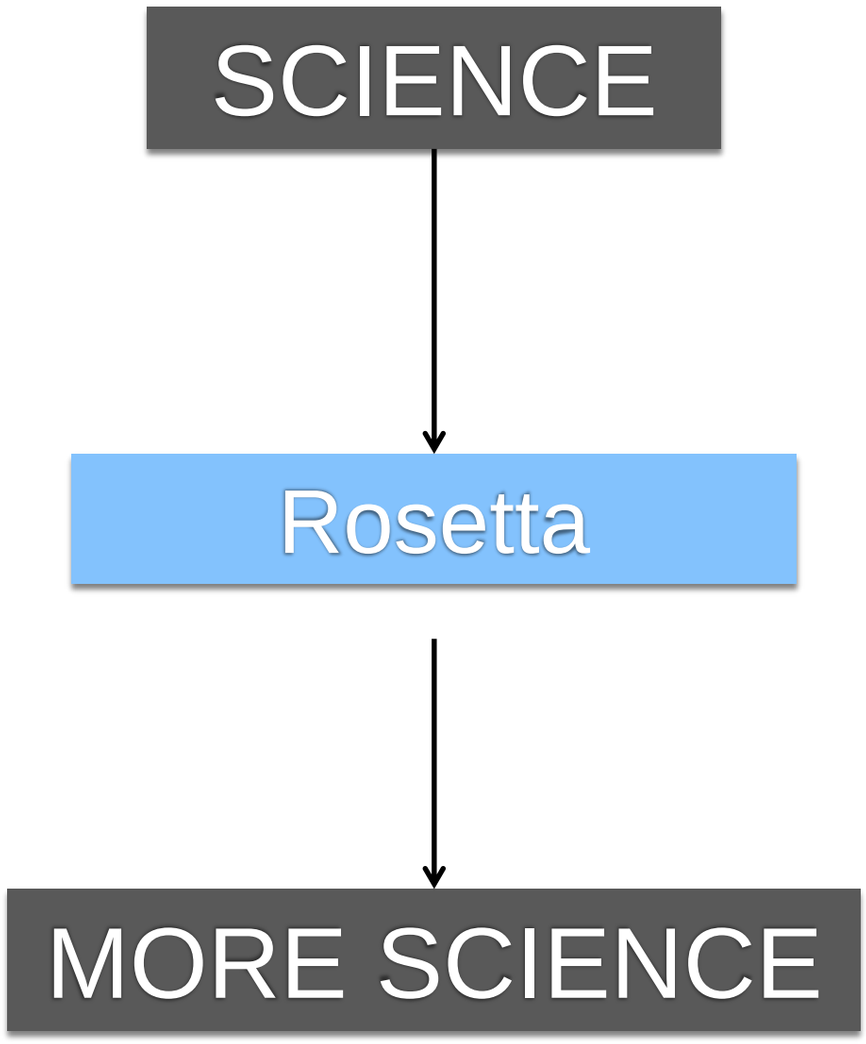
Rosetta Input/Output

Darwin Fu
darwinyfu@gmail.com
@EquationForLife

Rosetta Workshop
November 11th, 2015

What goes in, What comes out



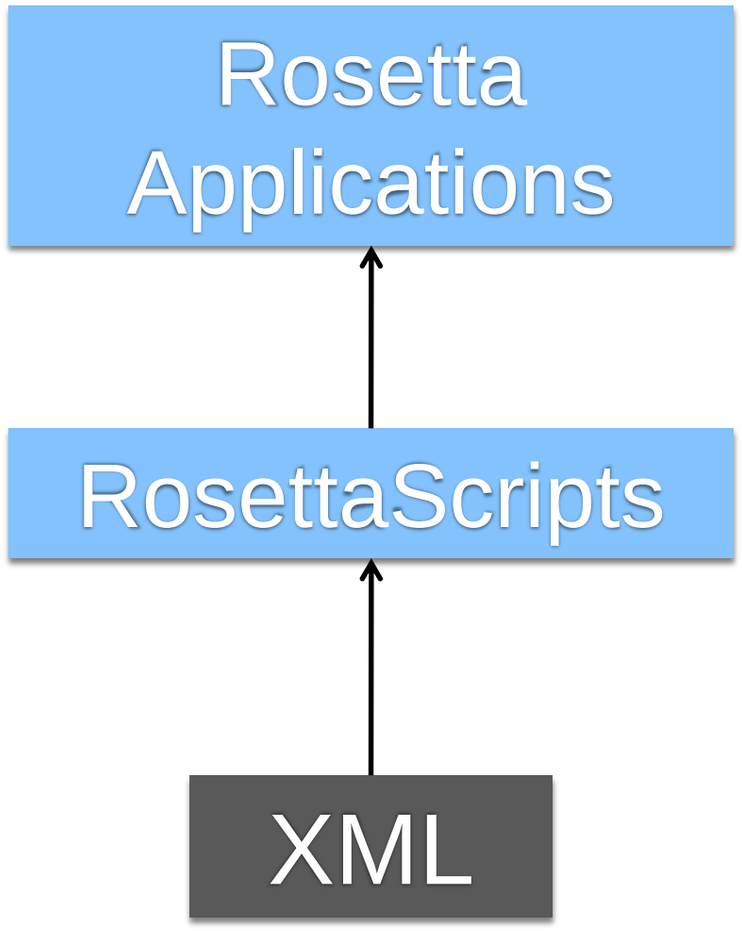
How do we get Rosetta?

- Since August 2013, Rosetta is on a rolling “weekly release” cycle
- New versions are released roughly every week
- Every revision passes scientific performance tests
- All tutorials use version 2015.31.58019 (September 2015)
- The latest versions can be downloaded here:
  - https://www.rosettacommons.org/software/
- Documentation:
  - https://www.rosettacommons.org/docs/latest/Home
- Glossary:
  - https://www.rosettacommons.org/docs/latest/rosetta_basics/Glossary/Glossary

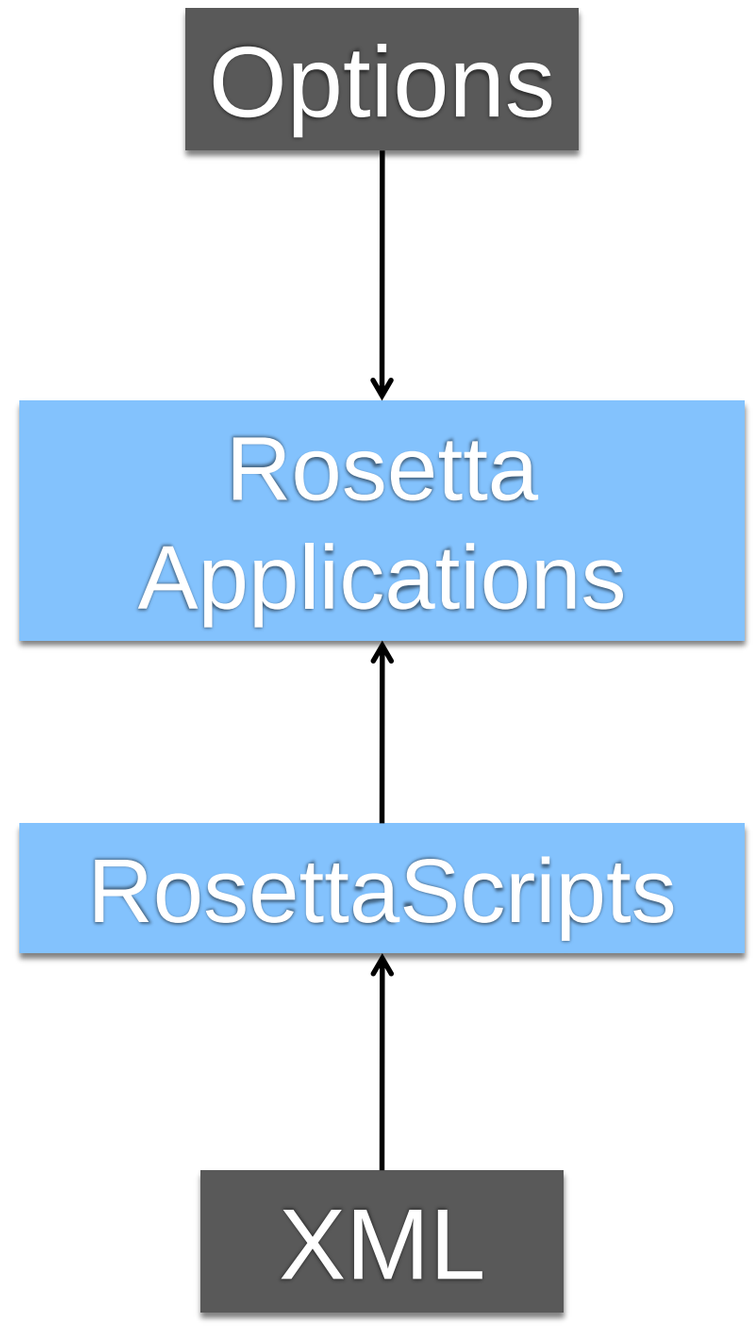
Rosetta is actually “suite”

- **Applications: “Rosetta”- user friendly protocols**
  - rosetta_scripts
  - score_jd2
  - relax
  - loop_model
- **Scripts: prepares/analyzes Rosetta IO files**
  - clean_pdb.py
  - pdb_renumber.py
  - molfile_to_params.py

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

- 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

How do we enter options?

<u>On the Command Line:</u>

score_jd2.default.linuxgccrelease -in:file:s test.pdb

<u>In an Options File</u>

score_jd2.default.linuxgccrelease @options.txt

<u>Options files are colon, space, or tab delimited</u>

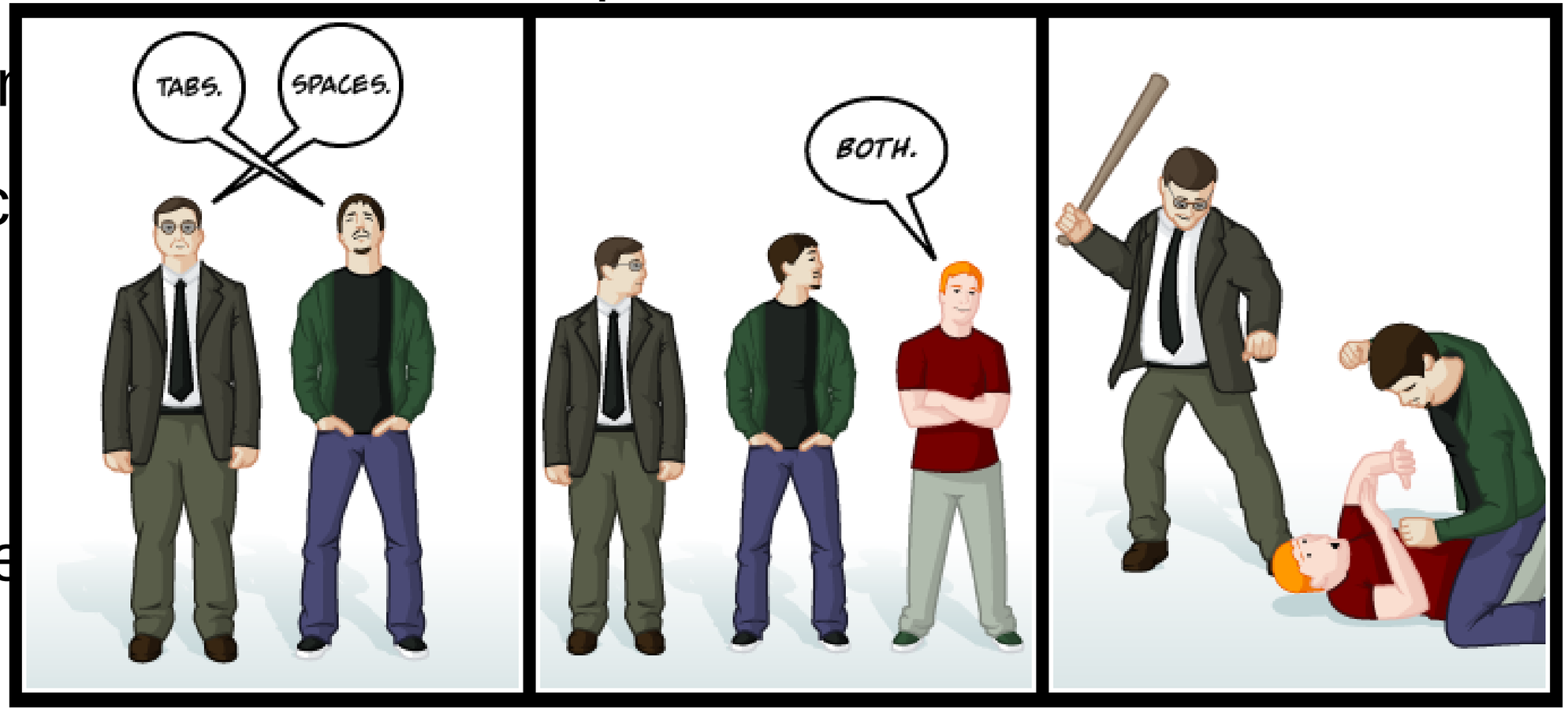
Color

Spac

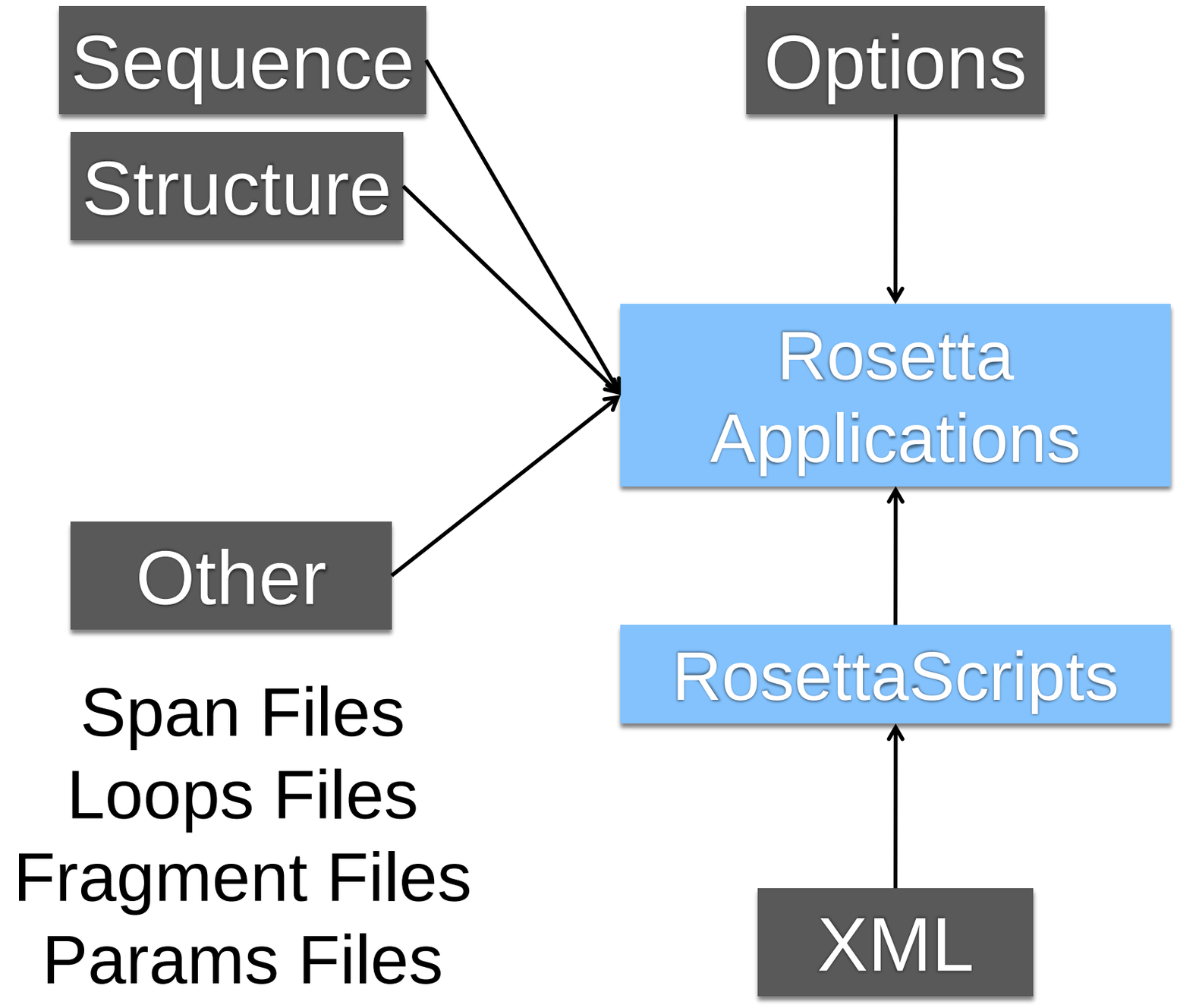
-in

-file

-s te



What goes in, What comes out



Span Files
Loops Files
Fragment Files
Params Files
Constraint Files

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
  - Can also be gzipped
- 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/documentation/file-format
- 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
- 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

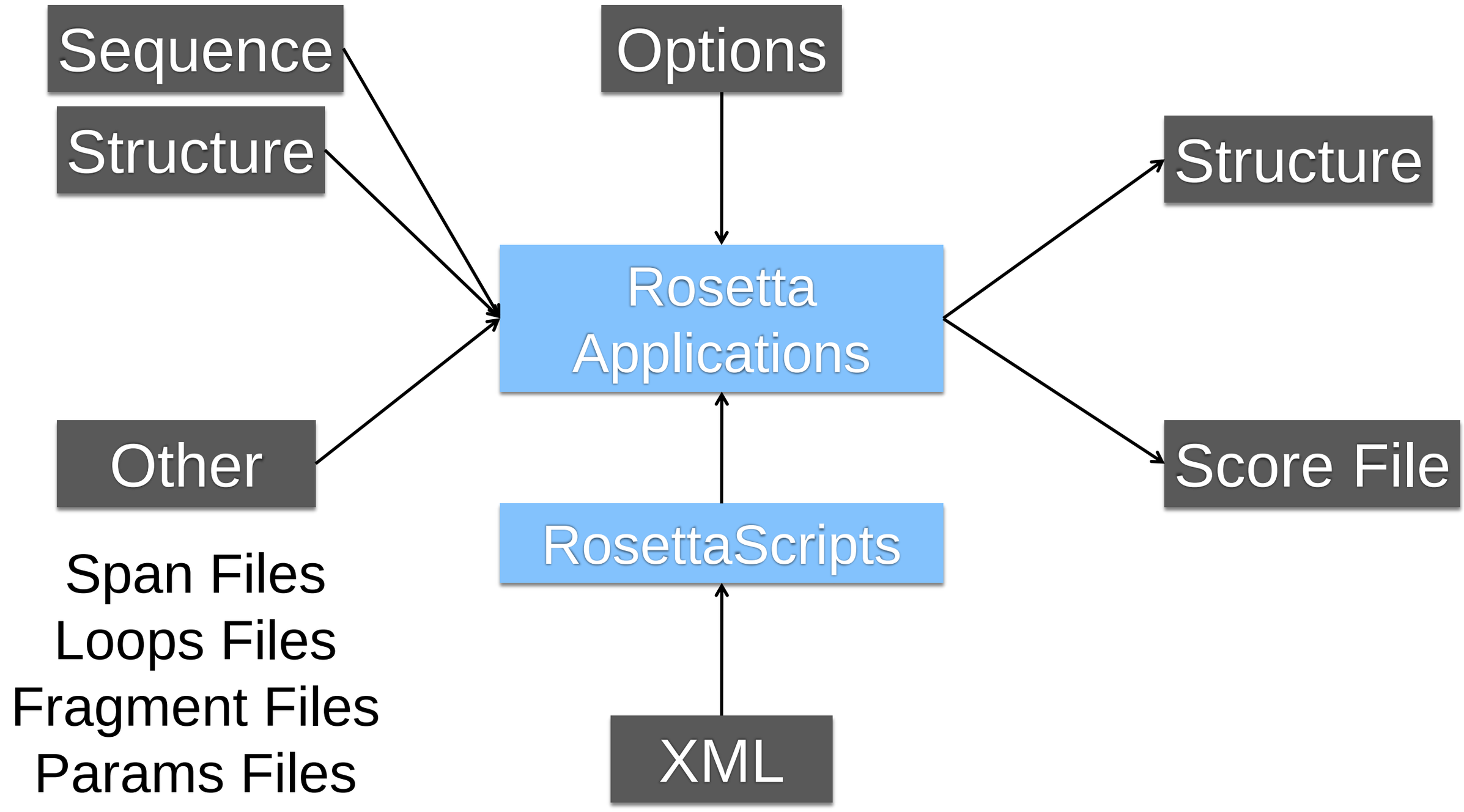
Protein and binary silent files

- Binary silent files are more compact
- Structures with non-ideal bond lengths must use the binary format
- Rosetta makes this decision automatically
- Details: https://www.rosettacommons.org/docs/latest/rosetta_basics/file_types/silent-file
- Extracted to PDBs with extract_pdb application

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
- Res Files: Indicates which residue positions should be designed

What goes in, What comes out



Span Files
Loops Files
Fragment Files
Params Files
Constraint Files

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

```
LYS_193 -3.59411 0.14513 2.02302 0.00582 0 -0.49267 0 -0.92698 0 0 0 -0.53918 0 0 0 0 0 0 0 0 -0.1299 0.54698 0.66528 -0.06428 -0.42 0 -2.78088
HIS_194 -4.55142 0.28903 2.04128 0.00303 0 -0.39232 0 -0.40256 0 0 0 -0.68768 0 0 0 0 0 0 0 0 0.05322 0.21978 0.2978 -0.36527 1.17 0 -2.32511
SER_195 -2.89552 0.35377 1.64498 0.00361 0 -0.03598 0 -0.07645 0 0 0 0 0 0 0 0 0 0 0 0 0.57819 0.00205 0.21641 -0.20511 0 0 -0.41407
GLN_196 -2.51513 0.10163 1.17276 0.00914 0 0.01537 0 -0.05807 0 0 0 0 0 0 0 0 0 0 0 0 -0.13144 0.30977 0.86969 -0.1211 -0.29 0 -0.63737
PHE_197 -1.73186 0.13025 0.4893 0.0245 0 0.05242 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -0.1163 0.16687 0.08488 -0.2005 1.88 0 0.77955
ILE_198 -2.83632 0.36351 0.8091 0.02149 0 0.01948 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -0.19486 0.23841 0.58581 0.05351 -0.73 0 -1.66986
GLY_199 -1.1838 0.09252 0.46713 4e-05 0 -0.06454 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.47472 0.11544 0 -0.16396 -0.55 0 -0.81244
TYR_200 -6.03075 0.59533 2.29882 0.02331 0 -0.35986 0.00108 0 0 0 0 -0.65917 0 0 0 0 0 0 0 0 -0.12857 0.03684 0.23138 -0.24871 0.81 0 -3.43031
PRO_201 -1.91179 0.17002 0.77817 0.00173 0 -0.14682 0.00213 0 0 0 0 0 0 0 0 0 0 0 0 0 -0.16557 0.0231 0.32465 -0.57603 -0.33 0 -1.83041
ILE_202 -2.82048 0.16481 0.38553 0.01518 0 -0.01128 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -0.18177 0.02148 0.13276 -0.60588 -0.73 0 -3.62966
THR_203 -2.75792 0.04526 1.45067 0.00865 0 -0.29027 0 -0.099 0 0 0 0 0 0 0 0 0 0 0 0 0.0033 0.10332 0.02485 -0.03009 -0.57 0 -2.11122
LEU_204 -3.57508 0.34458 0.55075 0.00955 0 -0.05076 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -0.06701 0.00178 0.43736 -0.17841 -0.66 0 -3.18725
PHE_205 -4.22608 0.24299 1.30269 0.02966 0 -0.09395 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -0.18615 0.05662 1.31119 0.08333 1.88 0 0.4003
VAL_206 -0.92973 0.01534 0.6203 0.01359 0 0.05242 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.02686 0.04282 0.06012 0.04544 -0.57 0 -0.62284
GLU:CtermProteinFull_207 -1.65508 0.01501 1.49732 0.00713 0 -0.3596 0 -0.30872 0 0 -0.34864 0 0 0 0 0 0 0 0 0 0 0.91444 0 -0.27 0 -0.50815
LIG_208 -5.68364 0.80302 2.88679 0.0179 0.00024 -0.2396 0 0 0 0 0 -0.84105 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -3.05635
#END_POSE_ENERGIES_TABLE 2JJC.hsp90_99.2JJC.relaxed_ligand_0500.pdb.gz
Transform_accept_ratio 0.734
angle_constraint 0
atom_pair_constraint 0
chainbreak 0
coordinate_constraint 1.39621
dihedral_constraint 1.45342e-07
dslf_ca_dih 0
dslf_cs_ang 0
dslf_ss_dih 0
dslf_ss_dst 0
fa_atr -794.182
fa_dun 95.1801
fa_elec -53.7925
fa_intra_rep 1.7716
fa_intra_rep_xover4 0.000240991
fa_pair -37.6743
fa_rep 73.6057
fa_sol 360.403
```

Rosetta score summary file

```
SEQUENCE:
SCORE: total_score Transform_accept_ratio angle_constraint atom_pair_constraint chainbreak coordinate_constraint
SCORE:   -1011.119                 0.522            0.000                0.000      0.000                 4.837
SCORE:   -1050.267                 0.602            0.000                0.000      0.000                 1.081
SCORE:   -1007.125                 0.562            0.000                0.000      0.000                14.511
SCORE:   -1040.489                 0.546            0.000                0.000      0.000                 1.526
SCORE:   -1017.127                 0.492            0.000                0.000      0.000                 3.638
SCORE:   -1021.812                 0.550            0.000                0.000      0.000                 6.087
SCORE:   -1012.409                 0.632            0.000                0.000      0.000                 4.739
SCORE:   -1044.905                 0.610            0.000                0.000      0.000                 1.090
SCORE:   -1049.928                 0.592            0.000                0.000      0.000                 1.620
SCORE:   -1019.955                 0.580            0.000                0.000      0.000                 1.369
SCORE:   -1046.579                 0.572            0.000                0.000      0.000                 1.569
SCORE:   -1044.294                 0.508            0.000                0.000      0.000                 0.555
SCORE:   -1040.868                 0.558            0.000                0.000      0.000                 1.074
SCORE:   -1029.726                 0.574            0.000                0.000      0.000                 4.628
SCORE:   -1032.794                 0.506            0.000                0.000      0.000                 1.075
SCORE:   -1050.539                 0.594            0.000                0.000      0.000                 0.800
SCORE:   -1051.451                 0.552            0.000                0.000      0.000                 0.898
SCORE:   -1026.903                 0.628            0.000                0.000      0.000                 4.695
SCORE:   -1014.123                 0.524            0.000                0.000      0.000                 3.922
SCORE:   -1039.412                 0.626            0.000                0.000      0.000                 0.908
SCORE:   -1032.375                 0.614            0.000                0.000      0.000                 0.854
SCORE:   -1042.355                 0.588            0.000                0.000      0.000                 0.739
SCORE:   -1047.374                 0.662            0.000                0.000      0.000                 0.929
SCORE:   -1011.842                 0.524            0.000                0.000      0.000                 5.148
SCORE:   -1053.325                 0.602            0.000                0.000      0.000                 1.142
SCORE:    -989.985                 0.556            0.000                0.000      0.000                14.164
```