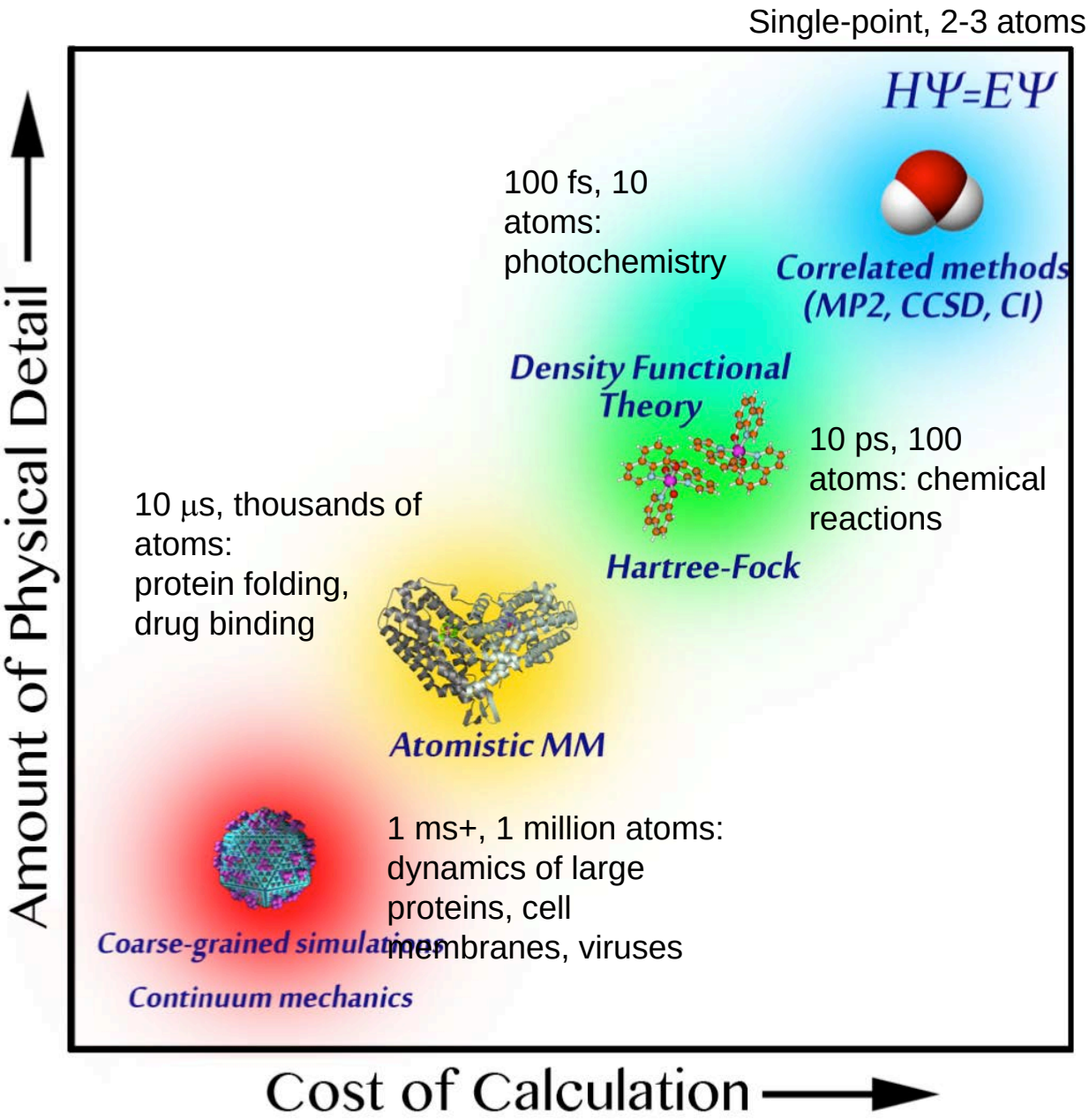
Systematic force field optimization for more accurate molecular simulations

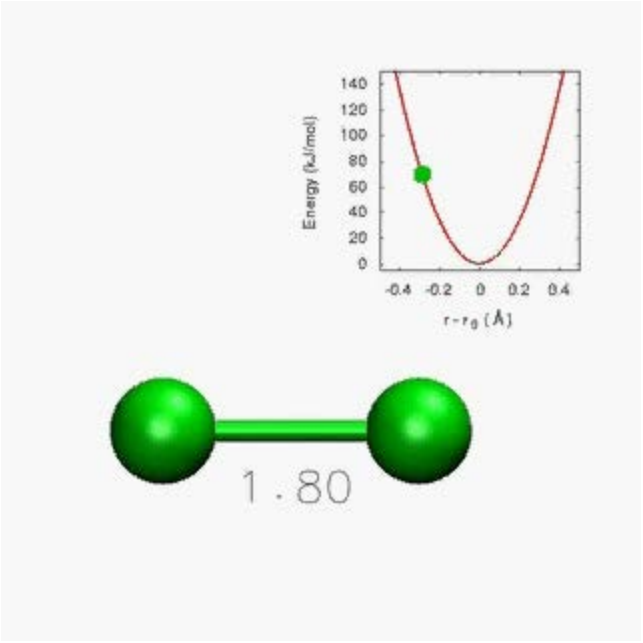
Lee-Ping Wang
Stanford Department of Chemistry
OpenMM Workshop, Stanford University
March 2013

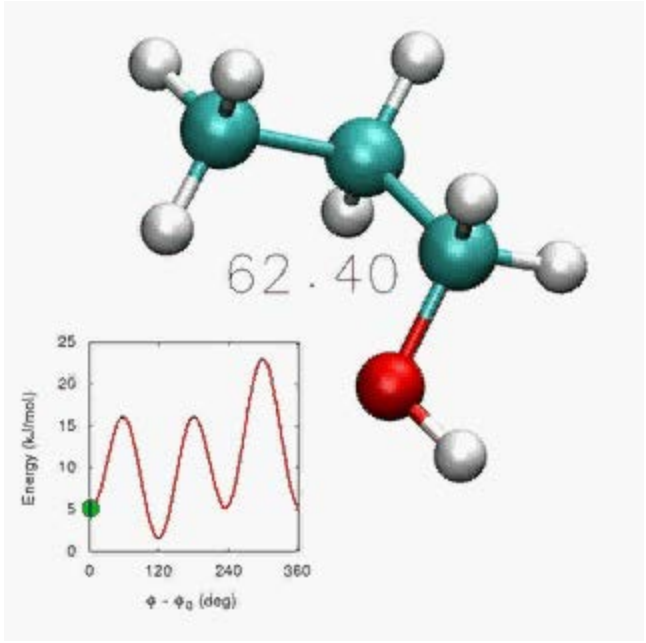
Introduction: A wide range of simulation domains

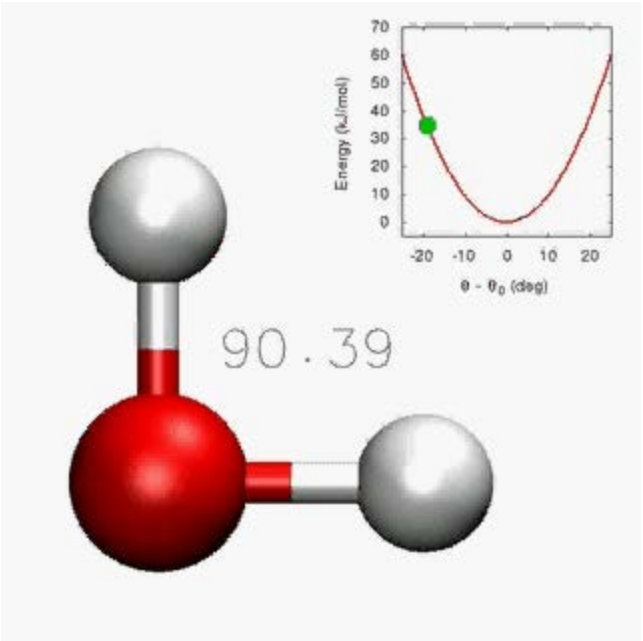


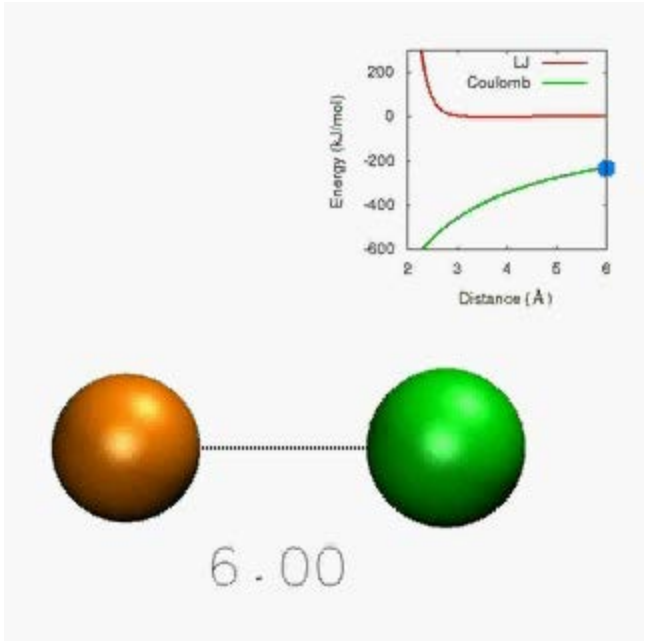
- Computer simulations of atoms and molecules span a vast range of detail
- More detailed theories can describe complex phenomena and offer higher accuracy
- Less detailed theories allow for simulation of larger systems / longer timescales
- In molecular mechanics simulation, the potential energy of molecules is

Introduction: Force Fields









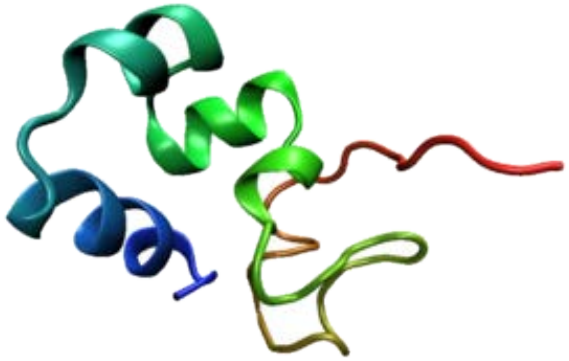
- **Force fields** are built from *functional forms* and empirical *parameters*
- Interactions include bonded pairwise, 3-body, and 4-body interactions…
- … as well as non-bonded pairwise interactions
- Simulation accuracy depends critically on choice of parameters

Creating a Force Field

Your project may require you to build a force field or obtain parameters from the literature.

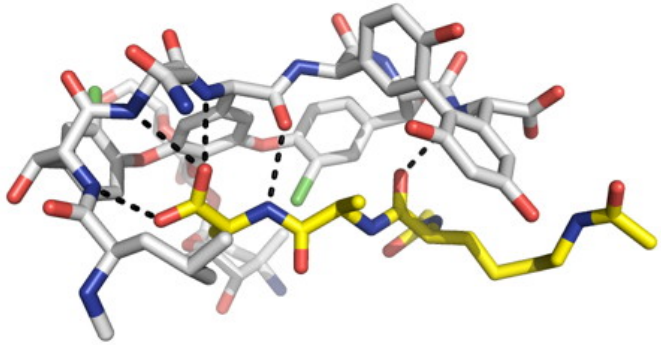
Common:
Biomolecules (e.g. villin)

Ships with OpenMM and most MD codes



Uncommon:
Organic molecules (e.g. vancomycin)

General force field procedures (GAFF, CGenFF) available



Rare:
Inorganic molecules (e.g. Photosystem II OEC)

Consult the literature

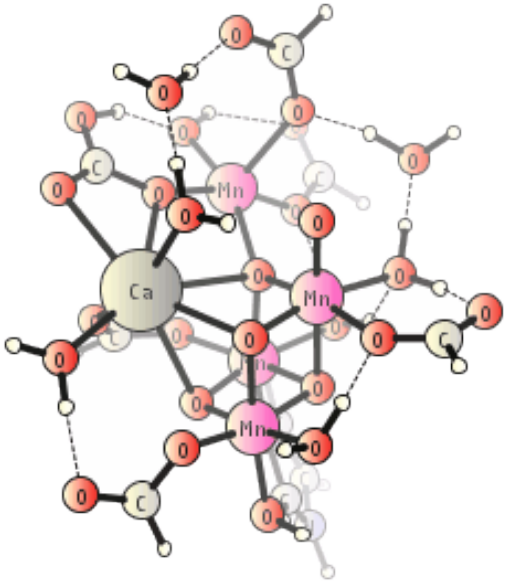


Image sources:
http://en.wikipedia.org/wiki/Villin
http://www.georgian.edu/chemistry/
http://qc.physto.se/~ps/biochem.html

Creating a force field: Functional form

Step 1: Choose a functional form to represent the potential energy surface, or design your own.

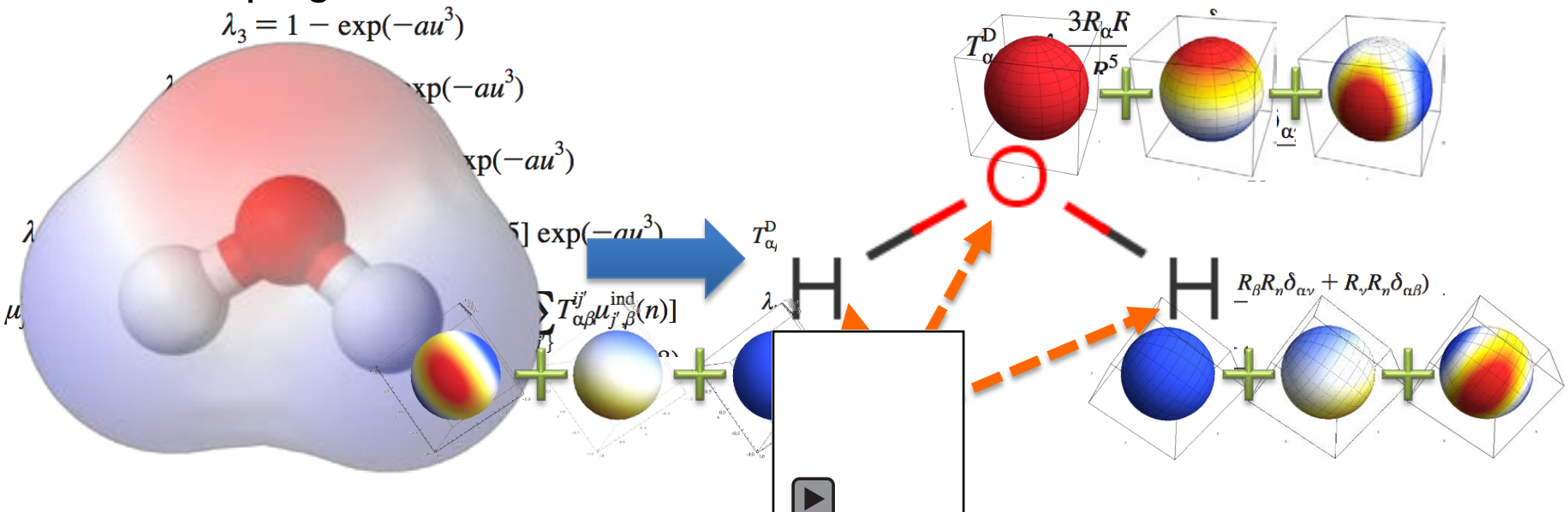
AMBER fixed-charge force field:

$$\sum_{i<j} \frac{q_i q_j}{r_{ij}}$$

- Point charge on each atom

AMOEBA polarizable force field:

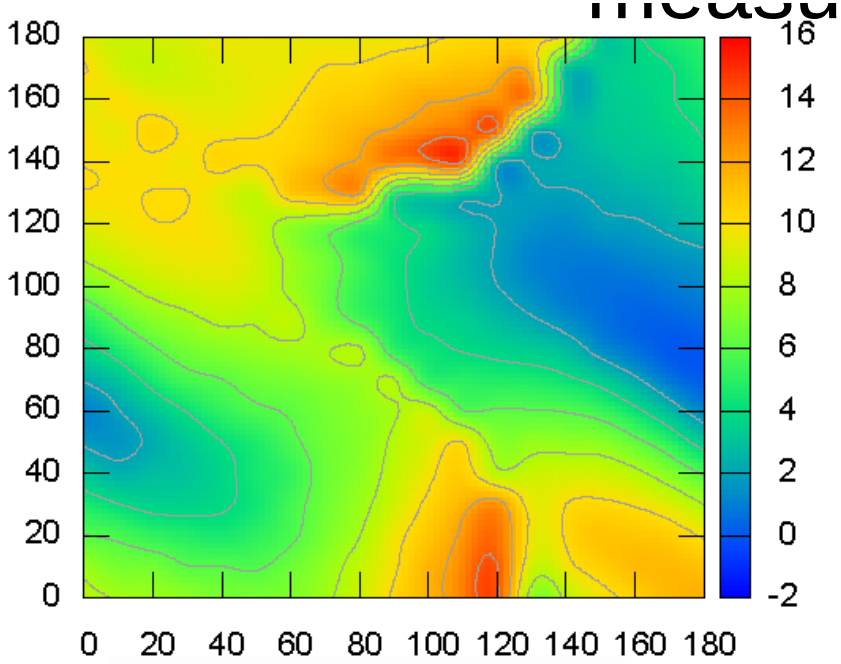
- Point charge, dipole, and quadrupole on each atom
- Polarizable point dipole on each atom with short-range damping



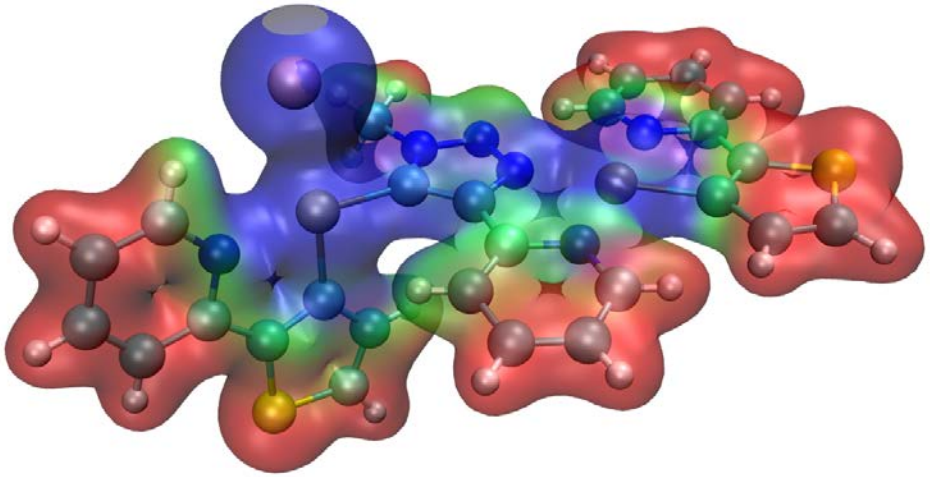
Creating a force field: Reference data

Step 2: Create a reference data set from theoretical calculations or experimental

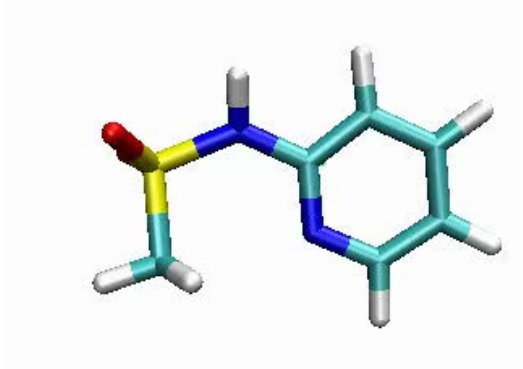
Energy scan across 2 dihedral angles

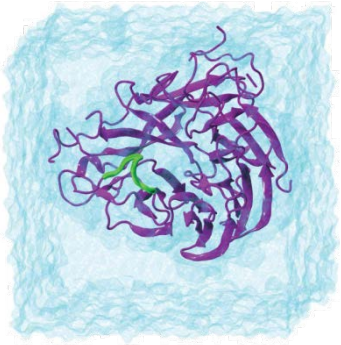


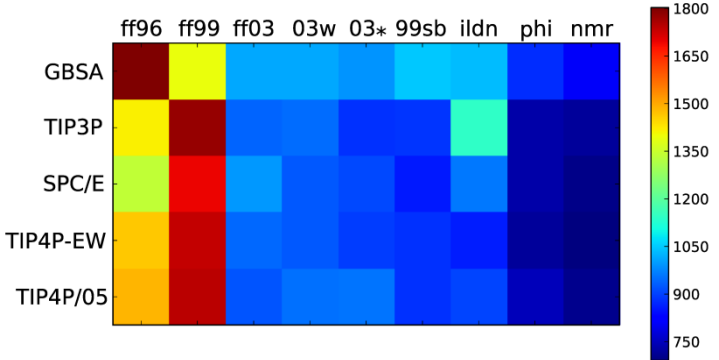
Electrostatic potential on a molecular surface (red = positive, blue = negative)



Simulated vs. experimental NMR chemical shifts for proteins (red = bad, blue = good)







Creating a force field: Optimization method

Step 3: Construct an objective function and apply an optimization method to minimize it.

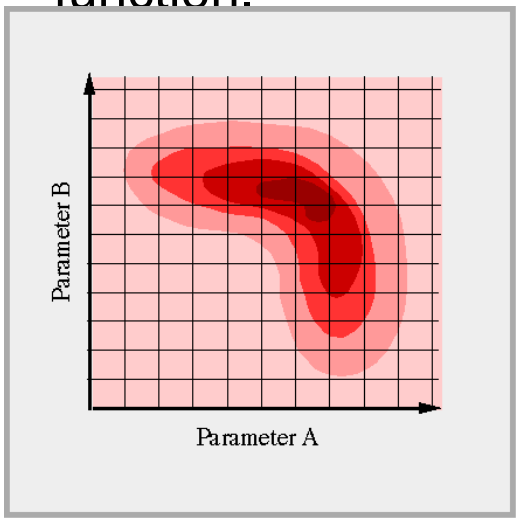
- The **objective function** measures the disagreement between the reference data and corresponding simulation result.
- An **optimization algorithm** searches for parameters that minimize the objective function.

$R = \text{Reference Data}$

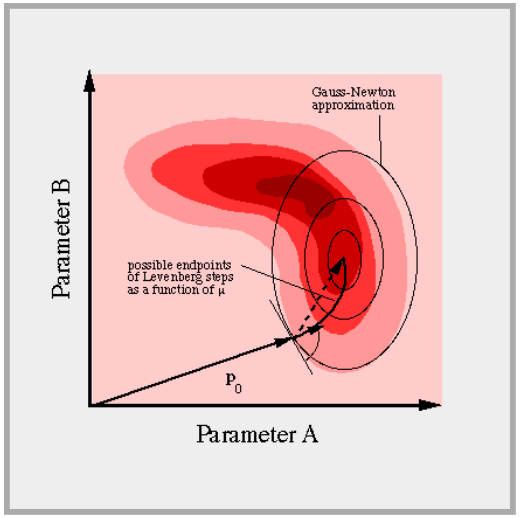
$S = \text{Simulation Result}$

$$\chi^2(\mathbf{k}) = (R - S(\mathbf{k}))^2$$

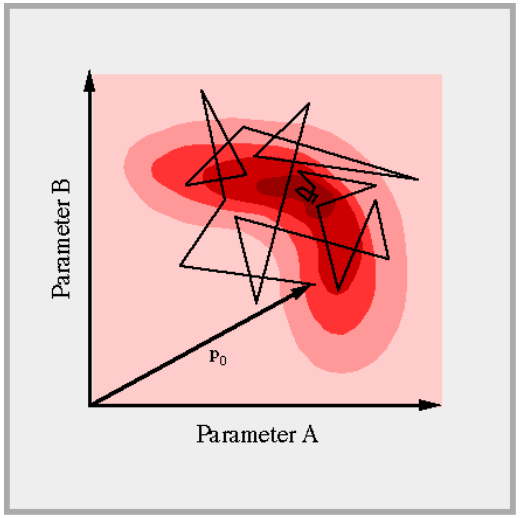
$$\mathbf{k}_{opt} = \min_{\mathbf{k}} \chi^2(\mathbf{k})$$



Grid Scan



Newton-Raphson

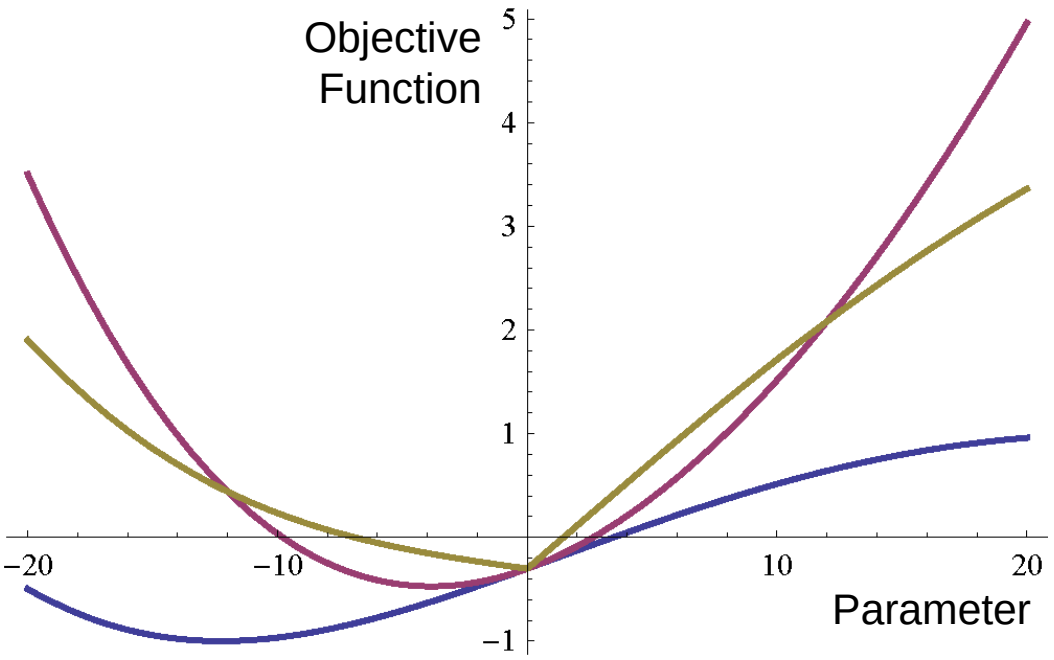


Simulated Annealing

Bayesian regularization

Optimizations with hundreds of parameters are made possible through strict regularization.

- We address overfitting issues by applying a penalty function, which may be interpreted as a Bayesian *prior distribution*.
- Different types of priors (Gaussian, Laplacian) have various impacts on the optimization behavior



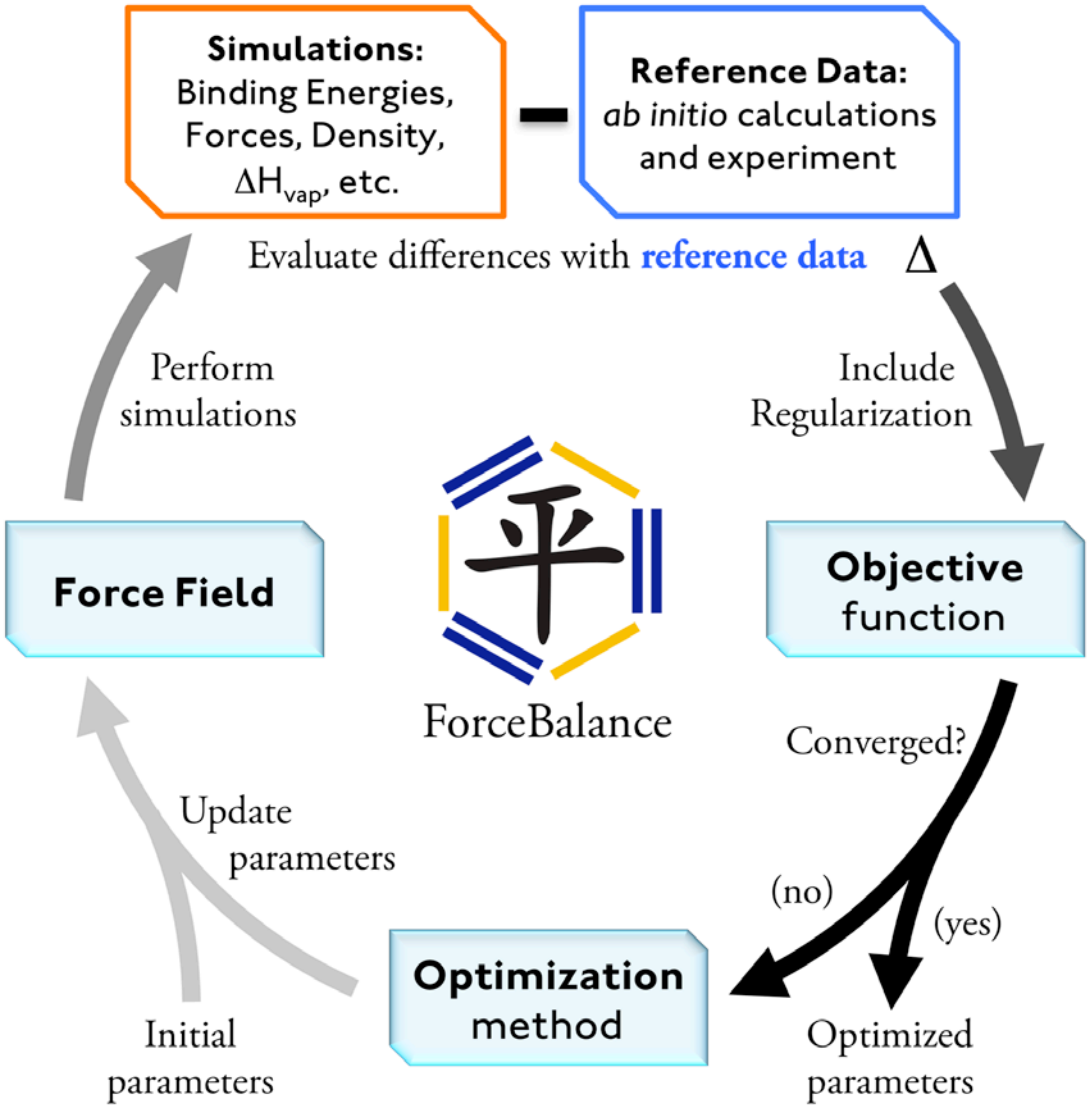
Gaussian prior (L2 regularized):
Large movements penalized

Laplacian prior (L1 regularized):
Some parameters don't change

No regularization:
Prone to overfitting

Introducing ForceBalance

ForceBalance is free software for creating force fields



- Written in Python
- Natural interface with OpenMM
- ForceBalance handles execution of simulations, building the objective function and derivatives, and nonlinear optimization

http://simtk.org/home/forcebalance