Science magazine

HELP   SUBSCRIPTIONS   FEEDBACK   SIGN IN

SEARCH   BROWSE

functional genomics resources

featured scientific sites

Site of the Month

IMB Jena Image Library of Biological Macromolecules

An almost awe-inspiring, extremely well-ordered collection of information and, especially, images of 3D structures for proteins, nucleic acids, carbohydrates, and other bio-involved molecules, from the Institute of Molecular Biotechnology in Jena, Germany. The resource is built around a large collection of individual pages devoted to particular molecules, each with basic PDB structure information and with a variety of interactive 3D visualization options, including scripts for the Chime and RasMol plug-ins and a (relatively slow-loading) Web-based alternative called WebMol. The expert can quickly find what he or she wants, through a variety of search and browse options. But there's plenty here for the newcomer to protein science as well, including a rich collection of pages with "Basic Information on Biological Macromolecules" that can act as a veritable tutorial on protein structure for the uninitiated. And the site has an appealing sense of whimsy, with offerings including a virtual "advent calendar" built around "Molecules of Light" such as rhodopsin and fluorescent proteins, and images of a human homodimer structure "dancing" to a MIDI rendering of "Swan Lake." (SW)

Gateways to general Web resources in genomics and post-genomics:

- Nine great general biology, biochemistry, and bioinformatics sites
- Selected genome research centers
- Nucleotide, genome, and protein sequences
- Expression data and microarrays
- Protein structure
- Proteomics
- Medical resources

Note: The links featured on this page will change periodically, so come back frequently. And be sure to visit our special pages on model organisms, educational resources, genome maps, and the biotech business.

Navigation

- home
- news
- research
- special issues
- human genome
- scientific resources
- biotech
- education
- about the site

Browse other functional-genomics resources:

Featured sites
Research centers
Model organisms
Gene sequences
Gene expression
Protein structure
Proteomics
Medical resources
Genome maps

The biotech business
Educational resources

Search the Web with
Google

GoClear Form

Upcoming Meetings:

- By discipline
- By date

Database QuickSearch:

Integrated databases

Sequence Retrieval System (EBI/LION Bioscience)
Entrez (NCBI)

Sequences

GenBank nucleotides
GenBank genomes

Nine Great General Biology, Biochemistry, and Bioinformatics Sites

- The Bio-Web
- BioChemLinks
- BioPlanet
- Art's Biotech Resource
- BioTech
- Molecular Biology Shortcuts
- Cell and Molecular Biology Online
- BioNetbook
- Interactive Biochemistry Web Links

Selected Research Centers

Brutlag Bioinformatics Group

The Brutlag Bioinformatics Group is a Stanford-based organization with a focus on predicting protein structure and function from primary sequence. Products developed by the group (EMOTIF, EMATRIX, and 3MOTIF) are used in assigning functions to unidentified genomic sequences. Other software products developed include LOCK and 3DSEARCH, which are used for comparing protein structures and searching structural databases.

Munich Information Center for Protein Sequences (MIPS)

MIPS is a bioinformatics center at the Max Planck Institute involved in numerous bioinformation and sequencing projects. The center contributes to the International Protein Sequence Databank, the European Yeast Functional Analysis Program, the European Molecular Biology Network (EMBNET), PEDANT (proteomics tool for genomic analysis), and Protfam (protein family database). Genome sequencing projects include *Arabidopsis thaliana* and *Neurospora crassa*.

National Center for Biotechnology Information (NCBI)

The NCBI is a bioinformatics-based center of the NIH. Its major efforts are creation of public databases, research in computational biology, software for analyzing genomic data, and dissemination of biomedical information. Through Entrez, the NCBI provides access to

Sequence homology

NCBI BLAST server (using NCBI-BLAST 2.1):
Basic search
Advanced search
Blast two nucleotide or protein sequences
BLAST overview
BLAST tutorial

EMBL BLAST servers (using WU-BLAST 2.0):
EBI BLAST
Advanced BLAST search at EMBL
Ensembl BLASTView

Gene expression

NCBI:
Gene Expression Omnibus (GEO)
Serial Analysis of Gene Expression (SAGE)
UniGene

Other sources:
Stanford Microarray Database
ChipDB (Whitehead Institute)
Bodymap (human and mouse expression data)

Protein sequence/ proteomics

ExPASy Servers:
Sequence Retrieval System (SRS) (gateway to ExPASy databanks)
Swiss-Prot/TrEMBL (protein sequences)
Prosite (protein families and domains)
Swiss-2D PAGE (gel electrophoresis database)
ENZYME (enzyme nomenclature)

MIPS:
ATLAS Gateway (multiple protein sequence databases)
MIPS Sequence Retrieval System (SRS)
ProtFam (protein families

Genbank, protein sequences, completed genomes, structure databases, evolutionarily related sequences, taxonomy, and the Online Mendelian Inheritance in Man database of human genes involved in disease.

The Institute for Genomic Research (TIGR)
TIGR is a not-for-profit research institute that has had a significant role in revealing structural, functional, and comparative features of genomes and gene products in dozens of organisms, including humans. The TIGR site provides access to numerous genome sequences, software tools, and microarray information.

Links to other major research centers

Sequence Data

DNA Data Bank of Japan (DDBJ)
The lesser known of the the world public DNA sequence databases, DDBJ is the sole DNA databank in Japan certified to serve as an official repository of DNA information. While the DDBJ serves to collect information from Japanese sources, it also accepts data from other international contributors. DDBJ shares information with both Genbank and EMBL on a daily basis so that the same data can be viewed on any of the services virtually simultaneously.

NCBI-UniGene
NCBI's UniGene system is designed to partition GenBank sequences into a non-redundant set of gene-oriented clusters. The database uses both established genes and novel expressed sequence tag (EST) sequences in defining clusters. Clusters contain unique gene sequences, the tissue types in which the gene has been expressed, and their map locations.

Nucleic Acid Database (NDB)
NDB provides information about nucleic acid structures. The onsite Atlas of Nucleic Acid Containing Structures features forms of DNA (A, B, Z), RNA (A, tRNA, ribozymes), unusual nucleic acids, protein-DNA complexes, and protein-RNA complexes. Searches of NDB and of NMR structural information are provided, as well as the macromolecular Crystallographic Information File (mmCIF) and an extensive collection of links.

and superfamilies)

Other sources:
NCBI protein search
Proteome Databases (Proteome, Inc.)

Protein structure

RCSB Protein Data Bank
3-D protein structure at NCBI
Swiss-Model (automated comparative protein structure modeling from ExPASy)
Swiss-3DImage (automated comparative protein structure modeling from ExPASy
Other protein structure servers

Other databases

Online Mendelian Inheritance in Man (NCBI/Johns Hopkins)
Online Mendelian Inheritance in Animals (University of Syndey)

Auto-Alert Services:

Cite-Track
E-mail updates of new content in *Science* and other journals (*Science* Online)

PubCrawler
Scans daily for updates to NCBI's Medline and GenBank databases (Trinity College, Dublin)

Swiss-Shop
E-mail reports of new SWISS-PROT protein sequence entries relevant to specific fields of interest (ExPASy)

Automated Entrez Queries
User-customized daily, weekly, or monthly queries of various NCBI

Other sites focused on nucleotide, genome, or protein sequence data

---

Expression Data

BodyMap

> BodyMap offers a database of the expression of human and mouse genes, using data generated by random sequencing of cDNAs from carefully isolated tissues. Visitors to the site can select a tissue and get a list of the frequency of occurrence of each mRNA in it.

MIRAGE (Molecular Informatics Resource for the Analysis of Gene Expression)

> A self-described "experimental Web resource" at the Institute For Transcriptional Informatics (IFTI), MIRAGE provides several tools for studying gene expression. They include downloadable software products, an impressive (albeit somewhat disorganized) collection of links, and access to the object-oriented Transcription Factor Database (ooTFD). A particularly strong point of the site is its emphasis on transcription factors images.

MRC HGU Mouse Atlas and Gene Expression Home Page

> A different approach to describing gene expression is provided at the Mouse Atlas and Gene Expression site - 3D models. These are derived from measuring expression in anatomically-based mouse embryonic development. Though still in the construction phase, the site has the potential to expand the dimensions of understanding of gene expression in the mouse.

Stanford Microarray Database

> Microarray technology provides the most comprehensive methodology for studying genome-wide gene expression. The Stanford Microarray Database not only provides access not only to extensive array data in yeast, but is a good source of information about microarray use in general.

Other gene expression and microarray sites

---

databases (Arizona Research Laboratories)

Sequence Alerting System
E-mail updates on homologues of user-specified DNA and protein sequences (EMBL)

Journal Abstracts Delivered Electronically (JADE)
Weekly search of new Medline entries (NCEMI)

BioMail
New references from Medline to users' e-mail accounts, based on user-customized criteria (SUNY, Stony Brook)

Seen anything interesting on the Web in genomics or post-genomics lately? Send us the URL!

SubmitClear

Protein Structure

3DB Browser

Provides an excellent mechanism for accessing Protein Data Bank (PDB) structures. The user interface consists of a search engine with many user-specified criteria, such as keyword, ID, text, or author. Additional user-specified restrictions include the method of structure determination (X-ray, NMR, or theoretical), resolution, organism, deposit date, chain size, and data source. FASTA searches are also available.

CATH Protein Classification

The CATH database provides a view of protein structures in the Brookhaven crystallographic database organized in a hierarchical domain-based fashion. The hierarchy consists of sortings according to class (secondary structure composition), architecture (overall domain structure), topology (fold families), homologous superfamilies (protein domains with common ancestors), and sequence families (sequence identity relationships).

Database of Macromolecular Movements

What PDB is to protein structure and IMB-JENA is to macromolecular visualization, the Database of Macromolecular Movements is to macromolecular motion. Visitors to the Database of Molecular Movements can watch complex molecular motions (primarily in proteins), create movies with the online Morph tool, download software for plotting molecular geometries, and read informative papers on the subject.

PRINTS

As stated at the site, PRINTS is a database of protein fingerprints. Fingerprints comprise groups of conserved motifs in proteins that can be used to characterize a family. Several homology-based mechanisms of searching the database are provided, including BLAST, InterPro, SPRINT (search a PRINTS relational database), in addition to simple retrievals via accession number, text, title, sequence, or author.

SCOP (Structural Classification of Proteins)

The most broadly encompassing and widely used database of protein structure classes, SCOP relies on structural and evolutionary knowledge to identify structural domains. Numerous links and search options are available, including hierarchical searches by class, by entry of a PDB identifier, by sequences corresponding to SCOP domains, by superfamily searches, and others. A site used in virtually every protein structure

analysis.

Other sites on protein structure and function

Proteomics

BCM Search Launcher

Gateway to an extensive set of analysis tools for DNA and protein sequences. The centralized search launcher provides access to standard sequence searchers (multiple BLAST and FASTA algorithms), pattern searches (Prosite, Blocks, COGs), sequence alignments (CLUSTALW, CAP, PIMA), gene feature identification (exon/intron boundaries, promoter/transcription factor binding, open-reading-frame identification), secondary structure prediction (coiled-coils, transmembrane, hydrophilicity/hydrophobicity), and miscellaneous sequence utilities (FASTA format conversion, restriction cutters, etc.). A great one-stop shop.

EMBL DALI

The DALI server provides email- or Web-based querying of an unknown protein structure against the Protein Data Bank structures in 3D. Users submit the coordinates of an unknown and receive back multiple alignments of similar 3D homologs.

InterPro

InterPro attempts to unify information stored in structure databases, such as PROSITE, PRINTS, Pfam, and ProDom, into a format that can be accessed easily from a single site. Strong points include a simple interface, excellent information compilations from many sources, and very helpful hyperlinking. InterPro can be either downloaded by FTP or accessed via the Web at the address above.

NCBI HomoloGene

A variation on identifying orthologs (genes in different species with a common origin) is taken by HomoloGene. Human, mouse, rat, and zebrafish genes in the UniGene and Locuslink were compared for nucleotide sequence similarity. Putative orthologs were then identified as UniGene clusters corresponding to the two sequences that are each other's best match. Datasets from Homogene can be downloaded by FTP.

PredictProtein

PredictProtein provides a portal to common protein databases (EMBL, SWISS-PROT, TREMBL, and PDB) for users to submit unknown sequences for comparison against. Similarity matching methods provided by PredictProtein include BLAST, COILS, ProSite, ProDom, and others. A one-stop shop for working with unknown protein sequences.

Protein Topology (TOPS)

The TOPS language of describing protein structure provides the underpinnings of this site, which allows visitors to search structure databases for structural domains. In addition, users can supply a protein structure/domain and have the TOPS site compare it against all domains in the TOPS Atlas or against the entire PDB set of domains (over 24,000).

Links to other proteomics resources

Medical Resources

GeneCards (human genes, maps, proteins and diseases)

One of the most comprehensive sources of information about genes involved in disease, GeneCards is an elegantly designed site at the Weizmann Institute. The database is accessed by a simple text-based search engine on the opening page or by specification of chromosome region or marker name. Cards of information retrieved by the program contain text with hyperlinks to dozens of other Web analyses for protein structure, relation to disease, sequence, alignments, homologues, and references.

Human Chromosome Launchpad

Designed to provide an easy-to-use interface to human genetic information, the opening page of the Human Chromosome Launchpad presents hyperlinks for each numbered chromosome. Clicking on one of the links brings up an extensive set of hyperlinked information about sequences in that particular chrmosome. While the approach is rather broad-based for gene information, the site functions well at the chromosomal level due to its simple design.

Other resources on genomics and medicine

Page content:
Kevin Ahern,
Oregon State University/
DaVinci Press, Inc.