Hands-on: PDB Database Exploration

Search Strategies

- Keyword search: protein name
- Advanced search: filters
- Sequence similarity
- Structure similarity

Structure Quality

- Resolution: <2Å is high quality
- R-factor: fit to data
- Ramachandran plot
- Missing residues

Functional Analysis

- Active site identification
- Ligand binding
- Protein-protein interfaces
- Conformational changes

Integration with AlphaFold

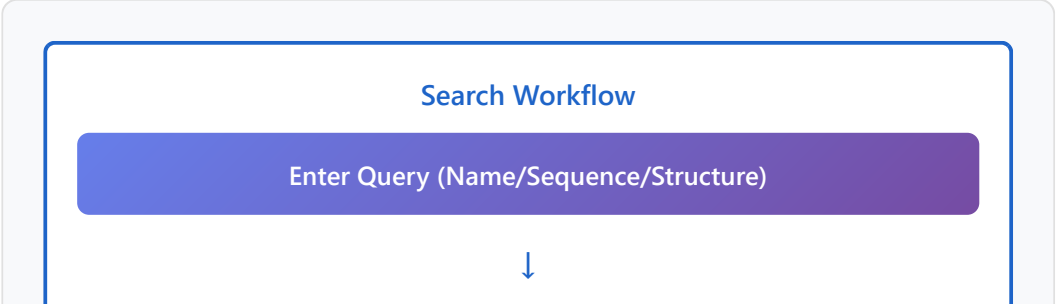
- Predicted structures available
- Confidence scores (pLDDT)
- Complement experimental data
- AlphaFold database

---

1 Search Strategies in PDB

Keyword Search

The simplest approach to find protein structures. Enter protein names, gene names, or biological terms directly into the search bar. For example, searching for "hemoglobin" returns all hemoglobin structures.



Advanced Search Filters

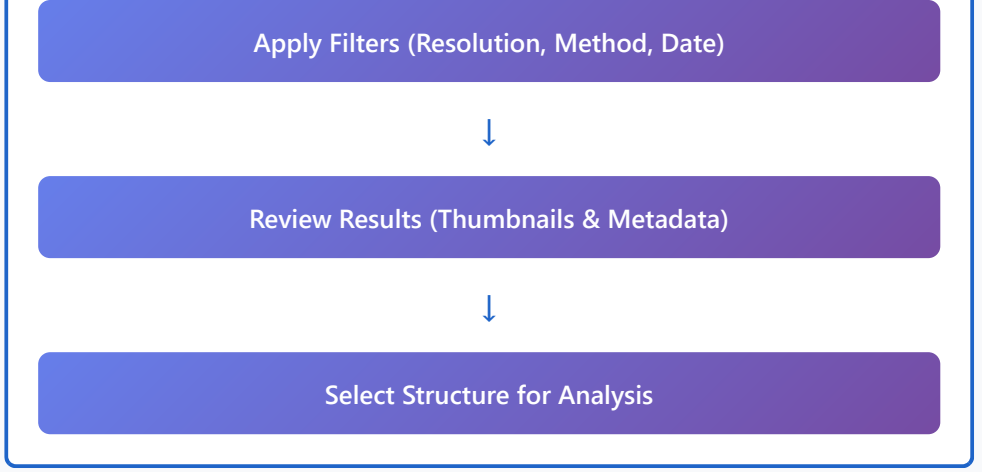
Refine results using multiple criteria including experimental method (X-ray, NMR, Cryo-EM), resolution range, organism source, molecular weight, and release date. Combine filters to find exactly what you need.

Sequence Similarity (BLAST)

Find structures of proteins with similar sequences. Upload your sequence or paste it directly. BLAST searches reveal homologous proteins that may share functional characteristics, even across different species.

Structure Similarity

Identify proteins with similar 3D folds regardless of sequence. This is powerful for discovering distant evolutionary relationships and functional analogs that wouldn't be detected by sequence alone.



Pro Tip: Use the "Advanced Search" option to combine multiple criteria. For instance, search for "kinase" + "X-ray" + "resolution < 2.0Å" + "Homo sapiens" to find high-quality human kinase structures.

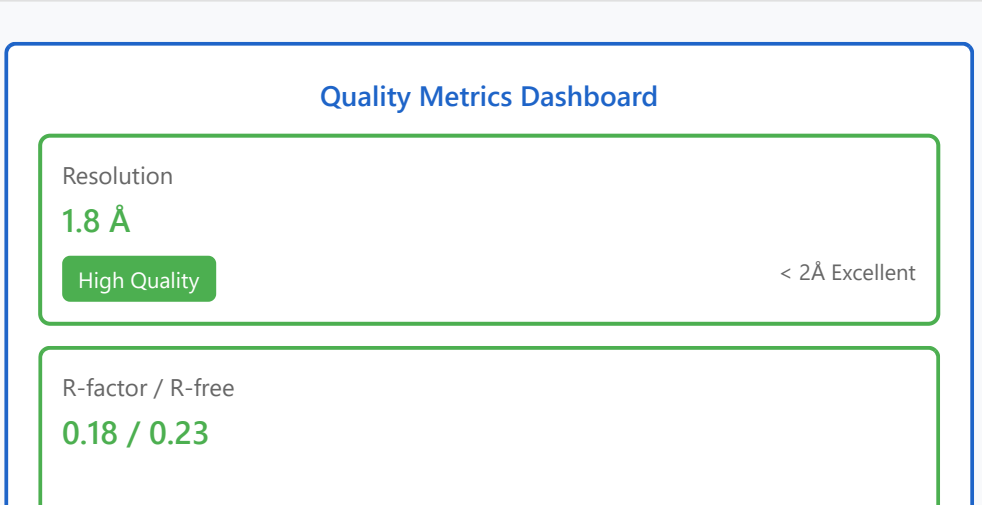
2 Assessing Structure Quality

Resolution

Measures the level of detail in the structure. Lower values indicate higher quality. Structures with resolution <2Å show clear atomic details, while >3Å may have ambiguous regions. X-ray crystallography typically achieves 1.5-3Å resolution.

R-factor (R-value)

Indicates how well the model fits the experimental data. Values range from 0 to 1, with lower being better. A good structure has R-factor <0.20 and R-free



<0.25. These metrics validate the reliability of the atomic coordinates.

Ramachandran Plot

Validates protein backbone geometry by showing phi-psi angles. Well-refined structures have >90% residues in favored regions. Outliers may indicate errors or unusual conformations requiring closer examination.

Missing Residues

Flexible or disordered regions may not be visible in electron density. Check the structure for gaps in the sequence. Missing termini or loops are common but may be functionally important regions.

Good Fit

< 0.25 Target

Ramachandran Favored

96.5%

Excellent

> 90% Expected

Quality Checklist: Always review these metrics before using a structure for modeling or drug design. High-resolution structures with good R-factors are most reliable for detailed analysis.

3 Functional Analysis Tools

Active Site Identification

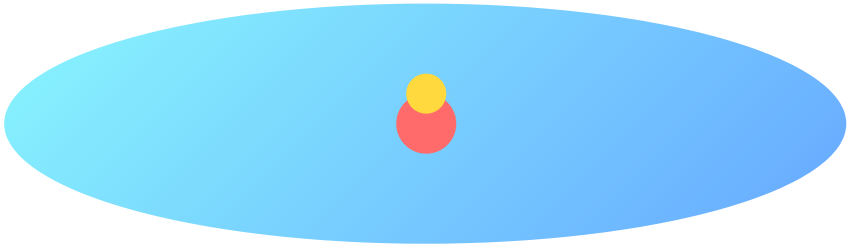
Locate catalytic residues and binding pockets using PDB annotations and computational tools. Active sites are often conserved across homologs and are critical for enzymatic function and drug targeting.

Ligand Binding Analysis

Examine how small molecules, substrates, or inhibitors interact with the protein. Analyze binding modes, contact residues, and hydrogen bonds. Co-crystallized ligands reveal key interactions for drug design.

Protein-Protein Interfaces

Protein-Ligand Interaction



Blue: Protein Surface
Red: Active Site
Yellow: Ligand

Study how proteins interact in complexes. Interface analysis reveals critical residues for binding and stability. Understanding these interactions is essential for studying signaling pathways and designing inhibitors.

Conformational Changes

Compare multiple structures of the same protein to observe conformational flexibility. Many proteins undergo dramatic shape changes during function. Identifying these movements reveals mechanisms of action.

```
# PyMOL Commands for Analysis select active_site, resi 50+52+100
show sticks, active_site show surface, protein color cyan,
protein color red, active_site
```

Analysis Tools: Use PyMOL, Chimera, or online tools like PDBsum to visualize and analyze functional features. Look for co-crystallized ligands to understand binding mechanisms.

---

4 Integration with AlphaFold

AlphaFold Predicted Structures

Access AI-predicted structures for millions of proteins through the AlphaFold Database. These predictions cover entire proteomes, including proteins without experimental structures. Download predictions directly from PDB or AlphaFold DB.

Confidence Scores (pLDDT)

AlphaFold provides per-residue confidence scores (0-100). Regions with pLDDT >90 are highly accurate, 70-90 are good, 50-70 are low confidence, and <50 should be treated with caution. Color coding helps identify reliable regions.

Complementing Experimental Data

Use AlphaFold predictions to fill in missing loops or disordered regions in experimental structures. Compare predictions with X-ray structures to validate

Experimental vs. Predicted Structures

PDB (Experimental)	AlphaFold (Predicted)
✓ High accuracy	✓ Full proteome
✓ Ligand binding	✓ Fast access
✗ Limited coverage	✗ No ligands/dynamics

pLDDT Confidence Scale

Low (<50) | Medium (50-70) | High (>90)

Best Practice: Start with experimental structures when available. Use AlphaFold predictions for proteins lacking experimental data or to model missing regions. Always check confidence scores before analysis.

models. Predictions are especially valuable for membrane proteins and large complexes.

Accessing the Database

Visit alphafold.ebi.ac.uk to search by UniProt ID or protein name. PDB now includes links to AlphaFold predictions. Download structures in PDB format for use in molecular dynamics, docking, or comparative analysis.