Common File Formats in Rosetta

Steven Combs

The Files

- Flags/Option files
- Resfiles
- Params
- PDB
- Silent
- Atom tree diffs

Flags / Options

- Paramaters for protocols
- Formatted in 3 different ways

Command Line

```
Fixbb.release –database <database> -s 1thfD.pdb –ex1 –ex2
–packing:repack_only –out:prefix repack_ -resfile 1thfD.resfile
```

Flags / Options

- Paramaters for protocols
- Formatted in 3 different ways

space / tab indented

Fixbb.release @options.txt

```
–database <database>
-in
 -file
  -s 1thfD.pdb
-out
 -prefix repack_
-packing
 -ex1
 -ex2
 -repack_only
```

Flags / Options

- Paramaters for protocols
- Formatted in 3 different ways

Mixed

Fixbb.release @options.txt

```
–database <database>
-in
 -file
  -s 1thfD.pdb
-out
 -prefix repack_
-packing:ex1
-packing:ex2
-packing:repack_only
```

Mixed

Fixbb.release –s 1thfD.pdb –ex1 –ex2 @options.txt

```
–database <database>
-out
 -prefix repack_
-packing:repack_only
```

<header> start <body>

Resfiles

- Defines which residues to pack/design
- Great documentation!
- http://www.rosettacommons.org/manuals/archive/rosetta3.2_user_guide/resfiles.html

Params Files

- Definition for residues
- Residues are defined by type (ie residue ala is typed ALA)
- Residue types found in: `database/chemical/residue_types/fa_standard/residue_types/l-caa/`
- Ligand residue types made through molfile_to_params.py

NAME GLY Residue type
IO_STRING GLY G 3 letter name and 1 letter name
TYPE POLYMER Defines type as polymer
AA GLY Defines as an amino acid gly
ATOM N Nbb NH1 -0.47 Atom name found in pdb
ATOM CA CAbb CT2 -0.02 Atom type used by Rosetta
ATOM C COb C 0.51 Molecular mechanics atom type used by Rosetta (mainly used in backrub application)
ATOM O OCbb O -0.51 Atom charge
ATOM H HNbb H 0.31
ATOM 1HA Hapo HB 0.09
ATOM 2HA Hapo HB 0.09
LOWER_CONNECT N What atom is the polymer connected to at the lower end of residue
UPPER_CONNECT C What atom is the polymer connected to at the upper end of residue

```
BOND N  CA
BOND N  H
BOND CA C
BOND CA 1HA
BOND CA 2HA
BOND C  O
```

Defines what atoms are bonded to each other

PROPERTIES PROTEIN Different properties, can be POLAR, CHARGED, AROMATIC, etc etc
NBR_ATOM CA Closest atom to the center of mass
NBR_RADIUS 3.4473 Radius of neighbor atom
FIRST_SIDECHAIN_ATOM NONE Where does the sidechain begin
ACT_COORD_ATOMS CA END Where the center of charge is. Used by fa_pair for polar amino acids

```
ICOOR_INTERNAL   N     0.000000   0.000000   0.000000  N   CA   C
ICOOR_INTERNAL   CA    0.000000 180.000000   1.458001  N   CA   C
ICOOR_INTERNAL   C     0.000000  68.799995   1.523259  CA  N    C
ICOOR_INTERNAL  UPPER 149.999969  63.800018   1.328685  C   CA   N
ICOOR_INTERNAL   O  -179.999985  59.200005   1.231015  C   CA  UPPER
ICOOR_INTERNAL   1HA 120.253975  70.928650   1.090168  CA  N   C
ICOOR_INTERNAL   2HA 119.463875  70.913704   1.089353  CA  N   1HA
ICOOR_INTERNAL  LOWER -150.000015  58.300003   1.328685  N   CA   C
ICOOR_INTERNAL   H   180.000000  60.850040   1.010000  N   CA  LOWER
```

Internal coordinate system. Defines how atoms are connected to each other in 3D space

PDB Files

- Holds 3D coordinates for proteins

Atom # | Atom name | Residue name | Chain ID | Residue # | X-coord | Y-coord | Z-coord | occupancy | B-factor | Element name

```
ATOM   1953  N   LEU D 253      32.832  26.636  -4.918  1.00 29.10           N
ATOM   1954  CA  LEU D 253      33.906  26.053  -4.131  1.00 30.18           C
ATOM   1955  C   LEU D 253      35.149  25.739  -4.939  1.00 31.08           C
ATOM   1956  O   LEU D 253      36.218  25.466  -4.351  1.00 32.40           O
ATOM   1957  CB  LEU D 253      33.495  24.738  -3.459  1.00 29.02           C
ATOM   1958  CG  LEU D 253      32.701  24.843  -2.152  1.00 28.80           C
ATOM   1959  CD1 LEU D 253      32.214  23.456  -1.719  1.00 27.59           C
ATOM   1960  CD2 LEU D 253      33.576  25.444  -1.045  1.00 28.02           C
TER    1961      LEU D 253
HETATM 1962  P   PO4 D 305      31.557  25.942  19.190  1.00  8.50           P
```

Atom lines

Ligand lines

Silent Files

- Specify by `-in:file:silent` and `-out:file:silent` (as opposed to `-in:file:s <pdb>` and `-out:pdb`)
- These are smaller files compared to PDBs. They store the internal coordinates of the model in terms of dihedral angles.
- If want to input/output silent files, also specify `-in:file:silent_struct_type binary` and `-out:file:silent_struct_type binary`
- Binary silent files store the same information in ASCII binary and are even more compressed and more robust in terms of extracting PDB coordinates from them.

Atom Tree Diffs

- Only used in ligand docking (output that contains models)
- No specific extension (ie. *.out)
- Allows for storing only coordinates that have changed between a reference model
- Extracted using extract_atomtree_diffs.release