



Tools

COURSE HOME

Computational Biology Tools and Resources

SYLLABUS

CALENDAR

THIS COURSE AT MIT

READINGS

LECTURE SLIDES

VIDEO LECTURES

RECITATIONS

ASSIGNMENTS

TOOLS	DESCRIPTIONS	REFERENCES
BLAST	Sequence alignments provide a powerful way to compare novel sequences with previously characterized genes. Both functional and evolutionary information can be inferred from well designed queries and alignments. The Basic Local Alignment Search Tool (BLAST) provides a method for rapid searching of nucleotide and protein databases. It is a sequence comparison algorithm optimized for speed used to search sequence databases for optimal local alignments to a query.	Altschul, S. F., W. Gish, et al. " Basic Local Alignment Search Tool ." <i>Journal of Molecular Biology</i> 215, no. 3 (1990): 403–10.
BIND	The Biomolecular Interaction Network Database (BIND) is a collection of records documenting molecular interactions. The contents of BIND include high-throughput data submissions and hand-curated information gathered from the scientific literature.	Bader, G. D., D. Betel, et al. " BIND: The Biomolecular Interaction Network Database ." <i>Nucleic Acids Research</i> 31, no. 1 (2003): 248–50.
Biology WorkBench	The Biology WorkBench is a web-based tool for biologists developed by the San Diego Supercomputer Center at the University of California San Diego. The WorkBench allows biologists to search many popular protein and nucleic acid sequence databases. Database searching is integrated with access to a wide variety of analysis and modeling tools, all within a point and click interface that eliminates file format compatibility problems.	
		Thompson, J. D., D. G. Higgins, et al. " CLUSTAL W: Improving the



ClustalW	ClustalW is a general purpose multiple sequence alignment program for DNA or proteins. It produces biologically meaningful multiple sequence alignments of divergent sequences.	Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-specific Gap Penalties and Weight Matrix Choice. <i>Nucleic Acids Research</i> 22, no. 22 (1994): 4673–80.
DALI	DALI stands for Distance mAtrix aLIgnment. The Dali server is an automatic service for the comparison of protein structure in 3D. You send the coordinates of a query structure and receive a multiple structure alignment in return. You can submit your coordinates either by electronic mail or interactively from the World Wide Web.	Holm, L., and C. Sander. " Mapping the Protein Universe. " <i>Science</i> 273, no. 5275 (1996): 595–602.
Deep View Swiss-Pdb Viewer Tutorial For Deep View (Swiss-PdbViewer)	The Deep View Swiss-PdbViewer is a software application with a user friendly interface that allows one to analyze several proteins at the same time. The proteins can be superimposed in order to deduce structural alignments and compare their active sites or any other relevant parts. Amino acid mutations, H-bonds, angles, and distances between atoms are easy to obtain thanks to the intuitive graphic and menu interface.	Guex, N., and M. C. Peitsch. " SWISS-MODEL and the Swiss-PdbViewer: An Environment for Comparative Protein Modeling. " <i>Electrophoresis</i> 18, no. 15 (1997): 2714–23.
DIPTM	DIP TM stands for Database of Interacting Proteins. The DIP TM database catalogs experimentally determined interactions between proteins. It combines information from a variety of sources to create a single, consistent set of protein-protein interactions.	Salwinski L., C. S. Miller, et al. " The Database of Interacting Proteins: 2004 Update. " <i>Nucleic Acids Research</i> 32, no. 1 (2004): D449–51.
Dot Matrix	Dot or matrix plots provide an easy and powerful means of sequence analysis for searching out regions of similarity in two sequences and repeats within a single sequence.	Maizel, J. V., and R. P. Lenk. " Enhanced Graphic Matrix Analysis of Nucleic Acid and Protein Sequences. " <i>Proceedings of the National Academy Sciences</i> 78, no. 12 (1981): 7665–9. Pustell, J., and F. C. Kafatos. " A High Speed, High Capacity Homology Matrix: Zooming through SV40 and Polyoma. " <i>Nucleic Acids Research</i> 10, no. 15 (1982): 4765–82.

		Quigley, G. J., L. Gehrke, et al. " Computer-aided Nucleic Acid Secondary Structure Modeling Incorporating Enzymatic Digestion Data ." <i>Nucleic Acids Research</i> 12, no. 1 (1984): 347–66.
Entrez NCBI	Entrez is a retrieval system designed for searching several linked databases at NCBI (National Center for Biotechnology Information).	
GENSCAN	GENSCAN predicts the locations and exon-intron structures of genes in genomic sequences from a variety of organisms. GENSCAN was developed by Prof. Chris Burge while he was in the research group of Samuel Karlin, Department of Mathematics, Stanford University.	Burge, C., and S. Karlin. " Prediction of Complete Gene Structures in Human Genomic DNA ." <i>Journal of Molecular Biology</i> 268, no. 1 (1997): 78–94.
Gibbs Motif Sampler	The Gibbs Motif Sampler will allow you to identify motifs, conserved regions, in DNA or protein sequences.	This software was developed by Eric C. Rouchka and Bill Thompson based on work by C. E. Lawrence, J. S. Liu, A. F. Neuwald and others (References) as part of the Bayesian Bioinformatics Program at the Biometrics Laboratory of Wadsworth Center.
MEME	Discover motifs (highly conserved regions) in groups of related DNA or protein sequences using MEME.	Bailey, Timothy L., and Charles Elkan. " Fitting a Mixture Model by Expectation Maximization to Discover Motifs in Biopolymers ." <i>Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology</i> 2 (1994): 28–36.
MODBASE	MODBASE is a database of annotated comparative protein structure models and associated resources.	Pieper U., N. Eswar, et al. " MODBASE: A Database of Annotated Comparative Protein Structure Models, and Associated Resources ." <i>Nucleic Acids</i>

		Research 34, no. 1 (2004): D291-5.
PDB	The Protein Data Bank (PDB) is the single worldwide repository for the processing and distribution of 3-D biological macromolecular structure data.	Berman, H. M., J. Westbrook, et al. " The Protein Data Bank. " <i>Nucleic Acids Research</i> 28, no. 1 (2000): 235-42.
PHYLIP	PHYLIP (the PHYLogeny Inference Package) is a package of programs for inferring phylogenies (evolutionary trees). It is available free over the Internet, and written to work on as many different kinds of computer systems as possible.	
NCBI	Established in 1988 as a national resource for molecular biology information, NCBI (National Center for Biotechnology Information) creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information—all for the better understanding of molecular processes affecting human health and disease. This web site provides access to a myriad of biological sequence databases, structural databases, bioinformatics tools, and literature search tools.	
Python Scripting Language	A scripting language widely used, along with PERL, for bioinformatics and computational biology.	
RasMol	RasMol is a molecular graphics program intended for the visualization of proteins, nucleic acids and small molecules. The program reads in molecular co-ordinate files and interactively displays the molecule on the screen in a variety of representations and colour schemes.	Sayle R., and E. James Milner-White. " RasMol: Biomolecular Graphics for All. " <i>Trends in Biochemical Sciences</i> 20, no. 9 (1995): 374-6.
Scansite	Scansite searches for motifs within proteins that are likely to be phosphorylated by specific protein kinases or bind to domains such as SH2 domains, 14-3-3 domains or PDZ domains.	Songyang, Z., S. Blechner, et al. " Use of an Oriented Peptide Library to Determine the Optimal Substrates of Protein Kinases. " <i>Current Biology</i> 4, no. 11 (1994): 973-82. Yaffe, M. B., G. G. Leparc, et al. " A Motif-based Profile Scanning Approach for Genome-wide Prediction of Signaling Pathways. " <i>Nature Biotechnology</i> 19, no. 4 (2001): 348-53.

		Obenauer, J. C., L. C. Cantley, et al. " Scansite 2.0: Proteome-wide Prediction of Cell Signaling Interactions using Short Sequence Motifs. " <i>Nucleic Acids Research</i> 31, no. 13 (2003): 3635–41.
SCOP	Nearly all proteins have structural similarities with other proteins and, in some of these cases, share a common evolutionary origin. The Structural Classification of Proteins (SCOP) database, created by manual inspection and abetted by a battery of automated methods, aims to provide a detailed and comprehensive description of the structural and evolutionary relationships between all proteins whose structure is known. As such, it provides a broad survey of all known protein folds, detailed information about the close relatives of any particular protein, and a framework for future research and classification.	Murzin, A. G., S. E. Brenner, et al. " SCOP: A Structural Classification of Proteins Database for the Investigation of Sequences and Structures. " <i>Journal of Molecular Biology</i> 247, no. 4 (1995): 536–40.
TMHMM	This software tool was developed by the Center for Biological Sequence Analysis at the Technical University of Denmark and is used to predict transmembrane helices in protein sequences.	Krogh, A., B. Larsson, B., et al. " Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes. " <i>Journal of Molecular Biology</i> 305, no. 3 (2001): 567–80.

FIND COURSES

- » [Find by Topic](#)
- » [Find by Course Number](#)
- » [Find by Department](#)
- » [Audio/Video Courses](#)
- » [Courses with Subtitles](#)
- » [Online Textbooks](#)
- » [New Courses](#)
- » [Most Visited Courses](#)
- » [OCW Scholar Courses](#)
- » [This Course at MIT](#)
- » [Supplemental Resources](#)

ABOUT

- » [About OpenCourseWare](#)
- » [Site Stats](#)
- » [OCW Stories](#)
- » [Media Coverage](#)
- » [New sletter](#)
- » [Press Releases](#)

DONATE

- » [Make a Donation](#)
- » [Why Donate?](#)
- » [Our Supporters](#)
- » [Other Ways to Contribute](#)
- » [Shop OCW](#)
- » [Become a Corporate Sponsor](#)

FEATURED SITES

- » [Highlights for High School](#)
- » [OCW Educator](#)
- » [MIT Crosslinks and OCW](#)
- » [MITx Courses on edX](#)
- » [Teaching Excellence at MIT](#)
- » [Open Education Consortium](#)

TOOLS

- » [Help & FAQs](#)

OUR CORPORATE SUPPORTERS



» Supplemental Resources

» Translated Courses

» View All Courses

» Help & Feedback

» Contact Us


» Advanced Search

» Site Map

» Privacy & Terms of Use

» RSS Feeds



Support for MIT OPENCOURSEWARE'S 15th anniversary is provided by 

ABOUT MIT OPENCOURSEWARE

MIT OpenCourseWare makes the materials used in the teaching of almost all of MIT's subjects available on the Web, free of charge. With more than 2,200 courses available, OCW is delivering on the promise of open sharing of knowledge. [Learn more »](#)



© 2001–2016
Massachusetts Institute of Technology

Your use of the MIT OpenCourseWare site and materials is subject to our [Creative Commons License](#) and other [terms of use](#).

[Need help with this content?](#) [Don't show me this again](#)