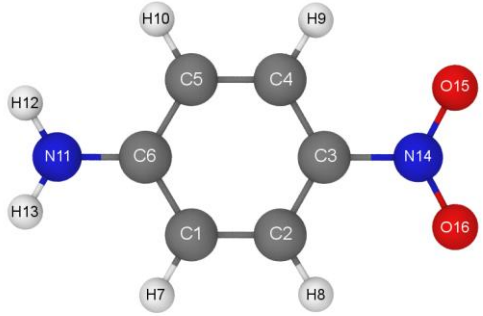
Force field parameterization for multiscale simulations

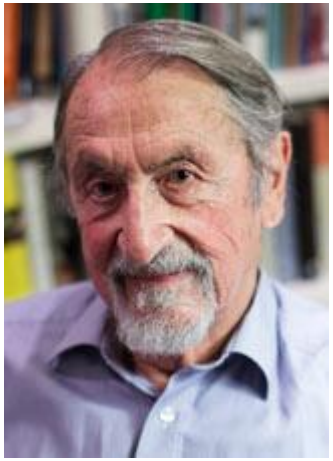
Xin Li
TheoChem&Bio, KTH, Sweden
lixin@theochem.kth.se



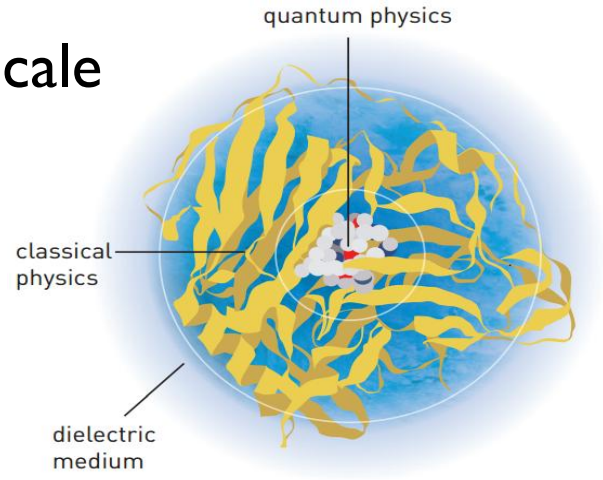
Wrocław, 2014-06-24

The Nobel Prize in Chemistry 2013

- Martin Karplus, Michael Levitt, Arieh Warshel

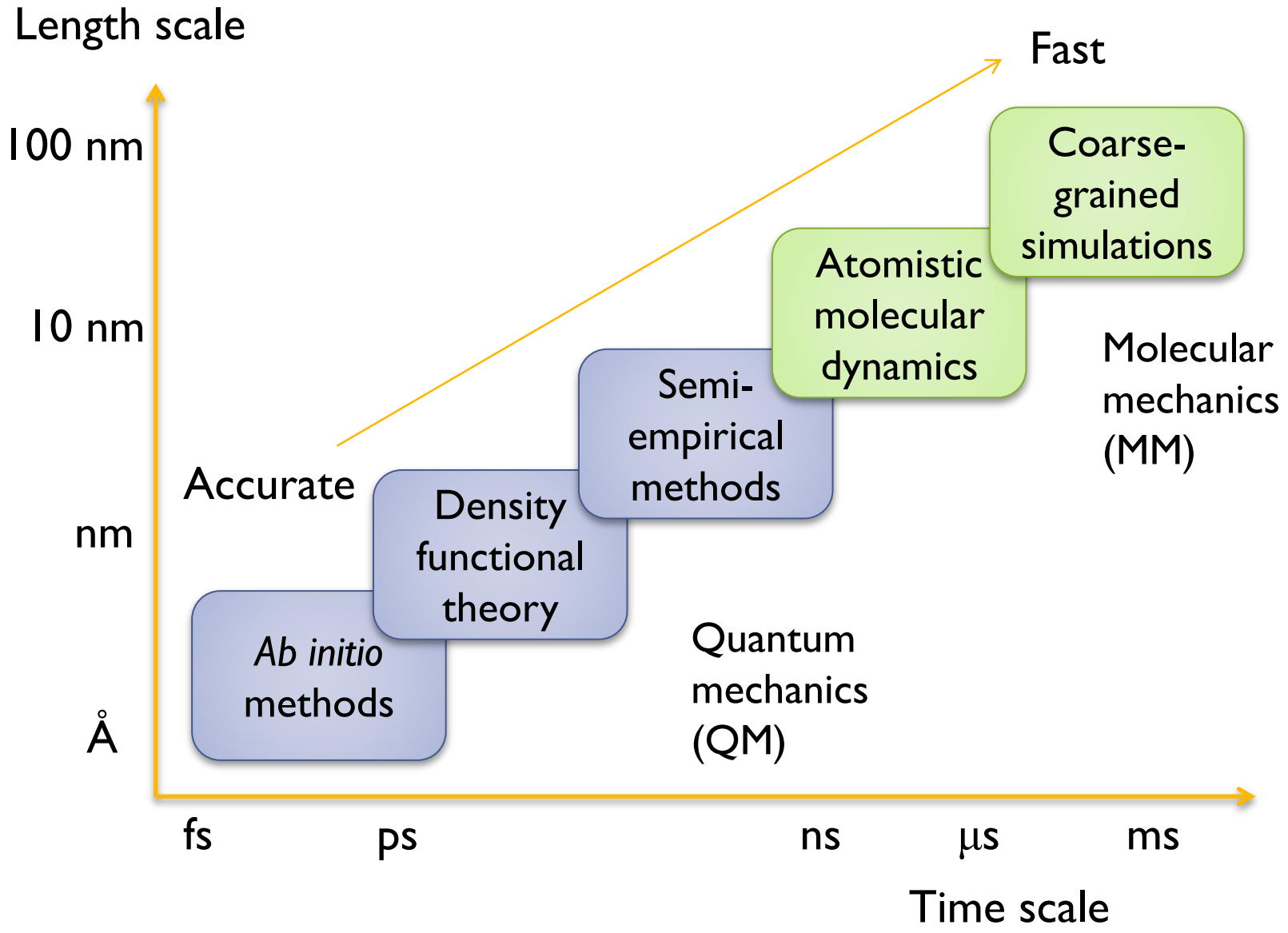


"For the development of multiscale models for complex chemical systems"



http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2013/press.html

Hierarchy of multiscale simulations



General procedure

Molecule

Gaussian09

Geometry and ESP charges

Amber Tools

Hessian and potential energy surface

GROMACS input files: .top and .gro

AMBER input files: .prmtop and .prmcrd

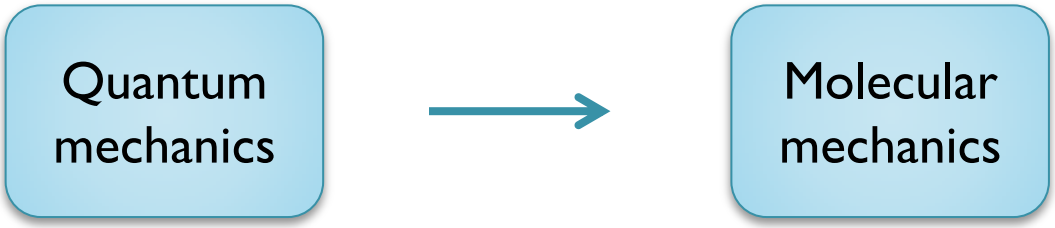
Refinement of force field parameters

MD trajectories

Property calculations

How to parameterize?

- Derive molecular mechanical force field parameters using quantum mechanical (mainly density functional theory) calculations



More accurate
Applicable to small systems

Less accurate
Applicable to large systems and long time scale

Step 1:

- Run necessary quantum chemical calculations.
- Generate initial force field parameters, for example, using GAFF.
- Get bond stretching and angle bending force constants from Hessian.
- Refine the dihedral angle force constant by fitting to QM potential energy surface.
- Run MD simulation with new set of parameters.

Quantum mechanics in chemistry

- Schrödinger equation

$$E\Psi = \hat{H}\Psi$$

- Born-Oppenheimer approximation

- Nucleus represented by point charge
- Electrons represented by wave function

Density functional theory

- Exchange-correlation approximation
  - A large number of functionals to choose from...
  - Local, GGA, meta-GGA, Hybrid, Double-hybrid, etc.

- Hybrid functionals are most widely used
  - B3LYP, PBE0, CAM-B3LYP, etc.

1993 1996 2004

Basis set

- A set of functions used to fit the electronic wave function
- Quality: double-zeta (DZ), triple-zeta (TZ), QZ, 5Z, etc.
- Polarization and diffusion functions: necessary for realistic simulations

- `B3LYP/6-311+G(d,p)`

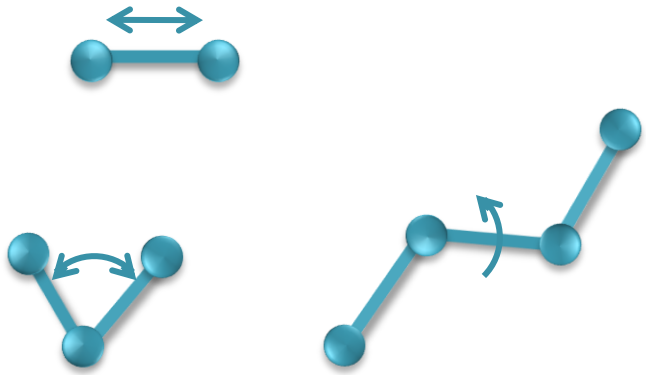
Step 2:

- Run necessary quantum chemical calculations.
- Generate initial force field parameters, for example, using GAFF.
- Get bond stretching and angle bending force constants from Hessian.
- Refine the dihedral angle force constant by fitting to QM potential energy surface.
- Run MD simulation with new set of parameters.

Force field

- Non-bonded interaction
  - Electrostatic interaction
  - Van der Waals interaction

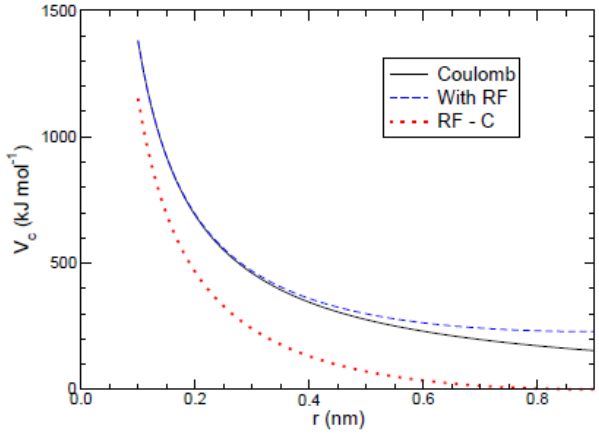
- Bonded interaction
  - Bond stretching
  - Angle bending
  - Dihedral angle



Non-bonded interaction

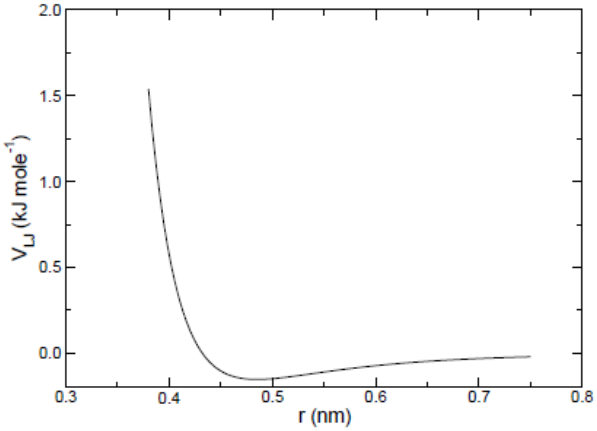
- Electrostatic interaction
  - Coulomb potential

$$E_{C,ij} = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$



- Van der Waals interaction
  - Lennard-Jones potential

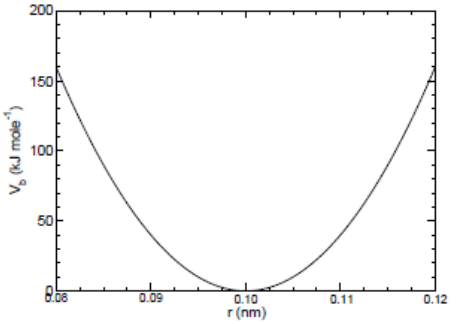
$$E_{LJ,ij} = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$$



Bonded interaction

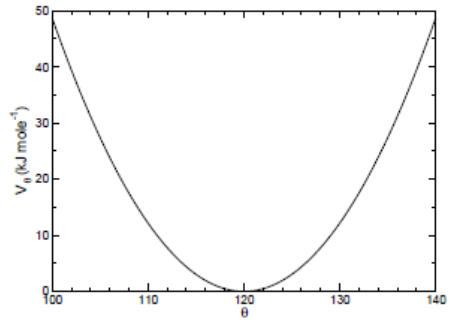
- Bond stretching

$$E_{b,ij} = \frac{1}{2} k_{b,ij} (r_{ij} - r_{ij,0})^2$$



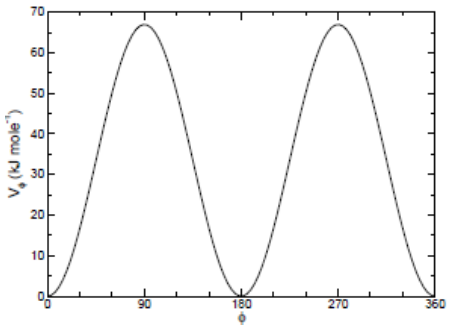
- Angle bending

$$E_{a,ijk} = \frac{1}{2} k_{a,ijk} (\theta_{ijk} - \theta_{ijk,0})^2$$



- Dihedral angle

$$E_{d,ijkl} = \sum k_{d,ijkl} [1 + \cos(n\phi_{ijkl} - \phi_{ijkl,0})]$$



Procedure

- Optimize the geometry and calculate the Hessian and ESP charges.
  - `g09 < molecule.inp > molecule.out`
  - `formchk molecule.chk`

```
%mem=1000MB
%chk=molecule.chk
#p B3LYP/6-311+G(d,p) NoSymm Opt Freq
SCRF(PCM,Solvent=Water) Pop=CHelpG

molecule

0 1
C 0.0408210174 1.2072071597 1.3657881129
......
O 0.0235063728 1.0862639647 -2.74056842
```

Procedure

- Generate AMBER input files using AmberTools.
  - `antechamber -i molecule.log -fi gout -o molecule.mol2 -fo mol2 -c esp`
  - `parmchk -i molecule.mol2 -o molecule.frcmod -f mol2`
  - `tleap -s -f leap.in`

```
source leaprc.gaff
mods = loadAmberParams molecule.frcmod
MOL = loadMol2 molecule.mol2
saveAmberParm MOL molecule.prmtop molecule.prmcrd
quit
```

Procedure

- Convert AMBER input files to GROMACS format
  - `perl amber_to_gmx_by_XL.pl -top molecule.prmtop -crd molecule.prmcrd -name molecule`

- Output files:
  - `gmx_molecule.top`
  - `gmx_molecule.gro`

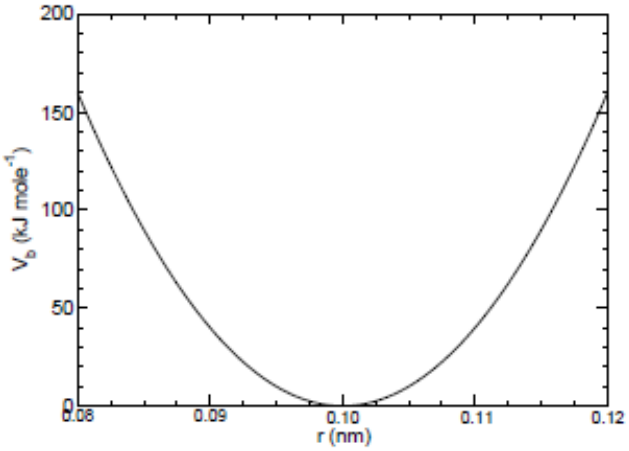
Step 3:

- Run necessary quantum chemical calculations.
- Generate initial force field parameters, for example, using GAFF.
- Get bond stretching and angle bending force constants from Hessian.
- Refine the dihedral angle force constant by fitting to QM potential energy surface.
- Run MD simulation with new set of parameters.

Why Hessian?

- Harmonic potential

$$E_{b,ij} = \frac{1}{2} k_{b,ij} (r_{ij} - r_{ij,0})^2$$



- First derivative

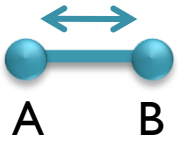
$$dE_{b,ij} / dr_{ij} = k_{b,ij} (r_{ij} - r_{ij,0})$$

- Second derivative

$$d^2E_{b,ij} / dr_{ij}^2 = k_{b,ij}$$

From Hessian to force constant

- For two atoms (A and B) connected by a chemical bond, the reaction force on atom A due to displacement of atom B is expressed as

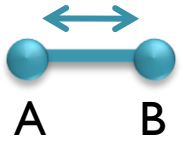


$$\begin{bmatrix} dF_{A,x} \\ dF_{A,y} \\ dF_{A,z} \end{bmatrix} = -\begin{bmatrix} \frac{\partial^2 E}{\partial x_A \partial x_B} & \frac{\partial^2 E}{\partial x_A \partial y_B} & \frac{\partial^2 E}{\partial x_A \partial z_B} \\ \frac{\partial^2 E}{\partial y_A \partial x_B} & \frac{\partial^2 E}{\partial y_A \partial y_B} & \frac{\partial^2 E}{\partial y_A \partial z_B} \\ \frac{\partial^2 E}{\partial z_A \partial x_B} & \frac{\partial^2 E}{\partial z_A \partial y_B} & \frac{\partial^2 E}{\partial z_A \partial z_B} \end{bmatrix} \times \begin{bmatrix} dx_B \\ dy_B \\ dz_B \end{bmatrix}$$

$$d\mathbf{F}_A = [k_{AB}] \times d\mathbf{r}_B$$

Bond stretching force constant

- The force constant matrix $[k_{AB}]$ has three eigenvalues λ_i^{AB} and three eigenvectors $\hat{v}_i^{AB}$ $(i = 1,2,3)$
- $\hat{u}^{AB}$ is unit vector from A to B
- The bond stretching force constant is

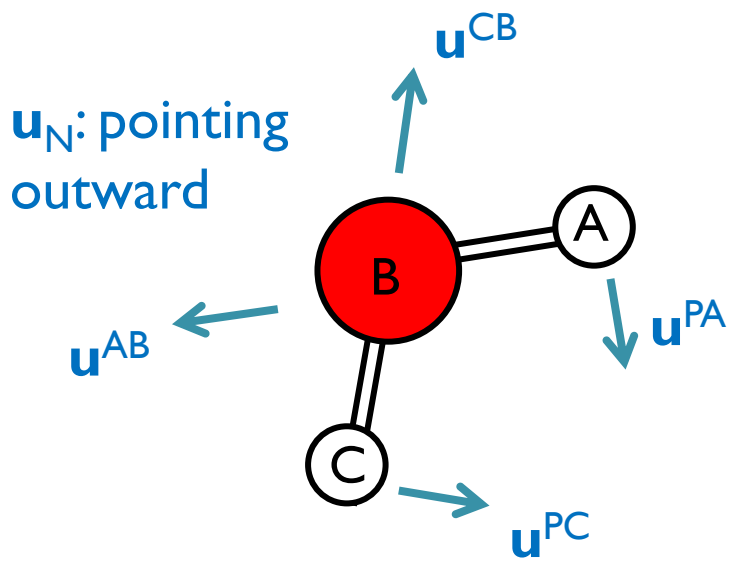


$$k_{bond} = \sum_{i=1}^{3} \lambda_i^{AB} \left| \hat{u}^{AB} \cdot \hat{v}_i^{AB} \right|$$

$$V_{bond} = \frac{1}{2} k_{bond} (r - r_0)^2$$

Angle bending force constant

- $\hat{u}_N$ is a unit vector perpendicular to the ABC plane
- The reaction forces on atoms A and C are perpendicular to AB and CB, respectively

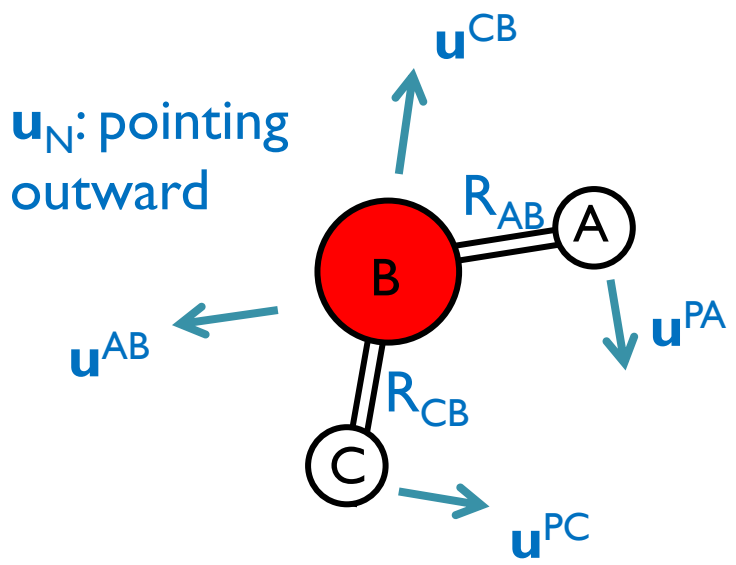


$$\hat{u}_N = \frac{\hat{u}^{CB} \times \hat{u}^{AB}}{\left| \hat{u}^{CB} \times \hat{u}^{AB} \right|}$$

$$\hat{u}^{PA} = \hat{u}_N \times \hat{u}^{AB}$$

$$\hat{u}^{PC} = \hat{u}^{CB} \times \hat{u}_N$$

Angle bending force constant



$$\hat{u}_N = \frac{\hat{u}^{CB} \times \hat{u}^{AB}}{\left|\hat{u}^{CB} \times \hat{u}^{AB}\right|}$$

$$\hat{u}^{PA} = \hat{u}_N \times \hat{u}^{AB}$$

$$\hat{u}^{PC} = \hat{u}^{CB} \times \hat{u}_N$$

$$\frac{1}{k_{angle}} = \frac{1}{R_{AB}^2 \sum_{i=1}^{3} \lambda_i^{AB} \left|\hat{u}^{PA} \cdot \hat{v}_i^{AB}\right|} + \frac{1}{R_{CB}^2 \sum_{i=1}^{3} \lambda_i^{CB} \left|\hat{u}^{PC} \cdot \hat{v}_i^{CB}\right|}$$

$$V_{angle} = \frac{1}{2} k_{angle} (\theta - \theta_0)^2$$

Coding tips

- The full Hessian matrix is saved in a lower triangle format in the fchk file, e.g.
  - $H_{00}, H_{10}, H_{11}, H_{20}, H_{21}, H_{22}, H_{30}, \ldots$
- The Hessian matrix is 3Nx3N and symmetric
- Use proper math libraries (e.g. GSL) to diagonalize the 3x3 matrix
- Be careful with unit conversion

Scripts

- Hess2FF.pl and getbonded.pl
  - Perl scripts to calculate bond/angle force constant from Hessian
  - Reference: J. M. Seminario, Int. J. Quantum Chem., 60, 1271-1277 (1996)

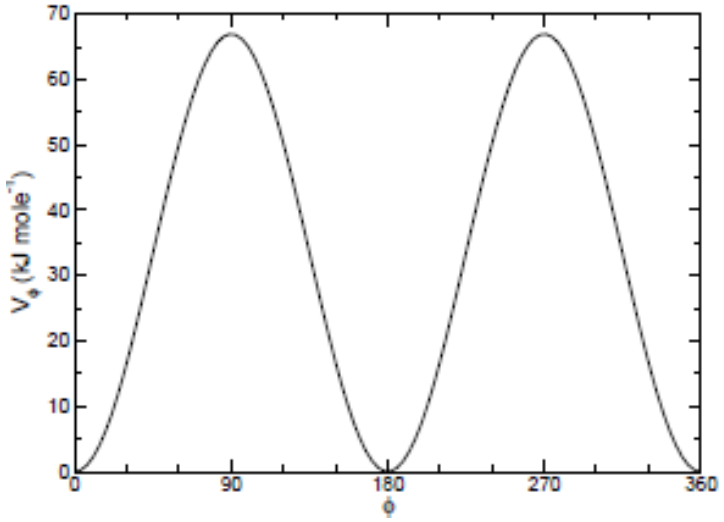
- Usage:
  - `perl getbonded.pl -top gmx_molecule.top -fchk molecule.fchk -name molecule`

Step 4:

- Run necessary quantum chemical calculations.
- Generate initial force field parameters, for example, using GAFF.
- Get bond stretching and angle bending force constants from Hessian.
- Refine the dihedral angle force constant by fitting to QM potential energy surface.
- Run MD simulation with new set of parameters.

Dihedral angle force constant

- Methodology: fitting molecular mechanical potential energy surface to quantum mechanical data



$$E_{d,ijkl} = \sum k_{d,ijkl}[1 + \cos(n\phi_{ijkl} - \phi_{ijkl,0})]$$

Relaxed potential energy scan

- Keyword in Gaussian 09:
  - Opt(Modredundant)

```
%mem=1000MB
%chk=scan.chk
#p B3LYP/6-311+G(d,p) Opt(Modredundant)
SCRF(PCM,Solvent=Water)

scan

0 1
C 0.0408210174 1.2072071597 1.3657881129
......
O 0.0235063728 1.0862639647 -2.74056842

1 5 12 13  S  9  10.0
```

Fitting MM energies to QM

- Extract the SCF energies and gather the geometries into a GROMACS gro file.
- Run MD simulation using the "-rerun" option to obtain the MM potential energy surface.
- Adjust the dihedral force constants so that the MM calculations could reproduce the QM results.

Step 5:

- Run necessary quantum chemical calculations.
- Generate initial force field parameters, for example, using GAFF.
- Get bond stretching and angle bending force constants from Hessian.
- Refine the dihedral angle force constant by fitting to QM potential energy surface.
- Run MD simulation with new set of parameters.

Run MD simulations

- Solvation
  - `genbox -ci gmx_molecule.gro -nmol 1 -box 2.2 2.2 2.2`
  - `genbox -cp out.gro -cs`
  - Edit the gmx_molecule.top file to include proper number of solvent molecules.

Run MD simulations

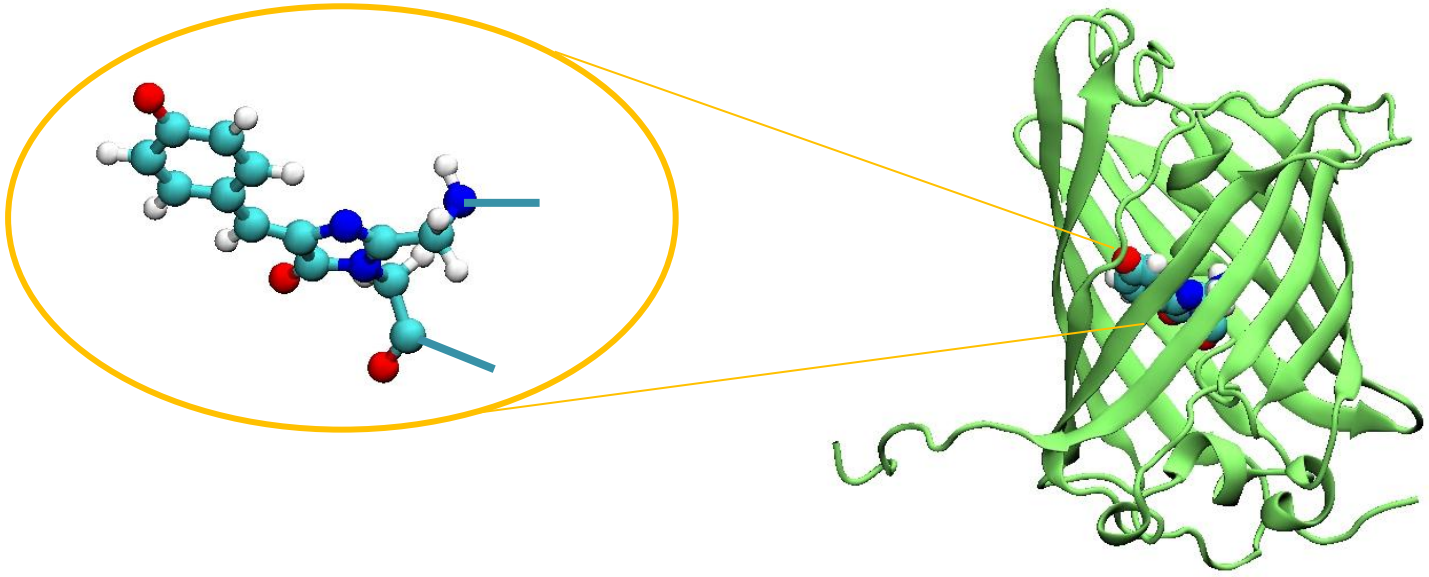
- Energy minimization
  - `grompp -f em.mdp -c gmx_molecule.gro -p gmx_molecule.top`
  - `mdrun -s topol.tpr`
  - `cp confout.gro after_em.gro`

Run MD simulations

- MD simulation
  - `grompp -f md.mdp -c after_em.gro -p gmx_molecule.top`
  - `mdrun -s topol.tpr`
  - Trajectory saved in traj.xtc
  - Energies saved in ener.edr

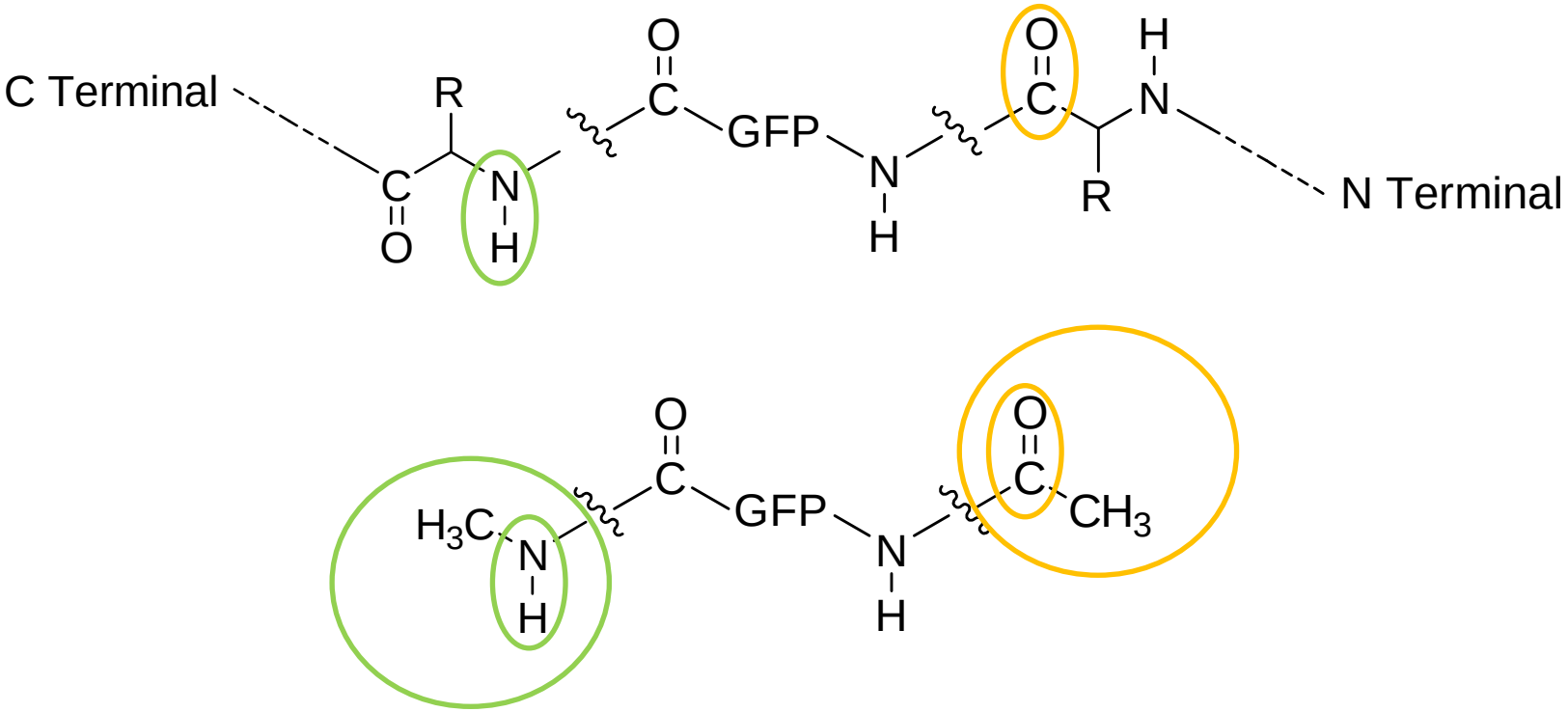
Further discussion: why do we need parameterization?

- Non-standard residue
  - Green fluorescent protein



Non-standard residue

- Derive charges for a non-standard residue



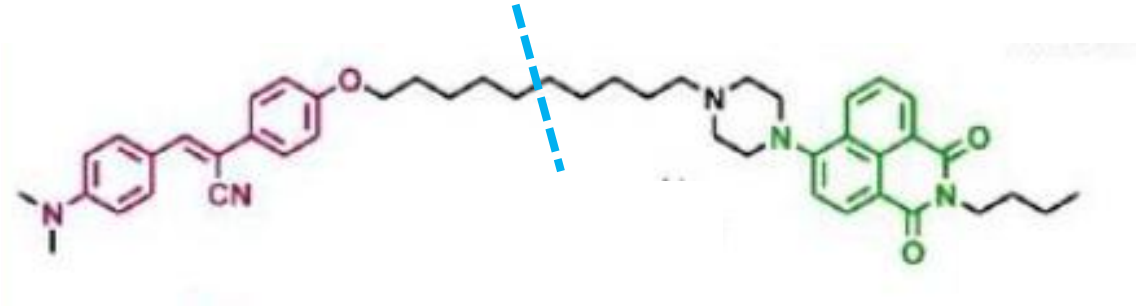
Non-standard residue

- respgen (AmberTools)
  - **-a** additional input data (predefined charges, atom groups, etc.)

```
//predefined charges
CHARGE -0.417500 7 N1
CHARGE 0.271900 8 H4
//charge groups
GROUP 4 0.00000
//atoms in the group
ATOM 7 N1
ATOM 8 H4
ATOM 9 C3
ATOM 10 H5
```

Non-standard residue

- Derive atomic partial charges for large molecule through fragmentation



Be aware: define proper fragments