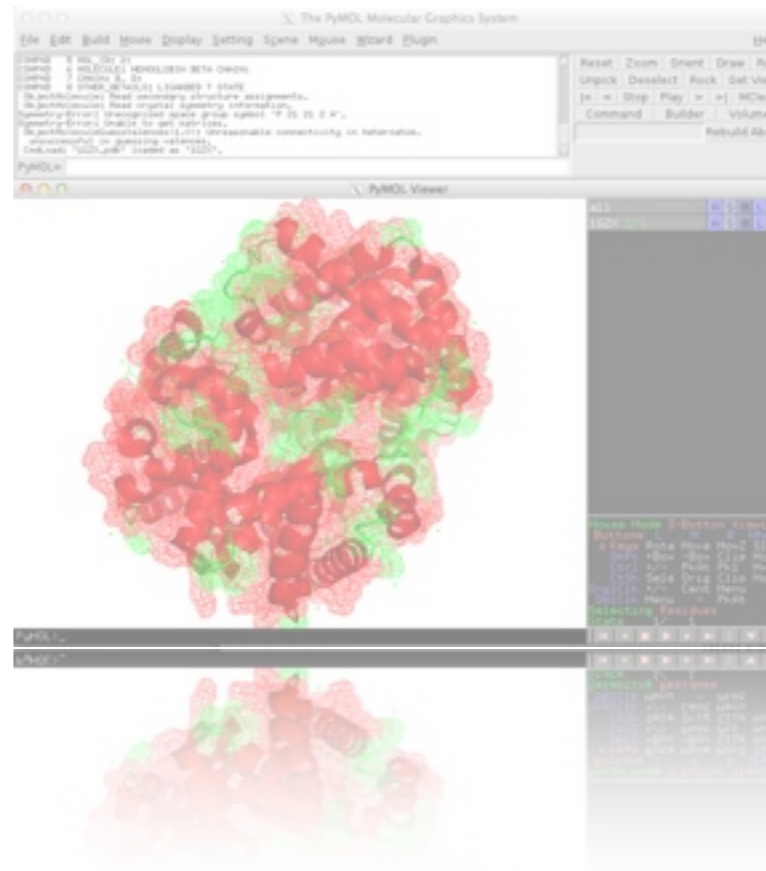


Structural Genomics

Protein data bases and viewers



Davide Baù

<http://sgu.bioinfo.cipf.es>

Structural Genomics Laboratory
Bioinformatics & Genomics Department
Prince Felipe Research Center (CIPF), Valencia, Spain

Outline

- ◆ PDB: the Protein Data Bank
- ◆ How to use the PDB
- ◆ PyMOL model viewer introduction
- ◆ PyMOL example

PDB

Protein Data Bank

The Protein Data Bank (PDB) is a repository for the three-dimensional structural data of large biological molecules, such as proteins and nucleic acids.

The data, typically obtained by X-ray crystallography or NMR spectroscopy and submitted by biologists and biochemists from around the world, are freely accessible.

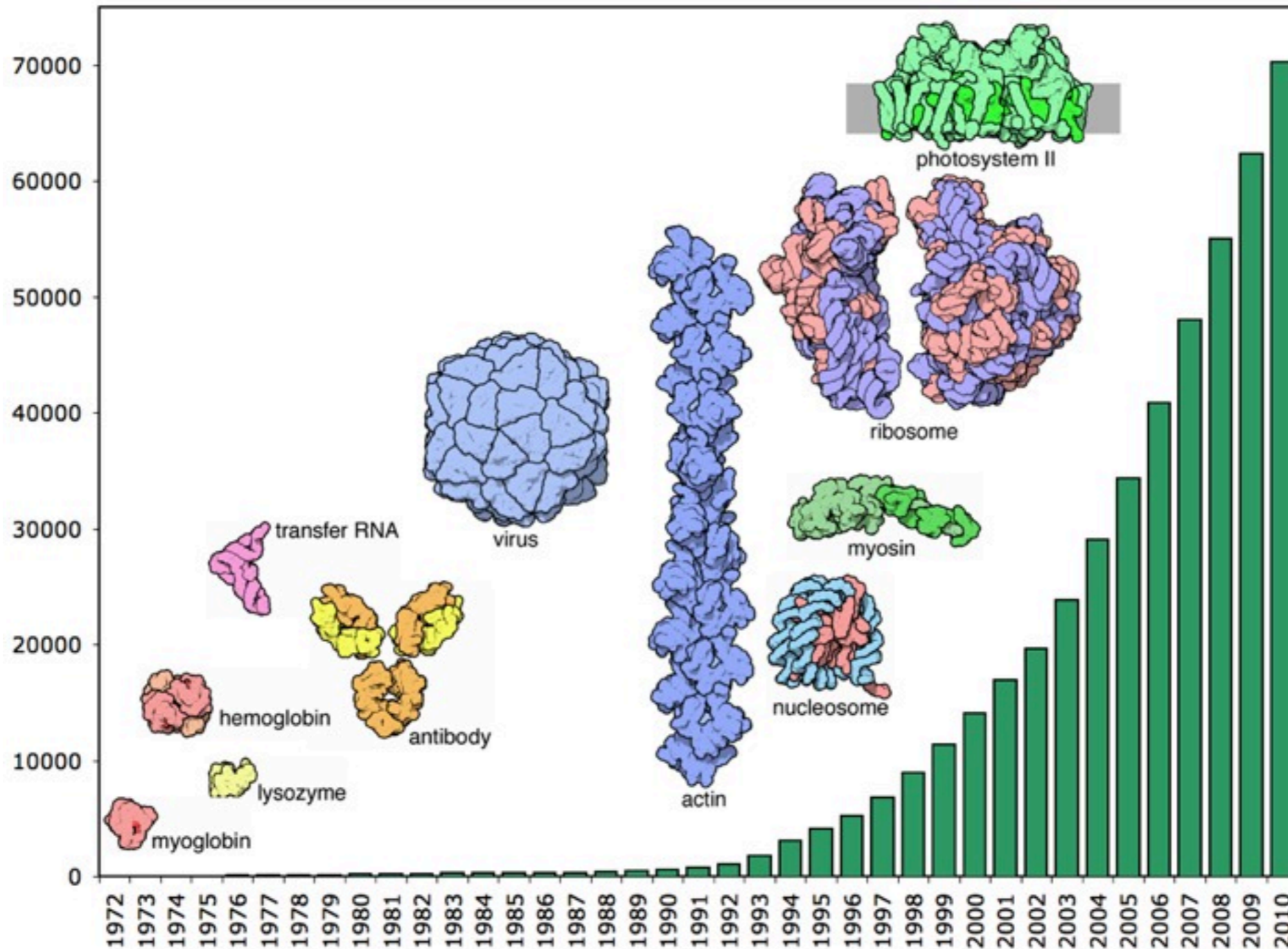
The PDB is a key resource in areas of structural biology, such as structural genomics.

The primary information stored in the PDB archive consists of coordinate files for biological molecules.

In a typical entry, you will find a diverse mixture of biological molecules, small molecules, ions, and water.

PDB

Yearly growth



PDB

Current status

Exp. Method	Proteins	Nucleic Acids	Protein/NA Complexes	Other	Total
X-RAY	61683	1308	2985	3	65979
NMR	7898	957	179	7	9039
Electron Microscopy	258	22	96	0	376
Hybrid	40	3	1	1	45
Other	133	4	5	13	155
Total	70010	2294	3266	24	75594

PDB

File format

```
HEADER      EXTRACELLULAR MATRIX                      22-JAN-98   1A3I
TITLE       X-RAY CRYSTALLOGRAPHIC DETERMINATION OF A COLLAGEN-LIKE
TITLE       2 PEPTIDE WITH THE REPEATING SEQUENCE (PRO-PRO-GLY)
...
EXPDTA      X-RAY DIFFRACTION
AUTHOR      R.Z.KRAMER,L.VITAGLIANO,J.BELLA,R.BERISIO,L.MAZZARELLA,
AUTHOR      2 B.BRODSKY,A.ZAGARI,H.M.BERMAN
...
REMARK 350 BIOMOLECULE: 1
REMARK 350 APPLY THE FOLLOWING TO CHAINS: A, B, C
REMARK 350   BIOMT1   1  1.000000  0.000000  0.000000          0.00000
REMARK 350   BIOMT2   1  0.000000  1.000000  0.000000          0.00000
...
SEQRES      1 A      9  PRO PRO GLY PRO PRO GLY PRO PRO GLY
SEQRES      1 B      6  PRO PRO GLY PRO PRO GLY
SEQRES      1 C      6  PRO PRO GLY PRO PRO GLY
...
ATOM        1  N      PRO A      1          8.316  21.206  21.530  1.00 17.44      N
ATOM        2  CA     PRO A      1          7.608  20.729  20.336  1.00 17.44      C
ATOM        3  C      PRO A      1          8.487  20.707  19.092  1.00 17.44      C
ATOM        4  O      PRO A      1          9.466  21.457  19.005  1.00 17.44      O
ATOM        5  CB     PRO A      1          6.460  21.723  20.211  1.00 22.26      C
...
HETATM     130  C      ACY      401          3.682  22.541  11.236  1.00 21.19      C
HETATM     131  O      ACY      401          2.807  23.097  10.553  1.00 21.19      O
HETATM     132  OXT   ACY      401          4.306  23.101  12.291  1.00 21.19      O
...

```

```

HEADER  EXTRACELLULAR MATRIX 20-100-00 1000
TITLE  X-RAY CRYSTALLOGRAPHIC DETERMINATION OF A COLLAGEN-LIKE
TITLE  2 PEPTIDE WITH THE REPEATING SEQUENCE (PRO-PRO-GLY)
...
EXPTNA  X-RAY DIFFRACTION
AUTHOR  R. J. ARAGÓN, L. VITELLO, J. BELLA, R. BERZIO, L. MAZZARELLA,
AUTHOR  F. B. ARAGÓN, A. SAGAN, J. R. BOWEN
...
REMARK 330  REMARKS: 1
REMARK 330  APPLY THE FOLLOWING TO CHAINS: A, B, C
REMARK 330  $$$$ 1 1.000000 0.000000 0.000000 0.000000
REMARK 330  $$$$ 1 0.000000 1.000000 0.000000 0.000000
...
SEQRES 1  A  N  PRO PRO GLY PRO PRO GLY PRO PRO GLY
SEQRES 1  B  N  PRO PRO GLY PRO PRO GLY
SEQRES 1  C  N  PRO PRO GLY PRO PRO GLY
...
ATOM   1  N  PRO  A  1  6.326  21.286  21.538  1.00  17.04  N
ATOM   2  CA  PRO  A  1  7.085  20.729  20.236  1.00  17.04  C
ATOM   3  C  PRO  A  1  6.467  20.787  19.692  1.00  17.04  C
ATOM   4  O  PRO  A  1  6.466  21.457  19.895  1.00  17.04  O
ATOM   5  CB  PRO  A  1  6.468  21.729  20.711  1.00  20.26  C
...
HETATM 138  C  ACY  481  3.682  22.340  11.226  1.00  21.29  C
HETATM 139  O  ACY  481  2.681  23.400  10.723  1.00  21.29  O
HETATM 140  OAT  ACY  481  6.388  23.181  12.071  1.00  21.29  O
...

```

PDB

File format

HEADER, TITLE and AUTHOR

Respectively: first line of the entry, contains PDB ID code, classification, and date of deposition; list of contributors; description of the experiment represented in the entry.

REMARK

General remarks about the data.

SEQRES

The primary structure (sequences) of the protein or DNA/RNA.

ATOM

Atomic the coordinates of the protein residues. Each line reports the atom number, name, amino acid, chain sequence number, coordinates, occupancy (1.0), B-factor (flexibility), and the element name.

HETATM

Atomic coordinates of small molecules, that is those non-standard atoms which are not part of the protein molecule.

Summary

The Protein Data Bank (PDB) is a repository for the three-dimensional structural data of large biological molecules, such as proteins and nucleic acids.

The data contained in the archive include atomic coordinates and experimental data.

Over 75000 structures deposited.

Freely accessible.

PDB Website

<http://www.pdb.org>

The screenshot shows the PDB website homepage with the following elements:

- Header:** RCSB PDB logo, "A MEMBER OF THE PDB", and a status message: "An Information Portal to Biological Macromolecular Structures. As of Tuesday Aug 30, 2011 at 5 PM PDT there are 75594 Structures | PDB Statistics".
- Navigation:** "Contact Us | Print" and a search bar with "PDB ID or Text" and "PDB ID lookup or Text search of the complete structure file" options.
- Left Sidebar:** "MyPDB" (Login, Register), "Home" (News, Policies, FAQ, etc.), "Deposition" (Services, X-ray, NMR, etc.), "Search" (Advanced, Latest, etc.), and "Tools" (Download, Compare, etc.).
- Main Content:**
 - A Resource for Studying Biological Macromolecules:** Introduction to the PDB archive and its tools.
 - Featured Molecules:** "Molecule of the Month: Rhomboid Protease GlpG" and "Protein Structure Initiative Featured System: Exploring the Secretome of Gut Bacteria".
 - Latest Structures:** "3rsx - Structure of Bace-1 (Beta-Secretase) in Complex with 6-(Thiophen-3-yl)quinolin-2-amine" and "From Fragment Screening to In Vivo Efficacy: Optimization of a Series of 2-Aminoquinolines as Potent Inhibitors of Beta-Site Amyloid Precursor Protein Cleaving Enzyme 1 (BACE1)".
- Right Sidebar:** "Customize This Page", "New Features", "wwPDB News" (PDB40 Symposium), and "RCSB PDB News".

PDB Website

<http://www.pdb.org>

1. Go to www.pdb.org in your web browser
2. Search for PDB ID **IT46**
3. Navigate the **Summary** and **Sequence** tabs (top menu)
4. Display the **FASTA** sequence and **PDB file** (top-right menu)
5. Download the PDB file in text format (top-right menu)
6. From the Summary tab, select the first PDB related entry (**IPKG**) and download it

PyMOL

<http://www.pymol.org>

PyMOL is an open-source, multi-platform program to visualize molecules and is available from the web.

The PyMOL GUI allows for an easy usage with no need to use the command line.

PyMOL allows to create high-quality images from 3D structures and to manipulate them. Some basic functions to analyze their chemical properties are also available.

The user has the possibility to write his/her own scripts and plugins.

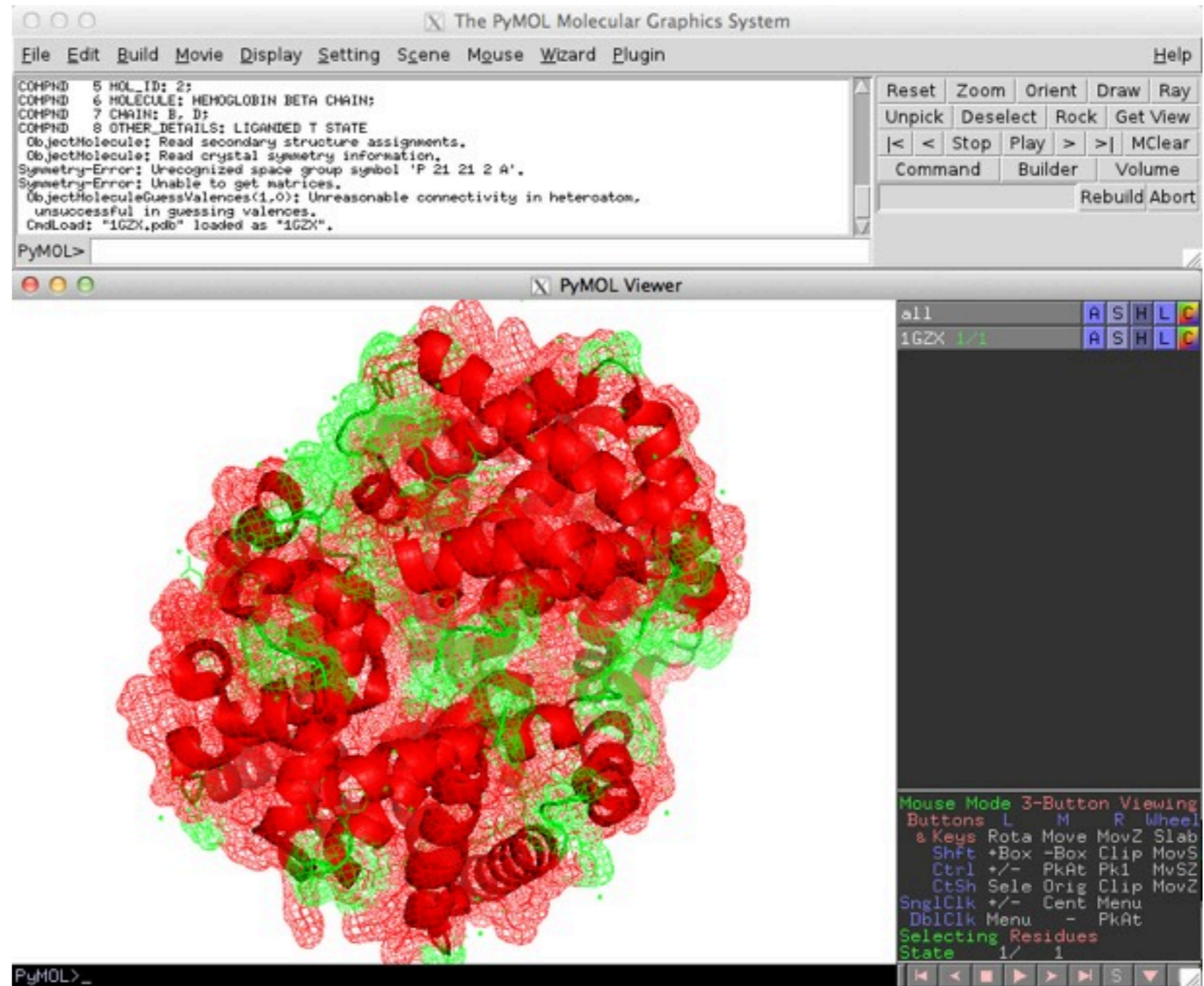
PyMOL

The graphical user interface (GUI)

PyMOL viewer: 3D molecules appear here.

External GUI: menu bar (File, Edit, etc), shortcut buttons, and the command line.

Object list: loaded molecules and user-defined selections.



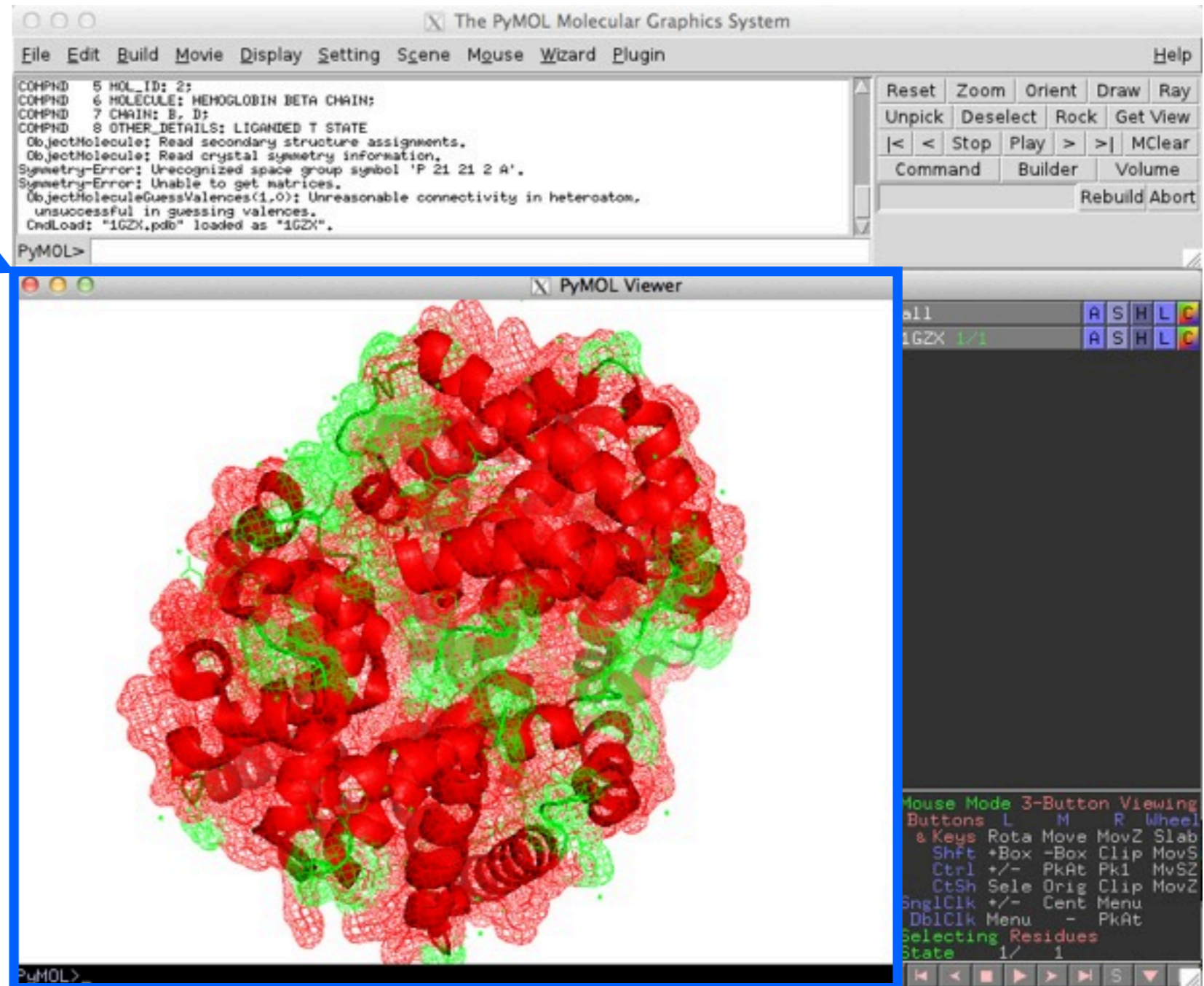
PyMOL

The graphical user interface (GUI)

PyMOL viewer: 3D molecules appear here.

External GUI: menu bar (File, Edit, etc), shortcut buttons, and the command line.

Object list: loaded molecules and user-defined selections.



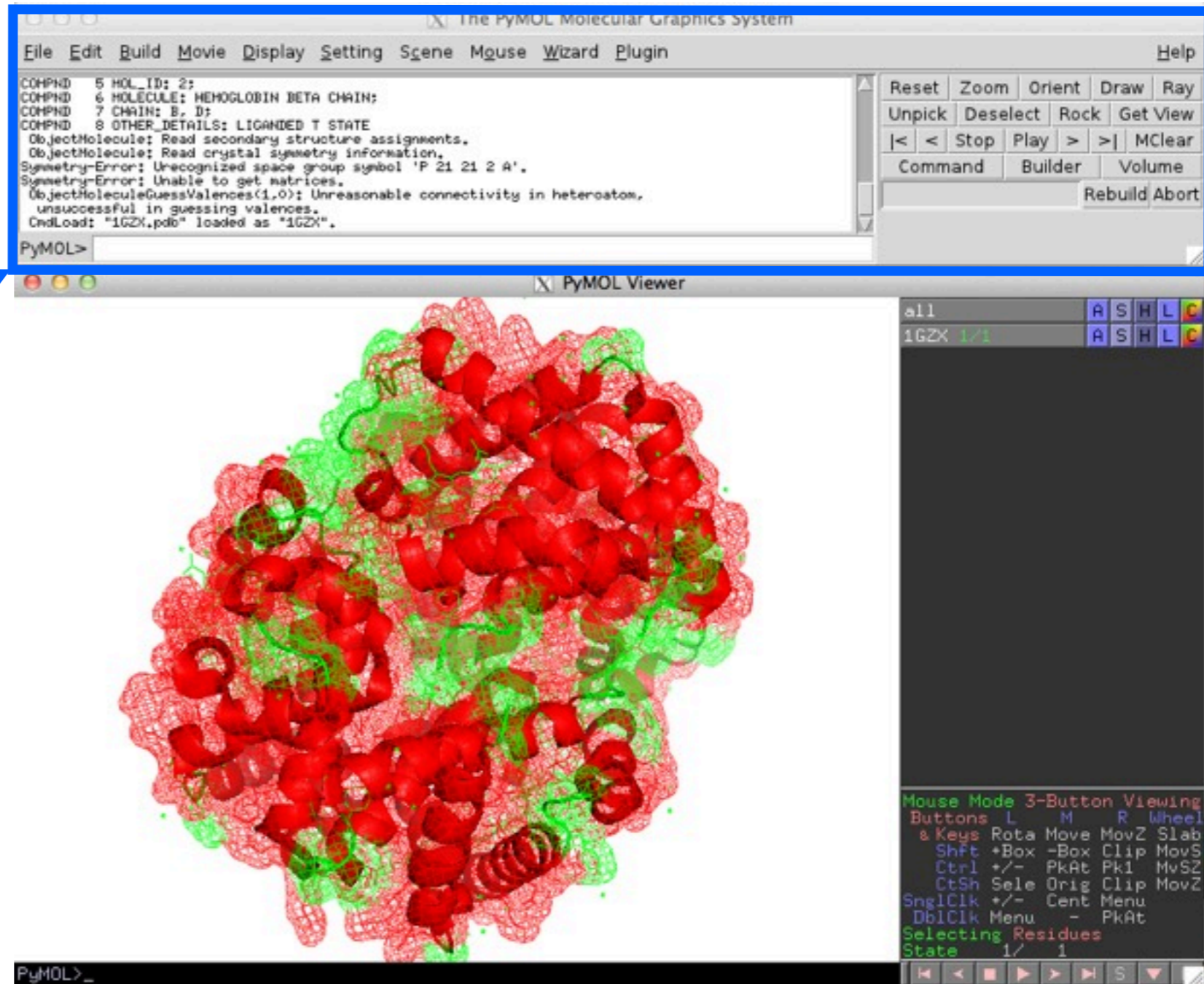
PyMOL

The graphical user interface (GUI)

PyMOL viewer: 3D molecules appear here.

External GUI: menu bar (File, Edit, etc), shortcut buttons, and the command line.

Object list: loaded molecules and user-defined selections.



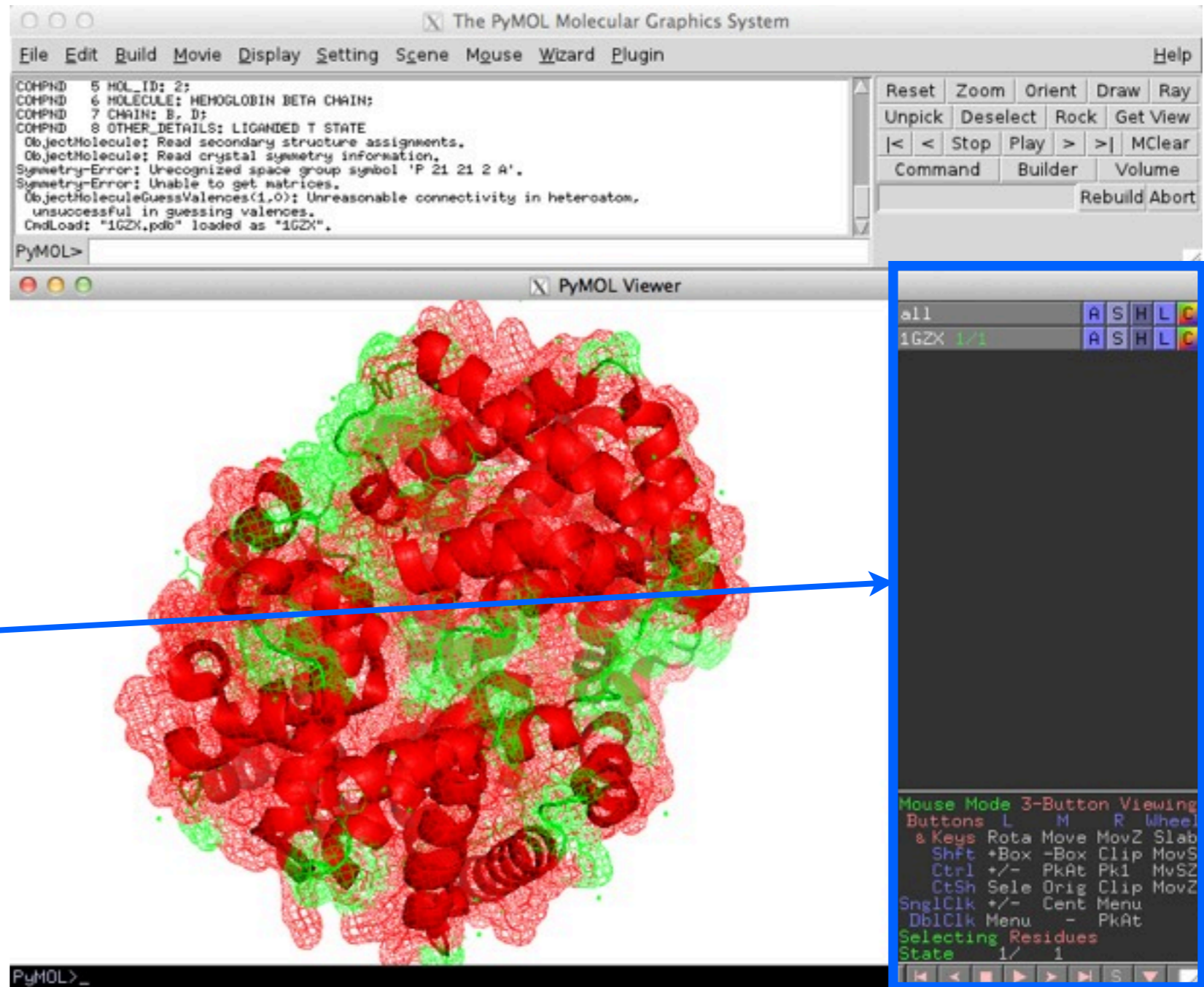
PyMOL

The graphical user interface (GUI)

PyMOL viewer: 3D molecules appear here.

External GUI: menu bar (File, Edit, etc), shortcut buttons, and the command line.

Object list: loaded molecules and user-defined selections.



PyMOL

The graphical user interface (GUI)

In the object list, after each object name there is a set of command buttons which control the object:

A - *Actions*: Rename, duplicate, remove, apply presets (like "ball-and-stick" or "publication"), perform computations

S - *Show*: Change the way things appear, eg change to stick or cartoon view.

H - *Hide*: Things that are shown using **S** accumulate, and don't automatically replace the last view. **H** is the opposite of **S** and hides unwanted representations.

L - *Label*: Label atoms, residues, etc.

C - *Color*: Change the color of atoms and groups.

PyMOL

Tutorial

Run PyMOL from the main menu:

1. Application ⇒ Science ⇒ PyMOL

Open a protein from the PyMOL menu:

2. File ⇒ open (select the previously downloaded IT46)
3. Use the mouse buttons to **rotate** (left), **move** (center) and **zoom** (right) the loaded structure.
4. A (*action*) ⇒ preset ⇒ ligands
5. IT46_pol_conts: S (*show*) ⇒ labels
6. IT46: A (*action*) ⇒ preset ⇒ default
7. Wizard ⇒ measurements

PyMOL

Tutorial

Open a protein from the PyMOL menu:

8. File ⇒ open (select the previously downloaded **IPKG**)

Now we have two proteins loaded!

9. **IT46: A** (*action*) ⇒ align ⇒ to molecule ⇒ **IPKG**

10. **IT46** and **IPKG: A** (*action*) ⇒ align ⇒ preset ⇒ publication

11. From the command line: **ray**

12. File ⇒ Save image as ⇒ PNG

Summary

PyMOL is an open-source, multi-platform program to visualize molecules and is available from the web:

<http://www.pymol.org>

The PyMOL GUI allows for an easy usage with no need to use the command line.

There are many other resources available online. A good one is the PyMOL wiki:

<http://www.pymolwiki.org>

Questions? (and coffee break!)

