PROSITE
Database of protein families and domains
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 18.34, of 19-Aug-2004 (contains 1277 documentation entries that describe 1736 different patterns, rules and profiles/matrices).

Access to PROSITE

Quick Search

in PROSITE by AC, ID or documentation text
Prefix and append wildcard '*' to words.

Browse PROSITE documentation entries
Search by author
Search by citation
Search by description
Search by full text search
SRS - Sequence Retrieval System
Download by FTP

Tools for PROSITE

Scan PROSITE patterns, profiles and rules with a Swiss-Prot/TrEMBL AC, ID or paste your own sequence in the box below (for more options, use the ScanProsite form):

ScanProsite - Scan a sequence against PROSITE or a pattern against Swiss-Prot or PDB and visualize matches on structures with graphical view and feature detection
ScanProsit

The ScanProsit tool [Help] allows to scan protein sequence(s) (either from Swiss-Prot or TrEMBL, or provided by the user) for the occurrence of patterns, profiles and rules (motifs) stored in the PROSITE database, or to search protein database(s) for hits by specific motif(s) [Reference / Download ps_scan, the standalone version]. The program PRATT can be used to generate your own patterns. You may either:

- Enter one or more PROSITE accession numbers and/or patterns [1 by line] to search the Swiss-Prot/TrEMBL and/or PDB databases, OR
- Enter one or more sequences [raw, Swiss_Prot or fasta format] and/or Swiss-Prot/TrEMBL accession numbers [1 by line] to be scanned with all patterns, profiles, rules in PROSITE, OR
- Fill in both fields to find all occurrences of a motif in a sequence.

### Protein(s) to be scanned:

Enter one or more Swiss-Prot/TrEMBL accession number(s) [AC] (e.g. P00747) and/or sequence identifier(s) [ID] (e.g. ENTK_HUMAN), and/or PDB identifier, and/or paste your own protein sequence(s) in the box below.

(leave this box blank to scan PROSITE entry(s) against selected protein databases)

### PROSITE pattern(s)/profile(s) to scan for:

Enter one or more PROSITE accession number(s) (e.g. PSS0240), and/or identifier(s) (e.g. CHEB), and/or type your pattern(s) in PROSITE format in the box below.

(leave this box blank to scan sequence(s) against the entire PROSITE database)
Pfam is a large collection of multiple sequence alignments and hidden Markov models covering many common protein domains and families. For each family in Pfam you can:

- Look at multiple alignments
- View protein domain architectures
- Examine species distribution
- Follow links to other databases
- View known protein structures

For more information on Pfam, on using this site, or on the changes between Pfam releases 14 and 15, click [here](http://www.sanger.ac.uk/Software/Pfam/).

Pfam can be used to view the domain organisation of proteins. A typical example is shown below. Notice that a single protein can belong to several Pfam families.

![Pfam protein domain example](http://www.sanger.ac.uk/Software/Pfam/)
Pfam is a database of protein families. It contains alignments and hidden Markov models for these families. For each family in Pfam you can:

1. **Protein name or sequence**
2. **Keyword**
3. **Domain query**
4. **DNA sequence**
5. **Taxonomy query**
6. Examine species distribution
7. Follow links to other databases
8. View known protein structures

For more information on Pfam, on using this site, or on the changes between Pfam releases 14 and 15, click here.

Pfam can be used to view the domain organisation of proteins. A typical example is shown below. Notice that a single protein can belong to several Pfam families.

**Voltage CLC**

[887 residues]

---

**Version 15.0**

- August 2004, 7503 families

**Pie Chart**

- Sequence coverage Pfam-A: 74%
- Sequence coverage Pfam-B: 24%
- Other

**Enter your keyword(s) here**
Pfam has pre-calculated the domain structure of the proteins in UniProt. If you know the name or accession number (e.g., P131_BOVIN or O91437) then you can see the Pfam domains on the sequence instantaneously.

**By Protein sequence**

**Single sequence searches**

If you don’t know the UniProt identifier for your sequence, you can perform a slower, HMM search by giving your sequence below.

Cut and Paste your sequence here. (This search will take 1-5 minutes)

**Pfam Search Options**

- **Search type:**
  - Both Global & Fragment Pfam search

- **Output format:**
  - Graphical output

- **E-value cutoff level:**
  - 1.0

**Or: Select the sequence file you wish to use**

For help on the scores in Pfam, and the difference between standard and fragment searches, click **here**
Pfam :: Home
The Pfam database of protein families and HMMs

Pfam 14.0 (June 2004, 7459 families)
Pfam is a large collection of multiple sequence alignments and hidden Markov models covering many common protein families. Pfam version 14.0 (June 2004) contains alignments and models for 7459 protein families, based on the Swissprot 43.2 and SP-TrEMBL 26.2 protein sequence databases.

HELP
More information on Pfam, using this site, and the changes between Pfam releases.

PROTEIN SEARCH
Analyze a protein query sequence to find Pfam family matches.

DNA SEARCH
Analyze a DNA query sequence to find Pfam family matches. (Uses the GeneWise server at the Sanger Centre.)

BROWSE PFAM
View Pfam annotation and alignments.

KEYWORD SEARCH
Query Pfam by keywords.

TAXONOMY SEARCH
Find Pfam families by taxonomy.

BROWSE SWISSPAM
View the domain organization of a SWISSPROT/TrEMBL sequence according to Pfam.
Pfam :: Protein Search
Analyze a query sequence using the Pfam HMM database

Analyze a query sequence by searching Pfam HMMs

Cut and paste your sequence here.
FASTA format or raw sequence are acceptable:

Or

Select the query sequence file you wish to use:

More advanced options
Molecular Visualization Freeware

Protein Explorer, Chime & RasMol

This is the RasMol Home Page visited by over 500,000 people from over 115 countries!

Contents  What's New?  Search  Job Opportunities!

Protein Explorer, a RasMol-derivative, is the easiest-to-use and most powerful software for looking at macromolecular structure and its relation to function. And it's free! It runs on Windows or Macintosh/PPC computers. (Linux users see below.) RasMol users will find its menus very familiar, and it understands RasMol commands. It is very fast: rotating a protein or DNA molecule shows its 3D structure. If you have never seen this, watch the image at the upper right of this page. (Click here to see another molecule rotate.) Look at our gallery to see still snapshots of other molecules. Also available here are Chime-based tutorials on

- DNA,
- Hemoglobin,
- Antibody,
- The Protein Morpher,
- Infrared spectra with animated molecular vibrations,
- Tutorials on many other popular molecules.

The above resources employ the Netscape plug-in Chime, freeware from MDL, and derived from RasMol. Here are reference materials and templates about how to create your own Chime websites.

Oxy-hemoglobin zooming in to oxy-heme (from 1hho.pdb by B. Shaban). This is an animated picture; unlike with Protein Explorer, you cannot move it with your mouse.
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 including preliminary analyses of the mouse and human genomes.

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.

InterPro is headed by Rolf Apweiler.