

An Introduction to Force Fields

Prof. Michael Shirts

University of Colorado Boulder

7th i-CoMSE Workshop: Molecular Dynamics

Boise State University, July 8, 2024

Licence: CC-BY 4.0



What is a Force Field?

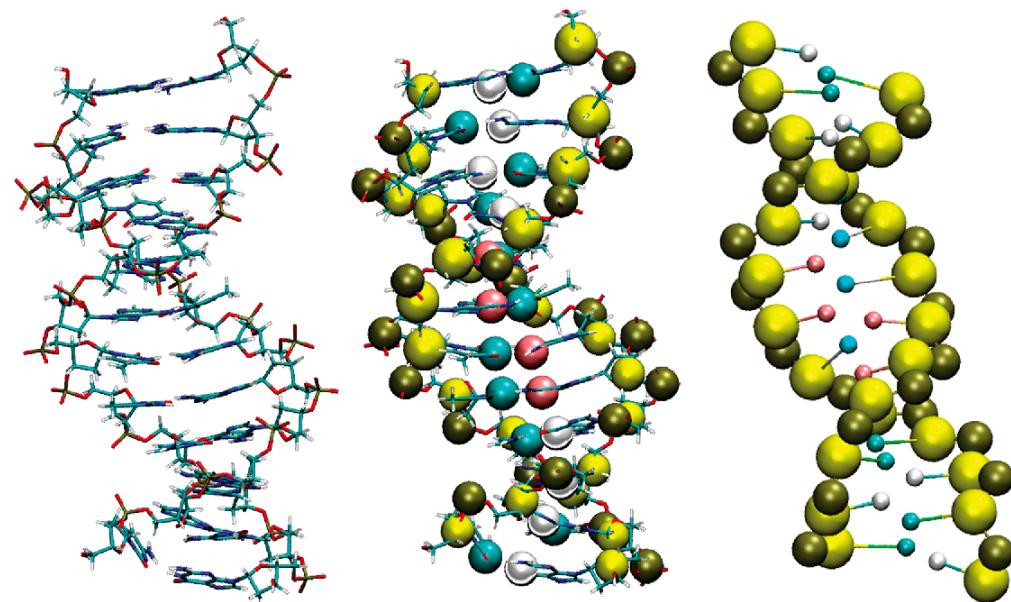
$$U(x_1, x_2, x_3, \dots, x_n)$$

- **Any way** to assign an energy to the coordinates of particles in the system
- You can compute the force and use for MD as long as the system has continuous derivatives.
- Pick a $U(x)$ **that is good enough for your purposes.**
- If you define a $U(x)$, you also define an $F(x)$, which assigns a vector (the force) to each configuration in space.
 - Thus, it is a **vector field**, where the vector is the force!

$$F(\vec{x}) = - \left[\frac{\partial U(\vec{x})}{\partial x_1}, \frac{\partial U(\vec{x})}{\partial x_2}, \dots, \frac{\partial U(\vec{x})}{\partial x_n} \right]$$

What are we trying to simulate?

- We have some **model** of our molecular system.
- We have the **hypothesis** that the physical behavior only depends on **some** of the physics of the molecule
- We test that hypothesis with a simulation
- This model can be atomistic, quantum mechanical, or coarse grained.
 - In CG, each particle isn't an atom, but a methane particle, or an amino acid in a protein

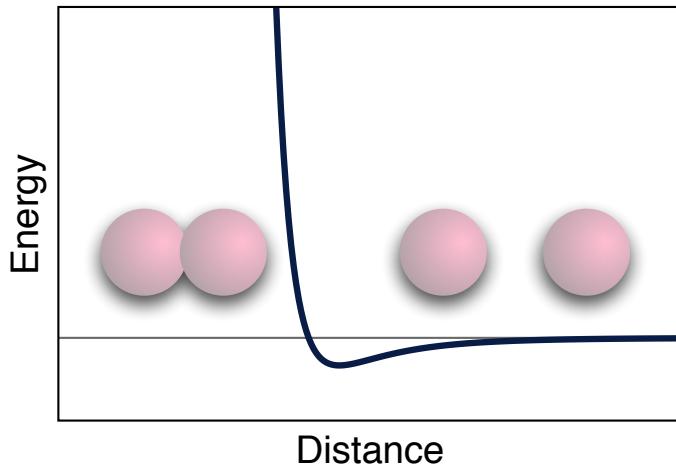
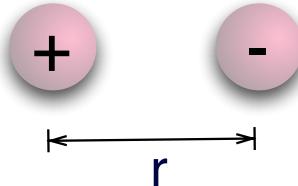


We break up the potential energy into classical terms, and fit the resulting parameters

Most common nonbonded force fields are **pair potentials**

$$V(r_1, r_2, \dots, r_n) = \sum_{i < j} V_{ij}(r_{ij})$$

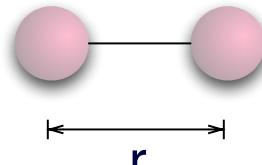
$$U(\mathbf{x}) = \sum_{\text{all pairs}} \frac{q_i q_j}{r} + \sum_{\text{all pairs}} \frac{A_{ij}}{r^{12}} - \frac{B_{ij}}{r^6}$$



"Bonded Terms"

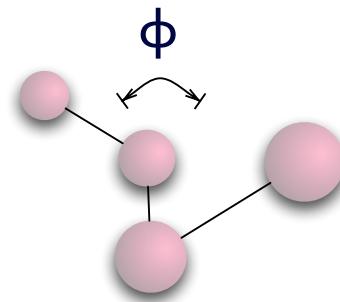
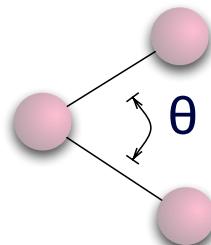
- We replace the electrons in orbitals with simple functional forms that are 10^6 times cheaper (but WAY more approximate).

$$+ \sum_{\text{all bonds}} \frac{1}{2} K_b (r - r_0)^2$$



$$+ \sum_{\text{all torsions}} \sum_n K_\phi [1 - \cos(n\phi + \phi_0)]$$

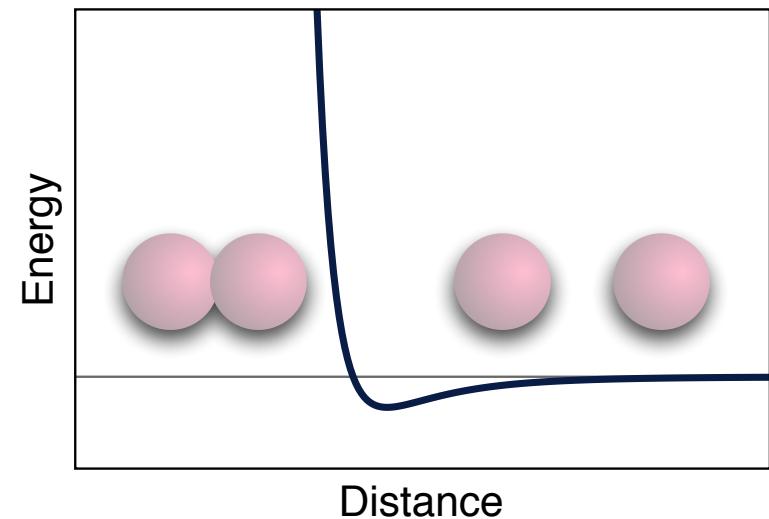
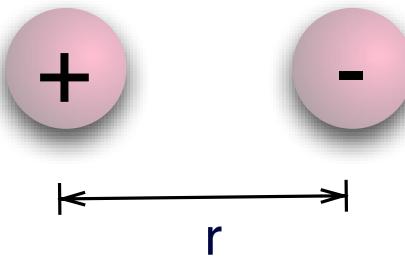
$$+ \sum_{\text{all angles}} \frac{1}{2} K_\theta (\theta - \theta_0)^2$$



"Nonbonded terms"

- We replace the electrostatic / dipole / induced dipole interactions with physically motivated (but not exact) functional forms.

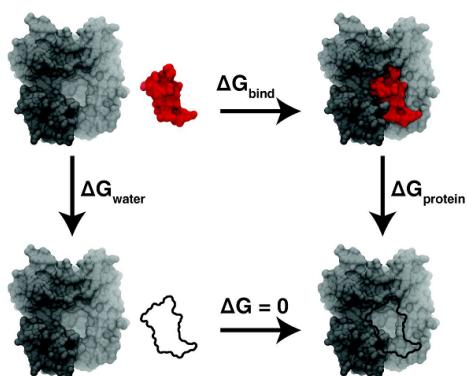
$$U(\mathbf{x}) = \sum_{\text{all pairs}} \frac{q_i q_j}{r} + \sum_{\text{all pairs}} \frac{A_{ij}}{r^{12}} - \frac{B_{ij}}{r^6}$$



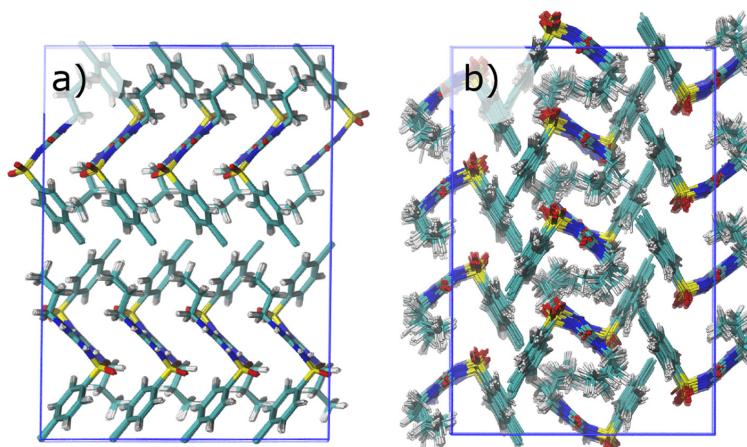
Small molecule force fields are widely-used tools in computational biophysics and soft matter systems

- Surprisingly, all-atom fixed charge force fields strike an appropriate balance between chemical accuracy and computation efficiency for many systems of interest.

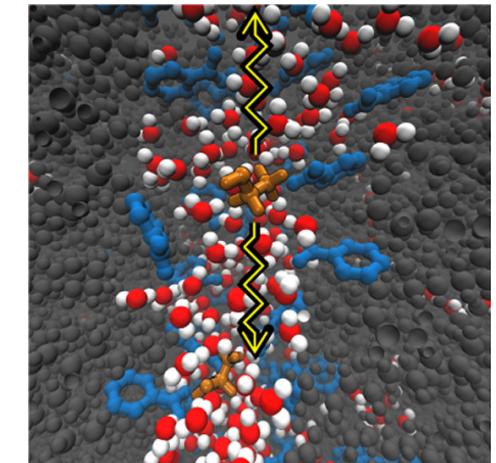
Protein-Ligand Binding



Drug molecule screening



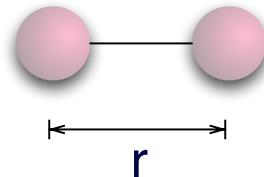
Molecular transport



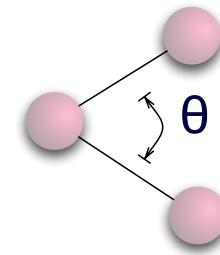
"Bonded Terms"

- We replace the energies of electrons jiggling in molecular orbitals with simple functional forms that are 10^6 times cheaper (but WAY more approximate).

$$+ \sum_{\text{all bonds}} \frac{1}{2} K_b (r - r_0)^2$$



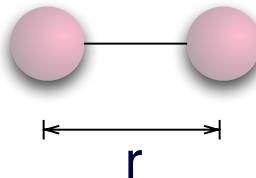
$$+ \sum_{\text{all angles}} \frac{1}{2} K_\theta (\theta - \theta_0)^2$$



- Most common are harmonic bonds and angles
- Pretty good approximation to quantum mechanics!
- Can get the parameters directly from QM of isolated molecules, as they are relatively independent of environment

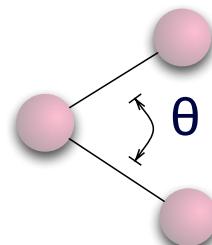
Bonds and Angles

$$+ \sum_{\text{all bonds}} \frac{1}{2} K_b (r - r_0)^2$$

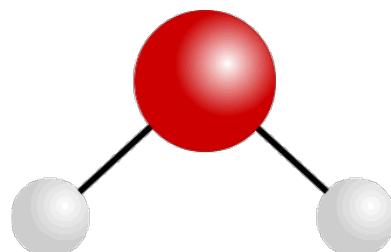


```
[ bondtypes ]
; i      j      func      b0          kb
CT      HC      1      0.10900    284512.0 ;
CA      CA      1      0.14000    392459.2 ;
CA      HA      1      0.10800    307105.6 ;
```

$$+ \sum_{\text{all angles}} \frac{1}{2} K_\theta (\theta - \theta_0)^2$$

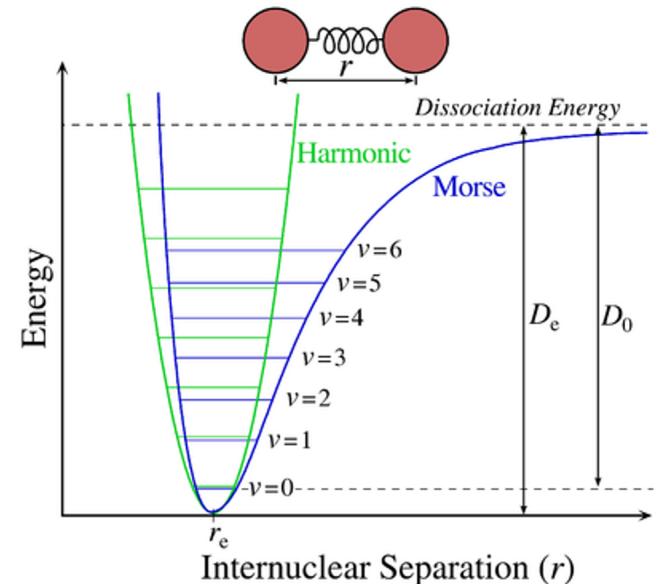


```
[ angletypes ]
; i      j      k      func      theta0      ka
HC      CT      HC      1      107.800    276.144 ;
CA      CA      CA      1      120.000    527.184 ;
CA      CA      HA      1      120.000    292.88  ;
CA      CA      CB      1      120.000    527.184 ;
```



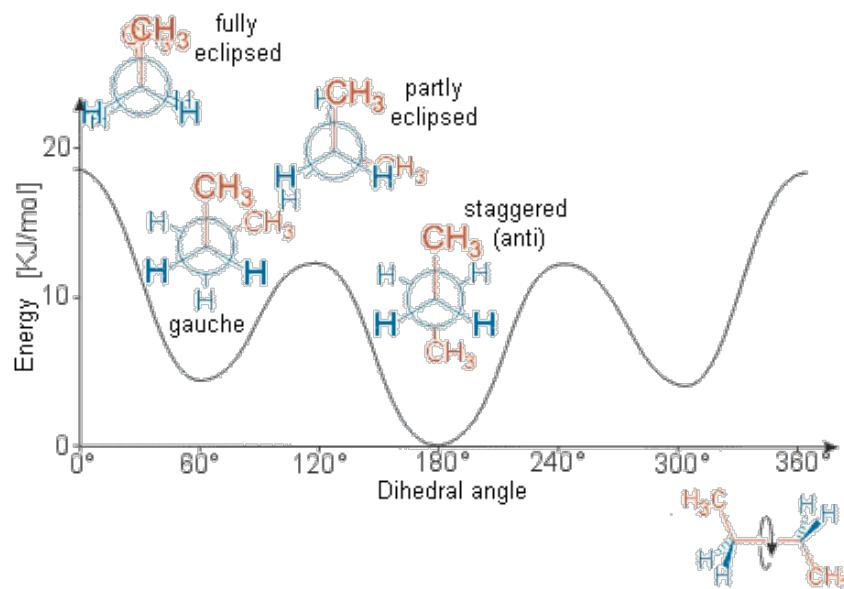
Other Bonds and Angles Options

- Bonds
 - Could use
$$U(x) = k_2(r - r_0)^2 + k_4(r - r_0)^4$$
 - Morse potentials, that allow for disassociation:
- Angles:
 - Could add Urey-Bradley term
 - $$U(x) = k_\theta(\theta - \theta_0)^2 + k_{UB}(r_{ik} - r_u b)^2$$
 - ML potentials

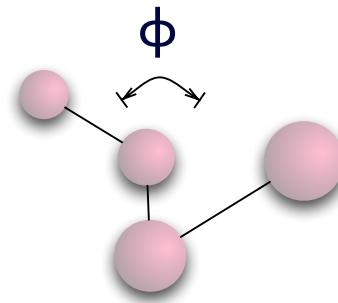


Torsions, or the junk bin of force fields

- Approximate QM with a sum of periodic trig functions



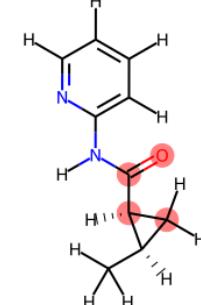
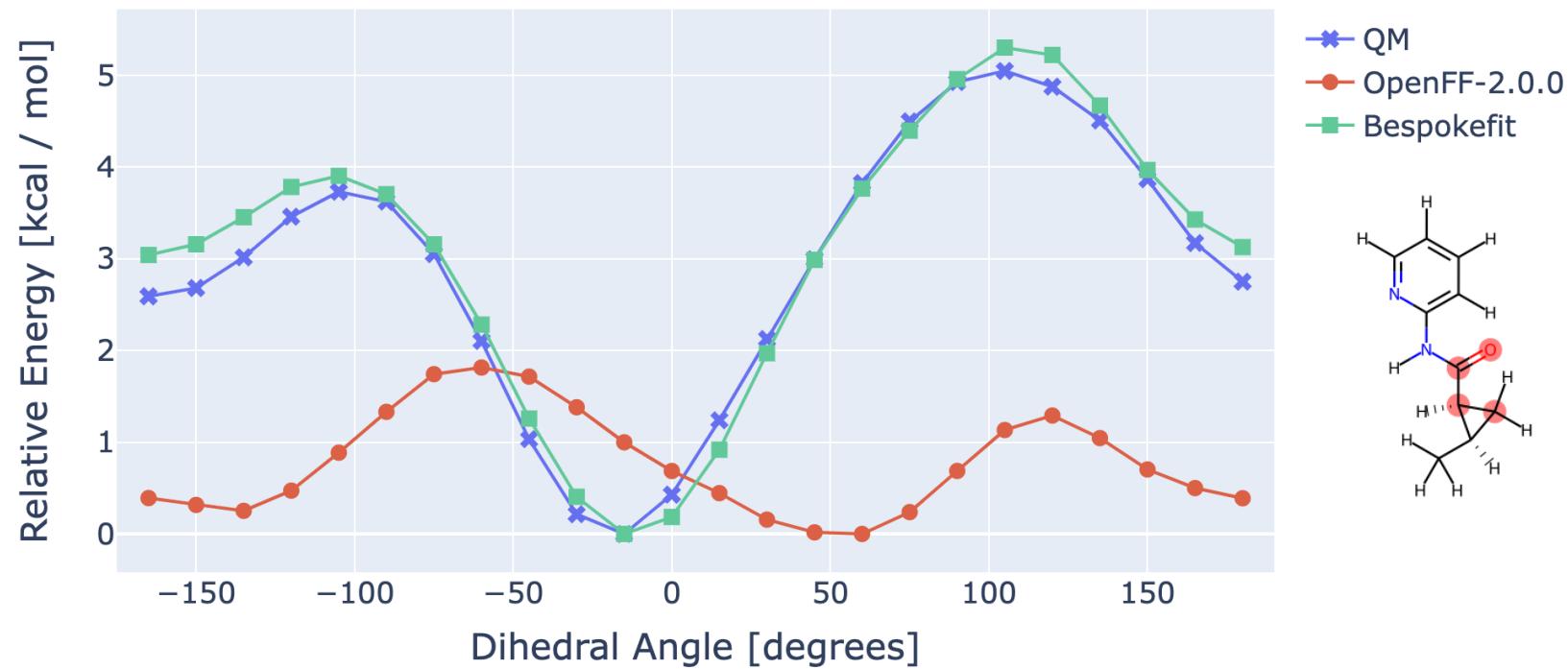
$$+ \sum_{\text{all torsions}} \sum_n K_\phi [1 - \cos(n\phi + \phi_0)]$$



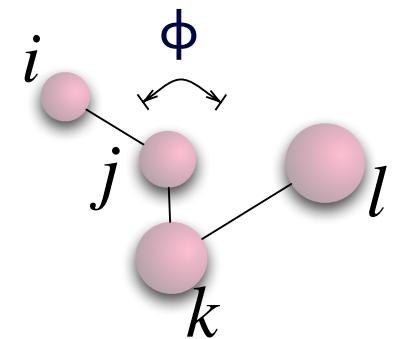
[dihedrals]						
;ai	aj	ak	al	funct	phi	k
51	54	57	60	1	0.000000000000	1.796983533954
51	54	57	60	1	180.000000000000	1.064375944885
51	54	57	60	1	180.000000000000	3.655209445001

- Even more annoyingly, there are at least 3 different equivalent representations of dihedrals

It can be difficult to get these right!

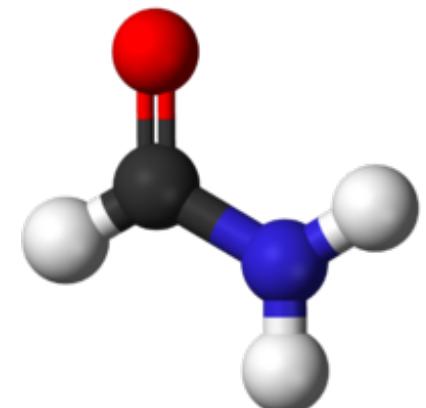
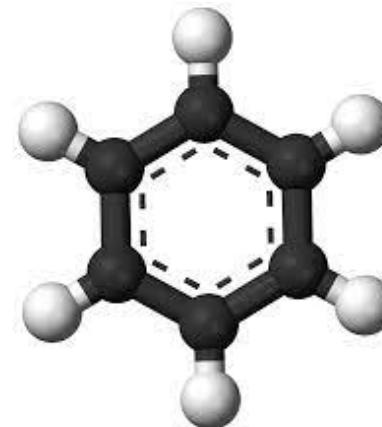
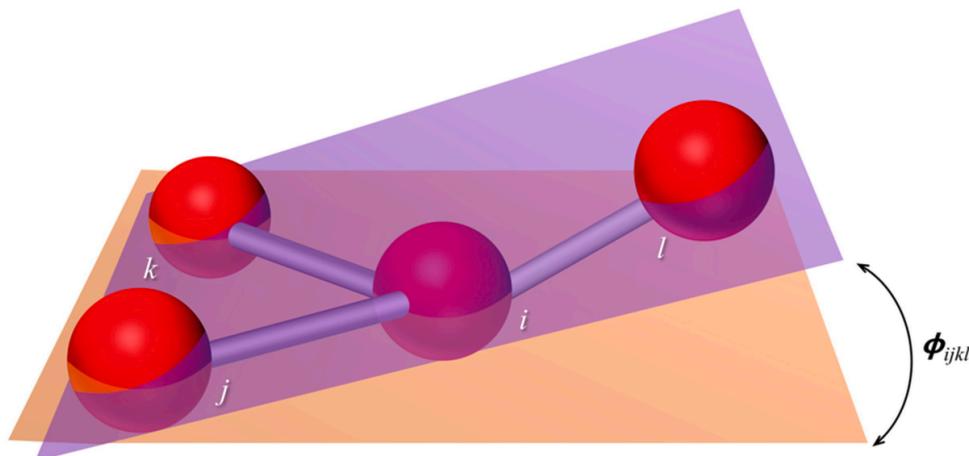


- NOTE: the energy of changing the torsion is not JUST the torsional potential, but involves the nonbonded interaction between atoms i and l .



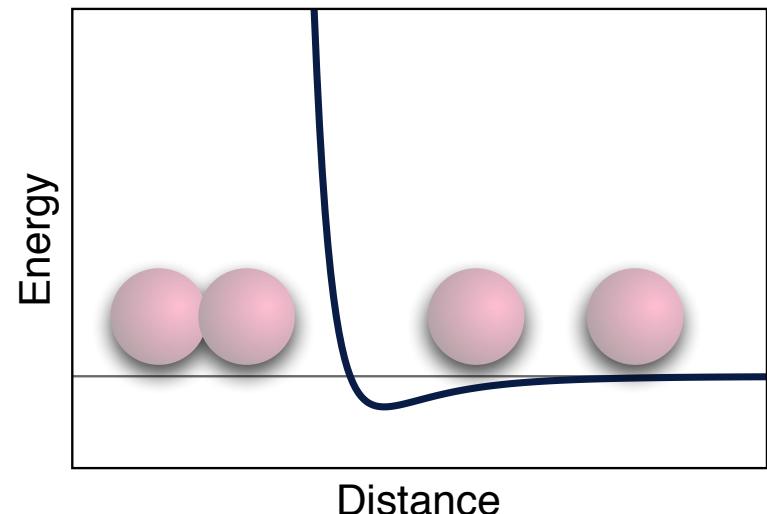
Improper torsions

- A way to keep molecular systems with conjugated double bonds planar
- Energy minimum is at 0 degrees



Lennard-Jones terms

$$U(r_{ij}) = 4\epsilon \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right)$$



```
[ atomtypes ]
; atom type name type      mass          sigma        epsilon
  opls_111      OW      15.99940    0.000      3.12171e-01  7.94960e-01
  opls_112      HW      1.00800    0.000      0.00000e-01  0.00000e-01
```

- This just has sigma and epsilon for O and H. What about the O-H interaction???

Combining rules

```
[ defaults ]
; nbfunc      comb-rule      gen-pairs      fudgeLJ  fudgeQQ
1           3               yes            0.5      0.5
```

Geometric $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$ $\sigma_{ij} = \sqrt{\sigma_i \sigma_j}$

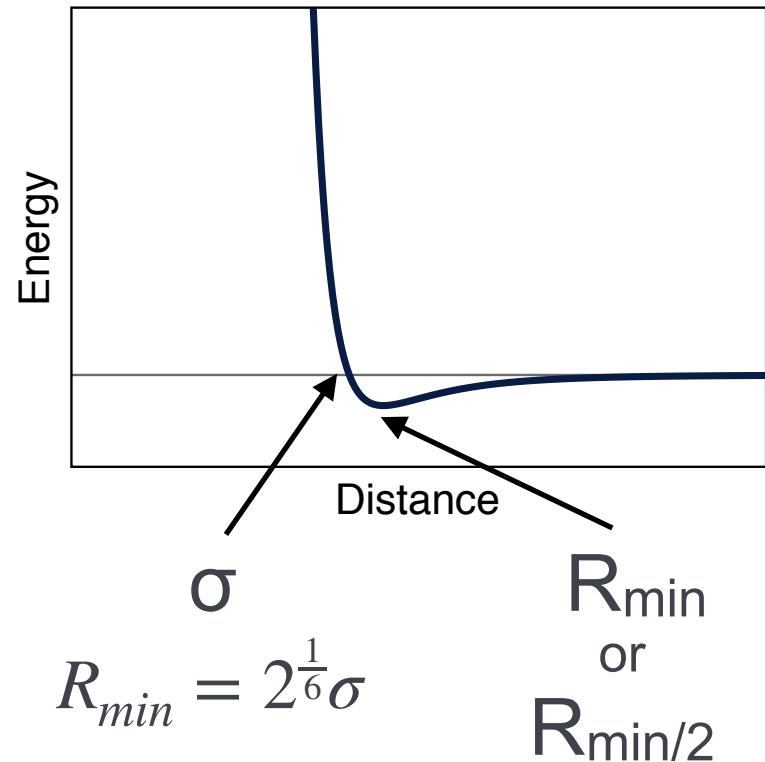
Lorentz-Berthelot $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$ $\sigma_{ij} = \frac{1}{2} (\sigma_i + \sigma_j)$

Works better than it should (but not perfect)

Lots of different, equivalent ways to write Lennard-Jones parameters!

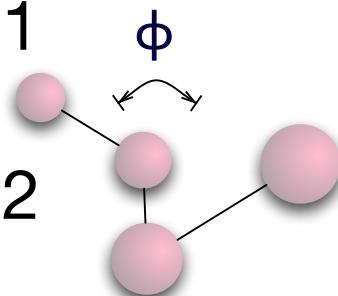
$$U(r_{ij}) = 4\epsilon \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right)$$

$$U(r_{ij}) = \frac{A_{ij}}{r^{12}} - \frac{B_{ij}}{r^6}$$



Exclusions?

- What about atoms separated by no bonds (1,2)?
- What about atoms separated by 1 bond (1,3)?
- What about atoms separated by 2 bonds (1,4)?



```
[ defaults ]
; nbfunc      comb-rule      gen-pairs
1           3                 yes
                                         fudgeLJ  fudgeQQ
                                         0.5      0.5
```

Other terms you could put in:

- Hydrogen bonding terms
- Polarizability
 - The ability for charge distributions to adjust to nuclear positions
 - Still classical
- Other many body terms
 - charge penetration
 - Force field that break bonds (ReaxFF)
 - Machine learning terms (ANI multibody neural net potentials, many others.
- BUT pairwise nonbonded potentials are usually faster than general multibody terms for MD

How do you choose force field parameters

- Bonded parameters:
 - Match to QM calculations
- Nonbonded parameters
 - Charges:
 - Match to electrostatics potential (RESP_)
 - AM1 with bond charge corrections (AM1-BCC)
- van der Waals
 - Match thermodynamic properties of liquids
 - Densities, enthalpies of vaporization or mixing + others

What sorts of force fields are out there?

- OPLS-AA / OPLS-UA
 - Small molecules, proteins
 - Evolved out of Bill Jorgensen's work
 - Now run by Schrödinger
 - Biomolecules and small molecule ligands
- TRAPPE
 - Coming mostly from Ilja Siepmann's group
 - Designed for phase equilibrium

Other Force Fields

- CHARMM
 - Evolved from the work in Martin Karplus's group at Harvard over ~50 years.
 - Alex MacKerrell (Maryland) and Benoit Roux (Chicago) are longest contributing developers but MANY more.
 - Proteins, lipids, nucleic acids small molecules
- AMBER
 - Evolved from the work in Peter Kollman's group at UCSF, and his academic descendants.
 - GAFF small molecule force field (Junmei Wang)
 - Proteins and nucleic acids

Other Force Fields

- Open Force Field

- <http://www.openforcefield.org>
- New effort, focused on open software as well as open force fields
- Small molecule based, in the process of adding biopolymers
- Can build from SMILES, sdf, mol2
- Difference is: it matches molecular parameters in a more general way and DOESN'T use atom types



Other Force Fields

- What about polymers?

- OPLS often used.
- COMPASS
- MMFF

- What about metals?

- Lots of metal specific functional forms.
- Generally requires more complicated interactions than the simple nonbonded potentials
- OpenKIM has cataloged these potentials
- Lammps supports many of them
-

But What Force Field Should I Use?

- The one that allows you to test the hypothesis that you want to

OR

- The one that is validated as an effective screening mechanism to predict the property of interest to other researchers.
- Has the force field I want to use been validated on sufficiently similar systems?
- "All models are wrong. Some models are useful."

Force Field Assignment Tools (with links)

- [AmberTools/tleap](#)
- [CHARMM-GUI](#)
- [CGENFF](#)
- [OpenFF toolkit](#)
- [foyer](#)
- [pysimm \(polymers - though OpenFF in beta\)](#)
- [OpenKIM \(metals\)](#)

How are parameters decided?

- Bonds, angles, torsions are fit to QM
- Charges are fit to QM electrostatic potential
- van der Waals are fit to experimental thermodynamics
- But what about for a new molecule that you haven't seen before?

Chemical perception

- Even if we haven't seen a molecule before, it's made up of local chemical environments we've (probably) seen before . . .

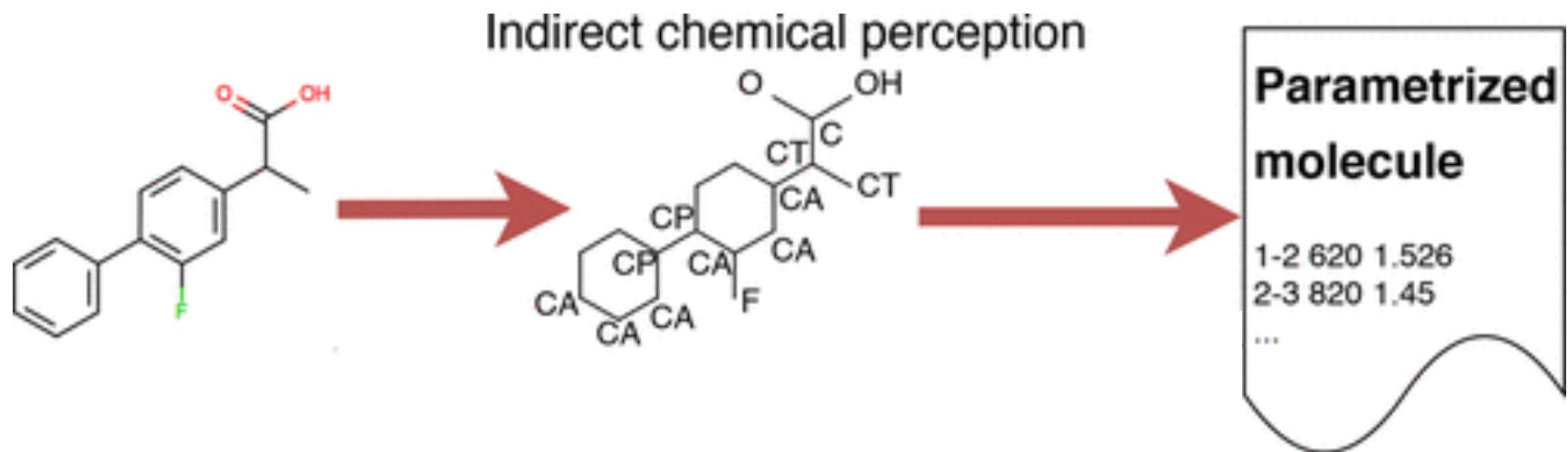
Atom Types

- Give a label to each atom depending on its environment
- Encode the chemical environment in the labels
- Assign parameters based on these labels

```
[ angletypes ]
; i      j      k      func    theta0      ka
HC      CT     HC      1      107.800   276.144 ;
CA      CA     CA      1      120.000   527.184 ;
CA      CA     HA      1      120.000   292.88  ;
CA      CA     CB      1      120.000   527.184 ;
```

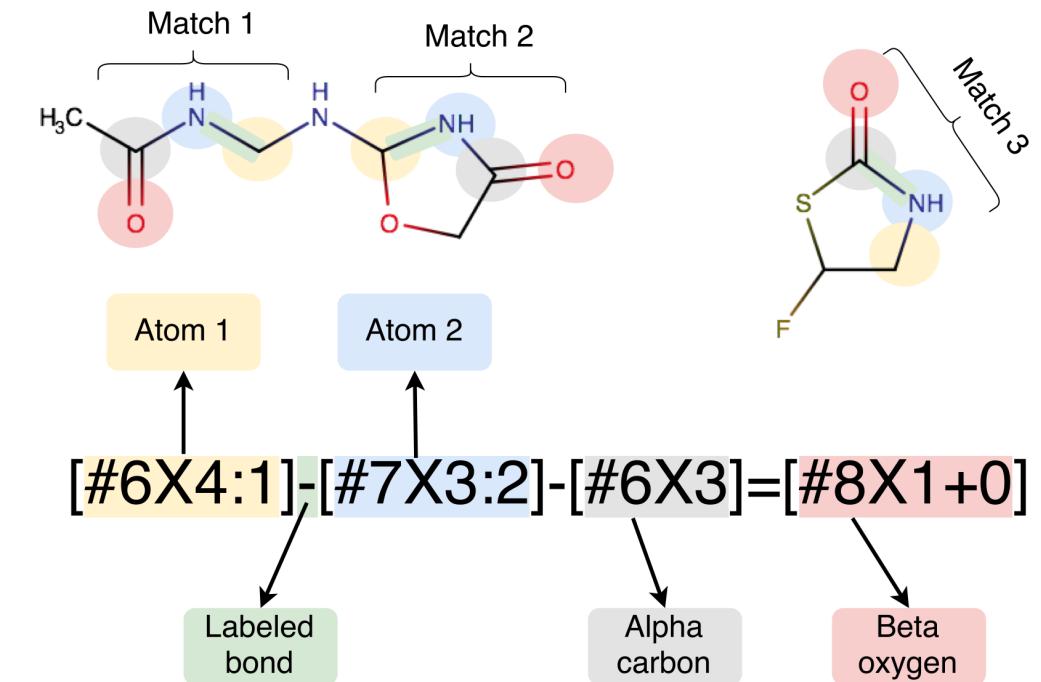
Category	Force Field	Atomic Type	Description (from AMBER/GAFF Parameters File)
Polarized	AMBER	C	sp2 C carbonyl group
	GAFF	c	sp2 C carbonyl group
Aromatic	AMBER	C*	sp2 arom. 5 memb.ring w/1 subst. (TRP)
	AMBER	CA	sp2 C pure aromatic (benzene)
	AMBER	CB	sp2 aromatic C, 5&6 membered ring junction
	AMBER	CC	sp2 aromatic C, 5 memb. ring HIS
	AMBER	CN	sp2 C aromatic 5&6 memb.ring junct.(TRP)
	AMBER	CR	sp2 arom as CQ but in HIS
	AMBER	CV	sp2 arom. 5 memb.ring w/1 N and 1 H (HIS)
	AMBER	CW	sp2 arom. 5 memb.ring w/1 N-H and 1 H (HIS)
	GAFF	ca	sp2 C in pure aromatic systems
	GAFF	cc	sp2 carbons in non-pure aromatic systems
Aliphatic	GAFF	cd	sp2 carbons in non-pure aromatic systems
	GAFF	cp	Head sp2 carbons connecting rings in bi-phenyls
	GAFF	cq	Head sp2 carbons connecting rings in bi-phenyls
	AMBER	CT	sp3 aliphatic C
	GAFF	c1	sp C
	GAFF	c2	sp2 C
	GAFF	c3	sp3 C
	GAFF	ce	Inner sp2 carbons in conjugated systems
	GAFF	cf	Inner sp2 carbons in conjugated systems
	GAFF	cg	Inner sp carbons in conjugated systems
	GAFF	ch	Inner sp carbons in conjugated systems
	GAFF	cu	sp2 carbons in triangle systems
	GAFF	cv	sp2 carbons in square systems
	GAFF	cx	sp3 carbons in triangle systems
	GAFF	cy	sp3 carbons in square systems

Atom Typing



Direct Chemical Perception

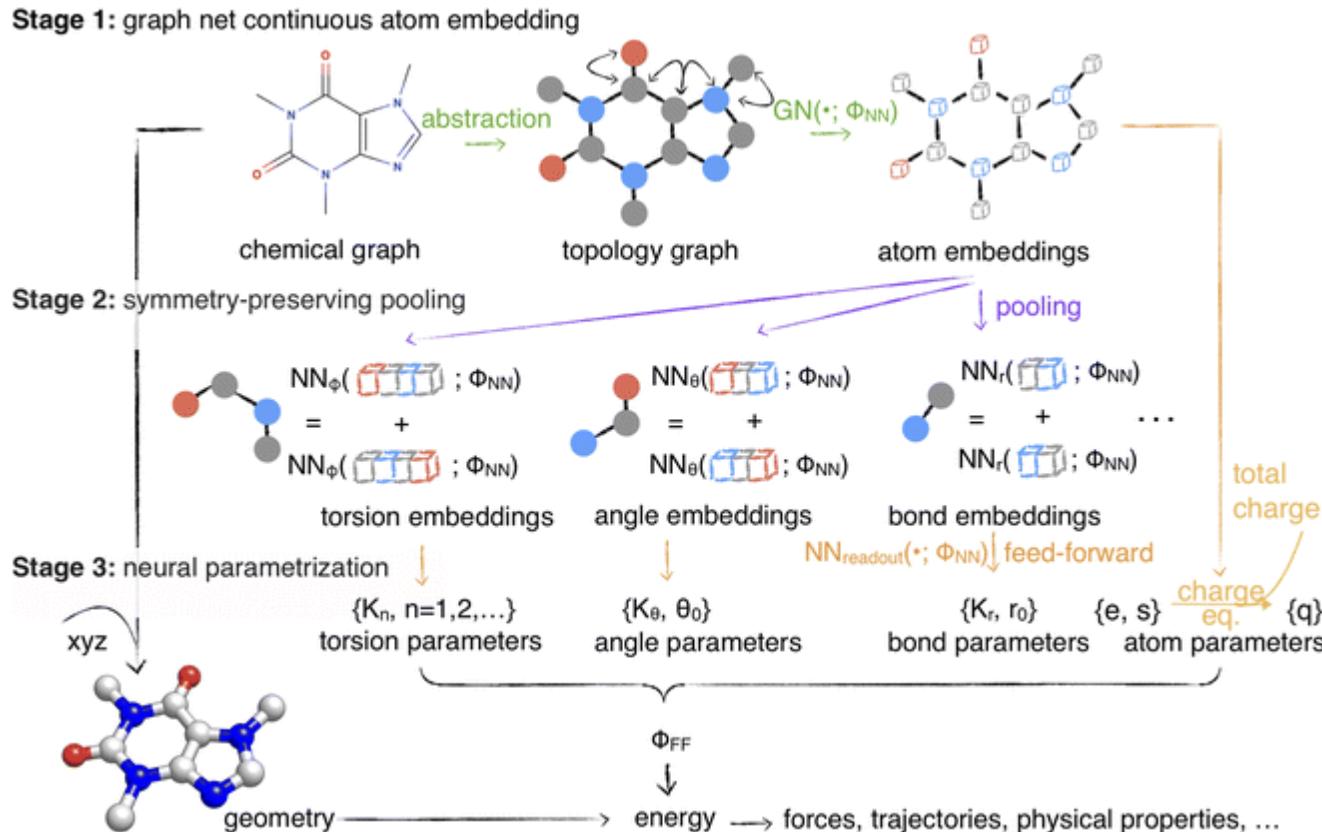
- The SMIRKS Native Open Force Field (SMIRNOFF)
- Use of industry-standard SMARTS/SMIRKS chemical perception greatly simplifies process



Charges

- RESP approach to fitting QM
- AM1-BCC
 - Use CHEAP QM, and adjust these charges empirically
 - Takes 2-3 seconds for small molecule; but 30 min for a lipid and crashes for a polymer
- ML approaches
 - TRAIN to RESP or AM1-BCC charges
 - Can run in seconds for proteins and polymers.
 - OpenFF implementation: NAGL

More general way to assign chemical environments: Espaloma



- Turns chemical environments into **vectors**
- Parameters are continuous functions of these environment vectors, not discrete choices
- Charges as well as other parameters