PARAMFIT:
A program for automated forcefield parameter generation using a genetic algorithm

Robin Betz and Ross C. Walker

UC San Diego, San Diego Supercomputer Center

INTRODUCTION

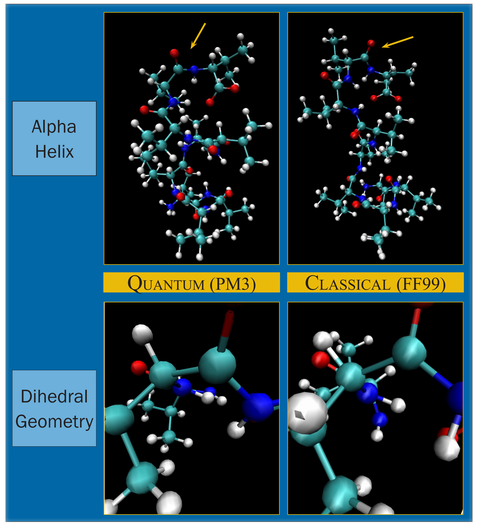
Classical molecular dynamics simulations rely on a forcefield that characterizes each type of bond, angle, and dihedral in the system. Without accurate forcefield parameters, simulations will predict incorrect results that do not match experimental data. Furthermore, when existing forcefields fail to provide an accurate description, the systematic improvement of the parameters is difficult.

Although AMBER's antechamber program can be used to assign atom types to systems and provide estimates for missing parameters, it often cannot provide even an estimation for missing data.

```
ca-ca-ca-ha    1    1.10        180.0          2.        calculated using empirical approach
ca-ca-ca-n     1    1.1         180.0          2.        general improper torsional angle
ct-c -n -h     1    0.00          0.0          2.        ATTN, need revision
X -C -CA-X     4   14.50        180.0          2.        intrpol.bsd.on C6H6
```

The lack of good parameters is especially problematic for protein or other simulations that rely on accurate peptide backbone dynamics. NMR calculations can be performed using molecular dynamics data, and can give incorrect ^{15}N shifts when lacking parameters. The equilibrium dynamics of a simulated system can also be used to make spectroscopic predictions, which will be similarly inaccurate without the correct parameters.

The optimized structure of the alpha helix demonstrates some of the shortcomings of a purely classical model. Alpha helices are an element of the secondary structure of a protein, which plays a large role in determining how it folds. The accurate modeling of these helices in crucial when conducting protein simulations. When comparing the two optimized structures below, differences in N-H geometry become especially evident.

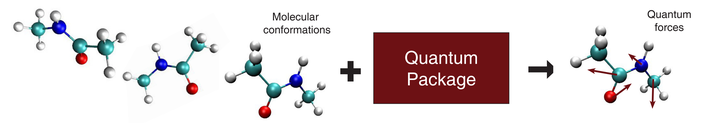


Ab-initio quantum methods provide much greater accuracy, but at great computational expense, rendering simulations unfeasible for all but the smallest of molecules. The greater accuracy of quantum methods can however be used to improve the accuracy of classical ones. When evaluated over a number of conformations of the desired molecule, the energy gradient and derivatives between quantum and classical calculations should be nearly identical when the classical method has correct parameters. A minimization algorithm can be used to fit forcefield parameters to quantum data, resulting in considerable improvements in simulation accuracy.

ALGORITHM

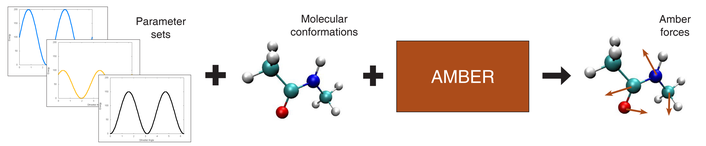
Quantum Calculations

A variety of conformations of the system are inputted into a quantum package to obtain values for the force on each atom using *ab-initio* methods.



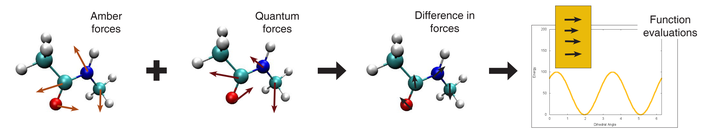
MD Calculations

Possible parameter sets are initially created randomly. The forces on each atom are calculated using the Amber molecular dynamics package with the given parameters as inputs to the Amber equation.



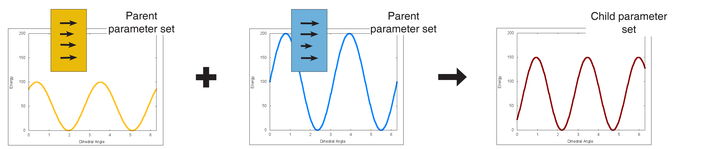
Function Evaluation

The difference between the MD and quantum methods is summed for each atom involved with the parameters to be optimized. The sum of the magnitudes of these vector differences is the function value.



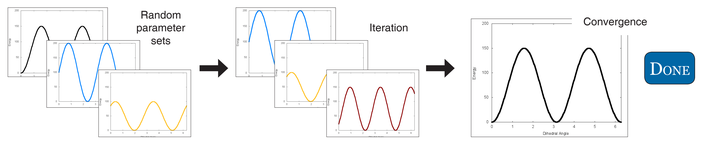
Genetic Algorithm

The genetic algorithm creates a large numbe of random parameter sets and recombines and mutates them in ways analagous to evolution, where fitter parameter sets have a lower function value.



Iteration

The genetic algorithm progresses until there is no improvement after many generations, after which it is considered converged, and the best set of parameters found is returned.

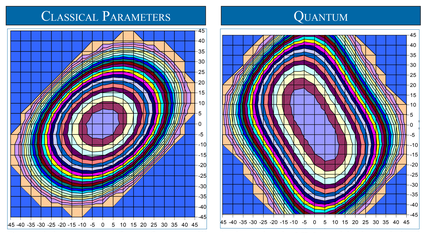


METHODOLOGY

Parameter fitting may be accomplished to a limited degree by using quantum programs to equilibrate the system and measuring parameters, but the generated values will not necessarily produce accurate results for non-equilibrium states. Furthermore, conducting quantum simulations is only feasible for small molecules, and is of little use for conducting protein or lipid simulations. Obtaining parameters experimentally is similarly unfeasible.

If the parameters of the classical forcefield are correct, the differences between the force on each atom obtained with classical and quantum methods should be near zero. In the equation, bond, angle, and dihedral energies are evaluated using the AMBER equation, which incorporates the parameters. K represents the intrinsic difference between the quantum and classical energies, and disappears when fitting forces, as they are the derivative of the equation.

$$f = \sum_{1}^{N} \left(E_{bonds}(n) + E_{angles}(n) + E_{dihedrals}(n) + E_{nb}(n) - E_{QM}(n) + K \right)^2$$



Potential energy surfaces for two dihedral angles in NMA with the original AMBER FF99 parameters and MP2 6-31++G** quantum evaluations.

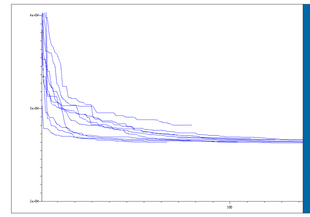
RESULTS

The parameter fitting algorithm was applied to n-methylacetamide (NMA), and the fitted parameters were used to produce potential energy surfaces that were compared with both the default AMBER forcefield and one obtained using a quantum package.

NMA was selected because it demonstrates the N-H out of plane bending at peptide group formations that is problematic to represent using the default parameters. In addition, it is a small molecule that can be simulated using quantum methods in a reasonable amount of time.

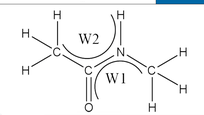
The potential energy surface between two dihedrals, shown as W1 and W2 on the figure, was investigated using all three methods: default parameters, quantum simulation, and the algorithm's fitted parameters. Input conformations representing a scan of each angle at five degree increments were inputted to the appropriate program, and the energy of the structure was plotted.

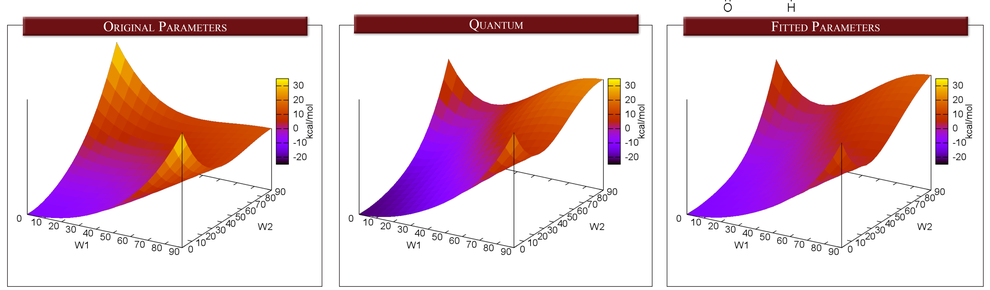
The quantum data were obtained using Gaussian with the MP2 functional and the 6-31++G** basis set, and the AMBER FF99SB forcefield parameters were used to illustrate the default classical parameters. The graphs show the relative differences in energy of each structure, with the structure W1=W2=45° defined as zero.



The genetic algorithm was very effective in locating the region of the search space with the correct parameters.

The algorithm's function evaluation reaches the same minimum regardless of starting point.





CONCLUSION

Generation of parameters for classical molecular dynamics simulations is difficult due to the many coupled degrees of freedom inherent in the problem. However, minimization of the energy derivatives between evaluations with test parameters and quantum data for a variety of input structures is an effective means of improving parameters. Implementing a genetic algorithm to accomplish the minimization resulted in a robust program that can fit many parameters at once with efficiency.

The algorithm handles the large, multi-dimensional search space with ease, provided that there are enough input conformations to provide an adequate sampling of the parameter space. Fitting of bond, angle, and dihedral parameters for a system may therefore be accomplished all at once, instead of using an incremental approach and attempting to only change one parameter at a time, and runs in a reasonable amount of time.

Parameterization of molecules using the algorithm follows a simple five-step process. The algorithm runs efficiently on a desktop computer, and has been parallelized to run on multi-core or multiple processors. High performance computing resources may be used for the quantum energy calculation and the final simulation.

1. Generate Input Conformations
2. Calculate Quantum Energies
3. Run Fitting Algorithm
4. Modify Classical Forcefield
5. Run Simulation

Thank you to Prof. Ross Walker for several graphics, the original parameter fitting code and many helpful explanations. Thanks also to Dr. Andreas Goetz and Dr. Romelia Salomon-Ferrer for clarifications and background information, and also to Ange Mason and the San Diego Supercomputer Center.

"A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules." Wendy D. Cornell, Piotr Cieplak, Christopher I. Bayly, Ian R. Gould, Kenneth M. Merz, David M. Ferguson, David C. Spellmeyer, Thomas Fox, James W. Caldwell, Peter A. Kollman, Journal of the American Chemical Society 1995 117 (19), 5179-5197