

The Molecular Modeling Database (MMDB) – Entrez Structures

<http://www.ncbi.nlm.nih.gov/Structure/MMDB/mmdb.shtml>

A subset of the Brookhaven Protein DataBank (PDB), MMDB contains thousands of three-dimensional biomolecular structures (primarily proteins) derived from X-ray crystallography and NMR-spectroscopy. All of the information entered into MMDB from PDB has been re-validated and reorganized to provide cross-referencing between the chemistry and the three-dimensional structure of macromolecules, through specification of a biopolymer's spatial structure, its chemical organization, and links between the two. Through the integration of chemical, sequence, and structure information, MMDB provides a resource for structure-based homology modeling and protein structure prediction. All MMDB records are stored in ASN.1 format, but can be graphically displayed with the Cn3D, Rasmol, or Kinemage viewers.

Because MMDB is built into the NCBI's Entrez Search System, it can be directly searched by many of the same methods used to search GenBank, PubMed, and OMIM. Queries can use keywords (text terms) that occur anywhere in the structure description record, specific fields such as author names, or PDB 4-character codes and numerical MMDB-Ids. Alternatively, because of the high level of integration among the Entrez databases, MMDB structures can be found through searches of the Entrez nucleotide and protein sections.

To search by protein similarity:

- Find the protein sequence of interest in the Entrez's protein database
- If the protein has a link to "Related Structures," click on it.
- Filter the list to show only proteins with links to structures. To do this, select "Structure Links" in the dropdown text-box next to the "Display" button, and then click on the "Display" button.
- If no structures are displayed, click on the "Related Sequences" link to retrieve a list of all similar sequences and repeat the prior two steps.
- Once a 3-D structure has been identified that is related to the protein of interest, use the NCBI's sequence-structure viewer, Cn3D to view the structure and import the sequence for comparison into the Cn3d sequence window.

A similar process can be used for nucleotide searches in GenBank.

The determination of structural homology is done using an algorithm called **VAST (Vector Alignment Search Tool)**. VAST identifies similar protein 3-D structures. Each of the more than 18,000 structural domains in MMDB has been automatically compared to every other one. From the MMDB summary page for each structure, a link to pre-determined structure neighbors is available.

Also available is the **VASTsearch** which allows scientists to compare a new structure with the MMDB database by entering a set of 3D-coordinates. This is used for newly determined protein structures that have not, yet, been entered into MMDB. Structure neighbors for proteins already in MMDB have been pre-computed and can simply be looked up as described above.

The Structure Summary Page

NCBI MMDB STRUCTURE SUMMARY Entrez ?

MMDB Id: [4185](#) PDB Id: [1EDN](#)

Protein Chains: [\(single chain\)](#)

MEDLINE: [PubMed](#)

Taxonomy: [Homo sapiens](#)

PDB Authors: B.A. Wallace & R.W. Janes

PDB Deposition: 19-Sep-94

PDB Class: Vasoconstrictor

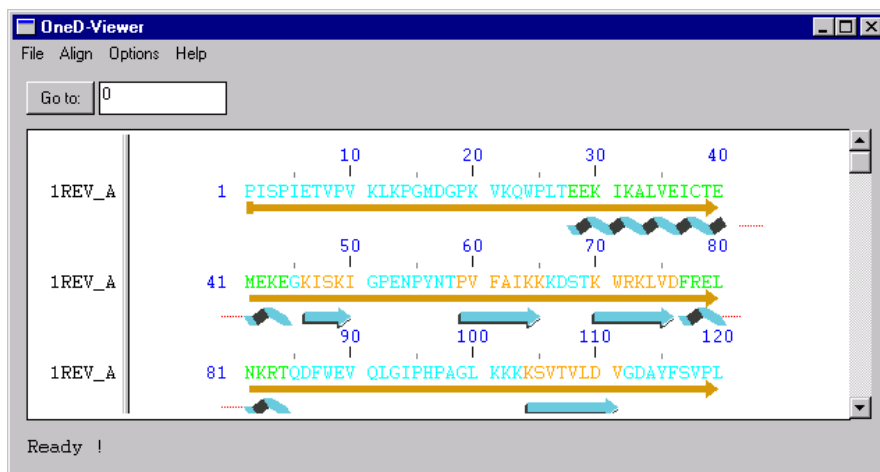
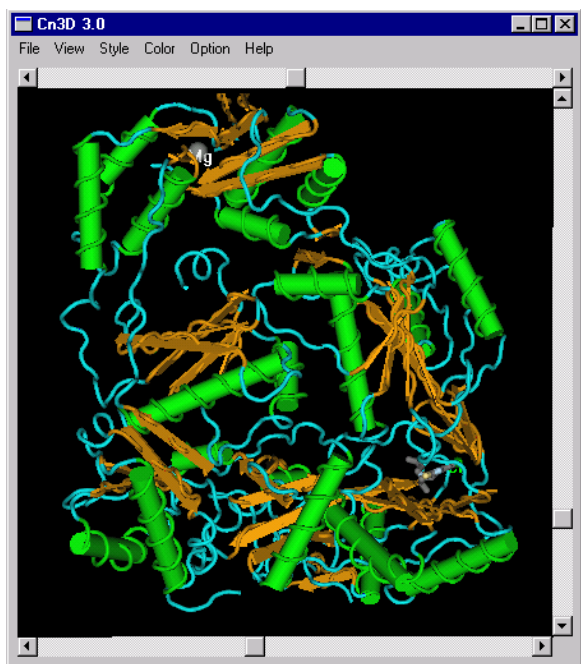
PDB Title: Human Endothelin-1

Sequence Neighbors: [\(single chain\)](#)

Contained on these pages are:

- MMDB identifier – a unique reference ID number given on entry into MMDB
- PDB identifier – the corresponding ID number from the original PDB record, if available
- Links to Entrez protein (GenPept), GenBank, MEDLINE and Taxonomy databases if available
- Authors (from the PDB record)
- Date of deposition into PDB
- PDB Classification - type of molecule
- PDB COMPND – macromolecular content of entry
- Sequence Neighbors – related sequences based on BLAST comparison
- Structure Neighbors – related structures based on VAST comparison
- Structure Viewing Options

While the actual structure record contains a long list of 3-D coordinates for each atom in the molecule, these are not very helpful to view. Therefore, specialized graphic viewer programs have been developed to allow users to see and manipulate the structure in 3-D. The NCBI provides one such structure viewer called Cn3D ('See in 3-D') for free download. It acts as a 'plug-in' to the browser program and is available for both Mac and PC platforms.



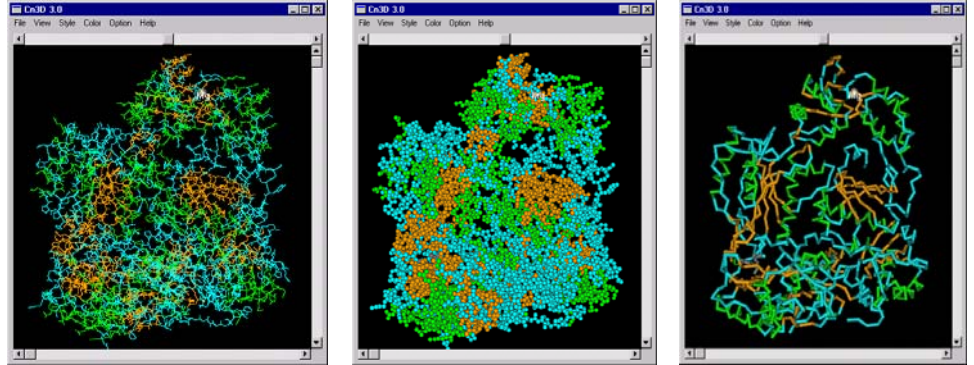
(The HIV-1 Reverse Transcriptase)

Cn3D provides 2 windows: one that shows the structure and another that shows the linear sequence. These are linked such that highlighting part of the sequence in the sequence window will automatically highlight the corresponding part of the structure.

The structure window is interactive, allowing the user to move the structure by clicking and dragging on the mouse. It can be rotated in all directions. The menus at the top of the window allow the form and color of the structure to be altered.

There are seven styles of display:

- Secondary structure (default)
- Neighbor
- Wire Frame
- Tubular
- Hierarchy
- Spacefill
- Ball and Stick



In Cn3D, color can be used to highlight various features of the sequence. Color can be assigned to units of secondary structure, individual residues, interesting parts of the sequence, levels of conservation of sequence, and levels of hydrophobicity.

Further, Cn3D can be used to display not just individual structures but also alignments of multiple structures based on VAST comparisons. Alignments can be created by using the 'structure neighbors' link in an MMDB structure summary or by directly importing additional sequences. The NCBI provides an excellent tutorial on the use of Cn3D at <http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3dtut.shtml>.