



University of
Zurich^{UZH}

Grid Computing Competence Center

Large Scale Computing Infrastructures

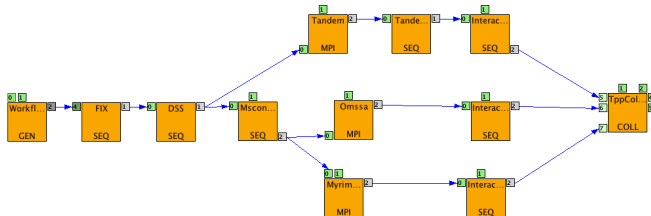
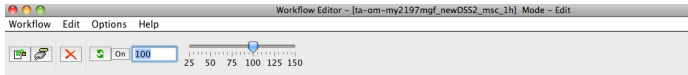
MINF 4526 HS201

Appendix to Lecture 4: High Throughput Scientific
usecases

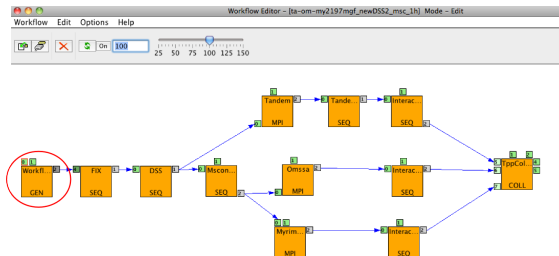
Sergio Maffioletti

Grid Computing Competence Center,
Organisch-Chemisches Institut,
University of Zurich

Mass Spectrometry Workflow

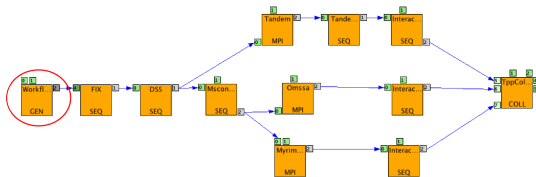


Mass Spectrometry Workflow. step 1



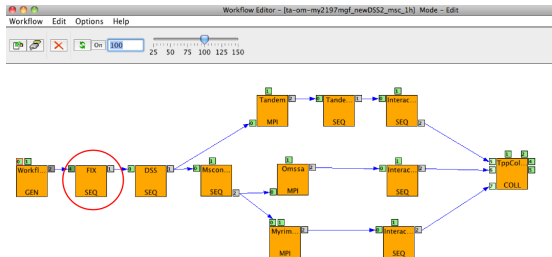
- **input:** 1 .ini file + 1 init folder (containing **N** input files)
- **output:** **N** folders with 1 copy of ini file per folder + 1 input file

Step 2: Start parallel workflow



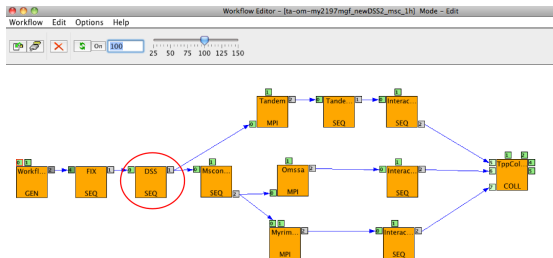
- for each folder, start one 'analysis sequence'

Step 3: analysis sequence FLX



- **input:** 1 .ini file + 1 input file
- **output:** converter input file (according to specs taken from .ini)
- **executable:** fix.bin (already deployed on site)

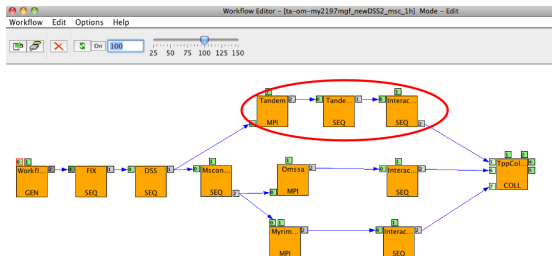
Step 4: analysis sequence DSS



- **input:** output file produced by *FIX* step
- **output:** converted input file
- **executable:** DSS.bin (already deployed on site)

Once **DSS** is terminated, start, in parallel, *Tandem* and *Mscan* sequences

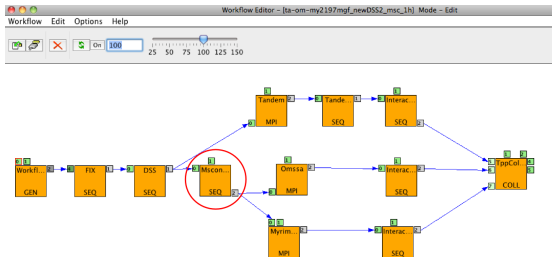
Step 5: Tandem sequence



- **input:** output file produced by DSS + .ini file
- **output:** Tandem.out

Run *Tandem* application twice. First run the *Tandem_mpi* then the result is passed as input to *Tandem*

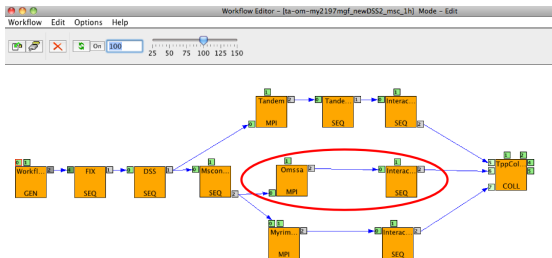
Step 6: Msccon sequence



- **input:** output file produced by DSS + .ini file
- **output:** MSccon.out
- **executables:** MSccon.bin (already deployed on site)

Once *MSccon* step is terminated, start in parallel *Omssa* and *Myrimatch* sequences.

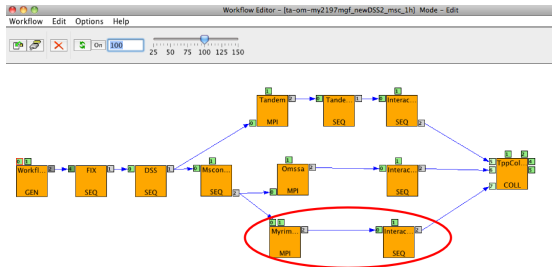
Step 7: Omssa sequence



- **input:** output file produced by Mscan + .ini file
- **output:** Omssa.out
- **executables:** Omssa.bin (already deployed on site)

Once *Omssa* application is terminated, a postprocess step is triggered *Intera*

Step 7: Myrimatch sequence



- **input:** output file produced by Mscan + .ini file
- **output:** Myrimatch.out
- **executables:** Myrimatch.bin (already deployed on site)

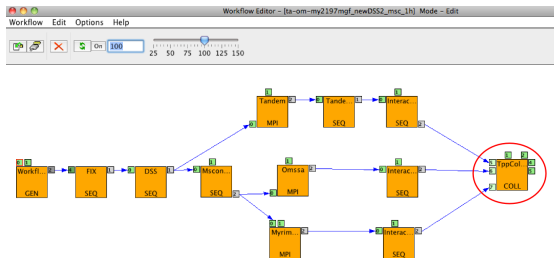
Once *Myrimatch* application is terminated, a postprocess step is triggered *Intera*

Step 7: Intera

- **input:** 1 .ini file + 1 input file (any format from Tandem, Myrimatch or Omssa sequence)
- **output:** tpp compatible format
- **executables:** python script (already deployed on site)

Note that *Intera* takes directives from .ini that has been modified by the previous step

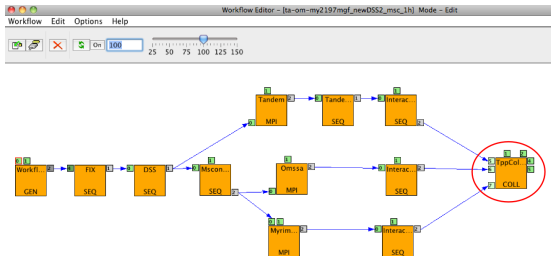
Step 8: TPP



- **input:** .ini file, an arbitrary number of TPP compatible input files
- **output:** *input_file_name_tpp.tgz*
- **executables:** tpp.bin (already deployed on site)

TPP aggregates the results from the three sequences

Mass Spectrometry Workflow



- *TPP* starts **only** when the three sequences are completed

Mass Spectrometry Workflow. handling output

- Each output produced needs to be stored in the same folder of the related input file.