ForceFit 2.0

Manual

(last updated July 2011)

Reference: Waldher, Benjamin[*]; Kuta, Jadwiga[+]; Chen, Samuel[#]; Henson, Neil; Clark, Aurora E. ForceFit: A Code to Fit Classical Force Fields to Quantum Mechanical Potential Energy Surfaces, *Journal of Computational Chemistry* **2010**, *31*, 2307-2316.

Table of Contents

Description of ForceFit

The ForceFit package has separate modules for reading in quantum mechanical data (scanReader subdirectory), creating MD input files (molecularDynamics subdirectory), minimization of the forces (minimizer subdirectory) and the gui (gui subdirectory). The gui is written in GTK. New modules can easily be built based upon the existing ones, allowing modifications for reading different QM code outputs or interfacing with new MD codes.

Compilation Instructions

Making ForceFit should be trivial, however errors will be encountered if you do not have all of needed libraries or compilers on your system. ForceFit requires C++/g++ compilers as well as the libglade and gtkmm libraries.

Creating a PES Database

The input files for ForceFit consist of either energies, gradients, or vibrational frequencies, or a combination of all. To date, most usage of the code has been to fitting of either energies or gradients. Energy, gradient, and vibrational data is obtained from either Gaussian03 or NWChem and may be performed at any level of theory desired. In our experience, some classical interatomic potentials are highly sensitive to the level of theory used for the PES database, while others are not. Benchmarking of this sensitivity is highly recommended.

The input files for Gaussian03 or NWChem must have specific keywords such that ForceFit can use their output during fitting of a potential.

Gaussian03

Input Files

The G03 input file will vary greatly depending on the types of calculations being performed with the exception of a few critical keywords needed by ForceFit later. ForceFit needs the atomic gradients to do the fitting, which can be done with the keyword “Freq” or “Force” in the route section. It is also important to not have symmetry, so the keyword “nosymm” in the route section should also be used. Either Cartesian coordinates or Z-matrix coordinates may be used. More information about how to create an input file for a specific job on Gaussian can be found in the Gaussian03 user’s manual.

Keywords: “frequency” and “nosymmetry”

```
%nproc=1
# b3lyp 6-31g* Integral=(Grid=UltraFine) SCF=(verytight) freq nosymm

1 Ang

0 1
 O                 -0.57417782    1.82968525    0.00000000
 H                  0.38582218    1.82968525    0.00000000
 H                 -0.89463240    2.73462108    0.00000000
 O                  0.24767802    2.39938077    0.00000000
 H                  1.20767802    2.39938077    0.00000000
 H                 -0.07277657    3.30431660    0.00000000
```

A sample Gaussian03 input file for two water molecules at 1 Angstrom separation

Output Files

The resulting G03 output file will contain a table of atomic gradients, the atomic charges, and the total energy. While fitting is generally performed to gradients (read from the table below), the total energy is generally used to create the initial guesses for the parameters of the potential that will be fitted by ForceFit.

```
---------------------------------------------------------------------
 Center     Atomic                   Forces (Hartrees/Bohr)
 Number     Number              X              Y              Z
 ---------------------------------------------------------------------
    1          8          -1.283447195   -0.672546284    0.000000000
    2          1           0.745097898   -1.695982469    0.000000000
    3          1          -0.311761399    0.158375575    0.000000000
    4          8           0.714437735    1.852019445    0.000000000
    5          1           0.206926289    0.089760100    0.000000000
    6          1          -0.071253327    0.268373633    0.000000000
 ---------------------------------------------------------------------
```

An X, Y, Z table of atomic gradients in the output file from Gaussian03

```
1\1\GINC-C0-16\Freq\RB3LYP\6-31G(d)\H4O2\ELIZA\12-Jul-2011\0\\# b3lyp
 6-31g* Integral=(Grid=UltraFine) SCF=(verytight) freq nosymm\\1 A\\0,1
 \O,-0.57417782,1.82968525,0.\H,0.38582218,1.82968525,0.\H,-0.8946324,2
 .73462108,0.\O,0.24767802,2.39938077,0.\H,1.20767802,2.39938077,0.\H,-
 0.07277657,3.3043166,0.\\Version=AM64L-G03RevE.01\HF=-151.2361898\RMSD
 =4.466e-09\RMSF=7.349e-01\ZeroPoint=0.0677824\Thermal=0.0711278\Dipole
```

The HF value in the output file from Gaussian03

```
Mulliken atomic charges:
1
1  O   -0.232334
2  H    0.208491
3  H   -0.040649
4  O   -0.022054
5  H    0.122425
6  H   -0.035878
Sum of Mulliken charges=   0.00000
```

Mulliken charges table in the output file from Gaussian 03

NWChem

Input Files

As in G03, the input file will look very different depending on what type of calculations are being performed, except for a few necessary keywords that ForceFit will need to use later. The keywords are "Mulliken" and the task should include "gradient." More information on how to create an input file for a specific job for NWChem can be found in the NWChem User's Manual.

Keywords: "gradient" and "Mulliken"

```
echo

title "1A"

start 1A

geometry complete
 O                 -0.57417782    1.82968525    0.00000000
 H                  0.38582218    1.82968525    0.00000000
 H                 -0.89463240    2.73462108    0.00000000
 O                  0.24767802    2.39938077    0.00000000
 H                  1.20767802    2.39938077    0.00000000
 H                 -0.07277657    3.30431660    0.00000000
 end

basis
oxygen library 6-31g*
hydrogen library 6-31g*
end

set geometry complete
dft
mulliken
mult 1
iterations 800
grid xfine
xc b3lyp
tolerances tight accCoul 12
decomp
end
task dft gradient
```

A sample NWChem input file for two water molecules at 1 Angstrom apart

Output Files

There are, three sections of note in the output files of NWChem. First, there are the DFT Energy Gradients table and the Mulliken Charges table, which are later used by ForceFit. Next is the total energy, which is used in generating initial variables.

```
                             DFT ENERGY GRADIENTS

 atom               coordinates                        gradient
           x          y          z          x          y          z
1 O      -0.134972   1.135870   0.000000   -0.098317  -1.445647   0.000000
2 H      -1.084095  -0.410176   0.000000    1.835177  -0.252314   0.000000
3 H       1.639219   0.757269   0.000000   -0.298080  -0.182831   0.000000
4 O      -0.030041  -0.750940   0.000000   -1.204550   1.577801   0.000000
5 H      -0.979164  -2.296987   0.000000    0.031765   0.223305   0.000000
6 H       1.744150  -1.129542   0.000000   -0.265995   0.079686   0.000000
```

Energy gradients table in the output file from NWChem

```
        Total DFT energy =     -151.236189255548
     One electron energy =     -341.319882448510
          Coulomb energy =      139.842103664852
        Exchange energy =      -18.427101182802
      Correlation energy =       -0.921256863633
Nuclear repulsion energy =       69.589947574545
```

Total energy in the output file from NWChem

```
-----------------------------------------
     Total Density - Mulliken Population Analysis
    -----------------------------------------

   Atom       Charge   Shell Charges
-----------   ------   -------------------------------------
   1 O    8     8.31   2.00  0.94  3.17  0.72  1.37  0.10
   2 H    1     0.77   0.57  0.20
   3 H    1     1.27   0.47  0.79
   4 O    8     8.09   2.00  0.89  3.44  0.47  1.07  0.22
   5 H    1     0.33   0.28  0.05
   6 H    1     1.23   0.48  0.76
```

Mulliken charges table in output file from NWChem

ForceFit uses the Mulliken charges by default to calculate the Coulombic energy from either NWChem or Gaussain. If the Mulliken charges are not the charges desired, then they will need to be changed manually in the output file for either Gaussian and NWChem.

Generating Variable Guesses

The next step in preparing data for ForceFit is to create initial guesses for the variables that will be fit within the potential. In the following example, an initial guess is created for epsilon and sigma within the Lennard-Jones potential using Excel.

Sheet 1: Coulombic Energy Calculations

If a Coulomb term will be used in the potential within the MD, then this term must be incorporated when determining the initial guesses for the variables. Below is a sample of calculating the Coulombic energy of one geometry. There will need to be one table like this for each geometry. This example is for the same two water molecules mentioned in the previous section, at one Angstrom separation.

Note: Formulas for each column need only be entered once and simply dragged down to the rest of the geometries entered.

Column A labels which geometry this section is, in this case 1 Angstrom apart, then lists the atoms in the order they were listed in the input file.

Column B has the charge for each atom. For this example the fixed charges from the SPC water model were used. The specific charges to use are up to the user's discretion.

Columns C-E are the X, Y, Z coordinates directly from the input files.

Columns G-K contain the Coulombic energy calculations. The energy calculations need to be calculated for each pair of atoms individually. The generic equation used is $charge_1$*$charge_2$/distance. For cell G4 the specific equation is **[=B3*B4/(SQRT ((C4-C3)^2+(D4-D3)^2+(E4-E3)^2))]**. That equation can be dragged down the entire column for all of the geometries, just delete the unnecessary cells. That will calculate the energy for all consecutive pairs. Column H calculates the energy for atom pairs 1 and 3, 2 and 4, 3 and 5, etc. Column I calculates the energy for atom pairs 1 and 4, 2 and 5, 3 and 6, etc. Calculate the remaining columns in this manner until all pairs have been calculated.

Finally, at the bottom of each geometry in *Column A* is a sum of all the energy for that geometry. It is this sum that will be used in the next spreadsheet.

	A	B	C	D	E	F	G	H	I	J	K
1		**Charge**	**X**	**Y**	**Z**						
2	**1 Ang.**										
3	O	-0.82	-0.57417782	1.82968525	0						
4	H	0.41	0.38582218	1.82968525	0		-0.350				
5	H	0.41	-0.8946324	2.73462108	0		0.107	-0.350			
6	O	-0.82	0.24767802	2.39938077	0		-0.282	-0.574	0.672		
7	H	0.41	1.20767802	2.39938077	0		-0.350	0.079	0.168	-0.180	
8	H	0.41	-0.07277657	3.3043166	0		0.107	-0.350	0.168	0.109	-0.216
9	**-1.241**										

Coulombic energy spreadsheet for one geometry

Sheet 2: Fitting

The purpose of this sheet is to get approximate values for Epsilon and Sigma in the Lennard-Jones equation that can go into ForceFit. We begin with some original values for Epsilon and Sigma from available research. These values don't need to be exact because they will get changed anyway, but they should be fairly close. If there is no research on the specific project, then values can be used from research on similar things.

Step 1

The first part of the fitting is simply getting the numbers ready to be fitted through conversions.

Column A has the distance between the two water molecules.

Column B is the total energy directly from each output file from Gaussian03 or NWChem.

Column C is the total energy from Column B converted into kcal/mol by multiplying Column B by 627.5. **[C7=B7*627.5]**

Column D is the energy from Column C scaled down. This is done by finding the smallest value in Column C and subtracting that number from the rest of the numbers in Column C (use $ around the letter in the cell name to indicate that that cell will not change as the formula is dragged down the column, for example **[D9=C9-$C$9]**).

Column E is the Coulombic energy. It is each sum at the bottom of Column A on Sheet 1. When copying and pasting this number be sure to specify to paste the value only.

Column F is the Coulombic energy from column E converted to Joules. **[F7=E7*(1.602177E-19)^2*8.987*10^19]**

Column G converts the energy from Joules to kcal/mol. **[G7=F7*0.000239005736*(6.022*10^23)]**

	A	B	C	D	E	F	G
1	**Original**						
2	**Sigma OO**	**3.166**					
3	**Epsilon OO**	**0.1553**					
4							
5							
6	**Distance OO**	**HF (au)**	**Au to kcal/mol**	**Scaled energy**	**Coul. Energy**	**Coul. to J**	**Coul. (J) to kcal/mol**
7	1	-151.2361859	-94900.70665	995.436131	-1.241494259	-2.86405E-18	-412.2204359
8	2	-152.7515673	-95851.60848	44.5343025	-1.223174915	-2.82179E-18	-406.1377595
9	3	-152.8225383	-95896.14278	0	-1.198562466	-2.76501E-18	-397.9655472
10	4	-152.8190605	-95893.96046	2.1823195	-1.191660319	-2.74909E-18	-395.6737881
11	5	-152.8172737	-95892.83925	3.3035365	-1.189119504	-2.74322E-18	-394.8301466
12	6	-152.8167397	-95892.50416	3.6386215	-1.187983911	-2.7406E-18	-394.4530893
13	7	-152.8164801	-95892.34126	3.8015205	-1.187403714	-2.73927E-18	-394.260443
14	8	-152.8163345	-95892.2499	3.8928845	-1.187077294	-2.73851E-18	-394.1520599
15	9	-152.8162466	-95892.19474	3.94804175	-1.186879867	-2.73806E-18	-394.0865071
16	10	-152.8161905	-95892.15954	3.9832445	-1.186753549	-2.73777E-18	-394.0445651
17							
18							

Part I of fitting in Excel

Step 2

At the top of this section of the spreadsheet are the same Sigma and Epsilon values from the last section. These should be copied over because these will be changed through the fitting.

Column H is a calculated Lennard-Jones potential from the original guess. For example the formula for cell H7 would be **[=4*$I$3*(($I$2/A7)^12-($I$2/A7)^6)]**. This version of the Lennard-Jones equation is used for DL_POLY, however another version of the equation may be used with other Molecular Dynamics programs.

Column I is the sum of the Lennard-Jones equation and Coulombic energy from Column G. **[I7=G7+H7]**

Column J is the previous column scaled to the minimum value, done in the same manner as Column D. **[J9=I9-$I$9]**

Column K is the difference between the two scaled columns, Column J and Column D. **[K7=J7-D7]** At the bottom of this column is the square root of the squared sum of this column. **[K18=SQRT (SUMSQ (K7:K16))]**

	H	I	J	K	L
1	**Excel Fit**				
2	**Sigma OO**	**3.166**			
3	**Epsilon OO**	**0.1553**			
4					

5					
6	**LJ**	**LJ+coul**	**Scaled LJ+coul**	**Diff. of scaled**	
7	629406.059	628993.8385	629391.4767	628396.0406	
8	144.0413192	-262.0964403	135.5417534	91.00745095	
9	0.327353517	-397.6381937	0	0	
10	-0.115181622	-395.7889697	1.849224033	-0.333095467	
11	-0.037457806	-394.8676044	2.77058935	-0.53294715	
12	-0.013119352	-394.4662086	3.171985117	-0.466636383	
13	-0.005271996	-394.265715	3.37247874	-0.42904176	
14	-0.002377307	-394.1544372	3.483756485	-0.409128015	
15	-0.001174946	-394.0876821	3.550511649	-0.397530101	
16	-0.00062497	-394.0451901	3.593003671	-0.390240828	
17					
18				**628396.0472**	

Part II of fitting in Excel, before Solver

Step 3

At this point the data can be fitted. The goal is to minimize the square root of the squared sum by changing the values for epsilon and sigma. To fit this data use the Solver utility, which can be downloaded for free from Frontline Solvers. Once the Solver is downloaded and opened, select a target cell. This cell will be K18, or the square root of the squared sum of the differences. Select "Minimize" and choose to change the cells that contain Sigma and Epsilon (I2 and I3). Now solve. After a few seconds it will give a prompt asking if the answer should be kept or rejected and select "yes" to keep the changes.

Finally, *Column L* is column K scaled down after the Solver has been used.

	H	I	J	K	L
1	**Excel Fit**				
2	**Sigma OO**	**2.893007094**			
3	**Epsilon OO**	**0.000735651**			
4					
5					
6	**LJ**	**LJ+coul**	**Scaled LJ+coul**	**Diff. of scaled**	**Scale of fitted K**
7	1009.681955	597.4615188	995.4275294	-0.008601649	52.47747981
8	0.219969963	-405.9177895	-7.951778959	-52.48608146	0
9	-0.000463332	-397.9660106	0	0	52.48608146
10	-0.000360897	-395.674149	2.291861606	0.109542106	52.59562357
11	-0.000106267	-394.8302528	3.135757737	-0.167778763	52.3183027
12	-3.65115E-05	-394.4531258	3.512884806	-0.125736694	52.36034477
13	-1.45905E-05	-394.2604576	3.705552995	-0.095967505	52.39011395
14	-6.56624E-06	-394.1520665	3.813944075	-0.078940425	52.40714103
15	-3.24261E-06	-394.0865104	3.879500202	-0.068541548	52.41753991
16	-1.72415E-06	-394.0445668	3.921443766	-0.061800734	52.42428073
17					
18				**52.48684348**	

Part II of fitting in Excel, after Solver

Necessary Files for MD Interface

ForceFit can interface with three different molecular dynamics programs, DL_POLY, LAMMPS, and TINKER. Each of these programs has specialized features that the others do not have, making one the more effective option for certain simulations. For the complete descriptions of all features for each program, consult their respective User Manuals. These programs will generate the energies and forces for the geometries inputted and read the forces back to ForceFit and the minimization algorithm. ForceFit will automatically change the variables, sigma and epsilon from the Lennard-Jones equation, in the files and the jobs will rerun until a suitable fit for the variables has been reached.

It is suggested that before each of these programs is used with ForceFit, that the programs are run by themselves. The input files should be checked before they are run with ForceFit and this provides an easier way than with ForceFit to do so. The files can be run interactively with the correct executable. When the program runs by itself there will need to be a table of forces or velocities for ForceFit to read. When the input files all work with the program then they can be tried with ForceFit. There are a few changes from regular files for each program and those files for ForceFit. The examples below are ready for ForceFit and will have to be modified slightly to work with the program alone. The most notable change is the variables, as explained below, that will have to be changed to the real numbers that would be entered into ForceFit.

Note: ForceFit is not able to perform a fitting for the previous example of two water molecules because it is too small of a system. However, the concepts in the previous examples are still easily applicable to larger systems. The examples for the remainder of the manual will consist of 8 water molecules. In each geometry, a water molecule is moved half an angstrom away from or toward the group, beginning at an equilibrium of 2.7 angstroms.

DL_POLY

ForceFit needs three different files in order to run with DL_POLY: the CONTROL file, the FIELD file, and the Executable file. Each file and its importance in ForceFit is explained below.

CONTROL File

The CONTROL file is where the calculations to be performed by DL_POLY are specified. The necessary keywords for ForceFit in the CONTROL file are “coul” in order to calculate the Coulombic energy for the fitting, and “steps 1” because only one step should be performed. The same CONTROL file can be used for a wide variety of ForceFit jobs because this file specifies calculations that need to be performed in order to get a fitting from ForceFit. More information on the syntax of and how to construct a CONTROL file can be found the in DL_POLY user’s manual.

```
(H2O)8 - single point for fitting

integrate leapfrog verlet

zero

ensemble         nve

timestep         0.0000010
steps           1
delr width       1.0000
cut              8.000
coul
rvdw cutoff      8.000
 job time         43200.
close time       100.00

finish
```

Sample CONTROL file for DL_POLY

FIELD File

The FIELD file contains all the information about the force fields for the specific potential to be fitted. The FIELD file needs to contain the directive "lj" in the VDW section. The variables in the Lennard-Jones equation can be entered directly into the ForceFit GUI, so instead of entering numbers for the variables, #A# and #B# can be entered. This will refer the variables to input that will be entered later in the ForceFit GUI. More information on how to create a FIELD file can be found in the DL_POLY user's manual. In the example file below, lines that are similar to those above and below it have been removed at the "…" in the interest of space.

```
newffm Generated Original Geometry
 UNITS kcal

 MOLECULES 1
 Water
 NUMMOLS 1
 ATOMS 24
  O      15.9994      -0.8200000     1
  O      15.9994      -0.8200000     1
  ...
  H       1.008        0.4100000     1
  H       1.008        0.4100000     1
  H       1.008        0.4100000     1
  H       1.008        0.4100000     1
...
 BONDS 16
  harm  1       9       1108.2698       1       0       0
  harm  1       10      1108.2698       1       0       0
  harm  2       21      1108.2698       1       0       0
  harm  2       22      1108.2698       1       0       0
...
ANGLES 8
 Harm       9       1       10      91.5392     109.47
 Harm       21      2       22      91.5392     109.47
...
 FINISH
 VDW  1
     O      O      lj      #A#      #B#   0.000000    0.000000  0.000000
 CLOSE
```

Sample FIELD file for DL_POLY (text removed at '…')

Executable File

The Executable is the DL_POLY program that will run the jobs and give results back to ForceFit. This file comes with the DL_POLY package and will not need to be user-modified.

LAMMPS

The LAMMPS package can perform some calculations that DL_POLY cannot. For those calculations in LAMMPS, the files ForceFit needs are the Input file, the Data file, and the Executable file. These have similar functions to the files used in DL_POLY. LAMMPS will do a simulation and ForceFit will use those results to do the fitting for the PES.

Input File

The Input file is where all the directions for a LAMMPS run are stored. There are a few keywords necessary for ForceFit in the Input file for LAMMPS. These are the "pair_style lj/cut/coul/cut" command, "run 1" to run for only one timestep, and "dump" of all forces. However, ForceFit will change the dump command to dump forces into a file called geometry.out when it runs. *Note: The change of the dump command is only in effect when running with ForceFit. For interactive runs, the dump command can to be to any file.* This same Input file can be used for many different fittings because it contains all the commands necessary to give ForceFit the data to perform the fit. More information on the sections and

commands shown below, the syntax, and how to construct an Input file can be found in the LAMMPS User Manual.

```
# 8-waters

units           real
dimension       3
boundary        p p p
atom_style      full

pair_style      lj/cut/coul/cut 8.0
pair_modify     mix arithmetic

bond_style      harmonic
angle_style     harmonic
dihedral_style  none
improper_style  none
kspace_style    none


read_data       water-eliza.data

neighbor        1.0 bin
neigh_modify    check yes

timestep        0.001
run_style       verlet

reset_timestep  0

fix             1 all nve
fix             2 all temp/rescale 1 0.0 0.0 1 0

dump            1 all custom 1 forces.out type x y z
run             1
```

Sample Input File for LAMMPS

Data File

The data file contains specifications about the PES and force-field parameters. This file is specific to each PES to be fitted. The pair, bond, and angle coefficients sections need to be in the Data file. If fitting for a Lennard-Jones potential, the pair coefficients section will contain the variables of sigma and epsilon. #A# and #B# will refer to the variables entered in the ForceFit GUI. More information on the commands and sections below and the syntax and construction of the Data file can be found in the LAMMPS user manual. In the example below, lines of data that are similar to those before and after it, were removed at the "…" in the interest of space.

```
LAMMPS data file for 8water

24 atoms
16 bonds
8 angles
0 dihedrals
0 impropers

2 atom types
1 bond types
1 angle types

-15.00000000    15.00000000     xlo xhi
-15.00000000    15.00000000     ylo yhi
-55.00000000    55.00000000     zlo zhi

Masses

1 16.00000
2 1.008000

Pair Coeffs

1 #A# #B#
2 0.0000000 0.00000000

Bond Coeffs

1 554.1349 1.0

Angle Coeffs

1 45.7696 104.47

Atoms

     1  1  1 -0.830000    -1.891805     0.464192     1.664909
     2  1  2  0.415000    -1.466638    -0.423623     1.804831
     3  1  2  0.415000    -2.096820     0.467729     0.712694
  ...
    22  1  1 -0.830000     1.857444    -0.450170    -1.249729
    23  1  2  0.415000    -0.762119    -2.615179    -1.756647
    24  1  2  0.415000    -1.091783    -1.141985    -1.312080

Bonds

     1     1     1     2
     2     1     1     3
  ...
    15     1    22    23
    16     1    22    24

 Angles

     1     1     2     1     3
   ...
     8     1    23    22    24
```

Sample Data File for LAMMPS (text removed at '…')

Executable File

The Executable file is the LAMMPS program that will run the jobs and give results back to ForceFit. This file comes with the LAMMPS package and will not need to be user modified.

TINKER

Because each MD package is specialized to do certain types of simulations, TINKER can be used for some simulations that DL_POLY or LAMMPS cannot do. The necessary files for TINKER are the XYZ file template, the PRM file, the Key file, and the Executable file.

XYZ File Template

The XYZ file is a file containing coordinates and bonds. There are no special keywords associated with this file that ForceFit will need. More information on how to construct an XYZ file can be found in the TINKER User's Guide. In the example below lines that were similar to those above and below were removed at the "…" in the interest of space.

```
24 Water
1  O    -1.891805        0.464192        1.664909  1  2  3
2  H     0.443003        1.899956        1.681447  2  1
3  H    -0.446238       -1.863644        1.736293  2  1
4  O     1.891808       -0.427194        1.711225  1  5  6
5  H    -0.439146       -1.881612       -1.213695  2  4
6  H    -1.853219        0.424415       -1.287902  2  4
...
22 O    -0.450874        1.478938        1.791561  1  23 24
23 H    -0.762119       -2.615179       -1.756647  2  22
24 H    -1.091783       -1.141985       -1.31208   2  22
```

Sample XYZ File for TINKER (text removed at '…')

PRM File

The Parameter file specifies the potential energy function and individual energy parameters for a particular force field. A parameter file is not strictly necessary to run ForceFit, depending on the types of calculations. This particular example of 8 water molecules does not require a parameter file, as is specified by "parameter none" in the key file. The parameter files come with the TINKER package and do not need to be user-modified for the most part. The user should read the documentation on each force field and select the appropriate one for the fitting in ForceFit.

Key File

The Key file is where all the commands and specifications are contained. ForceFit will need the "VDWTYPE Lennard-Jones" and "RADIUSTYPE Sigma" if the 4ε version of the LJ equation was used, as was the case in the sample Excel spreadsheets. The command "save-force" must also be used. Many of the numbers in the key file are related to specific parameter files and force fields, but as no parameter file was used in this simulation, the values are set to 1 and 2 in this example. More information about the commands used below and how to construct a Key file can be found in the TINKER User's Guide.

```
parameters      none
verbose

group-molecule

vdwtype                 LENNARD-JONES
radiusrule              GEOMETRIC
radiustype              SIGMA
radiussize              DIAMETER
epsilonrule             GEOMETRIC
dielectric              1.0
integrate               VERLET
thermostat              BERENDSEN
stepmax                 1.0

atom    1       O       "SPC Water O"     8     15.995      2
atom    2       H       "SPC Water H"     1     1.008       1

charge  2001    -0.820
charge  2002     0.410

vdw     1               #A#       #B#
vdw     2               0.0000    0.0000

bond    1         2             1000.000        1.000
angle   2         1     2       100.000         109.47

save-force
```

Sample Key file for TINKER

Executable File

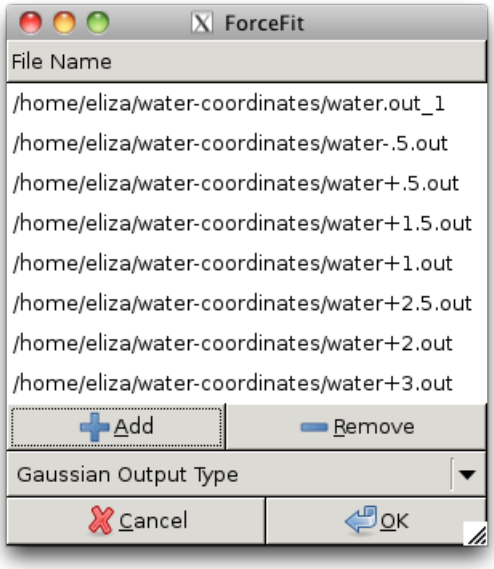
In TINKER there are many executables for various simulations, but the one that will be used for ForceFit is "dynamic." This executable comes with the TINKER package and will not need to be user modified.

The ForceFit GUI: Fitting and Output

At this point the geometry files from Gaussian or NWChem, initial variable guesses, and MD files for DL_POLY, LAMMPS, or TINKER are all in order and the potential energy can now be fitted using ForceFit. Open ForceFit in Unix using the command ForceFit.

Loading the PES Database Files

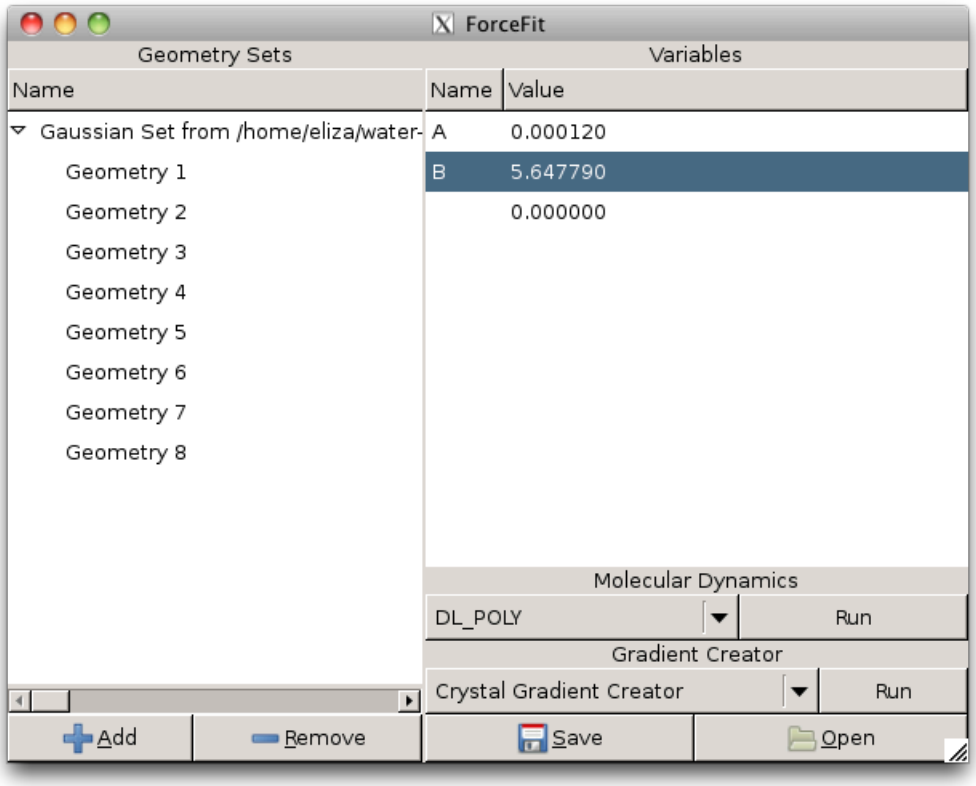
Once ForceFit is open, the first step is to load all of the geometries made from NWChem or Gaussian. To do so, click on "Add" in the bottom left corner of the ForceFit GUI. A window such as the one below will appear to enter geometries into. Before entering geometries, select the type of output file that will be entered, NWChem or Gaussian, from the drop-down menu at the bottom of the screen. Click on "Add" to browse directories to find the appropriate output files. Once all geometries are loaded click "OK." The files can be added one at a time, but they may also be added as one long file. All of the output files may be concatenated before opening ForceFit to add as one file.



The add geometries window

Depending on the output type selected, one of two things will happen. If NWChem output files were entered, the ForceFit GUI will return and you may proceed to the next step. However, if Gaussian output files were entered, a prompt on the Unix terminal will appear asking for a step size. This is a glitch in ForceFit that can easily be worked around. Enter any number for the step size. It will then ask for modes to be listed for that geometry. Do not enter a number, simply press Enter again. This same prompt will appear for each geometry that was loaded and the same steps may be taken again. Once this has been done for all geometries, return to the ForceFit screen.

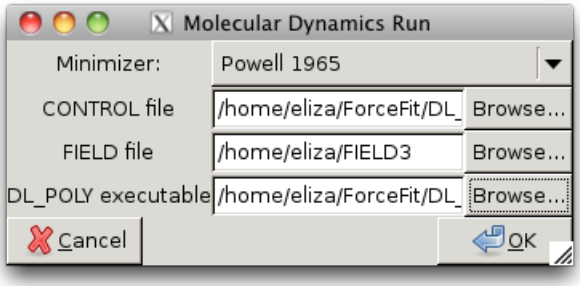
Now it is time for the variables to be entered. On the right side of the window there will be zeros. Click on the zeros to enter the variables. Click away and another row of zeros will appear for the next variable to be entered. The variables come from the Excel spreadsheet. Enter the second set of Epsilon and Sigma values or the values that the solver was applied to. Enter Epsilon as variable “A” and Sigma as variable “B.”



Add variables in the ForceFit GUI

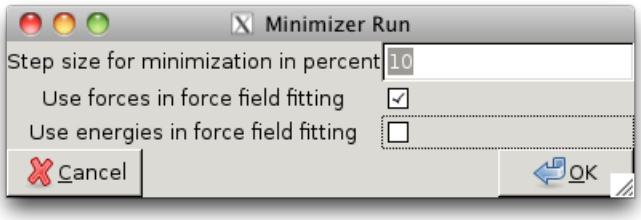
Choosing MD Code and Template Files

The next step is to choose which MD program to run with ForceFit. Choose one from the drop-down menu at the bottom right of the screen under Molecular Dynamics. A new window will appear where the necessary files can be put in. If using DL_POLY or LAMMPS all the files must be entered. However, if TINKER is being used, the PRM file may not apply to all calculations, so it can simply be left blank.



Entering necessary files for MD run

After all of the necessary files have been entered, another window will appear about minimization. The user may choose whether to use forces and/or energies as part of the fitting process. In this example the energies were not used. The default step size for minimization in percent is 10, which will work for most simulations.



Specifications for Minimization Run

Output

The fitting will run on the Unix terminal. When it stops running (it will take several seconds) two variables will appear in the last line on the screen. These are the sigma and epsilon variables from the spreadsheet. The "U" value next to the variables should be very small if the simulation was accurate. Because the example simulation was so small, the U value is quite large.

```
dl_polyRun/set_1_geometry_1
dl_polyRun/set_1_geometry_2
dl_polyRun/set_1_geometry_3
dl_polyRun/set_1_geometry_4
dl_polyRun/set_1_geometry_5
dl_polyRun/set_1_geometry_6
dl_polyRun/set_1_geometry_7
dl_polyRun/set_1_geometry_8
RMSD: 37.5784
X: 22.0486 Y: 21.6228 Z: 21.4115
dl_polyRun/set_1_geometry_1
dl_polyRun/set_1_geometry_2
dl_polyRun/set_1_geometry_3
dl_polyRun/set_1_geometry_4
dl_polyRun/set_1_geometry_5
dl_polyRun/set_1_geometry_6
dl_polyRun/set_1_geometry_7
dl_polyRun/set_1_geometry_8
RMSD: 37.5803
X: 22.0505 Y: 21.6241 Z: 21.4115
dl_polyRun/set_1_geometry_1
dl_polyRun/set_1_geometry_2
dl_polyRun/set_1_geometry_3
dl_polyRun/set_1_geometry_4
dl_polyRun/set_1_geometry_5
dl_polyRun/set_1_geometry_6
dl_polyRun/set_1_geometry_7
dl_polyRun/set_1_geometry_8
RMSD: 37.581
X: 22.0513 Y: 21.6245 Z: 21.4115
U: 8.51043e+06 Variables: 0.000122978, 5.63693
```

ForceFit Output

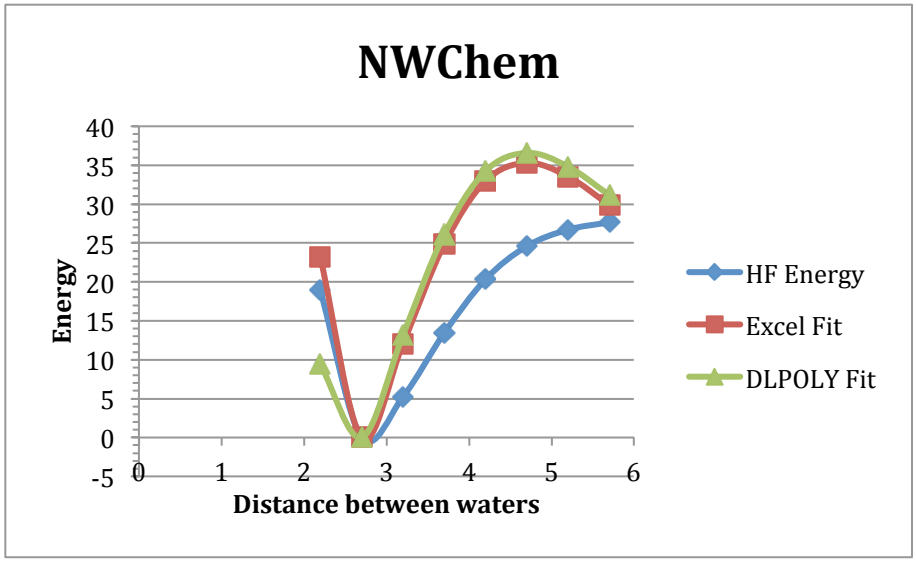
Excel

These two variables can be put back into Excel, starting a few new columns. The final calculations on these variables will be to calculate the Lennard-Jones equation with these two variables, add the Coulombic energy, and scale the sums. For more details on these steps, refer to the Generating Variable Guesses section of this manual, *Columns H-J*. It should be noted that the example spreadsheet and graph below are not extensions of the excel sheet found in the Generating Variable Guesses section of this manual. The numbers were derived from the second example of 8 water molecules.

	M	N	O
1	**Force Fit-DL_POLY**		
2	**Sigma**	**5.64779**	
3	**Epsilon**	**0.000120032**	
4			
5			
6	**LJ of DLP fit**	**DLP LJ + coul.**	**DLP LJ+coul scaled**
7	39.20193161	-2277.264138	14.98724127
8	3.329038262	-2292.251379	0
9	0.42411942	-2279.529197	12.72218157
10	0.070747595	-2266.650793	25.60058548
11	0.0139457	-2258.51822	33.73315894
12	0.002906787	-2256.195204	36.05617518
13	0.000505628	-2257.948444	34.3029345
14	-2.44046E-05	-2261.613917	30.63746216

After ForceFit in Excel

Finally, the original scaled energy (*Column D*), scaled LJ equation from the Excel fit (*Column J*), and the scaled LJ equation from ForceFit (*Column O*) can all be graphed over the distances (*Column A*). The graph should be a "Smooth Marked X Y Scatter Plot" and all three energies can be graphed on the same chart to easily compare them. The shapes of all three lines should be similar.



A graph of the three different energies in Excel

Errors

When an error message appears it will refer to a directory of files that are created during the run. The directories will be located in the same directory with the ForceFit program and be called dl_polyRun/, lammpsRun/, or tinkerRun/, depending on which MD program was used. Upon a successful fitting, there will be separate subdirectories for each geometry entered inside the appropriate directory. Inside any one of these subdirectories there will be the files entered and other files that were created during the fitting, such as geometry.out if LAMMPS was used.

If an error message directed you to one of these directories, open the first subdirectory (set_1_geometry_1) because ForceFit may not have fitted with the later geometries if an error occurred early on. The files are re-written for each run with ForceFit, so the later geometries may be from previous runs. In the output file inside this directory there will probably be an error message directing you to what may have gone wrong in the fitting.

Appendix

The example used throughout the User Manual changed halfway from two waters to eight waters. The following is from the second example of 8 water molecules that was not shown the section Generating Variable Guesses. To clarify the examples, the spreadsheets from Excel are contained in the appendix with the formulas used in each cell.

This table is the first five columns of the Coulombic Energy spreadsheet for the second example of 8 water molecules. As the formulas are simply dragged down to the rest of the column and not different for each cell in a column, the bottom of the columns are not shown.

distance separation	d from equilibrium			
2.2	-0.5			
O	-0.82	-1.891805	0.464192	1.664909
O	-0.82	0.443003	1.899956	1.681447
O	-0.82	-0.446238	-1.863644	1.736293
O	-0.82	1.891808	-0.427194	1.711225
O	-0.82	-0.439146	-1.881612	-1.213695
O	-0.82	-1.853219	0.424415	-1.287902
O	-0.82	0.435408	1.855324	-1.287178
O	-0.82	1.427718	-0.718014	-1.242987
H	0.41	-1.466638	-0.423623	1.804831
H	0.41	-2.09682	0.467729	0.712694

This table contains the next two columns of the Coulombic Energy spreadsheet in formula. Columns such as these two would be repeated until there is only one row in the column left to be calculated. The bottom of the columns are not shown.

=B3*B4/(SQRT((C4-C3)^2+(D4-D3)^2+(E4-E3)^2))	
=B4*B5/(SQRT((C5-C4)^2+(D5-D4)^2+(E5-E4)^2))	=B3+B5/(SQRT((C5-C3)^2+(D5-D3)^2+(E5-E3)^2))
=B5*B6/(SQRT((C6-C5)^2+(D6-D5)^2+(E6-E5)^2))	=B4+B6/(SQRT((C6-C4)^2+(D6-D4)^2+(E6-E4)^2))
=B6*B7/(SQRT((C7-C6)^2+(D7-D6)^2+(E7-E6)^2))	=B5+B7/(SQRT((C7-C5)^2+(D7-D5)^2+(E7-E5)^2))
=B7*B8/(SQRT((C8-C7)^2+(D8-D7)^2+(E8-E7)^2))	=B6+B8/(SQRT((C8-C6)^2+(D8-D6)^2+(E8-E6)^2))
=B12*B13/(SQRT((C13-C12)^2+(D13-D12)^2+(E13-E12)^2))	=B11+B13/(SQRT((C13-C11)^2+(D13-D11)^2+(E13-E11)^2))

This is the first section of the fitting spreadsheet from Excel in formula view. This is taken from the second example of 8 waters.

	A	B	C	D	E	F

1	Original					
2	Sigma OO	3.166				
3	Epsilon OO	0.1553				
4						
5	Gaussian					
6	Distance	HF	kcal/mol	scaled energy	coul energy	coul to J
7	2.2	-611.3956638	=B7*627.5	=C7-$C$8	-6.97655689869441	=E7*(1.602177E-19)^2*8.987*10^19
8	2.7	-611.4257993	=B8*627.5	=C8-$C$8	-6.91365507508133	=E8*(1.602177E-19)^2*8.987*10^19
9	3.2	-611.417456	=B9*627.5	=C9-$C$8	-6.86659055874346	=E9*(1.602177E-19)^2*8.987*10^19
10	3.7	-611.4043829	=B10*627.5	=C10-$C$8	-6.82674009965802	=E10*(1.602177E-19)^2*8.987*10^1
11	4.2	-611.3933845	=B11*627.5	=C11-$C$8	-6.80207595978868	=E11*(1.602177E-19)^2*8.987*10^1
12	4.7	-611.3866119	=B12*627.5	=C12-$C$8	-6.79504642943292	=E12*(1.602177E-19)^2*8.987*10^1
13	5.2	-611.3832289	=B13*627.5	=C13-$C$8	-6.80031947516649	=E13*(1.602177E-19)^2*8.987*10^1
14	5.7	-611.3817235	=B14*627.5	=C14-$C$8	-6.81135727058503	=E14*(1.602177E-19)^2*8.987*10^1

G
to kcal/mol
=F7*0.000239005736*(6.022*10^23)
=F8*0.000239005736*(6.022*10^23)
=F9*0.000239005736*(6.022*10^23)
=F10*0.000239005736*(6.022*10^23)
=F11*0.000239005736*(6.022*10^23)
=F12*0.000239005736*(6.022*10^23)
=F13*0.000239005736*(6.022*10^23)
=F14*0.000239005736*(6.022*10^23)

This is the extension of the previous spreadsheet. This is where the variables are fitted. The scale of K is not calculated due to redundancy of results.

	H	I	J	K	L
1	Excel Fit				
2	Sigma OO	5.64778552728727			
3	Epsilon OO	0.000147530552235198			
4					
5					
6	LJ	LJ+coul	scaled	diff of scaled	scale of K
7	=4*$I$3*(($I$2/A7)^12-($I$2/A7)^6)	=H7+G7	=I7-$I$8	=J7-D7	
8	=4*$I$3*(($I$2/A8)^12-($I$2/A8)^6)	=H8+G8	=I8-$I$8	=J8-D8	
9	=4*$I$3*(($I$2/A9)^12-($I$2/A9)^6)	=H9+G9	=I9-$I$8	=J9-D9	
10	=4*$I$3*(($I$2/A10)^12-($I$2/A10)^6)	=H10+G10	=I10-$I$8	=J10-D10	

11	=4*$I$3*(($I$2/A11)^12-($I$2/A11)^6)	=H11+G11	=I11-$I$8	=J11-D11	
12	=4*$I$3*(($I$2/A12)^12-($I$2/A12)^6)	=H12+G12	=I12-$I$8	=J12-D12	
13	=4*$I$3*(($I$2/A13)^12-($I$2/A13)^6)	=H13+G13	=I13-$I$8	=J13-D13	
14	=4*$I$3*(($I$2/A14)^12-($I$2/A14)^6)	=H14+G14	=I14-$I$8	=J14-D14	
15					
16				=SQRT(SUMSQ(K7:K14))	

These results are from fitting with ForceFit. This example is from DL_POLY, but all the different MD programs will give results that can be treated in the same way.

	M	N	O
1	Force Fit-DL_POLY		
2	sigma	5.64779	
3	epsilon	0.000120032	
4			
5			
6	LJ of DLP fit	DLP LJ + coul.	DLP LJ+coul scaled
7	=4*$N$3*(($N$2/A7)^12-($N$2/A7)^6)	=M7+G7	=N7-$N$8
8	=4*$N$3*(($N$2/A8)^12-($N$2/A8)^6)	=M8+G8	=N8-$N$8
9	=4*$N$3*(($N$2/A9)^12-($N$2/A9)^6)	=M9+G9	=N9-$N$8
10	=4*$N$3*(($N$2/A10)^12-($N$2/A10)^6)	=M10+G10	=N10-$N$8
11	=4*$N$3*(($N$2/A11)^12-($N$2/A11)^6)	=M11+G11	=N11-$N$8
12	=4*$N$3*(($N$2/A12)^12-($N$2/A12)^6)	=M12+G12	=N12-$N$8
13	=4*$N$3*(($N$2/A13)^12-($N$2/A13)^6)	=M13+G13	=N13-$N$8
14	=4*$N$3*(($N$2/A14)^12-($N$2/A14)^6)	=M14+G14	=N14-$N$8