

## Database ordered by group

### Genomdatenbanken

Enthalten komplette Genome (Contigs) [Human, Maus, c.Eleganz, Zebrafisch, Hefe, Drosophila], mit Annotationen (Exon, Intron, etc.)

Nützlich: für Targetsuche, Primerbau, Sequenzvergleiche, Alignments, BLAST(N)

### GENBANK

GenBank is the NIH (National Institutes of Health) genetic sequence database, an annotated collection of all publicly available DNA sequences. There are approximately 1,500,000,000 bases in 2,209,000 sequence records as of April 1998. GenBank is part of the International Nucleotide Sequence Database Collaboration, which is comprised of the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI. These three organizations exchange data on a daily basis.

### EMBL

The EMBL Nucleotide Sequence Database is a comprehensive database of DNA and RNA sequences collected from the scientific literature and patent applications and directly submitted from researchers and sequencing groups. Data collection is done in collaboration with GenBank (USA) and the DNA Database of Japan (DDBJ). The database currently doubles in size approximately every 12 months.

### Proteinsequenzdatenbanken

Enthalten AS-Sequenzen von Proteinen annotiert mit Domänen, Modifikationen und oftmals auch schon mit Funktionen

Nützlich: „Wozu passt meine gefundene cDNA ?“, Proteinvergleiche zwischen Spezies („Hat der Mensch ein ähnliches Protein wie Drosophila ?“, Grundlage für Strukturvorhersagen, Patentfragen, Phylogenetische Bäume

### TREMBL

Automatisch generiert aus vorhergesagten EMBL Sequenzen.

TREMBL is a protein sequence database supplementing the SWISS-PROT Protein Sequence Data Bank. TREMBL contains the translations of all coding sequences (CDS) present in the EMBL Nucleotide Sequence Database not yet integrated in SWISS-PROT. TREMBL can be considered as a preliminary section of SWISS-PROT. For all TREMBL entries which should finally be upgraded to the standard SWISS-PROT quality, SWISS-PROT accession numbers have been assigned.

TREMBL is split in two main sections: SP-TREMBL and REM-TREMBL:

**SP-TREMBL** (*SWISS-PROT TREMBL*) contains the entries which should be incorporated into SWISS-PROT. SWISS-PROT accession numbers have been assigned for all SP-TREMBL entries.

REM-TREMBL (*REMaining TREMBL*) contains the entries that are not wanted to be in SWISS-PROT.

### SWISSPROT

Hohe Qualität, da manuell gepflegt – überprüfen EMBL/TREMBL Daten in Literatur etc.  
Annotationen: Funktionen, Domänen, posttranslationale Modifikationen, ...

SWISSPROT is a protein database that contains sequences translated from the EMBL Nucleotide Sequence Database, prepared by the European Bioinformatics Institute. A small part of the information in SWISS-PROT was originally adapted from information

contained in the Protein Sequence Database of the Protein Information Resource (PIR) supported by the Division of Research Resources of the NIH, The SWISS-PROT protein sequence data bank and its new supplement TrEMBL.

## **PIR**

---

Manuell gepflegt

PIR - Protein Information Resource

The Protein Identification Resource consists of an integrated computer system composed of a number of protein and nucleic acid sequence databases and software designed for the identification and analysis of protein sequences and their corresponding coding sequences. The PIR serves the scientific community through on-line access, distributing magnetic tapes, and performing off-line sequence identification services for researchers.

Strukturdatenbanken (meist 3D) für Proteine Enthalten Strukturinformationen zu den Proteinen Nützlich: DrugDesign („Form“ des akt. Zentrums -> Ligandendesign)
----------------------------------------------------------------------------------------------------------------------------------------------------------------------

## **PDP**

---

Standarddatenbank für Proteinstrukturen; enthält Atomkoordinaten, Primär/Sekundärstruktur, Kristallographie und NMR Daten

## **DSSP**

---

Kann z.B. PDP Datei einlesen und daraus graphische Ausgabe des Proteins erzeugen.

The DSSP program was designed by Wolfgang Kabsch and Chris Sander to standardize secondary structure assignment. The DSSP database is a database of secondary structure assignments (and much more) for all of the entries in the Protein Data Bank (PDB).

## **HSSP**

---

HSSP (homology-derived structures of proteins) is a derived database merging structural (2-D and 3-D) and sequence information (1-D). For each protein of known 3D structure from the Protein Data Bank, the database has a file with all sequence homologues, properly aligned to the PDB protein. Homologues are very likely to have the same 3D structure as the PDB protein to which they have been aligned. As a result, the database is not only a database of sequence aligned sequence families, but it is also a database of implied secondary and tertiary structures. Likely secondary structure are carried over from the PDB protein to each homologous protein. Tertiary structure models can be built by fitting the sequence of the homologue as aligned into the 3D template of the protein of known structure.

## **PDBFINDER**

---

The PDBFINDER database is constructed using a PERL script from the PDB, DSSP and HSSP databases. Many of the fields contained in the PDBFINDER database are difficult to access from the original databases. Some Information is retrieved from the original literature.

Speziesspezifische
--------------------

## **MGD**

---

Mousegenomedatabase; enthält Kartierung (physische, genetische), Phänotypen

## **FlyBase**

---

A Database of the Drosophila Genome

FlyBase is a comprehensive database for information on the genetics and molecular biology of Drosophila.

## **ACEDB**

---

DB for c.Eleganz

## **SGD**

---

DB für "Bierhefe" (Sequenzen, Karten, ...)

Metabolische Datenbanken

Enthalten Enzyme mit Funktionen, CoFaktoren, verbundene Krankheiten

## **ENZYME**

---

The ENZYME data bank contains the following data for each type of characterized enzyme for which an EC number has been provided: EC number, Recommended name, Alternative names, Catalytic activity, Cofactors, Pointers to the SWISS-PROT entrie(s) that correspond to the enzyme, Pointers to disease(s) associated with a deficiency of the enzyme.

## **BRENDA**

---

Enzymstruktur, -stabilität, ...

## **KEGG, EMP**

---

Pathways, Orthologe Gruppen, Molekülkatalog, Genomkarten, Genkataloge

Mutations-Datenbanken

Enthalten: SNPs, Indels

Nützlich für: Markersuche

## **OMIM**

---

OMIM - Online Mendelian Inheritance in Man

This database is a catalog of human genes and genetic disorders authored and edited by Dr. Victor A. McKusick and his colleagues at Johns Hopkins and elsewhere, and developed for the World Wide Web by NCBI, the National Center for Biotechnology Information. The database contains textual information, pictures, and reference information. It also contains copious links to NCBI's Entrez database of MEDLINE articles and sequence information. For more refined maps of genes and DNA segments, consult the NCBI Entrez Genomes division and the Genome Data Base.

## **HGBASE**

---

the Human Genic Bi-Allelic Sequences Database

HGBASE is an attempt to summarize all known sequence variations in the human genome, to facilitate research into how genotypes affect common diseases, drug responses, and other complex phenotypes

Sequence variations are presented with details of how they are physically and functionally related to the closest neighbouring gene. Records include SNPs, Indels, simple tandem repeats, and other sequence alternatives, regardless of location, allele frequencies, or known affect upon phenotype. All records are highly curated and annotated, ensuring maximal utility and data accuracy.

Sekundäre oder andere Datenbanken  
Enthalten: Infos, Verknüpfungen zwischen anderen Datenbanken

---

## LITERATURDATENBANKEN

medline, bioinformatics-online, nature, science

---

## PROSITE

PROSITE is a method of determining what is the function of uncharacterized proteins translated from genomic or cDNA sequences. It consists of a database of biologically significant sites, patterns and profiles that help to reliably identify to which known family of protein (if any) a new sequence belongs.

---

## PRINTS

The PRINTS database is a compendium of protein motif fingerprints. Each fingerprint has been defined and iteratively refined using database scanning procedures within the ADSP sequence analysis package. Two types of fingerprint are represented in the database, i.e. they are either simple or composite, depending on their complexity: simple fingerprints are essentially single-motifs; while composite fingerprints encode multiple motifs. The bulk of the database entries are of the latter type because discrimination power is greater for multi-component searches, and results are consequently easier to interpret.

---

## PRODOM

Cluster von SWISSPROT & TREMBLE; multiple Alignments und Konsensussequenzen

---

## PFAM

Proteinalignments / Clustering (Konsensussequenzen, funktionale Einheiten, ...)

---

## INTERPRO

integriert PROSITE, PRINTS, PFAM & PRODOM; verknüpft mit SWISSPROT & TREMBLE liefert umfangreiche Informationen zu geg. Proteinsequenzen

---

## GENECARDS

Menschliche Gene; Genprodukte, bezug zu Krankheiten, Funktionsannotationen

---

## ENSEMBLE

Verarbeitet Daten aus anderen Datenbanken (EMBL, ...); Zieht Daten aus den anderen DBs, analysiert sie und fügt sie in eigenen Bestand ein (-> Interface für die anderen Datenbanken, Aufbereitung der Informationen / Annotation der Daten); Abfragen nach Contigs, Marker, Kartierungen, ...; Genvorhersagen; Repeats-Analyse, Homologie-Analyse

---

## DAS

Distributed annotation server  
Besteht aus „referenceserver“ (DNA / Sequenzdatenbank) und „Annotationsserver“ – gibt „jedem“ die Möglichkeit Informationen / Annotationen zu bestimmten Segmenten / Sequenzbereichen hinzuzufügen

---

## SRS

Sequence retrieval system  
Informationssystem für „alle“ Datenbanken; „kopiert“ und analysiert sich alle Datenbanken in Textfiles, die er dann durchsuchen und die Ergebnisse mit den original DB verlinken kann.