Proteomics and Protein Structure

Introduction to Bioinformatics
Dortmund, 16.-20.07.2007

Lectures:
Sven Rahmann

Exercises:
Udo Feldkamp, Michael Wurst
Overview

• Proteomics tools for almost all purposes
  – ExPASy Server (Expert Protein Annotation System)

• Structural Classification of Proteins
  – CATH and SCOP databases

• 3D protein structure information and visualization
  – wwPDB containing RCSB-PDB
  – different embedded visualization tools

• Specialized protein databases (e.g., GPCR DB)

• CASP: protein structure prediction contest
ExPASy: Expert Protein Analysis SYstem

- URL: http://expasy.org/

ExPASy Proteomics Server

The ExPASy (Expert Protein Analysis System) proteomics server of the Swiss Institute of Bioinformatics (SIB) is dedicated to the analysis of protein sequences and structures as well as 2-D PAGE (Disclaimer / References).

Databases
- UniProt Knowledgebase (Swiss-Prot and TrEMBL) - Protein knowledgebase
- PROSITE - Protein families and domains
- SWISS-2DPAGE - Two-dimensional polyacrylamide gel electrophoresis
- ENZYME - Enzyme nomenclature
- SWISS-MODEL Repository - Automatically generated protein models
- Links to many other molecular biology databases

Tools and software packages
- Proteomics and sequence analysis tools
  - Identification and characterization (Aldente, FindMod, Popitam, Phenyx, plMw, ProtParam...)
  - DNA -> Protein
  - Similarity searches (ELAST...)
  - Pattern and profile searches (ScanProsite...)
  - Post-translational modification and topology prediction
  - Primary structure analysis
  - Secondary and tertiary structure tools (Swiss-PdbViewer...)
  - Alignment and Phylogenetic analysis
- ImageMaster / Melanie - Software for 2-D PAGE analysis
- MSight - Mass Spectrometry Imager
- Roche Applied Science’s Biochemical Pathways
## ENZYME Database

### NiceZyme View of ENZYME: EC 1.1.1.25

<table>
<thead>
<tr>
<th>Official Name</th>
<th>Alternative Name(s)</th>
<th>Reaction catalysed</th>
<th>Comment(s)</th>
<th>Cross-references</th>
</tr>
</thead>
</table>
| Shikimate dehydrogenase. | 5-dehydroshikimate reductase. | Shikimate + NADP(+) <-> 3-dehydroshikimate + NADPH | • NAD(+) cannot replace NADP(+).  
• In higher organisms, this enzyme forms part of a multienzyme complex with EC 4.2.1.10. | Biochemical Pathways; map number(s) |

D3
Hundreds of Proteomics Tools

ExPASy Proteomics tools

The tools marked by 🚗 are local to the ExPASy server. The remaining tools are developed and hosted on other servers.

[Protein identification and characterization] [DNA -> Protein] [Similarity searches] [Pattern and profile searches]
[Post-translational modification prediction] [Topology prediction] [Primary structure analysis] [Secondary structure prediction] [Tertiary structure]
[Sequence alignment] [Phylogenetic analysis] [Biological text analysis]

<table>
<thead>
<tr>
<th>Protein identification and characterization with peptide mass fingerprinting data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aldente 🚗</strong> - Identify proteins with peptide mass fingerprinting data. A new, fast and powerful tool that takes advantage of Hough transformation for spectra recalibration and outlier exclusion.</td>
</tr>
<tr>
<td><strong>FindMod 🚗</strong> - Predict 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, and mass differences are used to better characterize the protein of interest.</td>
</tr>
<tr>
<td><strong>FindPept 🚗</strong> - Identify peptides that result from unspecific cleavage of proteins from their experimental masses, taking into account artefactual chemical modifications, post-translational modifications (PTM) and protease autolytic cleavage</td>
</tr>
<tr>
<td><strong>GlycoMod 🚗</strong> - Predict possible oligosaccharide structures that occur on proteins from their experimentally determined masses (can be used for free or derivatized oligosaccharides and for glycopeptides)</td>
</tr>
</tbody>
</table>

- **Mascot** - Peptide mass fingerprint from Matrix Science Ltd., London
- **PepMAPPER** - Peptide mass fingerprinting tool from UMIST, UK
- **PFMUTS** - Shows the possible single and double mutations of a peptide fragment from MALDI peptide mass fingerprinting
- **ProFound** - Search known protein sequences with peptide mass information from Rockefeller and NY Universities [or from Genomic Solutions]
- **ProteinProspector** - LC/MS tools for peptide mass data (MS Fit, MS Pattern, MS Digest, etc.)
Protein Domains

- “building blocks” of proteins
- functional units
- Major Domain Databases:
  - Pfam
  - SMART
Structural Classification: CATH

Welcome to SCOP: Structural Classification of Proteins.

1.71 release (October 2006)

27599 PDB Entries. 1 Literature Reference. 75930 Domains. (excluding nucleic acids and theoretical models).

Folds, superfamilies, and families statistics here
New folds, superfamilies, families.
List of obsolete entries and their replacements.

Authors: Alexey G. Murzin, John-Marc Chandonia, Antonina Andreeva, Dave Howorth, Loredana Lo Conte, Bartlett G. Ailey, Steven E. Brenner, Tim J. P. Hubbard, and Cyrus Chothia. scop@mrc-lmb.cam.ac.uk


Access methods

- Enter SCOP at the top of the hierarchy
- Keyword search of SCOP entries
- SCOP parsable files
- All SCOP releases and reclassified entry history
- SCOP domain sequences and pdb-style coordinate files (ASTRAL)
SCOP

• 4 Levels
  – Class (alpha; beta; alpha/beta; alpha+beta)
  – Fold (1000 folds)
  – superfamily (evolutionary relationship)
  – family

• Proteins in each family consist of the same domains
ALERT: Our data files are changing soon. Please see http://www.wwpdb.org for more details.

Welcome to the RCSB PDB

The RCSB PDB provides a variety of tools and resources for studying the structures of biological macromolecules and their relationships to sequence, function, and disease.

The RCSB is a member of the wwPDB whose mission is to ensure that the PDB archive remains an international resource with uniform data.

This site offers tools for browsing, searching, and reporting that utilize the data resulting from ongoing efforts to create a more consistent and comprehensive archive.

Information about compatible browsers can be found here.

A narrated tutorial illustrates how to search, navigate, browse, generate reports and visualize structures using this new site. [This requires the Macromedia Flash player download]

Comments? info@rcsb.org

Molecule of the Month: Thymine Dimers

Summer is here, and we're all heading
Viewing PDB information

1nyt  DOI 10.2210/pdb1nyt/pdb

Red - Derived Information

Title  SHIKIMATE DEHYDROGENASE AroE COMPLEXED WITH NADP+

Authors  Roszak, A.W., Lapthorn, A.J.


History  Deposition 2003-02-13  Release 2003-03-04

Experimental Method  Type  X-RAY DIFFRACTION  Data [ EDS ]

Parameters  Resolution[A]  1.50  R-Value  0.136 (obs.)  R-Free  0.169  Space Group  C 2 (C 1 2 1)

Images and Visualization

Biological Molecule / Asymmetric Unit

Display Options

KiNG  Jmol  WebMol  MBT Protein Workshop  QuickPDB

Start 3D structure viewers here
Specialized Protein DBs

- http://www.gpcr.org/

Molecular Class-Specific Information System (MCSIS) project

Available MCSIS

- The GPCRDB: a Molecular-Specific Information System for G Protein-Coupled Receptors (created in 1994)
  - The GPCRDB at the CMBI, the Netherlands

- The NuclearDB: a Molecular-Specific Information System for Nuclear Receptors (created in April 2000)
  - The NuclearDB at the CMBI, the Netherlands
  - Mirror site at UCSF, USA (no longer available)

- The PrionDB: a Molecular-Specific Information System for Prion proteins (created July, 21 2003)
  - The PrionDB at the CMBI, the Netherlands
  - Mirror site at UCSF, USA (no longer available)

- The KChannelDB: a Molecular-Specific Information System for potassium channels (created July, 25 2003)
  - The KChannelDB at the CMBI, the Netherlands
  - Mirror site at UCSF, USA (no longer available)

- The GPCRIPDB: a Molecular-Specific Information System for GPCR Interacting Partners (G proteins & RAMPs) (created May, 23 2005)
Welcome to the Protein Structure Prediction Center!

Our goal is to help advance the methods of identifying protein structure from sequence. The Center has been organized to provide the means of objective testing of these methods via the process of blind prediction. In addition to support of the CASP meetings our goal is to promote an evaluation of prediction methods on a continuing basis.

CASP experiments aim at establishing the current state of the art in protein structure prediction, identifying what progress has been made, and highlighting where future effort may be most productively focused. The organizers are thankful to CASP assessors for their valuable contribution to this field.

There have been seven previous CASP experiments.

Homology Search

- Structure Predication by Homology search:
  - find structure in PDB with similar sequence as query protein
- High sequence similarity implies evolutionary relation and structural similarity
- However: structural similarity does not imply high sequence similarity (even 20% is ok)
Threading

• Try to fit a given sequence on all known structures
• Background: Angles of bonds between different amino acids must be in a limited range
• “dual” to Homology modeling
Ab Initio Prediction

- Molecular Dynamics (“low level”)
  - propose a confirmation, compute its stability, modify
  - very compute-intensive
  - only for short peptides

- Fragment Assembly
  - Assemble predicted 3D-structures of short fragments into a larger 3D structure