# The PDB and Molecular Visualization

Biochemistry Boot Camp 2022
Session #9
Nick Fitzkee
nfitzkee@chemistry.msstate.edu

# Properties of PDB Files

- Experimental methodology:
  - X-Ray: Typically more precise
  - NMR: Need lots of “restraints;” sometimes hard to assess quality
  - Cryo-Electron Microscopy: Large complexes

- Most PDB files are solved using X-ray crystallography
  - 191,000 structures total
  - 166,000 are crystal structures

# Building an X-Ray Structure

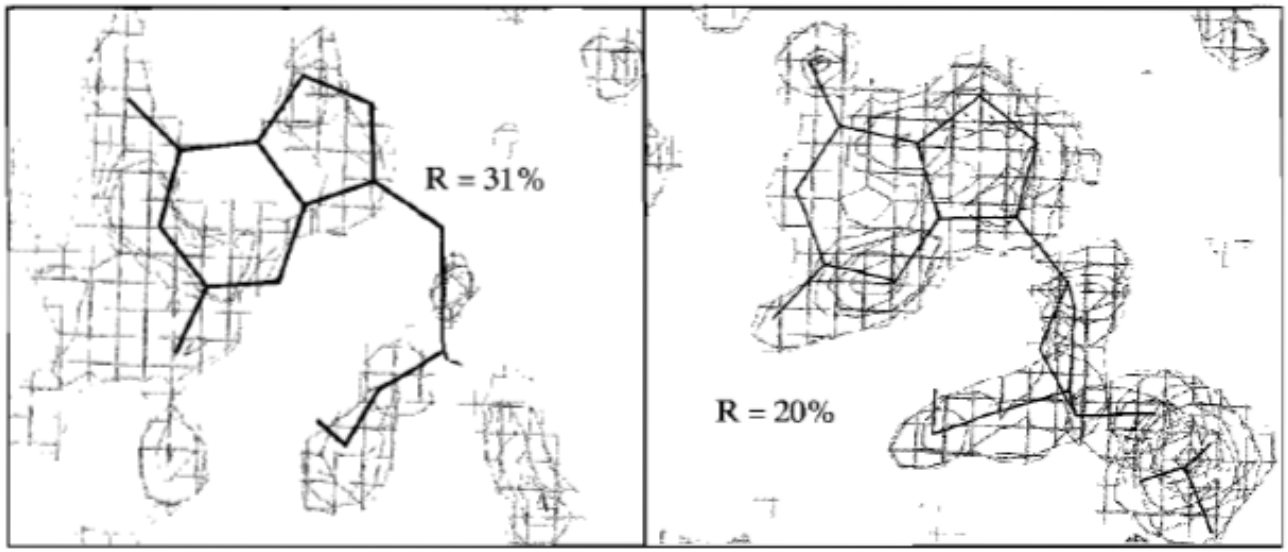



**Figure 6.31** Effect of refinement on structure. The guanine nucleotide of a DNA fragment is shown with its electron density map prior to refinement and after refinement. Prior to refinement, the *R* factor is 31%. The structure is refined against the data to an *R* factor of 20%, which is one criterion of a good fit of the model to the data.

- **At first:** look for gross structural features (helix, backbone), then add side chains
- Molecular mechanics are used to help refine positions

# Guidelines for X-Ray Quality

- **R-factor:** Less than 25% (or 0.25; ideally, less than 20%)
- **R-Free:** Bigger than R, but smaller than 25%
- **Resolution:** Less than 2.5 Å, but think about how much you need (1.5 Å usually very good)
  - At ~1 Å hydrogens become visible
- **Validation:** No clashes, good torsions, etc.

# Guidelines for X-Ray Quality

- **Not too Much Bound Water**:
  - Higher resolution typically means reduced molecular motions, more bound water will be visible
  - However, water molecules can also be used to "sop up" electron density of unknown origin

- **Rules of Thumb (Approximate)**

| 1.5 Å or better | Worse than 2.0 Å |
|---|---|
| 0.5-1.5 waters per AA | 0.2-0.8 waters per AA |
| 5-15 waters per kDa | 2-8 waters per kDa |

- **Example:** 7B4O ($R_{free}$ = 0.19, 1.25 Å): 564 water molecules for 153 x 4 = 612 AA total (65 kDa)

Nittinger, *et al.* (2015) *J. Chem. Inf. Model.* 55: 771−783

# Protein Data Bank Revisited

- http://rcsb.org/

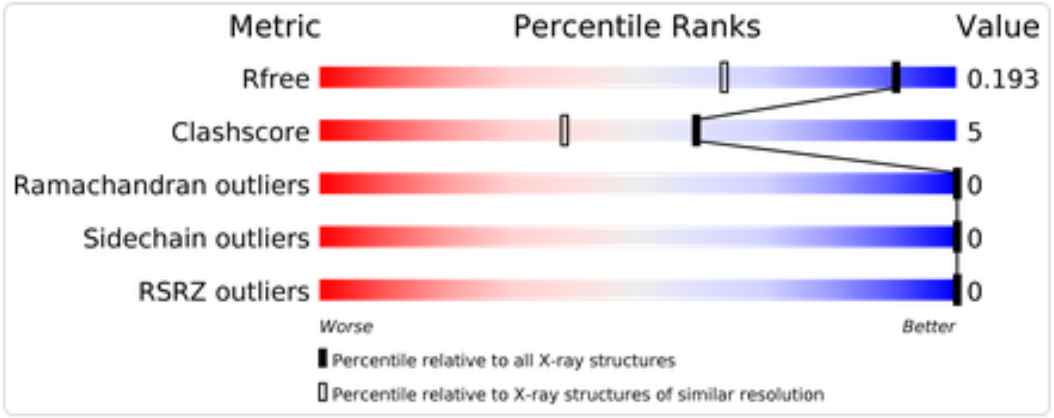
- **Input:** Protein name, PDB ID, authors, etc.

- **Output:** 3D coordinates of protein structures
  - Author information on methods
  - Cofactors and other information

# Assessing a Crystal Structure: 3TJW

## Structure Validation

View Full Validation Report or Ramachandran Plots




- PDB contains a lot of useful information for determining how good a crystal structure is
- **Things to look at:** R, $R_{free}$, resolution, structure validation

Experimental Data & Validation

### Experimental Data

**Method:** X-RAY DIFFRACTION
**Resolution:** 1.46 Å
**R-Value Free:** 0.190
**R-Value Work:** 0.150
**Space Group:** F 2 2 2

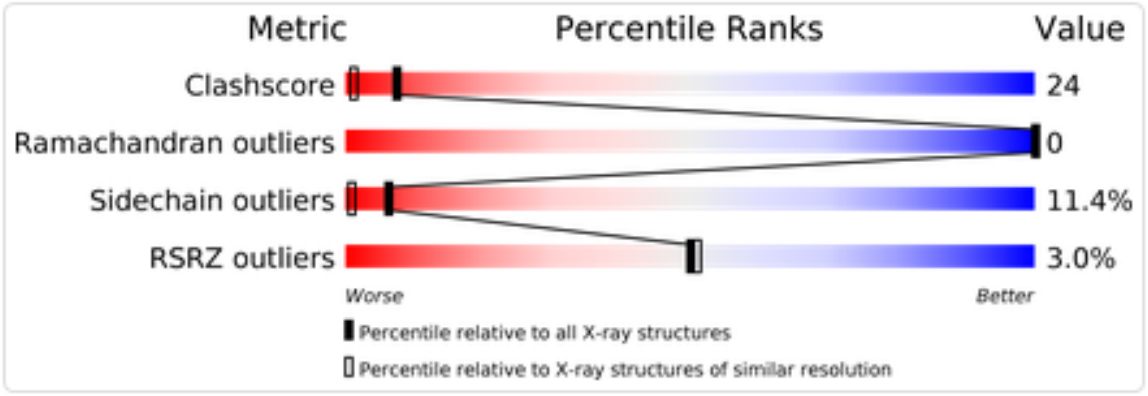
**Unit Cell**:

| Length (Å) | Angle (°) |
|---|---|
| a = 38.656 | α = 90.00 |
| b = 88.112 | β = 90.00 |
| c = 88.663 | γ = 90.00 |

# Assessing a Crystal Structure: 1SNC

## Structure Validation

View Full Validation Report or Ramachandran Plots




- Not every protein will have $R_{free}$

Experimental Data & Validation

### Experimental Data

**Method:** X-RAY DIFFRACTION
**Resolution:** 1.65 Å
**Space Group:** P $4_1$

**Unit Cell**:

| Length (Å) | Angle (°) |
|---|---|
| a = 48.000 | α = 90.00 |
| b = 48.000 | β = 90.00 |
| c = 63.500 | γ = 90.00 |

# PDB Files: A Closer Look

# Obtaining PyMOL

- Recent versions of educational PyMOL (EduPyMOL) are great, although time-limited.

- Older educational builds can be obtained by registering at:

  http://www.pymol.org/educational/

- **Mac or PC:** Fill in form and follow instructions for downloading; follow standard procedure

- **Linux:** Open source (full version) can be obtained through most distributions (e.g, apt, yum)

# Obtaining PyMOL

- PyMOL License File: http://pymol.org/ep
  Username: jun2021
  Password: betabarrel

- Download PyMOL from https://pymol.org/2/#download (install license file from above)

- Older versions for Windows XP, etc. (version 1.3r1) are available at:
  Windows: http://goo.gl/0mRH2O
  Mac: http://goo.gl/U68Hoc
  Linux: http://goo.gl/HMZvPU

# Obtaining PyMOL on Linux/Debian

- PyMOL is open source, and an alternative is downloading compiled versions on Linux

- Linux versions (i.e., versions compiled independently from Schrodinger) are not crippled, even the latest version. They do lack some Schrodinger-specific features.

- On Debian/Ubuntu (easy, but you need admin access):
  `apt-get install pymol`

# Obtaining PyMOL on CentOS

- On Redhat/Centos (harder, and you still need admin access):

```
1. sudo yum update
2. sudo yum install epel-release
3. sudo yum install subversion gcc gcc-c++
   kernel-devel python-devel tkinter python-pmw
   glew-devel freeglut-devel libpng-devel
   freetype-devel libxml2-devel apbs
```

- The commands above install the necessary libraries for compiling PyMOL
- The final command should be typed entirely on one line.

# Obtaining PyMOL: CentOS (cont.)

- Obtain the latest compressed zip file for PyMOL from
  https://github.com/schrodinger/pymol-open-source/archive/refs/heads/master.zip

- Extract the source code using the following command:
  `unzip pymol-open-source-master.zip`

- This will extract the source code to a subdirectory in the current working directory (e.g. `pymol-open-source-master`)

# Obtaining PyMOL: CentOS (cont.)

- Finally, compile and install python. Change to the python source directory and type (will take some time):

```
1. python3 setup.py build
2. python3 setup.py install --prefix=~/somedir
```

- You can then run PyMOL from the `~/somedir` directory (you can also create a link):

```
1. cd ~/somedir /bin
2. ./pymol
```

- Administrator privileges are ***not*** needed for a local installation; just for installing developer libraries (two slides ago)

# Alternatives to PyMOL

- UCSF Chimera (https://www.cgl.ucsf.edu/chimera/)
  - Absolutely free, although tricky to learn. A large and growing user base.

- VMD (https://www.ks.uiuc.edu/Research/vmd/)
  - Optimized for looking at MD simulations

- More information about history at http://www.umass.edu/microbio/rasmol/history.htm

# Quick Intro to PyMOL

# **PyMOL Tools:** Protein Alignment

- Open two PDB files within PyMOL: 1F8A and 1PIN (prolyl isomerase)

```
fetch 1f8a
fetch 1pin
```

- Align both structures using the following command (uses similar residues as cues):

```
align 1f8a and resi 80-160, 1pin and resi 80-160
```

# **PyMOL Tools:** Measurement

- Measure distances and dihedral angles using PyMOL (Wizard → Measurement)

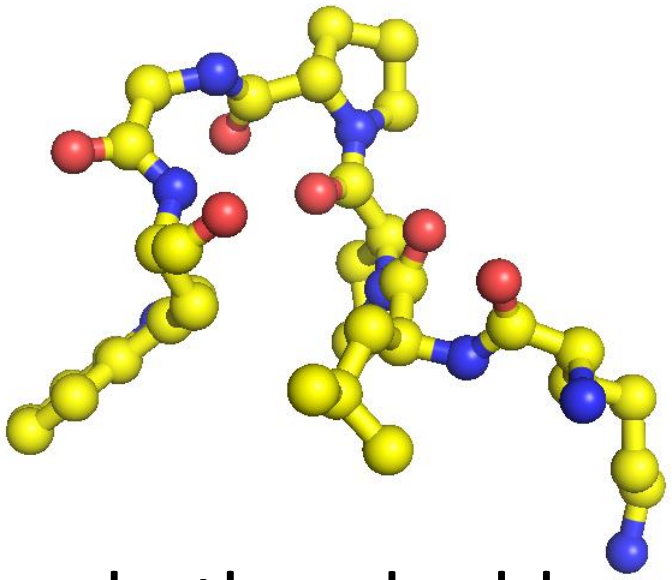
- Recall that phi (ϕ) is defined as the dihedral angle defined by:

$$C'_{i-1} - N_i - C_{\alpha,i} - C'_i$$

# **PyMOL Tools:** Ball-and Stick

- Combining small spheres with stick model can produce a pleasing result

```
show sticks
show spheres
set sphere_scale, 0.3
set valence, 0
```



- The final command toggles whether double bonds are shown

# **PyMOL Tools:** Transparent Surfaces

- Show surfaces using

  `show surface, selection`

- Transparency can be applied to different renderings (e.g. cartoon vs. spheres)

  `set sphere_transparency, 0.65`

- Surface transparency is a global property, so a separate object must be created if mixed results are desired (see next slide)

# **PyMOL Tools:** Transparent Surfaces

- **Example:** Cartoons under a semi-transparent surface
  1. First, create a duplicate object by clicking 1f8a (A) → Copy to Object → New
  2. Then show surface for the new object (obj01)
     ```
     show surface, obj01
     color grey, obj01
     ```
  3. Set transparency for the new object
     ```
     set transparency, 0.5, obj01
     ```
  4. Show cartoons for the original object
     ```
     hide everything, 1f8a
     show cartoon, 1f8a
     color yellow, 1f8a
     ```

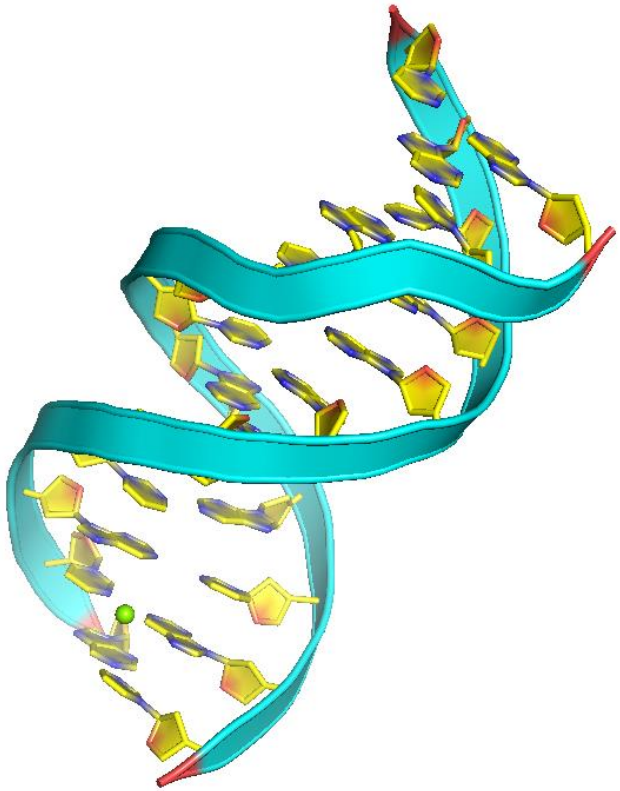
# **PyMOL Tools:** Awesome DNA

- Grab the Dickerson dodecamer

```
fetch 4c64
```

- Reasonable first attempt

```
set cartoon_ring_mode, 3
cartoon dumbbell
set cartoon_dumbbell_radius, 0.25
color cyan, elem P
```



- My go-to pages for DNA:

https://kpwu.wordpress.com/2006/10/22/pymol-fancy-dna-helix-and-filled-rings/
https://kpwu.wordpress.com/2011/07/26/pymol-example-of-dna/
https://pymolwiki.org/index.php/Examples_of_nucleic_acid_cartoons

# **PyMOL Tools:** Ray Tracing Tips

- **Ray Tracing:** Simulation of photon light paths, reflection, and scattering to give a photorealistic image

- New Draw/Ray Panel
  - Fine control over resolution/image size
  - Automatic removal of background (a bit buggy); for best results, maximize slab so no fading
  - Display → Background: Can set color

- Always set slab to largest view area before ray tracing (mouse wheel up until fog is gone)

- Save image: Use File → Export Image As… (or click Draw/Ray again)

# Summary

- PDB files are complex models, derived from experimental data
  - Need to assess these models, too
- Atom names and properties are stored in the PDB, can be used to investigate structures
- PyMOL can make interesting pictures, but it is also a very powerful analytical tool