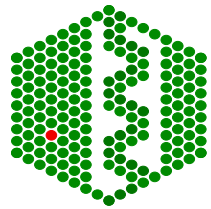
PROTEIN PATTERN DATABASES



PROTEIN SEQUENCES

SUPERFAMILY

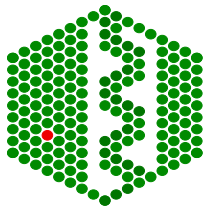
FAMILY

DOMAIN

MOTIF

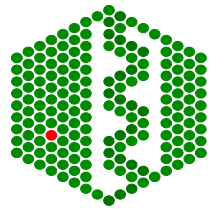
SITE

RESIDUE



BASIC INFORMATION COMES FROM SEQUENCE

- Multiple alignments of related sequences- can build up consensus sequences of known families, domains, motifs or sites.
- Pattern
- Matrix
- Profile
- HMM

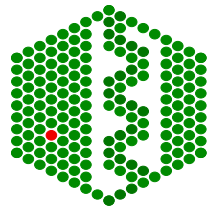


COMMON PROTEIN PATTERN DATABASES

- Prosite patterns
- Prosite profiles
- Pfam
- SMART
- Prints
- TIGRFAMs
- BLOCKS

Alignment databases

- ProDom
- PIR-ALN
- ProtoMap
- Domo
- ProClass



PROSITE Patterns and profiles

- http://www.expasy.ch/prosite/
- Building a pattern:

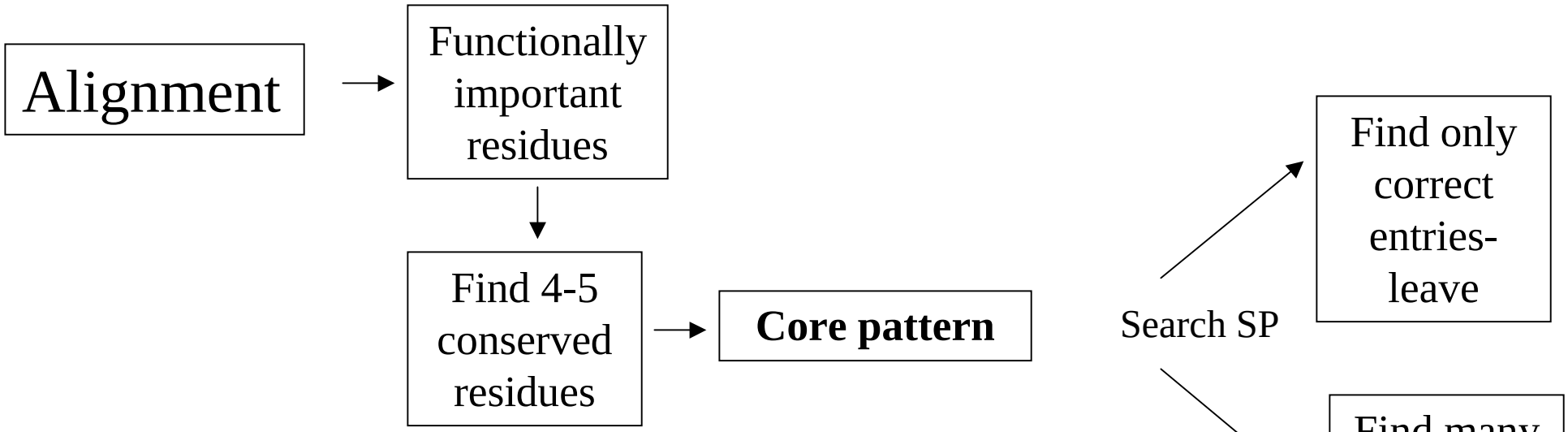
a.) from literature -test against SP, update if necessary

b.) new patterns:

Start with reviewed protein family, known functional sites:

- enzyme catalytic site,
- attachment site eg heme,
- metal ion binding site
- cysteines for disulphide bonds,
- molecule (GTP) or protein binding site

PROSITE PATTERNS



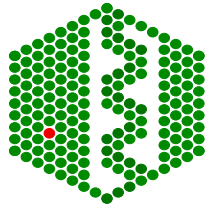
Pattern is given as regular expression:

[AC]-x-V-x(4)-{ED}

ala/cys-any-val-any-any-any-any-(any except glu or asp)

PROSITE PROFILES

- Not confined to small regions, cover whole protein or domain and has more info on allowed aa at each position
- Start with multiple seq alignment -uses a symbol comparison table to convert residue frequency distributions into weights
- Result- table of position-specific amino acid weights and gap costs- calculate a similarity score for any alignment between a profile and a sequence, or parts of a profile and a sequence
- Tested on SP, refined. Begin as prefiles then integrated



PROSITE
Database of protein families and domains

PROSITE is a database of protein families and domains. It consists of biologically significant sites, patterns and profiles that help to reliably identify to which known protein family (if any) a new sequence belongs [More details / References / Disclaimer].

Release 16.37, of 05-May-2001 (contains 1089 documentation entries that describe 1474 different patterns, rules and profiles/matrices).

Access to PROSITE

- by description
- by entry name or accession number (PSxxxxx or PDOCxxxxx number)
- by author
- by citation
- by full text search

- SRS - Sequence Retrieval System

Documents

- PROSITE user manual
- List of PROSITE documentation entries
- How to obtain PROSITE
- Document describing the syntax of profiles in PROSITE
- List of programs that make use of PROSITE
- List of abbreviations for journals cited
- List of on-line experts
- The optimal way to develop patterns

Tools for PROSITE

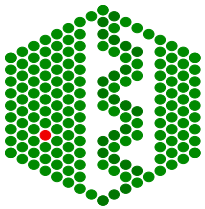
- ScanProsite - Scan a sequence against PROSITE or a pattern against SWISS-PROT
- ProfileScan - Scan a sequence against the profile entries in PROSITE
- Other pattern and profile search tools

Services

- Downloading PROSITE by FTP

Access to servers offering related services

- InterPro - Integrated Resource of ProteinDomains and Functional Sites
- BLOCKS from the Henikoff group at the FHCRC in Seattle (USA)
- DOMO from Jérôme Gracy at Infobiogen (France)
- **Pfam** from the Sanger Centre in Hinxton (UK) or from Washington University (USA)
- PRINTS from Terri Attwood at University of Manchester (UK)
- ProDom from Daniel Kahn at the INRA in Toulouse (France)

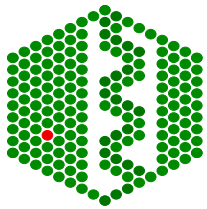


Search in PROSITE for: kinase

(Release 16.37, of 05-May-2001)

Please choose one of the following entries:

- PDOC00004 cAMP- and cGMP-dependent protein kinase phosphorylation site
- PDOC00005 Protein kinase C phosphorylation site
- PDOC00006 Casein kinase II phosphorylation site
- PDOC00007 Tyrosine kinase phosphorylation site
- PDOC00370 Hexokinases signature
- PDOC00099 Galactokinase signature
- PDOC00545 GHMP kinases putative ATP-binding domain
- PDOC00336 Phosphofructokinase signature
- PDOC00504 pfkB family of carbohydrate kinases signatures
- PDOC00490 Phosphoribulokinase signature
- PDOC00524 Thymidine kinase cellular-type signature
- PDOC00408 FGGY family of carbohydrate kinases signatures
- PDOC00100 Protein kinases signatures and profile
- PDOC01049 MAP kinase signature
- PDOC00845 Casein kinase II regulatory subunit signature
- PDOC00101 Pyruvate kinase active site signature
- PDOC00868 Shikimate kinase signature
- PDOC00820 Prokaryotic diacylglycerol kinase signature
- PDOC00710 Phosphatidylinositol 3- and 4-kinases signatures and profile
- PDOC00826 Acetate and butyrate kinases family signatures
- PDOC00102 Phosphoglycerate kinase signature
- PDOC00289 Aspartokinase signature
- PDOC00701 Glutamate 5-kinase signature
- PDOC00104 Adenylate kinase signature
- PDOC00409 Nucleoside diphosphate kinases active site
- PDOC00670 Guanylate kinase signature and profile
- PDOC01034 Thymidylate kinase signature
- PDOC00631 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase signature
- PDOC00421 Phosphoenolpyruvate carboxykinase (GTP) signature
- PDOC00460 Phosphoenolpyruvate carboxykinase (ATP) signature
- PDOC00212 Receptor tyrosine kinase class II signature
- PDOC00213 Receptor tyrosine kinase class III signature
- PDOC00629 Receptor tyrosine kinase class V signatures
- PDOC00728 Cyclin-dependent kinases regulatory subunits signatures
- PDOC00979 Glucokinase regulatory protein family signature



NiceSite view of PROSITE: PDOC00101 (documentation)

Pyruvate kinase active site signature

PROSITE cross-reference(s)

PS00110; PYRUVATE_KINASE

Documentation

Pyruvate kinase (EC 2.7.1.40) (PK) [1] catalyzes the final step in glycolysis, the conversion of phosphoenolpyruvate to pyruvate with the concomitant phosphorylation of ADP to ATP. PK requires both magnesium and potassium ions for its activity. PK is found in all living organisms. In vertebrates there are four, tissues specific, isozymes: L (liver), R (red cells), M1 (muscle, heart, and brain), and M2 (early fetal tissues). In Escherichia coli there are two isozymes: PK-I (gene pykF) and PK-II (gene pykA). All PK isozymes seem to be tetramers of identical subunits of about 500 amino acid residues.

As a signature pattern for PK we selected a conserved region that includes a lysine residue which seems to be the acid/base catalyst responsible for the interconversion of pyruvate and enolpyruvate, and a glutamic acid residue implicated in the binding of the magnesium ion.

Description of pattern(s) and/or profile(s)

Consensus pattern	[LIVAC]-x-[LIVM](2)-[SAPCV]-K-[LIV]-E-[NKRST]-x-[DEQHS]- [GSTA]-[LIVM] [K is the active site residue] [E is a magnesium ligand]
Sequences known to belong to this class detected by the pattern	ALL.
Other sequence(s) detected in SWISS-PROT	1.

Last update

July 1999 / Pattern and text revised.

References

[1]
Muirhead H.
Biochem. Soc. Trans. 18:193-196(1990).

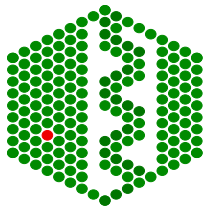
Copyright

View entry in original PROSITE document format

PROSITE: PDOC00101 (documentation)

View entry in NiceSite format

```
{PDOC00101}
{PS00110; PYRUVATE_KINASE}
{BEGIN}
****************************************
* Pyruvate kinase active site signature *
****************************************

Pyruvate kinase (EC 2.7.1.40) (PK) [1] catalyzes the final step in glycolysis,
the conversion of phosphoenolpyruvate to pyruvate with the concomitant
phosphorylation of ADP to ATP. PK requires both magnesium and potassium ions
for its activity. PK is found in all living organisms. In vertebrates there
are four, tissues specific, isozymes: L (liver), R (red cells), M1 (muscle,
heart, and brain), and M2 (early fetal tissues). In Escherichia coli there are
two isozymes: PK-I (gene pykF) and PK-II (gene pykA). All PK isozymes seem to
be tetramers of identical subunits of about 500 amino acid residues.

As a signature pattern for PK we selected a conserved region that includes a
lysine residue which seems to be the acid/base catalyst responsible for the
interconversion of pyruvate and enolpyruvate, and a glutamic acid residue
implicated in the binding of the magnesium ion.

-Consensus pattern: [LIVAC]-x-[LIVM](2)-[SAPCV]-K-[LIV]-E-[NKRST]-x-[DEQHS]-
                    [GSTA]-[LIVM]
                    [K is the active site residue]
                    [E is a magnesium ligand]
-Sequences known to belong to this class detected by the pattern: ALL.
-Other sequence(s) detected in SWISS-PROT: 1.
-Last update: July 1999 / Pattern and text revised.

[ 1] Muirhead H.
     Biochem. Soc. Trans. 18:193-196(1990).

+--------------------------------------------------------------------------+
| This PROSITE entry is copyright by the Swiss Institute of Bioinformatics |
| (SIB). There are no restrictions on its use by non-profit institutions as |
| long as its content is in no way modified and this statement is not      |
| removed. Usage by and for commercial entities requires a license agreement |
| (See http://www.isb-sib.ch/announce/ or email to license@isb-sib.ch).    |
+--------------------------------------------------------------------------+
{END}
```

NiceSite View of PROSITE: PS00110

General information about the entry	
Entry name	**PYRUVATE_KINASE**
Accession number	**PS00110**
Entry type	PATTERN
Date	APR-1990 (CREATED); JUL-1999 (DATA UPDATE); JUL-1999 (INFO UPDATE).
PROSITE documentation	PDOC00101
Name and characterization of the entry	
Description	Pyruvate kinase active site signature.
Pattern	[LIVAC]-x-[LIVM](2)-[SAPCV]-K-[LIV]-E-[NKRST]-x-[DEQHS]-[GSTA]-[LIVM].

Numerical results

- SWISS-PROT release number: **38**, total number of sequence entries in that release: **80000**.
- Total number of hits in SWISS-PROT: **57 hits in 57 different sequences**
- Number of hits on proteins that are known to belong to the set under consideration: **56 hits in 56 different sequences**
- Number of hits on proteins that could potentially belong to the set under consideration: **0 hits in 0 different sequences**
- Number of false hits (on unrelated proteins): **1 hits in 1 different sequences**
- Number of known missed hits: **2**
- Number of partial sequences which belong to the set under consideration, but which are not hit by the pattern or profile because they are partial (fragment) sequences: **9**
- Precision (true hits / (true hits + false positives)): **98.25 %**
- Recall (true hits / (true hits + false negatives)): **96.55 %**

Comments

- Taxonomic range: **Archaebacteria, Eukaryotes, Prokaryotes (Bacteria)**
- Maximum known number of repetitions of the pattern in a single protein: **1**
- 'Interesting' site in the pattern: **5,active_site**
- 'Interesting' site in the pattern: **7,magnesium**

Cross-references

SWISS-PROT

True positive hits:

KPY1_ECOLI (P14178), KPY1_FELCA (P11979), KPY1_HUMAN (P11974), KPY1_RABIT (P11974), KPY1_RAT (P11980), KPY1_SALTY (P11980), KPY1_SYNY3 (Q55863), KPY1_TRYBB (P30615), KPY1_YEAST (P30615), KPY2_ECOLI (P21599), KPY2_HUMAN (P14786), KPY2_MOUSE (P14786), KPY2_RABIT (O18919), KPY2_RAT (P11981), KPY2_SYNY3 (O18919), KPY2_TRYBB (P30616), KPY2_YEAST (P52489), KPY3_AGRVI (P52489), KPY4_AGRVI (Q44473), KPYA_RICCO (Q43117), KPYA_TOBAC (Q43117), KPYC_ARATH (O65595), KPYC_SOLTU (P22200), KPYC_SOYBN (P22200), KPYC_TOBAC (Q42954), KPYG_RICCO (P55964), KPYG_TOBAC (Q42954), KPYK_ASPNG (Q12669), KPYK_BACLI (P51181), KPYK_BACPY (P51181), KPYK_BACST (Q02499), KPYK_BACSU (P80885), KPYK_BORBU (O51323), KPYK_CHICK (P00548), KPYK_CORGL (Q46078), KPYK_DROME (O62619), KPYK_EIMTE (O44006), KPYK_EMENI (P22360), KPYK_HAEIN (P43924), KPYK_LACLA (Q07637), KPYK_LEIME (Q27686), KPYK_METEX (O05118), KPYK_METJA (Q57572), KPYK_MYCGE (P47458), KPYK_MYCIT (P94939), KPYK_MYCPN (P78031), KPYK_MYCTU (O06134), KPYK_SCHPO (Q10208), KPYK_TRIRE (P31865), KPYK_TRYBO (Q27788), KPYK_XENLA (Q92122), KPYK_YARLI (P30614), KPYK_CANFA (Q29536), KPYR_HUMAN (P30613), KPYR_MOUSE (P53657), KPYR_RAT (P12928)

False negative hits (sequences which belong to the set under consideration, but which have not been picked up by the pattern or profile):

KPYK_CHLTR (P94685), KPYK_LACDE (P34038)

'Potential' hits (sequences that belong to the set under consideration, but which were not picked up because the region(s) that are used as a 'fingerprint' (pattern or profile) is not yet available in the data bank (partial sequence)):

KPY1_PHOLE (O30853), KPY1_SPICI (P19680), KPYK_CANAL (P46614), KPYK_CLOAB (O08309), KPYK_CLOPA (P81344), KPYK_CLOPE (Q46289), KPYK_LEIBR (Q04668), KPYK_THEAC (P32044), KPYK_THELI (Q56301)

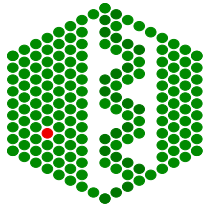
False positive hits (sequences which do not belong to the set under consideration):

DNAB_HAEIN (P45256)

1PKY; 1PYK; 1PKM; 1AQF; 1PKN; 1A5U; 1A3W; 1A3X; 1PKL;

Pfam

- http://www.sanger.ac.uk/Software/Pfam/index.shtml
- Database of HMMs for domains and families
- HMMs are built from HMMER2 (Bayesian statistical models), can use two modes ls or fs, all domains should be matched with ls
- Use Bits scores, thresholds are chosen manually using E-value from extreme fit distribution
- Two parts to Pfam:
  - ⇨ PfamA -manually curated
  - ⇨ PfamB -automatic clustering of rest of SPTR from ProDom using Domainer
- Use -looking at domain structure of SPTR protein or new sequence



Pfam

Protein families database of alignments and HMMs

Version 6.2, April 2001, **2773** families

Pfam is a large collection of multiple sequence alignments and hidden Markov models covering many common protein domains. For more information on Pfam, on using this site, or on the changes between Pfam releases 5 and 6, click here.

Interactive WWW access to Pfam

- KEYWORD SEARCH - Query Pfam by keywords.
- PROTEIN SEARCH - Find Pfam domain matches in your protein sequence.
- DNA SEARCH - Find Pfam domain matches in your DNA sequence.
- BROWSE PFAM - View Pfam annotation and alignments.
- TAXONOMY SEARCH - Find Pfam domain matches by organism.
- **More information and help on Pfam**

Pfam mirror servers worldwide

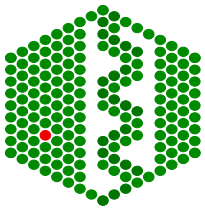
- **Sanger Centre (UK)**
- **St. Louis (USA)**
- **Karolinska Institutet (Sweden)**
- **Institut National de la Recherche Agronomique (France)**

Query Pfam text

Results for query 'kinase'

Matches to documentation in the selected databases with links back to Pfam

Family	Description
6PF2K	6-phosphofructo-2-kinase
aakinase	Amino acid kinase family
Acetate_kinase	Acetokinase family
adenylatekinase	Adenylate kinase
Anti_proliferat	BTG1 family
APH	Aminoglycoside phosphotransferase
APS_kinase	Adenylylsulfate kinase
arf	ADP-ribosylation factor family
ATP-gua_Ptrans	ATP:guanido phosphotransferase, C-terminal catalytic domain
ATP-gua_PtransN	ATP:guanido phosphotransferase, N-terminal domain
C2	C2 domain
cadherin	Cadherin domain
CDI	Cyclin-dependent kinase inhibitor
CheW	CheW-like domain
Choline_kinase	Choline/ethanolamine kinase
CK_II_beta	Casein kinase II regulatory subunit
CKS	Cyclin-dependent kinase regulatory subunit
CNH	CNH domain
CobU	Cobinamide kinase / cobinamide phosphate guanyltransferase
cyclin	Cyclin
Cytidylate_kin	Cytidylate kinase
DAG_PE-bind	Phorbol esters/diacylglycerol binding domain (C1 domain)
DAGK_prokar	Prokaryotic diacylglycerol kinase
DAGKa	Diacylglycerol kinase accessory domain (presumed)
DAGKc	Diacylglycerol kinase catalytic domain (presumed)
Dak1	Dak1 domain
death	Death domain
dNK	Deoxynucleoside kinase
DSPc	Dual specificity phosphatase, catalytic domain



Results for UserSeq

Trusted matches - domains scoring higher than the gathering threshold

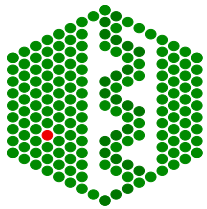
Domain	Start	End	Bits	Evalue	Alignment
efhand	114	142	18.10	0.21	Align
efhand	159	187	22.40	0.011	Align
DAG_PE-bind	206	253	73.30	5.2e-18	Align
DAG_PE-bind	270	319	47.10	3.8e-10	Align
DAGKc	376	500	236.20	4.7e-67	Align
DAGKa	520	701	406.60	2.3e-118	Align

Matches to Pfam-B

Domain	Start	End	Evalue	Alignment
Pfam-B_4919	1	108	1.4e-54	Align
Pfam-B_6031	270	334	9e-06	Align
Pfam-B_5578	702	734	4e-14	Align

[735 residues]

efhand 114-142
efhand 159-187
DAG_PE-bind 206-253
DAG_PE-bind 270-319
DAGKc 376-500
DAGKa 520-701



Alignments of Pfam-A domains to HMMs

Format for fetching alignments to seed: Hypertext linked to swisspfam

Alignment of efhand vs UserSeq/114-142

```
                  *->elkeaFkefDkDgDGkIsfeEfkaalkkl<-*
                     +l+  Fk++D+D++G ++  E  +++ ++
     UserSeq   114    KLEFTFKLYDTDRNGILDSSEVDKIILQM    142
```

Align to seed

Alignment of efhand vs UserSeq/159-187

```
                  *->elkeaFkefDkDgDGkIsfeEfkaalkkl<-*
                      l+e++ke+D Dg+G +s  E+++a +++
     UserSeq   159    ILQEMMKEIDYDGSGSVSQAEWVRAGATT    187
```

Align to seed

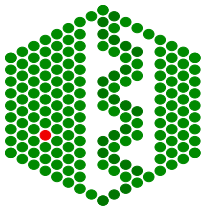
Alignment of DAG_PE-bind vs UserSeq/206-253

```
                  *->HrFkrttfyksptfCdhCgellwglakQGlkCsnCglnvHkrChekV
                     H++++++f ++p +C++C++ ++ + kQGl C  C+++vH++C++k+
     UserSeq   206    HMWRPKRF-PRPVYCNLCESSIGLG-KQGLSCNLCKYTVHDQCAMKA 250

                  ptnC<-*
                    +C
     UserSeq   251 L-PC    253
```

Align to seed

Alignment of DAG_PE-bind vs UserSeq/270-319

```
                  *->HrFkrttfyksptfCdhCgellwgla.kQGlkCsnCglnvHkrChek
                     H ++r  +    ++Cd C++ ++ +++ +Gl+C +C+l +H+ C+++
     UserSeq   270    HVWVRGGC--ESGRCDRCQKKIRIYHsLTGLHCVWCHLEIHDDCLQA 314

                  VptnC<-*
                  V  +C
     UserSeq   315 VGHEC    319
```

DAG_PE-bind

Figure 1: 1fbn
calcium-binding protein
nmr structure of a protein kinase c-g phorbol-binding domain, minimized average structure

Accession number: **PF00130**

Phorbol esters/diacylglycerol binding domain (C1 domain)

-!- This domain is also known as the Protein kinase C conserved region 1 (C1) domain.

INTERPRO description (entry IPR002219)

Diacylglycerol (DAG) is an important second messenger. Phorbol esters (PE) are analogues of DAG and potent tumor promoters that cause a variety of physiological changes when administered to both cells and tissues. DAG activates a family of serine/threonine protein kinases, collectively known as protein kinase C (PKC) . Phorbol esters can directly stimulate PKC. The N-terminal region of PKC, known as C1, has been shown to bind PE and DAG in a phospholipid and zinc-dependent fashion. The C1 region contains one or two copies (depending on the isozyme of PKC) of a cysteine-rich domain about 50 amino-acid residues long and essential for DAG/PE-binding. The DAG/PE-binding domain binds two zinc ions; the ligands of these metal ions are probably the six cysteines and two histidines that are conserved in this domain.

For additional annotation, see the PROSITE document PDOC00379 [Expasy|SRS-UK|SRS-USA]

To contribute to the annotation for this family (and win a T-shirt), click here

Alignment

Seed (40) Full (343)

Format Hypertext in Pfam format

Get alignment

Further alignment options here

Help relating to Pfam alignments here

Domain organisation

Seed (40) Full (343)

As a Graphic **As a Tree**

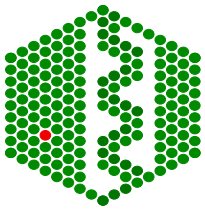
Zoom 0.5 pixels/aa. Bootstrap tree

View Graphic NIFAS Applet

To find out about the NIFAS tree-viewer, click here

Species Distribution

Tree depth Show all levels View Species Tree



Literature References

[1]
Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase.
Knighton DR, Zheng JH, Ten Eyck LF, Ashford VA, Xuong NH, Taylor SS, Sowadski JM;
Science. 1991;253:407-414.

Database References

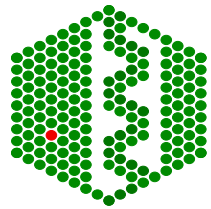
PDB You can find out how to set up Rasmol here	1tbn ; 102; 151; PDB 2 Pfam Rasmol (unix) Chime (pc) CATH-PDBSUM SCOP-UK SCOP-USA
PROSITE	PDOC00379 [Expasy|SRS-UK|SRS-USA]
PRINTS	PR00008
INTERPRO	IPR002219
PFAMB The following Pfam-B families contain sequences that according to Prodom are members of this Pfam-A family.	PB000018 PB006031 PB014315 PB044288 PB058467

Pfam specific information

Author of entry	Bateman A
Alignment method of seed	Manual
Source of seed members	Prosite
Gathering cutoff	10 3
Trusted cutoff	10.40 4.40
Noise cutoff	9.40 16.00
Build method of HMM	hmmbuild HMM SEED hmmcalibrate --seed 0 HMM

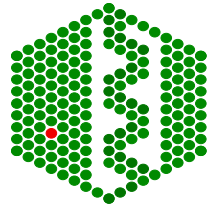
Additional features of Pfam

- PfamA has about 65% coverage of SPTR, rest is covered by PfamB
- Can search directly with DNA -Wise2 package
- Can view taxonomic range of each entry
- Can view proteins with similar domain structure and view of all family members
- Links to other databases including 3D structure
- Note: No 2 PfamA HMMs should overlap



SMART- Simple Modular Architecture Research Tool

- http://smart.embl-heidelberg.de/
- Relies on hand curated multiple sequence alignments of representative family members from PSI-BLAST- builds HMMs- used to search database for more seq for alignment- iterative searching until no more homologues detected
- Store Ep (highest per protein E-value of T) and En (lowest per protein E-value of N) values
- Will predict domain homologue with sequence if
  - Ep < E-value <En and E-value <1.0



SMART

Number of SMART HMMs
530

S imple
M odular
A rchitecture
R esearch
T ool

References
Schultz et al. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 5857-5864
Schultz et al. (2000) *Nucleic Acids Res* **28**, 231-234

SMART STATUS: RUNNING

HOME | FAQ | LITERATURE | ABOUT | GLOSSARY | WHAT'S NEW | FEEDBACK

SMART Software version 3.3 released. Check "What's new" page for details and send us your comments and suggestions!

Sequence analysis

You may use either the swissprot/sptrembl sequence identifier (ID) / accession number (ACC) or the protein sequence itself to request the smart service

Sequence ID or ACC

Sequence

Sequence SMART | Reset

HMMER searches of the SMART database occur by default. You may also find:

- Outlier homologues and homologues of known structure
- PFAM domains
- signal peptides and GPI anchors
- internal repeats

Architecture analysis

You can search for proteins with combinations of specific domains in different species or taxonomic ranges.

Domain selection

Example: TyrKc AND SH3 AND NOT SH2

Taxonomic selection

Select a taxonomic range via the selection box or type it into the text box below

Select:

All

Examples Dictyostelium discoideum
Porifera

Architecture SMART | Reset

Alert

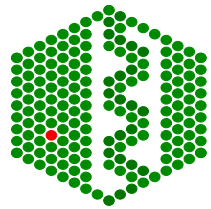
If you want to be automatically informed each time a new protein with a defined domain composition is deposited in databases, please use 'alert SMART' (this facility is also available following an architecture analysis query)

Domains detected by SMART

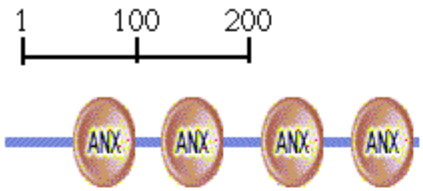
You can search for keywords in the domain annotation

Search Keywords

- Browse the database of all available domains in the SMART database
- See a list of recent domain changes



Domains within the query sequence of 345 residues



Mouse over domain / undefined region to see the limits; click on it to go to further annotation; right-click to save whole protein as PNG image

Transmembrane segments as predicted by the *TMHMM2* program (▬), coiled coil regions determined by the *Coils2* program (▬) and Segments of low compositional complexity, determined by the *SEG* program (▬)

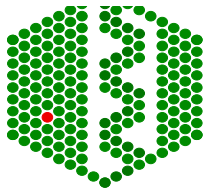
Architecture analysis

Display all proteins with similar domain organisation.
Display all proteins with similar domain composition.

The SMART diagram above represents a summary of the results shown below. Domains with scores less significant than established cutoffs are not shown in the diagram. Features are also not shown when two or more occupy the same piece of sequence; the priority for display is given by SMART > PFAM > PROSPERO repeats > Signal peptide > Transmembrane > Coiled coil > Low complexity. In either case, features not shown in the above diagram are marked 'hidden'

Confidently predicted domains, repeats, motifs and features:

name	begin	end	E-value
ANX	58	110	8.44e-22
ANX	130	182	8.70e-24
ANX	214	266	7.08e-15
ANX	289	341	2.46e-23

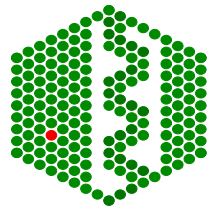


The following proteins have the same domain organisation as your query protein.

You can display the domain architecture of ALL (136) or selected (below) proteins.

To get an overview of the species distribution of these proteins, see the TAX BREAK

Protein	Description	Species
ANX2_HUMAN	ANNEXIN II (LIPOCORTIN II) (CALPACTIN I HEAVY CHAIN) (CHROMOBINDIN 8)(P36) (PROTEIN I) (PLACENTAL ANTICOAGULANT PROTEIN IV) (PAP-IV).	Homo sapiens
ANX9_HUMAN	ANNEXIN A9 (ANNEXIN 31) (ANNEXIN XXXI).	Homo sapiens
ANX1_RABIT	ANNEXIN I (LIPOCORTIN I) (CALPACTIN II) (CHROMOBINDIN 9) (P35)(PHOSPHOLIPASE A2 INHIBITORY PROTEIN).	Oryctolagus cuniculus
CAB94770	ANNEXIN A11.	Mus musculus
ANXB_XENLA	ANNEXIN II TYPE I (LIPOCORTIN II) (CALPACTIN I HEAVY CHAIN)(CHROMOBINDIN 8) (P36) (PROTEIN I) (PLACENTAL ANTICOAGULANT PROTEINIV) (PAP-IV).	Xenopus laevis
CAA13092	ANNEXIN V.	Mus musculus
Q9ZVJ7	PUTATIVE ANNEXIN.	Arabidopsis thaliana
ANX3_HUMAN	ANNEXIN III (LIPOCORTIN III) (PLACENTAL ANTICOAGULANT PROTEIN III)(PAP-III) (35-ALPHA CALCIMEDIN) (INOSITOL 1,2-CYCLIC PHOSPHATE 2-PHOSPHOHYDROLASE).	Homo sapiens
AAH05595	SIMILAR TO ANNEXIN A6.	Mus musculus
Q43864	ANNEXIN P35.	Zea mays
P93158	ANNEXIN (FRAGMENT).	Gossypium hirsutum
O93447	ANNEXIN MAX4.	Oryzias latipes
Q9NGU7	ANNEXIN (FRAGMENT).	Taenia solium
O65848	ANNEXIN.	Medicago truncatula
CAC34623	ANNEXIN A13 ISOFORM A.	Mus musculus
O70371	LIPOCORTIN V (FRAGMENT).	Rattus norvegicus
ANXB_MOUSE	ANNEXIN A11 (ANNEXIN XI) (CALCYCLIN-ASSOCIATED ANNEXIN 50) (CAP-50).	Mus musculus
ANX2_CHICK	ANNEXIN II (LIPOCORTIN II) (CALPACTIN I HEAVY CHAIN) (CHROMOBINDIN 8)(P36) (PROTEIN I) (PLACENTAL ANTICOAGULANT PROTEIN IV) (PAP-IV).	Gallus gallus
ANX6_RAT	ANNEXIN VI (LIPOCORTIN VI) (P68) (P70) (PROTEIN III) (CHROMOBINDIN 20)(67 KDA CALELECTRIN) (CALPHOBINDIN-II) (CPB-II) (CALCIUM-BINDINGPROTEIN CATA 65/67).	Rattus norvegicus
ANXD_CANFA	ANNEXIN A13 (ANNEXIN XIII) (ANNEXIN, INTESTINE-SPECIFIC) (ISA).	Canis familiaris
BAA07708	ANNEXIN V.	Rattus norvegicus
AAH00871	ANNEXIN A3.	Homo sapiens
AAH01748	ANNEXIN A2.	Homo sapiens



S imple
M odular
A rchitecture
R esearch
T ool

References

Schultz et al. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 5857-5864
Schultz et al. (2000) *Nucleic Acids Res* **28**, 231-234

signalling

14_3_3	4.1m	ADF	ANK	ARF	ArfGap	ARM	B41
B_lectin	BAG	BH4	BIR	BTK	C1	C2	Calx_beta
CARD	CASc	CH	CheW	CNH	cNMP	CULLIN	CYCc
DAGKa	DAGKc	DAX	DEATH	DED	DEP	DSPc	DUF1
DUF2	DYNc	EFh	EH	ENTH	FBOX	FCH	FF
FH2	FHA	FYVE	G_alpha	GAF	GAS2	GED	GGL
GoLoco	GRAM	GS	GuKc	HAMP	HATPase_c	HECTc	HisKA
HPT	HR1	ILWEQ	IPPc	IQ	ITAM	KISc	L27
LIM	LMWPc	MA	MeTrc	MIR	MYSc	MyTH4	OPR
PAC	PAS	PBD	PDZ	PH	PI3K_C2	PI3K_p85B	PI3K_rbd
PI3Ka	PI3Kc	PIPKc	PLAc	PLCXc	PLCYc	PLDc	PP2Ac
PP2Cc	PTB	PTBI	PTPc	PTPc_DSPc	PTPc_motif	PX	PXA
RA	RAB	RAN	RanBD	RAS	RasGAP	RasGEF	RasGEFN
RBD	REC	RGS	RHO	RhoGAP	RhoGEF	RIIa	RING
RIO	S_TK_X	S_TKc	SAM	SAR	SEC14	Sec7	SH2
SH3	small_GTPase	SOCS	SPEC	SPRY	START	STYKc	t_SNARE
TarH	TBC	TIR	TyrKc	UBA	UBCc	UBQ	UBX
VHP	VHS	VPS9	WD40	WH1	WH2	WW	ZnF_A20
ZnF_AN1	ZnF_RBZ	ZnF_UBP	ZnF_ZZ	ZU5			

nuclear

35EXOc	53EXOc	A1pp	ADEAMc	AP2	AP2Ec	AT_hook	AWS
AXH	BAH	BASIC	BBC	BBOX	BCL	BRCT	BRIGHT
BRLZ	BROMO	CBF	CHROMO	CSP	Cu_FIST	DEXDc	DM
DSRM	DWA	DWB	eIF5C	ENDO3c	ETS	EXOIII	FES
FH	G_patch	GIYc	H15	H2A	H2B	H3	H4
HALZ	HELICc	HhH1	HhH2	HLH	HMG	HMG17	HNHc
HNS	HOLI	HOX	HRDC	HSF	HTH_ARAC	HTH_ARSR	HTH_ASNC
HTH_CRP	HTH_DEOR	HTH_DTXR	HTH_GNTR	HTH_ICLR	HTH_LACI	HTH_LUXR	HTH_MARR
HTH_MERR	HTH_XRE	IENR1	IENR2	IPT	IRF	IRO	JmjN
KH	KRAB	Ku78	LEM	LER	LIGANc	MA3	MADS

POU	PreSET	PUA	RIBOc	RRM	S4	SAND	SANT
SAP	SET	SFM	Skp1	SMR	SRA	STE	TBOX
TEA	TFIIE	TFS2M	TFS2N	TOP1Ac	TOP1Bc	TOP2c	TOP4c
TOPEUc	TOPRIM	Ubox	XPGI	XPGN	Zalpha	ZnF_C2C2	ZnF_C2H2
ZnF_C2HC	ZnF_C3H1	ZnF_C4	ZnF_CHCC	ZnF_GATA	ZnF_NFX	ZnF_U1	

extracellular

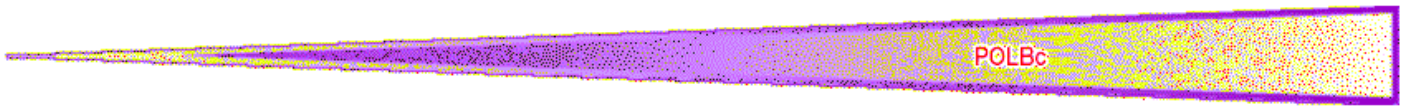
A4_EXTRA	AAI	ACR	ALBUMIN	ANATO	APPLE	B_lectin	BowB
BPI1	BPI2	btg1	C1Q	C4	CA	CALCITONIN	CBD_IV
CCP	ChtBD1	ChtBD2	ChtBD3	CLa	CLb	CLECT	CNX
COLFI	COLIPASE	CRF	CSF2	CT	CUB	CW	CY
CysPc	DEFSN	DISIN	DNaseIc	DSL	EGF	EGF_CA	EGF_Lam
EGF_like	END	EPEND	FA58C	FAS1	FBG	fCBD	FGF
FIMAC	FN1	FN2	FN3	FOLN	FRI	FTP	FU
Galanin	GASTRIN	GEL	GHA	GHB	GLA	GLECT	GLUCA
GPS	GRAN	HintC	HintN	HormR	HX	HYDRO	IB
IFabd	IG	IG_like	IGc1	IGc2	IGv	IL1	IL10
IL2	IL4_13	IL6	IL7	IlGF	INB	Int_alpha	KAZAL
Knot1	KR	KU	LamB	LamG	LamGL	LamNT	LCCL
LDLa	LIF_OSM	LINK	LRR	LRR_BAC	LRR_CC	LRR_PS	LRR_RI
LRR_SD22	LRR_TYP	LRRCT	LRRNT	LU	LY	LysM	LYZ1
LYZ2	MACPF	MAM	MATH	MD	NAT_PEP	NGF	NH
NIDO	NL	NMU	NTR	NUC	OLF	OSTEO	P
PA2c	PAH	PAN_AP	PBPb	PBPe	PDGF	PKD	PLP
PQQ	PRP	PSI	PTH	PTI	PTN	PTX	RICIN
RNAse_Pc	SAA	SAPA	SAPB	SCP	SCY	SEA	SERPIN
SF_P	SH3b	ShKT	SO	SOG	SR	STI	TGFB
THN	TK	TNF	TNFR	TR_FER	TR_THY	Tryp_SPc	TSP1
TSPc	TSPN	TY	UTG	VIT	VPS10	VWA	VWC
VWD	WAP	WIF	WNT1	WR1	WSC	XTALbg	ZnMc
ZP							

Others

AAA	acidPPc	ACTIN	alkPPc	ANX	BHL	BTB	CAD
CBS	ChSh	CLH	CUE	CYCLIN	DCX	DnaJ	EZ_HEAT
FAF	GAL4	GIT	GYF	HAT	HDc	IG_FLMN	JAB_MPN
LH2	LON	LRRcap	MIF4G	NDK	NEBU	PHB	PINT
PLEC	PlsC	PolyA	PP2C_SIG	PROF	Pumilio	PWI	PWWP
R3H	RHOD	S1	SAM_PNT	SNc	Sorb	SWIB	SynN
TAFH	TGc	THY	TPR	TUDOR	WSN	ZnF_TAZ	ZnF_UBR1

See Family Alignment in | CLUSTALW (CHROMA colored) format

See alignment consensus sequence



POLBc

DNA polymerase type-B family
DNA polymerase alpha, delta, epsilon and zeta chain (eukaryota), DNA polymerases in archaea, DNA polymerase II in e. coli, mitochondrial DNA polymerases and and virus DNA polymerases

Occurrence of POLBc domains / proteins with POLBc in nrdb	364 / 358

Evolution

Species distribution (number of POLBc domains / proteins detected in non-redundant sequence database (NRDB))

Eukaryota	Archaea	Bacteria	viruses
127 / 126	38 / 38	6 / 6	179 / 174
↓ ofthese			
Metazoa	**Fungi**	**Viridiplantae (Plants)**	
37 / 36	31 / 31	11 / 11	
↓ ofthese			
Chordata	**Arthropoda**	**Nematoda**	
22 / 22	9 / 9	6 / 5	

Number of POLBc domains / proteins in selected organisms (to see numbers in genomes, click on the species)

Homo sapiens	Mus musculus	Drosophila melanogaster
10 / 10	7 / 7	9 / 9
Caenorhabditis elegans	**Arabidopsis thaliana**	**Saccharomyces cerevisiae**
6 / 5	5 / 5	4 / 4

EMBL
European B

Localisation

Nr. of POLBc domains / proteins in SwissProt proteins of predicted localisation					
intracellular		**extracellular**		**membrane-associated**	
total	51 / 51	total	--	total	--
nuclear	51 / 51				
cytoplasmic	--				
ER-golgi	--	secreted	--	transmembrane	--
chloroplast	--				
mitochondrion	--				

Cellular role							
chromatin	**metab.**	**sign.**	**transp.**	**transl.**	**transcr.**	**repl.**	**interact.**
						X	

Binding / catalysis

Literature

Primary literature (below); Secondary (Automatically-derived) Literature

[1] Crystal structures of an NH2-terminal fragment of T4 DNA polymerase andits complexes with single-stranded DNA and with divalent metal ions.
Wang J, et al.
Biochemistry **35** (1996): 8110-9
medline: 0008679562

Structure

3D Structures of POLBc domains in PDB	1clq 1d5a 1gcx 1noy 1noz 1qe4 1qht 1qqc 1tgo 1waf 1wag 1wah 1wai 1waj

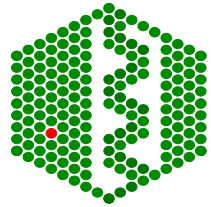
Links

PFAM	DNA_pol_B
PROSITE	DNA_POLYMERASE_B

EMBL
European

Additional features of SMART

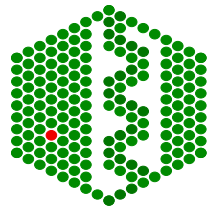
- Used for identification of genetically mobile domains and analysis of domain architectures
- Can search for proteins containing specific combinations of domains in defined taxa
- Can search for proteins with identical domain architecture
- Also has information on intrinsic features like signal sequences, transmembrane helices, coiled-coil regions and compositionally biased regions



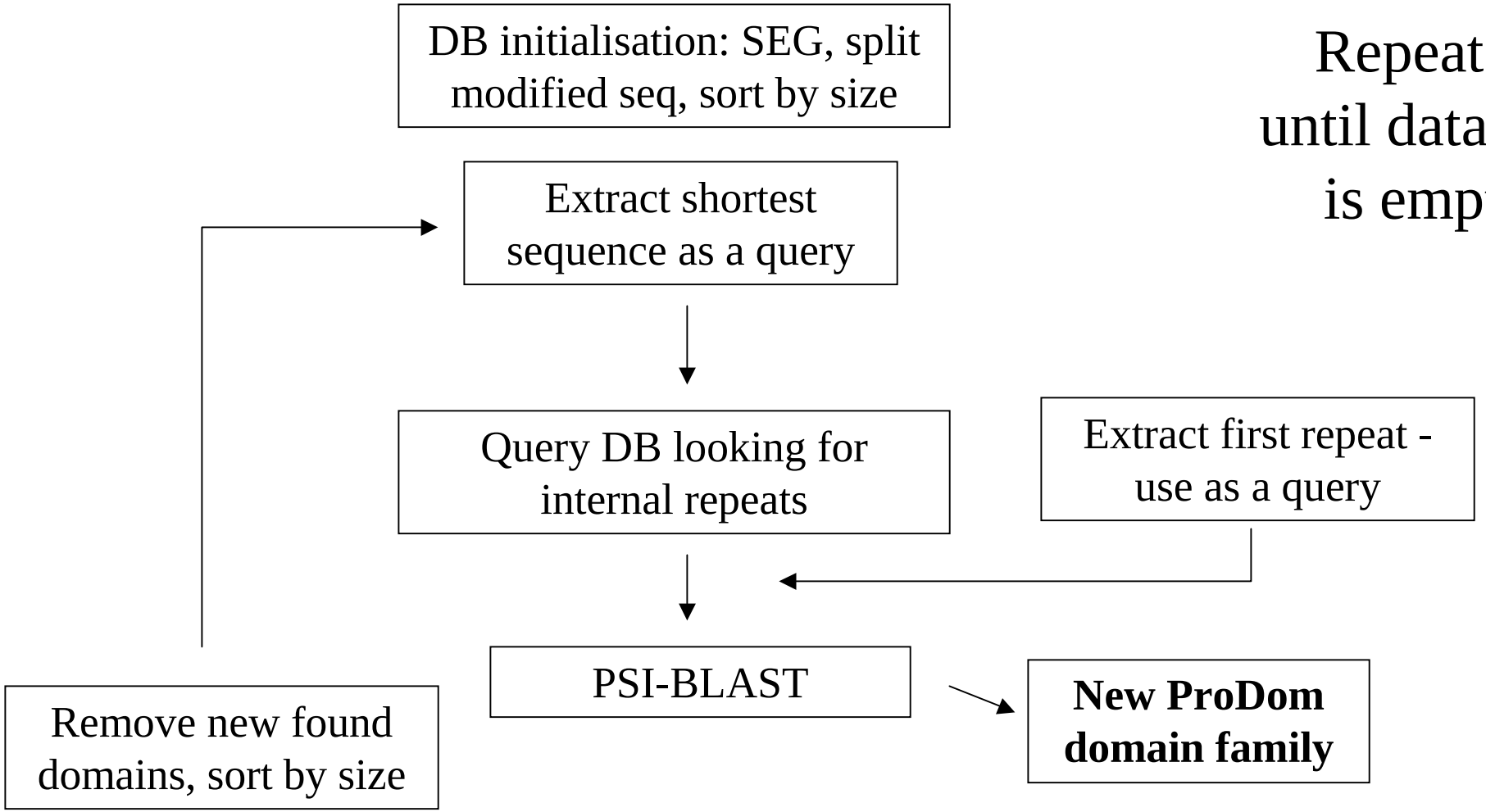
ProDom

The Protein Domain Database
ProDom

- http://www.toulouse.inra.fr/prodom.html
- Groups all sequences in SPTR into domains ->150 000 families
- Use automatic process to build up domains -DOMAINER
- For expert curated families, use PfamA alignments to build new ProDom families
- Use diameter (max distance between two domains in family) and radius of gyration root mean square of distance between domain and family consensus), both counted in PAM (percent accepted mutations (no per 100 aa) to measure consistency of a family, lower these values, more homogeneous family



Building of ProDom families



Repeated until database is empty

The Protein Domain Database

ProDom

WARNING: new procedure for ProDom construction

· **July 1998 (ProDom 35)**

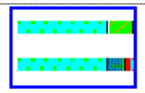
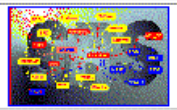
The ProDom protein domain database consists of an automatic compilation of homologous domains. Current versions of ProDom are built using a novel procedure based on recursive PSI-BLAST searches (Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ, 1997, *Nucleic Acids Res.*, **25**:3389-3402; Gouzy J., Corpet F. & Kahn D., 1999, *Computers and Chemistry* **23**:333-340.) Large families are much better processed with this new procedure than with the former DOMAINER program (Sonnhammer, E.L.L. & Kahn, D., 1994, *Protein Sci.*, **3**:482-492).

· **March 2001 (ProDom 2001.1)**

390 ProDom families were generated automatically using PSI-BLAST with a profile built from the seed aligments of Pfam-A 4.3 families.

Last ProDom update:	March 30th, 2001
Current ProDom release:	ProDom 2001.1 / (Statistics)
built from	non fragmentary sequences from SWISS-PROT 39 + TREMBL + TREMBL updates - December 8, 2000

Both the ProDom database and this server have been designed as a tool to help analyze domain arrangements of proteins and protein families *(Corpet F, Servant F, Gouzy J, Kahn D (2000) ProDom and ProDom-CG: tools for protein domain analysis and whole genome comparisons. Nucleic Acids Res. 28:267-269)*. Strong emphasis has been put on the graphical user interface which allows for interactive analysis of protein homology relationships. Here is a brief outline of what the ProDom server can do for you:

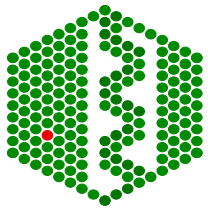
BLAST	Homology search against all domain sequences in ProDom
	ProDom domain arrangements of all proteins sharing homology with a given protein, or ProDom domain arrangements of all proteins containing a given domain
	Retrieval of ProDom multiple alignments and consensus sequences.
	Navigation between ProDom, SWISS-PROT, TrEMBL, PROSITE, PFAMA, INTERPRO and PDB

More info · References · History · Statistics · HOWTO link to this server · FTP · ProDom in SRS format · InterPro

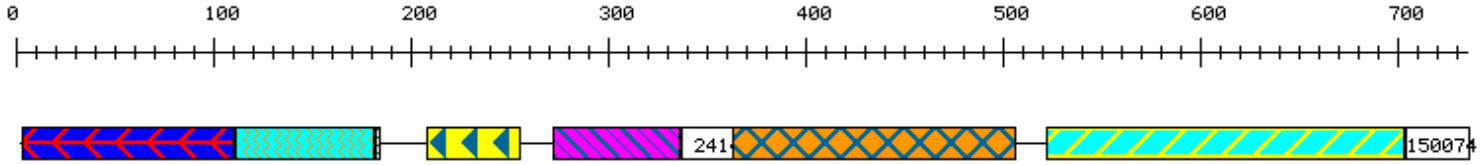
Your Query : "" vs ProDom release 2001.1

```
Warning: Original output has been filtered to yield non-redundant similarities

BLASTP 2.0.8 [Jan-05-1999]


Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer,
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.

Query=
         (735 letters)

Database: ProDom2001.1 multiple alignments
           1,057,487 sequences; 119,872,694 total letters

Searching.................................................done



ProDom domains producing High-scoring Segment Pairs:

  Position   ProDom domain                                        Score E value

     1-108   #PD010653  KINASE DIACYLGLYCEROL ALPHA TRANSFERASE     561  3e-57
   109-179   #PD009174  KINASE CALCIUM-BINDING CHANNEL POTASSIUM    348  3e-32
   112-182   #PD000012                                            101  0.003
   112-180   #PD024029                                            107  6e-04
   206-253   #PD000215                                            282  2e-24
   270-334   #PD011450  KINASE DIACYLGLYCEROL TRANSFERASE          386  1e-36
   335-380   #PD241406  KINASE ALPHA DIACYLGLYCEROL DAG            236  4e-19
   335-380   #PD343891  KINASE ALPHA DIACYLGLYCEROL DIGLYCERIDE     126  3e-06
   361-504   #PD002780  KINASE DIACYLGLYCEROL TRANSFERASE BINDING   644  6e-67
   520-701   #PD002939  KINASE DIACYLGLYCEROL TRANSFERASE          952  e-103
   702-734   #PD150074  KINASE DIACYLGLYCEROL ALPHA TRANSFERASE    190  1e-13

>PD002939 (Closest domain: KDGA_HUMAN 520-701)
Number of sequences in family: 44
Most frequent protein names:   KDGA(4) KDGL(3)
Commentary (automatic):        KINASE
            DIACYLGLYCEROL TRANSFERASE DIGLYCERIDE PHORBOL-ESTER DAG
            BINDING ALPHA DGK- MULTIGENE
            Length = 182
   Score = 952  (375 bits), Expect = e-103
  Identities = 182/182 (100%), Positives = 182/182 (100%)

 Query:   520 IINNYFSIGVDASIAHRFHIMREKYPEKFNSRMKNKLWYFEFATSESIFSTCKKLEESLT 579
             IINNYFSIGVDASIAHRFHIMREKYPEKFNSRMKNKLWYFEFATSESIFSTCKKLEESLT
 Sbjct:   520 IINNYFSIGVDASIAHRFHIMREKYPEKFNSRMKNKLWYFEFATSESIFSTCKKLEESLT 579
```

EMBL
European Bioinformatics



Clickable ncbi-blastp2 output (*Go to HSPs*)



Form to "Multalin" your sub sequence with ProDom domain

Domain ID	Sequence BEGIN	Sequence END	
#PD002939	520	701	Submit Query
#PD002780	361	504	Submit Query
#PD010653	1	108	Submit Query
#PD011450	270	334	Submit Query
#PD009174	109	179	Submit Query
#PD000215	206	253	Submit Query
#PD241406	335	380	Submit Query
#PD150074	702	734	Submit Query
#PD343891	335	380	Submit Query
#PD024029	112	180	Submit Query
#PD000012	112	182	Submit Query

Form to create a "Swiss-Model" for your sub sequence using ProDom domain homology

Domain ID	Sequence BEGIN	Sequence END	
#PD009174	109	179	Submit Query
#PD000215	206	253	Submit Query
#PD000012	112	182	Submit Query

Click here to compare your sequence to the NEW Database ProDom-CG

EMBL
European

ProDom Release 2001.1

Domain PD009174

Graphic representation of all proteins containing this domain.

Display the family as a tree.

MSF Alignment in MSF format.

Build an ESPript view

Run Predict Protein server with this domain

Go to 3D Structures

Most frequent protein names	KDGA(4) RECO(3) KDGB(2)
Commentary (automatic)	KINASE CALCIUM-BINDING CHANNEL POTASSIUM REPEAT DIACYLGLYCEROL MYRISTATE ALPHA TRANSFERASE DIGLYCERIDE
Alignment length	82
Number of sequences in family	72
Consistency indicator	DIAMETER: 330 PAM RADIUS OF GYRATION: 101 PAM SEQUENCE CLOSEST TO CONSENSUS: NCS1_

InterPro 2.0

ID	Accession number
"EF-hand"	IPR002048

Pfam-A

ID	Accession number
efhand	PF00036

PROSITE

Consensus position	PROSITE Pattern	PROSITE Entry Documentation
18-30	EF_HAND	PDOC00018

PROSITE

Consensus position	PROSITE Pattern	PROSITE Entry Documentation
18-30	EF_HAND	PDOC00018
70-82	EF_HAND	PDOC00018

Sample 3D Structures

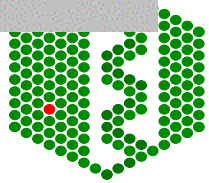
SwissProt ID	position	PDB Short	chain number	position	Entrez	Scop	Rasmol
RECO_BOVIN	92-171	1iku	1	92-171	Entrez	Scop	Rasmol
RECO_HUMAN	94-125	1jsa	1	94-125	Entrez	Scop	Rasmol
RECO_MOUSE	99-125	1rec	1	87-113	Entrez	Scop	Rasmol

Other descriptions of the family corresponding to this alignment 

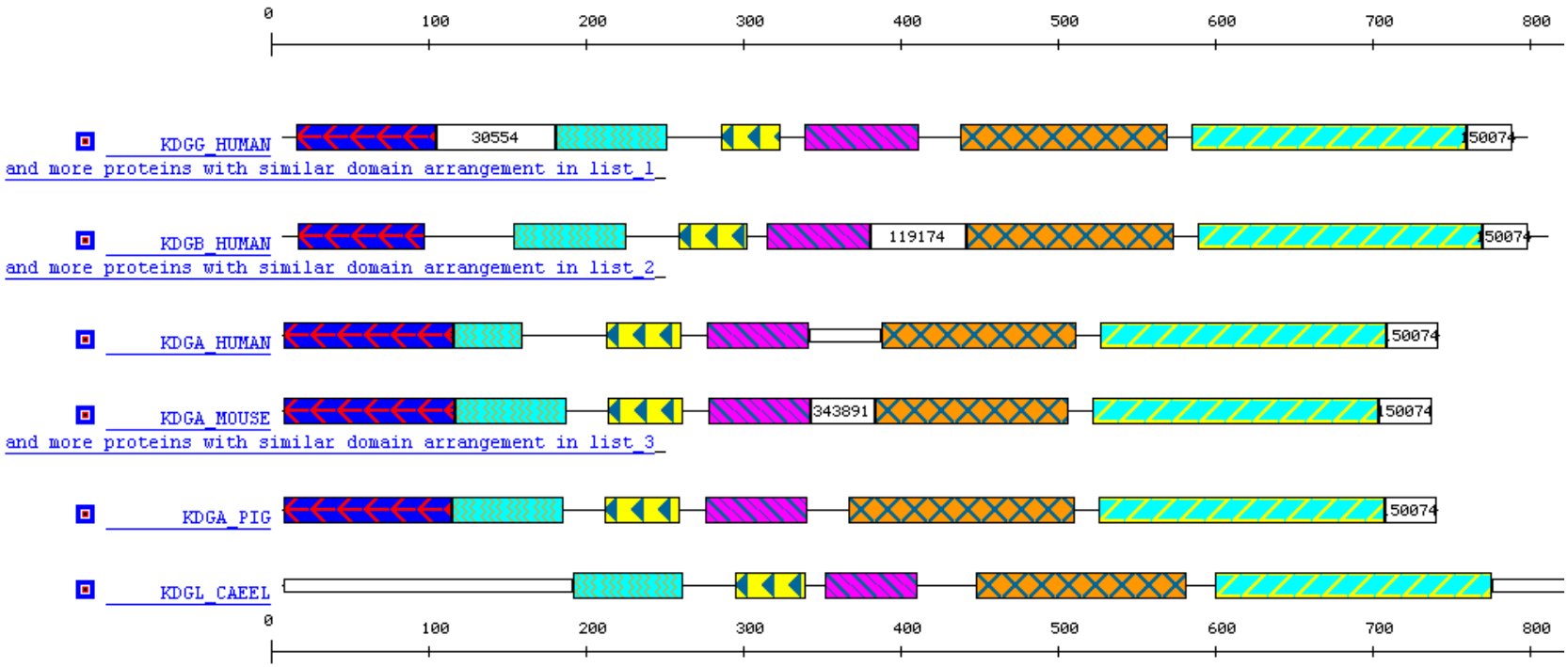
	Sequence ID	start	end	weight	10 20 30 40 50 60 70 80
5 ⊞	Q9TT98_PIG	.	.	5.51	...GSPEDKSRLMFTMYDLDENGFLSKDEFFTMMRSFIEI-------SNNC-LSKAQLAEVVESMFRESGFQDKEELTWEDF
1	Q9VQH2_DROME	826	899	4.34	.SRGKTDDKLRIIFDMCDNDRNGVIDKGELSEMMRSLVEI-------ARTTSLGDDQVTELIDGMFQDVGLEHKNHLTYQDF
9 ⊞	KDGB_RAT	.	.	11.71	..RGRPEDKLEFMFRLYDSDGNGFLDSSEIDKIISQMMHV-------AEYLEWDVTELRPILQEMMQEMDYDNDGFVSLDE.
1	O73763_RANPI	84	160	2.38	.AHGTPEDKLKWSFKLYDKDGDGAITRSEMLEIMRAVYKMSVVASL----TKVNPMTAEECTNRIFVRLDKDQNAIISLQEF
2 ⊞	GCA3_HUMAN	.	.	2.72	....KMEQKLKWYFKLYDADGNGSIDKNELLDMFMAVQAL----------NGQQTLSPEEFINLVFHKIDINNDGELTLEEF
3 ⊞	GCA2_RANPI	.	.	3.65	..RGKLEHKLKWTFKVYDKDGNGCIDKTELLEIVEGIYQLKKVCRRELETERGPLLTPEEVVDRIFQLVDENGDGQLSLDEF
1	Q9VW66_DROME	97	170	1.64	...GTPEEKLKWAFRMYDVDGNGVIDIQEMTKIVQAIYDM--LGA---CSSNRPADSAEERAKNIFAKMDENNDGQLTQDEF
1	NCS2_CAEEL	96	169	2.01	...GTPEQKLEWAFRMYDIDGNGTIDEKEMIKIIEAIYEM--LGP---EVTKSADDSPRKRAKMIFEKMDVNNDKELTLKEF
13 ⊞	SMOD_RANCA	.	.	11.38	TSAGKTNQKLEWAFSLYDVDGNGTISKNEVLEIVTAIFKM--IPPEDQKKLPEDENTPEKRADKIWAYFDKNDDDKLTEGEF
6 ⊞	VIS2_RAT	.	.	2.43	TSRGSFEQKLNWAFNMYDLDGDGKITRVEMLEIIEAIYKM--VGTVIMMKMNEDGLTPEQRVDKIFSKMDKNKDDQITLDEF
3 ⊞	Q9TZN4_CAEEL	.	.	4.32	TSRGNLDEHFAWAFKLYDVDNDGFITRDEMYDIVDAIYQM--VGNM--LPQPKDENTPQKRVDKIFTNMDKNHDGQLTREEF
27 ⊞	NCS1_YEAST	.	.	19.92	.SRGTVDEKLBWTFKLYDLNKDGYITWDEMFDIITSIYDM--MGKH--TYPHHTEEQPCEHVEQIFQKMDKNKDGVITIEEF
72	Consensus			72.01	TSRGTPEDKLEWAFKLYDVDGNGYIDKDEMLEIIKAIYEM--MGKH-AQYMPEDEDTPEERVDKIFQKMDKNNDGQITLEEF
2	PROSITE				---------- ----------

Minimal distance between sequences (in PAM)	20
Maximal number of clusters	12
If possible, clusterIDs should contain the following string (e.g.: human)	

To display a new alignment with these parameters, Submit Query
or click here to display all sequences in the family



For each protein you can retrieve "All proteins sharing a homologous domain" by clicking on the image [▣] adjacent to sequence name



Total : 9 protein(s)

Proteins having the same domain structure

list_1 >go back<

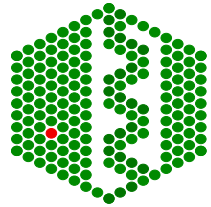
KDGG_HUMAN KDGG_RAT

list_2 >go back<

KDGB_HUMAN KDGB_RAT

PRINTS -Fingerprint DB

- http://www.bioinf.man.ac.uk/dbbrowser/PRINTS/
- Fingerprint- set of motifs used to predict occurrence of similar motifs in a sequence
- Built by iterative scanning of OWL database
- Multiple sequence alignment- identify conserved motifs- scan database with each motif- correlate hitlists for each- should have more sequences now- generate more motifs- repeat until convergence
- Recognition of individual elements in fingerprint is mutually conditional
- True members match all elements in order, subfamily may match part of fingerprint



PRINTS

Protein Fingerprint Database

PRINTS is a compendium of protein **fingerprints**. A fingerprint is a group of conserved motifs used to characterise a protein family; its diagnostic power is refined by iterative scanning of a *SWISS-PROT/TrEMBL* composite. Usually the motifs do not overlap, but are separated along a sequence, though they may be contiguous in 3D-space. Fingerprints can encode protein folds and functionalities more flexibly and powerfully than can single motifs, full diagnostic potency deriving from the mutual context provided by motif neighbours. **References**

New:

- SPRINT - *Search PRINTS-S (relational PRINTS)*
- InterPro - *Search the integrated InterPro family database*

Direct PRINTS access:

- By accession number
- By PRINTS code
- By database code
- By text
- By sequence
- By title
- By number of motifs
- By author
- By query language

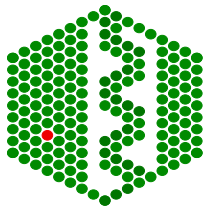
PRINTS search:

- Search PRINTS with NEW FingerPRINTScan
- FPScan
- GRAPHScan
- MULScan
- FingerPRINTScan binaries and source are available: contact scordis@bioinf.man.ac.uk

---

PRINTS BLAST search

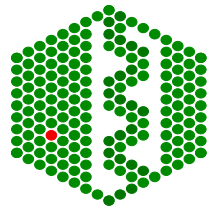
- Run a BLAST search of sequences in PRINTS



WORKLIST ENTRIES (99):

1433ZETA View alignment	14-3-3 protein zeta signature
6PFRUCTKNASE View alignment	6-phosphofructo-2-kinase family signature
ACETATEKNASE View alignment	Acetate kinase family signature
ACTIN View alignment	Actin signature
ACTIVIN2R View alignment	Activin type II receptor signature
ADENOKINASE View alignment	Adenosine kinase signature
ADENYLTKNASE View alignment	Adenylate kinase signature
ANNEXIN1 View alignment	Annexin type I signature
ANNEXIN2 View alignment	Annexin type II signature
BCTRLSENSOR View alignment	Bacterial sensor protein C-terminal signature
C2HCZNFINGER View alignment	C2HC-type zinc-finger signature
CALSEQUESTRN View alignment	Calsequestrin signature
CAMPKINASE View alignment	cAMP-dependent protein kinase signature
CASNKINASEII View alignment	Casein kinase II regulatory subunit family signature
CGMPKINASE View alignment	cGMP-dependent protein kinase signature
CLCHANNEL3 View alignment	CLC-3 chloride channel signature
CONNEXIN View alignment	Connexin signature
CONNEXIN43 View alignment	Gap junction alpha-1 protein (Cx43) signature
CYCLINKINASE View alignment	Cyclin-dependent kinase regulatory subunit signature
DAGPEDOMAIN View alignment	Diacylglycerol/phorbol-ester binding signature
EGFBLOOD View alignment	Type II EGF-like signature
FNTYPE1 View alignment	Fibronectin type I repeat signature
FNTYPE3 View alignment	Fibronectin type III repeat signature
GALCTOKINASE View alignment	Galactokinase signature
GLU5KINASE View alignment	Glutamate 5-kinase family signature
GPCRKINASE View alignment	GPCR kinase signature
GPROTEINB View alignment	Beta G-protein (transducin) signature
GPROTEINBRPT View alignment	G-protein beta WD-40 repeat signature
GPROTEING View alignment	Gamma G-protein (transducin) signature
HEATSHOCK70 View alignment	70kDa heat shock protein signature
HEATSHOCK90 View alignment	90kDa heat shock protein signature
HEXOKINASE View alignment	Hexokinase family signature
HISTRIAD View alignment	Histidine triad family signature
HOMSERKINASE View alignment	Homoserine kinase signature
HYETHTZKNASE View alignment	Hydroxyethylthiazole kinase family signature
INTRLEUKIN8R View alignment	Interleukin-8 receptor signature
INTRLEUKN8AR View alignment	Interleukin 8A receptor signature
INTRLEUKN8BR View alignment	Interleukin 8B receptor signature
KCHANNEL View alignment	Potassium channel signature
LEUZIPPRCREB View alignment	cAMP response element binding (CREB) protein signature
LMBPPROTEIN View alignment	LmbP protein signature
LPARECEPTOR View alignment	Lysophosphatidic acid receptor (EDG2) signature
MARCKS View alignment	MARCKS family signature
MEVGALKINASE View alignment	Mevalonate kinase family signature
MPIPHPHTASE View alignment	M-phase inducer phosphatase signature

EMBL
European Bioinformatics



ANNEXINI View alignment Annexin type I signature

Type of fingerprint: COMPOUND with 8 elements

Links:

PRINTS; PR00196 ANNEXIN; PR00198 ANNEXINII; PR00199 ANNEXINIII;
PRINTS; PR00200 ANNEXINIV; PR00201 ANNEXINV; PR00202 ANNEXINVI
INTERPRO; IPR002388
PDB; 1AIN 3Dinfo
SCOP; 1AIN
CATH; 1AIN

Creation date 26-OCT-1993; UPDATE 10-JUN-1999

1. BARTON, G.J., NEWMAN, R.H., FREEMONT, P.S. AND CRUMPTON, M.J.
Amino acid sequence analysis of the annexin super-gene family of proteins.
EUR.J.BIOCHEM. 198 749-760 (1991).

2. GEISOW, M.J.
Annexins-forms without function but not without fun.
TIBTECH 9 180-181 (1991).

The annexins are a family of proteins that bind to phospholipids in a calcium-dependent manner [1]. There are 11 distinct classes of annexin, each of which has an amino acid sequence consisting of an N-terminal `arm' followed by 4 or 8 copies of a conserved domain of 61 residues (only one of these residues, an arginine, is conserved between all copies): the calcium binding sites are found within the repeated domains [2]. Individual repeats (sometimes known as endonexin folds) consist of 5 alpha-helices wound into a right-handed superhelix.

Each annexin class is thought to have a specific function, although for some the precise role is unclear. The N-terminal residues are believed to confer the functional specificity that differentiates each class. Type I annexins inhibit phospholipase A2, either in response to inflammation, or following dephosphorylation by protein kinases involved in the signal transduction pathway. The protein may also associate with the cell cytoskeleton by binding to actin fibres.

ANNEXINI is an 8-element fingerprint that provides a signature for type I annexins. The fingerprint was derived from an initial alignment of 5 sequences: motif 1 encodes an N-terminal region; motifs 2 and 3 span the first repeat (cf. PROSITE pattern ANNEXIN (PS00223)); motifs 4-6 span the first half of 3 further repeats; and motifs 7 and 8 encode C-terminal regions. Motifs 2, 4, 5 and 6 include the conserved Arg, 4 and 6 also containing the GxG region associated with calcium binding. Two iterations on OWL21.1 were required to reach convergence, at which point a true set comprising 11 sequences was identified. Thirty nine partial matches were also found, corresponding to sequences from the remaining annexin classes.

SUMMARY INFORMATION

9	codes involving 8 elements
0	codes involving 7 elements
10	codes involving 6 elements
29	codes involving 5 elements
5	codes involving 4 elements
4	codes involving 3 elements
10	codes involving 2 elements

COMPOSITE FINGERPRINT INDEX

8|	9	9	9	9	9	9	9	9
7|	0	0	0	0	0	0	0	0
6|	0	10	10	10	10	10	0	10
5|	0	29	7	28	29	29	0	23
4|	0	4	1	5	5	5	0	0
3|	0	0	1	4	3	3	0	1
2|	0	9	1	1	0	1	0	8
--+								
|	1	2	3	4	5	6	7	8

True positives..

ANX1_HUMAN	ANX1_BOVIN	ANX1_CAVCU	ANX1_RAT
ANX1_RABIT	ANX1_MOUSE	AN12_COLLI	AN11_COLLI
ANX1_RODSP			

Subfamily: Codes involving 6 elements

Subfamily True positives..

093446	ANX2_HUMAN	ANX2_CHICK	ANX2_RAT
ANX2_BOVIN	ANX2_MOUSE	ANXB_XENLA	ANX2_XENLA
093444	ANX5_BOVIN		

Subfamily: Codes involving 5 elements

Subfamily True positives..

093447	ANX3_RAT	ANX5_CHICK	ANX6_MOUSE
035639	ANX4_MOUSE	ANX4_HUMAN	ANX4_RAT
ANXA_BOVIN	ANXB_BOVIN	ANX4_PIG	ANX4_BOVIN
ANXA_RABIT	ANX6_HUMAN	ANX4_CANFA	ANXA_HUMAN
ANX6_RAT	ANX5_RAT	ANX3_HUMAN	ANX5_MOUSE
ANXA_MOUSE	ANX5_HUMAN	ANXD_HUMAN	093445
ANX7_HUMAN	ANX7_MOUSE	ANX6_CHICK	ANXX_DROME
ANXD_CANFA			

Subfamily: Codes involving 4 elements

Subfamily True positives..

ANX8_HUMAN	035640	ANXC_HYDAT	ANX5_CYNPY
Q27512			

Subfamily: Codes involving 3 elements

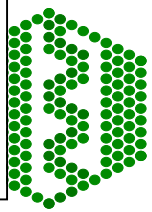
Subfamily True positives..

ANX7_XENLA	Q27473	ANX7_DICDI	059907

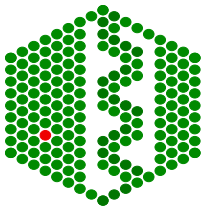
Subfamily: Codes involving 2 elements

Subfamily True positives..

Q27864	081536	081535	076027
Q43863	024131	Q42657	024132
082090	065848		



```
 Q27864          NEX1 ANNEXIN - CAENORHABDITIS ELEGANS.
 O81536          ANNEXIN P34 - LYCOPERSICON ESCULENTUM (TOMATO).
 O81535          ANNEXIN P35 - LYCOPERSICON ESCULENTUM (TOMATO).
 O76027          ANNEXIN 31 (ANNEXIN XXXI) - HOMO SAPIENS (HUMAN).
 Q43863          ANNEXIN P33 - ZEA MAYS (MAIZE).
 O24131          ANNEXIN - NICOTIANA TABACUM (COMMON TOBACCO).
 Q42657          ANNEXIN - CAPSICUM ANNUUM (BELL PEPPER).
 O24132          ANNEXIN - NICOTIANA TABACUM (COMMON TOBACCO).
 O82090          FIBER ANNEXIN - GOSSYPIUM HIRSUTUM (UPLAND COTTON).
 O65848          ANNEXIN - MEDICAGO TRUNCATULA (BARREL MEDIC).

SCAN HISTORY
 OWL21_1     2   100 NSINGLE
 OWL26_0     1   100 NSINGLE
 SPTR37_9f   2   122 NSINGLE

INITIAL MOTIF SETS

ANNEXINI1          Length of motif = 16   Motif number = 1
Annexin type I motif I - 1
                                   PCODE          ST    INT
FLKQAWFIENEEQEYV                   ANX1_HUMAN     6     6
FLKQARFLENQEQEYV                   ANX1_MOUSE     6     6
FLKQAYFIDNQEQDYV                   ANX1_CAVCU     7     7
FLKQAWFMENLEQECI                   ANX1_COLLI     7     7
FLKQACYIEKQEQEYV                     ANX1_RAT     6     6

ANNEXINI2          Length of motif = 23   Motif number = 2
Annexin type I motif II - 1
                                   PCODE          ST    INT
MVKGVDEATIIDILTKRNNAQRQ            ANX1_HUMAN     55    33
MVKGVDEATIIDILTKRTNAQRQ            ANX1_MOUSE     55    33
TVKGVDEATIIDILTKRNNAQRQ            ANX1_CAVCU     56    33
TAKGVDEATIIDIMTTRTNAQRP            ANX1_COLLI     51    28
MVKGVDEATIIDILTKRTNAQRQ              ANX1_RAT     55    33

ANNEXINI3          Length of motif = 17   Motif number = 3
Annexin type I motif III - 1
                                   PCODE          ST    INT
LKKALTGHLEEVVLALL                  ANX1_HUMAN     95    17
LRKALTGHLEEVVLAML                  ANX1_MOUSE     95    17
LKKALTGHLEEVVLALL                  ANX1_CAVCU     96    17
MKRVLKSHLEDVVVALL                  ANX1_COLLI     91    17
LKKALTGHLEEVVLAML                    ANX1_RAT     95    17

ANNEXINI4          Length of motif = 22   Motif number = 4
Annexin type I motif IV - 1
                                   PCODE          ST     INT
LRAAMKGLGTDEDTLIEILASR             ANX1_HUMAN     122    10
LRGAMKGLGTDEDTLIEILTTR             ANX1_MOUSE     122    10
LRAAMKGLGTDEDTLIEILVSR             ANX1_CAVCU     123    10
LRACMKGHGTDEDTLIEILASR             ANX1_COLLI     118    10
LRAAMKGLGTDEDTLIEILTTR               ANX1_RAT     122    10
```

Scan of sequence: USER_SEQUENCE

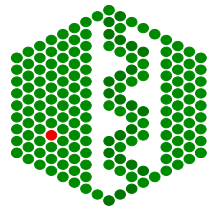
Highest scoring fingerprints for your query		
Fingerprint	**E-value**	**GRAPHScan**
ANNEXINI (relations)	7.394563e-92	Graphic
ANNEXIN (relations)	2.174040e-66	Graphic
ANNEXINII (relations)	3.575397e-54	Graphic
ANNEXINV (relations)	1.433121e-44	Graphic
ANNEXINIII (relations)	1.698603e-44	Graphic
ANNEXINVI (relations)	1.798458e-39	Graphic
ANNEXINIV (relations)	7.809646e-39	Graphic

for further information choose any of the following options

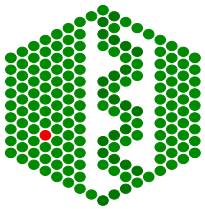
- Simple - Top Ten
- Detailed - Top Ten (detailed by motif)

Back to top

Ten top scoring fingerprints for your query								
Fingerprint	**No. of Motifs**	**SumId**	**AveId**	**PfScore**	**Pvalue**	**Evalue**	**GRAPHScan**	
ANNEXINI	8 of 8	7.1e+02	88	6586	2.9e-97	7.4e-92	IIIIIIII	Graphic
ANNEXIN	6 of 7	381.47	63.58	3841	8.5e-72	2.2e-66	IIIII.I	Graphic
ANNEXINII	6 of 8	408.76	68.13	4236	1.4e-59	3.6e-54	.IIII.II	Graphic
ANNEXINV	6 of 8	325.08	54.18	3484	5.6e-50	1.4e-44	I.III.II	Graphic
ANNEXINIII	6 of 8	343.25	57.21	3780	6.6e-50	1.7e-44	.III.III	Graphic
ANNEXINVI	6 of 8	317.61	52.94	3406	7e-45	1.8e-39	.IIII.II	Graphic
ANNEXINIV	6 of 8	337.19	56.20	3374	3e-44	7.8e-39	I..IIIII	Graphic
RSOLVASERUVC	2 of 5	58.57	29.29	414	5.3e-06	1.4	i..i.	Graphic
VD10PROTEIN	2 of 6	49.20	24.60	419	4.5e-05	12	..ii..	Graphic
JOSEPHIN	2 of 11	63.19	31.60	495	8.5e-05	22	.i.....I..	Graphic

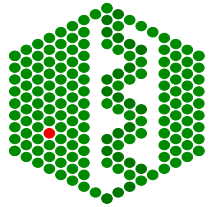


Ten top scoring fingerprints for your query. Detailed by motif									
FingerPrint Name	Motif Number	IdScore	PfScore	Pval	Sequence	Length	low	Pos	high
ANNEXINI	1 of 8	78.47	712	3.52e-10	FLKQAWFIENEEQEYV	16	6	6	7
	2 of 8	89.86	1001	2.22e-16	MVKGVDEATIIDILTKRNNAQRQ	23	51	55	56
	3 of 8	84.31	692	7.82e-12	LKKALTGHLEEVVLALL	17	91	95	96
	4 of 8	92.42	962	1.11e-16	LRAAMKGLGTDEDTLIEILASR	22	118	122	123
	5 of 8	86.42	1210	1.00e-16	LYEAGERRKGTDVNVFNTILTTRSYPQ	27	201	205	206
	6 of 8	89.42	923	4.66e-15	MKGVGTRHKALIRIMVSRSEI	21	280	285	286
	7 of 8	92.59	420	1.55e-05	ISLCQAILD	9	315	320	321
	8 of 8	93.65	666	5.87e-10	ETKGDYEKILVALC	14	324	329	330
ANNEXIN	1 of 7	60.79	647	1.23e-12	MVKGVDEATIIDILTKRNNAQRQ	23	26	55	384
	2 of 7	54.68	463	8.24e-10	LKKALTGHLEEVVLALL	17	66	95	424
	3 of 7	75.00	781	2.89e-15	LRAAMKGLGTDEDTLIEILASR	22	93	122	451
	4 of 7	57.75	737	5.55e-16	LYEAGERRKGTDVNVFNTILTTRSYPQ	27	175	205	534
	5 of 7	72.13	724	1.27e-13	MKGVGTRHKALIRIMVSRSEI	21	255	285	614
	7 of 7	61.11	489	4.13e-08	ETKGDYEKILVALC	14	299	329	658
ANNEXINII	2 of 8	68.32	786	2.14e-11	MVKGVDEATIIDILTKRNNAQRQ	23	46	55	47
	3 of 8	66.39	547	1.77e-09	LKKALTGHLEEVVLALL	17	86	95	87
	4 of 8	79.87	901	3.16e-13	LRAAMKGLGTDEDTLIEILASR	22	113	122	114
	5 of 8	56.08	724	9.41e-10	LYEAGERRKGTDVNVFNTILTTRSYPQ	27	197	205	198
	7 of 8	76.87	764	1.09e-12	MKGVGTRHKALIRIMVSRSEI	21	277	285	278
	8 of 8	61.22	514	1.13e-06	ETKGDYEKILVALC	14	321	329	322
ANNEXINV	1 of 8	52.17	570	7.59e-09	MVKGVDEATIIDILTKRNNAQRQ	23	26	55	31
	3 of 8	42.86	388	5.26e-06	LKKALTGHLEEVVLALL	17	66	95	71
	4 of 8	64.94	679	1.43e-11	LRAAMKGLGTDEDTLIEILASR	22	93	122	98
	5 of 8	48.15	574	4.54e-08	LYEAGERRKGTDVNVFNTILTTRSYPQ	27	175	205	181
	7 of 8	60.85	808	5.88e-13	MKGVGTRHKALIRIMVSRSEIDMNDIK	27	255	285	261
	8 of 8	56.12	465	3.66e-06	ETKGDYEKILVALC	14	299	329	305
ANNEXINIII	2 of 8	59.42	640	1.90e-09	MVKGVDEATIIDILTKRNNAQRQ	23	32	55	33
	3 of 8	50.98	433	1.04e-05	LKKALTGHLEEVVLALL	17	72	95	73
	4 of 8	66.67	759	4.94e-12	LRAAMKGLGTDEDTLIEILASR	22	99	122	100
	6 of 8	55.56	762	2.11e-09	LYEAGERRKGTDVNVFNTILTTRSYPQ	27	182	205	183
	7 of 8	65.38	760	2.10e-11	KGVGTRHKALIRIMVSRSEIDMNDIK	26	263	286	264
	8 of 8	45.24	426	1.52e-05	ETKGDYEKILVALC	14	306	329	307

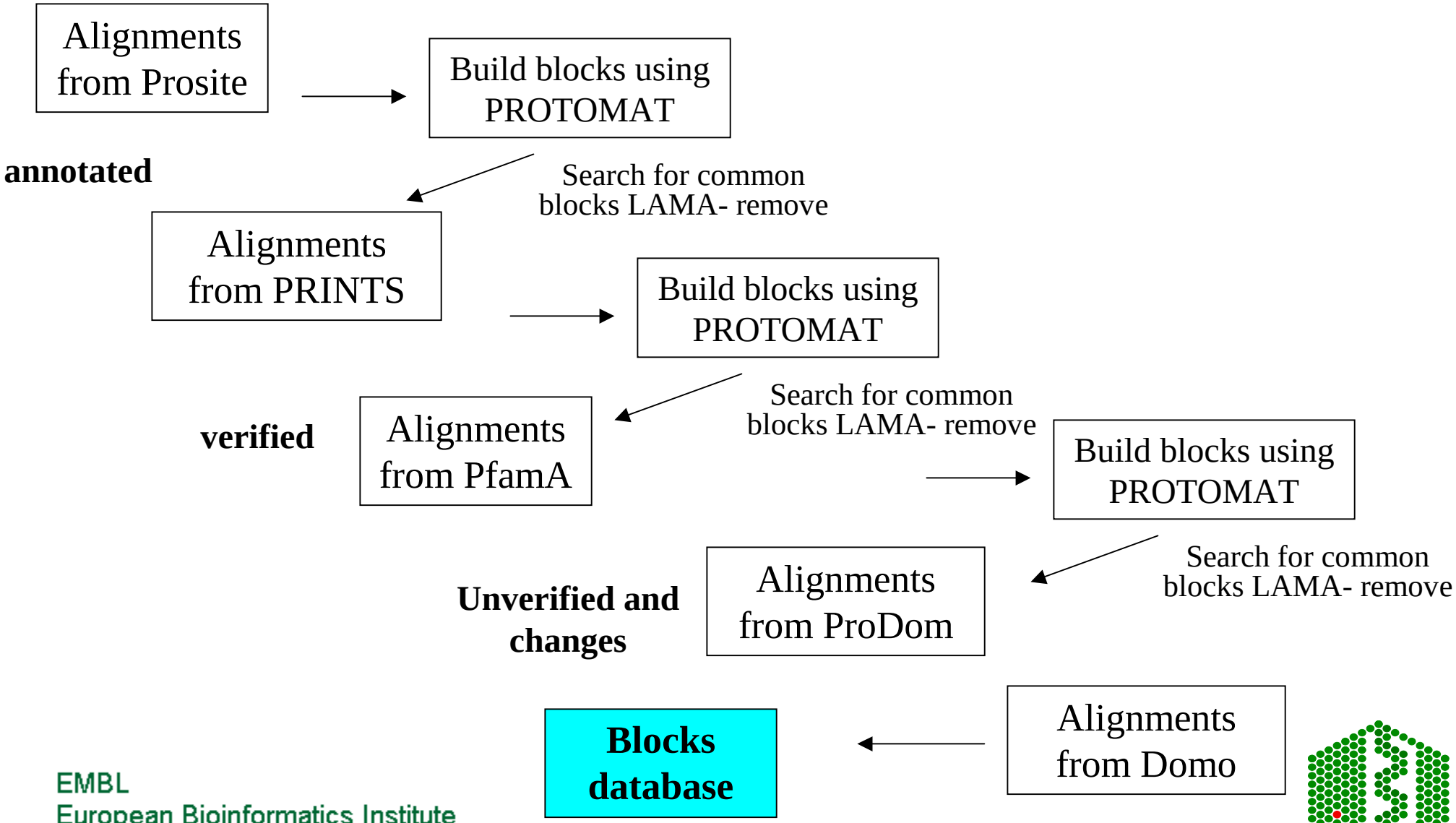


BLOCKS

- http://www.blocks.fhcrc.org/
- Multiply aligned ungapped segments corresponding to most highly conserved regions of proteins- represented in profile
- Built up using PROTOMAT (BLOSUM scoring model), calibrated against SWISS-PROT, use LAMA to search blocks against blocks
- Starting sequences from Prosite, PRINTS, Pfam, ProDom and Domo - total of 2129 families



Building of Blocks

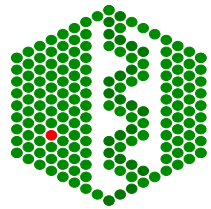


SEARCHING BLOCKS

- Compare a protein or DNA (1-6 frames) sequence to database of blocks
- Blocks Searcher- used via internet or email:

  First position of sequence aligned to first position of first block - score for that position, score summed over width of alignment, then block is aligned with next position etc for all blocks in database- get best alignment score. Search is slow (350 aa/2 min)

- Can search database of PSI-BLAST PSSMs for each blocks family using IMPALA



Blocks WWW Server

BLOCKS

A service for biological sequence analysis at the Fred Hutchinson Cancer Research Center in Seattle, Washington, USA.
Visit the Blocks mirror site at the Weizmann Institute of Science in Israel.

Contents:

- What's New as of 18 November 2000
- About Blocks
- Current Blocks Release
- Blocks Bibliography

- Get Blocks by key word
- Get Blocks by number

- Block Searcher to search a sequence vs Blocks
- Block Searcher Help

- Reverse PSI-BLAST Searcher to search a sequence vs Blocks using NCBI's RPS-BLAST program
- RPS-BLAST Help
- Impala Searcher to search a sequence vs Blocks using NCBI's IMPALA program
- IMPALA Help

- Block Maker to create Blocks
- Block Maker Help

- Multiple Alignment Processor to excise Blocks from multiple alignments

- LAMA Searcher to search Blocks vs Blocks
- LAMA Help

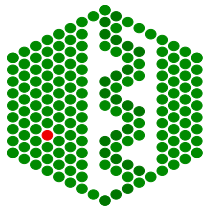
- COBBLER to search embedded Blocks vs sequence databases
- About COBBLER

- CODEHOP to design PCR primers from Blocks
- CODEHOP Help

- Biassed Block Checker

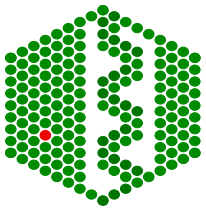
- Blocks FTP Site

- Guide to Sequence Analysis Tools
- Protein Family Sites and Resources



The following 51 item(s) match your query `**kinase**':

- DM00522
  DOMO2.0.DM00522 499 kw TRYPSIN KINASE KUNITZ PANCREATIC.
- PR00989
  ADENOKINASE Adenosine kinase signature
- PR01049
  PHOSPHBKNASE Phosphorylase kinase family signature
- PR00988
  URIDINKINASE Uridine kinase signature
- PR01099
  HYETHTZKNASE Hydroxyethylthiazole kinase family signature
- BP04294
  p99.1.4294 CYCLIN-DEPENDENT KINASE INHIBI
- PR00104
  CGMPKINASE cGMP-dependent protein kinase signature
- PF01504
  PIP5K Phosphatidylinositol-4-phosphate 5-Kinase
- PR00678
  PI3KINASEP85 PI3 kinase P85 regulatory subunit signature
- BP04994
  p99.1.4994 KINASE DEHYDROGENASE PYRUVATE L
- BP02447
  p99.1.2447 PROTEIN KINASE SENSOR SENSORY
- IPB000829
  DAGK_prokar Prokaryotic diacylglycerol kinase
- BP03894
  p99.1.3894 KINASE ACETYLGLUTAMATE DEHYDROGENASE T
- PF01633
  Choline_kinase Choline/ethanolamine kinase
- PR01469
  CARBMTKINASE Bacterial carbamate kinase signature
- PF01687
  FAD_Synth Riboflavin kinase / FAD synthetase
- PR01100
  SHIKIMTKNASE Shikimate kinase family signature
- PR00296
  CYCLINKINASE Cyclin-dependent kinase regulatory subunit signature
- IPB000623
  SKI Shikimate kinase
- PR00958
  HOMSERKINASE Homoserine kinase signature
- IPB000789
  CKS Cyclin-dependent kinase, regulatory subunit
- PR00472
  CASNKINASEII Casein kinase II regulatory subunit family signature
- PR00471
  ACETATEKNASE Acetate kinase family signature
- BP01519
  p99.1.1519 KINASE THYMIDINE TRANSFERASE DNA SYNTHESIS AT
- IPB000062
  THYMIDYLATE_KINASE Thymidylate kinase
- PR00991



IPB000829: DAGK_prokar

Prokaryotic diacylglycerol kinase

- Introduction
- Block number IPB000829A
- Block number IPB000829B
- Block number IPB000829C

- COBBLER sequence and BLAST searches [About COBBLER]
- LAMA search of all blocks vs a blocks database [About LAMA]
- MAST Search of all blocks vs a sequence database [About MAST]

- CODEHOP to design PCR primers from blocks [About CODEHOP]

- Logos.[About Logos]
  Select display format: [GIF] [PDF] [Postscript]

- Block Map [About Maps]

- Tree from Blocks alignment. [About Trees] [About ProWeb Display]
  Select display format: [Data] [XBitmap] [GIF] [PDF] [Postscript] [Newick] [ProWeb]

- InterPro entry IPR000829 (source of sequences used to make blocks)
- Additional Links

Introduction

```
          Blocks Database Version 12.0, June 2000
     Copyright 2000 by Fred Hutchinson Cancer Research Center
          1100 Fairview AV N, Seattle, WA 98109, USA
  Please cite: S Henikoff & JG Henikoff (1991) Automated assembly of
  protein blocks for database searching, Nucleic Acids Res. 19:6565-6572.
  Blocks made by PROTOMAT for protein families documented in InterPro 1.0
  and SWISS-PROT 38: PROSITE patterns were not used. ID, AC and DE are
  adapted from prosite.dat; BL is PROTOMAT information. For each segment,
  the sequence ID is followed by the position of the first residue in
  the segment. Segments are clustered if >=80% of aligned residues match
  between any pair of segments. Sequence weights are shown to the right of
  each segment. The higher the weight the more dissimilar the segment is
  from other segments in the block. These weights were obtained using the
  position-based method of S Henikoff & JG Henikoff (1994), JMB 243:574-578.
  ==========================================================================
```

Block IPB000829A

```
ID   DAGK_prokar; BLOCK
AC   IPB000829A; distance from previous block=(2,57)
DE   Prokaryotic diacylglycerol kinase
BL   GFR;  width=41; seqs=9; 99.5%=1974; strength=1159
KDGL_HELPY|P56411  (  15) RLFKALFYSKDGLKCAWIEESAFRQIVILALFCIVLASYLA  53
Q9ZLE0             (  15) RLFKALFYSKDGLKCAWAEESAFRQIVILALFCIVLASYLT  54

KDGL_PSEDE|P29945  (  23) RHLFAAASYSLGGAKRLIGEAAFRHELIAFAAAMIAFIIVG  65
KDGL_RHIME|Q06119  (  23) RHLFAAASYSFGGAKRLIGEAAFRHELIAFAVAMVAFMIVG  67

KDGL_ECOLI|P00556  (   9) RIIKAAGYSWKGLRAAWINEAAFRQEGVAVLLAVVIACWLD  72

KDGL_HAEIN|P44424  (   9) HLINSTKYSLQGLKSAFKNETAFRHECFLACILIPLTFFLG  82

KDGL_STRMU|Q05888  (  19) TLTSSLEFALTGIFTAFKEERNMKKHAVSALLAVIAGLVFK 100

KDGL_SYNY3|Q55143  (  58) NLLVSFRYAWAGVSYAFATQRNFRIHTFTGVAVITAASLLH  94

KDGL_BACSU|P19638  (   3) RFFKSFVHAGRGIWETARTERNFQFHAAAACAVLICGFLVE  99
//
```

[Return to top]

Block IPB000829B

```
ID   DAGK_prokar; BLOCK
AC   IPB000829B; distance from previous block=(18,19)
DE   Prokaryotic diacylglycerol kinase
BL   EED;  width=12; seqs=9; 99.5%=687; strength=1117
KDGL_HELPY|P56411  (  75) ELINSSIEKAVD  88
Q9ZLE0             (  75) ELINSSIEKAVD  88

KDGL_ECOLI|P00556  (  69) EILNSAIEAVVD  83
KDGL_HAEIN|P44424  (  69) ELLNSAVETVVD  96
KDGL_STRMU|Q05888  (  79) EIVNSAIENVVD 100
KDGL_SYNY3|Q55143  ( 118) ELLNTALESVVD  97
KDGL_BACSU|P19638  (  63) ELLNTAIEHTVD  91

KDGL_PSEDE|P29945  (  83) EAINTAIEEIVD  84
KDGL_RHIME|Q06119  (  83) EAINTAIEEIVD  84
//
```

[Return to top]

Block IPB000829C

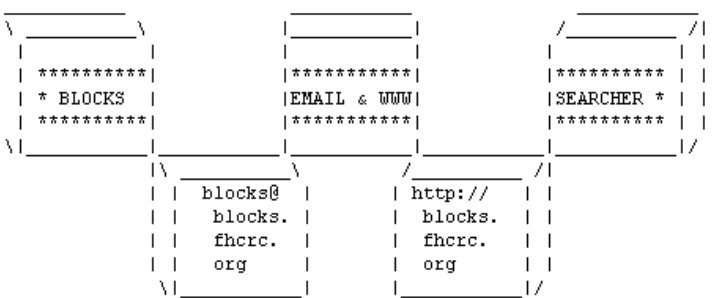
```
ID   DAGK_prokar; BLOCK
AC   IPB000829C; distance from previous block=(1,19)
DE   Prokaryotic diacylglycerol kinase
BL   GLA;  width=11; seqs=9; 99.5%=623; strength=919
KDGL_HELPY|P56411  (  89) GTEFHPLAKKA  36
Q9ZLE0             (  89) GTEFHPLAKKA  36
```

Block Searcher Results

Go to hits

Introduction

```
BLKPROB Version  5/21/00.1
Database=../data-blplus/blocks.dat
```



Here are your search results. The database searched was Blocks+ (15 Nov 00) which includes: BLOCKS 12.0 consisting of 4071 blocks representing 998 groups documented in InterPro 1.0/PROSITE, plus 3479 blocks from 553 groups documented in PRINTS 28.0 but not represented in BLOCKS, plus 2335 blocks from 559 groups

EMBL
European Bioinformatics Institute

Hits

```
Query=Unknown Unknown
Size=735 Amino Acids
Blocks Searched=11117
Alignments Done=         8443048
Cutoff combined expected value for hits=  1
Cutoff block expected value for repeats/other=  1
```

Family		Strand	Blocks	Combined E-value
PF00609	Diacylglycerol kinase accessory dom	1	8 of 8	4.3e-158
IPB002219	Phorbol esters/diacylglycerol bindi	1	4 of 5	1.5e-18
PR01362	Flagellar calcium-binding protein (	1	1 of 9	0.2
IPB001604	DNA/RNA non-specific endonuclease	1	1 of 5	0.57

>PF00609 8/8 blocks Combined E-value=4.3e-158: Diacylglycerol kinase accessory domain (presumed)

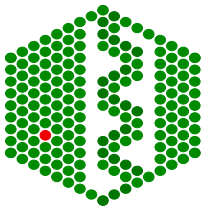
Block	Frame	Location (aa)	Block E-value
PF00609A	0	299-305	0.0011
PF00609B	0	319-331	5.9e-09
PF00609C	0	376-414	3.3e-33
PF00609D	0	428-442	3.5e-12
PF00609E	0	454-477	2.7e-23
PF00609F	0	522-558	9e-35
PF00609G	0	595-610	7.5e-10
PF00609H	0	694-720	1.9e-23

Other reported alignments:

```
                        |---  512 amino acids---|
          PF00609 A:..B:.................CC:DE::.....................FF::G:::..H
          Unknown                        ::::::::::::::::A:B::CC:D:E::FF::G::::H

PF00609A            <->A   (21,694):298
KDGA_HUMAN|P23743  299     HCVWCHL
                           |||||||
Unknown            299     HCVWCHL

PF00609B            A<->B  (11,56):13
KDGA_HUMAN|P23743  319     CDCGLLRDHILPP
                           |||||||||||||
Unknown            319     CDCGLLRDHILPP

PF00609C            B<->C  (21,341):44
KDGA_HUMAN|P23743  376     PLLVFVNPKSGGKQGQRVLWKFQYILNPRQVFNLLKDGP
                           |||||||||||||||||||||||||||||||||||||||
Unknown            376     PLLVFVNPKSGGKQGQRVLWKFQYILNPRQVFNLLKDGP

PF00609D            C<->D  (11,15):13
KDGA_HUMAN|P23743  428     RILVCGGDGTVGWIL
                           |||||||||||||||
Unknown            428     RILVCGGDGTVGWIL
```

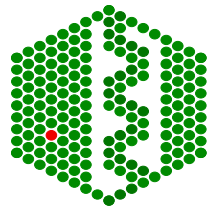
```
--------------------------------------------------------------------------------
>IPB002219 4/5 blocks Combined E-value= 1.5e-18: Phorbol esters/diacylglycerol binding domain
Block     Frame    Location (aa)      Block E-value
IPB002219A  0        206-222            9.4e-07
IPB002219B  0        231-246            9.3e-07
IPB002219C  0        432-442               0.33
IPB002219D  0        489-498            1.9e+02
Up to 50 repeats expected:
Other reported alignments:

                        |---  563 amino acids---|
          IPB002219 A.......B:::.............C:.........................D........E
            Unknown <::::::AB::::::::C::D

IPB002219A           <->A   (21,485):205
KDGA_HUMAN|P23743  206     HMWRPKRFPRPVYCNLC
                           |||||||||||||||||
Unknown            206     HMWRPKRFPRPVYCNLC

IPB002219B           A<->B  (5,160):8
KDGA_HUMAN|P23743  231     QGLSCNLCKYTVHDQC
                           ||||||||||||||||
Unknown            231     QGLSCNLCKYTVHDQC

IPB002219C           B<->C  (60,355):185
KDGA_MOUSE|088673  427     CGGDGTVGWVL
                           ||||||||| |
Unknown            432     CGGDGTVGWIL

IPB002219D           C<->D  (14,580):46
KDGA_HUMAN|P23743  489     LEMSKVVHMD
                           ||||||||||
Unknown            489     LEMSKVVHMD

--------------------------------------------------------------------------------
>PR01362 1/9 blocks Combined E-value=     0.2: Flagellar calcium-binding protein (calflagin) signature
Block     Frame    Location (aa)      Block E-value
PR01362H    0        164-180               0.21
Other reported alignments:

                        |---   79 amino acids---|
           PR01362 AAAAAAA:::BBBBB:CCCCCC:DDDDD:EEEEEEFFFFFF::GGGG:HHHHHIIIIII
           Unknown <::::::::::::::::::::::::::::::::::::::::::::::::HHHHH

PR01362H             <->H   (166,365):163
FCA1_TRYRA|Q27052  166     FKELDRNGSGSVTFDEF
                            || |  |||||   |
Unknown            164     mKEIDydGSGSVsqaEw

--------------------------------------------------------------------------------
>IPB001604 1/5 blocks Combined E-value=    0.57: DNA/RNA non-specific endonuclease
Block     Frame    Location (aa)      Block E-value
IPB001604A  0        541-567               0.57
Other reported alignments:
```

EMBL
European Bioinformatics In



TIGRFAMs

- http://www.tigr.org/TIGRFAMs
- Collection of protein families in HMMs built with curated multiple sequence alignments and with associated functional information
- Equivalog- homologous proteins conserved with respect to function since last ancestor (other pattern databases concentrate on related seq not function)
- > 800 non-overlapping families -can search by text or sequence
- Has information for automatic annotation of function, weighted towards microbial genomes



TIGR THE INSTITUTE FOR GENOMIC RESEARCH

tigr fams
tigr protein families

TIGR FAMs Page | Text Search | HMM Search | Download | CMR Home | Help

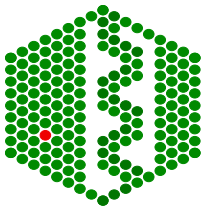
All Genomes in the CMR: TIGRFAM Search Results

TIGRFAMs/HMMs containing pattern "kinase"

The follwing table contains a list of all TIGRFAMs or HMMs with the pattern "kinase" in its name or accession number. Click on the accession number to get to the HMM Profile page on a TIGRFAM. Click on the link below the table to start a new search.

Name	Accession#	Common_Name
6PF2K	PF01591	6-phosphofructo-2-kinase
aakinase	PF00696	Amino acid kinase family
Acetate_kinase	PF00871	Acetokinase family
ackA	TIGR00016	acetate kinase
adenylatekinase	PF00406	Adenylate kinase
APS_kinase	PF01583	Adenylylsulfate kinase
apsK	TIGR00455	adenylylsulfate kinase
arcC	TIGR00746	carbamate kinase
argB	TIGR00761	acetylglutamate kinase
asp_kin_monofn	TIGR00656	aspartate kinase, monofunctional class
asp_kinases	TIGR00657	aspartate kinase
CDI	PF02234	Cyclin-dependent kinase inhibitor
cdk7	TIGR00570	cdk-activating kinase assembly factor (cdk7)
Choline_kinase	PF01633	Choline/ethanolamine kinase
CK_II_beta	PF01214	Casein kinase II regulatory subunit
CKS	PF01111	Cyclin-dependent kinase regulatory subunit
cmk	TIGR00017	cytidylate kinase
Cytidylate_kin	PF02224	Cytidylate kinase
DAGK_prokar	PF01219	Prokaryotic diacylglycerol kinase

Text search results



Example entry

HMM Profile Page
Accession #: TIGR00016 Name: ackA

Both TIGRFAMs and Pfams are displayed on this page. TIGRFAMs and Pfams are based on Hidden Markov Models or HMMs. An HMM is a statistical model for any system that can be represented as a succession of transitions between discrete states. Scores are reported both in bits of information and as an E-value. See below for more information on this TIGRFAM or Pfam and its HMM.

---

ackA Information: See below for detailed information on this family, including the cutoff score for inclusion in this family and the average score of genes/proteins in this family. To view all genes with the same EC number, click on the **EC Number** link. To view more information on the Role ID for this family, click on the **Role ID** link.

TIGRFAM Name: acetate kinase
Accession: TIGR00016
Trusted Cutoff: 300.00
EC Number: 2.7.2.1
Created: Apr 20 1999 2:05PM

Isology Type: equivalog
Author(s): Richardson DL
Avg. Score: 943.33 +/- 19.2
Role ID: 102
Last Modified: Mar 8 2001 2:32PM

HMM type: TOGA
HMM length: 416
Noise Cutoff: 30.00

References:
PROSITE: PDOC00826
Source: TIGR
Alignment source: clustalw

Comments: function: involved in the activation of acetate to acetyl CoA and the secretion of acetate. during an
growth of the organism, this enzyme is also involved in the synthesis of most of the ATP formed catabolically (
catalytic activity: ATP + acetate = ADP + acetyl phosphate. subcellular location: cytoplasmic.

Display Hits and Overlaps: To view all CMR proteins that are members of this HMM, click on **All CMR Hits**. T
overlapping HMMs, click on **Any overlapping HMMs?**.

All CMR Hits to TIGR00016 | Any Overlapping HMMs?

Alignment Display: View a multiple protein alignment display for this HMM. Choose to view the HMM alignme
FASTA or **MSF** format and then depress the submit button below. Depress the **JalView** button to start a prote
program. This program allows the user to view the alignment in different ways (i.e. highlight the identical and si
acids in the alignment).

MSF TIGRFAM Alignment | Submit

Jalview (viewer alignment program):

JalView

Members of the ackA HMM

elow are the current members of this HMM, including those not found in the CMR. Click on the **Protein ID** link to vie
ein report for a particular member of this family.

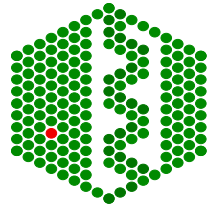
PROTEIN ID	COORDINATES	DATABASE
CMR: NT01BS3432	37-433	TIGR - CMR
SP: P37877	1-395	SWISS-PROT/TrEMBL
CMR: NT01EC2743	1-400	TIGR - CMR
CMR: HI1204	1-401	TIGR - CMR
GP: 1359437	1-400	GenBank
CMR: NT01SS0762	11-418	TIGR - CMR
SP: P73162	1-405	SWISS-PROT/TrEMBL
CMR: BB0622	6-410	TIGR - CMR
CMR: NT01MP0391	1-389	TIGR - CMR
SP: O51567	1-402	SWISS-PROT/TrEMBL
SP: P47599	1-389	SWISS-PROT/TrEMBL
SP: P11868	2-394	SWISS-PROT/TrEMBL
CMR: NT01EC3702	6-398	TIGR - CMR
GP: 10175815	1-391	GenBank
CMR: VC1098	1-397	TIGR - CMR
GP: 3006124	1-397	GenBank
SP: O59331	1-397	SWISS-PROT/TrEMBL
CMR: TM0274	1-399	TIGR - CMR

Sequence search result

```
hmmpfam - search a single seq against HMM database
HMMER 2.1.1 (Dec 1998)
Copyright (C) 1992-1998 Washington University School of Medicine
HMMER is freely distributed under the GNU General Public License (GPL).
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
HMM file:                 ALL_LIB_bin.HMM
Sequence file:            hmmpfam-search-23339-990483518.in
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Query:  query  sequence

Scores for sequence family classification (score includes all domains):
Model     Description                                   Score    E-value
--------  -----------                                   -----    -------
TIGR00016 ackA: acetate kinase                          963.8   4.5e-286
PF00871   Acetate_kinase: Acetokinase family            872.6   1.2e-258

Parsed for domains:
Model     Domain  seq-f seq-t    hmm-f hmm-t      score  E-value
--------  ------- ----- -----    ----- -----      -----  -------
PF00871     1/1       3   388 ..     1   397 []   872.6 1.2e-258
TIGR00016   1/1       1   395 []     1   416 []   963.8 4.5e-286

Alignments of top-scoring domains:
PF00871: domain 1 of 1, from 3 to 388: score 872.6, E = 1.2e-258
                   *->kiLviNaGSSSlKfqlfdakkaegeevlasGLaErigidnarilqkv
                      ki+ iNaGSSSlKfqlf++++   e+vl++GL+Erigi++++++++
       query     3    KIIAINAGSSSLKFQLFEMPS---ETVLTKGLVERIGIADSVFTIS- 45

                   nggkKieektaiadHeeAlkhilntLkesdfgvikdlseIdAvGHRvVhG
                   ++g+K++e t+i+dH+ A+k++ln+L+  +fg+ikdl+eId++GHRvVhG
       query    46 VNGEKNTEVTDIPDHAVAVKMLLNKLT--EFGIIKDLNEIDGIGHRVVHG 93

                   GekFtesvlitdevieaIkdlieLAPLHNpAniiGIeiadkllppvkdkn
                   GekF++svl+tde+i++I+d++eLAPLHNpAni+GI++++++lp+v+  +
       query    94 GEKFSDSVLLTDETIKEIEDISELAPLHNPANIVGIKAFKEVLPNVP--A 141

                   VAVFDTaFHqTmPeeAyLYalPyelyeehGiRRYGFHGTSHkYVaqraak
                   VAVFDTaFHqTmPe++yLY+lPye+ye++GiR+YGFHGTSHkYV++raa+
       query   142 VAVFDTAFHQTMPEQSYLYSLPYEYYEKFGIRKYGFHGTSHKYVTERAAE 191

                   iLgKpledlnlItcHLGNGaSiaAvknGksvDTSMGfTPLEGLvMGTRSG
                   +Lg+pl+dl+lI cHLGNGaSiaAv++Gks+DTSMGfTPL+G++MGTRSG
       query   192 LLGRPLKDLRLISCHLGNGASIAAVEGGKSIDTSMGFTPLAGVAMGTRSG 241

                   dIDPaivvyLaeteglSadEivnlLNKKSGlLGlsGlsSDlRdvedaiee
                   +IDPa ++y++e++g++adE+ n+LNKKSGlLG+sG+sSDlRd+++a++e
       query   242 NIDPALIPYIMEKTGQTADEVLNTLNKKSGLLGISGFSSDLRDIVEATKE 291

TIGR00016: domain 1 of 1, from 1 to 395: score 963.8, E = 4.5e-286
                   *->meskkiLviNaGSSSlKFalfdaenDRLgekvPePLlsGLvErifla
                        ++ki++iNaGSSSlKF+lf+++    +e+v   L++GLvEri++a
       query     1    --MSKIIAINAGSSSLKFQLFEMP----SETV---LTKGLVERIGIA 38

                   nariktvnenggkkeeellaiadHqeAvkfilntLtnqsdkkIikllseI
                   ++ ++t+ +++g+k++e+++i+dH++Avk++ln+Lt   +++Iik+l+eI
       query    39 DS-VFTI-SVNGEKNTEVTDIPDHAVAVKMLLNKLT---EFGIIKDLNEI 83

                   dlIGHRVVHGGekftdSViitdevikkIkdiseLAPLHNpAhldGIEaal
                   d+IGHRVVHGGekf+dSV++tde+ik+I+diseLAPLHNpA++GI+a++
       query    84 DGIGHRVVHGGEKFSDSVLLTDETIKEIEDISELAPLHNPANIVGIKAFK 133

                   klkvlpkaKnVAVFDTaFHqTiPeeaYLYAlPyswYkehGiRRYGFHGTS
                     +vlp++++VAVFDTaFHqT+Pe++YLY+lPy++Y+++GiR+YGFHGTS
       query   134 --EVLPNVPAVAVFDTAFHQTMPEQSYLYSLPYEYYEKFGIRKYGFHGTS 181

                   HkYvtqraaklLNKpldd LnLIvCHLGNGASvcAvkNGkSiDTSMGfTPL
                   HkYvt+raa+lL++pl+dL+LI+CHLGNGAS++Av++GkSiDTSMGfTPL
       query   182 HKYVTERAAELLGRPLKDLRLISCHLGNGASIAAVEGGKSIDTSMGFTPL 231

                   EGLvMGTRSGDIDPAIisylaetlgmSaddientLNKkSGLLGisGlSSD
                   +G++MGTRSG+IDPA+i+y++e++g++ad+++ntLNKkSGLLGisG+SSD
       query   232 AGVAMGTRSGNIDPALIPYIMEKTGQTADEVLNTLNKKSGLLGISGFSSD 281

                   lRdiedayeEgneqAklAikvYvhRiakYIGsYiAsLegnrlDaiVFTGG
                   lRdi++a++Egne+A++A++v+++Ri+kYIGsY+A+++g  +Dai+FT+G
       query   282 LRDIVEATKEGNERAETALEVFASRIHKYIGSYAARMSG--VDAIIFTAG 329

                   IGENaaevReLvleglevLGlelDpelNnaaqrsgkesvIStpnSkvkil
                   IGEN++evRe+vl+gle++G+++Dp+lNn   ++g+e++IS+p+S+vk++
       query   330 IGENSVEVRERVLRGLEFMGVYWDPALNN---VRGEEAFISYPHSPVKVM 376

                   vipTNEElmIaeDairLtk<-*
                   +ipT+EE+mIa+D++rL+k
       query   377 IIPTDEEVMIARDVVRLAK    395
//
```

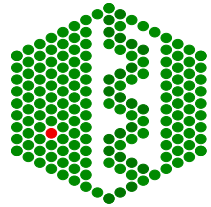
PIR-ALN

- http://www-nbrf.georgetown.edu/pirwww/search/textpiraln.html
- Database of annotated protein sequence alignments derived automatically from PIR PSD
- Includes alignments at superfamily (whole sequence), family (45% identity) and domain (in more than one superfamily) levels
- 3983 alignments, 1480 superfamilies, 371 domains
- Can search by protein accession number or text



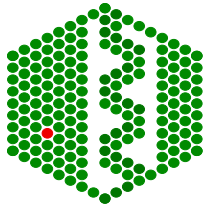
PROTOMAP

- http://www.protomap.cs.huji.ac.il
- Automatic classification of all SWISS-PROT proteins into groups of related proteins (also including TrEMBL now)
- Based on pairwise similarities
- Has hierarchical organisation for sub- and super-family distinctions
- 13 354 clusters, 5869 ≥ 2 proteins, 1403 ≥ 10
- Keeps SP annotation eg description, keywords
- Can search with a sequence -classify it into existing clusters



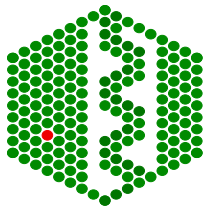
DOMO

- http://www.infobiogen.fr/srs6bin/cgi-bin/wgetz?-page+LibInfo+-lib+DOMO (SRS)
- Database of gapped multiple sequence alignments from SWISS-PROT and PIR
- Domain boundaries inferred automatically, rather than from 3D data
- Has 8877 alignments, 99058 domains, and repeats
- Each entry is one homologous domain, has annotation on related proteins, functional families, evolutionary tree etc



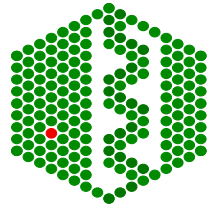
ProClass

- http://pir.georgetown.edu/gfserver/proclass.html
- Non-redundant protein database organized by family relationships defined by Prosite patterns and PIR superfamilies.
- Facilitates protein family information retrieval, domain and family relationships, and classifies multi-domain proteins
- Contains 155,868 sequence entries



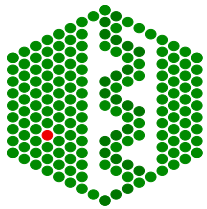
SBASE (Agricultural Biotechnology Centre)

- http://sbase.abc.hu/main.html
- Protein domain library from clustering of functional and structural domains
- SBASE entries - grouped by Standard names (SN groups) that designate various functional and structural domains of protein sequences- relies on good annotation of domains
- Detects subclasses too
- Can do similarity search with BLAST or PSI-BLAST



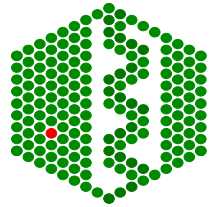
Integrating Pattern databases

- MetaFam
- IProClass
- CDD
- InterPro



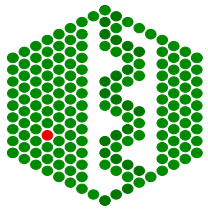
METAFAM

- http://metafam.ahc.umn.edu/
- Protein family classification built with Blocks+, DOMO, Pfam, PIR-ALN, PRINTS, Prosite, ProDom, SBASE, SYSTERS
- Automatically create supersets of overlapping families using set-theory to compare databases-reference domains covering total area
- Use non-redundant protein set from SPTR & PIR

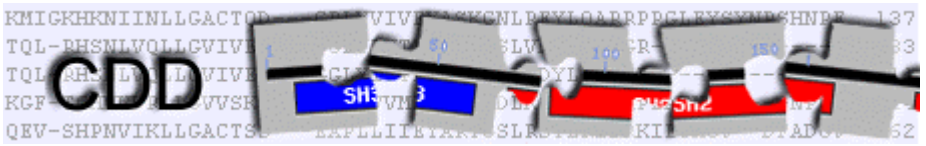


IProClass

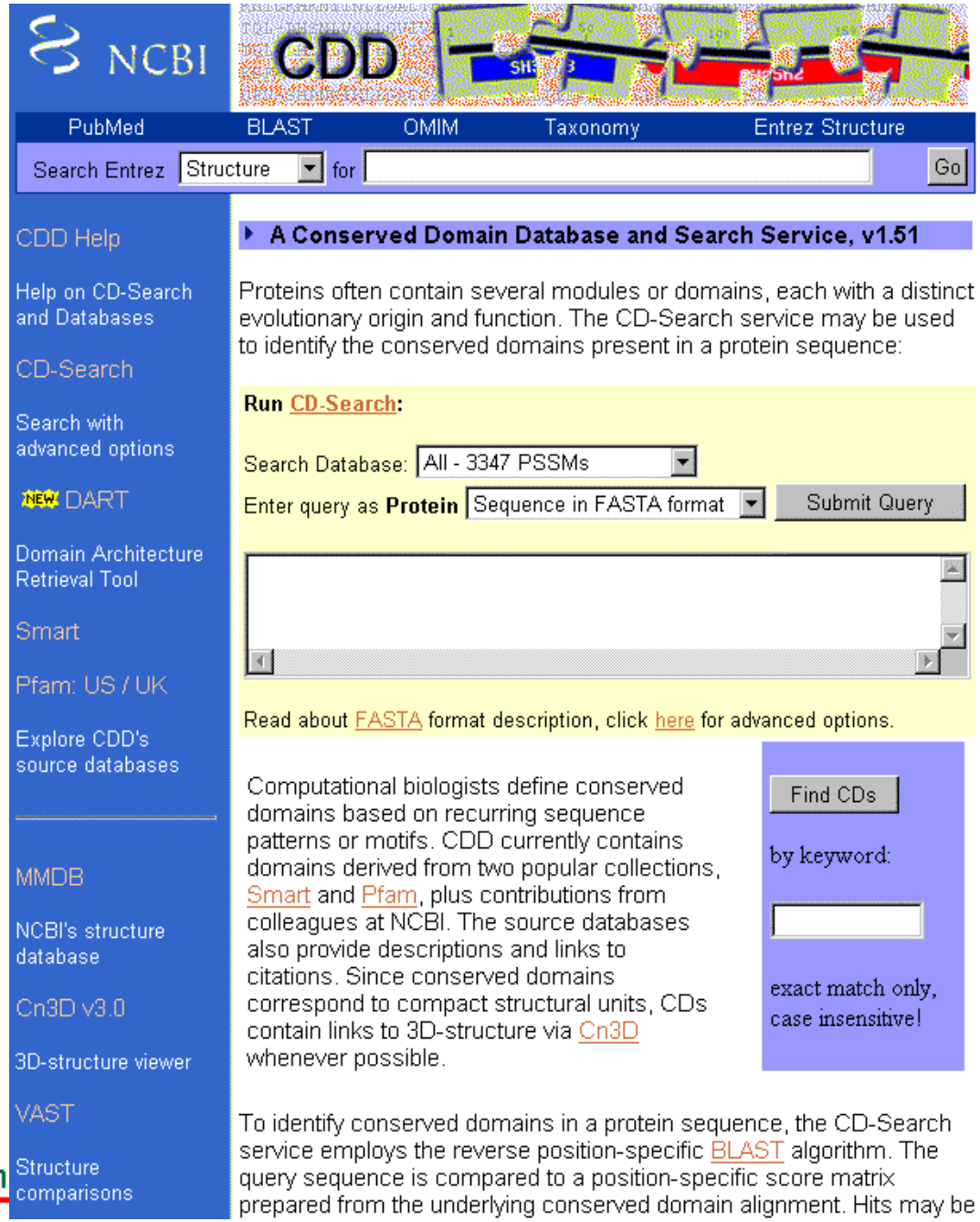
- http://pir.georgetown.edu/iproclass/
- Integrated database linking ProClass, PIR-ALN, Prosite, Pfam and Blocks
- Contains >20000 non-redundant SP & PIR proteins, 28000 superfamilies, 2600 domains, 1300 motifs, 280 PTMs
- Can be searched by text or sequence



CDD Conserved Domain Database



- http://www.ncbi.nlm.nih.gov:80/Structure/cdd/cdd.shtml
- Database of domains derived from SMART, Pfam and contributions from NCBI (LOAD)
- Uses reverse position-specific BLAST (matrix)
- Links to proteins in Entrez and 3D structure
- Stand-alone version of RPS-BLAST at: ftp://ncbi.nlm.nih.gov/toolbox



CDD homepage

EMBL European

CDD Search result

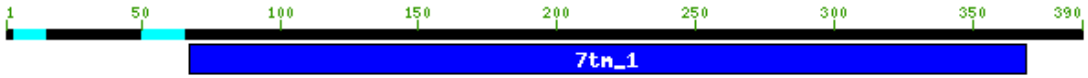
NCBI **CD-Search** Entrez ?

RPS-BLAST 2.2.1 [Apr-13-2001]

Query= local sequence:
(390 letters)

Database: oasis_sap.v1.51
3347 PSSMs; 659,460 total columns

Mouse-over boxes to display more information



NEW Show other proteins containing these domains

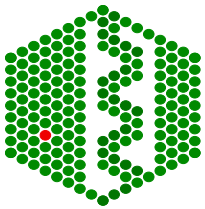
PSSMs producing significant alignments:		Score (bits)	E value
gnl|Pfam|pfam00001	7tm_1, 7 transmembrane receptor (rhodopsin family)	184	7e-48

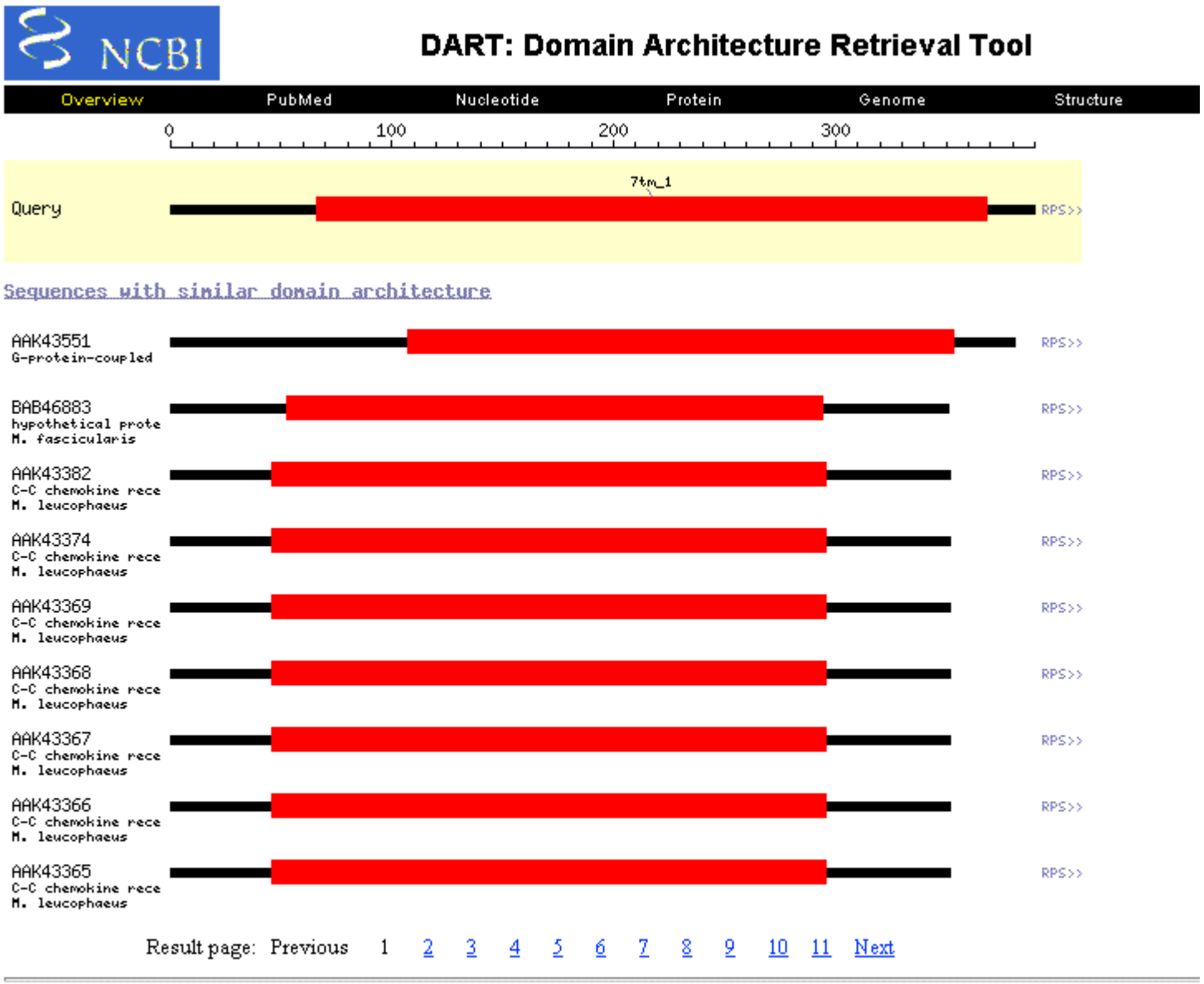
gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family).

Add query to multiple alignment, display up to 10 sequences most similar to the query

CD-Length = 254 residues, 99.6% aligned
Score = 184 bits (467), Expect = 7e-48

```
Query:  67   NAFVIATVYRTRKLHTPANYLIASLAVTDLLVSILVMPISTMYTVTGRWTLGQVVCDFWL  126
Sbjct:  2    NLLVILVILRTKKLRTPTNIFLLNLAVADLLFLLTLPPWALYYLVGGDWVFGDALCKLVG  61

Query:  127  SSDITCCTASILHLCVIALDRYWAITDAVEYSAKRTPKRAAVMIALVWVFSISISLPPFF  186
Sbjct:  62   ALFVVNGYASILLLTAISIDRYLAIVHPLRYRRIRTPRRAKVLILLVWVLALLLSLPPLL  121

Query:  187  --WRQAKAEEEVSECVVNTDHILYTVYSTVGAFYFPTLLLIALYGRIYVEARSRILKQTP  244
Sbjct:  122  FSWLRTVEEGNTTVCLIDFPEESVKRSYVLLSTLVGFVLPLLVILVCY----TRILR---  174

Query:  245  NRTGKRLTRAQLITDSPGSTSSVTSINSRVPDVPSESGSPVYVNQVKVRVSDALLEKKKL  304
Sbjct:  175  ----------------------------------------------TLRKRARSQRSLKR  188

Query:  305  MAARERKATKTLGIILGAFIVCWLPFFIISLVMPICKDACW-FHLAIFDFFTWLGYLNSL  363
Sbjct:  189  RSSSERKAAKMLLVVVVVFVLCWLPYHIVLLLDSLCLLSIWRVLPTALLITLWLAYVNSC  248
```




NCBI
DART: Domain Architecture Retrieval Tool
Overview
PubMed
Nucleotide
Protein
Genome
Structure
0
100
200
300
7tm_1
Query
RPS>>
Sequences with similar domain architecture
AAK43551
G-protein-coupled
BAB46883
hypothetical prote
M. fascicularis
AAK43382
C-C chemokine rece
M. leucophaeus
AAK43374
C-C chemokine rece
M. leucophaeus
AAK43369
C-C chemokine rece
M. leucophaeus
AAK43368
C-C chemokine rece
M. leucophaeus
AAK43367
C-C chemokine rece
M. leucophaeus
AAK43366
C-C chemokine rece
M. leucophaeus
AAK43365
C-C chemokine rece
M. leucophaeus
Result page: Previous 1 2 3 4 5 6 7 8 9 10 11 Next
Subset by selected domains:
pfam00001 7tm_1, 7 transmembrane receptor (rhodopsin family).

DART

NCBI **CD-Browser** Entrez ?

CD: ACTIN

Description: Actin; ACTIN subfamily of ACTIN/mreB/sugarkinase/Hsp70 superfamily

CD status: Full-length sequences, including 3D structure if known. Alignment from source, reindexed to representative

Source: Smart

Created: 24-Apr-2001

Taxonomy spanned: Eukaryota -> [3] from Alveolata
[1] from Diplomonadida
[20] from Fungi/Metazoa group
[1] from Mycetozoa
[2] from Viridiplantae

Aligned sequences: 28

Representative: Consensus sequence:

Aligned range: 1-374

Sequence:
```
TPAIVIDNGSGTIKAGFAGEDFPQVVFPSIVGRPKDGKGMVGDAKDTFVGDEAQEKRGGL
ELKYPIEHGIVENWDDMEKIWDYTFFNELKVEPEEHPVLLTEPPLNPKSNREKILEIMFE
EFNFPALYIAIQAVLSLYASGGRTTGLVIDSGDGVTHVVPVVDGYVLPHAIKRIDIAGRD
LTDYLKELLSERGYQFNSSAEFEIVREIKEKLCYVAEDFEKEMKKARESSESSKLTKTYE
LPDGNTIKVGNERFRIPEILFSPELIGLEQKGIHELVYESIQKCDIDVRKDLYENIVLSG
GSTLIPGFGERLEKELKRLAPKKLKVKVIAPPDRKYAVWLGGSILASLSTFEDMWISKKE
YEEHGSQIVERKCF
```

This CD alignment includes 3D structure. To display structure, download **Cn3D v3.0**!

View Alignment showing up to 10 top listed sequences Reset

Aligned chains / All chains / Virtual Bonds / All Atoms

Launch Cn3D / HTML Display / Text Display

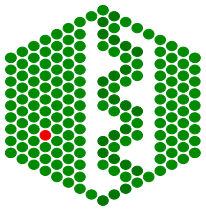
FASTA with gaps / Phylip format / See ASN.1 file

Conservation color threshold: 2.0 bits

Alignment width: 60

pick aligned sequences (will be added to selection above)			
	3D	**PDB-Id/gi**	**Definition**
☐	◉	1DGA A	Chain A, Structure Of Dictyostelium Discoideum Actin Complexed With Mg Atp And Human Gelsolin Segment 1. [CD]
☐		6322380	Saccharomyces cerevisiae chromosome X, complete chromosome sequence [CD]
☐		113295	ACTIN, CYTOPLASMIC [CD]
☐		1168336	ACTIN II (CENTRACTIN-LIKE PROTEIN) [CD]
☐		1703155	ACTIN [CD]

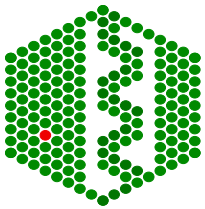
CDD example entry



1: NP_012454 **54.8 kDa actin-related protein; Arp4p [Saccharomyces cerevisiae]**

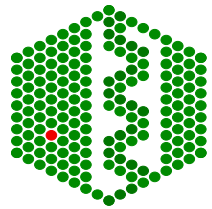
```
LOCUS       NP_012454     489 aa                    PLN       30-APR-2001
DEFINITION  54.8 kDa actin-related protein; Arp4p [Saccharomyces cerevisiae].
ACCESSION   NP_012454
PID         g6322380
VERSION     NP_012454.1  GI:6322380
DBSOURCE    REFSEQ: accession NC_001142.1
KEYWORDS    .
SOURCE      baker's yeast.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
            Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (residues 1 to 489)
  AUTHORS   Galibert,F., Alexandraki,D., Baur,A., Boles,E., Chalwatzis,N.,
            Chuat,J.C., Coster,F., Cziepluch,C., De Haan,M., Domdey,H.,
            Durand,P., Entian,K.D., Gatius,M., Goffeau,A., Grivell,L.A.,
            Hennemann,A., Herbert,C.J., Heumann,K., Hilger,F., Hollenberg,C.P.,
            Huang,M.E., Jacq,C., Jauniaux,J.C., Katsoulou,C.,
            Karpfinger-Hartl,L. et al.
  TITLE     Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
            X
  JOURNAL   EMBO J. 15 (9), 2031-2049 (1996)
  MEDLINE   96208490
REFERENCE   2  (residues 1 to 489)
  AUTHORS   Goffeau,A., Barrell,B.G., Bussey,H., Davis,R.W., Dujon,B.,
            Feldmann,H., Galibert,F., Hoheisel,J.D., Jacq,C., Johnston,M.,
            Louis,E.J., Mewes,H.W., Murakami,Y., Philippsen,P., Tettelin,H. and
            Oliver,S.G.
  TITLE     Life with 6000 genes
  JOURNAL   Science 274 (5287), 546 (1996)
  MEDLINE   97002444
REFERENCE   3  (residues 1 to 489)
  AUTHORS   Saccharomyces Genome Database (yeast-curator@genome.stanford.edu).
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-1999) Department of Genetics, Stanford
            University, Saccharomyces Genome Database, Stanford, CA 94305-5120,
            USA
COMMENT     REFSEQ: This reference sequence was provided by the Saccharomyces
            Genome Database (SGD).
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..489
                     /organism="Saccharomyces cerevisiae"
                     /strain="S288C"
                     /db_xref="taxon:4932"
                     /chromosome="X"
     Protein         1..489
                     /product="54.8 kDa actin-related protein"
                     /note="Arp4p"
     CDS             1..489
```

PIR link from CDD



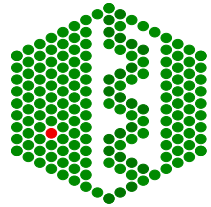
INTERPRO

- http://www.ebi.ac.uk/interpro
- Integration of different signature recognition methods (PROSITE, PRINTS, PFAM, ProDom and SMART)



InterPro release 3

- Built from PROSITE, PRINTS, Pfam, ProDom, SMART, SWISS-PROT and TrEMBL
- Contains 3915 entries encoded by 7714 different regular expressions, profiles, fingerprints, Hidden Markov Models and ProDom domains
- InterPro provides >1 million InterPro matches hits against 532403 SWISS-PROT + TrEMBL protein sequences (68% coverage)
- Direct access to the underlying Oracle database
- A XML flatfile is available at ftp://ftp.ebi.ac.uk/pub/databases/interpro/
- SRS implementation
- Text- and sequence-based searches



InterPro Home
Text Search
Databases
Documentation
FTP Site
Sequence Search

InterPro Home

Upcoming events, new documents.

- InterPro Workshop Announcement
- InterPro Paper in NAR (Adobe PDF format)
- List of InterPro to GO mappings

InterPro release 3.1 (May 2001) was built from Pfam 6.0, PRINTS 30.0, PROSITE 16.35, ProDom 2001.1, SMART 3.1 and the current SWISS-PROT + TrEMBL data. This release of InterPro contains 3915 entries, representing 991 domains, 2845 families, 64 repeats and 15 post-translational modification sites.

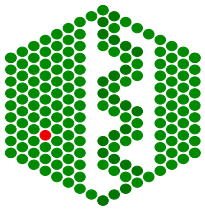
InterPro is a useful resource for whole genome analysis and has already been used for the proteome analysis of a number of completely sequenced organisms. A *preliminary* proteome analysis was also produced for the human genome. Please refer to the proteome analysis pages.

Further information on InterPro can be found in the Documentation page, which includes links to the release notes, the user manual, a list of deleted InterPro entries, the dataflow scheme of the database, a fully annotated sample entry and references for the member databases. For publications please cite:

R.Apweiler, T.K.Attwood, A.Bairoch, A.Bateman, E.Birney, M.Biswas, P.Bucher, L.Cerutti, F.Corpet, M.D.R.Croning, R.Durbin, L.Falquet, W.Fleischmann, J.Gouzy, H.Hermjakob, N.Hulo, I.Jonassen, D.Kahn, A.Kanapin, Y.Karavidopoulou, R.Lopez, B.Marx, N.J.Mulder, T.M.Oinn, M.Pagni, F.Servant, C.J.A.Sigrist, E.M.Zdobnov.
The InterPro database, an integrated documentation resource for protein families, domains and functional sites, Nucleic Acids Research vol 29(1):37-40.

For information, comments and / or suggestions on the InterPro database, please email us at interhelp@ebi.ac.uk

EMBL
European



InterPro

- InterPro Home
- **Text Search**
- Databases
- Documentation
- FTP Site
- Sequence Search

InterPro Text Searches

This page allows you to search the InterPro database by keyword. If you are looking for somewhere to locate patterns in your protein sequence, use the 'Sequence Search' link to the left.

Oracle Search - Simple, but Fast!

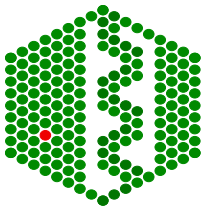
You can use this search box to search for InterPro, Pfam, PRINTS, Prosite, ProDom, SWISS-PROT, TrEMBL accession numbers and names, database names, and entry_types. You may combine more than one search term with 'AND', '&', 'OR', '|', 'NOT' and '!'; you may also use wildcarded expressions (eg. bar*).

Enter search terms here :

human transporter [Search]

SRS Searches

The SRS search system allows more complex queries, but will take longer to return results to you. Follow this link to go to the main SRS page.



InterPro

InterPro Simple Search

You can use this page to search for InterPro, Pfam, PRINTS, Prosite, SWISS-PROT, TrEMBL accession numbers and names, database names, and entry_types. You may combine more than one search term with 'AND', '&', 'OR', '|', 'NOT' and '!'; you may also use wildcarded expressions (eg. bar*).

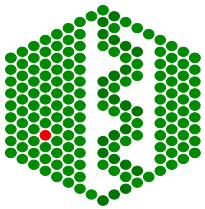
Enter search terms here...

human transporter [Search]

Search results for 'human transporter'

Click on the links below to jump to individual InterPro entries.

Entry	Entry name
IPR000076	K-Cl co-transporter
IPR000622	K-Cl Co-transporter type 1 (KCC1)
IPR000803	Facilitated glucose transporter family
IPR000849	GlpT family of transporters
IPR001066	Sugar transporter
IPR001204	Phosphate transporter family
IPR001902	Sulfate transporter
IPR002259	Delayed-early response protein/equilibrative nucleoside transporter
IPR002293	Permease for amino acids and related compounds, family I
IPR002435	Noradrenaline neurotransmitter transporter
IPR002436	Dopamine neurotransmitter transporter
IPR002437	Serotonin (5-HT) neurotransmitter transporter

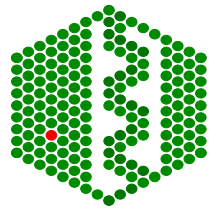


InterPro

InterPro Entry IPR000803

Facilitated glucose transporter family

Database	InterPro
Accession	IPR000803; Gluc_transporter (matches 65 proteins)
Name	Facilitated glucose transporter family
Type	Family
Dates	08-OCT-1999 (created) 02-MAR-2000 (last modified)
Signatures	PR00172; GLUCTRNSPORT (91 proteins)
Parent [tree]	IPR003663; Sugar transporters (250 proteins)
Children [tree]	IPR002439; Glucose transporter type 1 (GLUT1) (8 proteins) IPR002440; Glucose transporter type 2 (GLUT2) (5 proteins) IPR002441; Glucose transporter type 4 (GLUT4) (9 proteins) IPR002442; Fructose transporter (GLUT5) (5 proteins) IPR002945; Glucose transporter type 3 (GLUT3) (7 proteins)
Process	transport (GO:0006810)
Function	glucose transporter (GO:0005355)
Component	membrane fraction (GO:0005624)
Abstract	The ability to transport glucose across the plasma membrane is a feature common to nearly all cells, from simple bacteria through to highly specialised mammalian neurones. Facilitative glucose (and fructose) transport is mediated by members of the GLUT transporter family. These are glycosylated transmembrane (TM) proteins that transport glucose in a passive (i.e., energy-independent) manner. In consequence, they can only transport glucose down its concentration gradient. Currently, five such mammalian transporters have been cloned and functionally characterised [1, 2, 3]. Four of these transport glucose (GLUT1-4), whereas GLUT5 preferentially transports fructose. A sixth cDNA, encoding an apparent glucose transporter, was cloned but was found to be a pseudo-gene (GLUT6) [4]. Similarly, another cDNA thought to encode a glucose transporter that was targeted to the endoplasmic reticulum was eventually realised to be an experimental cloning artefact (GLUT7) [5]. The five confirmed isoforms are expressed in a tissue and cell-specific manner, and have been found to exhibit distinct kinetic and regulatory properties, presumably reflecting their specific functional roles in these locations. Hydropathy analysis reveals they have 12 presumed TM domains, and that they belong to a much larger 'major facilitator superfamily' of 12 TM transporters that are involved in the transport of a variety

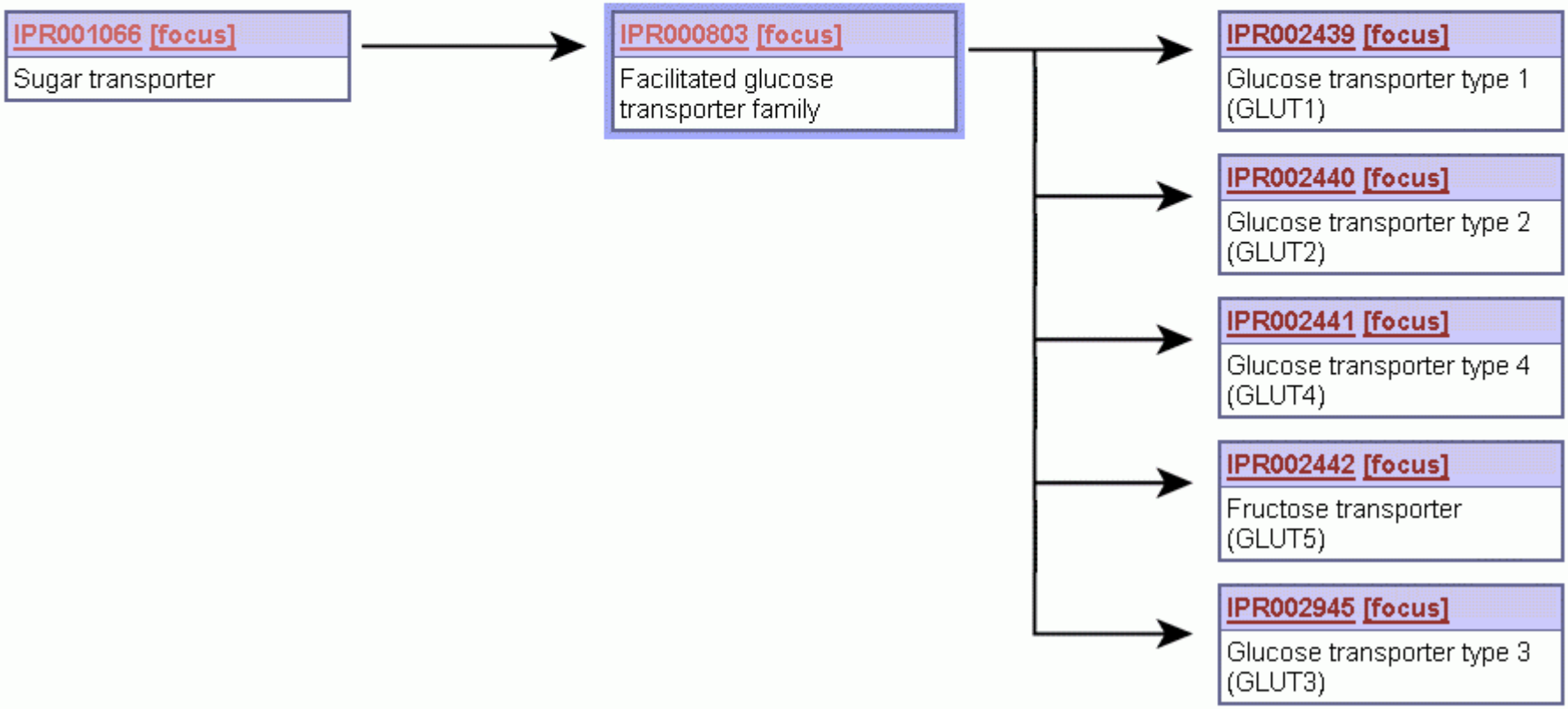


InterPro

Tree display for IPR000803

The tree below shows the selected InterPro entry, the path to the root of the tree, the immediate children and the immediate children of the selected entry's parent (i.e. the entry's siblings).

To return to the full entry for this accession number, click here.



The five confirmed isoforms are expressed in a tissue and cell-specific manner, and have been found to exhibit distinct kinetic and regulatory properties, presumably reflecting their specific functional roles in these locations. Hydropathy analysis reveals they have 12 presumed TM domains, and that they belong to a much larger 'major facilitator superfamily' of 12 TM transporters that are involved in the transport of a variety of hexoses and other carbon compounds, including: bacterial sugar-proton symporters (H^+/xylose and H^+/arabinose); bacterial transporters of carboxylic acids and antibiotics; and sugar transporters in various yeast, protozoa and higher plants. Nevertheless, amino acid identity within the superfamily may be as low as ~25% [6, 7]. Besides the 12 presumed TM domains, the most characteristic structural feature of the superfamily is the presence of a five residue motif (RXGRR, where X is any amino acid). In the GLUT transporters, this motif is present in the presumed cytoplasmic loops connecting TM domains 2 with 3, and also 8 with 9. The 12 TM transporter superfamily appears to be structurally unrelated to the Na^+-coupled, Na^+/glucose co-transporters (SGLT1-3) found in the intestine and kidney, which are able to transport glucose against its concentration gradient [8].

Comparison of the hydropathy profiles for GLUT1-5 reveals that they are virtually superimposable, despite the fact that their primary structures may differ by up to 60%. Of the presumed TM domains, the fourth, fifth and sixth are the most highly conserved, and conserved residues are also found in the short exofacial loops joining the putative TM regions. The presumed cytoplasmic N- and C-termini, and the extracellular loop between the first and second TM domains, show the greatest divergence, both in terms of primary structure and size.

Examples

- P46896 GTR1_CHICK
- Q01440 GTR1_LEIDO

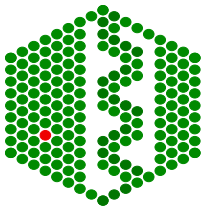
View examples

References

1. Gould G.W., Bell G.I.
   Facilitative glucose transporters: an expanding family.
   Trends Biochem. Sci. 15: 18-23(1990). [MEDLINE:90194363] [PUB00005353]
2. Bell G.I., Burant C.F., Takeda J., Gould G.W.
   Structure and function of mammalian facilitative sugar transporters.
   J. Biol. Chem. 268(26): 19161-19164(1993). [MEDLINE:93374885] [PUB00006044]
3. Mueckler M., Caruso C., Baldwin S.A., Panico M., Blench I., Morris H.R., Allard W.J., Liender G.E., Lodish H.F.
   Sequence and structure of a human glucose transporter.
   Science 229: 941-945(1985). [MEDLINE:85272595] [PUB00005096]
4. Kayano T., Fukumoto H., Eddy R.C., Fan Y., Byers M.G., Shows T.B., Bell G.I.
   Evidence for a family of human glucose transporter-like proteins.
   J. Biol. Chem. 263: 15245-15248(1988). [MEDLINE:89008414] [PUB00002464]
5. Burchell A.
   A re-evaluation of GLUT 7.
   Biochem. J. 331: 973(1998). [MEDLINE:99004677] [PUB00006047]
6. Maiden M.C.J., Davis E.O., Baldwin S.A., Moore D.C.M., Henderson P.J.F.
   Mammalian and bacterial sugar transport proteins are homologous.
   Nature 325: 641-643(1987). [MEDLINE:87115869] [PUB00003999]
7. Marger M.D., Saier M.H.Jr.
   A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport.
   Trends Biochem. Sci. 18: 13-20(1993). [MEDLINE:93174460] [PUB00005398]
8. Hediger M.A., Coady M.J., Ikeda T.S., Wright E.M.
   Expression cloning and cDNA sequencing of the Na+/glucose co-transporter.
   Nature 330(6146): 379-381(1987). [MEDLINE:88065856] [PUB00006048]

Matches

Table all Graphical all Condensed graphical view



InterPro

InterPro - Proteins matching IPR000803

Table **Graphical**

Grid shows 10aa intervals, first mark at position 0. Move the mouse over a match to see more information in the status line of your browser window.

Item 21-40 of 91 < 1 **2** 3 4 5 >

Protein	Match Display
SWISS-PROT GTR2_HUMAN P11168	IPR000803 PR00172 — GLUCTRNSPORT IPR001066 PS00216 — SUGAR_TRANSPORT IPR001066 PS00217 — SUGAR_TRANSPORT IPR001066 PR00171 — SUGRTRNSPORT IPR001066 PF00083 — sugar_tr IPR002440 PR01191 — GLUCTRSPORT2
SWISS-PROT GTR3_HUMAN P11169	IPR000803 PR00172 — GLUCTRNSPORT IPR001066 PS00216 — SUGAR_TRANSPORT_1 IPR001066 PS00217 — SUGAR_TRANSPORT_2 IPR001066 PR00171 — SUGRTRNSPORT IPR001066 PF00083 — sugar_tr

Help for : graphic key - Netscape

Graphical match display legend

The table below shows the colour coding used in the graphical match display. The extent of the bars denotes the region on the protein sequence that the selected method matches.

	True	Unknown
PRINTS		
PROSITE *pattern*		
PROSITE *profile*		
PFAM		
ProDom		n/a

See also :

InterPro

InterPro - Proteins matching IPR001066

Table Graphical

Item 401-420 of 1177 < Previous **21** 22 23 24 25 Next >

	PS00216	PS00217	PR00171	PF00083
P39637 YWFA_BACSU				19-406 T
P39843 BMR2_BACSU	65-81 T			17-398 T
P39850 CAPA_STAAU		175-200 F		
P39924 HXTC_YEAST	370-387 T	169-194 T	68-78 T 164-183 T 328-338 T 423-444 T 446-458 T	60-521 T
P39932 STL1_YEAST	347-364 T	N		30-488 T
P40441 YIR0_YEAST	263-280 T	62-87 T		2-416 T
P40474 YIM1_YEAST	117-133 F			61-539 T
P40475 YIM0_YEAST	125-141 F			71-547 T
P40862 PROP_SALTY	P	P		
P40885 HXT9_YEAST	373-390 T	172-197 T	72-82 T 167-186 T 331-341 T	64-526 T

Help for : table legend - Netscape

Tabular match display legend

The single letter codes after the amino acid ranges in this table denote the status of each individual match. Possible values are shown in the table below :

T True
F False Positive
N False Negative
P Partial
? Unknown

InterPro

Release 3.1

Proteins belonging to InterPro entry
IPR003662(IPR000803)

To view the complete output click here

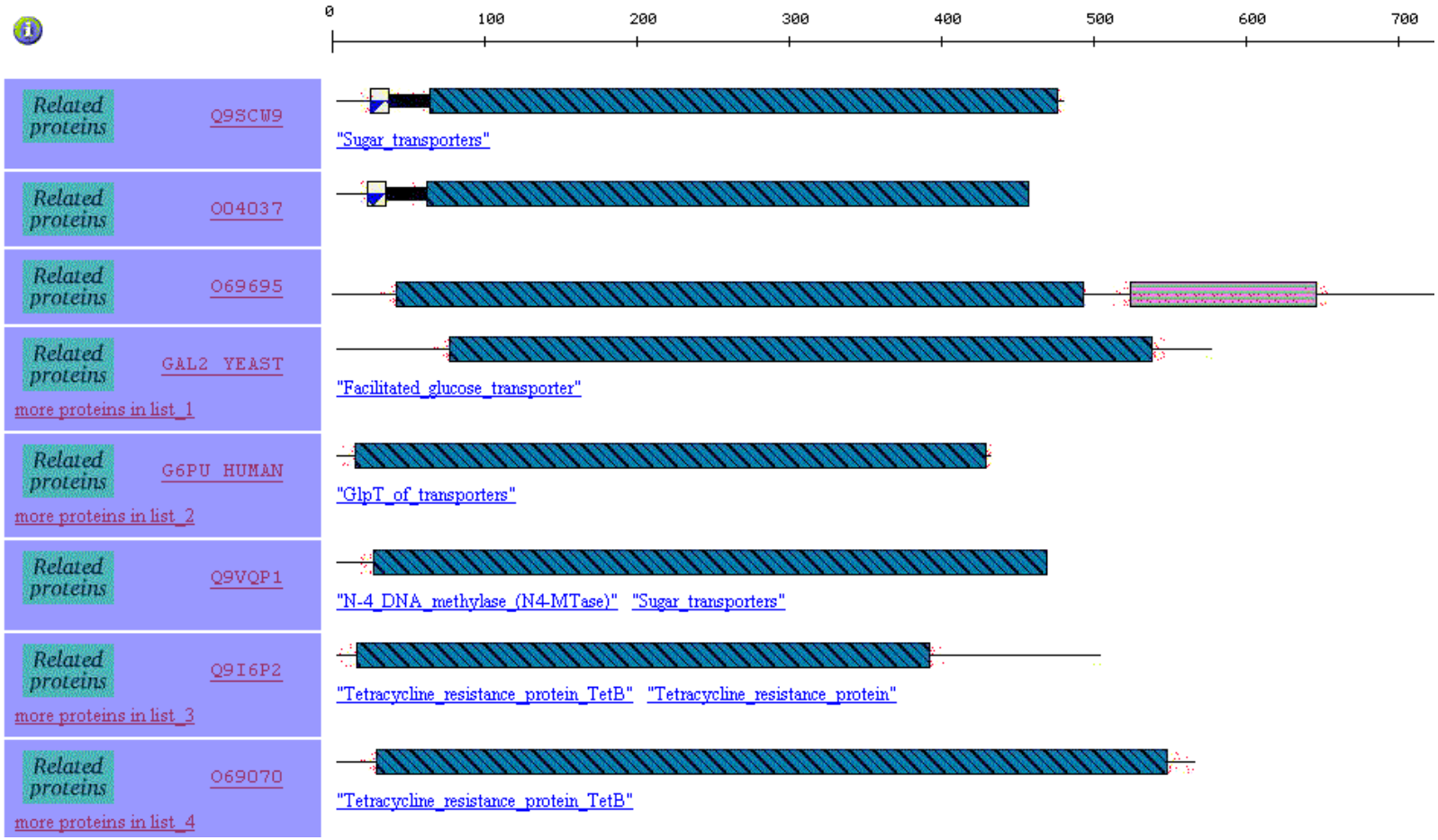
EMBL
European Bio

InterPro

Release 3.1

Proteins sharing an InterPro "Domain" with GTR1_BOVIN (P27674)

To view the complete output click here



InterPro

- InterPro Home
- Text Search
- Databases
- Documentation
- FTP Site
- **Sequence Search**

InterProScan

This form allows you to query your protein sequence against InterPro. If you wish to use this facility during a course, or if you have any problems or suggestions, then please contact at support@ebi.ac.uk.

1. Protein Sequence

Please either enter (or cut and paste) your protein sequence into the text box below, or, if you have the sequence in a file on your computer, click the 'Browse' button to upload it directly (you will be given a file selection window if you choose this option). If you need help on sequence formats, this page details various common formats.

Enter or cut and paste protein sequence here.

...or upload sequence from a local file [Browse...]

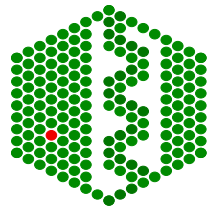
2. Query Mode

You can either wait for the search results to be returned in the web browser window, or choose to have them sent to your email address on completion. The latter may be useful, as some searches will take a considerable time to complete.

EMB
Euro

InterProScan

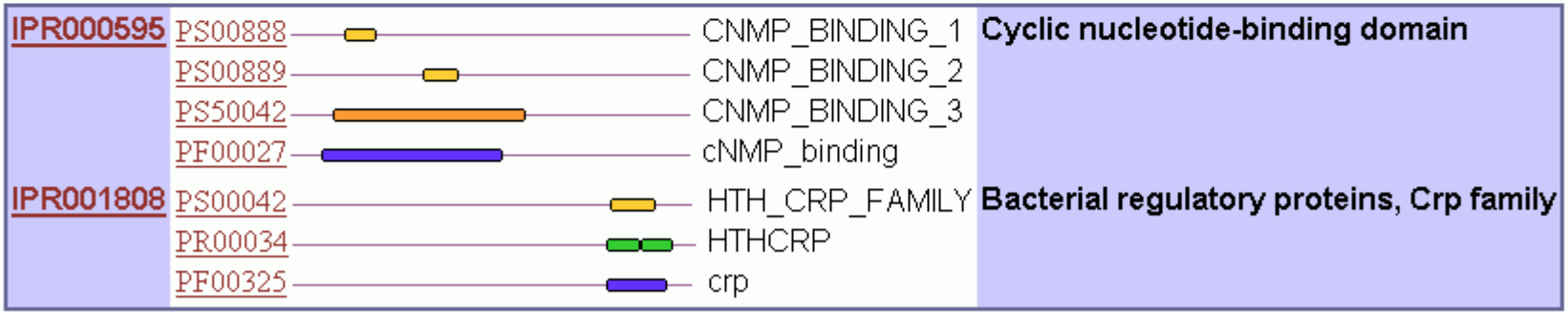
- PROSITE patterns: ppsearch
- PROSITE profiles: pfscan
- PFAM HMMs: hmmpfam
- PRINTS fingerprints: fpscan
- ProDom
- SMART
- eMotif derived PROSITE pattern
- TMHMM
- SignalP



InterPro search Results.

1 Query Sequence submitted Length 210 aa.

InterPro	Results of PPsearch against PROSITE	Results of PFScan against PROSITE	Results of FingerPRINTScan against PRINTS	Results of HMMDecypher against PFAM-A
IPR000595 Cyclic nucleotide-binding domain	PS00888 [30-46] PS00889 [71-89]	PS50042 [24-124]		PF00027 [18-112]
IPR001808 Bacterial regulatory proteins, Crp family	PS00042 [168-191]		PR00034 [166-183] [182-198]	PF00325 [166-197]



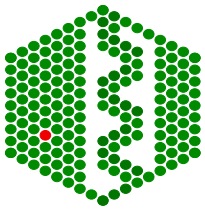
XML / TXT formatted.

PRINTS detailed results

ANX3_MOUSE

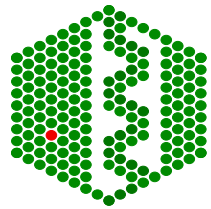
Annexin type III

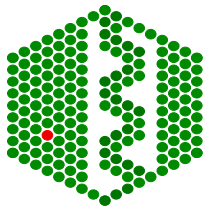
```
1TBS
1TBH ANNEXINIII      4.257972e-88   Annexin type III signature
1TBH ANNEXIN         1.017764e-65   Annexin family signature
1TBH ANNEXINV        8.020715e-56   Annexin type V signature
1TBH ANNEXINIV       4.202039e-55   Annexin type IV signature
1TBH ANNEXINVI       2.077635e-46   Annexin type VI signature
1TBH ANNEXINI        3.704047e-46   Annexin type I signature
1TBH ANNEXINII       1.263206e-42   Annexin type II signature
1TBF
2TBS
2TBT FingerPrint     No.Motifs SumId    AveId    ProfScore  Ppvalue     Evalue      GraphScan
2TBH ANNEXINIII      8  of  8  7.4e+02  92       7153       5.3e-93     4.3e-88     IIIIIIII
2TBH ANNEXIN         6  of  7  362.54   60.42    3781       1.3e-70     1e-65       IIIII.I
2TBH ANNEXINV        6  of  8  358.81   59.80    4121       1e-60       8e-56       I.III.II
2TBH ANNEXINIV       7  of  8  415.08   59.30    4449       5.3e-60     4.2e-55     II.IIIII
2TBH ANNEXINVI       7  of  8  375.22   53.60    3923       2.6e-51     2.1e-46     IIIII.II
2TBH ANNEXINI        6  of  8  333.30   55.55    3478       4.6e-51     3.7e-46     .IIIII.I
2TBH ANNEXINII       6  of  8  345.13   57.52    3496       1.6e-47     1.3e-42     .IIII.II
2TBN HEATSHOCK70     2  of  9  61.81    30.91    425        0.00013     10          ...iI....
2TBN NAHEXCHNGR3     2  of  16 68.16    34.08    432        0.00037     29          ........i.I.....
2TBN CCYSTOKNINAR    2  of  7  60.29    30.15    346        0.0013      1e+02       ..i.I..
2TBF
3TBS
3TBT MotifName       No.Mots   IdScore PfScore Pvalue    Sequence
3TBH ANNEXINIII      1  of  8  83.33   512     2.59e-06  WVGPRGTIKD
3TBH ANNEXINIII      2  of  8  92.75   1069    2.22e-16  RGLGTDEKTLINILTERSNAQRQ
3TBH ANNEXINIII      3  of  8  96.08   856     2.72e-12  LKGDLSGHFEHVMVALV
3TBH ANNEXINIII      4  of  8  92.42   992     4.44e-16  LKKSMKGTGTDEDALIEILTTR
3TBH ANNEXINIII      5  of  8  90.48   346     5.76e-04  YTVYKKS
3TBH ANNEXINIII      6  of  8  92.59   1436    1.00e-16  LYNAGENKWGTDEDKFTEVLCLRSFPQ
3TBH ANNEXINIII      7  of  8  98.72   1253    1.00e-16  KGAGTDEFTLNRIMVSRSEIDLLDIR
3TBH ANNEXINIII      8  of  8  92.86   689     1.33e-09  DTSGDYRTVLLKIC
3TBB
3TBH ANNEXIN         1  of  7  60.39   719     3.47e-14  RGLGTDEKTLINILTERSNAQRQ
3TBH ANNEXIN         2  of  7  51.85   447     1.86e-09  LKGDLSGHFEHVMVALV
3TBH ANNEXIN         3  of  7  65.40   716     6.27e-14  LKKSMKGTGTDEDALIEILTTR
3TBH ANNEXIN         4  of  7  53.16   720     1.33e-15  LYNAGENKWGTDEDKFTEVLCLRSFPQ
3TBH ANNEXIN         5  of  7  69.31   683     8.14e-13  LKGAGTDEFTLNRIMVSRSEI
```



SUMMARY

- Many different protein signature databases from small patterns to alignments to complex HMMs
- Have different strengths and weaknesses
- Have different database formats
- **Therefore:** best to combine methods, preferably in a database with them already merged for simple analysis with consistent format





Protein Secondary Structure

- **CATH** (Class, Architecture,Topology, Homology) http://www.biochem.ucl.ac.uk/dbbrowser/cath/
- **SCOP** (structural classification of proteins) -hierarchical database of protein folds http://scop.mrc-lmb.cam.ac.uk/scop
- **FSSP** Fold classification using structure-structure alignment of proteins http://www2.ebi.ac.uk/fssp/fssp.html
- **TOPS** Cartoon representation of topology showing helices and strands http://tops.ebi.ac.uk/tops/