Torsion Fitting Update

Yudong Qiu, 06/18/2019

Background

This is the initial approach to build and test the fitting pipeline.

Target Molecules

We mainly used the 468 proposed molecules provided by Roche. More molecules are under preparation to:

1. Provide full coverage of the SMIRNOFF parameters.
2. Identify and improve on differences between SMIRNOFF and other models (GAFF, etc)

Fitting Data Generation Pipeline

The pipeline for generating torsion fitting data is working:

1. Process input molecules, write one molecule per file. (Fragmentation can be done here in the future).
2. For each molecule, identify interested torsions for scanning.
   - Group 1: center bond not in ring, all heavy atoms, one torsion per center bond
3. Submit torsion scan to QCFractal server, run torsiondrive.
   - Early benchmark suggests B3LYP-D3BJ/DZVP method.
4. Download data from QCFractal server.
5. Format downloaded data into ForceBalance fitting targets. Data includes QM relative energies and gradients along torsion rotation profile.

Fitting Data

- Group 1 torsions: 468 molecules, 819 one dimensional torsion profiles, each has 24 torsion angles evenly spaced between 0-360 degree.
- QM energies and gradients are included for all torsion profiles of all molecules.

ForceBalance Fitting Pipeline

1. Select fitting parameters by annotating the offxml file. (automated)
2. Organize all fitting targets. (automated)
3. Select fitting options for optimizer and fitting targets. (manual, based on experience)
4. Run ForceBalance. Performance improved by over 10x after hacking openforcefield toolkit.
5. Analyze fitting results.

ForceBalance Fitting Results

All scripts, data files, input/output files are versioned and detailed steps are documented in the published release: https://github.com/lpwgroup/forcebalance-qcarchive/releases

A. Fit only torsion force constants:

- Parameter coverage: 131/258
- ForceBalance objective function decrease: 1280 --> 970
- ForceBalance number of non-linear steps to covergence: 1
- Notes: Agreement between MM and QM torsion profile varies among fitting targets. Some matches exactly and some are completely off. Further investigation on the problematic targets is needed. For each problematic target, we can identify relevant parameters and run short fittings, and try with increased number of phases.

B. Fit all valence parameters (bonds, angles & torsions)

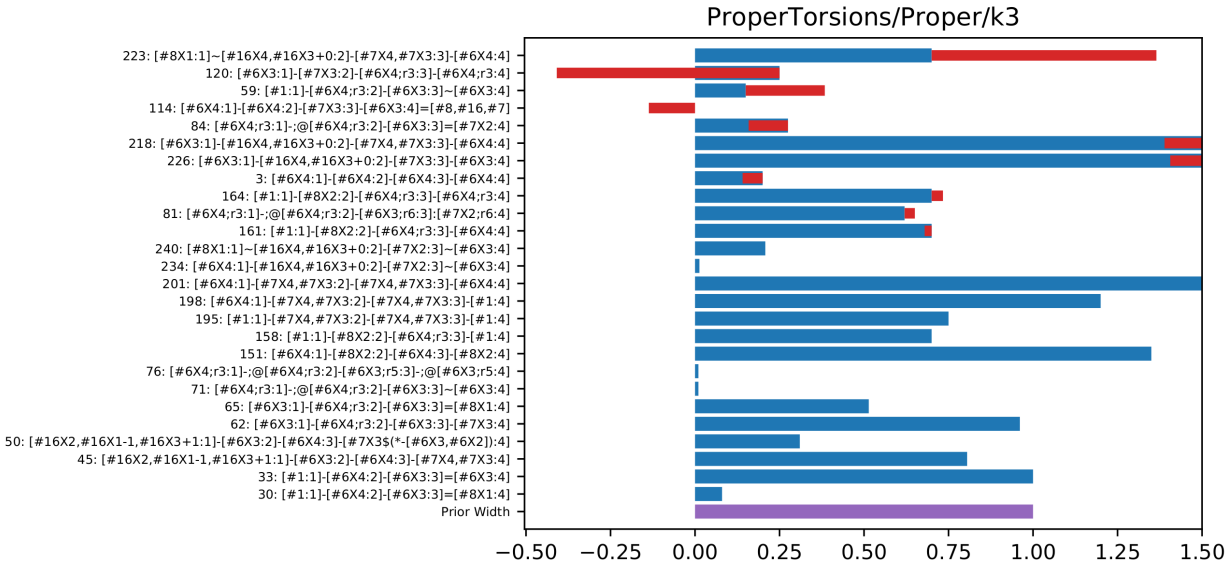
- Parameter coverage: 261/508
- ForceBalance objective function decrease: 1280 --> 480
- ForceBalance number of non-linear steps to covergence: 6
- Notes: The objective function contribution from gradients decreased. Since the QM gradients are all near zero, the MM force constants are reduced substantially in fitting. This indicates that the QM gradients from the torsion profiles are not sufficient for fitting MM valence force constants.

ForceBalance Fitting Result Analysis

Two types of analysis are implemented and plots are ready for interpretation.

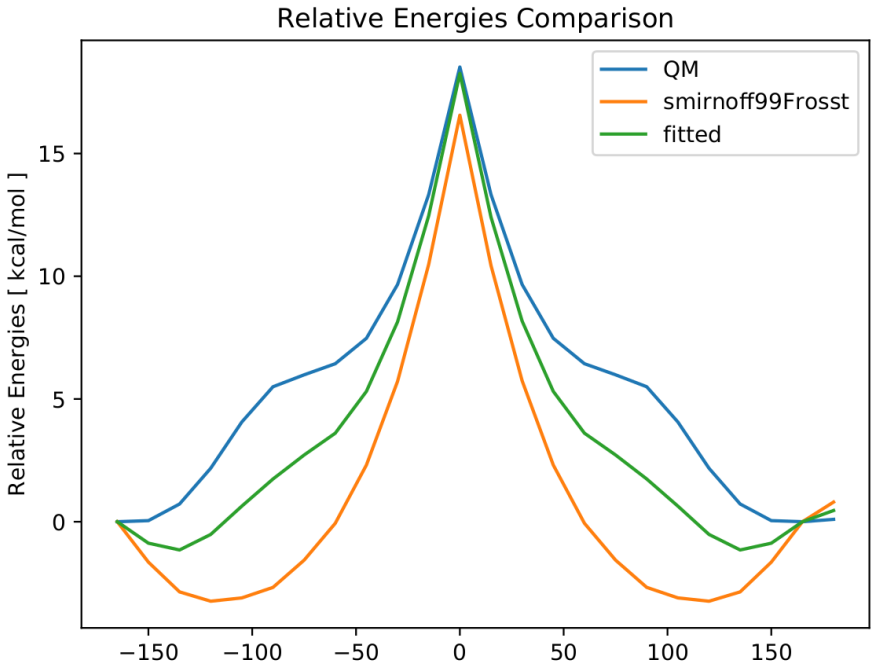
A. Force field parameter changes

- Example



B. QM vs MM torsion profile relative energies comparison

- Example



Ongoing Projects

1. Data curation
   - 22 out of 819 torsion scans are still "incomplete", after new keywords are set. These scans has been verified to be successful in my local server. More discussion with Daniel Smith is needed on fixing them.
   - 3 out of 468 molecules have -CN4 ring that current SMIRNOFF does not support. New SMIRKS is on the way to solve this.
2. Better torsion fitting
   - Identify problematic torsions and improve torsion definition.
3. Better valence fitting
   - New fitting targets are under preparation, for better fitting the bond and angle parameters.
     - Optimized geometries. Internal coordinates such as bond length, bond angle, dihedral angle are fitted individually between QM and MM geometries. The implementation is done, the data is almost ready. (865 complete, 72 error)
     - Vibrational frequencies or hessian on optimized geometries. The implementation is done, the calculation tasks are yet to be submitted.
4. Better ForceBalance interface for interpretation and analysis
   - The ForceBalance web-interface was originally developed with support from MOLSSI Software Fellowship in 2018.
   - The web-interface is continued to be improved, now supporting fitting the SMINOFF parameters to *ab initio* QM energies and gradients.
   - The web-interface provides naturally a more intuitive way to explore fitting data and visualize the results.
   - Analysis plots will be built into the web interface.