**Instructions to calculate molecular descriptors:**

**SMILES to .log Workflow**

**Generate SMILES strings then convert to .sdf files for conformational searching**

Requirements: open babel, RDkit (installed on your own machine)

1. Generate .txt file with list of SMILES strings for molecules of interest e.g., Using Reaxys, hand-drawn in ChemDraw, etc.
2. Paste SMILES strings into excel file named smiles.xls
3. Assign ID to molecules (keep it simple, something like An1, An2, ….An5000 for anilines)

# e.g py0001, pyridine: py…

1. Open terminal, navigate to folder/directory with smiles.xls and smiles\_to\_sdf.py
2. Locally run conversion script in your terminal

python smiles\_to\_sdf.py

**Conformational searching on cluster using MacroModel**

1. To the folder containing the .sdf files add the following:
   1. “mm\_reference.com”
   2. Folder named “output” containing “convert\_gauss\_UW\_v2.py”
   3. Folder named “cluster\_fail”
2. Transfer this folder to the cluster using scp or client server (i.e., FileZilla)
3. In terminal, ssh into cluster/kestrel

ssh netID@kestrel.chem.wisc.edu

1. Navigate to directory with .sdf and “mm\_reference.com” files (example: folder name “descriptor\_test”)

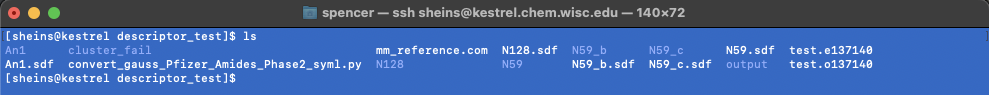
cd /home/sheins/descriptor\_test

**Conformational searching and clustering for >20 conformations using MacroModel**

1. This should run “prep\_MM\_npsh\_UW.sh”, “MM\_clustering\_and\_converting\_UW.sh”, and “check\_manual\_clustering\_UW.sh” scripts

qdesc1 jobname

A text file named “jobname.okestreljob#” is produced:

****

**Figure 1**: Red box shows the text files produced after step 10, where the jobname from step 10 is “test” and the kestrel job# is 137140. Green box is the output folder.



**Figure 2:** Example of text file created after conformational searching and clustering. The last lines indicate which files didn’t cluster properly. In this example, N59\_b, N59\_c, and N59 did not cluster properly.

The text file lists compounds that were converted and clustered. Listed at the end are molecules that were not clustered properly. An example is shown in Figure 2, indicating that N59, N59\_b and N59\_c were not clustered, but should have been. Remove these files before proceeding. These must be manually clustered in the Schrodinger user interface.

An “output” folder containing .com files for individual and clustered conformers, and original .sdf files is also produced (Figure 1, green box). In “output” folder is a new folder named “clustered”, containing .sdf files of molecules that should have been clustered (i.e., molecules that did and did not cluster correctly).

1. Save all files at this point onto a local machine or secure location.
2. Move files and folders (.sdf files and new folders created in directory you started in) for molecules that did not cluster properly into “cluster\_fail” folder created in step 7. In this example, they would be N59, N59\_b, N59\_c. Do this using mv command or client server.

**Prepare Gaussian input files**

1. Start interactive session

qsub -I

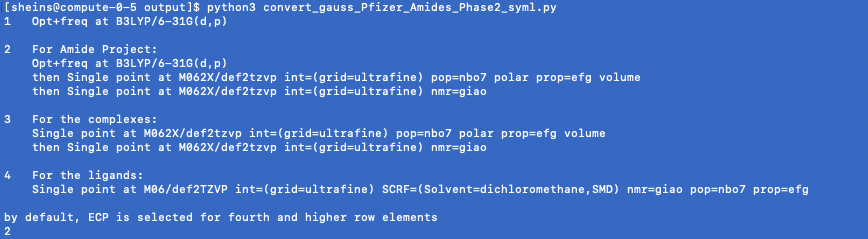
1. Navigate to output folder

cd /home/sheins/descriptor\_test/output

1. In same directory/folder with conformations and .com files (i.e., output folder) run python script:

python3 convert\_gauss\_Pfizer\_Amides\_Phase2\_syml.py

when prompted, type “2”, then hit enter

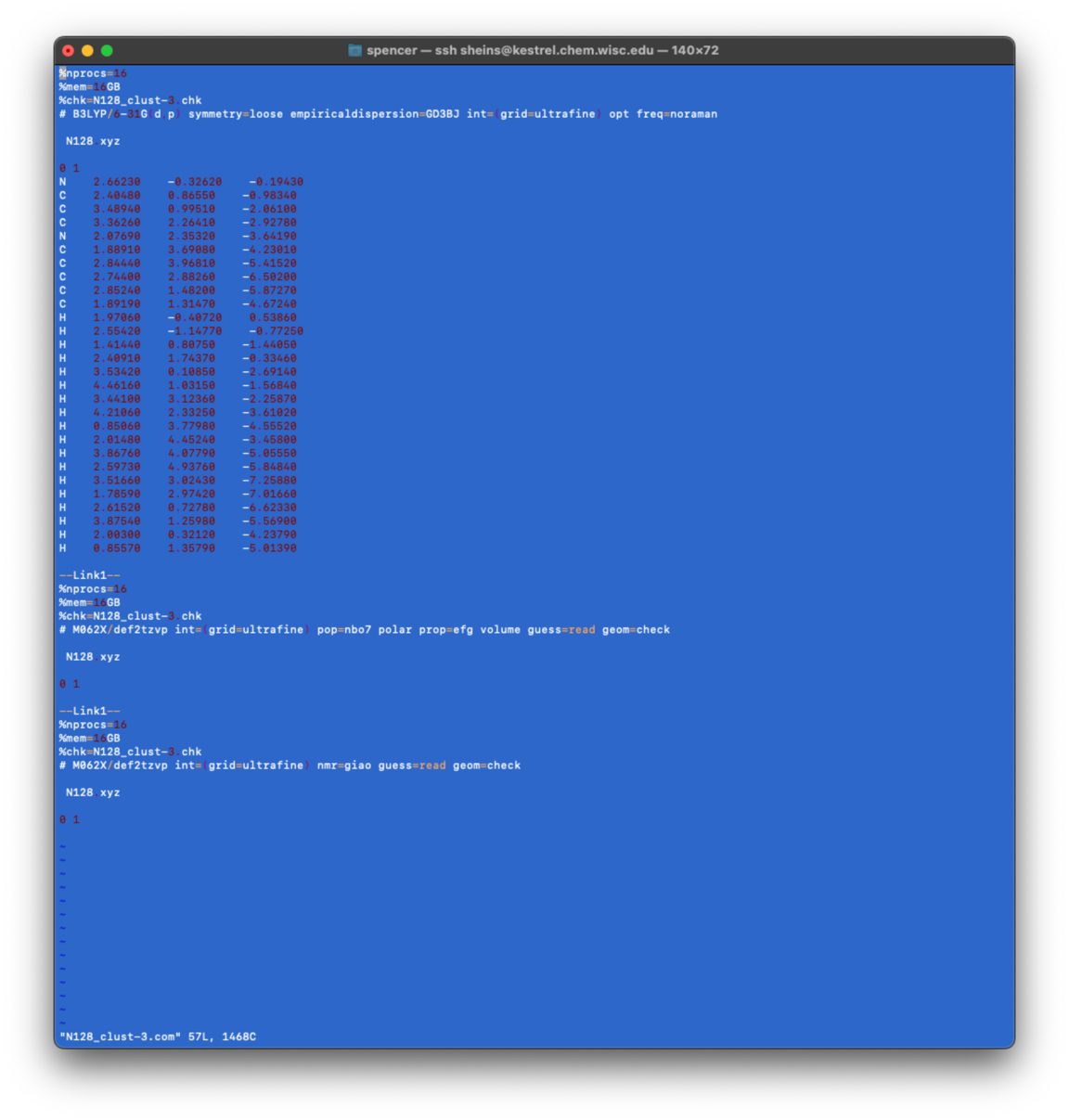


**Figure 3**: Example of prompt asking how to generate the Gaussian input file. In this case, option 2 was used.

1. quit interactive session

exit

The .com files in the output folder should now contain information for a proper gaussian job. An example is shown in Figure 4:



**Figure 4**: Example Gaussian input file for N128, containing geometry optimization, frequency calculations, cartesian coordinates, and two linked jobs.

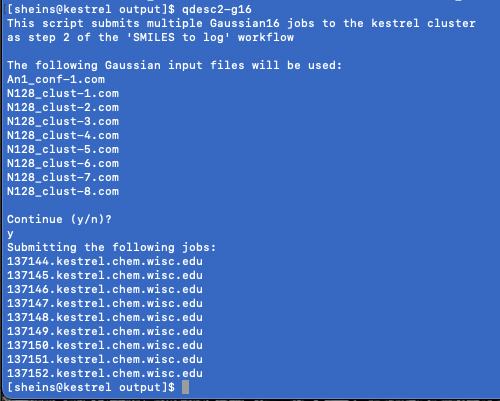
**Submit all jobs to Gaussian**

1. In output folder, submit gaussian jobs

qdesc2-g16

(answer “y” to submit jobs)

wait for jobs to finish



**Figure 5**: Example of submission of all jobs in the output folder

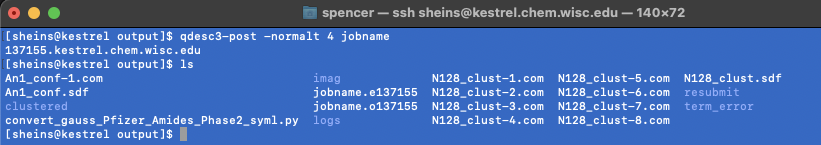
**Check .log files for errors**

Checks .log files for normal termination, no imaginary frequencies, and that every .com file has a corresponding .log file. Makes intuitive directories to sort various errors. Uses "log\_check\_for\_post\_processing.py” and “post\_processing\_v1\_1\_0\_UW.sh "

1. In output folder, run post processing command. Example shown below with four “normal termination” expected from the Gaussian calculation (i.e., the number of jobs that should have ended normally). 4 is a typical value: Opt + Freq = 2; Each linked job = 1. For the example inpute shown in Figure 2, the number of Gaussian jobs submitted using the .com file is 4.

qdesc3-post -normalt 4 jobname

You should now have four new folders: imag, logs, resubmit, term\_error (Figure 6). The resubmit folder contains the molecules that need to be revisited for calculations.



**Figure 6**: Files and folder in output folder following post-processing step (step 19).

1. Save all data to your computer or other secure location.

**Dealing with errors**

Imaginary frequencies:

* Caused by vibrational frequencies being calculated for a conformation that is not at its minimum on the potential energy surface.
* Frequencies are in cm-1
* Steps to remedy
  + Open .log file (vi editor)
  + Go to bottom of file (shift+g)
  + Search for “freq” (?freq)
  + Look at imaginary frequency and its magnitude. If it is small (1-10) maybe ignore.
  + Next calculate the force constants during the geometry optimization and resubmit the jobs
    - Create folder to move files to store originals, and move logs out of way
      * mkdir tmp
      * cp \*.com
      * mv \*.log
    - Replace opt with “opt=calcfc” by doing the following:
      * In director with .com’s run:
        + sed -i “s\opt\opt=calcfc\g” \*.com
    - Resubmit jobs
    - When complete repeat post processing
* If it fails again
  + Open log in gv
  + Selected the freq job
  + Results ->vibrations-> animate
  + Select frequency line of interest (i.e., negative)
  + Select manual displacement
  + Displace/vibrate the mode you selected
  + Save structure
  + Copy new coordinates into com file and resubmit

Termination error

* Check that there was not a step error, i.e., When iterations of the (self-consistent field) SCF process converge, the energy is stationary with respect to infinitesimal variations in the orbitals. Max number of steps reached before this convergence occurred leads to a step error
* Open log file in vi editor and search for steps from the bottom of the log file (shift+g command goes to bottom)
  + Vi editor command to search for “step”: ?steps
* If you see a message saying the maximum number of steps were reached (N is based on number of atoms in molecule) then change .com file to use a different scf algorithm (xqc)
  + Add scf=xqc to each com file by:
  + sed -i “s\freq=noraman\freq=noraman scf=xqc\g” \*.com
  + \*\*or use opt=calcfc as above\*\*
* resubmit jobs
* when complete repeat post processing

error in dihedral -> switch to opt=cartesian (look up cartesian option)

Others:

https://docs.alliancecan.ca/wiki/Gaussian\_error\_messages#Link\_9999