

# Rosetta Input/Output

Darwin Fu

[darwinyfu@gmail.com](mailto:darwinyfu@gmail.com)

Rosetta Workshop

April 29, 2014

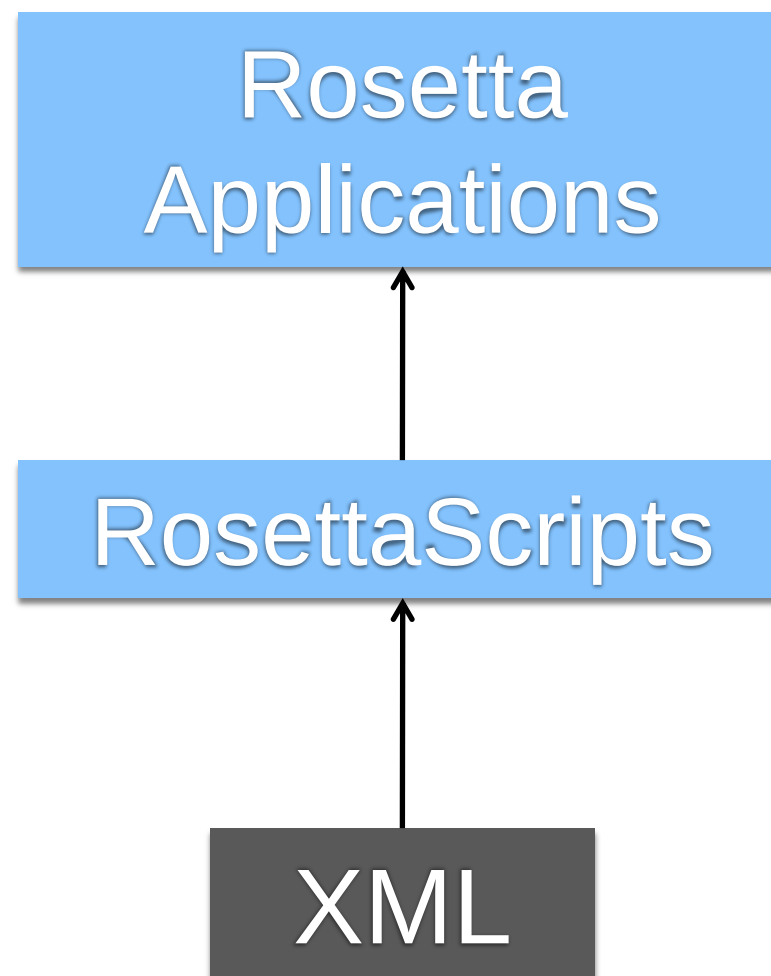
# How do we get Rosetta?

- Since August 2013, Rosetta is on a rolling “weekly release” cycle
- New versions are released roughly every week
- Every revision passes basic scientific performance tests
- All tutorials use version 2013wk52 (Dec 2013)
- The latest versions can be downloaded here:
  - <https://www.rosettacommons.org/software/>

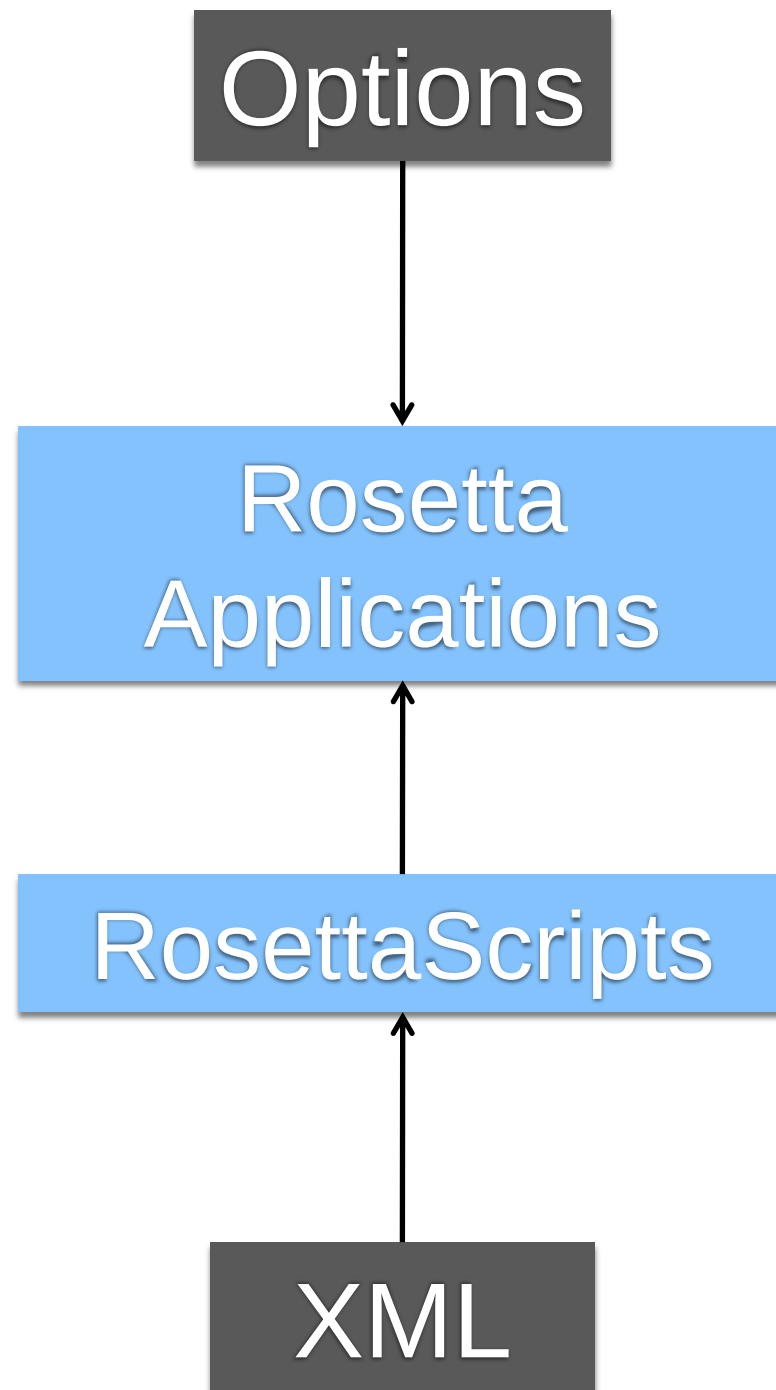
# What goes In, What Comes out

Rosetta  
Applications

# What goes In, What Comes out



# What goes In, What Comes out



# Types of Options

Full List of Options can be found at:

<https://www.rosettacommons.org/docs/latest/full-options-list.html>

- Input
  - Location of Specific Files
  - Structure
- Output
  - Filename and Filetypes
  - Number of Structures
- Rotamer Library Options
- Application Specific Options

# How do we enter options?

## On the Command Line:

score\_jd2.default.linuxgccrelease -in:file:s test.pdb

## In an Options File

score\_jd2.default.linuxgccrelease @options.txt

## Options files are colon, space, or tab delimited

Colon: -in:file:s test.pdb

Space delimited:

-in

-file

-s test.pdb

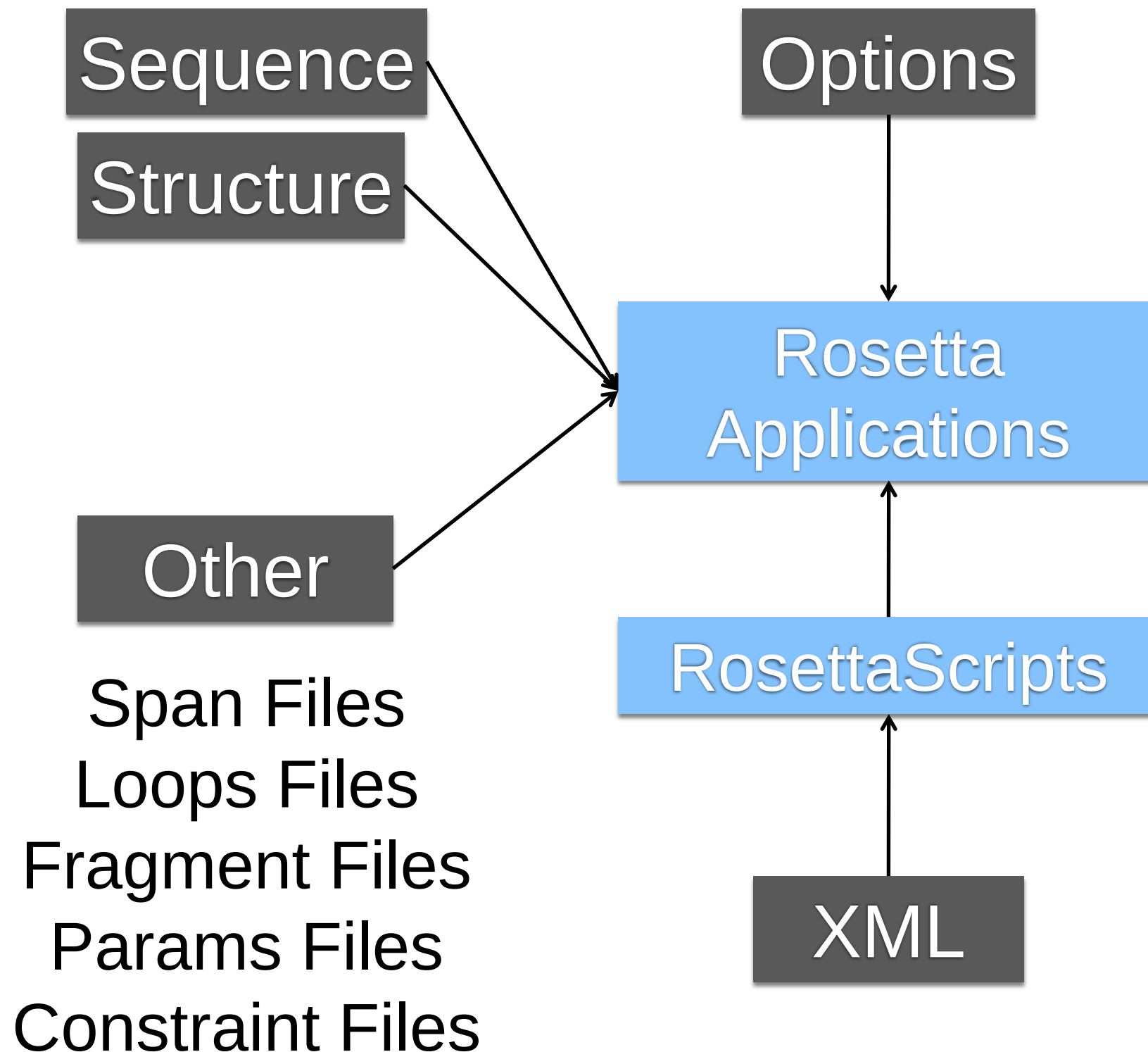
Tab delimited:

-in

-file

-s test.pdb

# What goes In, What Comes out



# Rosetta has two main formats for inputting structures

- PDB Files
  - International standard
  - Readable by PyMOL, MOE, etc.
- Silent Files
  - Rosetta specific
  - compact, fast to output

# When should each format be used?

- PDB files
  - Small numbers of output structures
  - Interacting with an existing processing pipeline that requires PDB files
  - Proteins with bound Ligands
- Silent files
  - Archiving large numbers of structures

# What's in a PDB file?

| Record Name | Atom Number | Atom Name | Alternate Location | Residue Name | Chain ID | Residue Number | Insertion Code | x coordinate | y coordinate | z coordinate | Occupancy | Temperature | Element | Charge |
|-------------|-------------|-----------|--------------------|--------------|----------|----------------|----------------|--------------|--------------|--------------|-----------|-------------|---------|--------|
| ATOM        | 556         | N         | LEU                | A            | 71       |                |                | 32.710       | 35.821       | 23.137       | 1.00      | 13.56       | N       |        |

- Full specification is documented here:  
<http://www.wwpdb.org/docs.html>
- Column delimited!
  - Field alignment matters
- Editing by hand is unpleasant, use a tool:
  - <http://biopython.org/wiki/Biopython>
  - [http://www.bioperl.org/wiki/Main\\_Page](http://www.bioperl.org/wiki/Main_Page)
- Rosetta has a number of useful scripts for processing PDBs
  - Clean\_pdb.py

# How Rosetta Reads a PDB

- Rosetta only reads ATOM/HETATM
- Residues are connected in the order they appear in the PDB file
- It is often convenient to number residues from 1 and without gaps
- Hydrogens and missing side-chain Atoms are added as needed
- Residues with Zero Occupancy Backbone Atoms are removed

# Protein and binary silent files

- Protein and Binary
  - Binary silent files are more compact, Protein is used infrequently
  - Structures with non-ideal bond lengths must use the binary format
  - Rosetta makes this decision automatically

# Anatomy of an Protein silent file

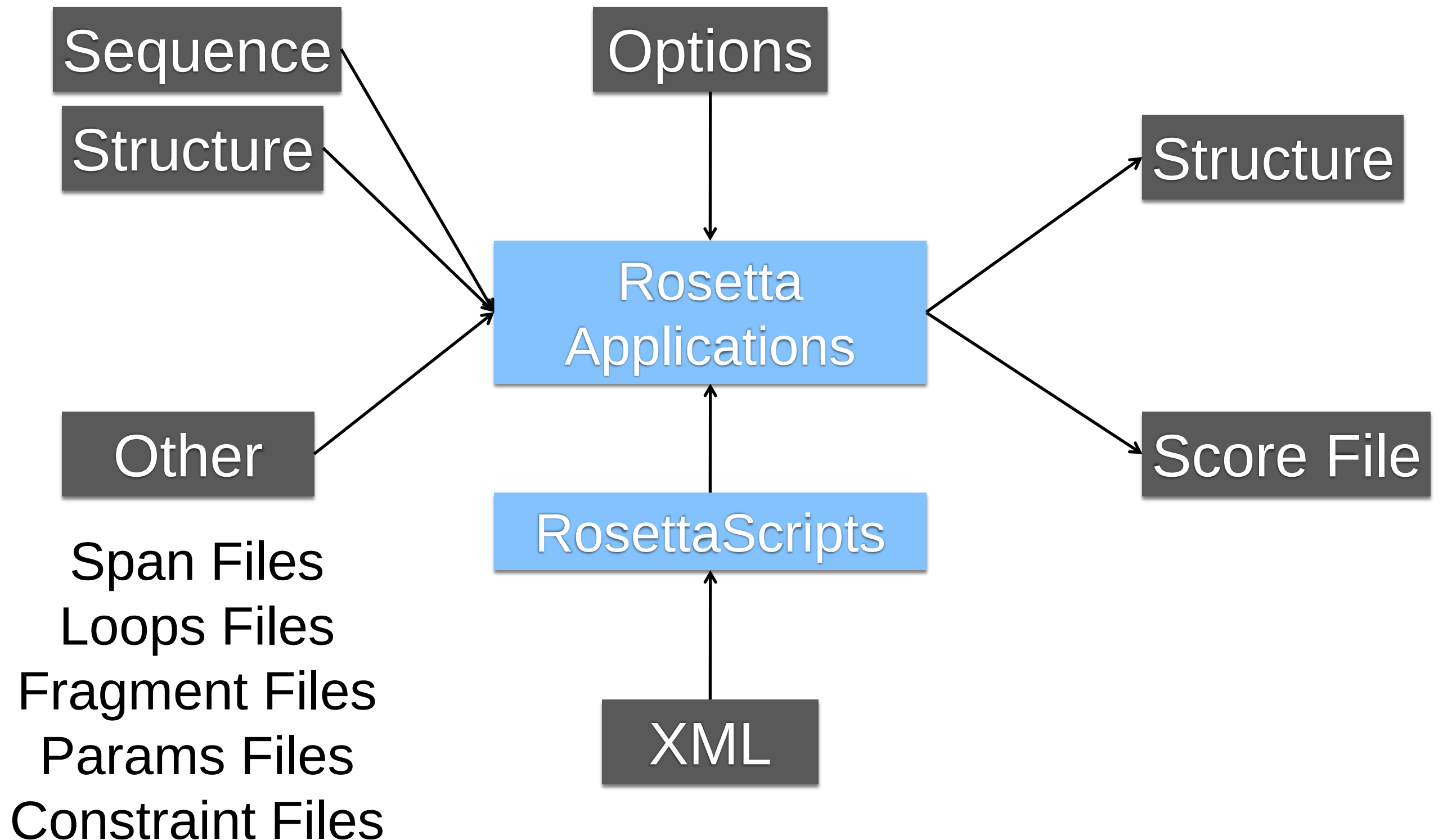
```
SEQUENCE: MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVL
SCORE:      score      env      pair      vdw      hs      ss      sheet      cb      rsigma hb_srbb hb_lrbb      rg
co contact      rama rms rms_min mxn mxrms mxlgE %cont mm11 mm22 mm33 mm43 mm74 gdtmm description
SCORE:      -95.27 -32.30 -20.00      2.43      0.62 -27.65      0.34      29.55 -18.72 -23.85 -8.55      13.29
23.63      0.00 -37.02 9.52 6.40 33 3.53 -3.5 -0.08 0.239 0.310 0.324 0.338 0.465 0.335 S_0001_1234
  1 E -97.266 124.773 -180.144      0.000      0.000      0.000 S_0001_1234
  1 E -97.266 124.773 -180.144      0.000      0.000      0.000 S_0001_1234
  2 E -111.820 110.440 -178.678      0.549      0.239      3.755 S_0001_1234
  3 E -97.729 112.193 178.878 -1.235 -2.392      5.841 S_0001_1234
[.....]
70 E -119.861 122.508 -178.150      4.782 -7.654      10.297 S_0001_1234
71 E -114.760 112.405 -180.314      3.854 -8.493      13.887 S_0001_1234
```

- SEQUENCE: the protein sequence
- SCORE: the score term names and scores for the protein
- coordinates (whitespace delimited):
  - residue number, secondary structure, phi, psi, omega, chi angles tag

# Application Specific Input Files

- Span File
  - Defines which residues are in the membrane
- Loops File
  - Identifies the loop residues for loop closure
- Params File
  - Custom parameters for small molecules or non-canonical amino acids
- Constraint File
  - Experimentally derived restraints
- Fragment File
  - Short protein segments used for comparative modeling and de novo folding

# What goes In, What Comes out



# How Rosetta Outputs a Structure

- PDB or Silent File
- Output Formatting Options
  - **Numbering** - GPCR\_model\_mGlu\_loops\_**0001**.pdb
  - **Prefixes** - **GPCR\_model**\_mGlu\_loops\_0001.pdb
  - **Suffixes** - GPCR\_model\_mGlu\_**loops**\_0001.pdb
- Score Table Appended At End of PDB
  - Per residue scoring terms
  - Whole structure scoring terms

# Score Table Appended At End of PDB

```
PHE_422 -7.11646 0.76578 2.36204 0.02356 0 0 0 0 -0.64839 0 0 0 0 0 0.12472 0.03945 2.06507 -0.15901 0.63 -1.91324
ARG_423 -4.25187 0.17744 2.56634 0.00827 0 0.03156 0 0 -0.13207 0 0 0 0 0 -0.01483 0.10937 3.61511 -0.03132 -0.98
LYS_424 -3.44202 0.52021 1.57514 0.00285 0 -0.14517 0 0 0 0 0 0 0 0 -0.08194 0.13082 1.53361 -0.04597 -0.65 -0.602
ALA_425 -3.48954 0.12844 1.51765 0.00078 0 0 0 0 0 0 0 0 0 0 -0.23778 0.0249 0 -0.09255 0.16 -1.98812
PHE_426 -5.91744 1.05261 1.54983 0.02913 0 0 0 0 0 0 0 0 0 0 -0.0011 0.11344 4.56074 -0.02141 0.63 1.9958
LEU_427 -2.50776 0.22554 1.20618 0.00421 0 0 0 0 0 0 0 0 0 0 -0.17378 0.02456 3.03518 -0.01717 -0.1 1.69695
LYS_428 -2.62602 0.26559 1.37597 0.00417 0 -0.22518 0 0 0 0 0 0 0 0 -0.19508 0.04368 4.05294 -0.05444 -0.65 1.9916
ILE_429 -2.87381 0.78189 1.06933 0.02789 0 0 0 0 0 0 0 0 0 0 0.0117 0.00476 4.79591 0.0973 0.24 4.15496
LEU_430 -1.86222 0.1049 0.89651 0.00961 0 0 0 0 0 0 0 0 0 0 -0.1717 0.05245 3.05486 -0.07557 -0.1 1.90885
SER_431 -1.4716 0.23417 0.9824 0.0078 0 0 0 0 0 0 0 0 0 0 -0.24175 0.24861 2.92084 -0.14419 -0.37 2.16627
CYS_p:CtermProteinFull_432 -1.25209 0.1101 0.96155 0.00192 0 0 0 0 0 0 0 0 0 0 0 0 0 2.43129 0 1.7 3.95278
ETQ_433 -11.6925 1.71806 5.14068 0.10181 0 0 0 0 0 -0.20613 0 0 0 0 0 0 0 0 0 0 -4.93808
#END_POSE_ENERGIES_TABLE 3PBL_A_ETQ_0500.pdb
ClassicGrid_grid_X -47
Transform_accept_ratio 0.394
angle_constraint 0
atom_pair_constraint 0
chainbreak 0
coordinate_constraint 3.94918
dihedral_constraint 0
dslf_ca_dih 0.323363
dslf_cs_ang 1.9574
dslf_ss_dih 15.9255
dslf_ss_dst -3.6307
fa_atr -1701.37
fa_dun 543.365
fa_elec -26.6126
fa_pair -34.1045
fa_rep 310.972
fa_sol 769.672
```

# Rosetta Summary Score File

```
SEQUENCE:
SCORE:  total_score  ClassicGrid_grid_X  Transform_accept_ratio
SCORE:      -730.214          -47.000          0.300
SCORE:      -735.299          -47.000          0.346
SCORE:      -740.139          -47.000          0.320
SCORE:      -739.866          -48.000          0.316
SCORE:      -736.618          -48.000          0.282
SCORE:      -730.437          -47.000          0.356
SCORE:      -733.917          -48.000          0.382
SCORE:      -739.471          -47.000          0.410
SCORE:      -733.131          -47.000          0.336
SCORE:      -736.054          -48.000          0.278
SCORE:      -737.242          -47.000          0.346
SCORE:      -736.981          -47.000          0.298
SCORE:      -731.206          -48.000          0.356
SCORE:      -731.542          -48.000          0.352
SCORE:      -703.123          -47.000          0.282
SCORE:      -735.296          -47.000          0.342
SCORE:      -734.978          -48.000          0.304
SCORE:      -727.645          -48.000          0.292
SCORE:      -732.402          -48.000          0.338
SCORE:      -736.152          -48.000          0.340
SCORE:      -702.435          -48.000          0.360
SCORE:      -735.285          -47.000          0.354
SCORE:      -734.744          -48.000          0.322
SCORE:      -727.960          -48.000          0.362
SCORE:      -734.144          -47.000          0.308
```