Hands-on: PyMOL Molecular Visualization

A Comprehensive Guide to Protein Structure Visualization and Analysis

Getting Started

- Install PyMOL (open source available)
- Load PDB files: fetch 1AKE
- Basic navigation: mouse controls
- Command line interface

Visualization Options

- Cartoon: secondary structure
- Sticks: detailed bonds
- Surface: molecular surface
- Ribbon: protein backbone

Analysis Tools

- Distance measurements
- Hydrogen bond identification
- Surface area calculations
- Electrostatic potentials

Creating Figures

- Ray tracing for publication
- Color schemes
- Label atoms/residues
- Export high-resolution images

1. Getting Started with PyMOL

Installation and Setup

PyMOL is a powerful molecular visualization system used by researchers worldwide. There are two main versions available:

- **Open-Source PyMOL:** Free version available through package managers or compilation from source
- **Incentive PyMOL:** Commercial version with additional features and support

```
# Install via conda (recommended) conda install -c conda-forge pymol-open-source # Or via pip pip install pymol-open-source
```

> 💡 **Pro Tip:** For beginners, using conda provides the easiest installation with all dependencies automatically handled.

PyMOL Interface Screenshot

Main PyMOL window showing the viewer panel, command line, and object list

Loading Molecular Structures

PyMOL can load structures from various sources:

Loading a Structure Example

Example of fetching PDB 1AKE (Adenylate Kinase) from the Protein Data Bank

```
# Fetch from PDB database fetch 1AKE # Load local file load /path/to/protein.pdb # Load with custom name load myprotein.pdb, protein1
```

Common File Formats:

- PDB (.pdb) - Protein Data Bank format
- mmCIF (.cif) - Macromolecular Crystallographic Information
- MOL2 (.mol2) - Tripos molecule format
- SDF (.sdf) - Structure Data File

Basic Navigation Controls

Mastering mouse controls is essential for efficient work in PyMOL:

- **Left Mouse Button:** Rotate the view around the molecule
- **Middle Mouse Button:** Move/translate the molecule in the viewing plane
- **Right Mouse Button:** Zoom in and out (move up/down)
- **Scroll Wheel:** Alternative zoom control

> ⚠ **Important:** These controls can be customized in Edit → Mouse → 3 Button Viewing for different mouse configurations.

Mouse Control Guide

Visual guide showing left/middle/right mouse button functions

```
# Useful view commands reset # Reset view to original zoom # Zoom to fit all objects center # Center on selection orient # Orient selection to standard view
```

Command Line Interface Basics

The PyMOL command line is powerful for automation and precise control:

```
# Selection examples select chain A # Select chain A select resi 10-50 # Select residues 10 to 50 select name CA # Select all alpha carbons # Basic commands hide everything # Hide all representations show cartoon # Show cartoon representation color blue, chain A # Color chain A blue bg_color white # Set background to white
```

2. Visualization Options

Cartoon Representation

The cartoon representation is ideal for visualizing secondary structures (α-helices, β-sheets, loops) in proteins. It provides a clear overview of the protein's overall fold while maintaining clarity.

Key Features:

- α-helices displayed as spirals/cylinders
- β-sheets shown as arrows/ribbons
- Loops represented as tubes or thin lines
- Excellent for publication figures

```
# Display cartoon show cartoon # Customize cartoon appearance set cartoon_fancy_helices, 1 set cartoon_smooth_loops, 1 set cartoon_tube_radius, 0.5 # Color by secondary structure color red, ss h # Helices red color yellow, ss s # Sheets yellow color green, ss l # Loops green
```

Cartoon Representation

Protein shown in cartoon mode highlighting helices, sheets, and loops

Stick Representation

Stick Representation

Stick representation displays individual atoms and bonds, perfect for examining active sites, ligand binding, and detailed molecular interactions.

Best Uses:

- Visualizing ligand-protein interactions
- Examining active site residues
- Showing detailed chemical bonds
- Highlighting specific residues of interest

Active site residues shown as sticks with detailed chemical bonds

```
# Display sticks for selection show sticks, resi 50-75 # Adjust stick properties set stick_radius, 0.15 set stick_ball, on set stick_ball_ratio, 1.5 # Combine with cartoon show cartoon show sticks, resi 100-110 and name CA+CB+CG
```

💡 **Pro Tip:** Combine cartoon and stick representations to show overall structure while highlighting specific regions of interest.

Surface Representation

Surface representation displays the molecular surface, showing the accessible surface area and revealing binding pockets and cavities.

Surface Types:

- **Surface:** Quick surface calculation
- **Mesh:** Wireframe surface
- **Dots:** Dot surface representation

```
# Show molecular surface show surface # Adjust transparency set transparency, 0.5 # Surface quality settings set surface_quality, 2 set surface_type, 2 # 0=solid, 1=mesh, 2=dots #
```

Surface Representation

Molecular surface showing protein topology and binding pockets

```
Color by hydrophobicity color yellow,
hydrophobic color red, polar
```

Ribbon Representation

Ribbon view showing protein backbone connectivity

Ribbon Representation

Ribbon representation traces the protein backbone (Cα atoms), providing a simplified view of protein topology and fold.

Applications:

- Showing backbone trace for complex structures
- Highlighting domain organization
- Comparing multiple structures
- Creating simplified overviews

```
# Display ribbon show ribbon # Ribbon
customization set ribbon_width, 3.0 set
ribbon_smooth, 1 set ribbon_trace_atoms, 1 #
Putty representation (B-factor coloring) show
cartoon set cartoon_putty_radius, 0.4 spectrum
b, rainbow, minimum=10, maximum=50
```

3. Analysis Tools

Distance Measurements

Measuring distances between atoms is crucial for understanding molecular interactions, validating models, and analyzing structural features.

Types of Measurements:

- **Distance:** Linear distance between two atoms
- **Angle:** Angle between three atoms
- **Dihedral:** Torsion angle between four atoms

```
# Measure distance between atoms distance dist1, resi 25 and name CA, resi 50 and name CA # Measure all distances in selection distance all_contacts, chain A, chain B, 4.0 # Measure angle angle ang1, resi 10 and name CA, resi 11 and name CA, resi 12 and name CA # Customize measurement display set dash_radius, 0.15 set dash_color, red set label_size, 20
```

⚠ **Note:** Typical hydrogen bond distances range from 2.5-3.5 Å, while van der Waals contacts are typically 3.5-4.5 Å.

Distance Measurements

Example showing distance measurements between key residues with labeled values

Hydrogen Bond Identification

Hydrogen Bond Network

Hydrogen bonds displayed with dashed lines showing interaction network

Hydrogen bonds are crucial for protein stability, ligand binding, and catalytic function. PyMOL provides tools to identify and visualize these interactions.

Criteria for H-bonds:

- Distance: 2.5-3.5 Å between donor and acceptor
- Angle: Donor-H-Acceptor angle > 120°
- Typically between O, N, and sometimes S atoms

```
# Find hydrogen bonds distance hbonds, chain A, chain B, 3.2, mode=2 # Show polar contacts set h_bond_cutoff_center, 3.6 set h_bond_cutoff_edge, 3.2 # Advanced H-bond visualization distance hb1, (resi 50 and name O), (resi 75 and name N), 3.5 color yellow, hb1 set dash_gap, 0 # Label H-bonds with distances set label_position, (0, 0, 5)
```

💡 **Pro Tip:** Use "mode=2" in distance command to automatically filter for potential hydrogen bonds based on geometry.

Surface Area Calculations

Surface area analysis helps understand protein-protein interactions, ligand binding sites, and structural accessibility.

Key Concepts:

- **Solvent Accessible Surface Area (SASA):** Area accessible to solvent
- **Buried Surface Area (BSA):** Area hidden upon complex formation
- **Surface Charge Distribution:** Electrostatic surface properties

```
# Calculate solvent accessible surface area
get_area selection # Example: Calculate total
protein area get_area protein # Calculate buried
surface area get_area chain A and chain B
buried_area = get_area(chainA) +
get_area(chainB) - get_area(complex) # Show
accessible surface show surface set
surface_mode, 1 # 0=solid, 1=surface set
surface_quality, 2
```

Surface Area Visualization

Protein surface colored by accessible surface area or hydrophobicity

Electrostatic Potentials

Electrostatic potential maps reveal charged regions important for ligand binding, protein-protein interactions, and catalysis. This requires APBS (Adaptive Poisson-Boltzmann Solver).

Applications:

- Identifying binding sites

Electrostatic Potential Map

Surface colored by electrostatic potential (red=negative, blue=positive)

- Understanding substrate specificity
- Analyzing protein-protein interfaces
- Drug design and docking studies

```
# Using APBS plugin (if installed) # Plugin → APBS Tools # Manual approach: color by atom charge color red, formal_charge < 0 color blue, formal_charge > 0 # Color surface by electrostatics
util.protein_vacuum_esp(selection, mode=2) # Simple charge visualization select acidic, resn ASP+GLU select basic, resn ARG+LYS+HIS color red, acidic color blue, basic
```

⚠ **Important:** For publication-quality electrostatic maps, install APBS and use the APBS Tools plugin with proper parameter settings.

4. Creating Publication-Quality Figures

Ray Tracing for Publication

Ray tracing produces high-quality, photorealistic images suitable for publications and presentations by simulating realistic lighting

and shadows.

Benefits of Ray Tracing:

- Smooth, anti-aliased edges
- Realistic shadows and depth perception
- Professional appearance
- Adjustable quality and resolution

```
# Basic ray tracing ray # High-resolution ray tracing ray 1200, 1200 # Width, height in pixels # Ray tracing settings set ray_trace_mode, 1 # 0=fast, 1=normal, 3=quantized set ray_shadows, on set ray_trace_fog, 0 set antialias, 2 set orthoscopic, on # Save ray-traced image png output.png, width=2400, height=2400, dpi=300, ray=1
```

💡 **Pro Tip:** For publications, use at least 1200x1200 pixels or 300 DPI resolution. Ray tracing may take several minutes for complex structures.

Ray Traced Image

Comparison of standard view vs ray-traced rendering showing improved quality

Color Schemes and Customization

Effective color schemes enhance clarity and highlight important features in molecular structures.

Common Coloring Strategies:

Color Scheme Examples

Various coloring schemes: by chain, by element, by B-factor, rainbow spectrum

- **By Chain:** Different colors for each protein chain
- **By Element:** CPK (carbon=cyan, oxygen=red, nitrogen=blue)
- **By Secondary Structure:** Helices, sheets, loops
- **By Property:** B-factor, hydrophobicity, charge
- **Rainbow/Spectrum:** N-terminus to C-terminus

```
# Color by chain util.cbc # Color by chain (automatic) color red, chain A color blue, chain B # Color by element (CPK) util.cnc # Color by element # Rainbow coloring (N to C terminus) spectrum count, rainbow, selection # Color by B-factor spectrum b, blue_white_red, minimum=10, maximum=50 # Custom colors set_color mycolor, [0.5, 0.8, 0.3] # RGB values color mycolor, resi 50-75
```

Labeling Atoms and Residues

Labels help identify specific residues, atoms, or features in molecular structures, making figures more informative and accessible.

Labeling Options:

- Residue names and numbers

- Atom names
- Custom text labels
- Distance measurements

```
# Label residues label resi 50 and name CA, "%s %s" % (resn, resi) # Label specific atoms label name CA and resi 25+50+75, "%s%s" % (resn, resi) # Customize label appearance set label_size, 20 set label_color, black set label_position, (0, 0, 3) set label_font_id, 7 # Font selection # Remove labels label all, "" # Label by chain label chain A and name CA and resi 1, "Chain A"
```

⚠ **Tip:** Keep labels minimal and strategic. Too many labels can clutter the image and reduce clarity.

Labeled Structure

Protein structure with residue labels highlighting active site residues

Exporting High-Resolution Images

Proper image export ensures your figures meet publication requirements and maintain quality across different media.

Export Formats:

- **PNG:** Best for publications (lossless, transparency support)
- **PDF/EPS:** Vector format for scalable graphics
- **TIFF:** High-quality raster format

Export Options

High-resolution PNG export at 300 DPI for publication

- **PDB/PSE:** Save PyMOL session for later editing

```
# High-quality PNG export png figure1.png, width=2400, height=2400, dpi=300, ray=1 # Quick preview (no ray tracing) png preview.png, width=800, height=800 # Export with transparent background set ray_opaque_background, off png transparent.png, dpi=300, ray=1 # Save session for later save mysession.pse # Export coordinates save output.pdb, selection
```

Publication Standards:

- Minimum 300 DPI for print publications
- Minimum 1200x1200 pixels for single-column figures
- Use white or light background for print
- Include scale bars or size references when appropriate

```
# Complete publication workflow example bg_color white set ray_shadows, on set antialias, 2 set orthoscopic, on ray 2400, 2400 png publication_figure.png, dpi=300
```

Additional Resources and Best Practices

Workflow Tips for Efficient Visualization

🎯 **Best Practices:**

- Always start with structure validation (check for missing atoms, unusual geometries)
- Use consistent color schemes across related figures
- Save PyMOL sessions (.pse files) regularly for reproducibility
- Document your visualization commands in scripts for publication methods
- Test different viewing angles before finalizing figures
- Use high-quality settings only for final renders (saves time during editing)

Common PyMOL Shortcuts

```
# Essential keyboard shortcuts reset # Reset view zoom # Zoom to fit center sele # Center on selection orient sele # Orient selection hide everything # Clean slate as cartoon # Show as cartoon (shorthand) as sticks # Show as sticks util.cbc # Color by chain util.cnc # Color by element clip # Toggle clipping planes
```

Recommended Learning Resources

- **PyMOL Wiki:** Comprehensive documentation and tutorials
- **PyMOL Users Mailing List:** Community support and discussions
- **Online Tutorials:** Video tutorials on YouTube and molecular visualization courses

- **PDB Education:** Resources from RCSB Protein Data Bank
- **Scientific Papers:** Methods sections often include PyMOL visualization details

> 📚 **Further Learning:** Practice with diverse structures from the Protein Data Bank (www.rcsb.org). Start with well-characterized proteins and gradually work with more complex systems like membrane proteins or large assemblies.