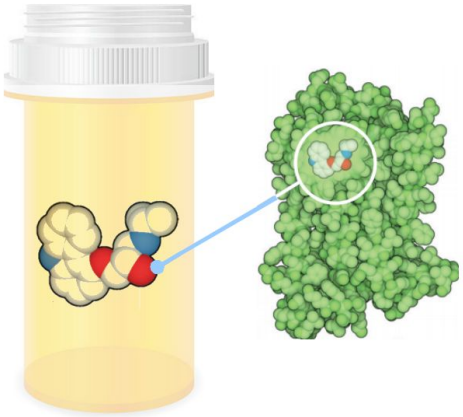
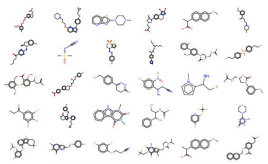
Jessica Maat

University of California, Irvine

Data-Driven Force Field Development for Drug Discovery

Openforcefield.org



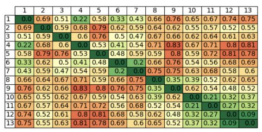


Stage 1

Parse inputs & parameter usage

Parse input molecule & expand tautomeric and isomeric state.

Determine the bond, angle and torsion parameters from .offxml that the input molecule utilizes.



Stage 3

Molecule Selection

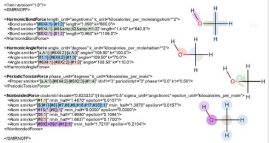
Using randomization, select a molecule from each cluster that is representative of the chemistry.

Generate QM data in QCA.

Starting Datasets

1. Roche set
2. Coverage set
3. eMolecules Discrepancy set
4. Pfizer Discrepancy set
5. Bayer set

Stage 2

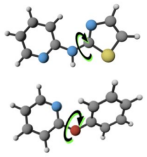


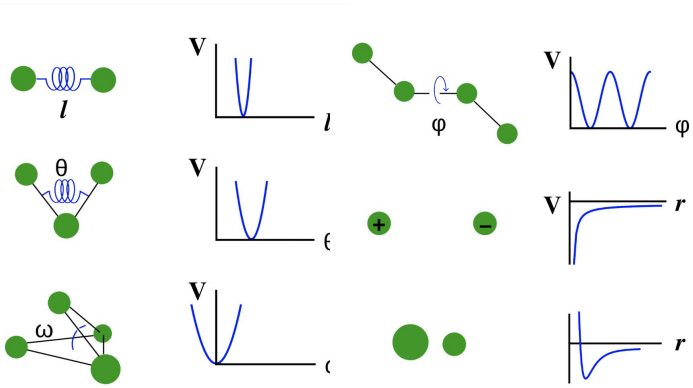
Determine fingerprints & Cluster

Calculate the MACCS key for each molecule.

Cluster the molecule based off the MACCS key using DBSCAN.

Stage 4





Jessica Maat
University of California, Irvine

In my research I deal with large quantum chemistry training datasets for force field optimization, the following topics would benefit my research:

- GPU Programming
- Big data processing
- Machine learning
- Workflow management