Info Up Highlight

Mapping and Sequencing Tracks

[Base Position](#) on
[Chromosome Band](#) dense
[STS Markers](#) dense
[FISH Clones](#) hide
[GenMapDB Clones](#) hide
[Recomb Rate](#) hide
[Map Contigs](#) hide
[Assembly](#) hide
[Gap](#) dense
[Coverage](#) hide
[BAC End Pairs](#) hide
[Fosmid End Pairs](#) hide
[GC Percent](#) hide
[BLAT Sequence](#) full

Genes and Gene Prediction Tracks

[Known Genes](#) pack
[RefSeq Genes](#) hide
[Acembly Genes](#) dense
[Twinscan](#) dense
[SGP Genes](#) dense
[Fgenesh++ Genes](#) hide
[Geneid Genes](#) hide
[Genscan Genes](#) pack

mRNA and EST Tracks

[Human mRNAs](#) pack
[Spliced ESTs](#) dense
[Human ESTs](#) hide
[Nonhuman mRNA](#) dense
[Nonhuman EST](#) hide
[UniGene](#) hide
[Gene Bounds](#) hide

Expression and Regulation

[CpG Islands](#) hide

Comparative Genomics

[Fugu Blat](#) dense
[Mouse Cons](#) full
[Tight Mouse](#) hide
[Best Mouse](#) dense
[Blastz Mouse](#) hide
[Mouse Synteny](#) full
[Rat Synteny](#) hide

Variation and Repeats

[Overlap SNPs](#) dense
[Random SNPs](#) dense
[RepeatMasker](#) dense
[Simple Repeats](#) hide

Internet

UCSC Genome Browser v17 - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://genome.ucsc.edu/cgi-bin/hgTracks?position=chr2:76697948-76698808&db=hg13&ss=../trash/hgS>

Home BLAT DNA Tables Convert Map View PDF/PS Guide

UCSC Genome Browser on Human Nov. 2002 Freeze

move <<< << < > >> >>> zoom in 1.5x 3x 10x zoom out 1.5x 3x 10x

position size 861 image width jump

Base Position	76698100 76698200 76698300 76698400 76698500 76698600 76698700 76698800
Chromosome Band	Chromosome Bands Localized by FISH Mapping Clones
STS Markers	STS Markers on Genetic (blue) and Radiation Hybrid (black) Maps
Gap	Gap Locations
PNP	Your Sequence from BLAT Search
Assembly Genes	Known Genes based on SWISS-PROT, TrEMBL, mRNA, and RefSeq Assembly Gene Predictions With Alt-splicing
Twinscan	Twinscan Gene Predictions Using Mouse/Human Homology
SGP Genes	SGP Gene Predictions Using Mouse/Human Homology GENSCAN Gene Predictions
Spliced ESTs	Human mRNAs from Genbank Human ESTs That Have Been Spliced
Nonhuman mRNA	Nonhuman mRNAs from Genbank
Fugu Blat	Takifugu rubripes Translated Blat Alignments
Mouse Cons	Human/Mouse Evolutionary Conservation Score (std units)
Best Mouse	BLASTZ Mouse(Feb02) Best in Genome Alignments
chr5	Human/Mouse Synteny using blastz single coverage (100k window)
Overlap SNPs	Single Nucleotide Polymorphisms (SNPs) from Clone Overlaps
Random SNPs	Single Nucleotide Polymorphisms (SNPs) from Random Reads
RepeatMasker	Repeating Elements by RepeatMasker

move start < 2.0 > Click on a feature for details. Click on base position to zoom in around cursor. Click on left mini-buttons for track-specific options

move end < 2.0 >

Internet



Gene Information

✓ BLink

- BLAST Hits

✓ Domains

- Protein domains

✓ Links

- Varies with available information

✓ LinkOut

- “Custom” links to other relevant databases

NCBI Sequence Viewer - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Copy Paste

Address http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=protein&list_uids=130377&dopt=GenPept Go Links »

NCBI Entrez Protein

PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books

Search Protein for Go Clear

Limits Preview/Index History Clipboard Details

Display default Show: 20 Send to File Get Subsequence

1: P00491. Purine nucleoside...[gi:130377] [BLink, Domains, Links](#)

LOCUS P00491 289 aa linear PRI 15-JUN-2002

DEFINITION Purine nucleoside phosphorylase (Inosine phosphorylase) (PNP).

ACCESSION P00491

VERSION P00491 GI:130377

DBSOURCE swissprot: locus PNP_HUMAN, accession P00491;
class: standard.
extra accessions:Q15160,created: Jul 21, 1986.
sequence updated: Jul 21, 1986.
annotation updated: Jun 15, 2002.
xrefs: gi: [35564](#), gi: [35565](#), gi: [190150](#), gi: [387033](#), gi: [190147](#),
gi: [190148](#), gi: [190149](#), gi: [66583](#), gi: [230387](#), gi: [230388](#)
xrefs (non-sequence databases): Aarhus/Ghent-2DPAGE2108, MIM
[164050](#), InterProIPR001369, PfamPF00896, PROSITEPS01240

KEYWORDS Transferase; Glycosyltransferase; Polymorphism; Disease mutation;
3D-structure.

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

REFERENCE 1 (residues 1 to 289)
AUTHORS Williams,S.R., Goddard,J.M. and Martin,D.W. Jr.
TITLE Human purine nucleoside phosphorylase cDNA sequence and genomic
clone characterization
JOURNAL Nucleic Acids Res. 12 (14), 5779-5787 (1984)
MEDLINE [84272252](#)
PUBMED [6087295](#)
REMARK SEQUENCE FROM N.A.

REFERENCE 2 (residues 1 to 289)
AUTHORS Williams,S.R., Gekeler,V., McIvor,R.S. and Martin,D.W. Jr.
TITLE A human purine nucleoside phosphorylase deficiency caused by a

Done Internet


[BLAST](#) [Protein](#) [Structure](#) [PubMed](#) [Taxonomy](#)
[Genome](#) [Nucleotide](#) [3D-Domains](#) [Books](#) [Help](#)

Query: [gi|130377](#) purine-nucleoside phosphorylase (EC 2.4.2.1) [validated] - human
 Matching gi: [35565](#), [4557801](#), [66583](#), [230387](#), [230388](#)

[COG0005](#) assigned by Cognitor (35 best hits)




















[Best hits](#) [Common Tree](#) [Taxonomy Report](#) [3D structures](#) [CDD-Search](#) [GI list](#)

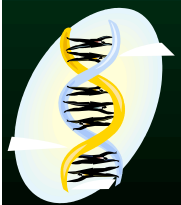
148 BLAST hits to 98 unique species [Sort by taxonomy proximity](#)

22 [Archaea](#) 79 [Bacteria](#) 42 [Metazoa](#) 2 [Fungi](#) 0 [Plants](#) 0 [Viruses](#) 3 [Other Eukaryotae](#)

Keep only Cut-Off [Select](#) [Reset](#)

289 aa


	SCORE	P	ACCESSION	GI	PROTEIN DESCRIPTION
	1515	27	AAA36460	387033	purine nucleoside phosphorylase [Homo sapiens]
	1501	27	BAC05327	21758578	unnamed protein product [Homo sapiens]
	1341	21	1B8NA	4558113	Chain A, Purine Nucleoside Phosphorylase
	1341	21	P55859	3287982	Purine nucleoside phosphorylase (Inosine phospho
	1335	21	1FXUA	11514560	Chain A, Purine Nucleoside Phosphorylase From Ca
	1334	21	AAB34886	1042206	purine nucleoside phosphorylase, PNP, purine nuc
	1332	21	1A9Q	3402089	Chain , Bovine Purine Nucleoside Phosphorylase (
	1331	21	1VFN	2624420	Chain , Purine Nucleoside Phosphorylase
	1329	21	1A9T	3318947	Chain , Bovine Purine Nucleoside Phosphorylase (
	1329	21	1PBN	1311143	Chain , Purine Nucleoside Phosphorylase
	1324	21	1A9Q	3402091	Chain , Bovine Purine Nucleoside Phosphorylase (
	1290	21	CAA39888	53750	purine-nucleoside phosphorylase [Mus musculus]
	1290	21	AAC37635	388921	purine nucleoside phosphorylase
	1287	21	AAA39835	200098	purine nucleoside phosphorylase
	1282	21	AAC37706	388923	purine nucleoside phosphorylase
	1267	21	BAB25491	12842148	unnamed protein product [Mus musculus]
	1001	21	XP_214155	27674996	similar to purine-nucleoside phopshorylase [Mus r
	814	8	EAA11700	21299555	agCP6049 [Anopheles gambiae str. PEST]
	760	8	AAF47654	7292245	CG16758-PB [Drosophila melanogaster]



NCBI CD Summary - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?INPUT_TYPE=precalc&SEQUENCE=130377 Go Links »

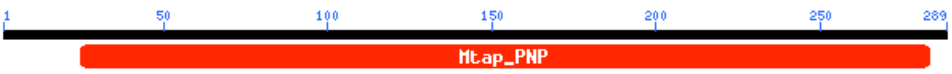
 **NCBI Conserved Domain Summary**

[New Search](#) [PubMed](#) [Nucleotide](#) [Protein](#) [Structure](#) **CDD** [Taxonomy](#) [Help?](#)

Query= [gi|130377|sp|P00491|PNPH_HUMAN](#) Purine nucleoside phosphorylase (Inosine phosphorylase) (PNP)
(289 letters)

Database: cdd.v.1.60

[gnl|CDD|4371 pfam00896, Mtap_PNP, Phosphorylase... S= 295 E=3e-](#)



[Show](#) Domain Relatives [gnl|CDD|4371 pfam00896, Mtap_PNP, Phosphorylase family 2](#) details

[Help](#) | [Disclaimer](#) | [Write to the Help Desk](#)
[NCBI](#) | [NLM](#) | [NIH](#)

NCBI Sequence Viewer - Microsoft Internet Explorer

Address http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=protein&list_uids=130377&dopt=GenPept

NCBI Entrez Protein

PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books

Search Protein for Go Clear

Limits Preview/Index History Clipboard Details

Display FASTA Show: 20 Send to File Get Subsequence

1: P00491. Purine nucleoside...[gi:130377]

LOCUS P00491 289 aa linear PRI 15-J

DEFINITION Purine nucleoside phosphorylase (Inosine phosphorylase) (PNP)

ACCESSION P00491

VERSION P00491 GI:130377

DBSOURCE swissprot: locus PNP_HUMAN, accession P00491;
class: standard.
extra accessions:Q15160,created: Jul 21, 1986.
sequence updated: Jul 21, 1986.
annotation updated: Jun 15, 2002.
xrefs: gi: [35564](#), gi: [35565](#), gi: [190150](#), gi: [387033](#), gi: [190147](#),
gi: [190148](#), gi: [190149](#), gi: [66583](#), gi: [230387](#), gi: [230388](#)
xrefs (non-sequence databases): Aarhus/Ghent-2DPAGE2108, MIM
[164050](#), InterProIPR001369, PfamPF00896, PROSITEPS01240

KEYWORDS Transferase; Glycosyltransferase; Polymorphism; Disease mutation;
3D-structure.

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



REFERENCE 1 (residues 1 to 289)
AUTHORS Williams,S.R., Goddard,J.M. and Martin,D.W. Jr.
TITLE Human purine nucleoside phosphorylase cDNA sequence and genomic
clone characterization

Links
 ▶ Related Sequences
 ▶ Domain Relatives
 ▶ OMIM
 ▶ PubMed
 ▶ Taxonomy
 ▶ LinkOut

OMIM - NUCLEOSIDE PHOSPHORYLASE; NP - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=164050> Go Links »

 **OMIM**  Johns Hopkins University
Online Mendelian Inheritance in Man

PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM

Search for Go Clear

Limits Preview/Index History Clipboard Details

Display Show: Send to

***164050** Links

NUCLEOSIDE PHOSPHORYLASE; NP

Alternative titles; symbols

**PURINE-NUCLEOSIDE:ORTHOPHOSPHATE RIBOSYLTRANSFERASE; PNP
NUCLEOSIDE PHOSPHORYLASE DEFICIENCY, INCLUDED
ATAXIA WITH DEFICIENT CELLULAR IMMUNITY, INCLUDED**

Gene map locus [14q13.1](#)

TEXT

[Edwards et al. \(1971\)](#) described electrophoretic variants of nucleoside phosphorylase ([EC 2.4.2.1](#)), the enzyme that catalyzes the phosphorolytic cleavage of inosine to hypoxanthine. The enzyme appeared to be a trimer. Family studies indicated autosomal codominant inheritance of the variants. [Zannis et al. \(1978\)](#) and [Williams et al. \(1984\)](#) demonstrated that human PNP is a symmetric trimer composed of 3 identical 32,153-Da subunits, each with a substrate-binding site. PNP reversibly catalyzes the phosphorolysis of the purine nucleosides, (deoxy)inosine and (deoxy)guanosine, to their respective purine bases and the corresponding ribose-1-phosphate. 💡

NCBI

MIM *164050
Text
Allelic Variants
• View List
See Also
References
Contributors
Creation Date
Edit History

• Clinical Synopsis
• Gene map


LocusLink
N Nomenclature
R RefSeq
G GenBank
P Protein
U UniGene

LinkOut
...CCR
...HGMD

OMIM - NUCLEOSIDE PHOSPHORYLASE; NP - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=164050> Go Links »

 NCBI

MIM *164050

Text

Allelic Variants

- View List

See Also

References






Contributors

Creation Date



Edit History

- Clinical Synopsis
- Gene map

LocusLink

-  Nomenclature
-  RefSeq
-  GenBank
-  Protein
-  UniGene

LinkOut

-  CCR
-  HGMD

Deficiency of nucleoside phosphorylase results in defective T-cell immunity ([Giblett et al., 1975](#)). This may not be surprising since deficiency of adenosine deaminase, the next enzyme in the pathway, results in combined immune deficiency disease ([102700](#)). Absence of red cell NP was observed in a child with severe T-cell immunodeficiency. The parents were consanguineous and showed less than half the normal activity of the enzyme in their red cells ([Berglund et al., 1975](#)). In a patient with deficiency of nucleoside phosphorylase, [Cohen et al. \(1976\)](#) found severe hypouricemia and hypouricosuria, but excessive amounts of purines (mainly inosine and guanosine) in the urine. The immune defect was thought to be related to inhibition of adenosine deaminase by inosine. [Mitchell et al. \(1978\)](#) found that deoxyadenosine and deoxyguanosine are particularly toxic to T cells but not to B cells. Addition of deoxycytidine or dipyridamole prevented deoxyribonucleoside toxicity. [Stoop et al. \(1977\)](#) studied a 15-month-old girl, 2 sisters of whom had died of immunodeficiency. NP was lacking from red cells and lymphocytes. The parents and a normal brother had intermediate levels. Both T cells and B cells were normal at birth, but thereafter a gradual decrease in T-cell immunity occurred. The patient showed high inosine and guanosine levels in the blood, as well as hypouricemia and hypouricosuria. Spastic tetraparesis was present. In one patient with severely defective T-cell function and normal B-cell function, [Osborne et al. \(1977\)](#) found no detectable red cell NP and no detectable immunologically reactive material. The parents, second cousins, had less than half the normal enzyme activity. Two patients in a second family had 0.5% residual enzyme activity and about half-normal immunologically reactive material. The parents, who were not related, showed electrophoretically different mutant enzymes that were also different from those in the first family. Thus the affected children in the second family were genetic compounds, not true homozygotes. In T cells, the absence of PNP activity is thought to lead to an accumulation of deoxyguanosine triphosphate, which inhibits the enzyme ribonucleotide reductase ([Mitchell et al., 1978](#); [Ullman et al., 1979](#)). This inhibition blocks DNA synthesis, thereby preventing the cellular proliferation required for an immune response. ☹


The immune defect from NP deficiency is often accompanied by a neurologic disorder. [Watson et al. \(1981\)](#) reported the case of a 2.5-year-old boy who died of malignant lymphoma of the B-immunoblastic type. He had spastic tetraplegia also. [Rijksen et al. \(1987\)](#) described a case in a 3-year-old boy who was admitted for investigation of a behavior disorder and spastic diplegia. Severe lymphopenia was found; however, clinical symptoms of immune deficiency did not become apparent until the age of 4 years. [Stephenson and Tolmie \(1990\)](#) informed me that the family reported by [Graham-Pole et al. \(1975\)](#) as having 'familial dysequilibrium-diplegia with T-lymphocyte deficiency' ([209000](#)) turned out to have PNP deficiency. The condition was diagnosed retrospectively from stored fibroblasts from an affected child and from demonstration that both parents had half-normal activity of PNP. [Stephenson and Tolmie \(1990\)](#) were prompted to restudy this family after diagnosing PNP deficiency in a young girl who presented with dysequilibrium syndrome with pyramidal signs (extensor plantar responses and exaggerated reflexes but not prominent spasticity) very similar to the neurologic picture in the family reported by [Graham-Pole et al. \(1975\)](#). The child had defective cell-mediated immunity and died of lymphoma shortly after her third birthday. ☹

Although early studies suggested that B-cell function is normal or even increased in PNP deficiency, later studies showed that B-cell function can be disrupted as well ([Markert, 1991](#)). This was the case in a patient in whom the nature of the molecular defects was demonstrated by [Aust et al. \(1992\)](#): she had normal B-cell counts but significantly depressed immunoglobulin levels. ☹

OMIM - NUCLEOSIDE PHOSPHORYLASE; NP - Microsoft Internet Explorer






File Edit View Favorites Tools Help



Address <http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=164050> Go Links »

 **NCBI**

MIM *164050
Text
Allelic Variants
• View List
See Also
References
Contributors
Creation Date
Edit History

• Clinical Synopsis
• Gene map

LocusLink
 Nomenclature
 RefSeq
 GenBank
 Protein
 UniGene

LinkOut
 CCR
 HGMD

ALLELIC VARIANTS (selected examples)

.0001 NUCLEOSIDE PHOSPHORYLASE DEFICIENCY [NP, GLU89LYS]

[Williams et al. \(1987\)](#) cloned the mutant gene from an NP-deficient patient who was the offspring of a consanguineous mating. A single base difference was found in the coding region of the mutant gene, a G-to-A transition in the third exon. This single base mutation altered the codon at position 89 from glu-to-lys, a result consistent with previously published peptide mapping data. The patient was demonstrated to be homozygous for the single base mutation on the basis of hybridization of synthetic oligomers to genomic DNA digests. 💡

.0002 NUCLEOSIDE PHOSPHORYLASE DEFICIENCY [NP, ALA174PRO]

[Markert and Barrett \(1989\)](#) demonstrated a G-to-C change of nucleotide 520, resulting in a substitution of proline for alanine as amino acid 174. The other allele carried the mutation described by [Williams et al. \(1987\)](#), namely, a G-to-A change of nucleotide 265, resulting in a glu-to-lys change in amino acid 89 ([164050.0001](#)). [Markert \(1992\)](#) indicated that when site-directed mutagenesis was used to create this mutation and the mutant allele was expressed in COS cells, it was found to have normal function. The possibility remains, however, that the mutation was the cause of the patient's clinical disorder, with an abnormality in protein stability or other posttranscriptional stages. 💡


.0003 NUCLEOSIDE PHOSPHORYLASE DEFICIENCY [NP, ASP128GLY]

In a patient with nucleoside phosphorylase deficiency, [Aust et al. \(1992\)](#) found an asp128-to-gly substitution in the maternal allele and an arg234-to-pro mutation ([164050.0004](#)) in the paternal allele. In addition, the patient was homozygous for a ser51-to-gly substitution ([164050.0005](#)), which is a polymorphism. In order to prove that the 2 mutations were responsible for the disease state, each of the 3 mutations was constructed separately by site-

OMIM - NUCLEOSIDE PHOSPHORYLASE; NP - Microsoft Internet Explorer






File Edit View Favorites Tools Help



Address <http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=164050> Go Links »

 NCBI

MIM *164050
Text
Allelic Variants
• View List
See Also
References
Contributors
Creation Date
Edit History

• Clinical
Synopsis
• Gene map

LocusLink
 Nomenclature
 RefSeq
 GenBank
 Protein
 UniGene

LinkOut
 CCR
 HGMD

REFERENCES

1. Aitken, D. A.; Ferguson-Smith, M. A. :
Regional assignment of nucleoside phosphorylase by exclusion to 14q13. *Cytogenet. Cell Genet.* 22: 490-492, 1978.
PubMed ID : [110525](#)
2. Allderdice, P. W.; Miller, O. J.; Miller, D. A.; Klinger, H. P. :
Spreading of inactivation in an (X;14) translocation. *Am. J. Med. Genet.* 2: 233-240, 1978.
PubMed ID : [263441](#)
3. Aust, M. R.; Andrews, L. G.; Barrett, M. J.; Norby-Slycord, C. J.; Markert, M. L. :
Molecular analysis of mutations in a patient with purine nucleoside phosphorylase deficiency. *Am. J. Hum. Genet.* 51: 763-772, 1992.
PubMed ID : [1384322](#)
4. Berglund, C.; Ammann, A. J.; Giblett, E. R. :
Characteristics of nucleoside phosphorylase in the parents of a child with deficiency of the enzyme. (Abstract) *Am. J. Hum. Genet.* 27: 17A only, 1975.
5. Carapella De Luca, E.; Stegagno, M.; Dionisi Vici, C.; Paesano, R.; Fairbanks, L. D.; Morris, G. S.; Simmonds, H. A. :
Prenatal exclusion of purine nucleoside phosphorylase deficiency. *Europ. J. Pediat.* 145: 51-53, 1986.
PubMed ID : [3089796](#)
6. Cohen, A.; Doyle, D.; Martin, D. W., Jr.; Ammann, A. J. :
Abnormal purine metabolism and purine overproduction in a patient deficient in purine nucleoside phosphorylase. *New Eng. J. Med.* 295: 1449-1454, 1976.
PubMed ID : [825775](#)



BLAST	Protein	Structure	PubMed	Taxonomy
Genome	Nucleotide	3D-Domains	Books	Help

Query: gi|130377 purine-nucleoside phosphorylase (EC 2.4.2.1) [validated] - human
 Matching gi: [35565](#), [4557801](#), [66583](#), [230387](#), [230388](#)

[COG0005](#) assigned by Cognitor (35 best hits)

- Best hits
- Common Tree
- Taxonomy Report
- 3D structures
- CDD-Search
- GI list

148 BLAST hits to 98 unique species [Sort by taxonomy proximity](#)

22 Archaea 79 Bacteria 42 Metazoa 2 Fungi 0 Plants 0 Viruses 3 Other Eukaryotae

Keep only Cut-Off

289 aa

	SCORE	P	ACCESSION	GI	PROTEIN DESCRIPTION
=====	1515	27	AAA36460	387033	purine nucleoside phosphorylase [Homo sapiens]
=====	1501	27	PAC05327	21758578	unnamed protein product [Homo sapiens]
=====	1341	21	1B8NA	4558113	Chain A, Purine Nucleoside Phosphorylase
=====	1341	21	F35639	3267982	Purine nucleoside phosphorylase (inosine phosphorylase)
=====	1335	21	1FXUA	11514560	Chain A, Purine Nucleoside Phosphorylase From Calf Sple
=====	1334	21	AAB34886	1042206	purine nucleoside phosphorylase, PNP, purine nucleoside
=====	1332	21	1A9O	3402089	Chain , Bovine Purine Nucleoside Phosphorylase Complex
=====	1331	21	1VFN	2624420	Chain , Purine Nucleoside Phosphorylase
=====	1329	21	1A9T	3318947	Chain , Bovine Purine Nucleoside Phosphorylase Complex
=====	1329	21	1PBN	1311143	Chain , Purine Nucleoside Phosphorylase
=====	1324	21	1A9Q	3402091	Chain , Bovine Purine Nucleoside Phosphorylase Complex
=====	1290	21	CAA39888	53750	purine-nucleoside phosphorylase [Mus musculus]
=====	1290	21	AAC37635	388921	purine nucleoside phosphorylase
=====	1287	21	AAA39835	200098	purine nucleoside phosphorylase
=====	1282	21	AAC37706	388923	purine nucleoside phosphorylase
=====	1267	21	BAB25491	12842148	unnamed protein product [Mus musculus]

NCBI Sequence Viewer - Microsoft Internet Explorer

Address: <http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val=4558113>

NCBI Entrez Protein

PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books

Search **Nucleotide** for

Limits Preview/Index History Clipboard Details

Display **default** Show: **1** **File**

1: 1B8NA. Chain A, Purine N...[gi:4558113]

LOCUS 1B8N_A 284 aa linear MAM 02

DEFINITION Chain A, Purine Nucleoside Phosphorylase.

ACCESSION 1B8N_A

VERSION 1B8N_A GI:4558113

DBSOURCE pdb: molecule 1B8N, chain 65, release Feb 2, 1999;
deposition: Feb 2, 1999;
class: Transferase;
source: Mol_id: 1; Organism_scientific: Bos Taurus;
Organism_common: Bovine; Organ: Spleen;
Exp. method: X-Ray Diffraction.

KEYWORDS .

SOURCE Bos taurus (cow)

ORGANISM [Bos taurus](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
Bovidae; Bovinae; Bos.

REFERENCE 1 (residues 1 to 284)

AUTHORS Mao,C., Cook,W.J., Zhou,M., Federov,A.A., Almo,S.C. and Ealick,S.E.

TITLE Calf spleen purine nucleoside phosphorylase complexed with
substrates and substrate analogues

JOURNAL Biochemistry 37 (20), 7135-7146 (1998)

MEDLINE [98254498](#)

Links
 ▶ Related Sequences
 ▶ 3D Domains
 ▶ Domain Relatives
 ▶ PubMed
 ▶ **Structure**
 ▶ Taxonomy
 ▶ LinkOut


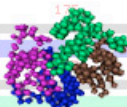
javascript:PopUpMenu2_Set(Menu4558113,"","","");

Internet

Structure Summary - Microsoft Internet Explorer

File Edit View Favorites Tools Help

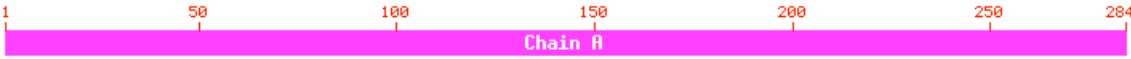

Address <http://www.ncbi.nlm.nih.gov/Structure/mmdb/mmdbsrv.cgi?form=6&db=t&Dopt=s&uid=13072> Go Links »

 **MMDB**
Structure Summary 

PubMed BLAST Structure Taxonomy OMIM **Help?** Cn3d

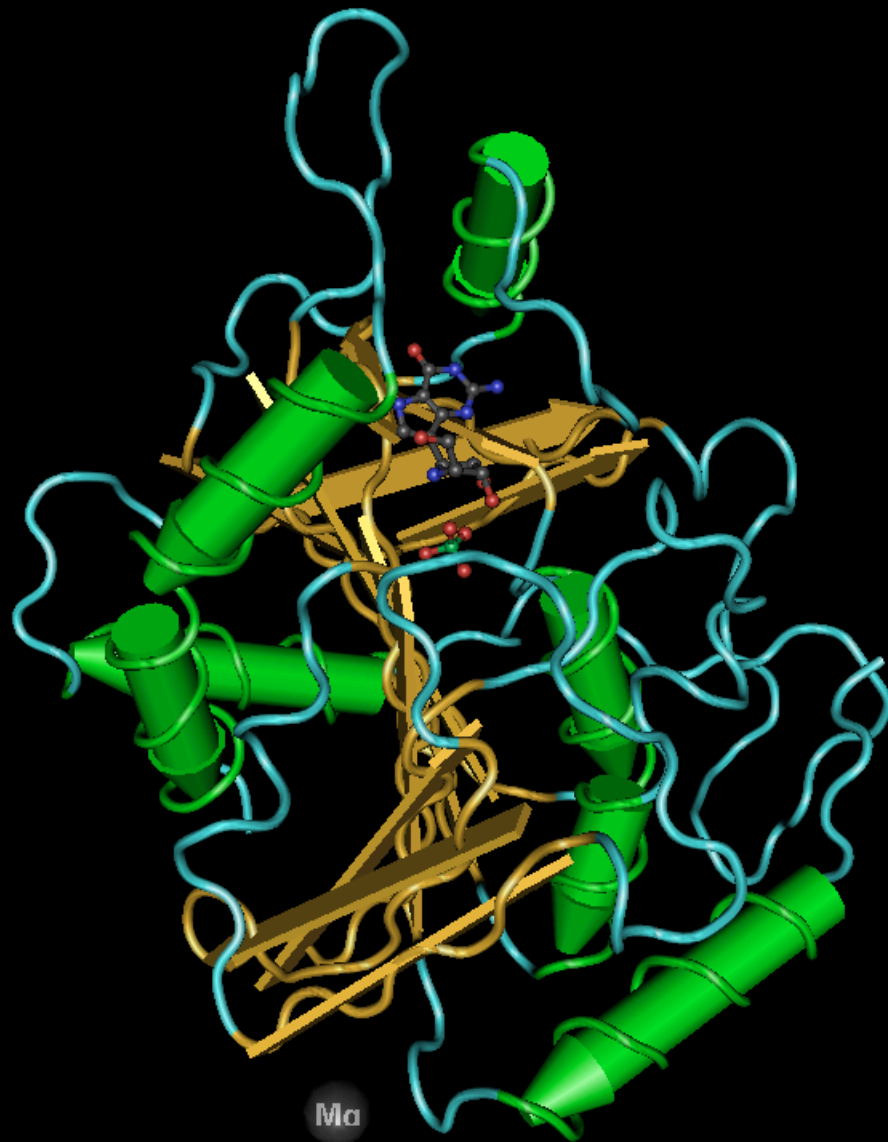
Description: Purine Nucleoside Phosphorylase.
Deposition: A.A.Fedorov, G.A.Kicska, E.V.Fedorov, B.V.Strokopytov, P.C.Tyler, R.H.Furneaux, V.L.Schramm & S.C.Almo, 2-Feb-99
Taxonomy: [Bos taurus](#)
Reference: [PubMed](#) MMDB: [13072](#) PDB: [1B8N](#)

View 3D Structure of with NEW [Get Cn3D 4.1!](#)

[Protein](#)  Chain A
[CDs](#)  Htap_PNP

[Disclaimer](#) | [Write to the Help Desk](#)
[NCBI](#) | [NLM](#) | [NIH](#)

Done Internet





LocusLink - Microsoft Internet Explorer

File Edit View Favorites Tools Help

← Back → Forward Stop Home Search Favorites Media Print Copy Paste Links »

Address [http://www.ncbi.nlm.nih.gov/LocusLink/list.cgi?V=0&Q=4557801\[pgi\]](http://www.ncbi.nlm.nih.gov/LocusLink/list.cgi?V=0&Q=4557801[pgi]) Go

[PubMed](#) [Entrez](#) [BLAST](#) [OMIM](#) [Map Viewer](#) [Taxonomy](#) [Structure](#)

Search: Display: Organism:

Query:

[LocusLink Home](#) [Help](#)

LocusID	Org	Symbol	Description	Position	Links
<input type="checkbox"/> 4860	<i>Hs</i>	NP	nucleoside phosphorylase	14q13.1	P O R G P H U V

[Homogene data](#)

Questions or Comments?
 Write to the [NCBI Service Desk](#)
[Disclaimer](#) [Privacy statement](#)

<http://www.ncbi.nlm.nih.gov/HomoloGene/homolquery.cgi?TEXT=4860> Internet




http://www.ncbi.nlm.nih.gov/HomoloGene/homolquery.cgi?TEXT=4860[loc]&TAXID=9606 - ...

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Copy Paste

Address http://www.ncbi.nlm.nih.gov/HomoloGene/homolquery.cgi?TEXT=4860[loc]&TAXID=9606 Go Links

 **HomoloGene**

PubMed Entrez BLAST OMIM Taxonomy Structure

Search All organisms Go

5 records satisfy the query 4860[loc]


- ▶ ESTs, Highly similar to PNP_HUMAN Purine nucleoside phosphorylase (Inosine phosphorylase) (PNP) [H.sapiens]-Bt.3800
- ▶ ESTs, Weakly similar to PHHUPN purine-nucleoside phosphorylase (EC 2.4.2.1) [validated] - human [H.sapiens]-Dr.3216
- ▶ purine-nucleoside phosphorylase-Pnp
- ▶ nucleoside phosphorylase-NP
- ▶ SPARC-like 1-Sparc11

HOMOLOGENE: [HomePage](#) | [Details](#) | [Release statistics](#) | [Download](#)

GENE INFORMATION: [Locus Link](#) | [UniGene](#)

GENOME RESOURCES [MapViewer Home](#) | [NCBI Entrez Genomes](#)

[Homo sapiens](#) | [Mus musculus](#) | [Rattus norvegicus](#) | [Danio rerio](#)

 [NLM](#) | [NIH](#) | [UniGene](#) | [Privacy Statement](#) | [Disclaimer](#) | [NCBI Help](#)


Internet

http://www.ncbi.nlm.nih.gov/HomoloGene/homol.cgi - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Copy Paste

Address http://www.ncbi.nlm.nih.gov/HomoloGene/homol.cgi Go Links >>



HomoloGene

PubMed
Entrez
BLAST
OMIM
Taxonomy
Structure

Search

HOMOLOGENE ENTRY

Mus musculus	Pnp	purine-nucleoside phosphorylase MapViewer UniGene LocusLink MGI
---------------------	------------	--

POSSIBLE HOMOLOGOUS GENES

Rattus norvegicus	Sparcl1	SPARC-like 1 UniGene
Drosophila melanogaster	CG16758	Drosophila melanogaster CG16758 gene FlyBase UniGene
Danio rerio	Dr.3216	ESTs, Weakly similar to PHHUPN purine-nucleoside phosphorylase (EC 2.4.2.1) [validated] - human [H.sapiens] UniGene
Xenopus laevis	XL.16206	ESTs, Weakly similar to PNP_HUMAN Purine nucleoside phosphorylase (Inosine phosphorylase) (PNP) [H.sapiens] UniGene
Homo sapiens	NP	nucleoside phosphorylase MapViewer LocusLink
Bos taurus	Bt.3800	ESTs, Highly similar to PNP_HUMAN Purine nucleoside phosphorylase (Inosine phosphorylase) (PNP) [H.sapiens] UniGene

Done Internet

http://www.ncbi.nlm.nih.gov/HomoloGene/homol.cgi - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Copy Paste

Address http://www.ncbi.nlm.nih.gov/HomoloGene/homol.cgi Go Links

CALCULATED ORTHOLOGS

Listed below are the nucleotide sequence comparisons used in determining homology. The pairs below represent reciprocal best hits; each alignment is the best one for both organisms. The percent ID below represents identity over an aligned region. The value is a hyperlink to Pairwise Blast, comparing the the indicated sequences. When present, red arrows (→) point out a group of sequence matches which are part of a triplet, being consistent between more than two organisms.

Organism-Gene	Sequence	Percent ID	Sequence	Organism-Gene
→ M.musculus -Pnp	NM_013632.1	92.9	BF543947.1	R.norvegicus - Sparcl1
→ M.musculus -Pnp	NM_013632.1	86.5	BF230403.1	B.taurus - Bt.3800
→ M.musculus -Pnp	NM_013632.1	85.0	NM_000270.1	H.sapiens - NP
→ M.musculus -Pnp	NM_013632.1	73.1	BJ060275.1	X.laevis - Xl.16206
→ M.musculus -Pnp	NM_013632.1			D.melanogaster - CG16758
→ M.musculus -Pnp	NM_013632.1	68.7	BF717799.1	D.rerio - Dr.3216

ADDITIONAL CALCULATED ORTHOLOGS

→ B.taurus -Bt.3800	BF230403.1	90.6	NM_000270.1	H.sapiens - NP
→ R.norvegicus -Sparcl1	BF543947.1	87.6	NM_000270.1	H.sapiens - NP
→ B.taurus -Bt.3800	AV606522.1	84.9	BF543947.1	R.norvegicus - Sparcl1
→ H.sapiens -NP	NM_000270.1			D.melanogaster - CG16758
→ R.norvegicus -Sparcl1	BF543947.1			D.melanogaster - CG16758
→ X.laevis -Xl.16206	BI676077.1	73.3	BF543947.1	R.norvegicus - Sparcl1
→ B.taurus -Bt.3800	AV603243.1			D.melanogaster - CG16758
→ B.taurus -Bt.3800	AV667617.1	71.3	BF717799.1	D.rerio - Dr.3216
→ D.rerio -Dr.3216	BF717799.1	71.2	NM_000270.1	H.sapiens - NP
→ X.laevis -Xl.16206	BJ060275.1	70.5	NM_000270.1	H.sapiens - NP
→ B.taurus -Bt.3800	BE481281.1	70.3	BJ060275.1	X.laevis - Xl.16206
→ X.laevis -Xl.16206	BI676077.1			D.melanogaster - CG16758

Done Internet


```
2:gcg.genome.uab.edu - GCG - SSH Secure Shell
File Edit View Window Help
Quick Connect Profiles
Welcome to the WISCONSIN PACKAGE
Version 10.3-UNIX
Installed on solaris

Copyright (c) 1982 - 2001, Accelrys Inc.
A wholly owned subsidiary of Pharmacoepia, Inc. All rights reserved.

Published research assisted by this software should cite:
Wisconsin Package Version 10.3, Accelrys Inc., San Diego, CA
Databases available:
GenBank Release 133.0 (12/2002)
EMBL (Abridged) Release 73.0 (12/2002)
GenPept Release 133.0 (12/2002)
PIR-Protein Release 74.05 (12/2002)
NRL_3D Release 28.0 (01/2001)
SWISS-PROT Release 40.38 (12/2002)
SP-TREMBL Release 22.0 (10/2002)
PROSITE Release 17.31 (12/2002)
Pfam Release 7.8 (12/2002)
Restriction Enzymes (REBASE) (12/2002)

Technical support see: http://www.accelrys.com/support/
Or call Elliot at (205) 934-1946 or e-mail him at ElliotL@uab.edu
Online help: % genhelp or http://genome.microbio.uab.edu/GCG/

*****
UAB Molecular and Genetic Bioinformatics Facility
For help contact Elliot Lekfowitz
(205) 934-1946
ELLIOTL@uab.edu
or visit our web site: http://www.genome.uab.edu/
*****
> |
```

```
2:gcg.genome.uab.edu - GCG - SSH Secure Shell
File Edit View Window Help
Quick Connect Profiles
> blast sw:pnph_human

BLAST searches one or more nucleic acid or protein databases
for sequences similar to one or more query sequences of any
type. BLAST can produce gapped alignments for the matches it
finds.

          Begin (* 1 *) ?
          End (* 289 *) ?

Search for query in what sequence database:

1) pir      p Protein Information Resource
2) nrl_3d   p NRL_3D Protein Sequence-Structure Database
3) swplus   p SWISS-PROT + SP-TREMBL
4) genembl  n GenBank + EMBL (HTGs Removed)
5) htg      n High Throughput Genomes (HTG from GenBank and EMBL)
6) est_human n Human Expressed Sequence Tags (GenBank and EMBL)
7) est_mouse n Mouse Expressed Sequence Tags (GenBank and EMBL)
8) est_other n All Other Expressed Sequence Tags (GenBank and EMBL)
9) htc      n High Throughput cDNA (GenBank and EMBL)
10) gss     n Genome Survey Sequences (GSS from GenBank and EMBL)
11) genpept p GenPept (Translated GenBank)

Please choose one (* 1 *): 3

Connected to gcg.genome.uab.edu  SSH2 - 3des-cbc - hmac-md5 - 77x36
```

2:gcg.genome.uab.edu - GCG - SSH Secure Shell

File Edit View Window Help

Quick Connect Profiles

Sequences producing significant alignments:

	Score (bits)	E Value
SW:PNPH_HUMAN Begin: 1 End: 289		
!P00491 homo sapiens (human). purine nucleoside pho...	590	e-168
SP_HUM:Q8N7G1 Begin: 1 End: 293		
!Q8n7g1 homo sapiens (human). hypothetical protein ...	582	e-165
SW:PNPH_BOVIN Begin: 1 End: 284		
!P55859 bos taurus (bovine). purine nucleoside phos...	521	e-147
SW:PNPH_MOUSE Begin: 1 End: 289		
!P23492 mus musculus (mouse). purine nucleoside pho...	500	e-141
SP_RO:Q9D8C9 Begin: 16 End: 301		
!Q9d8c9 mus musculus (mouse). adult male small intes...	492	e-138
SP_IN:Q9W004 Begin: 66 End: 350		
!Q9w004 drosophila melanogaster (fruit fly). cg16758...	297	9e-80
SP_IN:Q95TC1 Begin: 72 End: 353		
!Q95tc1 drosophila melanogaster (fruit fly). gh03781...	295	3e-79
SP_IN:Q9BMI9 Begin: 10 End: 285		
!Q9bmi9 schistosoma mansoni (blood fluke). purine-nu...	279	2e-74
SW:PNPH_YEAST Begin: 29 End: 311		
!Q05788 saccharomyces cerevisiae (baker's yeast). p...	265	3e-70
SP_BAP:Q8XHI8 Begin: 8 End: 272		
!Q8xhi8 clostridium perfringens. purine-nucleoside ...	257	8e-68
SP_FUN:Q9UTG1 Begin: 40 End: 312		
!Q9utg1 schizosaccharomyces pombe (fission yeast). ...	256	2e-67
SP_BAP:Q9X259 Begin: 29 End: 292		
!Q9x259 thermotoga maritima. purine nucleoside phos...	251	6e-66
SP_BAP:Q92A55 Begin: 21 End: 271		
!Q92a55 listeria innocua. pnp protein. 10/2002	248	5e-65
SW:PUNA_BACST Begin: 7 End: 271		
!P77834 bacillus stearothermophilus. purine nucleos...	248	5e-65
SP_BAP:Q9KCN8 Begin: 21 End: 269		
!Q9kcn8 bacillus halodurans. purine nucleoside phos...	246	2e-64
SP_BAP:Q8Y5V2 Begin: 21 End: 271		
!Q8y5v2 listeria monocytogenes. pnp protein. 6/2002	245	3e-64

Connected to gcg.genome.uab.edu

SSH2 - 3des-cbc - hmac-md5 - 77x36


```
2:gcg.genome.uab.edu - GCG - SSH Secure Shell
File Edit View Window Help
Quick Connect Profiles
> pileup @pnp.list

FileUp creates a multiple sequence alignment from a group of related
sequences using progressive, pairwise alignments. It can also plot a
tree showing the clustering relationships used to create the alignment.

1      PNPH_HUMAN      289 aa
2      PNPH_BOVIN    284 aa
3      PNPH_MOUSE    289 aa
4      Q9W004         285 aa
5      PNPH_YEAST    283 aa

What is the gap creation penalty (* 8 *) ? 5

What is the gap extension penalty (* 2 *) ? 1

This program can display the clustering relationships graphically.
Do you want to:

  A) Plot to a FIGURE file called "pileup.figure"
  B) Plot graphics on HP7550 attached to /dev/tty15
  C) Suppress the plot

Please choose one (* A *): c

What should I call the output file name (* pnp.msf *) ?

Determining pairwise similarity scores...

1  x  2  4.72
1  x  3  4.45
1  x  4  2.69
1  x  5  2.52
2  x  3  4.45
2  x  4  2.70
```

```

1                                         50
PNPH_HUMAN  MENGTYTYEDY KNTAEWLLSH TKHRPQVAII CGSGLGGLTD KLTQ...AQ
PNPH_BOVIN  MQNGYTYEDY QDTAKWLLSH TEQRPOVAVI CGSGLGGLVN KLTQ...AQ
PNPH_MOUSE  MENEFTYEDY ETTAKWLLQH TEYRPOVAVI CGSGLGGLTA HLKE...AQ
Q9W004      ~EDTYPYEVV EELADFITKG SGMRPKIGII CGSGLGSLAD MIQD...PK
PNPH_YEAST  ~~~~~~F KNTTNF... ..EPPRTLII CGSGLGGIST KLSRDNPPPV

51                                         100
PNPH_HUMAN  IFDYSEIPNF PRSTVPGHAG RLVFGFLNGR ACVM.MQGRF HMYEGYPLWK
PNPH_BOVIN  TFDYSEIPNF PESTVPGHAG RLVFGILNGR ACVM.MQGRF HMYEGYPPWK
PNPH_MOUSE  IFDYNEIPNF PQSTVQGHAG RLVFGLLNGR CCVM.MQGRF HMYEGYSLSK
Q9W004      IFEYEKIPNF PVSTVEGHAG RLVVGTLEG. ATVMAMQGRF HFYEGYPLAK
PNPH_YEAST  TVPYQDIPGF KKSTVPGHSG TLMFGSMNGS PVVL.MNGRL HGYEGNTLFE

101                                        150
PNPH_HUMAN  VTFPVRVFHL LG.VDTLVVT NAAGGLNPKF EVGDI MLIRD HINLPGFSGQ
PNPH_BOVIN  VTFPVRVFRL LG.VETLVVT NAAGGLNPNF EVGDI MLIRD HINLPGFSGE
PNPH_MOUSE  VTFPVRVFHL LG.VETLVVT NAAGGLNPNF EVGDI MLIRD HINLPGFCGQ
Q9W004      CSMPVRVMKL CG.VEYLFAT NAAGGINPRF AVGDI MLIMHD HVNMLGFAGN
PNPH_YEAST  TTFPIRVLNH MGHVRNLIVT NAAGGINAKY QACDI MLCIYD HLNIPGLAGQ

151                                        200
PNPH_HUMAN  NPLRGPNDER FGDRFPAMD AYDRTRMORA LSTWKQMG EQ RELQEGTYVM
PNPH_BOVIN  NPLRGPNEER FGVRFPAMD AYDRDMRQKA HSTWKQMG EQ RELQEGTYVM
PNPH_MOUSE  NPLRGPNDER FGVRFPAMD AYDRDMRQKA FTAWKQMG EQ RKLQEGTYVM
Q9W004      SPLQGPNDER FGPRFPALVN SYNKDLINKA IEIAKAMGIE SNIHVGVSYS
PNPH_YEAST  HPLRGNLDE DGPRFLALSD AYDLELRKLL FKKWKELKIQ RPLHBTYTF

201                                        250
PNPH_HUMAN  VAGPSFETVA ECRVLQKLG A DAVGMSTVPE VIVARHCGLR VFGFSLITNK
PNPH_BOVIN  LGGPNFETVA ECRLLRNLG A DAVGMSTVPE VIVARHCGLR VFGFSLITNK
PNPH_MOUSE  LAGPNFETVA ESRLKMLG A DAVGMSTVPE VIVARHCGLR VFGFSLITNK
Q9W004      LGGPNYETIA ELKALRMMG V DAVGMSTVHE VITARHCDMK VFAFSLITNK
PNPH_YEAST  VSGPTFETRA ESKMIRMLG G DAVGMSTVPE VIVARHCGR VLALSLITNT

251                                        300
PNPH_HUMAN  VIMDY..... .ES...L..E KANHEEVLA A GKQAAQKLEQ FVSILMASIP
PNPH_BOVIN  VIMDY..... .ES...Q..G KANHEEVLEA GKQAAQKLEQ FVSLLMASIP
PNPH_MOUSE  VVM DY..... .EN...L..E KANHMEVLDA GKAAAQTLER FVSILMESIP
Q9W004      CATEY..... .SD...KKDD EANHDEVMAV AKNRQACCE LVSRLIREIH
PNPH_YEAST  CVVDS PASAL DESPVLE RG KATHAEVLEN GKIASNDVQN LIAAVMGEEL~

301
PNPH_HUMAN  LPDKAS
PNPH_BOVIN  V~~~~~
PNPH_MOUSE  LPDRGS
Q9W004      L~~~~~
PNPH_YEAST  ~~~~~~

```




Conserved Domain Database

PubMed Nucleotide Protein Structure **CDD** Taxonomy Help?

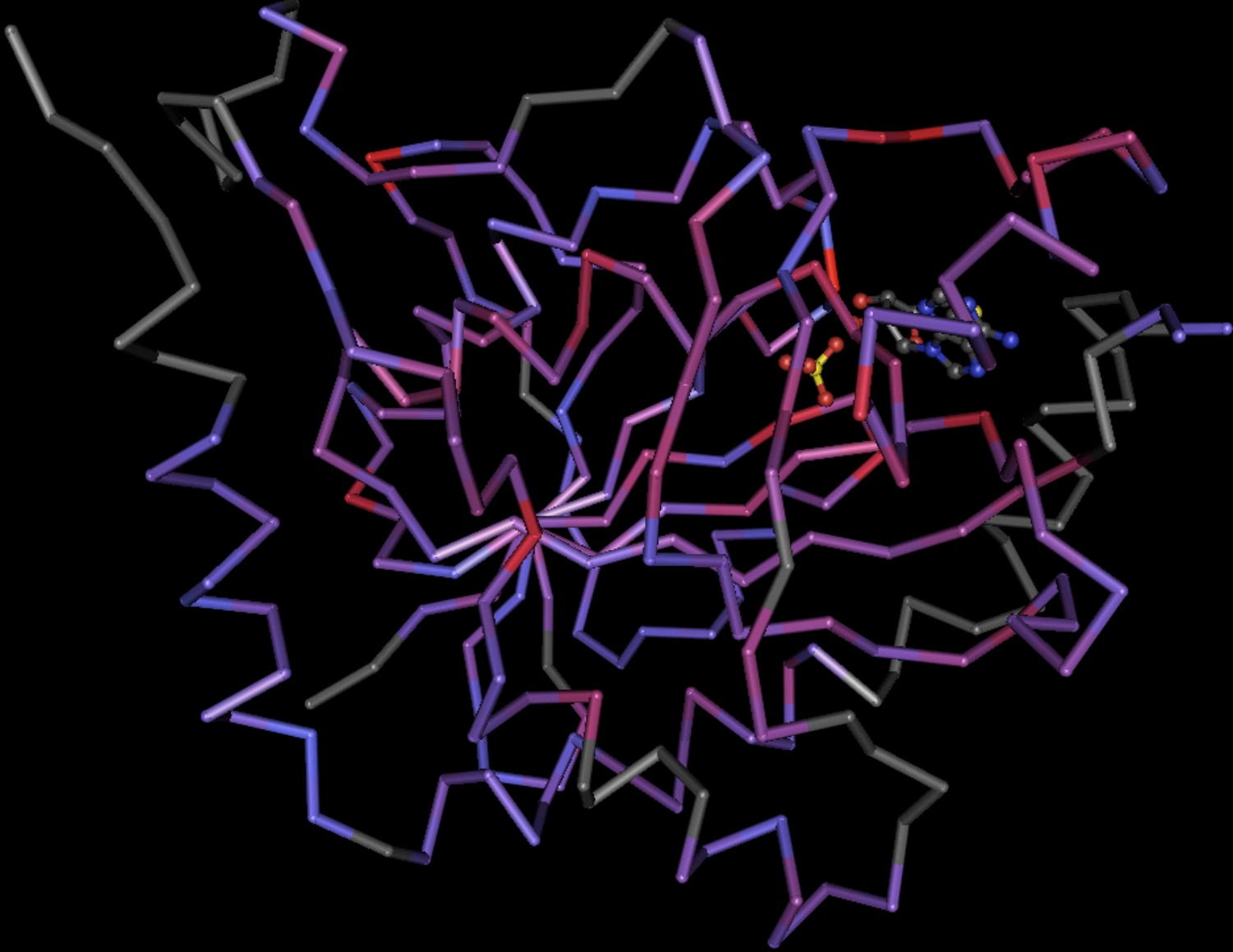
CD: [pfam00896.6](#), [Mtap_PNP](#), Query added PSSM-Id: 4371 Source: [Pfam\[US\]](#), [Pfam\[UK\]](#)

Description: Phosphorylase family 2.
 Taxa: [cellular organisms](#) Related: [COG0005](#), [COG0813](#)
 Status: Alignment from source Created: 24-Sep-2002
 Aligned: 14 rows PSSM: 270 columns Representative: Consensus
 Proteins: [Click here for CDART summary of Proteins containing pfam00896](#)

View 3D Structure with using (To display structure, download [Cn3D](#))
 View Alignment as width color at
 Subset Rows

		10	20	30	40	50	60																											
	*.....*.....*.....*.....*.....*.....																																
consensus	1	PPMIGIIGGSGGLGDLADLLE	REV----	V-PYS	DT	TPY	GK	PSTVIGHAGTLVGG--VGVAV 52																										
query	24	RPQVAIICGSGGLGGLTDKLT	QAQ----	I	F	DY	SEI	PNFPRSTVPGHAGRLVFGfINGRAC 78																										
1CG6_A	9	AVKIGIIGGTGLDDPEILEGR	TE----	K--	Y	VD	TP	FGKPSDALILG-KIKN--VDCVL 57																										
1B80_A	24	RPQVAVICGSGGLGLVNKLT	QAQt----	Fd-	Y	SEI	PNF	PESVPGHAgRLVFGiINGRAC 78																										
1I80_A	27	EHDVAVVLGSGWLPVAVALGS	PTt---	vLp-	Q	AEL	PGF	VPPTAAGHAgELLSvpiGAHRV 82																										
qi_2113997	5	GRMLGVIGGSGFYTFFGSDTR	TV-----		N	S	D	TPY	QPSAPITIG--TIGv--HDVAF 52																									
qi_2622720	1	--MIGIIGGTGIYEMAEGRL	LER----	Rg-	S	L	I	T	PY	GKTPEISVFK--LHG---RRVAF 47																								
qi_2506345	25	TPRVAFILGSGLGALADQIEN	AVa----	Is-	Y	E	K	L	P	GF	PVSTVHGHA	gELVLghlQGV	PV 79																					
qi_2494051	37	PPRTLIIICGSGLGGISTKLS	RDNpppv	tVp-	Y	Q	D	I	P	G	F	K	S	T	V	P	G	H	S	g	T	L	M	F	g	s	m	N	G	S	P	V 95		
qi_1169270	19	SPKIGLILGSGLGILADEIEN	PV	k----	Lk-	Y	E	D	I	P	E	F	F	V	S	T	V	E	G	H	A	g	Q	L	V	L	g	t	L	E	G	V	S	V 73

		70	80	90	100	110	120	
	*.....*.....*.....*.....*.....*.....						
consensus	53	LARHGRGHDYEPHKVNYRAN-	IRALKALG-	VERLIVT	NAAGSLR	EELEPGDLV	VPDDHIN 110	
query	79	VMMQGRFHYEGYPLWKVTFp	VRVFHLLG-	VDTLVVT	NAAGGLN	PKFEVGDIML	IRDHIN 137	
1CG6_A	58	LARHGRQHTIMP	SKVNYQAN-	IWALKEEG-	CTHVIVTT	ACGSLREI	IQPGDIVIIDQFID 115	
1B80_A	79	VMMQGRFHYEGYPPFWKVT	FpVRVFRLLG-	VELVVT	NAAGGLN	PNFEVGDIML	IRDHIN 137	
1I80_A	83	LVLAGRIHAYEGHDLRYV	VVhPVRAARAAG-	AQIMVLT	NAAGGLR	ADLQVGQPV	LISDHLN 141	
qi_2113997	53	LPRHGAAHQYSAHAVPYRAN-	MWALRALG-	VRVFG	PCAVGSLD	PELEPGAVV	VPDQLVD 110	
qi_2622720	48	IPRHS	PGHDKPPH	MVNYRAN-	IWALKELG-	VRQIIAT	NAVGLKRSIGPGDFVVP	PHDFLD 105
qi_2506345	80	VCMKGRGHFYBGRGMTIMT	daIRTFKLLG-	CELLFCT	NAAGSLR	PEVGAGSLV	ALKDHIN 138	
qi_2494051	96	VLNNGRLHG	YEGNTL	FETTFp	IRVLNHMG	hVRNLIVT	NAAGGINAKYQACDLMCIYDHLN 155	
qi_1169270	74	IAMQGRFHYEGYSMEKVT	FpVRVMKALG-	VEALIVT	NAAGGVN	TEFRAGDLMI	ITDHIN 132	





Only One Final Word of Wisdom...

- ✓ “...although the computer is a wonderful helpmate for the sequence searcher and comparer, biochemists and molecular biologists must guard against the blind acceptance of any algorithmic output; given the choice, think like a biologist and not a statistician.”

– - Russell F. Doolittle, 1990



Farewell!