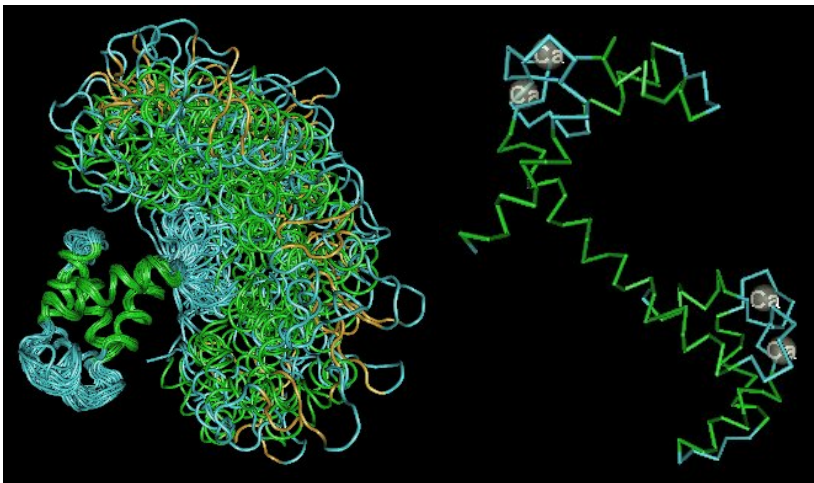


Molecular Structure Viewing NCBI's MMDB and Cn3D

Proteins fulfill a wide range of biological functions which depend upon their 3-D structures. The structures of proteins can be experimentally determined at the atomic level by two different methods, X-ray crystallography and NMR (Nuclear Magnetic Resonance) Spectroscopy. Both techniques have allowed scientists to create atomic models for protein structures that can be viewed and manipulated. Also, computer methods exist for predicting a structure from the amino acid sequence, however, they are not yet reliable. The most effective form of computational analysis models a protein structure based on similarity to another experimentally-derived protein structure.

X-Ray Crystallography - Provides the highest resolution. Requires crystallization of protein and usually gives only one model of structure. May be automated in future.
NMR - Allows structure determination of protein in solution. Variability of solution conditions possible. Provides characterization of intrinsic protein motion in solution.
Computation - Simulates the action of the forces that act on each atom in a molecule of known composition and approximates structure. Produces non-experimental models. Fast, but presently least reliable.



Left: Structure of Calmodulin (1DMO) showing 30 models.

Right: X-Ray Crystallography structure (3CLN) of Calmodulin.

Regardless of the method used, no protein structure is properly represented by a single conformation. Every protein molecule has intrinsic motion and will shift and change based on molecular properties and environmental conditions. When you view a single structure, you are actually viewing a single snapshot of the protein in time.

Introducing Entrez Structures

Entrez Structures is the section of the Entrez system that allows the user to specifically search for molecular structure records in the Molecular Modeling DataBase (MMDB).

Access:

- From NCBI home page - click on 'structures' in top bar.
- From site map - click on structures, then click on structures home
- Links from other database records
 - Entrez system: structures link option
 - Entrez system: drop-down menu structures link

Searching:

- Keyword
- Fields: accession number, all fields, author name, EC/RN number, filter, issue, journal name, modification date, organism, page number, publication date, substance name, text word, UID, and volume
- Limits screen
- PDB/MMDB Ids

Introducing MMDB

The MMDB (Molecular Modeling DataBase) contains experimentally-determined 3-dimensional molecular structures, primarily proteins, also including nucleic acids. It is a curated subset of a database called PDB (Protein DataBank). All MMDB records have been checked, annotated, and stored in an ASN.1 format. The structure coordinates in PDB/MMDB records have been obtained experimentally by scientists using X-Ray Crystallography and NMR. The structural data in MMDB has been cross-linked with bibliographic information, the sequence databases, and the NCBI taxonomy.

Each protein structure in the MMDB database is represented by an MMDB Structure Summary page. This page provides general information about the sequence including references, taxonomy, protein domains, links to sequences, access to sequence and structural neighbors, and visualization options.

The screenshot shows the NCBI MMDB Structure Summary page for Calmodulin. The page includes the NCBI logo, navigation tabs (PubMed, BLAST, Structure, Taxonomy, OMIM, Help?, Cn3d), and a description: "Calmodulin, Nmr, 30 Structures." The deposition information is "M.Zhang, T.Tanaka & M.Ikura, 24-Apr-96" and the taxonomy is "Xenopus laevis". The reference is "PubMed MMDB: 4925 PDB: 1DMO". There are buttons for "View 3D Structure" and "Display". Below the navigation is a protein domain diagram showing two chains (1 and 2) with four ehand domains each. The protein sequence is shown as a bar with residue numbers 1, 20, 40, 60, 80, 100, 120, and 140.

PDB - The Protein Database (<http://www.rcsb.org/pdb/>)

PDB (Protein DataBank) is an international database of 3-D biological macromolecular structures. It is maintained by a nonprofit organization, the Research Collaboratory for Structural Bioinformatics (RCSB), associated with Rutgers University, San Diego Supercomputer Center, and the Biotechnology Division of the National Institute of Standards and Technology. There are multiple mirror sites worldwide. This is a public free-access database that contains molecular structures, proteins and nucleic acids, primarily structures experimentally-derived by X-Ray crystallography and NMR. Data submitted to PDB is validated prior to complete entry.

How do MMDB and PDB differ?

- MMDB is a subset of PDB that undergoes additional curation and modification
- MMDB uses the ASN.1 format only. PDB supports several data formats.
- PDB can export a file in PDB or mmCIF formats. MMDB exports ASN.1 as well as PDB or Kinemage.
- MMDB is integrated (interlinked) with Entrez; PDB is not.
- PDB includes some theoretical models, though they are not encouraged. MMDB does not include any theoretical models.
- PDB files do not contain explicit bond information for biopolymers but MMDB does.

Introducing Cn3D

Cn3D is a helper application for a web browser (ie Netscape or Internet Explorer) that allows viewing of 3-D structures from NCBI's structure database. It is available for free download from the NCBI web site:

<http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml>

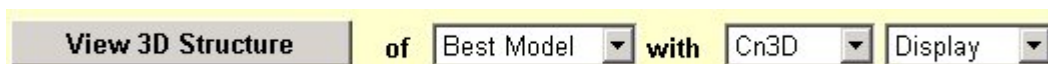
Installing Cn3D

1. Choose the correct operating system for your computer and click on the appropriate link: PC, Mac or UNIX.
2. Follow the directions to download the file and install it.
3. Save the file to your browser directory.
4. Locate it with your file manager program
5. Activate the file (usually by double-clicking on it).
6. Cn3D should automatically install itself and connect to your browser

The latest version of Cn3D available is version 4.0 which was first made available in June 2002.

How to view a structure with Cn3D

In the MMDB structure summary, select the appropriate viewing options in the dropdown menus and then click on the "View 3D Structure" button.



- The first dropdown menu allows you to select whether to view a single or multiple model of the protein structure. The options are: Best Model, Virtual Bond, All Models.
 - The second menu allows you to select the structure viewing software. The choices are Cn3D, RasMol, and MAGE. The appropriate software must be already loaded on your computer.
 - The third menu allows you to choose between viewing the 3D structure, viewing the computer file that contains the 3D atom coordinates, or saving the file.
- The default settings are Best Model, view with Cn3D and Display 3D structure.

Manipulating the Structure Viewer Window

A large number of options are available for manipulating the structure viewer. Here are a few key points:

- **Rotating and Moving** - The structure can be rotated around its center simply by clicking and dragging on the mouse. Speed of rotation will vary depending on the power of the machine, the display format and the structure size. The structure can be moved within the window by holding down the 'shift' key while clicking and dragging on the mouse.
- **Zooming** - Zooming in can be done either using the View - Zoom In menu option or by using the "z" key to zoom in and "x" key to zoom out. Another alternative zooming method is to hold down the 'ctrl' key and click/drag the mouse.
- **Highlighting** - The two windows (structure and sequence) are interactive. Highlighting in one window will also highlight the corresponding sequence/structure in the other window.
- **Restore and Reset** - Restore returns the display to the last saved version of the structure. If no changes have been saved since opening the file, it will return to the original display. Reset will bring the entire structure back into view.

The View Menu

The View Menu contains the following options. Note that many of the options have keyboard shortcuts available.

- Zoom in "z"/Zoom out "x"
- Restore and Reset
- Animation Options including next frame, previous frame, first frame, last frame and all frames. These allow the user to move between individual models of a protein structure, one per frame.
- Animation Options including play frames, spin, stop and set delay. Play frames will show the models of a protein structure in a continuous, cycling movie-like form. Spin will spin the shown structure model in a continuous circle. Stop will freeze the play or spin. Set delay allows you to slow down or speed up the rotation rate.

Cn3D Show/Hide Menu

- Pick Structures: Opens a dialog box that allows you to select which protein chains are viewed.
- Show Everything: Makes all parts of a structure visible
- Show Aligned Domains: Shows protein domains that contain aligned residues. Useful when viewing multidomain structures.
- Show Aligned Residues: Shows only the aligned residues. Useful for viewing the conserved part of an alignment.
- Show Selected Residues: Shows only highlighted parts of the alignment
- Unaligned Residues: Shows only the unaligned regions. Useful for viewing the more variable/unconserved parts of the alignment. You can choose to view unaligned domains or unaligned residues.
- Select by Distance: Allows you to highlight residues based on distance between atoms. Enter a distance and it will only show those residues within that distance from a designated residue.

Cn3D Style Menu

The Style Menu changes the spatial appearance of the current structure. It includes a variety of rendering and color options.

- Edit Global Style: Use this to edit the overall style that applies to all imported structures. You can use it to change the background color, the default rendering and color schemes, and what elements of the structure are initially shown. The label tab can be used to apply labels to amino acids, and the details tab controls the sizes of structure elements. Click on 'apply' or 'done' to view the changes made.
- Favorites: This allows you to save global styles that you have created with the global style editor above. Once saved, they can be re-applied to a new structure at any time.
- Rendering shortcuts: Use this to select one of the following drawing styles: worms, tubes, wire, ball and stick, space fill.
- Color shortcuts: This provides access to a wide variety of coloring options including element, molecule (chain), domain, temperature, hydrophobicity, secondary structure, charge, and alignment.
- Annotate: This allows you to create specific annotation for chosen areas of the structure. The part of the structure to be annotated must be selected (highlighted in yellow) prior to creating the annotation. Click on 'new', then give the annotation a name and description. Then use the 'edit style' button to create the annotation. The speed of the animation is dependent on the complexity of the scene that Cn3D has to draw. For example, wireframe is fast but tubular representations are slower.

Cn3D Sequence Window

Cn3D's second window contains the sequence of the displayed structure. The structure and sequence windows are linked, such that highlighting residues in one window causes the same residues to be highlighted in the other. A second function of the sequence

window is to hold sequences that have been aligned with the sequence of the displayed structure or the sequences that correspond with aligned structures. Multiple sequences can be displayed and manipulated simultaneously.



Cn3D Import Window

New in Version 4.0 of Cn3D, the Import Window provides a wealth of options for importing, aligning, and manipulating sequences within Cn3D. Multiple sequences can be imported and then aligned using either the standard BLAST algorithm or using a new threading algorithm that combines stored structural information with sequence information to make an alignment. Once the alignments are created, they can then be moved to the sequence and structure windows for viewing and further manipulation.



Viewing alignments

Both sequence and structure alignments can be created and viewed in Cn3D. Sequence and structure alignments are different. An individual protein may exist in more than one structural conformation based on environmental conditions or binding to other molecules. Two proteins may have a similar sequence but different structures. Two proteins may have a similar structure but different sequences.

Sequence alignments are based on BLAST comparisons of sequence identity and similarity.
Structure alignments are based on VAST comparisons of structural (spatial) elements.

Making a Sequence Alignment in Cn3D

The following is the basic procedure for making an alignment of a single sequence against the sequence of a displayed structure.

1. Display the structure in Cn3D.

2. Select the "Show Imports" option from the Imports Menu of the Sequence Window.
3. Select "Import Sequences" option from the Edit Menu of the Import Window
4. Choose the type of import - from a FASTA file or from Entrez with a gi or accession number. Note that if the molecule has multiple protein chains, you must first select which chain will be used for the alignment.
5. Upload the sequence either by finding the file or by entering the gi or accession number.
6. Once you see the sequence imported, go to the Algorithms menu and select a method of alignment.
7. Click on the imported sequence. Once it is aligned, it should show capital letters.
8. Select "Merge Single" from the Alignments menu and then click on the imported sequence. The Import Window should become empty and the alignment should be displayed in the Sequence and Structure windows.

Making a Sequence Alignment Using BLINK

This is based on the ability of BLINK (BLAST Link) to display the sequences within a precomputed BLAST search that have known structures.

1. Locate the GenPept or LocusLink record for the protein of interest (for which a structure is not yet been solved).
2. Click on the "BL" link to open BLINK.
3. Click on the "3D Structures" button to display only those matched sequences that have known structures.
4. Locate an interesting match that has an associated structure and click on the blue dot for that match.
5. Cn3D will automatically open and display the alignment.

About Structure Alignments

Structural alignments can only be constructed and viewed if the structures of both molecules have already been experimentally determined and entered into MMDB. A VAST comparison is automatically done when a structure is entered into MMDB which compares the new structure to all others in MMDB and the results are stored. Cn3D and VAST do not model/predict a 3-D structure from a protein sequence. Structural alignments can't be constructed 'on-the-fly' the way that BLAST can do a sequence alignment on order. All VAST structural alignments have been pre-computed and stored in the database to be retrieved. The exception is VASTSearch which only applies to new sequences that have not yet been entered into MMDB.

Remember: Only if both structures have been experimentally solved is it possible to view a true structural alignment in Cn3D. To view a true structure alignment, it is necessary to use the structure neighbors option and locate the second structure in the list of structure neighbors.

About VAST (Vector Alignment Search Tool)

VAST is a software program that compares three dimensional molecular structures for structural similarity. This is different from BLAST which does linear sequence alignments. As already noted, VAST comparisons are automatically done for every structure in MMDB and that information is made available through a link in each MMDB structure summary, the 'structure neighbors' link. Note that VAST can only compare known (experimentally-determined) 3D structures. VAST does NOT do structure predictions for linear protein sequences.

Structure Neighbors

From MMDB's structure summary pages, lists of structure neighbors are available by clicking on the protein chain in the graphic portion of the MMDB structure summary. The VAST-result pages then present a tabulated listing of the neighboring structures indexed by their four-character PDB-Identifiers and sorted by VAST similarity scores.

Viewing Sequence Neighbors/VAST results

- The VAST-result pages display a tabulated listing of the neighboring structures.
- The matches are sorted by their VAST similarity score.
- Each match is identified by its PDB and MMDB numbers and a textual description.
- Individual structure neighbors can be selected by clicking in the check-boxes at the left margin.
- If the button labelled 'View/Save Alignments' is clicked, structural superpositions of the parent protein with selected neighbors are displayed using either Cn3D or Mage as mime-typed viewers. Rasmol will not do this.
- These alignments can be saved to disk.

Manipulating VAST Results

The listing of VAST results can be modified by these options:

- Display Subset - specifies what level of sequence redundancy should be allowed in defining neighbors. As you go down the list of available display subsets, you will display more sequences. To find a specific structure, it is usually most efficient to choose the 'All of MMDB' option and then scroll through the results pages to find it. See <http://www.ncbi.nlm.nih.gov/Structure/VAST/vasthelp.html#HowNR> for more detailed information.
- Sort - Specifies the criterion for ordering the structure neighbors in the table. The default is to sort by VAST alignment score.
- Column Format - specifies which columns will be included in the table. Adjust this to see additional columns or remove columns.

About VAST Search

VAST Search allows the user to submit a structure to NCBI's structure-structure similarity search service. It compares 3D coordinates of a newly determined protein structure to those in the MMDB/PDB database and computes a list of structure neighbors based on a set of structure coordinates supplied by the user. VastSearch will not predict a structure for a linear protein sequence. It requires the user to submit the full 3-D coordinates for an experimentally-resolved molecular structure and will take an extensive time period relative to a BLAST search. The amount of time needed to complete a search depends on the size of the query protein and the number of potential structure neighbors of the query protein.

VASTSearch supplies the user with an ID and password in order to protect the privacy of individual searches. This VAST ID and password must be used to retrieve VASTsearch results.

The submitted structure must be in PDB format. PDB format can be generated using X-PLOR or Brookhaven PDB. Approximately 80% of PDB files have been generated with X-PLOR. The Brookhaven PDB format is slightly different from X-PLOR PDB format. X-PLOR can be obtained free online.

Other Structure Viewers

There are several other molecular structure viewer programs available besides Cn3D. The most well known of these are

- RasMol <http://www.umass.edu/microbio/rasmol/>
- Kinemage <http://kinemage.biochem.duke.edu/kinemage/kinemage.html>
- CHIME
[http://www.mdli.com/cgi/dynamic/product.html?uid=\\$uid&key=\\$key&id=6](http://www.mdli.com/cgi/dynamic/product.html?uid=$uid&key=$key&id=6)
- MolMol <http://www.mol.biol.ethz.ch/wuthrich/software/molmol/>
- Visual Molecular Dynamics (VMD) <http://www.ks.uiuc.edu/Research/vmd/>
- JAVA Molecular Viewer <http://www.ks.uiuc.edu/Development/jmv/>

These software programs are available free online. In general, these programs all do the same thing. They differ much like Netscape and Internet Explorer differ from each other - they use different commands and menus to accomplish the same or very similar tasks. Some have unique features that may make them particularly useful in a certain situation. The advantage to using Cn3D is that it is designed to interface well with the NCBI Entrez system.