

S1 (box) | Web resources for peptide sequencing

Gene and protein databases

- Ensembl (<http://www.ensembl.org/>). Annotation project of the human and other genomes. This database is excellently organized with links for each gene to a genome viewer that integrates cross-species, haplotype, gene and protein information. It also contains many other resources.
- GenBank (<http://www.ncbi.nlm.nih.gov/Sitemap/index.html#GenBank>). Entries are guaranteed not to change. However, many entries for the same protein and the large size of this database make it unwieldy.
- International Protein Index (IPI; <http://www.ebi.ac.uk/IPI/IPIhelp.html>). Minimal, non-redundant database that is composed of Ensembl and Swiss-Prot entries. Excellent for peptide-sequencing projects, but the protein entries are still in flux.
- National Center for Biotechnology Information (NCBI; <http://www.ncbi.nlm.nih.gov/>). Central entry point to PubMed, Entrez and the many other essential services that are offered by the NCBI.
- Swiss-Prot (<http://us.expasy.org/sprot/>). Excellent protein database that is annotated by hand.
- Unigene (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene>). Partitions GenBank sequences into non-redundant clusters.
- Uniprot (<http://www.pir.uniprot.org>). A combination of Swiss-Prot, IPI and the Protein Identification Resource (PIR). Looks set to become the standard database for proteomics projects.

Protein-identification software

- Genomics solutions (ProFound and Sonar ms/ms; <http://65.219.84.5/ProteinId.html>). These are commercial, database identification programmes. However, small searches can be carried out through the internet for free.
- Mascot (<http://www.matrixscience.com/>). Popular, probability-based, peptide-identification software, which works with most mass-spectrometry formats. Small searches can be carried out through the internet for free.
- MS BLAST by the Shevchenko group at the European Molecular Biology Laboratory (EMBL; <http://dove.embl-heidelberg.de/Blast2/msblast.html>). BLAST- and peptide-sequence-tag-based internet tools that allow proteins that are not in the other protein databases to be identified on the basis of homology.
- MSQuant (<http://Msquant.sourceforge.net>). Open-source software for peptide validation and quantification from Matthias Mann's laboratory. It is based on output from the Mascot search engine and on the SILAC (stable-isotope labelling with amino acids in cell culture) technology for quantification.
- PeptideSearch from the EMBL Bioanalytical Research Group (<http://www.mann.embl-heidelberg.de/GroupPages/PageLink/peptidesearchpage.html>). Original implementation of the peptide-sequence-tag algorithm.
- Sequest (<http://fields.scripps.edu/sequest/>). Peptide-identification software that is based on autocorrelation and is mainly used for ion-trap data. This software is available for ThermoFinnigan instruments.
- University of California, San Francisco (UCSF) ProteinProspector (<http://prospector.ucsf.edu/>). A free site with many useful tools (for example, a tool for assigning disulphide bridges on the basis of mass-spectrometry data). However, it cannot be used for searching large data sets.

De novo sequencing software

- Lutfisk (<http://www.hairyfatguy.com/Lutfisk>). *De novo* sequencing software by the Richard Johnson group. This software has limited automation.
- PEAKS (<http://www.bioinformaticssolutions.com/products/peaksstudio.php>). Commercially available *de novo* sequencing software.
- SPIDER (<http://proteome.sharcnet.ca:8080/spider.htm>). Another freely available *de novo* sequencing software.

Statistical treatment of peptide-sequencing results

- Institute for Systems Biology (ISB) Protein Informatics (<http://www.systemsbiology.org/Default.aspx?pagename=proteomicssoftware>). This site provides tools for the statistical treatment of peptide-sequencing results, as well as other freely available tools for proteomics research by the ISB.