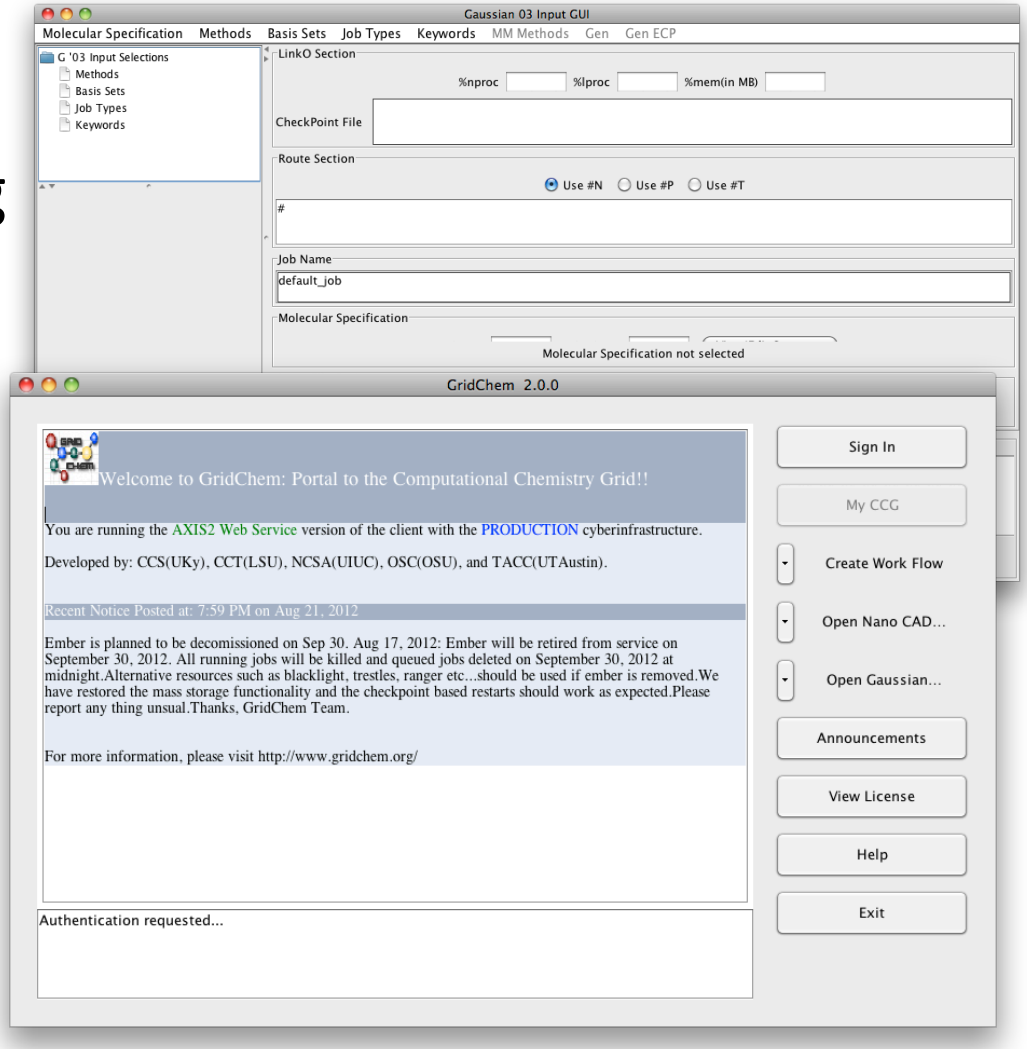
GridChem and ParamChem: Science Gateways for Computational Chemistry (and More)

Marlon Pierce, Suresh Marru
Indiana University
Sudhakar Pamidighantam
NCSA

Computational Chemistry Grid ("GridChem")

- Science Gateway that provides access to high performance computing resources for chemistry and engineering.
  - Java desktop client
  - Service oriented middleware
- "Deep" usage of XSEDE and campus resources
- www.gridchem.org



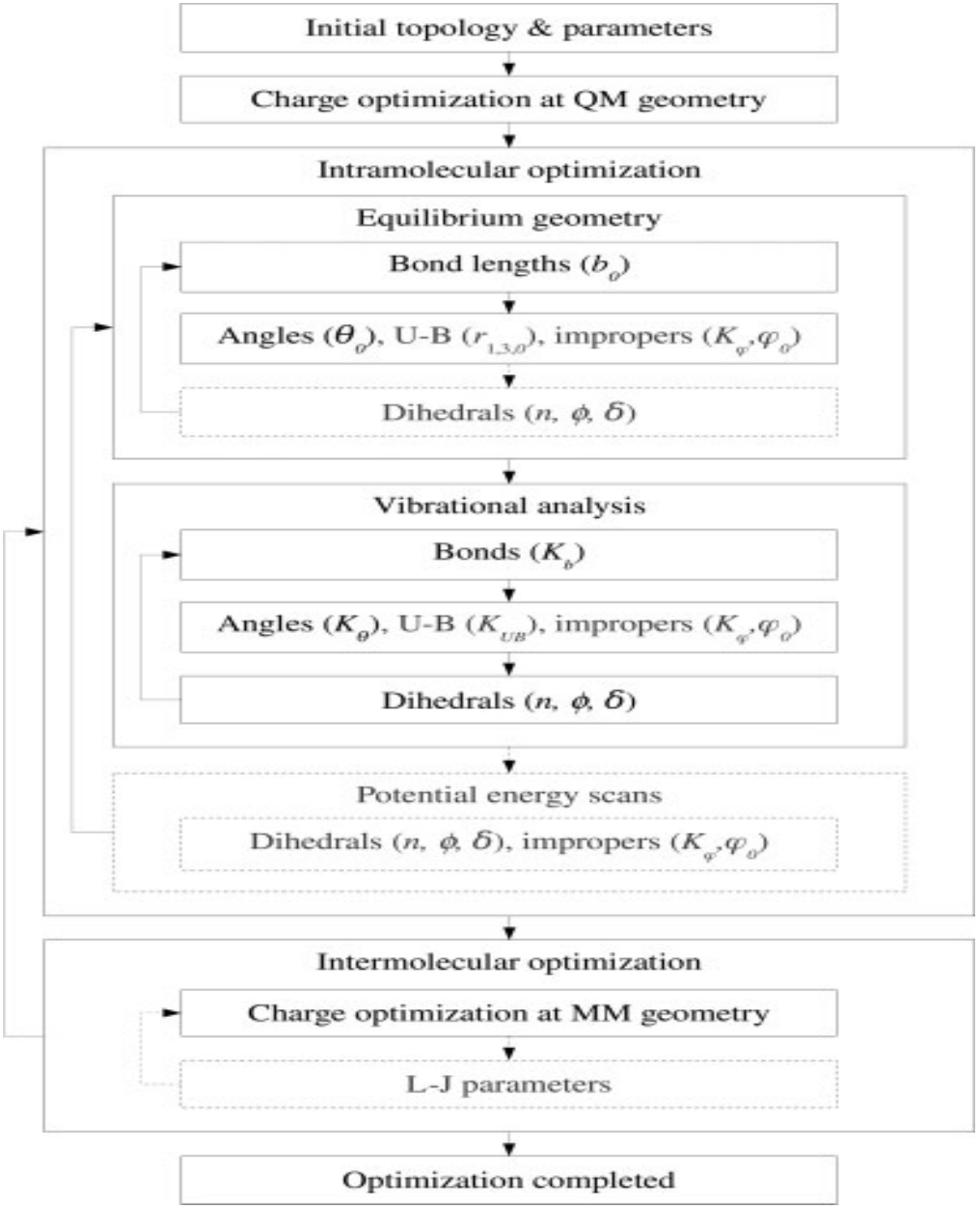
GridChem Stats

Attribute	Value
User Community	> 400 registered users
Available applications	Gaussian, GAMESS, NWChem, LAMMPS, Amber, GROMACS, NAMD, Molpro, DMol3, DDSCAT, CHARMM, Abaqus, Fluent
Available Resources	Ember (NCSA), Trestles (SDSC), Blacklight (PSC), Gordon (SDSC), and Forge (NCSA)
Usage	> 7000 jobs consuming > 194,000 SUs since 31 Aug 2010
Scientific Papers	> 50 publications, 4 Theses
Scientific results managed	> 40,000 experiments stored in the GridChem DB

ParamChem
www.paramchem.org

- ParamChem builds on GridChem/CCG infrastructure to support investigation in the ***molecular force field parameterization problem***.
- This is a multi-step process (workflow)
  - Requires some interactivity, human review and decisions at intermediate steps.
- ParamChem uses Apache Airavata as the workflow tool.

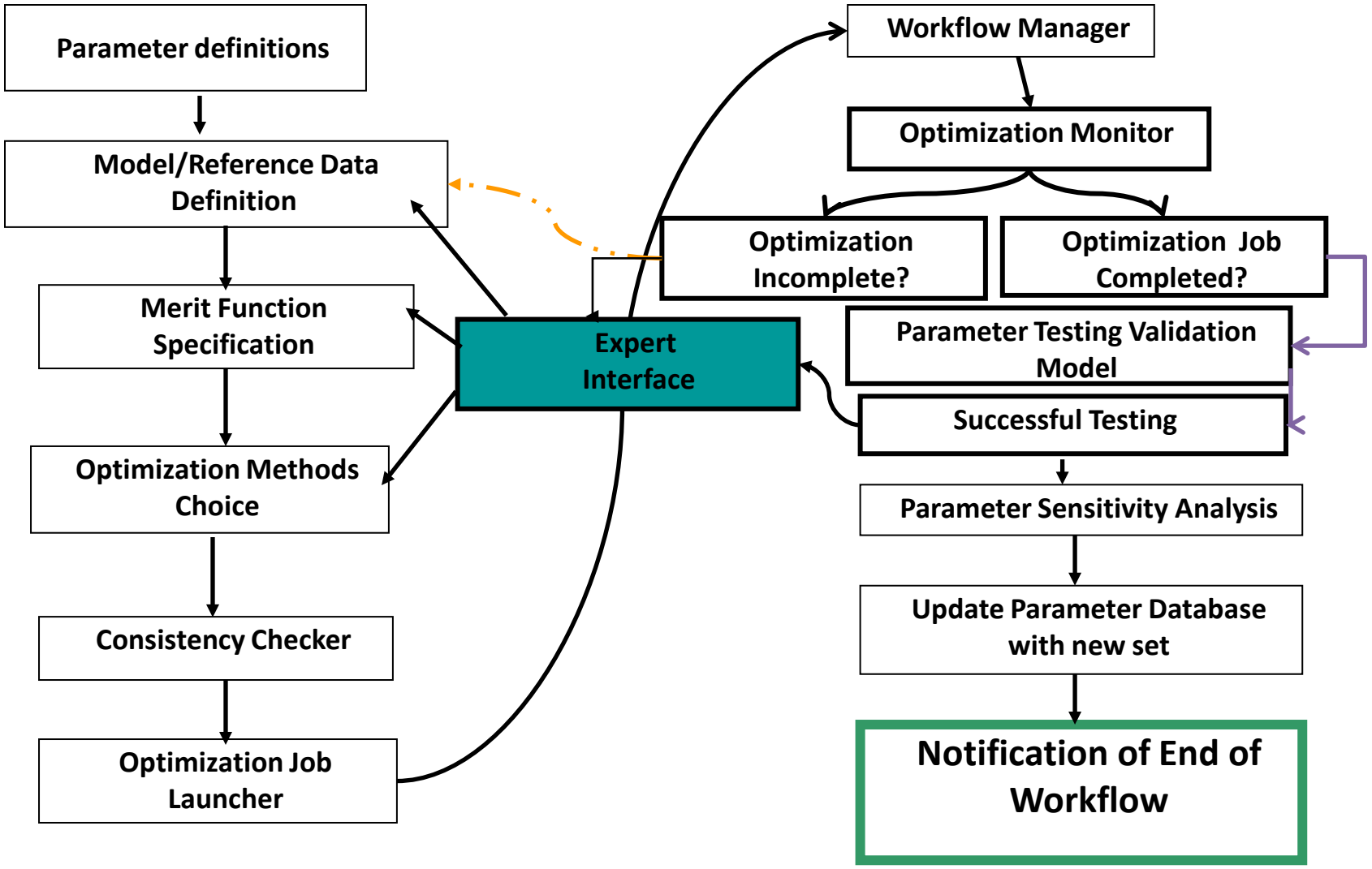
Parameterization Process



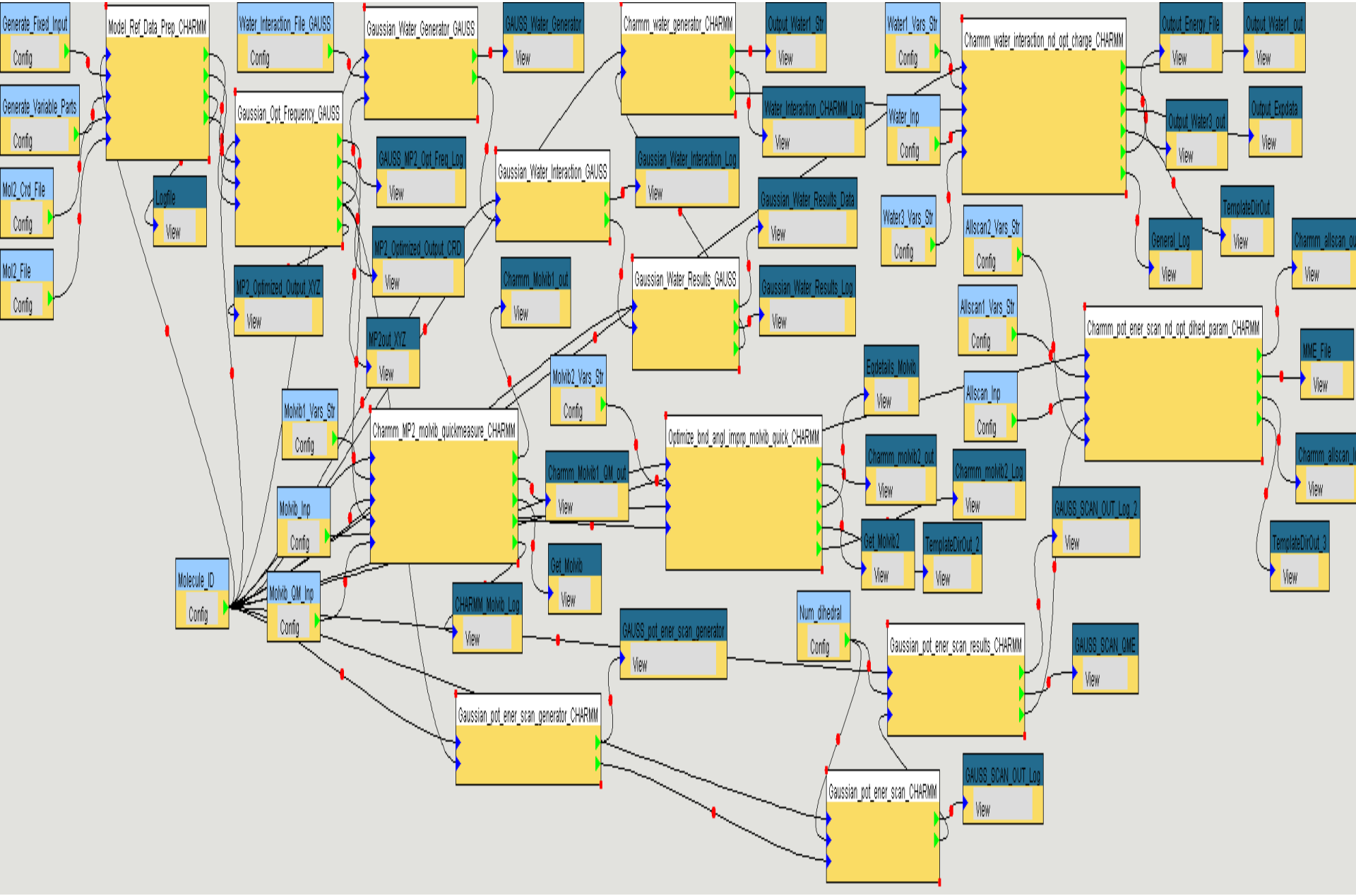
Vanommeslaeghe et al.
J. Comp.Chem 2010, **31**, 671-690

Molecular Force Field Cyberenvironments

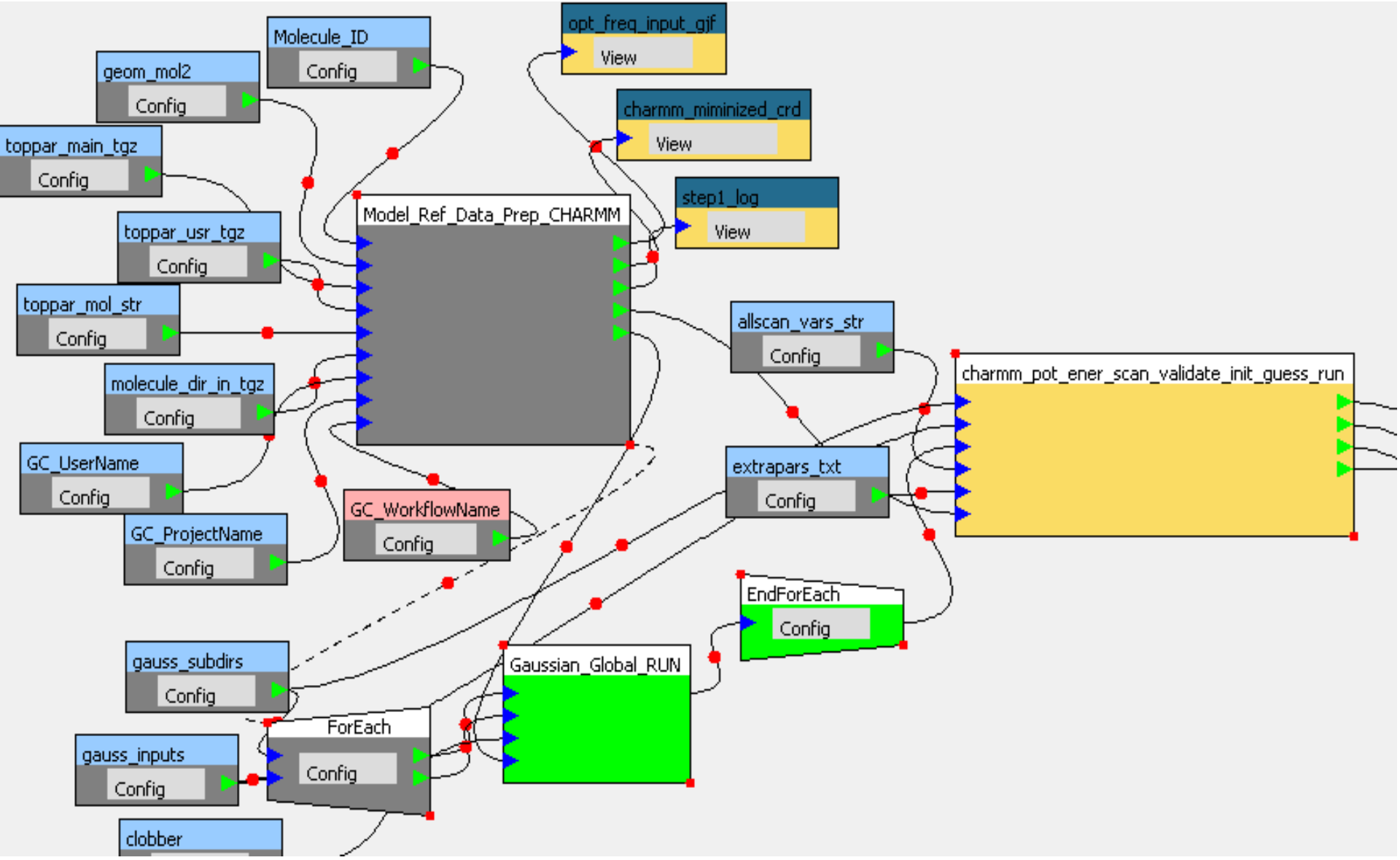
Parameter Initialization and optimization workflow



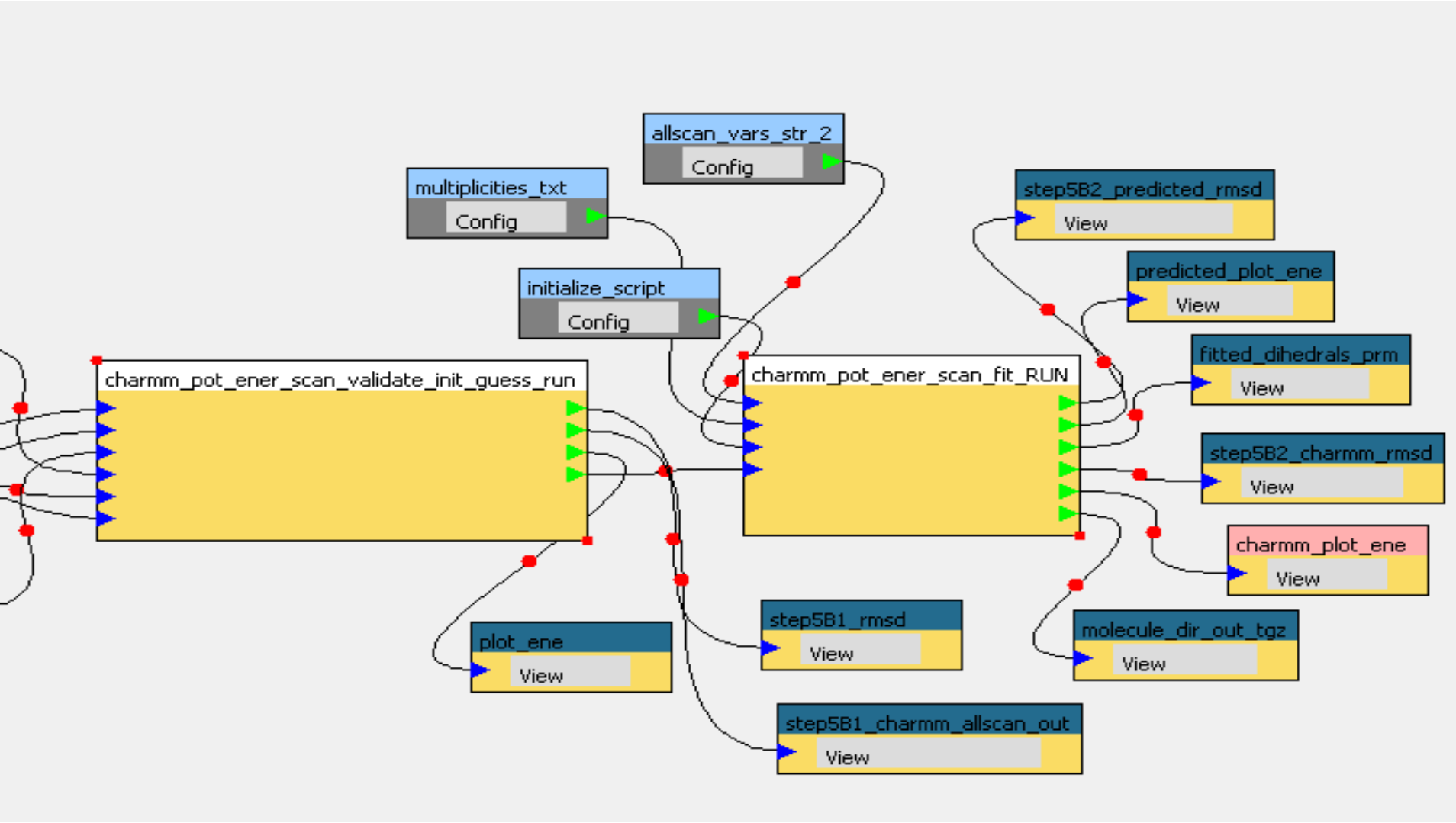
ParamChem Workflow in Airavata



Workflow steps 1-3 in Dihedral Parametrization



Steps 3-4 in Dihedral Optimization and Validation



Acknowledgements

- Support for GridChem/ParamChem integration with Apache Airavata from NSF OCI 1032742 - SDCI NMI Improvement: Open Gateway Computing Environments
- Thanks to Sudhakar Pamidighantam (NCSA) and team for help with slides and demo.

Backup Slides

Force Field Parameterization

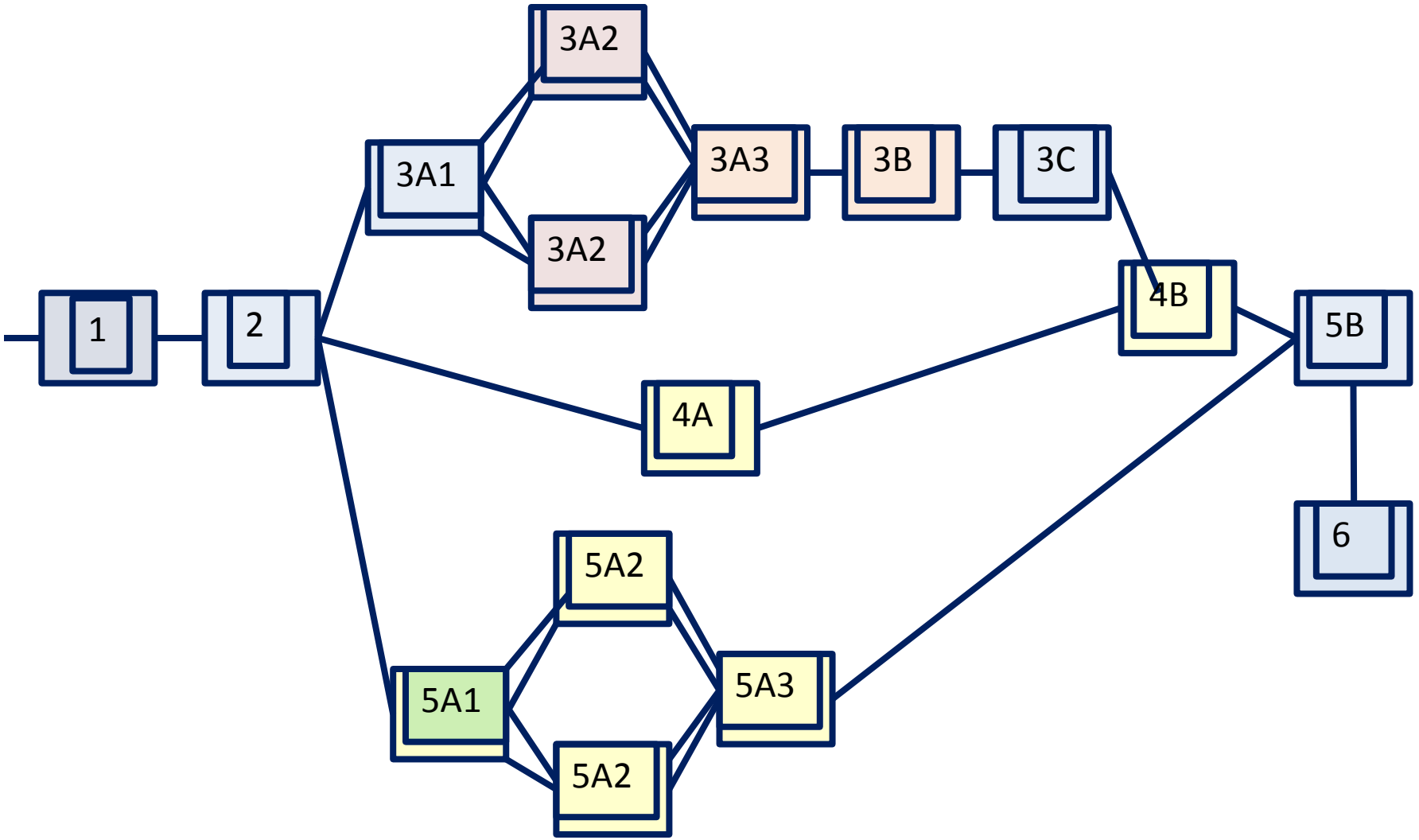
- Molecular Force Fields require constant improvement as new reference data becomes available
- New molecular systems become amenable for computational analysis
- New models/potential energy functions/Hamiltonians for force are established
- Coverage of force fields should constantly be extended to cover new fields of research/new functionality

Paramchem CGenFF Workflow Steps Schematic

1: Generate structure & input for Gaussian MP2 optimization and frequencies
2: Run Gaussian MP2 optimization and frequencies
3A1: Generate inputs for Gaussian HF water interactions
3A2: Run Gaussian water interaction (in parallel)
3A3: Collect results of Gaussian water interactions
3B: Generate input for CHARMM water interaction
3C: Run CHARMM water interactions, optimize charges

4A: Prepare and run CHARMM and MP2 molvib and quick measurements
4B: Optimize bond, angle and improper parameters using molvib and quick measurements
5A1: Setup Gaussian potential energy scans parallel
5A2: Run Gaussian potential energy scans in
5A3: Collect results of Gaussian potential energy scans
5B: Run CHARMM potential energy scans and optimize dihedral parameters
6: Re-optimize charges

Paramchem CGenFF Workflow Steps Schematic as a DAG

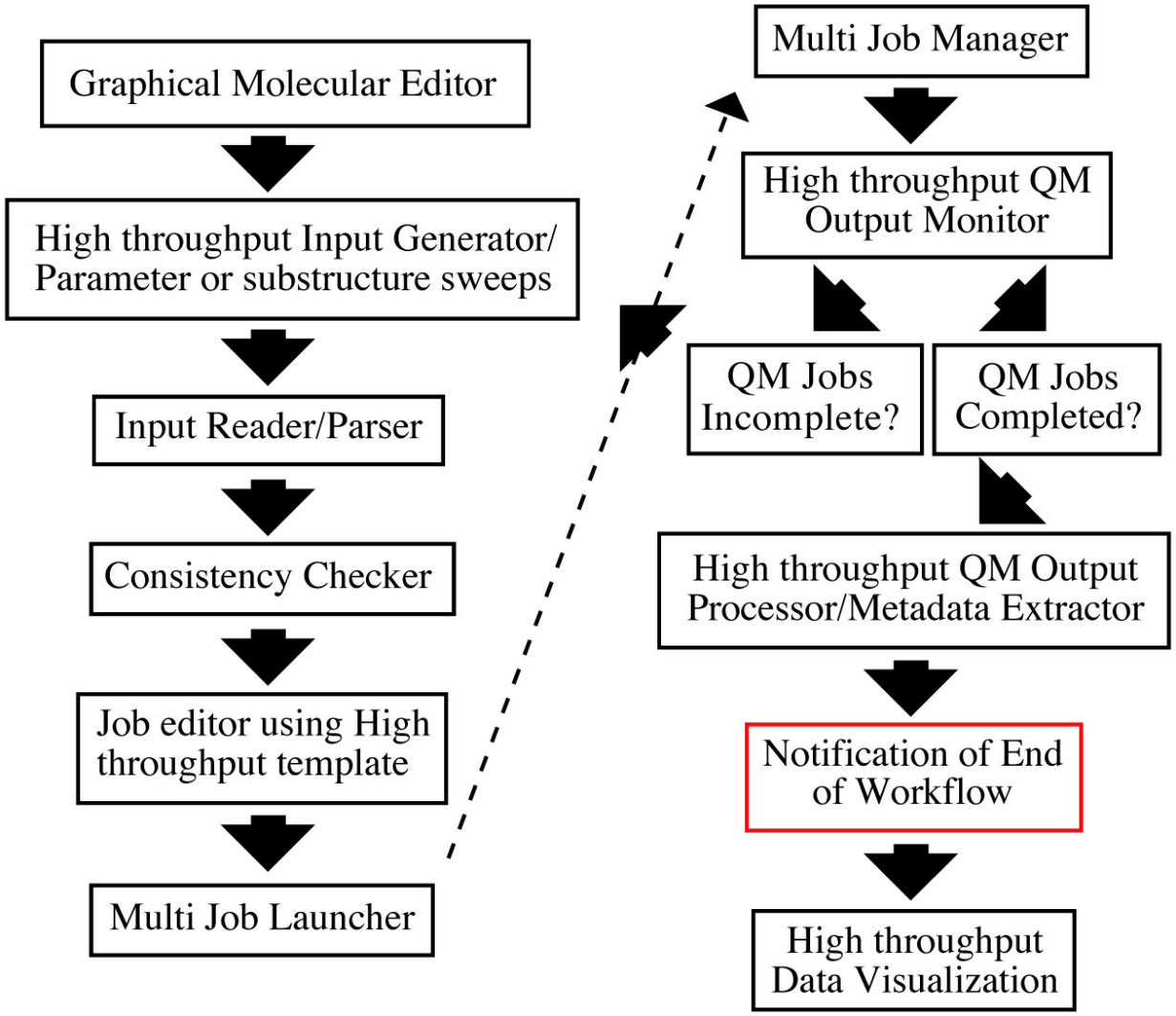


Cyberenvironments for Molecular Force Fields

- Extension of currently available models, with the resulting parameters sets to be made available publicly
- Databases of experimental and quantum mechanical reference data to be used in the parameterization process
- Integration of computational resources for data acquisition, automation of QM reference data generation
- Automation of Extensible infrastructure for parameterization management for rapid and systematic parameterization of existing and novel Hamiltonians (empirical and semi-empirical)
- Systematic improvement of parameter optimization processes

Cyberenvironments for Parameterization

Computational Reference Data Generation



Initial Parameter Assignment

- Parameter Assignment by analogy
  – For a given molecule, generate a list of missing parameters
  – In the case of a missing angle parameter ABC, take one of the outer atoms types (A or C), and look at different branches
  – continue moving up tree
    ⇒ gradual degradation of quality; quality/penalty score
- Assignment of charges by analogy
  – bond charge increment and/or electro-negativity equalization
  – use the existing set of model compounds as a training set

Generate Initial Guess for Optimization