

# Molekularbiologische Datenbanken

Übungen  
Sommersemester 2003



Silke Trißl  
Ulf Leser



Wissensmanagement in der  
Bioinformatik

## Organisatorisches

- Mittwoch 13 – 15 Uhr, RUD25 I.011
  - Mi, 02. Juli 2003 entfällt
  - Ende: 16. Juli 2003
- Anmeldung für die Übung über Goya
  - Übungsschein
  - 2 Übungsscheine (MDB + DWH) = Seminarschein
- Web-page zur Übung:
  - <http://www.informatik.hu-berlin.de/wbi/teaching/sose03/mdb/uebung/index.html>

# Organisatorisches – 2 –

- Sprechstunden:
  - Silke Trißl: Mo + Di 10 – 12 Uhr  
RUD IV.102  
oder nach Vereinbarung
  
  - Ulf Leser: Fr 13 – 15 Uhr  
RUD IV.103

## Ablauf

- 5 Aufgaben
  - Aufgabe 1: jeder für sich
  - Aufgaben 2 – 5: in Gruppen von 2 – 3 Personen
  
- Abgabe
  - Über Goya oder in RUD25 IV.102
  - Vorstellung der Ergebnisse jeder Gruppe in den Übungen
  - Ergebnisse der Übungen über Goya
  - Einsehen der Lösung 1 Woche nach Abgabe

# Aufgaben

## 1. Datenbanksuche

1 Woche 30.04. – 07.05.

## 2. Microarray-DB-Modellierung

2 Wochen 07.05. – 21.05.

## 3. Laden von Rohdaten

2 Wochen 21.05. – 04.06.

## 4. Integrieren von SWISS-PROT-Daten

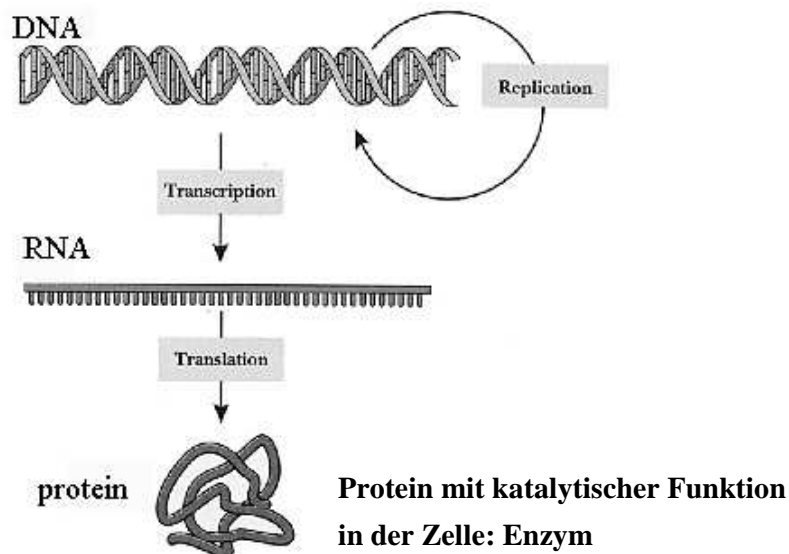
3 Wochen 04.06. – 25.06.

## 5. Integrieren von GO und GOA

3 Wochen 25.06. – 16.07.

# Aufgabe 1 - Datenbanksuche

„Central Dogma in Molecular Biology“



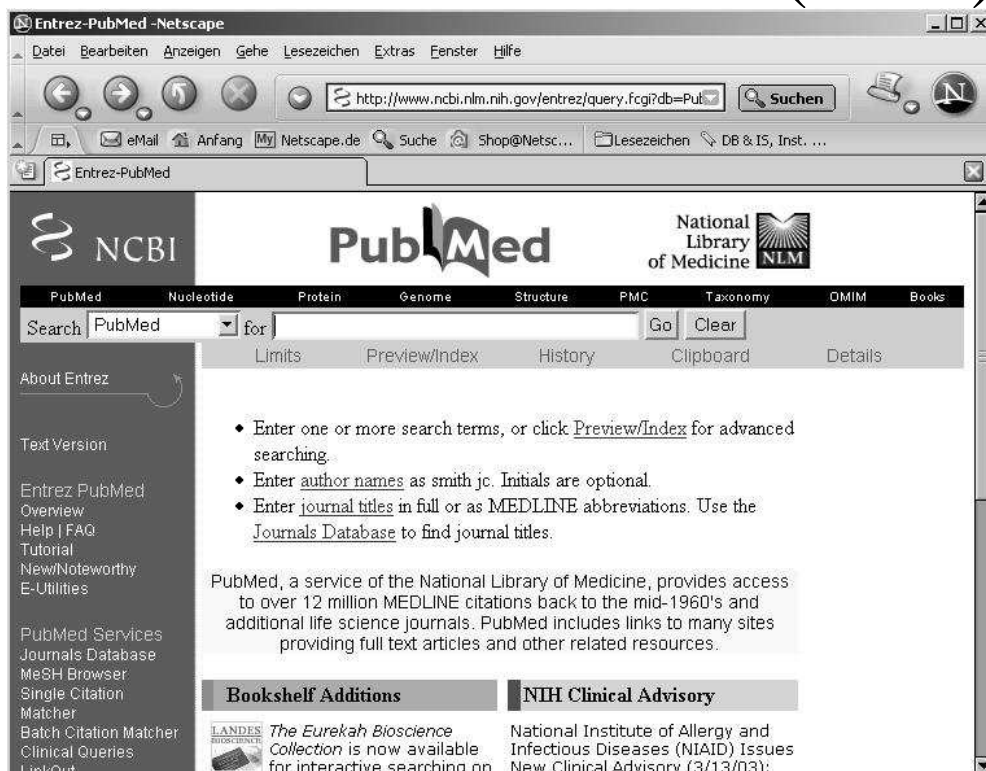
Quelle: <http://cellbio.utmb.edu/cellbio/ribosome.htm>

# Datenbanken - Literatur

http://www.informatik.hu-berlin.de/wbi/teaching/sose03/mdb/uebung/Aufgabe1/datenbanken.html

Literature search	
<a href="#">Entrez PubMed</a>	PubMed, a service of the National Library of Medicine, provides access to over 12 million MEDLINE citations back to the mid-1960's and additional life science journals. PubMed includes links to many sites providing full text articles and other related resources.
<a href="#">MeSH - Medical Subject Headings</a>	It is designed to help quickly locate descriptors of possible interest and to show the hierarchy in which descriptors of interest appear.
<a href="#">OMIM - Online Mendelian Inheritance in Man</a>	This database is a catalog of human genes and genetic disorders.

## Datenbanken – Literatur (cont.)



The screenshot shows the Entrez-PubMed website in a Netscape browser window. The browser's address bar displays the URL: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>. The website header includes the NCBI logo, the PubMed logo, and the National Library of Medicine (NLM) logo. Below the header, there is a search bar with the text "Search PubMed for" and buttons for "Go" and "Clear". A navigation menu includes links for "Limits", "Preview/Index", "History", "Clipboard", and "Details". The main content area features a list of search tips:

- Enter one or more search terms, or click [Preview/Index](#) for advanced searching.
- Enter author names as smith jc. Initials are optional.
- Enter journal titles in full or as MEDLINE abbreviations. Use the [Journals Database](#) to find journal titles.

Below the search tips, there is a paragraph describing PubMed: "PubMed, a service of the National Library of Medicine, provides access to over 12 million MEDLINE citations back to the mid-1960's and additional life science journals. PubMed includes links to many sites providing full text articles and other related resources."

At the bottom of the page, there are two sections: "Bookshelf Additions" and "NIH Clinical Advisory". The "Bookshelf Additions" section mentions "The Eureka Bioscience Collection is now available for interactive searching on". The "NIH Clinical Advisory" section mentions "National Institute of Allergy and Infectious Diseases (NIAID) Issues New Clinical Advisory (3/13/03):".

# Datenbanken - Sequenzen

## DNA sequence databases

<a href="#">GenBank</a>	GenBank® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences.
<a href="#">EMBL über SRS@EBI</a>	The EMBL Nucleotide Sequence Database constitutes Europe's primary nucleotide sequence resource. Main sources for DNA and RNA sequences are direct submissions from individual researchers, genome sequencing projects and patent applications.

## Protein sequence databases

<a href="#">SWISS-PROT</a>	The Swiss-Prot Protein Sequence Database is a database of protein sequences. It contains high-quality annotation, is non-redundant, and cross-referenced to several other databases, notably the EMBL nucleotide sequence database, PROSITE pattern database and PDB.
<a href="#">PIR</a>	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.



Top Page Query Results Projects Views Databanks

Reset  Quick Search

show all  collapse all

- Literature, Bibliography and Reference Databases**
  - [MEDLINE](#)  [MEDLINE \(Updates\)](#)  [MEDLINE \(Main Release\)](#)  [OMIM](#)
  - [TAXONOMY](#)  [GENETICCODE](#)
- Nucleotide sequence databases**
  - [EMBL](#)  [EMBL \(Release\)](#)  [EMBL \(Updates\)](#)  [EMBL \(WGS\)](#)
  - [EMBL \(TPA\)](#)  [EMBL \(Contig\)](#)  [REFSEQ](#)  [ENSEMBL HUMAN](#)
  - [ENSEMBL MOUSE](#)  [ENSEMBL FLY](#)  [ENSEMBL FISH](#)  [IMGTHLA](#)
  - [IMGT/LIGM-DB](#)  [PATENT DNA](#)
- Protein sequence databases**
  - [SWALL \(SPTR\)](#)  [Swiss-Prot](#)  [SpTrEMBL](#)  [TrEMBL \(Updates\)](#)
  - [IPI](#)  [RemTrEMBL](#)  [PIR](#)  [REFSEQP](#)
  - [PATENT PRT](#)  [JPO PRT](#)  [USPO PRT](#)  [MHCBN](#)
  - [SWISSCHANGE](#)
- Nucleotide related databases**
- Protein function databases**
- Protein structure databases**
- Enzymes, reactions and metabolic pathway databases**
- Mutation and SNP databases**

Query forms: [Standard](#) [Extended](#)

Browse Databanks

Quick Sequence Search:  [FastA](#)

Applications

- ▶ bookmark this [link](#) to return to your project
- ▶ [BookmarkLet](#) for:
  - ♦ [Protein Seq](#)
  - ♦ [DNA/RNA Seq](#)
  - ♦ [Structures](#)



# Datenbanken - Proteine

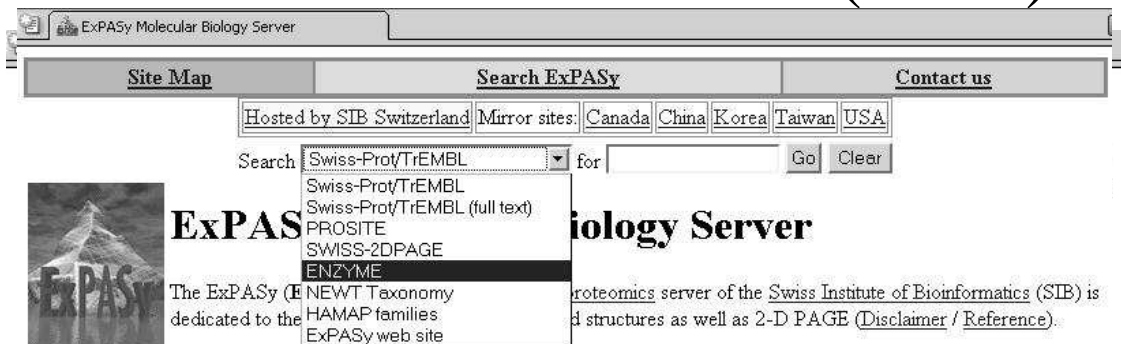
## Protein Structure databases

<a href="#">PDB - Protein Database</a>	The Protein Data Bank (PDB) is an archive of experimentally determined three-dimensional structures of biological macromolecules, serving a global community of researchers, educators, and students. The archives contain atomic coordinates, bibliographic citations, primary and secondary structure information, as well as crystallographic structure factors and NMR experimental data.
--	---

## Enzymes, reactions and metabolic pathway databases

<a href="#">KEGG</a>	KEGG is a suite of databases and associated software, integrating our current knowledge on molecular interaction networks in biological processes (PATHWAY database), the information about the universe of genes and proteins (GENES/SSDB/KO databases), and the information about the universe of chemical compounds and reactions (COMPOUND/REACTION databases).
<a href="#">ENZYME</a>	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 entry(s) that correspond to the enzyme, Pointers to disease(s) associated with a deficiency of the enzyme.
<a href="#">Roche Applied Science "Biochemical Pathway"</a>	This page gives access to the digitized version of the Roche Applied Science "Biochemical Pathways" wall chart.

# Datenbanken – Proteine (cont.)



The screenshot shows the ExPASy Molecular Biology Server interface. At the top, there are navigation links for 'Site Map', 'Search ExPASy', and 'Contact us'. Below these, there are links for 'Hosted by SIB Switzerland' and 'Mirror sites' including Canada, China, Korea, Taiwan, and USA. A search bar is visible with a dropdown menu showing options like 'Swiss-Prot/TrEMBL', 'PROSITE', 'SWISS-2DPAGE', and 'ENZYME'. The 'ENZYME' option is currently selected. To the left of the search bar is the ExPASy logo, and to the right is the text 'Molecular Biology Server'. Below the search bar, there is a brief description of the server as a proteomics server of the Swiss Institute of Bioinformatics (SIB).

[\[Announcements\]](#) [\[Job opening\]](#) [\[Mirror Sites\]](#)

Databases	Tools and software packages
<ul style="list-style-type: none"> <li>• <a href="#">Swiss-Prot and TrEMBL</a> - Protein knowledgebase</li> <li>• <a href="#">PROSITE</a> - Protein families and domains</li> <li>• <a href="#">SWISS-2DPAGE</a> - Two-dimensional polyacrylamide gel electrophoresis</li> <li>• <a href="#">ENZYME</a> - Enzyme nomenclature</li> <li>• <a href="#">SWISS-3DIMAGE</a> - 3D images of proteins and other biological macromolecules</li> <li>• <a href="#">SWISS-MODEL Repository</a> - Automatically generated protein models</li> <li>• <a href="#">CD40Lhase</a> - CD40 ligand defects</li> </ul>	<ul style="list-style-type: none"> <li>• <a href="#">Proteomics and sequence analysis tools</a> <ul style="list-style-type: none"> <li>◦ <a href="#">Proteomics</a> [PeptIdent, PeptideMass, ...]</li> <li>◦ <a href="#">DNA -&gt; Protein</a> [Translate]</li> <li>◦ <a href="#">Similarity searches</a> [BLAST]</li> <li>◦ <a href="#">Pattern and profile searches</a> [ScanProsite]</li> <li>◦ <a href="#">Post-translational modification and topology prediction</a></li> <li>◦ <a href="#">Primary structure analysis</a> [ProtParam, pI/MW, ProtScale]</li> <li>◦ <a href="#">Secondary and tertiary structure prediction</a></li> </ul> </li> </ul>

# Fragen?

- Aufgabe 1 bei Goya oder auf der Web-page
- Datenbank-Katalog auf der Web-page
- Lösung bis 06.05., 17 Uhr

