Web alert

Protein structures

Chemistry & Biology June 1998, 5:R149–R150

 ISSN 1074-5521

For many years a protein's atomic structure could only be visualised by looking at the images found in journal articles, or using specialised molecular graphics workstations. Both chemists and biologists frequently find themselves studying processes that involve proteins, and there are numerous resources on the web that are useful not only for X-ray crystallographers, nuclear magnetic resonance (NMR) spectroscopists and molecular modellers, but also for nonspecialists who are interested in a given protein's structure.

For the nonspecialist, there are collections on the web of static protein images, some with detailed annotations, but in addition the web now offers everyone the opportunity to view the structures of proteins interactively, using a protein's atomic coordinates to produce images that show, for example, its secondary structure or its backbone atoms, and these images can be rotated and zoomed as the user wishes. For this the user requires both the three-dimensional atomic coordinates for the protein, as determined by X-ray crystallography or NMR, and display software which can be used on a desktop computer, often in conjunction with your web browser. In addition, interlinking of web sites means that a search for a protein by name or sequence will produce links to databases of literature references, sequences, medical information, metabolic pathways and protein structure classification, as well as data on related proteins.

Sources of protein structure information

This list provides a starting point for exploration of protein structure on the web. There are hundreds of sites, some providing a broad spectrum of information and others dealing with particular types of protein, such as proteases or receptors.

The Brookhaven Protein Data Bank (PDB; http://www.pdb.bnl.gov) is the primary store of experimentally determined atomic coordinates of proteins. The PDB website allows for searching using, for example, the molecule's name or its PDB code (a unique four-character identifier). The complete entry, including coordinates and annotations, can be retrieved in 'PDB format', or using several modes of visualisation, including VRML (virtual reality markup language), Chime and RasMol. There are a growing number of links to the relevant entries in other databases — of sequences (Swiss-Prot) or domain definitions (3-Dee), and CATH (the CATH protein structure classification), for example. The PDB has mirror sites worldwide.

The NDB Atlas of Nucleic Acid Containing Structures (http://ndb-mirror-2.rutgers.edu/NDB/NDBATLAS/index.html) contains three-dimensional coordinates of structures containing nucleic acids, including protein–DNA and protein–RNA complexes together with searching tools and links to other macromolecular databases.

SCOP (Structural classification of proteins; http://scop.mrc-lmb.cam.ac.uk) is a database of protein structural domains. The SCOP site allows a protein structure to be downloaded in the original PDB format for use locally, but it also allows the structure to be viewed using Chime, RasMol, or the PDB3D Java viewer, VRML. The SCOP site also allows the searching for proteins with similar amino-acid sequences using the Basic alignment search tool (BLAST) algorithm.

FSSP (Fold classification based on structure–structure alignment of proteins; http://croma.ebi.ac.uk/dali/fssp/fssp.html) is based on an all-against-all comparison of structures in the PDB. The results of a search (by protein name or PDB code) include the ability to view structure neighbours, sequence neighbours and similar structures superimposed in three dimensions.

The National Center For Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov) at the National Institutes of Health (NIH) is home to many databases that can all be searched using the ENTREZ engine. In this way, protein sequence data, protein structures, and MEDLINE bibliographic data are all integrated. The Molecular Modeling Database (MMDB) provides a three-dimensional structural viewer (Cn3D) for viewing structures within ENTREZ. The NCBI site also provides the Vector Alignment Tool (VAST), allowing you to find and view similar structures, and BLAST to find similar sequences.

ExPASy (http://expasy.hcuge.ch/www/expasy-top.html) is the website of the Swiss Institute of Bioinformatics, and it covers all aspects of protein sequence and structure. The SWISS-PROT annotated sequence database, the SWISS-3DIMAGE database of macromolecules and the SWISS-MODEL repository of automatically generated protein models are some of the possible starting points for browsing molecular data. Online tools include a protein structure display program (Swiss-PdbViewer, for Macintosh or PC) and an automated knowledge-based protein modelling server.

The Center for Molecular Modelling at the NIH (http://cmm.info.nih.gov/modeling/) provides links to resources for molecular modelling, such as research tools on the web, information on hardware and software and links to related modelling sites. The 'Molecules R Us' utility allows text searches of PDB.

The IMB Jena Biocomputing Group Image Library of Biological Macromolecules (http://www.imb-jena.de/IMAGE.html) is an expanding image library that contains at least one automatically generated image, as well as RasMol and WebMol interfaces, for every entry in the PDB.

The Biomolecular Structure and Modelling Group at UCL (http://www.biochem.ucl.ac.uk/bsm) brings together people who derive protein and carbohydrate structure information using crystallography or NMR, those who manage databases of this information, and analysts and modellers who examine these structures to try to derive principles of protein folding. The group's structural databases include the CATH protein structure classification and an associated glossary of terms used in protein structure description.

The Protein Society (http://www.faseb.org/protein) has links to many web sites of interest to protein scientists, and it houses a large collection of kinemages. The kinemage database can be searched against a protein's name or function.

The CMS–SDSC Molecular Biology Resource (http://www.sdsc.edu/projects/ResTools/). The San Diego Computer Center's large listing of tools for bioscientists, including on-line molecular viewing, and protein structure. An excellent starting point for exploring web resources for molecular biology and modelling.

Technology

As well as databases and collections of links, the web can provide programs that can be downloaded and used for viewing molecules. This list is by no means exhaustive; there are several Java viewers, for example, and programs for displaying VRML files.

RasMol (http://www.umass.edu/microbio/rasmol/) is a freely available program that produces interactive images from PDB coordinate files, and runs on Macintosh, PC and Unix systems. In addition to rotating and zooming into regions of interest, RasMol allows users to choose between several different depictions of proteins, including ribbon diagrams, Richardson-type diagrams, and space-filling models. RasMol can be run as a stand-alone program, or web browsers can be configured to use RasMol as a helper application, and sites such as SCOP serve RasMol scripts which contain instructions to highlight the particular domain of interest. The RasMol homepage also gives links to sites from which coordinate files for macromolecules and small molecules can be obtained.

Chime (http://www.mdli.com/tech/chemscape.html) is a plug-in for web browsers that allows interactive RasMol-like images to be embedded within web pages, and is a popular tool for displaying molecular structure on the web. The plug-in can be downloaded free from MDL Information Systems, Inc., or from many web sites including the RasMol homepage.

The kinemage (http://www.faseb.org/protein/kinemages/kinpage.html) is a form of interactive protein display pioneered by the journal *Protein Science* that allows for flexible display and annotation of protein features. Versions of both the viewer (for most platforms) and a program to create kinemages, are free, and instructions on how to configure web browsers to make use of the viewer are offered.

Matthew Day, Garland Publishing, 1 Gunpowder Square, London, EC4A 3DE, UK; matthew.day@tandf.co.uk. Judith Murray-Rust, Department of Crystallography, Birkbeck College, London, WC1 7HX, UK; ubcg09j@mail.cryst.bbk.ac.uk.