Intro to Molecular Dynamics (MD) Simulation

using CHARMM

DAVID CHATFIELD, FIU DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY

WORKSHOP ON MACROMOLECULAR MODELING

FIU, APRIL 8-9, 2017

Overview

Principles of MD

MD Program Structure

System Preparation

Running MD

Analyzing Results

A. Force Fields

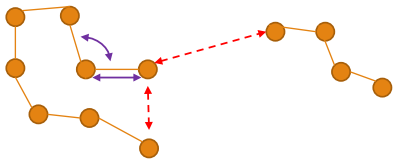
Quantum mechanics (QM): too expensive for macromolecules

Force fields: method of choice

- Constructed from experimental data and QM calculations on small molecules
- Principle of transferability → atom type

Bonded and Non-bonded Terms

$$V = E_{bonded} + E_{non-bonded}$$



E_{bonded}

$$E_{bonded} = E_{bond-stretch} + E_{angle-bend} + E_{dihedral} \; (+ \; \ldots)$$

$$E_{bond-stretch} = \sum_{bonds} K_b(b - b_0)^2$$

$$E_{angle-bend} = \sum_{angles} K_\theta(\theta - \theta_0)^2$$

$$E_{dihedral} = \sum_{dihedrals} K_\varphi\big(1 + \cos(n\varphi - \delta)\big)^2$$

A picture is worth a thousand words

E_{bonded}

Occasionally a few extra terms are included, e.g.

- $E_{impropers}$ (to maintain chirality & optimize fit to experiment)
- $E_{Urey-Bradley}$ (to optimize ..)
- CMAP (numerical correction and optimize ..)

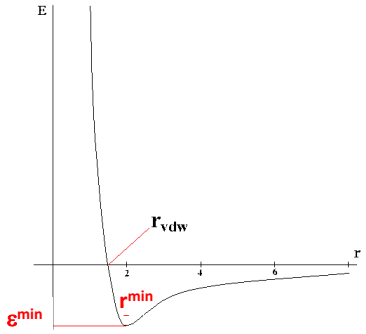
$$E_{impropers} = \sum_{impropers} K_\omega(\omega - \omega_0)^2$$

$$E_{Urey-Bradley} = \sum_{Urey-Bradley} K_{UB}(S - S)^2$$

$$E_{CMAP} = \sum_{residues} U_{CMAP}(\varphi, \psi)$$

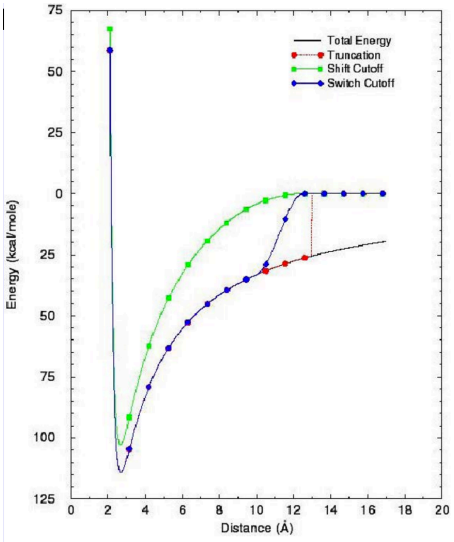
$E_{non-bonded}$

$$E_{non-bonded} = E_{van\ der\ Waals} + E_{electrostatic}$$



$$E_{vdW} = \sum_{\substack{non-bonded \\ atom\ pairs}} \varepsilon_{ij}^{min} \left[\left(\frac{r_{ij}^{min}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}^{min}}{r_{ij}} \right)^{6} \right]$$

$$E_{electrostatic} = \sum_{\substack{non-bonded \\ atom\ pairs}} \frac{q_i q_j}{4\pi\varepsilon_o \varepsilon r_{ij}}$$



Cut-offs

- Reduce computational demands
- Makes $E_{non-bond}$ precisely zero beyond pre-defined interatomic distance
- Shift method: modifies entire PES (downside: equilibrium distances slightly decreased)
- Switch method: modifies PES only over "window" (downside: too small a "window" can introduce strong, non-physical forces)
- Van der Waals: always used
- Electrostatics: sometimes used

Atom types

Atoms classified by TYPE

- E.g. sp^2 and sp^3 carbons: different TYPES

Parameters defined by TYPE

Compromise between:

- Different properties for every conceivable environment (impractical)
- Only one property set per element only (too inflexibe)

Carbon atom types (Parm22)

```
MASS  20 C    12.01100 C ! carbonyl C, peptide backbone
MASS  21 CA   12.01100 C ! aromatic C
MASS  22 CT1  12.01100 C ! aliphatic sp3 C for CH
MASS  23 CT2  12.01100 C ! aliphatic sp3 C for CH2
MASS  24 CT3  12.01100 C ! aliphatic sp3 C for CH3
MASS  25 CPH1 12.01100 C ! his CG and CD2 carbons
MASS  26 CPH2 12.01100 C ! his CE1 carbon
MASS  27 CPT  12.01100 C ! trp C between rings
```

Atom name

Atom type

Topology file entry

```
RESI ALA      0.00
GROUP
ATOM N     NH1    -0.47
ATOM HN    H       0.31
ATOM CA    CT1     0.07
ATOM HA    HB      0.09
GROUP                        !     |
ATOM CB    CT3    -0.27      !  HN-N
ATOM HB1   HA      0.09      !     |     HB1
ATOM HB2   HA      0.09      !     |    /
ATOM HB3   HA      0.09      !  HA-CA--CB-HB2
GROUP                        !        \
ATOM C     C       0.51      !        HB3
ATOM O     O      -0.51      !   O=C
BOND CB CA  N  HN  N  CA     !     |
BOND C  CA  C  +N  CA HA  CB  HB1  CB  HB2  CB  HB3
DOUBLE O  C
IMPR N -C CA HN  C CA +N O
DONOR HN N
ACCEPTOR O C
```

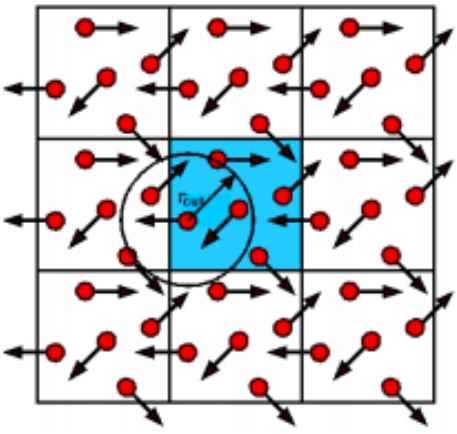
Parameter file entry

```
HA    CT3   322.000   1.1110
```

K_b (kcal/mol-Å^2)

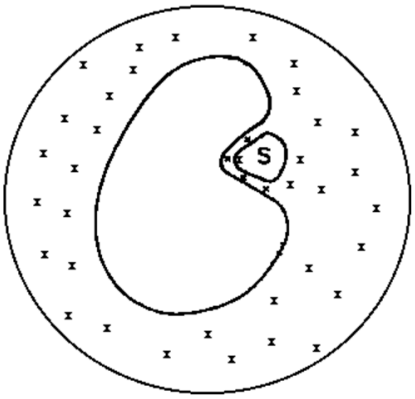
r_e (Å)

B. Solvation
Method 1: Periodic Boundary Conditions



- Central box surrounded by neighbors
- Only particles in central box modeled explicitly
- Particles in surrounding boxes are images
- Forces on primary particles include interactions with images
- Several lattices available: cubic, truncated octahedral, …
- $E_{electrostatics}$ can be calculated without cut-offs using clever series expansion: EWALD method
- Cut-offs needed for E_{VDW}

Method 2: Solvation Shell ("droplet")



- Sphere of surrounding waters added to macromolecule
- Boundary potential (BP) restrains waters
- Trade-off: BP too small → waters "evaporate"
- .. .. large → extraneous forces

C. MD Algorithms

Begin with Newton's equations of motion (1-D):

$$F = ma \qquad F = -\frac{dV}{dx}$$

Integrate to solve for $x(t)$ and $v(t)$:

$$v = at + v_0$$
$$x = vt + x_0$$

Discretize to create computer algorithm (Leap-Frog):

$$x(t + \delta t) = x(t) + v\left(t + \frac{1}{2}\delta t\right)\delta t$$
$$v\left(t + \frac{1}{2}\delta t\right) = v\left(t - \frac{1}{2}\delta t\right) + a(t)\delta t$$

Principles of MD

MD Program Structure

System Preparation

Running MD

Analyzing Results

Essential MD data structures

Parameter file

- Contains the force field: all bond, angle parameters etc.

Topology file

- Contains data for essential moieties (amino acid residues, DNA bases, small molecules etc.): what is bonded to what, partial atomic charges etc.

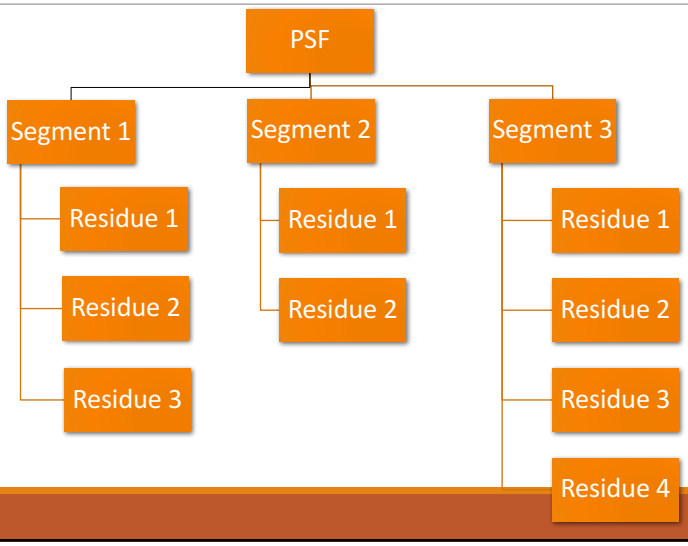
PSF

- Contains data for ENTIRE SYSTEM (amino acid residue or nucleotide sequence, solvent molecules; list of all bond, angle terms etc.)

Coordinate file

- Contains x,y,z coordinates of all atoms in system

MD structure hierarchy for physical system



MD anachronisms

PSF originally stood for Protein Structure File (the early development of MD codes focused on proteins).

But now used to describe any complete physical system.

Residue originally indicated amino acid residue.

But now refers more generally to substructures from which segments are made.

Nucleotide in DNA

Individual lipid molecule in lipid bilayer

Individual water in set of solvating waters

Principles of MD

MD Program Structure

System Preparation

Running MD

Analyzing Results

Structure of macromolecule

X-ray crystallography
- Have to build some missing atoms, including almost all hydrogens

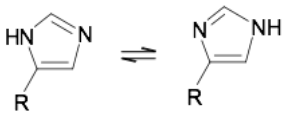
Neutron diffraction
- Expensive (uses synchrotron radiation) but resolves the hydrogens

Homology modeling

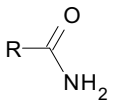
May need to complete or fix structure: Go through check list

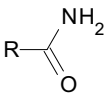
- ✓Add missing hydrogens (most not resolved in X-ray diffraction)
- ✓Add missing heavy atoms (occasionally not resolved)
- ✓Choose between duplicate sets of atoms (crystal structure disorder)
- ✓Adjust protonation states for pH being simulated

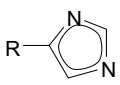
- ✓Adjust location of proton to create correct hydrogen bonding patterns

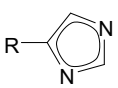


- ✓Where isoelectronic groups create experimental ambiguity, adjust for hydrogen bonding (affects His, Asn, Gln: 180° rotation may be needed)









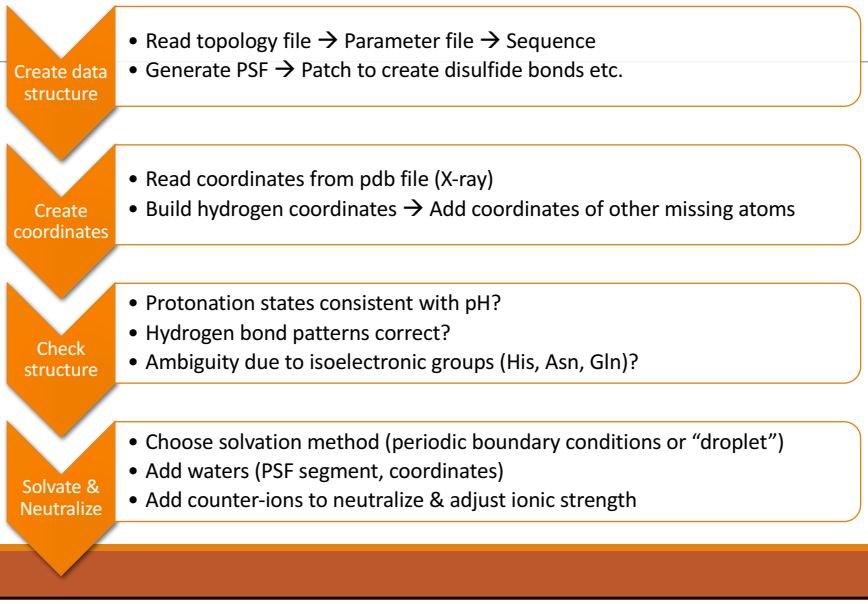
Finish preparing system

Solvate (add waters)

Make neutral (add counter-ions)

Adjust ionic strength? (more counter-ions)

Summary of preparation steps



Principles of MD

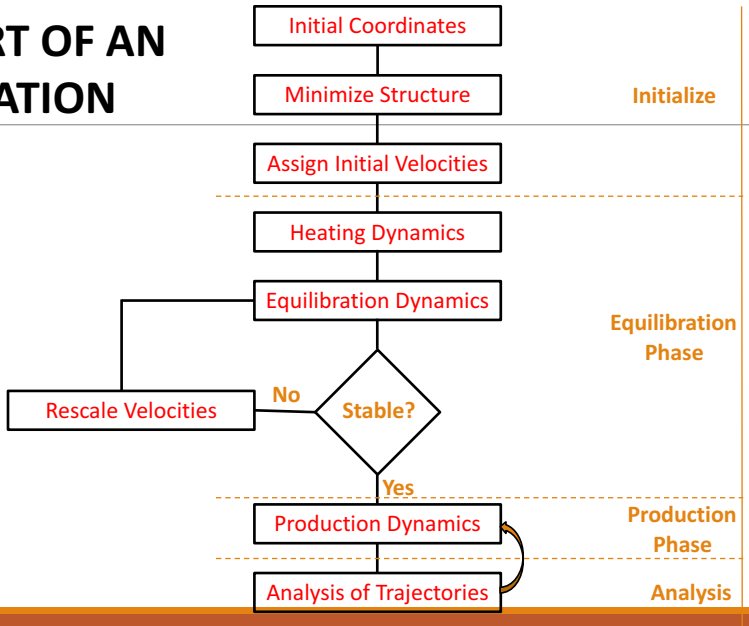
MD Program Structure

System Preparation

Running MD

Analyzing Results

FLOWCHART OF AN MD SIMULATION



Judging stability during equilibration

Temperature steady or drifting?

Energy fluctuations small (relative fluctuations < 10^{-4})?

RMSD of coordinates steady or drifting?

Principles of MD

MD Program Structure

System Preparation

Running MD

Analyzing Results

Analysis

Whatever you want!

- Equilibrium structure
- Sampling of hydrogen bonding patterns
- Loop conformations
- Stability of secondary structure
- Opening or closing of channels
- Migration of small ligands
- …