Recent KNIME Developments At Vernalis
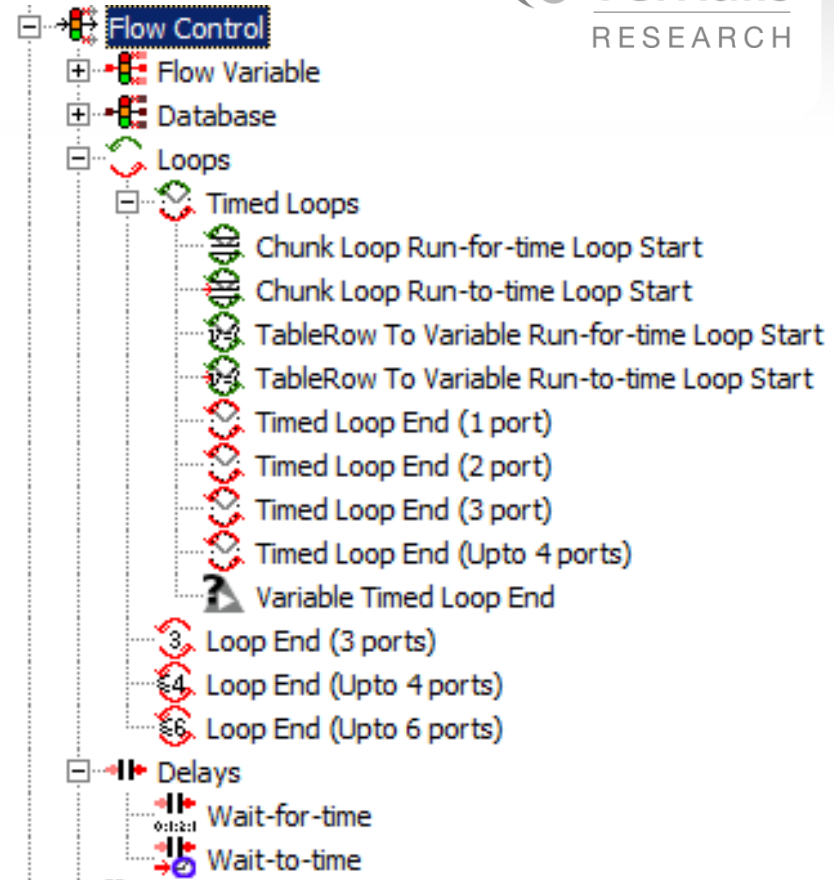
Stephen Roughley

s.roughley@vernalis.com

UKKIUGM 29th September 2014

THE QUEEN'S AWARDS
FOR ENTERPRISE:
2014

Timing Nodes

- Delays
  - Wait-for-time
  - Wait-to-time

- Loops
  - Chunk Loop & TableRow to Variable Starts
    - Run-for-time and Run-to-time variants
  - Timed Ends for 1, 2, 3 and <=4 ports, and Variable Loop end
    - Unprocessed rows go to last output port

Run-TO-Time

Run-FOR-Time

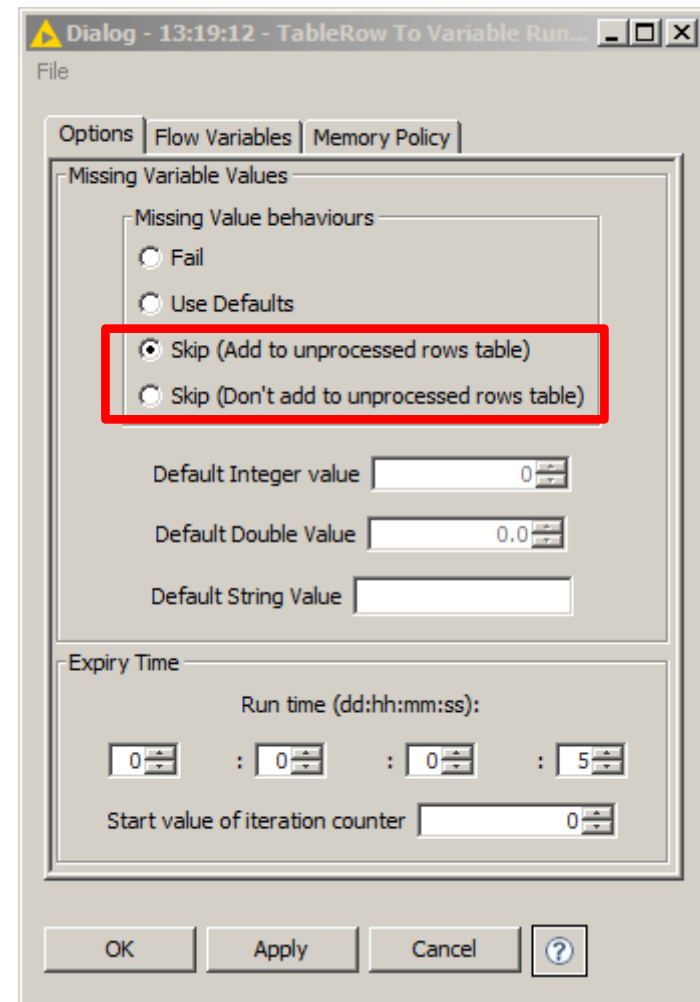# Example – Overnight Conformer Generation
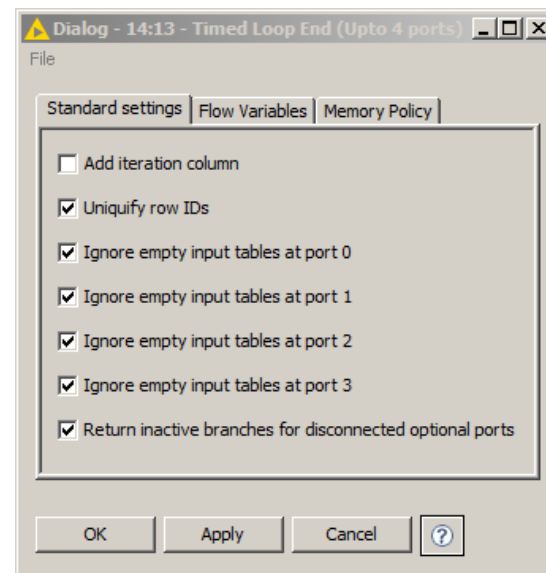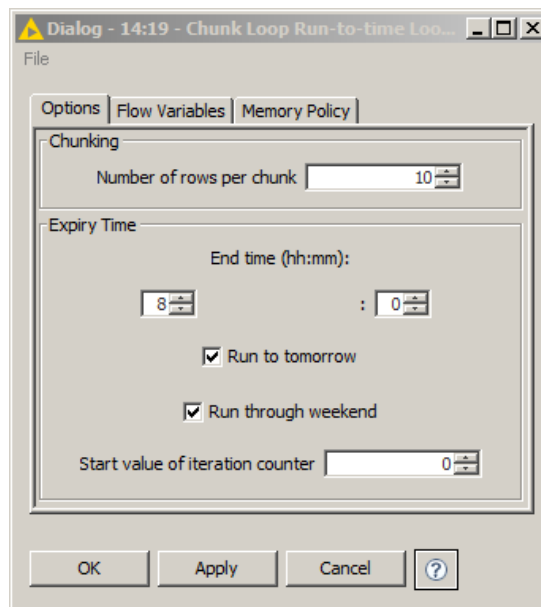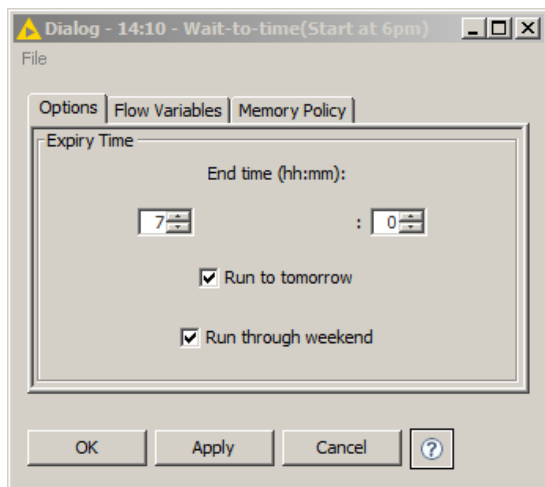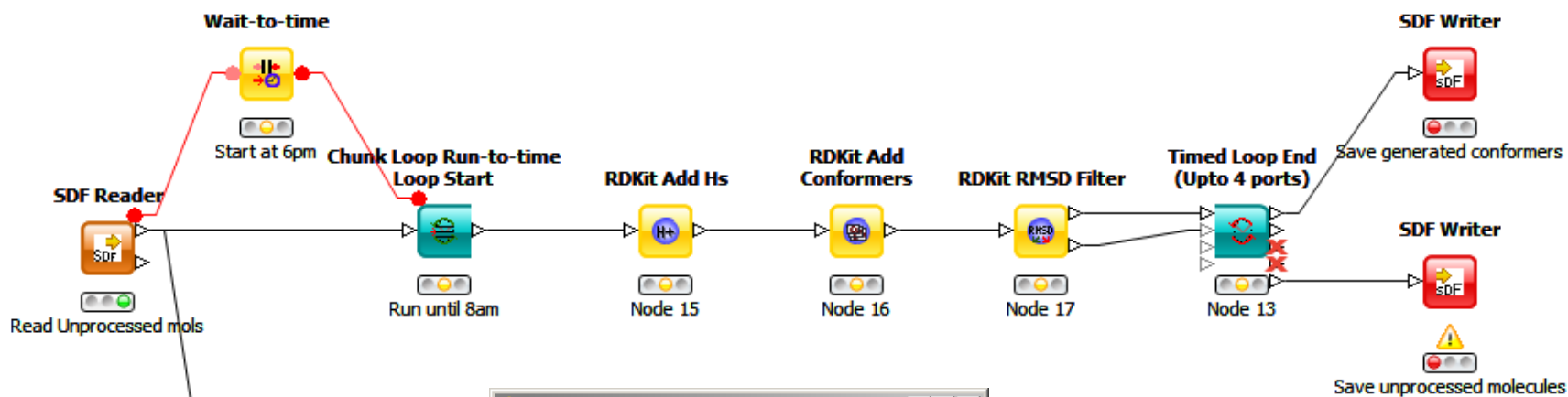
If you are feeling brave…
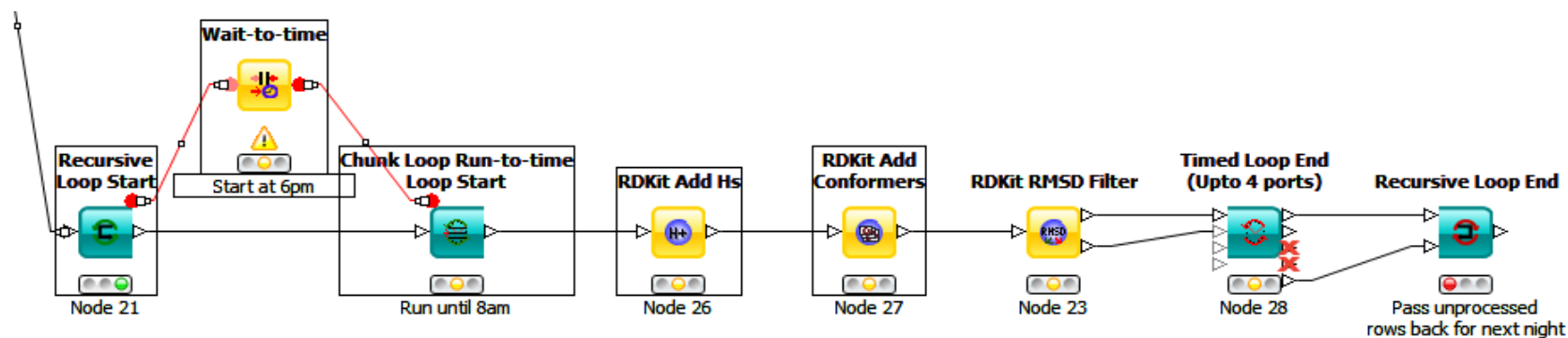
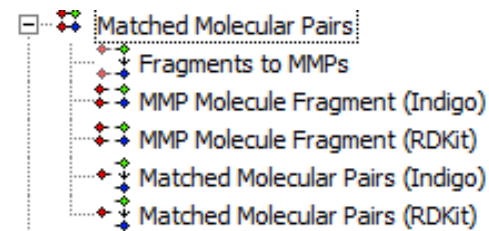- Embed in a Recursive loop and send unprocessed rows back to the beginning

Variable Loop start

- Has same options as TableRow to Variable loop start node
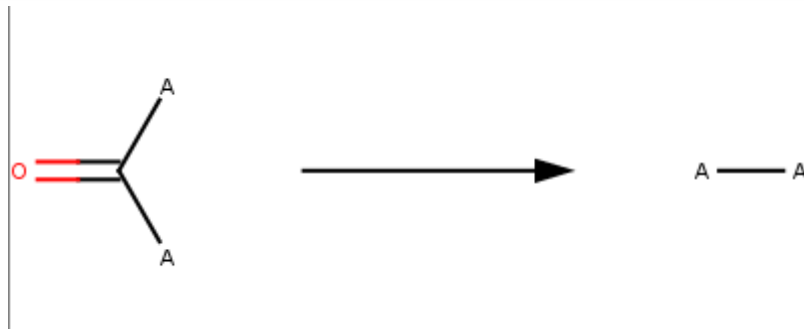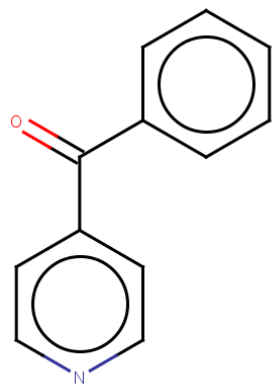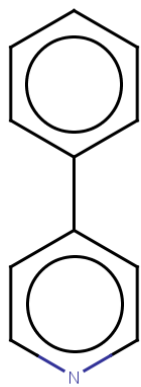 - ADDITIONAL option for skipped rows
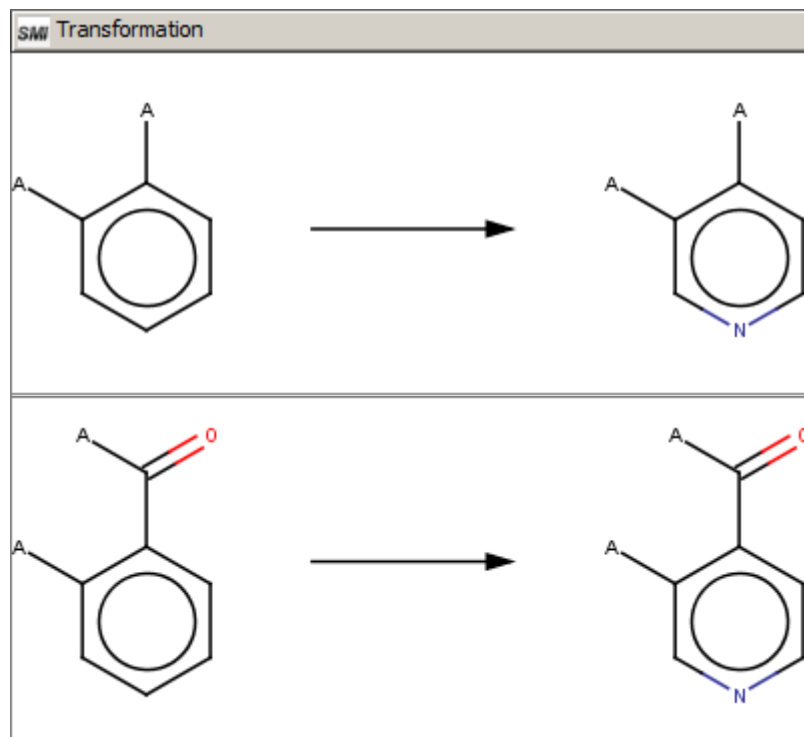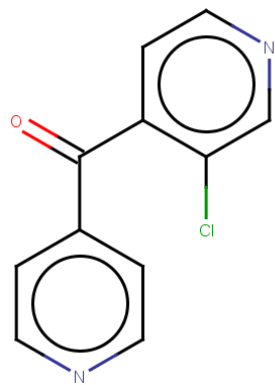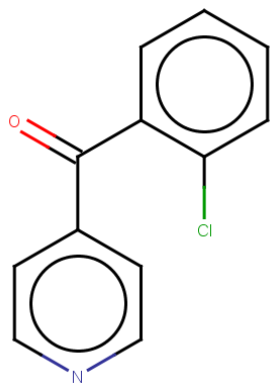
Matched Molecular Pairs

- Re-implemented in RDKit and Indigo toolkits
  - RDKit much faster
  - Upto 10 bond cuts
  - Any acyclic single bond or only acyclic single bonds to a ring atom
- Fragmentation and Pair generation optionally separated
  - Allows databasing of fragmentations and quick MMP generation for new compounds
  - Fragment to MMPs can use pre-sorted Keys
    - Less memory-hungry
    - Can check for sorting errors!
- Constraints on maximum changing Heavy Atoms and Ratio of Changing Heavy Atoms
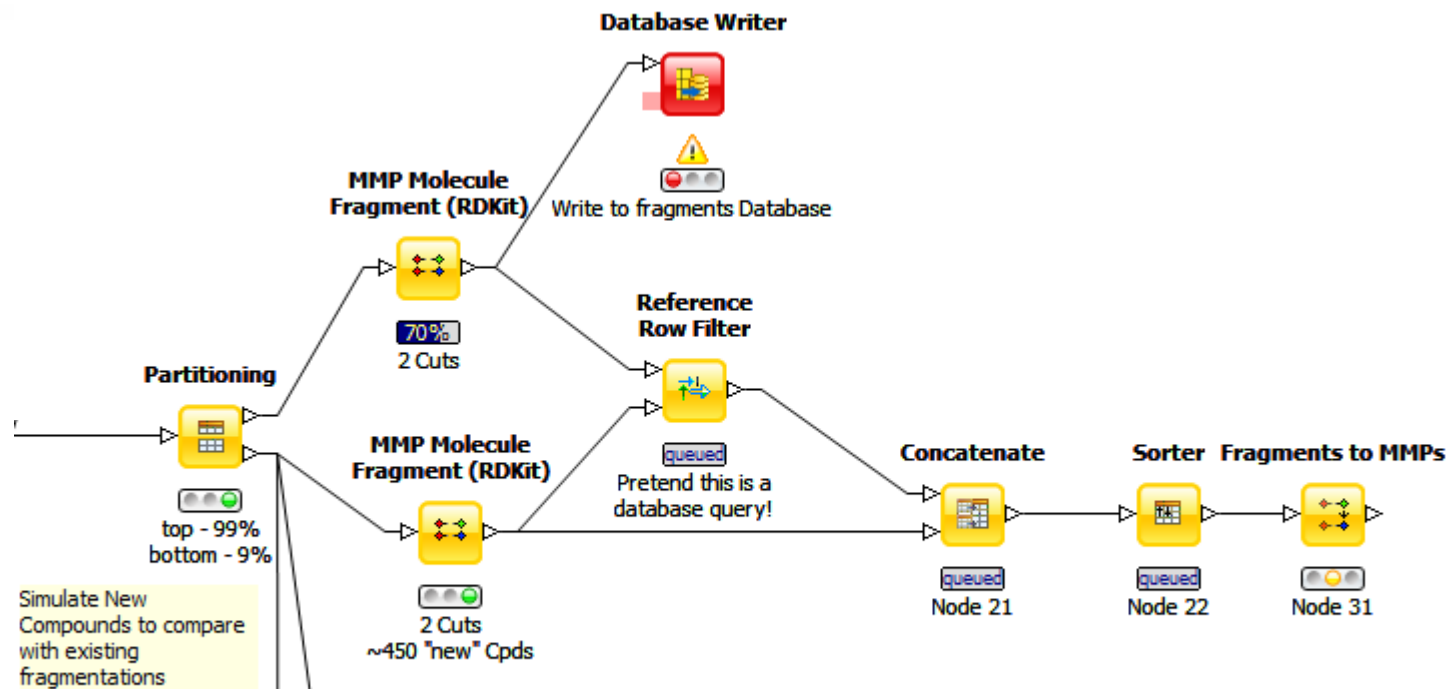- No distance constraints (Yet!)

2-Bond example

- NB 'bond' is a viable replacement

Example Use case

Example Use case

What Else?

- Ertl Scaffold Keys (Counts and Fingerprint)

- Co-ordinate manipulations (RDKit)
  - Rotations, translations etc

- Read / Write Variables
  - Directly analogous to 2.9 Read/Write Table nodes

- Generic Internal Webservices Implementation
  - Abstract NodeModel Class calls service and parses JSON return object
    - ColumnRearranger implemented, uses an Abstract function for return row
  - NodeFactory supplies output column names and types, Service URL, and implements abstract method to convert extract output cells from JSON object

Jenkins

- Build System

- Automatically rebuild, test and deploy new internal update site

- Eventually also run tests nightly internally
  - Increasing test coverage to all nodes - 35/75 so far…

- Currently hogging memory on production knime server
  - Builds against 2.9.4 at present
  - Will update to 2.10.2 once majority of users have migrated
  - Possibly maintain 2.9 site for legacy users

- Thanks to Thorsten@KNIME for the many emails of advice!