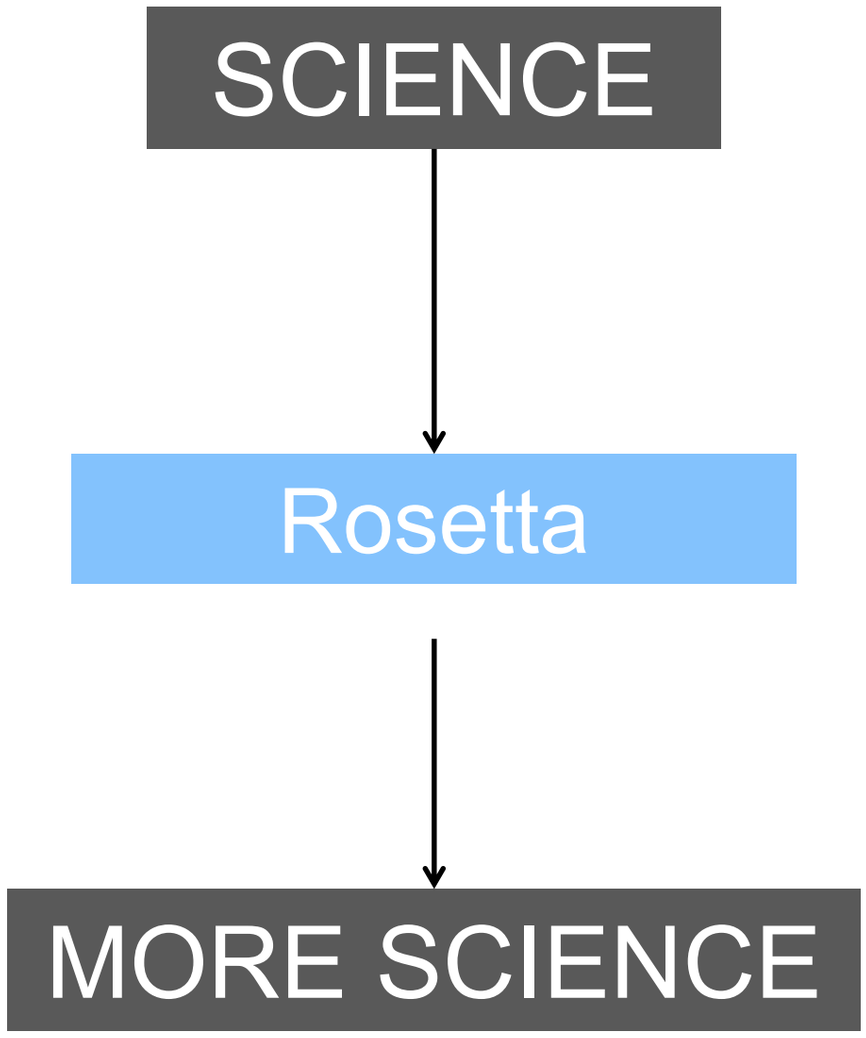
Rosetta Input/Output

Rocco Moretti
rocco.moretti@vanderbilt.edu

What goes in, What comes out



How do we get Rosetta?

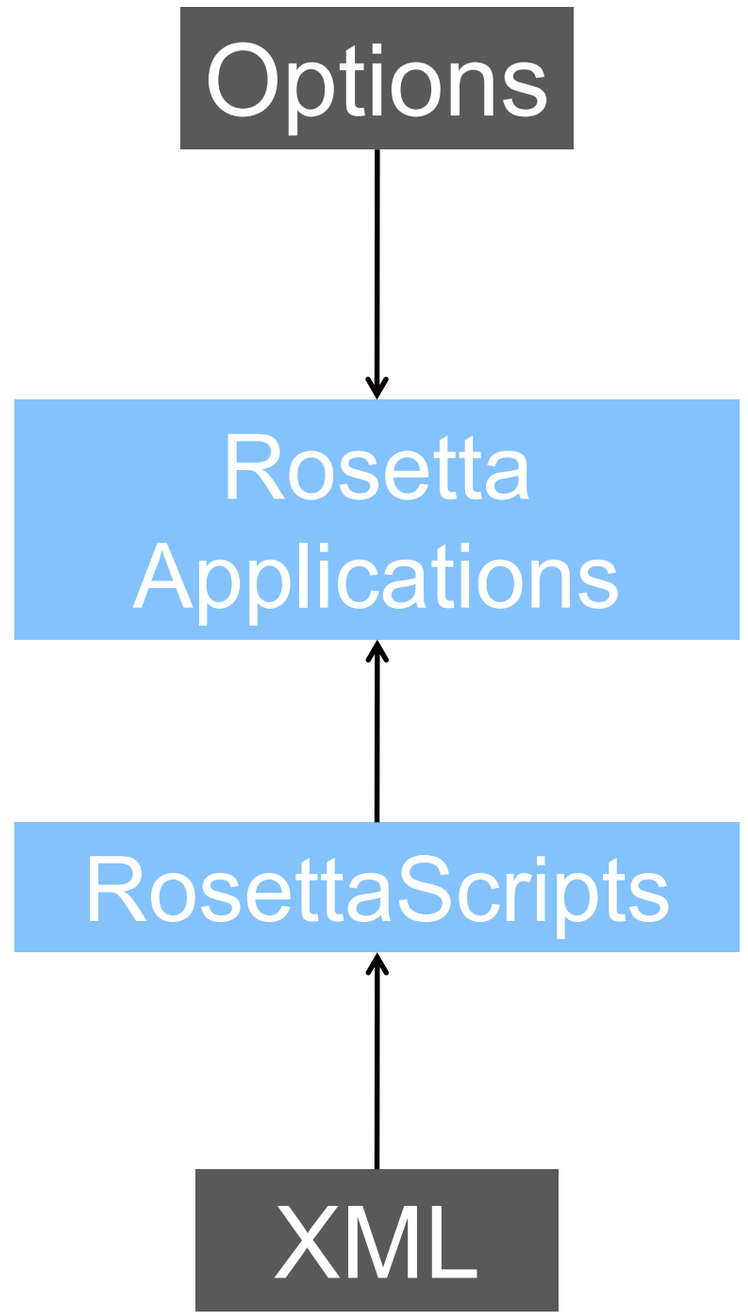
- https://www.rosettacommons.org/software/license-and-download
- Weekly Releases: (e.g. “2016.37”)
  - Latest version of the code, released roughly every week
  - Every revision passes scientific performance tests
- Numbered Releases (e.g. “3.7”)
  - A weekly release that’s relabeled, released roughly every 6 months
- All tutorials use version 3.7 (September 2016)

- Documentation & Demos:
  - https://www.rosettacommons.org/docs/latest/Home
  - https://www.rosettacommons.org/demos/latest/Home

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



Types of options

Full List of Options can be found at:

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

3500+ options

- Input
  - Location of Specific Files (*e.g.* “-in:file:fasta”)
  - Structure (*e.g.* “-in:file:s”)
- Output
  - Filename and Filetypes (*e.g.* “-out:path:pdb”)
  - Number of Structures (*e.g.* “-out:nstruct”)
- Rotamer Library Options (*e.g.* “-packing:ex1”)
- 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:

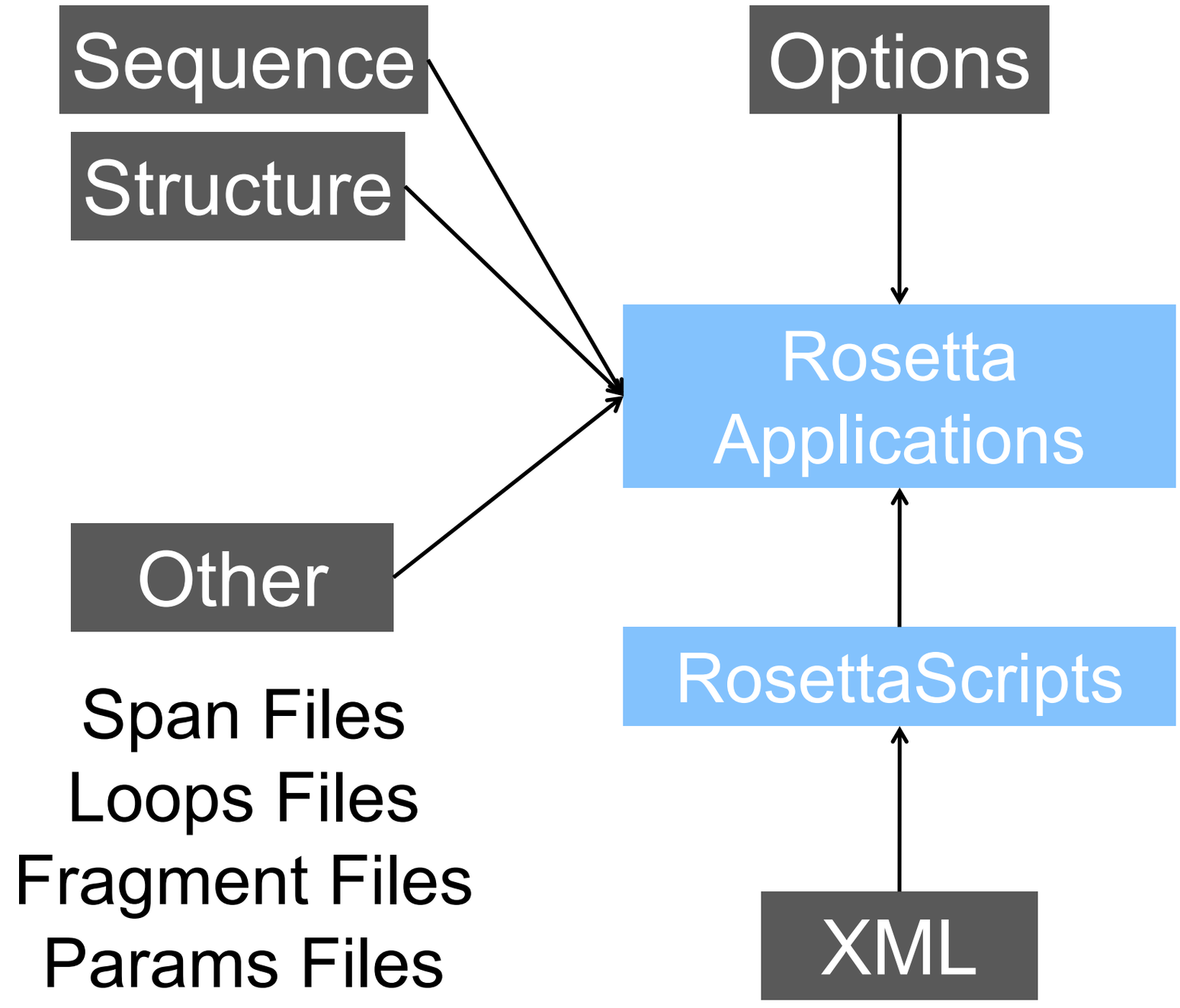
-in

-file

-s test.pdb

Don’t mix tabs and spaces.

What goes in, What comes out



Span Files
Loops Files
Fragment Files
Params Files
Constraint Files

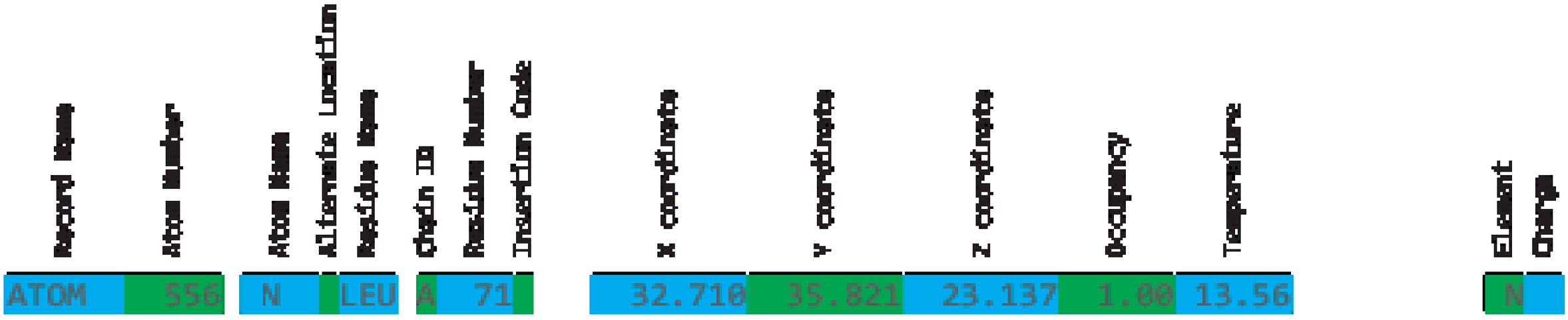
Rosetta has two main formats for inputting structures

- PDB Files
  - International standard
  - Readable by PyMOL, MOE, etc.
- Silent Files
  - Rosetta specific
  - Compact, fast to output
  - Doesn’t fill your directories

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?



- 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 (mostly)
- 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

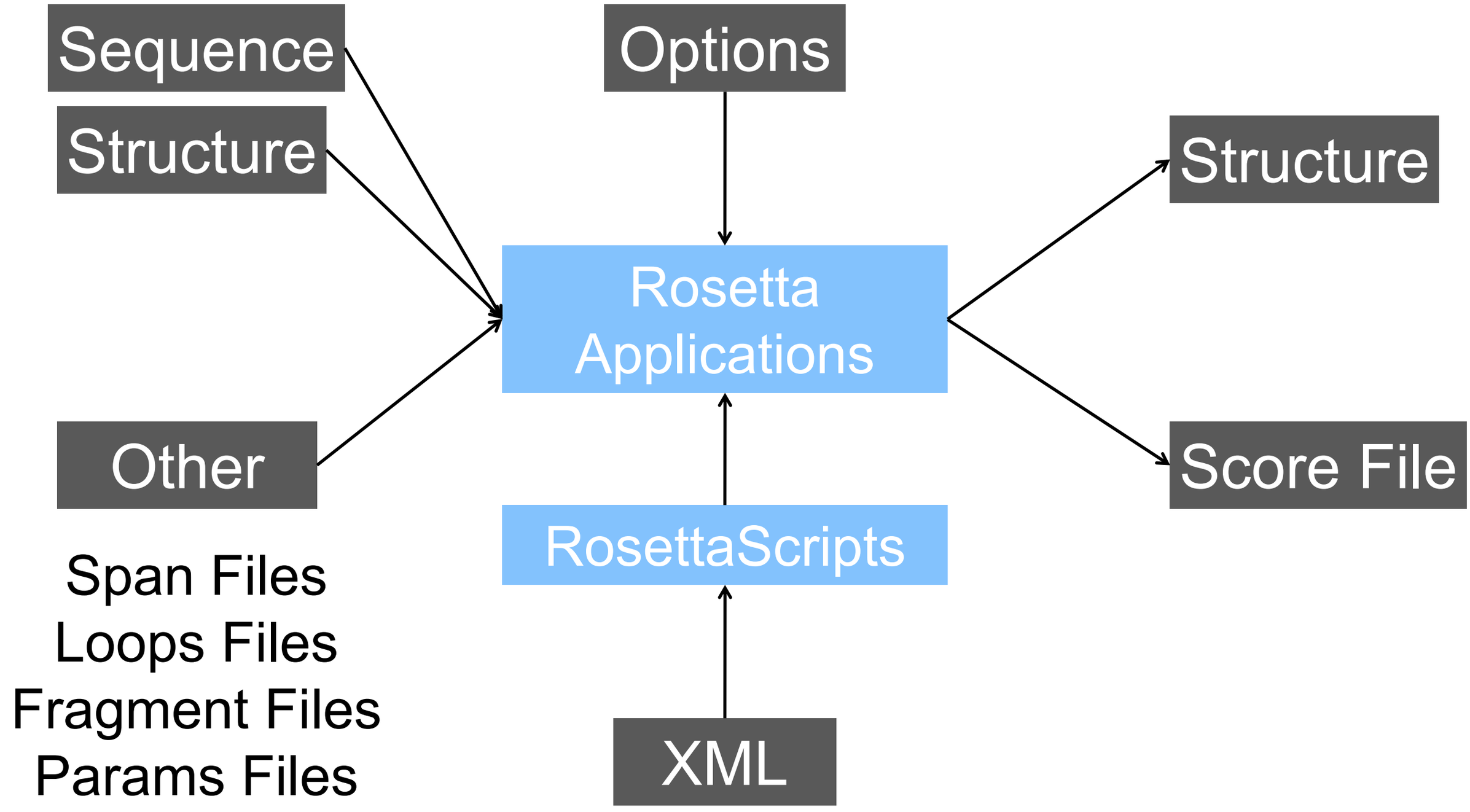
Protein and binary silent files

- 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



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

```
ATOM   3378  HB  THR L 227     -36.166  22.580  28.848  1.00  0.00           H
ATOM   3379  HG1 THR L 227     -34.994  19.987  29.136  1.00  0.00           H
ATOM   3380 1HG2 THR L 227     -34.138  22.579  30.246  1.00  0.00           H
ATOM   3381 2HG2 THR L 227     -35.593  22.831  31.238  1.00  0.00           H
ATOM   3382 3HG2 THR L 227     -34.799  21.240  31.213  1.00  0.00           H
TER
# All scores below are weighted scores, not raw scores.
#BEGIN_POSE_ENERGIES_TABLE 3gbn_Ab_0005.pdb
label fa_atr fa_rep fa_sol fa_intra_rep fa_elec pro_close hbond_sr_bb hbond_lr_bb hbond_bb_sc hbond_sc dslf_fa13 rama omega fa_dun p_aa_pp yhh_planarity ref total
weights 1 0.55 0.9375 0.005 0.875 1.25 1.17 1.17 1.17 1.1 1.25 0.25 0.625 0.7 0.4 0.625 1 NA
pose -994.338 137.719 561.027 2.15688 -112.612 19.7634 -12.2069 -79.5391 -24.1449 -23.4441 -1.15166 -7.47192 71.8572 276.633 -29.8673 0.09431 13.9828 -201.541
GLU:NtermProteinFull_1 -1.58225 0.53996 1.45283 0.00353 0.06909 0 0 0 0 0 0 0 0.01109 6.53174 0 0 -1.96094 5.06505
VAL_2 -3.34255 0.45648 1.46378 0.01322 -0.08167 0 0 0 0 0 0 -0.16095 0.87346 0.30715 0.39992 0 0.97964 0.90848
GLN_3 -2.79445 0.10936 1.74929 0.00451 -0.52743 0 0 0 -0.35772 0 0 -0.09682 0.35321 2.59775 0.02034 0 -1.51717 -0.45911
LEU_4 -5.13483 0.73792 1.6574 0.00685 -0.16379 0 0 0 0 0 0 0.06265 0.2281 2.29891 -0.1217 0 0.76113 0.33264
VAL_5 -2.72905 0.12167 1.72074 0.00789 -0.45069 0 0 0 0 0 0 -0.27382 0.01969 0.02557 -0.49649 0 0.97964 -1.07485

...

TYR_205 -9.68927 0.6599 4.04398 0.02918 -1.11866 0 0 0 -0.69927 -0.4197 0 -0.03371 0.01485 1.86931 -0.07409 0.02555 0.1625 -5.22942
TYR_206 -8.27907 0.53027 3.95082 0.03533 -0.75969 0 0 0 0 0 0 -0.10219 0.77773 2.1235 -0.06518 0.00994 0.1625 -1.61604
CYS:disulfide_207 -4.8779 0.37363 2.6835 0.00127 -0.70663 0 0 0 -0.56501 0 -0.29622 -0.22449 0.0286 1.4135 -0.08388 0 0.44379 -1.80984
ALA_208 -4.48831 0.2957 1.92707 0.00107 -0.48653 0 0 0 0 0 0 0.05006 0.28785 0 -0.24738 0 0.77374 -1.88673
THR_209 -5.73046 0.81574 3.59921 0.00829 -1.2848 0 0 0 -0.46497 0 0 0.23162 0.01866 0.16545 -0.28654 0 0.20134 -2.72646

...

THR_224 -6.10085 0.26779 3.79175 0.00865 -0.91162 0 0 0 -0.45219 -0.61107 0 0.02255 0.07826 0.06259 -0.14086 0 0.20134 -3.78367
LYS_225 -3.90687 0.70976 2.6297 0.00615 -0.81559 0 0 0 0 0 0 -0.04424 0.00136 0.98794 0.09041 0 -0.35857 -0.69994
LEU_226 -5.97934 0.46605 2.34857 0.02903 -0.32933 0 0 0 0 0 0 -0.08426 0.1675 2.62906 -0.15244 0 0.76113 -0.14405
THR:CtermProteinFull_227 -3.2336 0.44967 3.66532 0.00842 -0.88667 0 0 0 -0.13737 0 0 0 0 0.062 0 0 0.20134 0.1291
#END_POSE_ENERGIES_TABLE 3gbn_Ab_0005.pdb
bref_irms 0.0777809
cen_irms 0.0777809
final_chainbreak 0.143091
final_looprelax_score -201.541
irms 0.0630764
```

Rosetta score summary file

SEQUENCE:

SCORE:	score	fa_atr	fa_rep	fa_sol	fa_intra_rep	fa_elec	...	omega	fa_dun	p_aa_pp	ref	description
SCORE:	-1217.209	-2778.696	266.309	1545.149	5.900	-301.320	...	63.032	684.989	-109.110	-32.534	3gbm_HA_3gbn_Ab_full_0011
SCORE:	-1217.028	-2792.422	263.906	1549.738	5.867	-295.799	...	66.036	682.694	-108.402	-32.534	3gbm_HA_3gbn_Ab_full_0012
SCORE:	-1204.280	-2760.354	259.175	1534.072	5.913	-293.050	...	65.391	674.840	-108.393	-32.534	3gbm_HA_3gbn_Ab_full_0013
SCORE:	-1207.127	-2768.191	260.443	1541.857	5.881	-301.847	...	67.951	686.381	-110.919	-32.534	3gbm_HA_3gbn_Ab_full_0014
SCORE:	-1208.390	-2769.872	262.398	1539.668	5.879	-297.571	...	64.073	681.731	-109.633	-32.534	3gbm_HA_3gbn_Ab_full_0015
SCORE:	-1214.352	-2763.475	252.977	1531.781	5.920	-295.620	...	65.883	676.240	-109.111	-32.534	3gbm_HA_3gbn_Ab_full_0016
SCORE:	-1219.655	-2789.629	264.179	1549.529	5.885	-298.401	...	63.134	682.856	-108.031	-32.534	3gbm_HA_3gbn_Ab_full_0017
SCORE:	-1186.718	-2801.397	279.206	1561.481	5.934	-295.056	...	67.359	681.528	-105.381	-32.534	3gbm_HA_3gbn_Ab_full_0018
SCORE:	-1215.503	-2758.727	255.909	1532.410	5.903	-299.673	...	66.402	678.984	-111.856	-32.534	3gbm_HA_3gbn_Ab_full_0019
SCORE:	-1217.883	-2782.742	259.590	1550.245	5.889	-299.331	...	63.274	680.028	-108.206	-32.534	3gbm_HA_3gbn_Ab_full_0020
SCORE:	-1198.782	-2777.296	263.448	1541.979	5.968	-297.280	...	68.332	687.685	-109.097	-32.534	3gbm_HA_3gbn_Ab_full_0021
SCORE:	-1209.557	-2779.964	261.169	1556.555	5.846	-302.588	...	63.881	681.855	-109.275	-32.534	3gbm_HA_3gbn_Ab_full_0023
SCORE:	-1224.909	-2776.609	259.690	1543.340	5.874	-298.333	...	63.170	675.545	-110.378	-32.534	3gbm_HA_3gbn_Ab_full_0024
SCORE:	-1202.468	-2771.633	263.260	1547.001	5.909	-298.973	...	65.652	683.601	-109.045	-32.534	3gbm_HA_3gbn_Ab_full_0026
SCORE:	-1216.261	-2775.841	257.923	1542.257	5.900	-299.608	...	66.459	682.661	-109.199	-32.534	3gbm_HA_3gbn_Ab_full_0027
SCORE:	-1209.306	-2779.799	262.576	1547.355	5.901	-297.681	...	65.276	682.562	-109.749	-32.534	3gbm_HA_3gbn_Ab_full_0028
SCORE:	-1221.261	-2783.969	262.440	1549.503	5.855	-302.005	...	64.895	684.862	-108.807	-32.534	3gbm_HA_3gbn_Ab_full_0029
SCORE:	-1202.380	-2774.027	259.636	1537.434	5.915	-292.278	...	66.615	682.911	-108.091	-32.534	3gbm_HA_3gbn_Ab_full_0030
SCORE:	-1201.713	-2771.522	259.725	1540.804	5.904	-295.524	...	67.195	681.868	-108.624	-32.534	3gbm_HA_3gbn_Ab_full_0031
SCORE:	-1195.317	-2786.103	273.621	1547.012	5.911	-292.683	...	68.143	680.827	-107.152	-32.534	3gbm_HA_3gbn_Ab_full_0032
SCORE:	-1209.027	-2766.003	256.911	1536.377	5.926	-298.778	...	64.564	685.533	-109.246	-32.534	3gbm_HA_3gbn_Ab_full_0033
SCORE:	-1206.444	-2780.299	265.236	1543.936	5.886	-297.829	...	64.436	682.950	-107.925	-32.534	3gbm_HA_3gbn_Ab_full_0035
SCORE:	-1209.361	-2774.235	259.464	1541.286	5.888	-296.928	...	67.864	679.797	-109.848	-32.534	3gbm_HA_3gbn_Ab_full_0036
SCORE:	-1209.428	-2774.683	264.095	1545.204	5.899	-301.183	...	65.623	683.858	-107.948	-32.534	3gbm_HA_3gbn_Ab_full_0037
SCORE:	-1209.120	-2784.944	266.037	1543.980	5.879	-295.598	...	65.160	685.062	-109.819	-32.534	3gbm_HA_3gbn_Ab_full_0038
SCORE:	-1204.681	-2773.288	264.079	1538.727	5.878	-297.499	...	65.903	683.761	-108.878	-32.534	3gbm_HA_3gbn_Ab_full_0039
SCORE:	-1210.196	-2771.337	263.743	1538.364	5.901	-299.579	...	65.708	683.140	-109.715	-32.534	3gbm_HA_3gbn_Ab_full_0041
SCORE:	-1211.073	-2776.715	259.882	1547.542	5.863	-303.553	...	67.405	685.284	-108.160	-32.534	3gbm_HA_3gbn_Ab_full_0042
SCORE:	-1214.648	-2780.396	261.300	1544.135	5.889	-300.337	...	66.263	685.718	-109.501	-32.534	3gbm_HA_3gbn_Ab_full_0043
SCORE:	-1216.096	-2779.447	260.645	1545.806	5.872	-299.687	...	66.893	679.556	-108.300	-32.534	3gbm_HA_3gbn_Ab_full_0044
SCORE:	-1227.359	-2790.062	261.244	1549.745	5.894	-299.397	...	63.919	682.009	-109.613	-32.534	3gbm_HA_3gbn_Ab_full_0045
SCORE:	-1224.560	-2787.163	263.018	1541.368	5.879	-297.468	...	65.586	683.252	-109.746	-32.534	3gbm_HA_3gbn_Ab_full_0046
SCORE:	-1207.967	-2773.475	261.557	1546.365	5.920	-301.947	...	66.652	682.649	-109.787	-32.534	3gbm_HA_3gbn_Ab_full_0047
SCORE:	-1210.521	-2773.902	257.256	1543.895	5.887	-296.454	...	66.742	675.904	-107.866	-32.534	3gbm_HA_3gbn_Ab_full_0048
SCORE:	-1203.967	-2767.613	263.371	1541.055	5.876	-297.629	...	64.674	681.230	-110.923	-32.534	3gbm_HA_3gbn_Ab_full_0049
SCORE:	-1206.728	-2759.167	259.854	1529.948	5.906	-297.720	...	65.498	683.356	-110.453	-32.534	3gbm_HA_3gbn_Ab_full_0050

Tracer Output

```
core.init: Rosetta version exported  from http://www.rosettacommons.org
core.init: command: /dors/meilerlab/apps/rosetta/rosetta_2016.08.58479/main/source/bin/rosetta_scripts.default.linuxgccrelease @docking.options -parser:protocol docking
core.init: 'RNG device' seed mode, using '/dev/urandom', seed=1059677151 seed_offset=0 real_seed=1059677151
core.init.random: RandomGenerator:init: Normal mode, seed=1059677151 RG_type=mt19937
core.init: Resolved executable path: /dors/meilerlab/apps/rosetta/rosetta_2016.08.58479/main/source/build/src/release/linux/2.6/64/x86/gcc/5.2/default/rosetta_scripts.d
core.init: Looking for database based on location of executable: /dors/meilerlab/apps/rosetta/rosetta_2016.08.58479/main/database/
protocols.jd2.PDBJobInputter: Instantiate PDBJobInputter
protocols.jd2.PDBJobInputter: PDBJobInputter::fill_jobs
protocols.jd2.PDBJobInputter: pushed 3gbm_HA_3gbn_Ab.pdb nstruct indices 1 - 50
protocols.evaluation.ChiWellRmsdEvaluatorCreator: Evaluation Creator active ...
protocols.jd2.JobDistributor: Parser is present.  Input mover will be overwritten with whatever the parser creates.
protocols.jd2.PDBJobInputter: PDBJobInputter::pose_from_job
protocols.jd2.PDBJobInputter: filling pose from PDB 3gbm_HA_3gbn_Ab.pdb
core.chemical.ResidueTypeSet: Finished initializing fa_standard residue type set.  Created 384 residue types
core.chemical.ResidueTypeSet: Total time to initialize 0.43 seconds.
core.conformation.Conformation: Found disulfide between residues 7 461
...
protocols.rosetta_scripts.ParsedProtocol.REPORT: ============End report for ==================
protocols.rosetta_scripts.ParsedProtocol.REPORT: ============Begin report for ==================
protocols.rosetta_scripts.ParsedProtocol.REPORT: ============End report for ==================
protocols.jd2.JobDistributor: 3gbm_HA_3gbn_Ab_full_0050 reported success in 381 seconds
protocols.jd2.JobDistributor: no more batches to process...
protocols.jd2.JobDistributor: 50 jobs considered, 50 jobs attempted in 16297 seconds
~
~
~
```

Options to control tracer output:

- Silence certain tracers:

  `-mute core.chemical.ResidueTypeSet`

- Change verbosity level (Error/Warning/Info/Debug/Trace)

  `-out:levels all:Warning core.init:Info`