Workflow

Identify Possible Substrates

Reaxys®

Acids: 8,676

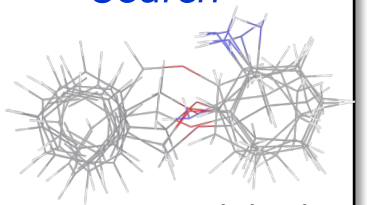
Export SMILES to Excel

SMILES to .sdf Files

SMILES_to_sdf.py

Open Babel

MacroModel Conformational Search



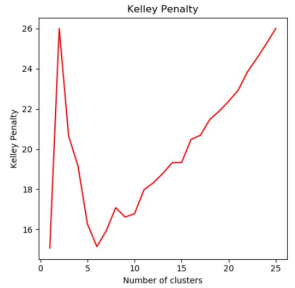
prep_MM_npsh.bash
mm_reference.com

MacroModel Clustering + .maegz to .sdf Files

cluster based on Kelley penalty value if >20 confs.

MM_clustering_and_converting.bash

Confirm clustering



check_manual_clustering.bash

Generate job files

one job per conformer

Group script: Convert Gaussian

Gaussian Geometry Optimization + Single Point

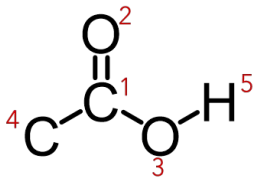
Opt: B3LYP/6-31G(d,p)
SP: M062X/def2tzvp

Check .log Files

normal job termination
+ no imaginary frequencies

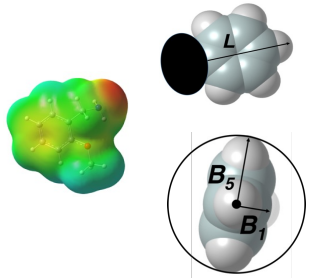
post_processing.bash
log_check_for_processing.py

Atom Number Collection



get_properties_notebook.ipynb

Get Properties



get_properties_notebook.ipynb

Automated Get_Properties_Notebook

Implemented Descriptors

- energies (goodvibes)
- nbo
- nmr
- angle
- dihedral angle
- distance
- plane angle
- total time
- frontier molecular orbitals
- Volume
- polarizability
- dipole
- Sterimol (morfeus)
- buried Sterimol (morfeus)
- buried volume (morfeus)
- buried volume scan (morfeus)
- pyramidalization (morfeus)
- SASA & sphericity (morfeus)
- Sterimol (dbstep)
- Sterimol2Vec (dbstep)
- Hirshfeld charges
- ChelpG

Partially Implemented / Not Generalizable

- IR stretching frequency – *works for one stretch in the input range*

- Ask Melissa or Brittany for a sample if you want to modify these to your needs:
  - Buried Volume: hemispheres, quadrants, and octants (morfeus)
  - Nborbs – *can be customized depending the types of orbital occupancy/energies you want out*

Not Implemented

- e, thermos, efg, tz_e - *use goodvibes instead*
- imaginary frequencies - *should be done in log check*
- qpole
- method
- route lines
- nmr tensors
- dorbs
- buried volume (dbstep) - *takes significantly longer than morfeus*
- Vol2Vec (dbstep) - *takes significantly longer than morfeus Vbur scan*
- cavity *(volume & surface area from solvation model)*