Creating custom protocols with ROSETTA

Presented by: Hope Woods

hope.woods@vanderbilt.edu
Custom Protocols

- Rosetta applications cover most common tasks

- Sometimes you want to do something different
  - Modify an existing protocol
  - Combine two protocols
  - Make an entire novel protocol

- Or if you just want to run multiple protocols from the same application
How to Make Custom Protocols

• C++ - Directly modify the Rosetta source code

• PyRosetta – Python bindings for directly interacting with Rosetta functions (http://www.pyrosetta.org/)

• RosettaScripts – XML based interface for creating protocols
• XML “eXtensible Markup Language”
• Consists of large level tags and sub-tags
• Everything not in brackets < > is a comment
• Read from the bottom up

Tip: Run rosetta_scripts without options to get template
Breaking down a tag

<MOVERS>
  <PackRotamersMover name="repack1" scorefxn="ref2015" />
</MOVERS>

- Name of mover used
- Name assigned to specific version (can be referenced elsewhere in XML)
- Custom settings

Most tags have required settings or default values, always check documentation
Movers

- Movers are the basic building blocks of a RosettaScripts protocol
  - Most modify the pose
    - Some compute metrics instead

```xml
<MOVERS>
  <PackRotamersMover name="repack1" scorefxn="ref2015" taskoperations="ifcl,rtrp" />
  <InterfaceAnalyzerMover name="iface" scorefxn="ref2015" fixedchains="A,B" />
</MOVERS>
```
Movers

• Movers are the basic building blocks of a RosettaScripts protocol

• Most modify the pose
  • Some compute metrics instead

• A single mover can be used more than once
Protocols

<PROTOCOLS>
    <Add mover="Repack1"/>
    <Add mover="Repack2" filter="avg_deg"/>
    <Add mover="iface"/>
</PROTOCOLS>

• Movers are executed in the order specified in PROTOCOLS

• Movers can be combined with filters

• Movers can be used more than once in a protocol
Score Functions

Different parts of a protocol can use different score functions.

Standard score functions can be modified.
Filters

<FILTERS>
  <ScoreType name="score_type_filter" scorefxn="ref2015" score_type="total_score"
  Threshold="-500" />
  <AverageDegree name="avg_deg" threshold="8" distance_threshold="10"
  task_operations="rtiv" />
</FILTERS>

- Can pass/fail an output structure
  - Stop a run earlier if the output will be bad.
- Also can be used to compute protein metrics, but Simple Metrics is recommended for this
Filters

• Can pass/fail an output structure
  • Stop a run earlier if the output will be bad.

• Also can be used to compute protein metrics, but Simple Metrics is recommended for this

```
<FILTERS>
  <ScoreType name="score_type_filter" scorefxn="ref2015" score_type="total_score"
  Threshold="-500" />
  <AverageDegree name="avg_deg" threshold="8" distance_threshold="10"
  task_operations="rtiv" />
</FILTERS>

<MOVERS>
  ...

  <IteratedConvergence name="iter_conv" mover="repack" filter="score_type_filter"
  delta="0.1" cycles="10" maxcycles="100" />
</MOVERS>
```
Simple Metrics

- Used to compute protein metrics
- Can be ran at different points during a protocol and output to a score file
- Use in combination with RunSimpleMetrics mover

```
<SIMPLE_METRICS>
  <RMSDMetric name="rmsd" rmsd_type="rmsd_protein_bb_heavy" residue_selector="L1" use_native="1" />
  <TotalEnergyMetric name="total_energy" residue_selector="L1"/>
</SIMPLE_METRICS>
```
Simple Metrics

- Used to compute protein metrics
- Can be ran at different points during a protocol and output to a score file
- Use in combination with RunSimpleMetrics mover

```xml
<SIMPLE_METRICS>
  <RMSDMetric name="rmsd" rmsd_type="rmsd_protein_bb_heavy" residue_selector="L1" use_native="1" />
  <TotalEnergyMetric name="total_energy" residue_selector="L1" />
</SIMPLE_METRICS>

<MOVERS>
  ...<RunSimpleMetrics name="run_metrics1" metrics="total_energy" prefix="m1_" />
</MOVERS>
```
Simple Metrics

```xml
<SIMPLE_METRICS>
  <RMSDMetric name="rmsd" rmsd_type="rmsd_protein_bb_heavy" residue_selector="L1" use_native="1"/>
  <TotalEnergyMetric name="total_energy" residue_selector="L1"/>
</SIMPLE_METRICS>

<MOVERS>
  ...
  <RunSimpleMetrics name="run_metrics1" metrics="total_energy" prefix="m1_"/>
</MOVERS>

<PROTOCOLS>
  ...
  <Add mover="Repack1"/>
  <Add mover="run_metrics1"/>
  ...
</PROTOCOLS>
```
Residue Selectors

- Selects a subset of the system for Rosetta to operate on
- There are overlaps with other XML parts (example: a mover may define residues in its own way)
Task Operations

<TASKOPERATIONS>
    <ReadResfile name="rrf" filename="resfile" />
    <RestrictToRepacking name="rtrp" />
    <RestrictResidueToRepacking name="restrict_Y100" resnum="100" />
</TASKOPERATIONS>

...  

    <PackRotamersMover name="repack1" taskoperations="rtrp" />

• Select residues
• Specify how to allow side chain movement
• Specify which residues to allow to design
Output

<OUTPUT scorefxn="ref2015" />

- Specifies the score function used for the final output model and in the scorefile

- If you use multiple score functions in a protocol or use a non-default score function – make sure to flag this
Running Rosetta Scripts

rosetta_scripts.linuxgccrelease -parser:protocol protocol.xml

The application

The actual protocol

Runs whatever procedure is dictated in the XML file

The file that describes your experimental steps

Not seen: @options file and command line options
Useful Features

- Rewrite old Rosetta XML scripts
  - tools/xsd_xrw/rewrite_rossetta_scripts.py

- Validate your XML scripts
  - [https://www.rosettacommons.org/docs/latest/application_documentation/rosetta_scripts/validate_rosetta_script](https://www.rosettacommons.org/docs/latest/application_documentation/rosetta_scripts/validate_rosetta_script)

- Variable substitutions
  - `-parser:script_vars repeat=5 cutoff=10.0` on command line changes every `%%%repeat%%%` to 5 and `%%%cutoff%%%` to 10.0 in XML
Documentation

RosettaScripts documentation
https://www.rosettacommons.org/docs/latest/scripting_documentation/RosettaScripts/RosettaScripts

Possible Movers**
https://www.rosettacommons.org/docs/latest/scripting_documentation/RosettaScripts/Movers/Movers-Movers-RosettaScripts

Original reference

**Best place to start