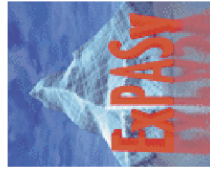


# A Quick Guide To The ExPASy proteomics server

<http://www.expasy.org>



This is an introduction to the ExPASy server, which has been serving protein and proteomics databases and analysis tools to the scientific community since 1993. It is maintained by a pluridisciplinary team at the Swiss Institute of Bioinformatics (<http://www.isb-sib.ch>).

## Databases

ExPASy is home to a number of databases specializing in different aspects of protein molecular biology and proteomics.

*Swiss-Prot* and *TrEMBL* - the *UniProt knowledgebase* (<http://www.expasy.org/sprot>)

The Swiss-Prot knowledgebase is a curated protein sequence database which strives to provide high-quality annotation, a minimal level of redundancy and a high level of integration with other databases. Swiss-Prot is supplemented by TrEMBL, which contains computer-annotated entries for all sequences not yet integrated in Swiss-Prot. Swiss-Prot and TrEMBL are maintained collaboratively by the SIB and the European Bioinformatics Institute (<http://www.ebi.ac.uk>). Note: The Swiss-Prot, TrEMBL and PIR protein database activities have united to form the Universal Protein Knowledgebase (UniProt) consortium.

*PROSITE* (<http://www.expasy.org/prosite>)

The PROSITE database of protein domains and families contains biologically significant sites, patterns and profiles that help to reliably identify to which known protein family a new sequence belongs.

*SWISS-2DPAGE* (<http://www.expasy.org/ch2d>)

SWISS-2DPAGE is a database of proteins identified on two-dimensional polyacrylamide gel electrophoresis (2D PAGE), and contains data from a variety of human and mouse samples, as well as from *A. thaliana*, *E. coli*, *S. aureus*, *S. cerevisiae* and *D. discoideum*.

*ENZYME* (<http://www.expasy.org/enzyme>)

This is a repository of information relative to the nomenclature of enzymes.

## *SWISS-MODEL repository*

(<http://swissmodel.expasy.org/repository>)

The SWISS-MODEL repository is a database of structural protein models, automatically generated using homology modelling.

## Download

For all the ExPASy databases, data and associated documentation files can be copied locally by anonymous FTP (<ftp.expasy.org>). In particular, the different download options for the Swiss-Prot and TrEMBL databases, including the different available subsections, release frequencies and data formats, are documented at <http://www.expasy.org/sprot/download.html>.

## Software tools

The ExPASy tools page (<http://www.expasy.org/tools>) contains links to many useful sequence analysis and proteomics tools. Some of these point to tools developed and maintained by the ExPASy team (see list below), and many others refer to services on web sites around the world.

### *Sequence analysis tools*

- **BLAST** provides very fast similarity searches of a protein sequence against a protein or nucleotide database. The ExPASy BLAST service is maintained in collaboration with the Swiss EMBnet node on dedicated hardware. The native output of BLAST is extended with several original features.
- **ScanProsite** scans a sequence against all the patterns, profiles and rules in PROSITE or scans a pattern, profile or rule against all sequences in Swiss-Prot, TrEMBL and/or PDB.
- **SWISS-MODEL**, an automated knowledge-based protein modelling server, is able to build models for the three-dimensional structure of proteins whose sequence is closely related to that of proteins with known 3D structure.
- **ProtParam** calculates physico-chemical parameters of a protein sequence such as the amino-acid composition, the pI, the atomic composition, the extinction coefficient, etc.
- **ProtScale** computes and represents the profile produced by any amino-acid scale on a selected protein. Some 50 predefined scales are available, such as the Doolittle and Kyte hydrophobicity scale.
- **RandSeq** generates a random protein sequence, based on a user-specified amino acid composition and sequence length.
- **Myristoylator** predicts N-terminal myristoylation of proteins by neural networks.

- **Sulfinator** predicts tyrosine sulfation sites within protein sequences.
- **Translate** translates a nucleotide sequence into a protein in six reading frames.

### *Proteomics tools*

- **AACompliment** identifies a protein by its amino-acid composition.
  - **AACompSim** for a given Swiss-Prot entry, finds the database entries which have the most similar amino-acid composition.
  - **Compute pI/MW** computes the theoretical isoelectric point (pI) and molecular weight (MW) from a Swiss-Prot or TrEMBL entry or for a user sequence.
  - **FindMod** predicts potential protein post-translational modifications and potential single amino acid substitutions in peptides. Experimentally measured peptide masses are compared with the theoretical peptides calculated from a specified Swiss-Prot entry or from a user-entered sequence. Mass differences are used to better characterize the protein of interest.
  - **FindPept** identifies peptides resulting from unspecific cleavage of proteins by their experimental masses, taking into account artefactual chemical modifications, post-translational modifications and protease autolytic cleavage.
  - **GlycanMass** calculates the mass of an oligosaccharide structure.
  - **GlycoMod** predicts possible oligosaccharide structures that occur on proteins from their experimentally determined masses. The mass of a potential glycan is compared to a list of pre-computed masses of glycan compositions.
  - **PeptideCutter** predicts potential protease cleavage sites and sites cleaved by chemicals in a given protein sequence.
  - **PeptideMass** calculates the theoretical masses of peptides generated by the chemical or enzymatic cleavage of proteins so as to assist in the interpretation of peptide mass fingerprinting.
  - **PeptIdent, TagIdent, MultiIdent** identify proteins using a variety of experimental information such as the pI, the MW, the amino acid composition, partial sequence tags and peptide mass fingerprinting data.
- ExPASy proteomics tools (such as PeptIdent, TagIdent, MultiIdent, PeptideMass, FindPept, or FindMod), when performing their computations and predictions, use the annotations relevant to post-translational modifications and processing, as well as splice variants documented in the Swiss-Prot feature tables.

## ExPASy as a portal to other life science resources

The mass of information available to life scientists on the Web has created many opportunities, but also brought new dangers. One of the most critical problems, the difficulty to distinguish useful and up-to-date sources of information from sites that provide either 'fossilized' or low-quality data, is addressed by the following series of lists and tools on ExPASy:

*Amos' WWW links* (<http://www.expasy.org/allinks.html>)

A page that contains links to over a thousand information and resources for the life sciences, updated very frequently and organized in a number of sections corresponding to specific topics.

*WORLD-2DPAGE*

(<http://www.expasy.org/ch2d/2d-index.html>)

A list of all known 2-D PAGE database WWW servers and related services.

*BioHunt* (<http://www.expasy.org/BioHunt>)

A service to help search the Internet for molecular biology information. BioHunt is built by Marvin, a software robot, which automatically roams the web to search and index life science and bioinformatics information.

*2DHunt* (<http://www.expasy.org/ch2d/2dHunt>)

A specialized index for 2-D PAGE-related sites.

## Other interesting ExPASy features

*Biochemical pathways*

(<http://www.expasy.org/tools/pathways>)

An indexed, digitized and clickable version of the Boehringer Mannheim's 'Biochemical Pathways' poster, allowing the user to navigate through the graphical representation of metabolic pathways, linked to the ENZYME database.

*DeepView (Swiss-PdbViewer)*

(<http://www.expasy.org/spdbv>)

An application running under Microsoft Windows, Mac, SGI and Linux, offering a wide range of options to visualize and manipulate protein structures. It can also be used as a WWW helper application for the display of PDB formatted entries. Swiss-Pdb Viewer complements the SWISS-MODEL homology-modeling tool.

*2-D PAGE* (<http://www.expasy.org/ch2d>)

A wide variety of information concerning 2-D PAGE is available from ExPASy. This includes the full description of experimental protocols; a 2-D gel viewer is available for download.

*Protein Spotlight* (<http://www.expasy.org/proteinspotlight>)

A periodical review centered on a specific protein or group of proteins.

*Swiss-Quiz* (<http://www.expasy.org/swiss-quiz>)

Gives you a chance to win some Swiss chocolate (real, not virtual!) after having successfully answered a molecular biology quiz.

*ExpASYbar* (<http://expasybar.mozdev.org>)

A useful navigation bar to the most important databases and tools on ExPASy is an add-on to the free Mozilla web browser (<http://www.mozilla.org>), and can be downloaded from <http://expasybar.mozdev.org>.

## Mirror sites

ExPASy mirror sites are computers that host exact copies of the information available from the Geneva ExPASy server, updated at the same frequency. They can help users to access the ExPASy databases and tools more rapidly in locations that do not have a fast connection to Switzerland. As of today, eight mirror sites are operational, in the following countries:

Australia: <http://au.expasy.org>

Bolivia: <http://bo.expasy.org>

Brazil: <http://br.expasy.org>

Canada: <http://ca.expasy.org>

China: <http://cn.expasy.org>

South Korea: <http://kr.expasy.org>

Taiwan: <http://tw.expasy.org>

United States: <http://us.expasy.org>

## How to cite ExPASy

If you want to cite ExPASy in a publication, please use the following reference:

Gasteiger E., Gattiker A., Hoogland C., Ivanyi I., Appel R.D., Bairoch A.

ExPASy: the proteomics server for in-depth protein knowledge and analysis:

*Nucleic Acids Res.* **31**:3784-3788(2003).

This document was written and designed by Elisabeth Gasteiger from the Swiss Institute of Bioinformatics and distributed by the P&PR Publications Committee of EMBnet.

EMBnet - European Molecular Biology network - is a network of bioinformatics support centres situated primarily in Europe. Most countries have a national node which can provide training courses and other forms of help for users of bioinformatics software.

<http://www.embnet.org/>

A Quick Guide To ExPASy

First edition © 2004

# A Quick Guide EXPASY

EMBnet

