ParFit

Version 0.9

June 30, 2015

Contents

Introduction

ParFit is a program designed to calculate molecular mechanics (MM), MM3 and MMFF94, fitting parameters for dihedral angles, referred to as torsions in this document. The parameters are fit to energies derived from quantum mechanics (QM) and checked with engine molecular mechanics software.

This manual is intended to get you started on running ParFit. ParFit is written in Python and uses libraries that must be installed separately, please see "INSTALL" for instructions.

Conventions

Pay close attention to commas. If a comma is needed, it will be included in the example. [ ]: square brackets include descriptive information that should be entered (without the brackets). Values enclosed by brackets and separated by "/" are keywords. Choose one. `format`: fixed width format indicates a command to be entered in the terminal command line.

Getting started

In order to run ParFit, you will need to have Python 2.7.0 or higher (but not Python 3) installed. You will also need the following libraries:

Numpy	http://www.numpy.org/
Scipy	http://scipy.org/scipylib/index.html
Pyevolve	http://pyevolve.sourceforge.net/
*Matplotlib	http://matplotlib.org *optional for plotting curves with Python.

MM Engine (http://serenasoft.com/) is needed to run molecular mechanics calculations and GAMESS (http://www.msg.ameslab.gov/gamess/) for quantum mechanics calculations.

ParFit File Structure

The ParFit parent directory contains the subdirectories, ParFit, GAMESS, ENGINE, Doc, and Utils. The ParFit directory contains the executable file, ParFit.py and input file, dih_scan_inp. The quantum mechanics calculation data, in the form of GAMESS output files or a condensed file, should be put in the GAMESS directory when running a "full" or "comp" ParFit run type, respectively, see run type information below. The ENGINE directory contains MM Engine input ("coengine"), output files (.pcm), and parameter files (.prm files) from engine runs executed by ParFit. Detailed manual and documentation is found in the Doc subdirectory. One can find various utilities described in detail below, in the Utils directory. For more information regarding the input and output files from MM Engine or GAMESS codes, please visit dedicated websites.

Running ParFit

Navigate to the Parfit subdirectory where ParFit.py file is contained. From that directory, run ParFit using the command: `./ParFit.py`

The program takes its directives from an input file named "dih_scan_inp" located in the same directory. One should create an input file with a descriptive name and then copy it to "dih_scan_inp" before running ParFit. To generate an input file, run PFinp.py located in the Utils directory, or generate one using a text editor.

There are 3 ParFit run types:

- Full: reads QM data from a series of GAMESS output files that each contain geometry and energy information.
- Compact: reads QM data from a condensed data file containing all geometry and energy information.
- Ginp: generates GAMESS input files for each torsion.

The input file structure will depend on the run type chosen. A detailed input file description for each run type follows.

ParFit Input Files

Each run type will have particular lines in the input file. To create a ParFit input file one can run the interactive PFinp.py program or a user may choose to create or modify an input file using a text editor.

The input file has the form:

- Line 1 -
  [full/comp/ginp*], [filename root], [atom index 1] [atom index 2] [atom index 3] [atom index 4], [starting torsion] [final torsion] [torsion step size]
- Line 2 -
  [engine executable path]
- Line 3 -
  [mm3/mmff94]
- Line 4 -
  [ga/fmin]
- Line 5, repeated for each parameter to be fit. -
  [line number of parameter to fit] [c/p] [c/p] [c/p]
- Last line -
  [csv_on/csv_off]

*Notice: If run type “ginp” is chosen, only line 1 is needed.

On line 1, choose between “comp,” “full,” or “ginp” run types. The “filename root” should be the root filename of the QM data file or files. The atom indices are simply the atom numbers defining the torsion. Starting, final and step size of the torsion should be determined based on symmetry of the molecule. Line 2 contains the entire engine executable file path. The MM type is indicated in line 3, mm3 or mmff94. On line 4, one can choose between using ga, a genetic algorithm or fmin, a Nedler-Mead simplex algorithm for fitting the parameters. Line 5 and on, instructs ParFit on which parameters to modify in order to fit the QM data. Each MM parameter has three phase values that can be modified. Define a parameter to be fit by indicated the line in which it appears the original “.prm” parameter file, located in ParFit_root_dir/ENGINE/. Choose which of the three phases ParFit is to modify by indicating “p” for phases that can change and “c” for those that should remain constant. The final line toggles the printing of a comma separated value (csv) file containing the energy comparison of QM derived energies and the MM energies after the requested parameters have been fit.

ParFit Data Files

ParFit fits MM parameters to QM potential energy data located in the [ParFit root dir]/GAMESS directory. Each run type will require a particular data format and is outlined below for each run type.

Full

If you are running run type "full," ParFit expects full GAMESS output files in the ~/ParFit_root_directory/GAMESS directory. There should be one output file for each torsion that is used for parameter fitting and the file names must follow the format:

[file_name_root]-[torsion value].log

Compact

Compact runs require that the data from quantum mechanics calculations be condensed into a single file containing geometry, energy and torsion for each torsion on the potential energy surface. The file name format should be:

[file name root]-scan

and the file should have the following format:

- Line 1 -
  [atom index 1] [atom index 2] [atom index 3] [atom index 4]
- Line 2 -
  [initial torsion] [final torsion] [torsion step size]
- Line 3 -
  [number of structures included in the comp file]
- Line 4 -
  [file name root] [number of atoms in the structure] [torsion] [structure energy]
- Line 5 -
  [atom] [x-coordinate] [y-coordinate] [z-coordinate] [atomic number]
- Repeat line 5 for each set of coordinates in the molecule–
- Repeat lines 4 and 5 for each torsion value

Ginp

The ginp run creates a series of GAMESS input files using the values from the ParFit input file. It does not read data from any files.

Utilities

A group of scripts that may be useful for evaluating ParFit runs or facilitating the collection of data needed to run ParFit. Below is a list of utilities.

PFinp.py: an interactive ParFit input file generator. Run by using command: `./PFinp.py` from the Utils directory.

opt.csv: a file containing comma separated values. Column 1 is the torsion, column 2 contains the mp2 energy, and column 3 is the MM energy calculated using the newly fit parameters.

submit.py: a template program used to submit and run the series of GAMESS input files.

rename.py: a template program used to remove the “-dlc” from the GAMESS name root.
If matplotlib is installed and you’d like to use it to plot the fit, you can use these templates:
mp2_mm3_OPMMPH.py
mp2_mm3_OPMMPH_csv.py

Outputs

ParFit generates files based on the run type chosen.

Run Type	Location	Output files
Full & compact	[parfit_root]/ENGINE/	add_[MM]_PF.prm (MM=mm3 or mmff94) [filename root]_[torsion]_inp.pcm and [filename root]_[torsion]_out.pcm
	[parfit_root]/Utils/	opt.csv (optional, prints if input file contains “csv_on”)
ginp	[parfit_root]/GAMESS/	[filename root]_[torsion].inp

The resulting parameters fit by ParFit can be found in the “add_[MM]_PF.prm” where MM is the MM type chosen in the input file and can be used when running MM calculations. Plotting the resulting MM energies and QM energies, found in the opt.csv file, one is able to check the fit quality.

Contacts

Theresa L. Windus, Principal Investigator twindus@iastate.edu
Federico Zahariev, Developer fzahari@iastate.edu
Marilu Dick-Perez, Tester marilu@iastate.edu