phenix.pdbtools commandline tutorial

phenix.pdbtools is a powerful commandline tool for simple tasks editing PDB and mmCIF files.

PDB tools is also available in the Phenix GUI under Models -> Modification, minimization, and dynamics. For single modifications, the GUI tool is recommended.

Many commands in phenix.pdbtools make use of a selection syntax, documented here: https://phenix-online.org/documentation/reference/atom_selections.html

More complete and traditional documentation for pdbtools is here: https://phenix-online.org/documentation/reference/pdbtools.html

Setting up the tutorial data

In your terminal, cd to your preferred working space.

```
mkdir pdbtools_tutorial
cd pdbtools_tutorial
phenix.fetch_pdb 2re8
```

This is a (glutaminyl) tRNA sythethase protein with its corresponding tRNA.

You can also get the mmCIF version by adding the -c flag. `phenix.fetch_pdb -c 2re8`. We will use PDB format in this tutorial because it is more human-readable.

You will want to have a text editor and/or molecular viewer handy to view the results of your commands. You can also use the command "less <your filename>" to view a file in the terminal. Use up and down arrows to scroll' press q to quit and return to the commandline.

Discovering options

pdbtools contains many possible options. To see a comprehensive display of the options and their default values, use the flag --show-defaults. This flag can be used with any Phenix commandline tool.

```
phenix.pdbtools --show-defaults
```

A copy of this output is included at the end of this document for reference.

Simple edits

Chain Selection

Most commands can be written in a shorter form, if the option names are unique. This tutorial will use the verbose forms, for maximum clarity.

2re8.pdb contains two chains; chain A is protein, and chain B is RNA. You might reasonably want to extract just one of these chains. Use modify.keep, followed by a selection, to choose a chain.

```
phenix.pdbtools 2re8.pdb modify.keep="chain B"
```

This will create a file named 2re8_modified.pdb.

Any future commands on 2re8.pdb will overwrite this automatically-named output. We will use explicit output names to avoid this behavior.

```
phenix.pdbtools 2re8.pdb modify.keep="chain B and resseq 950:960"
output.filename=2re8_chainB_950-960.pdb
```

pdbtools can make use of certain keywords in selections. Try the following:

```
phenix.pdbtools 2re8.pdb modify.keep="rna"
output.filename=2re8_rna_only.pdb
```

How does the 2re8_rna_only.pdb file differ from the 2re8_modified.pdb file created by selecting "chain B"?

Dangerous power

Commandline tools are very powerful, and they can let you do strange things.

```
phenix.pdbtools 2re8.pdb modify.remove="element O"
output.filename=2re8_cannot_breathe.pdb
```

The resulting file contains no oxygens of any kind! There's probably no reason to do this, but you can! So be careful what you ask for – you'll probably get it.

Water removal

A more reasonable way to remove a lot of oxygens from a file is to strip out all the waters.

```
phenix.pdbtools 2re8.pdb modify.remove="resname HOH"
output.filename=2re8_no_swimming.pdb
```

If you are doing structural studies, rather than solving structures, selecting out certain kinds of residues may be useful:

```
phenix.pdbtools 2re8.pdb modify.keep="resname U"
output.filename=2re8_u_are_the_only_one.pdb
```

(Selecting ligands is probably a more reasonable thing to do)

Fancier edits

Truncate to poly-alanine.

Multiple options can be combined in a single command. This command will return just the protein chain, with all the sidechains trimmed down to their CB atoms.

```
phenix.pdbtools 2re8.pdb modify.keep="chain A"
modify.truncate_to_polyala=True output.filename=2re8_aaaaaa.pdb
```

"Within"

To select something of interest plus its local environment, use the "within" selection. Useful for contacts from individual ligands to entire interfaces. Syntax is within(distance value , regular selection)

```
phenix.pdbtools 2re8.pdb modify.keep="within(7, chain A and resseq
260)" output.filename=2re8_closefriends.pdb
```

This command keeps only atoms within the specified radius, which results in partial residues. To keep all of each residue with at least one atom within the radius, use "residues_within" instead.

```
phenix.pdbtools 2re8.pdb modify.keep="residues_within(7, chain A and
resseq 260)" output.filename=2re8_allmyclosefriends.pdb
```

Shake

You can introduce a little randomness into a model with the Shake option. It will randomly change atomic coordinates of the selection within an rmsd value. Careful! Passing shake a large value will disintegrate your model!

```
phenix.pdbtools 2re8.pdb modify.selection="chain A"
modify.sites.shake=0.2 output.filename=2re8_rattleroll.pdb
```

Move (Translate)

Sometimes it is useful to move models into a standard reference frame. This command will translate the model through space. The translation vector in this example send the first CA of the protein chain to the origin (0,0,0).

```
phenix.pdbtools 2re8.pdb modify.keep="chain A"
modify.sites.translate="-21.463 -28.009 -14.045"
output.filename=2re8_ca_to_origin.pdb
```

Move (Rotate)

Our final and most difficult command. This will take all the atoms in the file, and rotate them 90 degrees around the y axis (0 0 0 1 0 0).

```
phenix.pdbtools 2re8_ca_to_origin.pdb
modify.rotate_about_axis.atom_selection="all"
modify.rotate_about_axis.axis="0 0 0 1 0 0"
modify.rotate_about_axis.angle=90 output.filename=2re8_roundabout.pdb
```

Note that these translation and rotation operations only affect the model. If you want to move the map as well, other tools are necessary. Check “Superpose Maps” for the most likely use case.

phenix.pdbtools --show-defaults

```
modify {
  remove = None
  keep = None
  put_into_box_with_buffer = None
  selection = None
  flip_symmetric_amino_acids = False
  adp {
    atom_selection = None
    randomize = False
    set_b_iso = None
    convert_to_isotropic = False
    convert_to_anisotropic = False
    shift_b_iso = None
    scale_adp = None
  }
  sites {
    atom_selection = None
    shake = None
    switch_rotamers = max_distant min_distant exact_match fix_outliers
    translate = 0 0 0
    rotate = 0 0 0
    euler_angle_convention = *xyz zyz
  }
  occupancies {
    atom_selection = None
    randomize = False
    set = None
  }
  rotate_about_axis {
    axis = None
    angle = None
    atom_selection = None
  }
  change_of_basis = None
  renumber_residues = False
  increment_resseq = None
  truncate_to_polyala = False
  truncate_to_polygly = False
  remove_alt_confs = False
  always_keep_one_conformer = False
  set_chemical_element_simple_if_necessary = None
  set_seg_id_to_chain_id = False
  clear_seg_id = False
  convert_semet_to_met = False
  convert_met_to_semet = False
  rename_chain_id {
```

```
    old_id = None
    new_id = None
  }
  set_charge {
    charge_selection = None
    charge = None
  }
  neutralize_scatterers = False
  remove_fraction = None
  random_seed = None
  move_waters_last = False
}
output {
  prefix = None
  suffix = _modified
  serial = None
  overwrite = True
  filename = None
  file_name = None
  serial_format = "%03d"
}
job_title = None
gui {
  output_dir = None
}
```