ALPHAFOLD AND PYMOL ON NEGISHI AND GILBRETH

Charles Christoffer, Graduate Research Assistant

PURDUE UNIVERSITY | Rosen Center for Advanced Computing

Contents

Outline

- Introduction to AlphaFold

- Running AlphaFold on Negishi
- Running AlphaFold on Gilbreth

- Attendee Example Launching

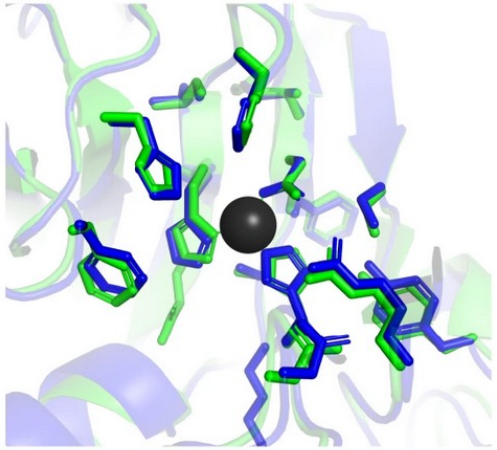
- Running PyMOL on Negishi Gateway
- Visualizing AlphaFold Models in PyMOL
- Rendering Publication-Quality Images in PyMOL

- Attendee Example Workshopping

DeepMind's AlphaFold

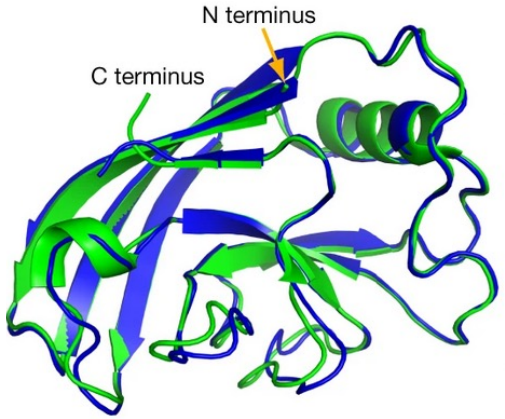
AlphaFold Predicts Protein Structure

- Classical prediction methods require structure templates (e.g. MODELLER, I-TASSER).
- Modern deep learning methods can predict structure using only sequence information.
- AlphaFold is the current gold standard for protein structure prediction.



AlphaFold Experiment
r.m.s.d. = 0.59 Å within 8 Å of Zn

PDB 6YJ1
Staph. phage 2638A
Endolysin M23 domain
(zinc binding site)



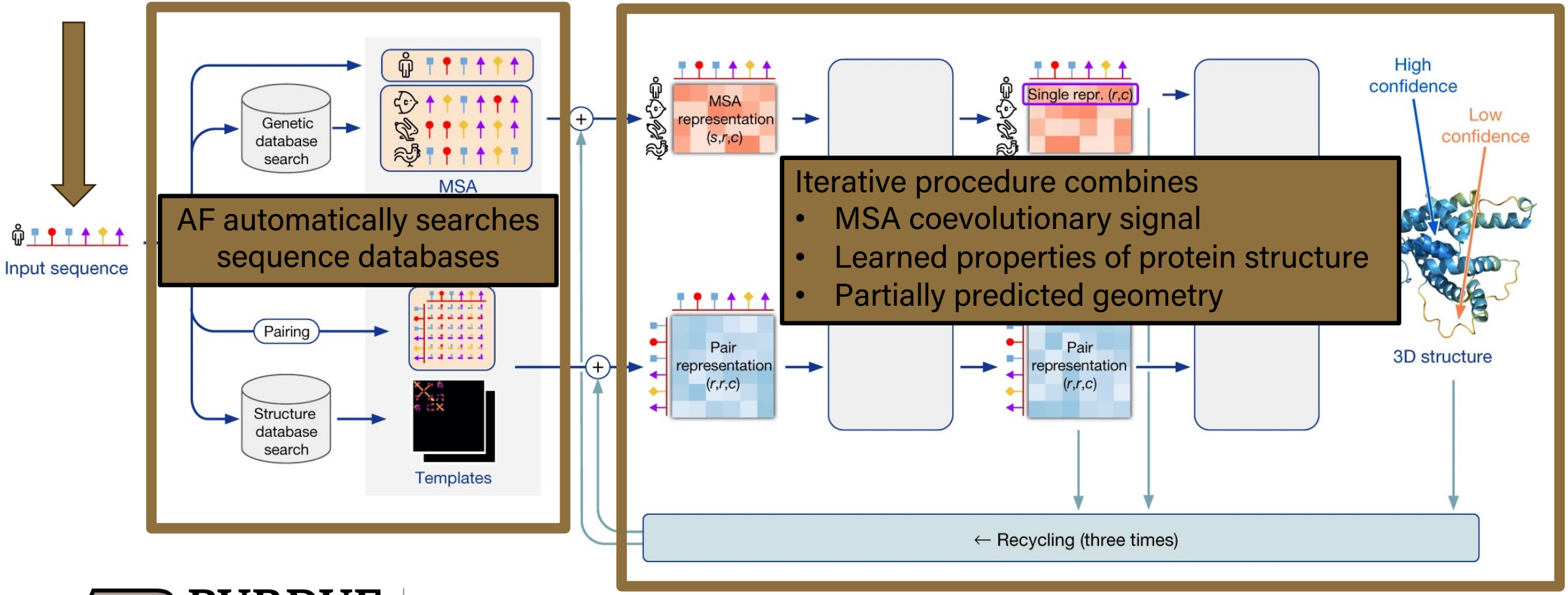
AlphaFold Experiment
r.m.s.d.$_{95}$ = 0.8 Å; TM-score = 0.93

PDB 6Y4F
P. mirabilis
Fimbrial adhesin

(Jumper, Evans et al. 2021)

Architecture of AlphaFold

AlphaFold takes only sequences from the user

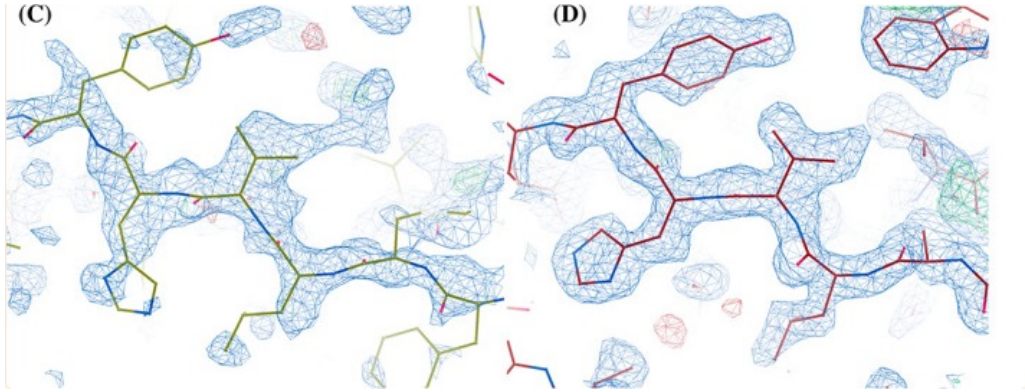


(Jumper, Evans et al. 2021)

Utility of AlphaFold

AlphaFold represents the state of the art

- Thoroughly validated in competition, but not perfect.
- AF models good for molecular replacement phasing in crystallography (Millan, Keegan et al. 2021).
- Not reliable when:
  - Too-sparse or no MSA available
  - Sequence not evolutionary
    - Antibody-antigen interface
    - Point mutation studies
  - Large state-dependent structure differences



(C) Phased w/ baseline (D) Phased w/ AF

(Millan, Keegan et al. 2021)

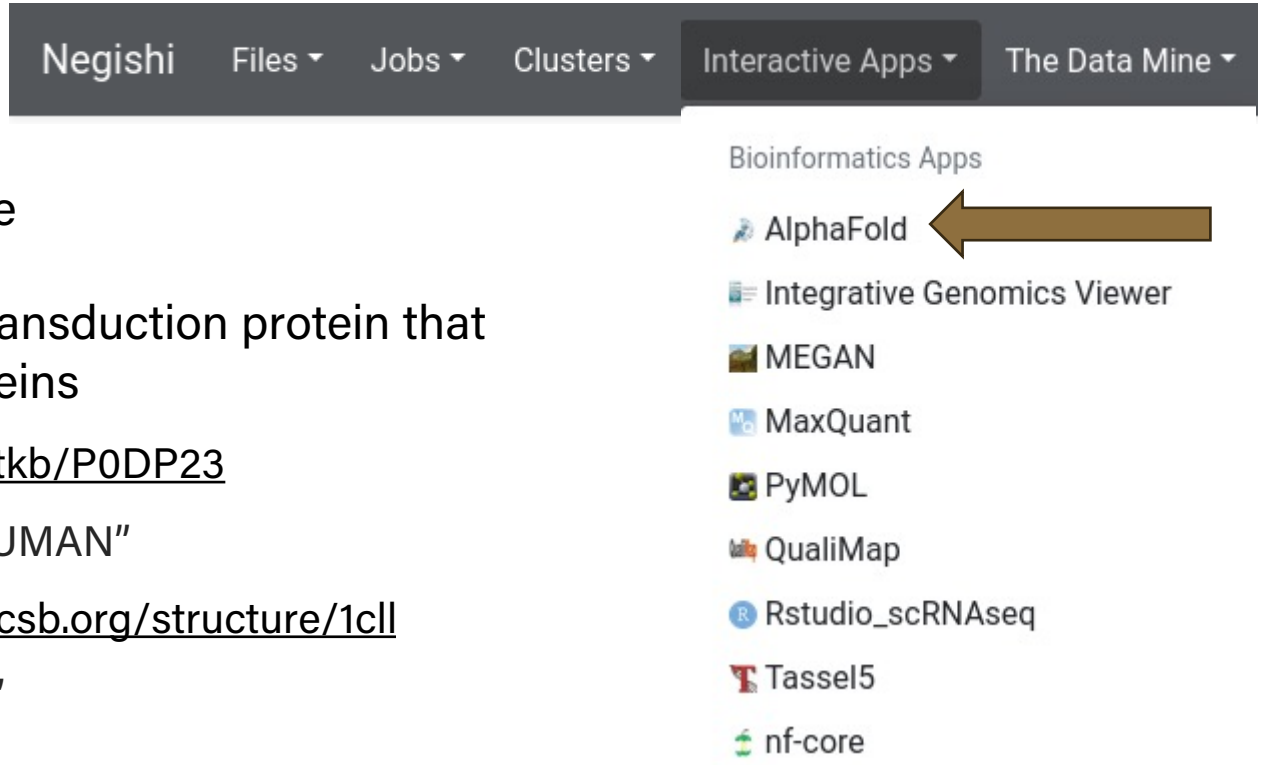
AlphaFold on RCAC Community Clusters

AlphaFold is available on all Community Clusters

- `biocontainers` module suite includes AlphaFold for command line and shell script use
- Negishi Gateway provides a web interface
  - https://gateway.negishi.rcac.purdue.edu
  - Coming soon to Gilbreth
- Databases are provided by RCAC
- Users do not need to install or download anything to go from a protein sequence to predicted protein structure models!

Negishi Gateway Web Interface

https://gateway.negishi.rcac.purdue.edu



- Let's start with a small example
- Calmodulin is a 149aa signal transduction protein that interacts with many other proteins
  - https://www.uniprot.org/uniprotkb/P0DP23
    - Or Google search "CALM1_HUMAN"
  - Crystal structure: https://www.rcsb.org/structure/1cll
    - Or Google search "PDB 1CLL"

Gateway Job Submission

Paste or upload your sequence to run from the browser

AlphaFold

This app will launch AlphaFold. More information about AlphaFold can be found here (https://github.com/deepmind/alphafold).

Queue

standby (Max 4.0 hours)

Use dedicated queue for large

Please select a queue from the drop-down and enter the number of hours below (up to the max listed above).

Wall Time in Hours

4

Use longer time limit for large

Number of hours you are requesting for your job.

Cores

128

Max 128 cores, no multi-node support

You will receive ~2GB of memory per core requested.

PURDUE UNIVERSITY | Rosen Center for Advanced Computing

Output directory Name

alphafold.out

Output PDB files will generate into here

Where the results will be going to (relative to the working directory field above).

Example: alphafold.out

fasta_paths

Fill filename here if uploading

The fasta files containing amino acid sequence(s) to fold. If there are more multiple files, please separate them using comma(e.g. seq1.fasta,seq2.fasta)

Amino acid sequence_1

Paste sequence here if using clipboard

Enter the 1st amino acid sequence(s) to fold:

model_preset

monomer

Change to “multimer” if you have a complex

Select to run the monomer or multimer model for sequences.

Command Line Usage

Modules available on all Community Clusters

- In a terminal, allocate a GPU node on Gilbreth

```
sinteractive -A standby –N1 –n8 --gpus-per-node=1 --time 4:00:00
module load biocontainers alphafold

infile=calm.fasta
outdir=alphafold.out
flagfile=/depot/itap/datasets/alphafold/full_db_20230311_monomer.ff
maxdate=1970-01-01

run_alphafold.sh --fasta_paths=$infile --output_dir=$outdir \
  --flagfile=$flagfile \
  --use_gpu_relax=True --max_template_date=$maxdate
```

Meanwhile...

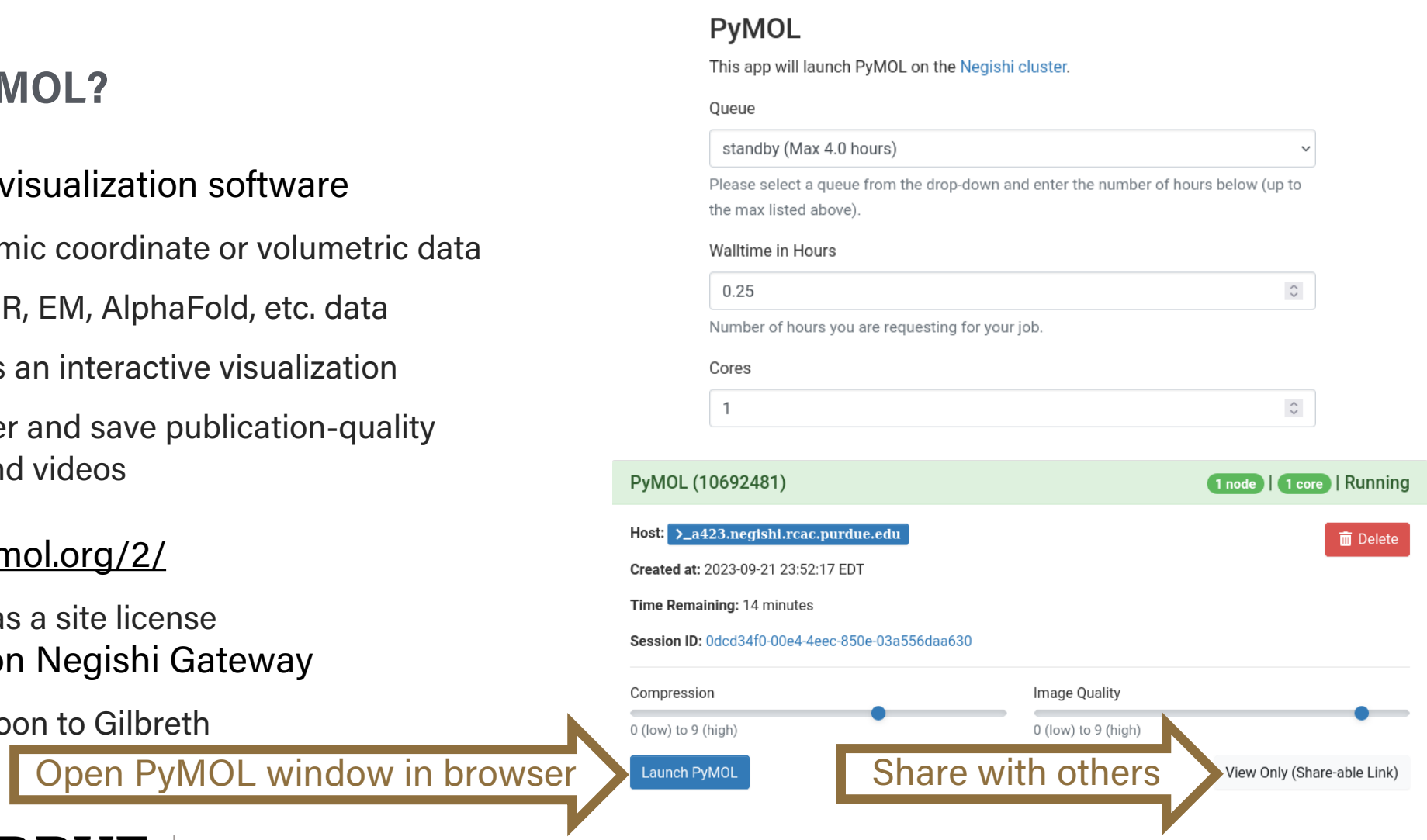
PyMOL

Let’s talk about visualizing protein structures while AlphaFold runs...

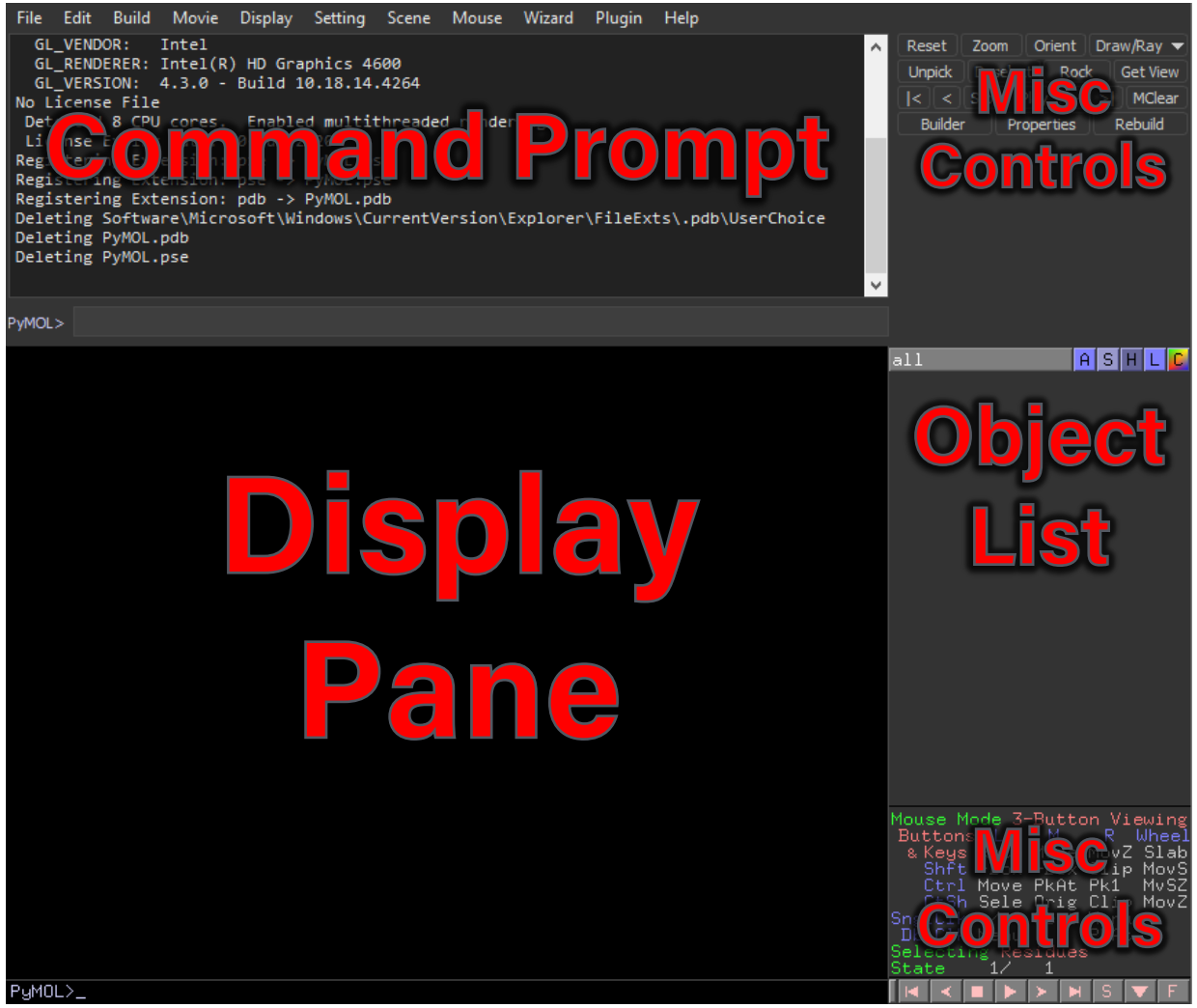
DeLano and Schrodinger's PyMOL

What is PyMOL?

- Molecular visualization software
  - Given atomic coordinate or volumetric data
  - X-ray, NMR, EM, AlphaFold, etc. data
  - Generates an interactive visualization
  - Can render and save publication-quality images and videos
- https://pymol.org/2/
  - Purdue has a site license
- Available on Negishi Gateway
  - Coming soon to Gilbreth



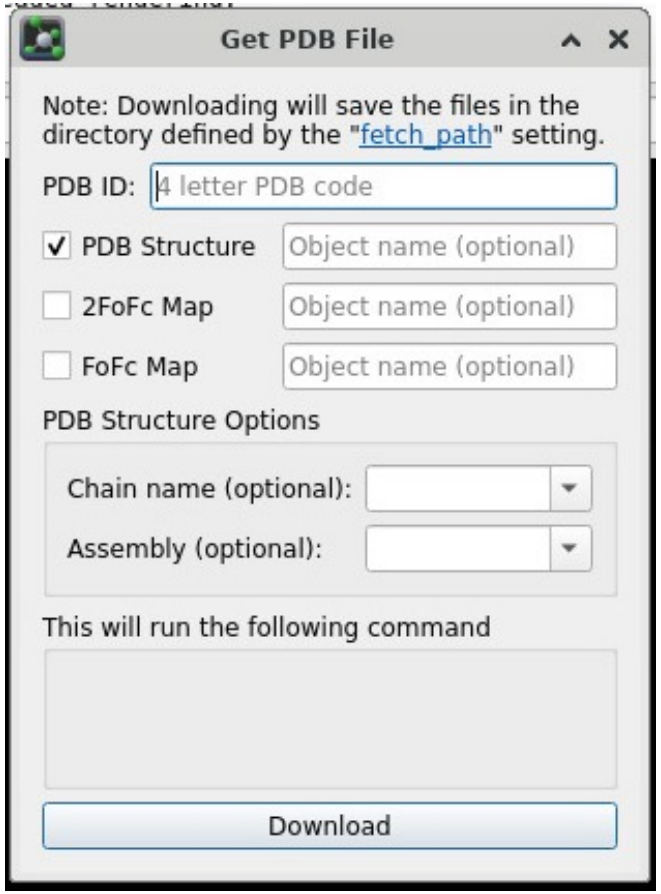
Main Window


File Edit Build Movie Display Setting Scene Mouse Wizard Plugin Help
GL_VENDOR: Intel
GL_RENDERER: Intel(R) HD Graphics 4600
GL_VERSION: 4.3.0 - Build 10.18.14.4264
No License File
Registering Extension: pdb -> PyMOL.pdb
Deleting Software\Microsoft\Windows\CurrentVersion\Explorer\FileExts\.pdb\UserChoice
Deleting PyMOL.pdb
Deleting PyMOL.pse
PyMOL>
Reset Zoom Orient Draw/Ray
Unpick Rock Get View
MClear
Builder Properties Rebuild
Command Prompt
Misc Controls
Display Pane
all
Object List
Mouse Mode 3-Button Viewing
Misc Controls
Selecting residues
State 1/ 1
PyMOL>_

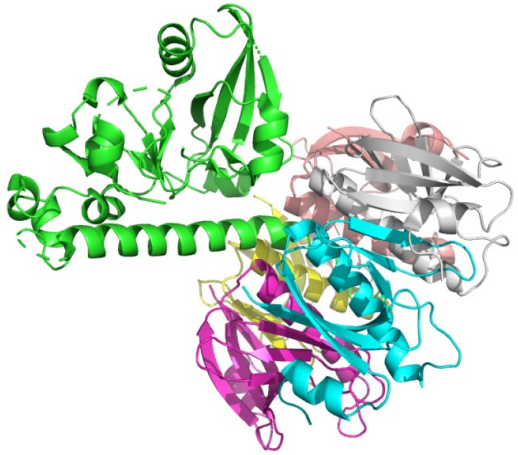
Loading Data

PyMOL handles PDB, mmCIF, MRC, SITUS, and more!

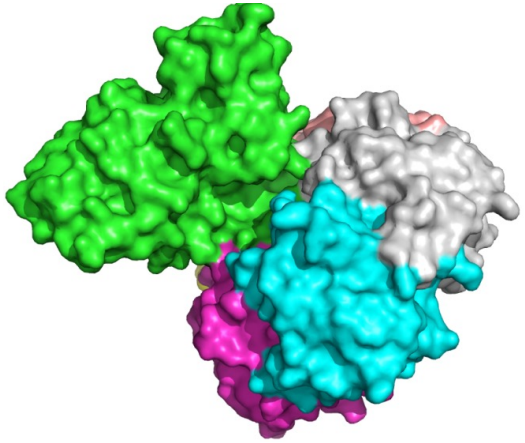
- Can open files on your computer
  - File → Open
  - `load <path to file>`
- Can download directly from PDB
  - File → Get PDB
  - `fetch <PDB code>`
- Check Setting → mmCIF... → Load Assembly
  - Loads structure without crystal partners, etc.



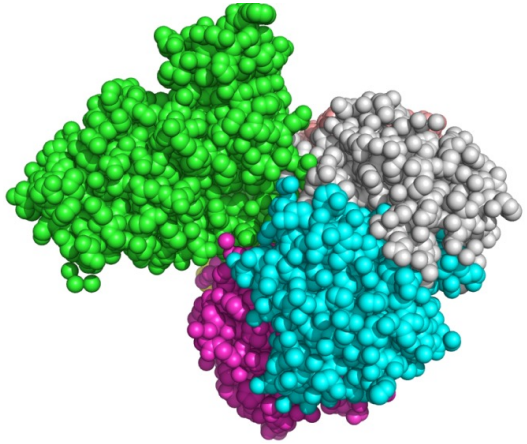
Representations for Atomic Coordinate Data



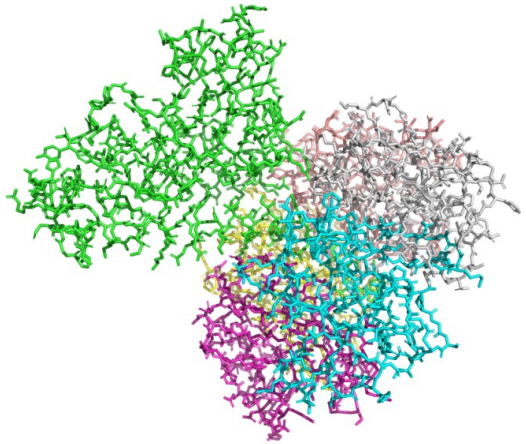
Cartoon



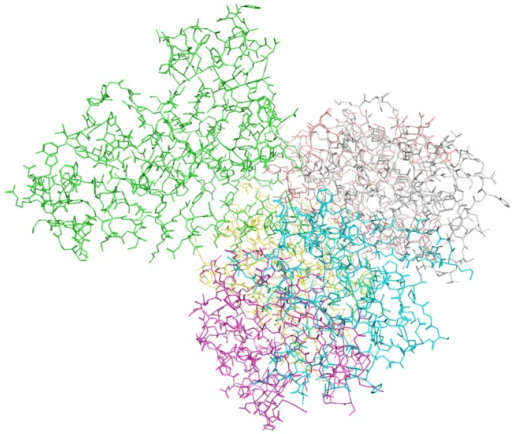
Surface



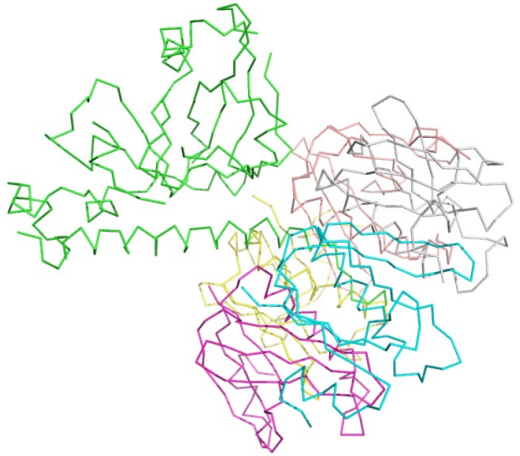
Spheres



Sticks



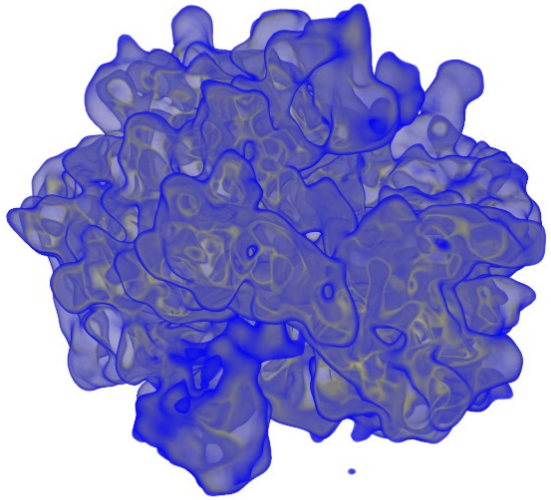
Lines



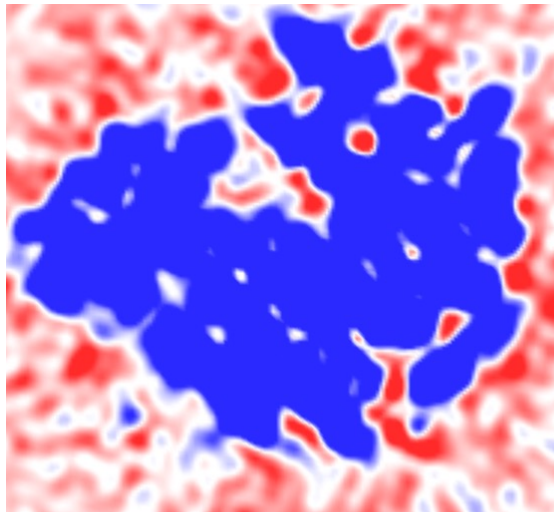
Ribbon

Representations for Volume Data

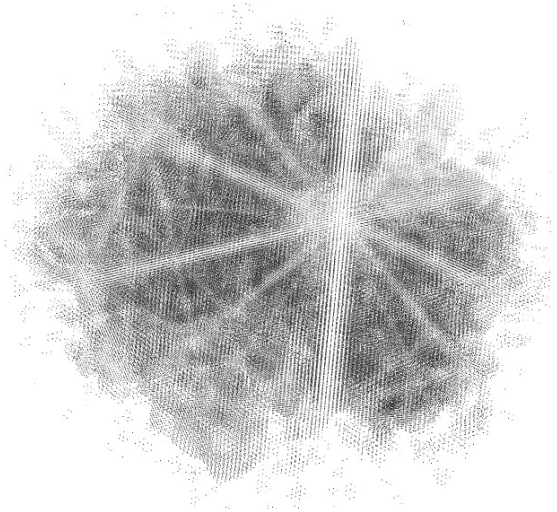
Surface



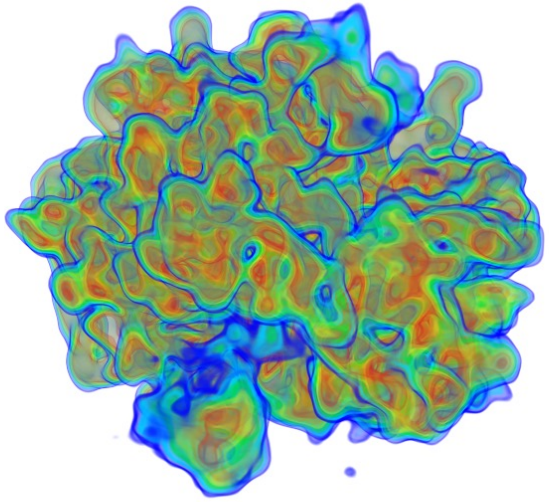
Slice



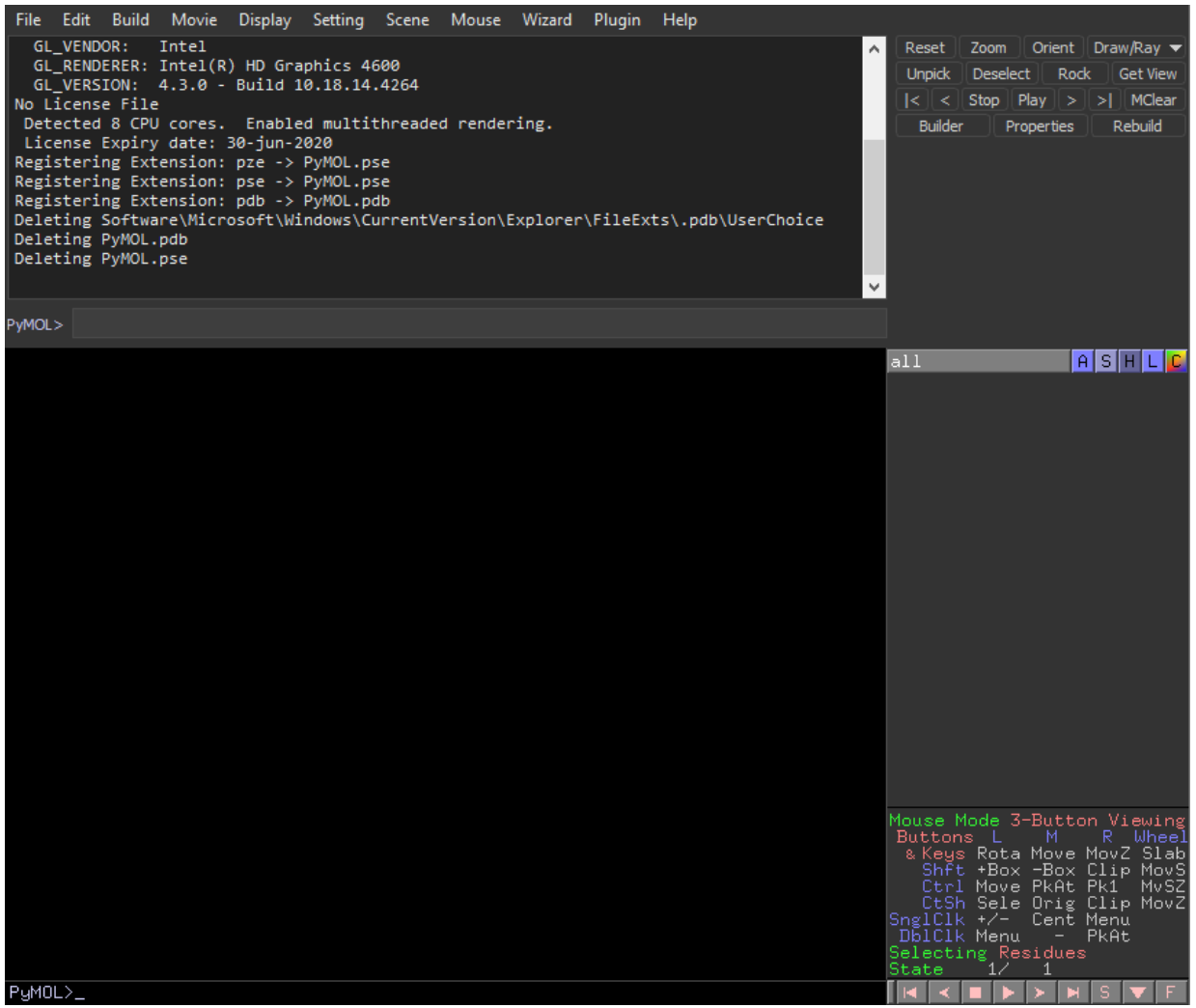
Dots



Volume (rainbow)



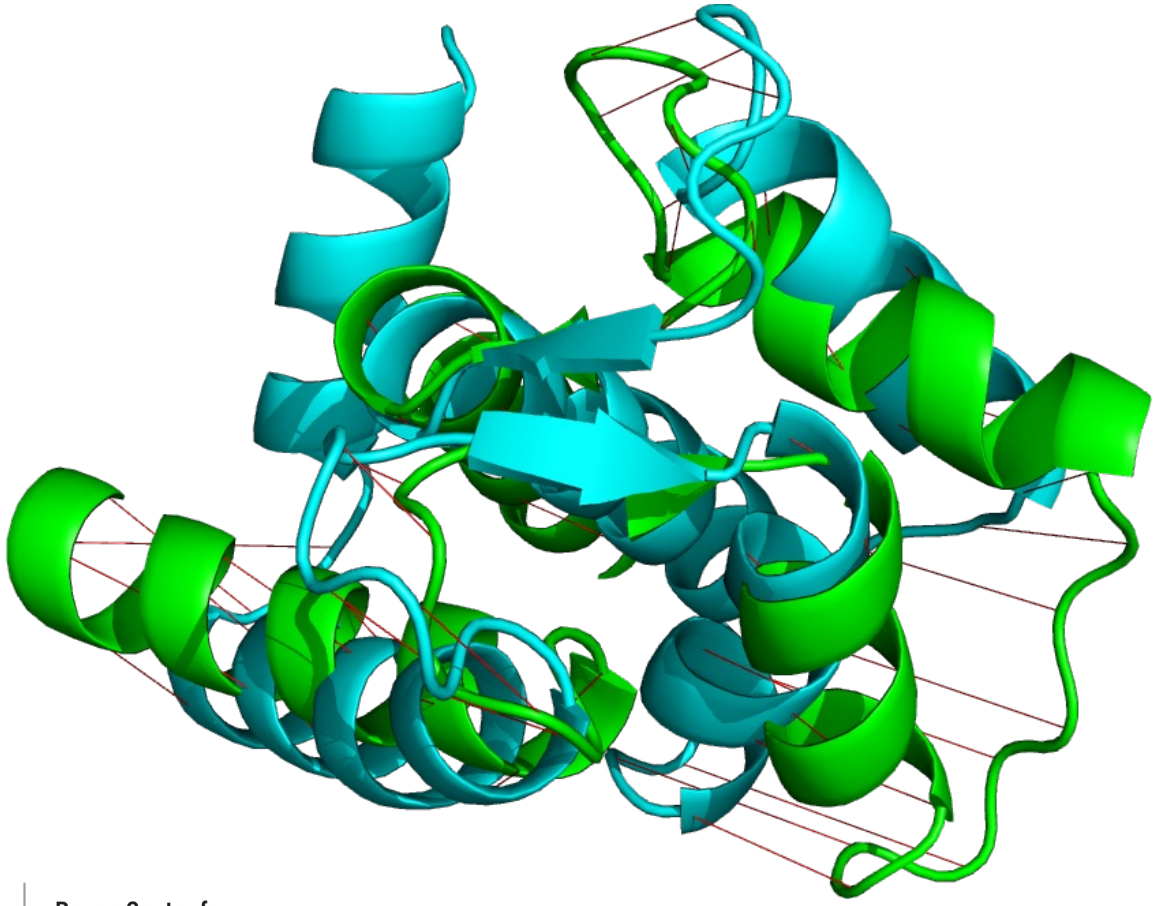
Rendering Publication-Quality Images



Ray tracing render menu

Structural Alignment

```
  16    21    26    31    36    41    46    51    56    61    66    71    76    81    86
LSEEMIAEFKAAFDMFDADGGGDISTKELGTVMRMLGQNPTKEELDAIIEEVDEDGSGTIDFEEFLVMMVRQMKE
  6     11    16    21    26    31    36    41    46    51    56    61    66    71    76
LTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNPTEAELQDMINEVDADGNGTIDFPEFLTMMARKMKD
```



Structural Alignment Methods

PyMOL provides several different alignment algorithms

- `align mobile, target`
  - Fast, assumes sequence similarity (~homology)
- `super mobile, target`
  - Expensive, assumes structural similarity
- `cealign fixed, mobile`
  - Most expensive, use for partial structural similarity
- Root-mean-square deviation (RMSD)

Selection Algebra

PyMOL provides a rich syntax for specifying what you interact with

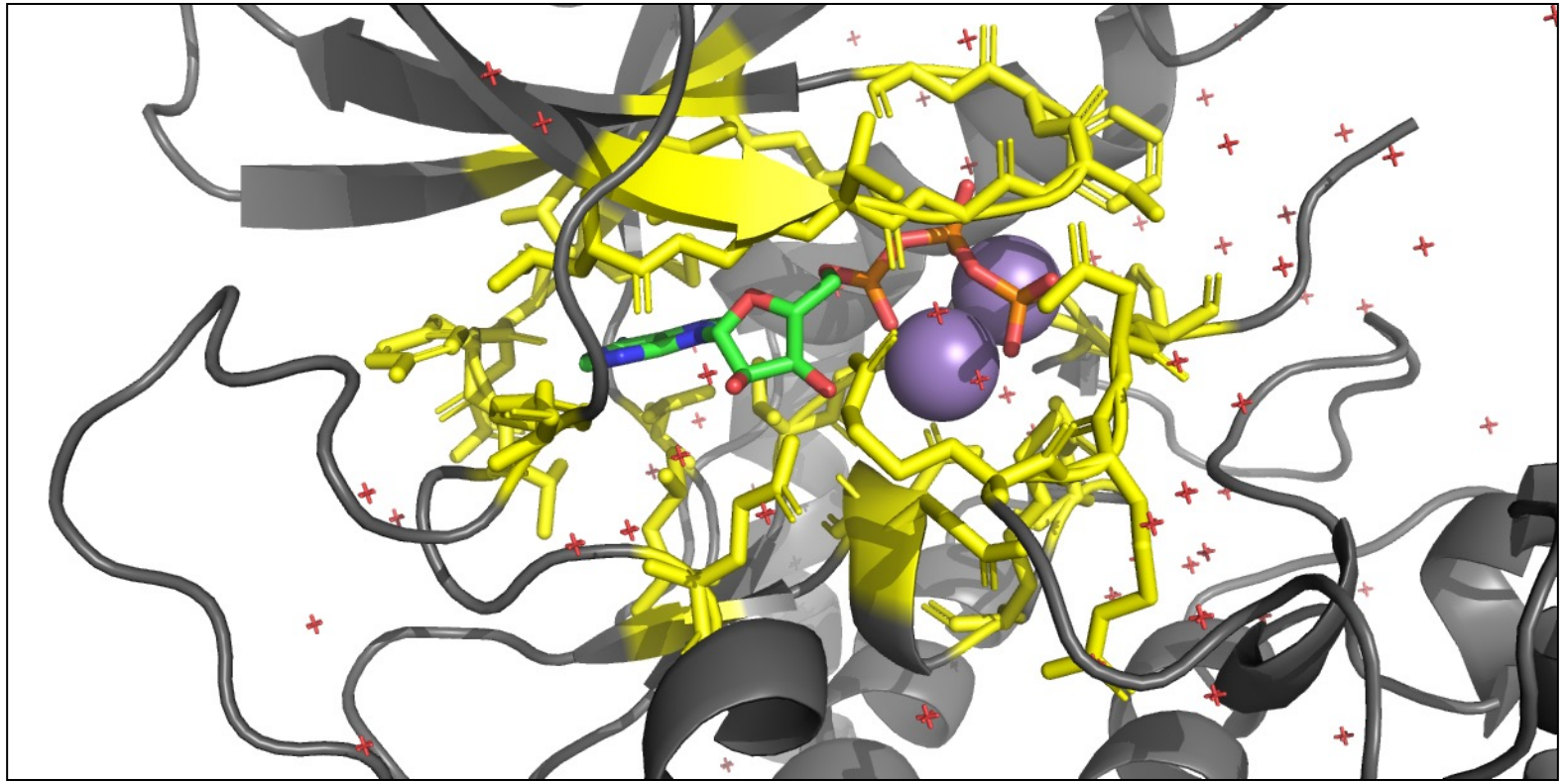
- Can select sets of atoms based on predicates
  - Boolean operators (and, or)
  - Identifiers (object name, chain, residue name, residue number, atom name)
  - Membership (is part of an object)
  - Proximity (within distance, expand by distance)
  - Many, many others…
- https://pymolwiki.org/index.php/Selection_Algebra
- Can use selections in place of objects most of the time (e.g. aligning, coloring)

Example 1

Highlight interface between PKA C-alpha and ATP in PDB 1ATP

- Download 1ATP from PDB
  - `fetch 1atp`
- Color the protein neutrally, since we'll want to focus on ATP and the interfacial residues
  - `color gray40, polymer`
- Find all polymer atoms within 5.0Å of residues named "ATP", expand that to include whole residues, and name the selection object "atp_iface"
  - `select atp_iface, br. (polymer within 5.0 of (resn atp))`
- Color the interfacial residues yellow
  - `color yellow, atp_iface`
- Show stick representations of the interfacial residues
  - `show sticks, atp_iface`

Example 1 Rendered

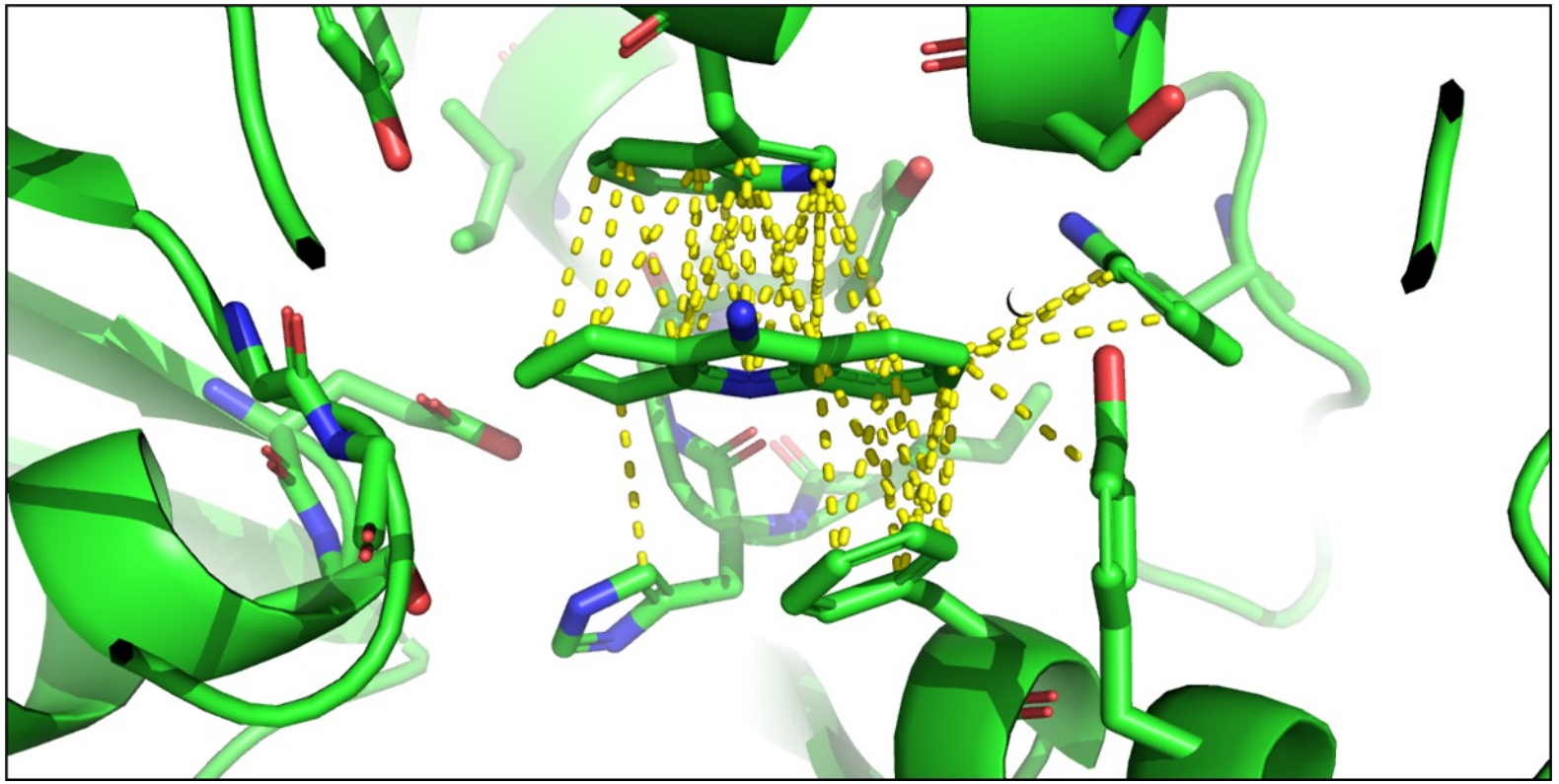


Example 2

Investigate stacking interactions between AChE and tacrine, a drug approved for treatment of Alzheimer's disease

- Download 1ACJ from PDB
  - `fetch 1acj`
- Select the interfacial residues, etc
  - `select iface, br. polymer within 5.0 of resn tha`
  - `show sticks, iface`
- Select rings in the interfacial residues and tacrine
  - `select iface_rings, byring iface`
  - `select tac_rings, byring resn tha`
- Create distance objects between ring atoms closer than 4.0Å
  - `distance mydist, iface_rings, tac_rings, 4.0`
- Hide the distance labels
  - `hide labels, mydist`

Example 2 Rendered



PyMOL commands can be loaded from a file

- In the command prompt, run:
  - `@<path to script file>`
- Best practice is to save the commands you used to a script
  - Ensures that you can reproduce a rendering later
  - Same applies to e.g. alignment or selection commands used for analysis
- Advanced usage
  - In your own Python environment, can `import pymol` and use commands
  - Can run scripts without opening the GUI
  - See the PyMOL wiki at https://pymolwiki.org

Movies

PyMOL commands can be loaded from a file

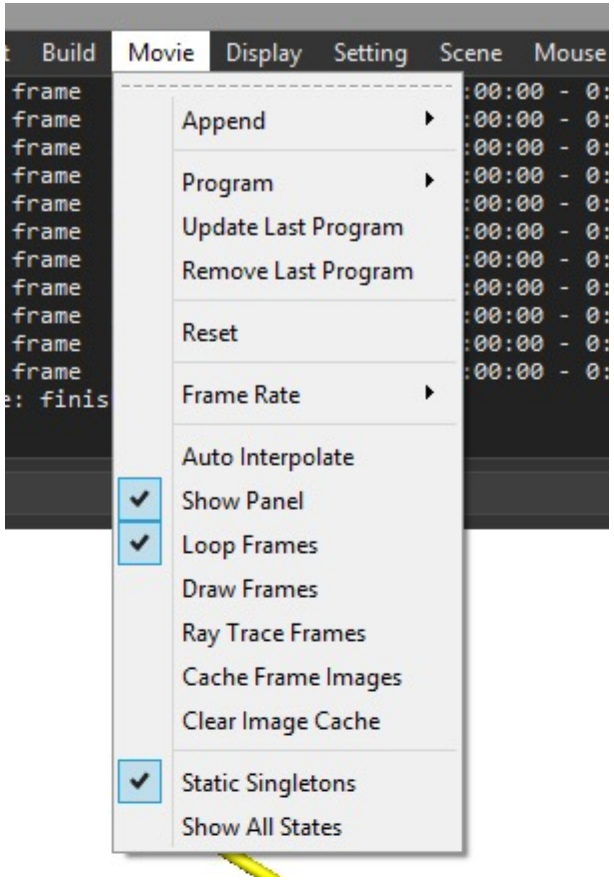
- Store scenes and use

  Program → Scene Loop

  To generate simple movies

- More complicated movies can be generated using scripts

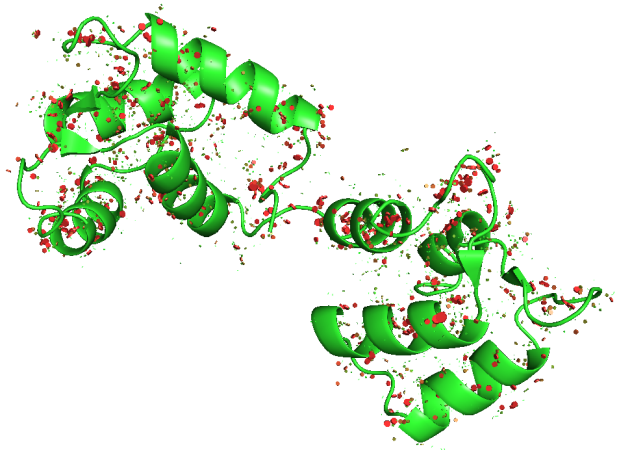
- PyMOL wiki contains advanced moviemaking instructions
  - https://pymolwiki.org/index.php/MovieSchool



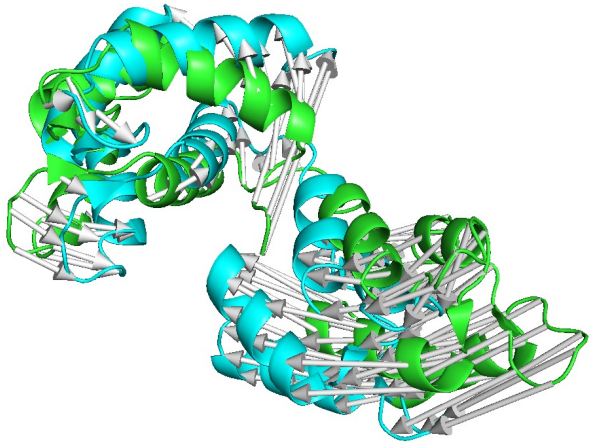
Plugins

Plugins can be installed through the built-in plugin manager

- `show_bumps` highlights areas of high van der Waals strain



- `modevectors` draws arrows in the direction of motion between states



USER EXAMPLES & QUESTIONS