A systematic force field optimization method

Lee-Ping Wang, Troy Van Voorhis†, Vijay S. Pande and Todd J. Martinez

Department of Chemistry, Stanford University, Stanford CA 94305; †Department of Chemistry, MIT, Cambridge MA 02139.

Introduction

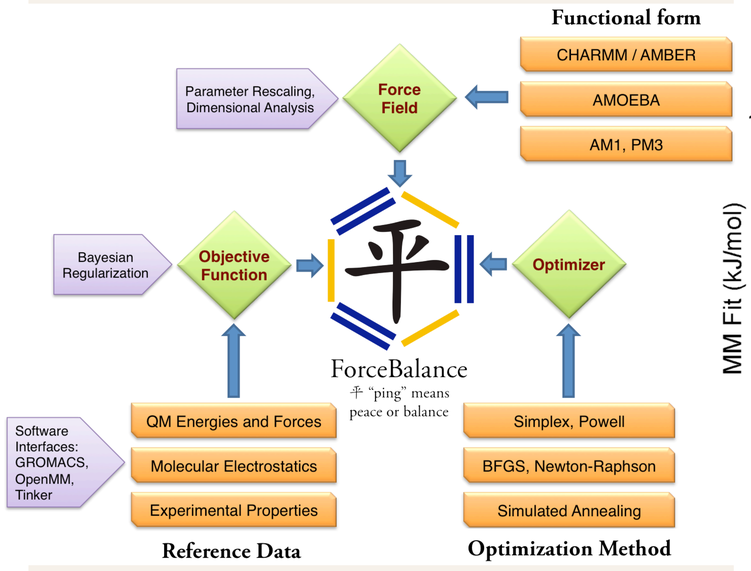
The key to a realistic, accurate, and simple simulation of any molecular system is to answer the following critical question:

What are the most essential physical interactions?

In molecular mechanics (MM) simulations, the physics comes from the empirical potential energy function (*force field*). Force fields are often used out of the box, but often we need to examine their basic approximations and assumptions (e.g. if a simulation is giving an unexpected result.)

Toward this end we developed **ForceBalance**, a systematic and highly general force field optimization method. It is highly flexible regarding the three main components of force field optimization: the *functional form*, the *optimization method*, and the *reference data.*

As a demonstration of the capabilities of ForceBalance we present the automatic parameterization of a polarizable water model using quantum mechanical (QM) reference data, and test the model by simulating some experimental bulk properties.



Method

$$\mathbf{M}(\mathbf{r}_s,\vec{k}) \equiv \begin{pmatrix} E_{\mathrm{MM}} \\ \mathbf{F}_{\mathrm{MM}} \\ \mathrm{etc...} \end{pmatrix} \qquad \chi^2(\mathbf{r}_s,\vec{k}) = \int d^3\mathbf{r}\, P(\mathbf{r}) \left| \mathbf{M}(\mathbf{r},\vec{k}) - \mathbf{Q}(\mathbf{r}) \right|^2 \equiv \left\langle \left| \mathbf{X}(\mathbf{r},\vec{k}) \right|^2 \right\rangle$$

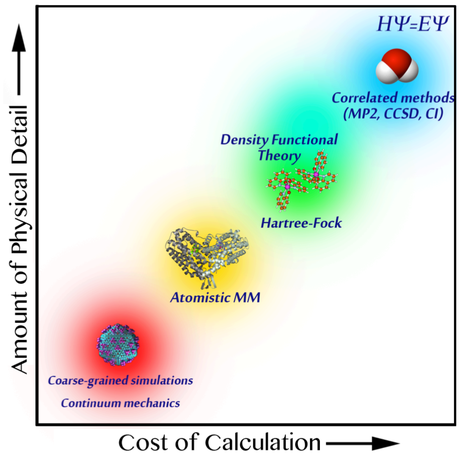
The force field is parameterized by minimizing an *objective function* in the parameter space. The objective function contains the force field errors with respect to theoretical or experimental reference data. For example, we may evaluate the errors in atomistic forces compared to to high-level QM calculations or reproduce an experimentally known structure.

In constructing and minimizing our objective function we address some well-known issues:

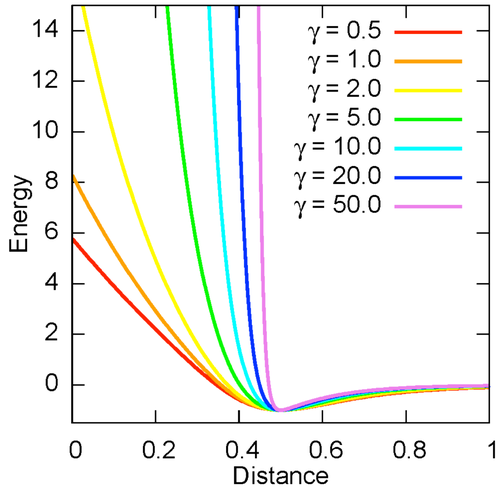
- Sample from both the reference (QM) and fitting (MM) Boltzmann distributions when performing energy and force matching.
- Prevent overfitting by applying an L2-norm penalty function.
- Use dimensional analysis to combine different physical quantities.

The robustness of the final model is improved by increasing the size and diversity of the reference data set.

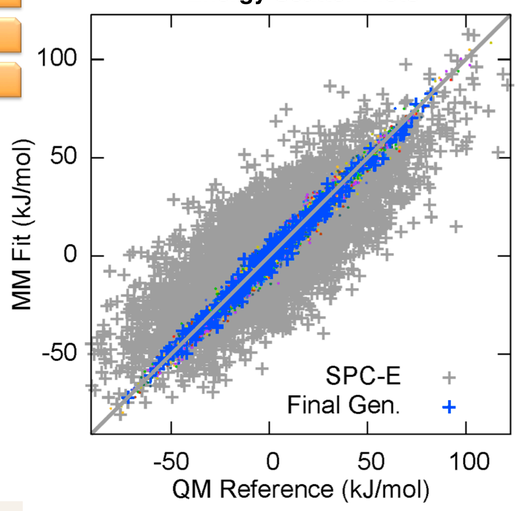
Survey of theoretical methods



Improved van der Waals potential



Energy Scatter Plots



3-D Plot of Atomistic Forces

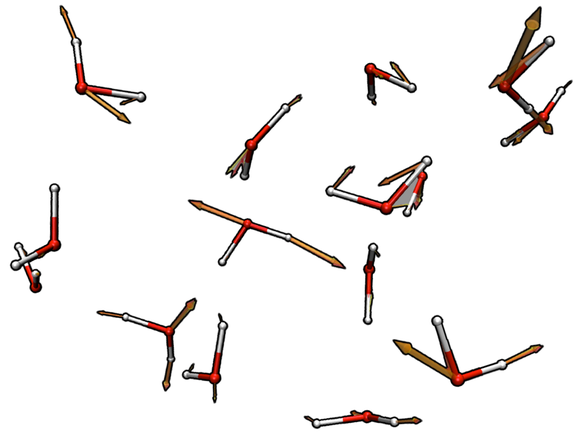


Figure 1: Upper left: Comparison of simulation methods. Upper right: New buffered exp-6 interaction. Lower left: Scatter-plot of QM vs. MM energies for SPC/E (gray) and the ForceBalance-generated model (blue). Lower right: Visualization of QM forces (yellow) and MM forces (red) illustrating quality of fit.

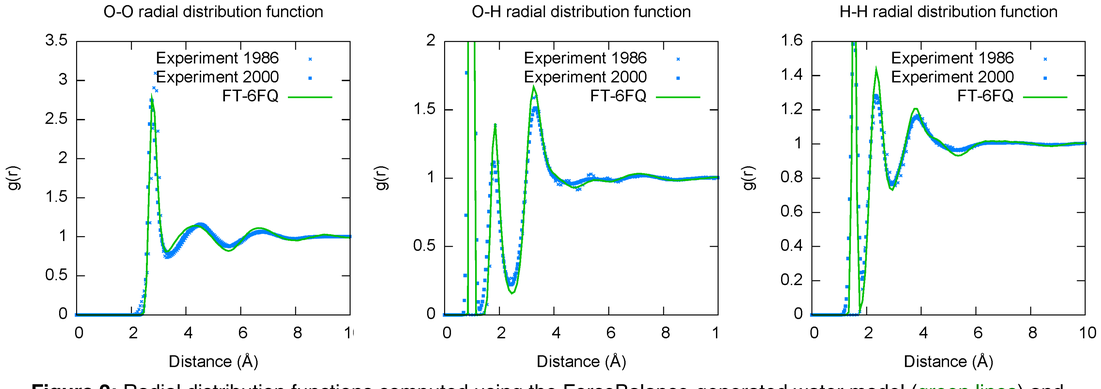


Figure 2: Radial distribution functions computed using the ForceBalance-generated water model (green lines) and comparison to two experimental measurements by Soper (blue dots). The agreement is generally good but some overstructuring is present, possibly due to the neglect of nuclear quantum effects.

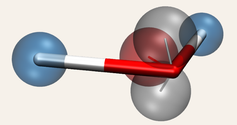
Application: Polarizable Water Model

To demonstrate the capabilities of ForceBalance, we developed a fluctuating-charge model of liquid water. The functional form of the 29-parameter force field is given by:

$$E = \sum_{(\mathrm{i,j})\in \mathrm{Bonds}} D_{ij}\left(1-e^{-a_{ij}\left(r-r_{ij}^0\right)}\right)^2 + \sum_{(\mathrm{i,j,k})\in \mathrm{Angles}} \left(k_{ijk}^{\theta}\left(\theta_{ijk}-\theta_{ijk}^0\right)^2 + k_{ijk}^{r}\left(r_{ik}-r_{ik}^0\right)^2\right) +$$

$$\min_{\{q_i\}}\left[\sum_i \left(\bar{\chi}_i q_i + \eta_i q_i^2\right) + \sum_{i<j} q_i q_j J_{ij}\right] + \sum_{\mathrm{mol}_i \neq \mathrm{mol}_j} \frac{2\varepsilon}{1-\frac{3}{\gamma+3}} \frac{\sigma^6}{\sigma^6+r^6}\left(\frac{3}{\gamma+3}e^{\gamma\left(1-\frac{r}{\sigma}\right)}-1\right)$$

The terms are: Morse bonds, Urey-Bradley angles, QTPIE (fluctuating charge) on five interaction sites, and a new buffered exp-6 interaction. The positions of virtual sites were optimized as well. The new exp-6 interaction is improved from the Buckingham interaction; it has no singularity at r=0.



Left: Illustration of polarizable water model. Fluctuating-charge sites (transparent spheres) colored to reflect charge at E=0. The out-of-plane site positions have been optimized; note their unusual positions.

The reference data contains:

- Energies and atomistic forces computed at the MP2 / aug-cc-pVTZ level for water clusters of four sizes (6, 9, 12, and 15) sampled using the force field. The sampling was self-consistently updated until convergence.
- Multipole moments up through second order and the dipole polarizability tensor of the water molecule were also included.

Property	Calculated	Experiment
Density (kg m^{-3})	1040 ± 10	1000
Dielectric constant	85 ± 10	78.4
Diffusion constant (10^{-5} cm^2 s-1)	1.5 ± 0.2	2.3
Density maximum (ºC)	<6ºC	4ºC
Solution phase dipole moment (D)	2.6	2.3 – 2.9

Table 1: Bulk properties of the ForceBalance-generated water model.

Discussion

The properties of water were simulated using a 512-molecule cubic box. At each step the fluctuating charges were solved using iterative conjugate gradients. Over 5.0 ns of dynamics were generated in total.

Radial distribution functions and various other comparisons to experiment are given in Figure 2 and Table 1. The performance is comparable to the newest parameterizations of TIP4P and many polarizable water models (e.g. TIP4P-FQ, AMOEBA, SWM4-DP, TTM2-F). Our results are notable because *none of the experimental properties were fitted directly*; only computed atomistic properties were considered and the parameterization procedure was fully automatic. This project demonstrates the capabilities of ForceBalance as a highly general methodology for force field creation.

Future applications involve the application of ForceBalance to examine the approximations made in protein force fields, especially for small peptide structures and protein-ligand binding free energies.

Resources

ForceBalance is open-source and available for free on the web at **https://simtk.org/home/forcebalance** .