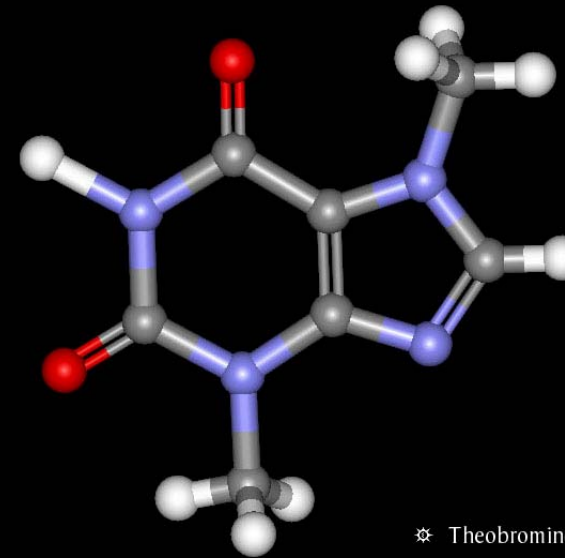# Bioengineering 215

An Introduction to Molecular Dynamics for Biomolecules

David Parker

May 18, 2007

Caffeine ☼

☼ Theobromin

Introduction

A principal tool to study biological molecules is molecular dynamics simulations (MD).

MD studies both thermodynamic properties and/or time dependent (kinetic) phenomenon of interacting atoms.

MD simulations permit the study of complex, dynamic processes that occur in biological systems on a scale too small or too fast to observe experimentally

- Protein stability

- Conformational changes

- Protein folding

- Molecular recognition: proteins, DNA, membranes, complexes

- Ion transport in biological systems

Molecular Mechanics

Simulating electron and nuclei motion using quantum mechanics is the most accurate for computing all properties from first principles (ab initio) but …

- Computationally expensive (number of particles, time scale 10^{-16})
- Don't need this level of detail for many applications
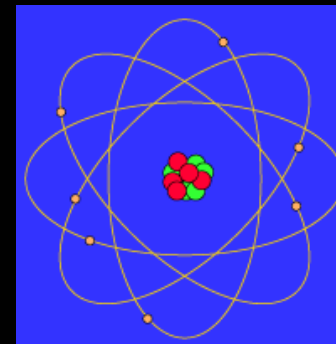
Molecular mechanics uses classical models of forces, geometry and motions of atoms

- Model atom as a sphere without considering electrons explicitly
- Treat a molecule as a collection of interconnected masses
  - masses represent the nuclei
  - connections represent bonds between atoms
- Energy and geometry of a molecule are approximated using a force field
  - collection of atom types
  - parameters (for bond lengths, bond angles, etc.)
  - equations (to calculate the energy between atoms)
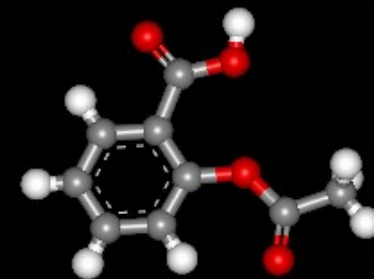
Molecular Dynamics

Molecular dynamics is based on Newton's second law

- The equation of motion for each atom in a system is $\mathbf{F}_i = m_i \mathbf{a}_i$

- Knowing the force on each atom you can compute its acceleration

- Integration of the equations of motion yields a trajectory that describes the positions, velocities and accelerations of the atoms as they vary with time

- Can convert this microscopic information to macroscopic observables such as pressure, energy, heat capacities, etc., using statistical mechanics (averaging)

The force $F(\mathbf{R})$ on each atom is computed from a potential function $U(\mathbf{R})$

- Function of the atomic positions, R, of all the atoms in the system $\mathbf{R} = \{ \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N \}$

- Sum of bonded and non-bonded terms (assume pair-additive)

- The force on atom i is the gradient of the potential function $\mathbf{F}_i = -\text{grad}_i\, U(\mathbf{R})$

- Form and parameters for U(R) is defined by a particular MD force field

MD simulations are computationally expensive

- Small time step (2 fs) limits simulations to ns range
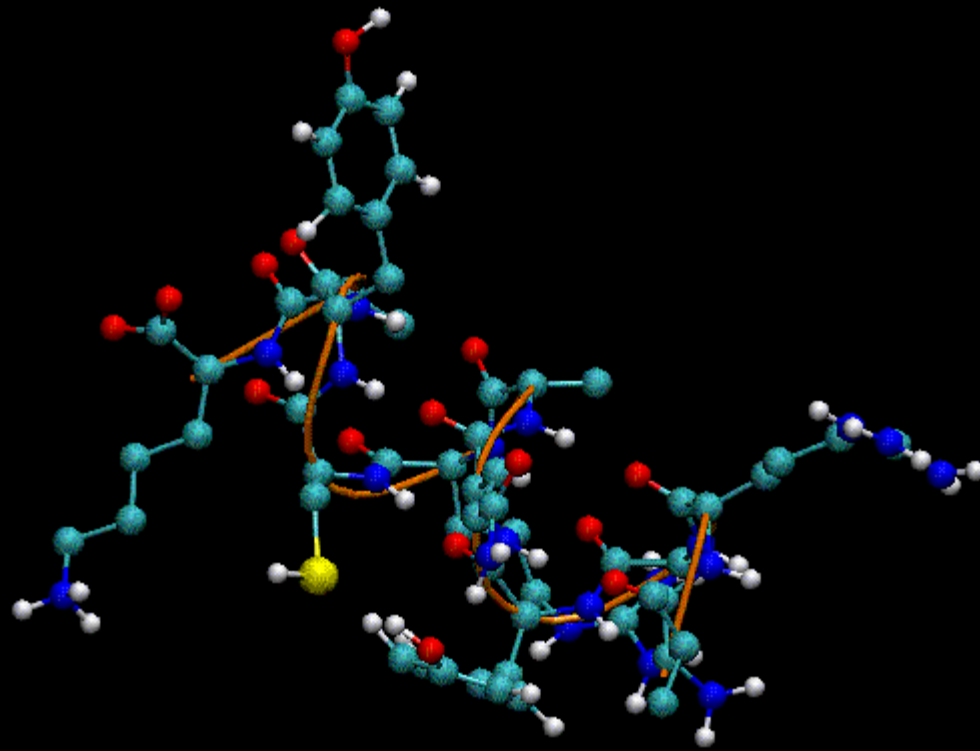
Molecular Dynamics

Example: 12 ns MD simulation of calmodulin binding to a target peptide

- 177 amino acids (1827 atoms)

- 13,833 water molecules (41,784 atoms)

- Compute time: 4 days on 6 CPU machine (2.8 ns / day)

- Actual binding time ~ 2.8 ms   (1,000,000 days = 2810 years to compute)
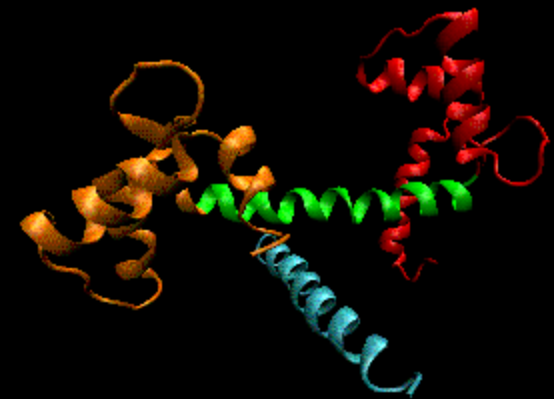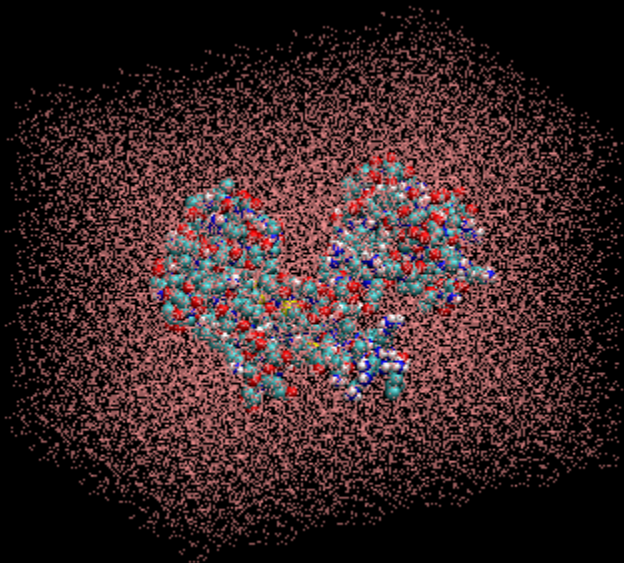
Molecular Dynamics

Example: 100 fs MD simulation of 11 amino acid helix.

Time and Length Scales

Biological molecules exhibit a wide range of time scales over which specific processes occur. Motions are chaotic due to thermal forces.

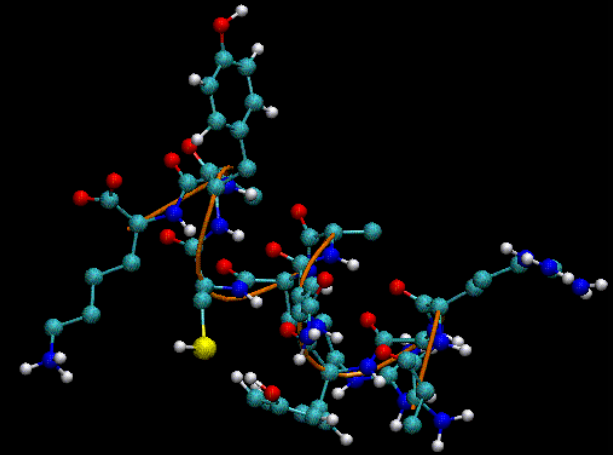Local Motions (0.01 to 5 Å, 10^{-15} to 10^{-1} s)

- Atomic fluctuations
- Sidechain Motions
- Loop Motions

Rigid Body Motions (1 to 10Å, 10^{-9} to 1s)

- Helix Motions
- Domain Motions (hinge bending)
- Subunit motions

Large-Scale Motions (> 5Å, 10^{-7} to 10^4 s)

- Helix coil transitions
- Dissociation/Association
- Folding and Unfolding

Units

Basic units

Length: nanometer (nm) = 10^{-9} m

Mass: atomic mass unit (amu) = $1.6605402 \times 10^{-27}$ kg

Time: picosecond (ps) = 10^{-12} s

Charge: electronic charge = $1.60217733 \times 10^{-19}$ C

Temperature: K

Derived units

Energy: kJ mol^{-1} = 7 pN nm = 7×10^{-21} N m

Force: kJ mol^{-1} nm^{-1}

Velocity: nm ps^{-1} = 1000 m/s

Thermal energy = $k_B T$ = 4 pN nm at 298K

Potential Energy Function

Potential energy is a function of atoms, from:

- Bonded neighbors (covalent bonds)
- Non-Bonded atoms
    - other atoms in the same molecule
    - atoms from different molecules (e.g. water)

The total potential energy U is a sum over appropriate atom groups

$$U = U_{bonded} + U_{non\text{-}bonded}$$

$$U_{bonded} = U_{bond\text{-}stretch} + U_{bond\text{-}bending} + U_{bond\text{-}rotation}$$

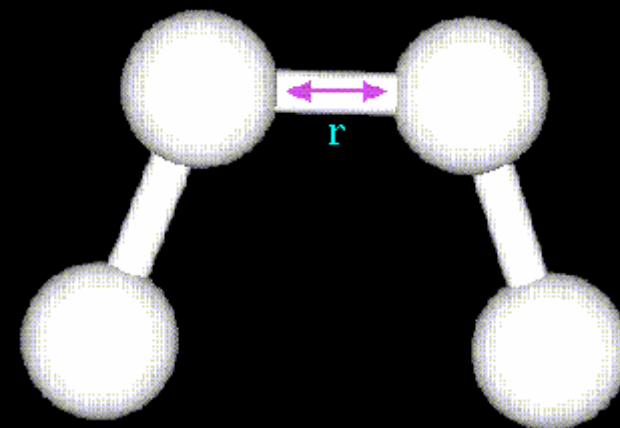$$U_{non\text{-}bonded} = U_{electrostatic} + U_{van\text{-}der\text{-}Waals}$$

Potential Energy Function: Bond Stretch

The bond stretching energy equation is based on Hooke's law

$$U_{bond\text{-}stretch} = \tfrac{1}{2} K_b (r - r_0)^2$$

K_b = bond stiffness
r_o = reference length

Example: sp3 C – sp3 C

K_b = 310.0 kcal/(mol Å^2)

r_o = 1.526 Å

Source: experiment

Reality

Optimum
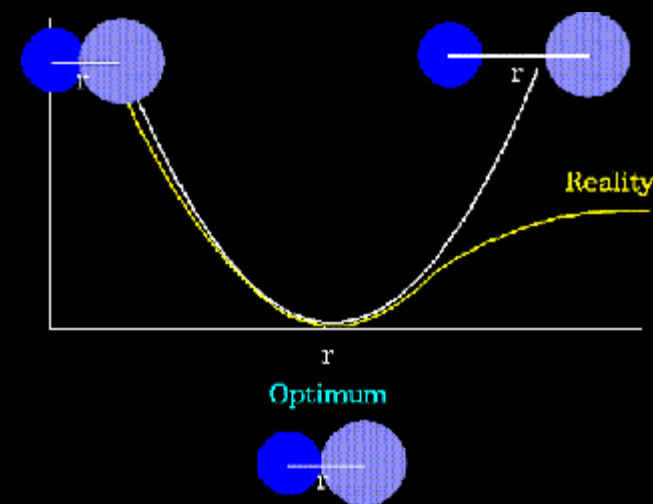
Potential Energy Function: Bond Bending

The bond bending energy equation is based on Hooke's law

$$U_{bond-bend} = \tfrac{1}{2} K_\theta (\theta - \theta_0)^2$$
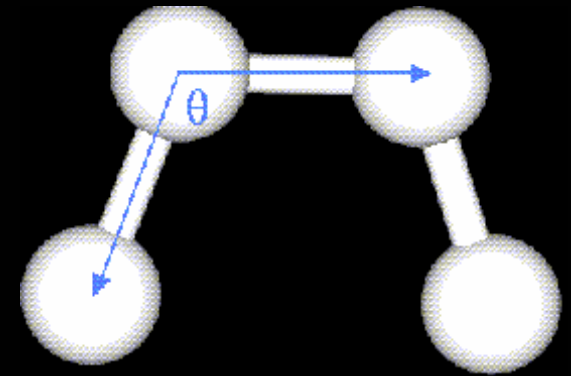
K_θ = bond stiffness
θ_0 = reference angle

Example: sp3 C – sp3 C – sp3 C

$\theta_0 = 109.50°$

$K_\theta = 40.0$ kcal/(mol radian^2)

Source: experiment
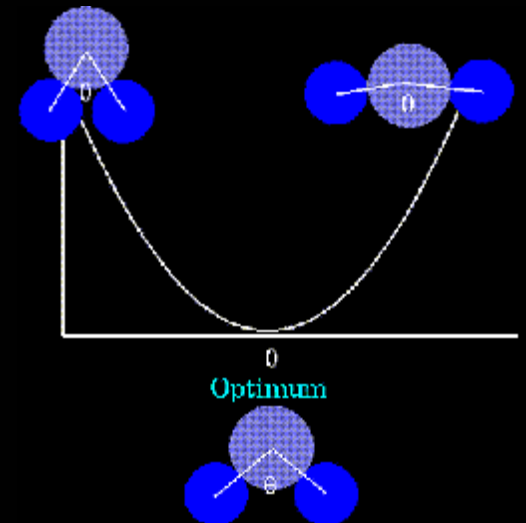
Potential Energy Function: Bond Rotation

The bond rotational energy equation is modeled as a periodic function

$$U_{\text{bond-rotation}} = \frac{1}{2} K_\tau [1 + \cos(n\tau - \varphi)]$$

K_τ = barrier to rotation
τ = torsion angle
n = number of minima
φ = phase angle

Example: CH3-CH3 (ethane)

K_τ = 1.4 kcal/mol

n = 3

φ = 0

Source: experiment
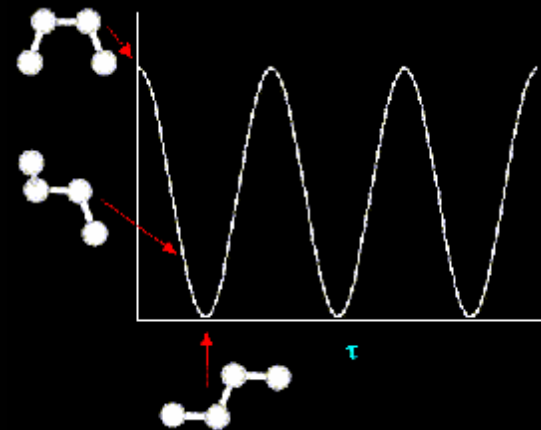
Potential Energy Function: van der Waals

Atoms with no net electrostatic charge attract each other at short distances

- Fluctuation of electrons create charge separation (dipole)
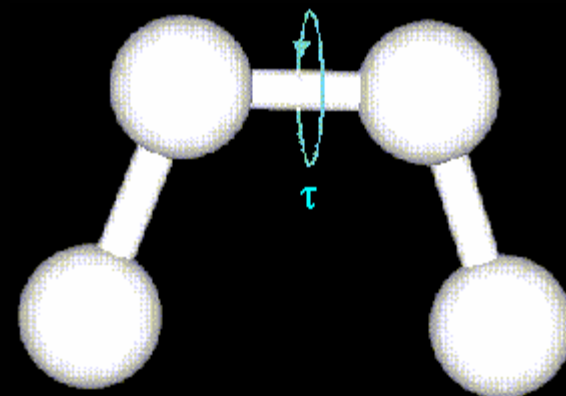- This can create an induced dipole in an approaching atom
- Atoms will be attracted to each other
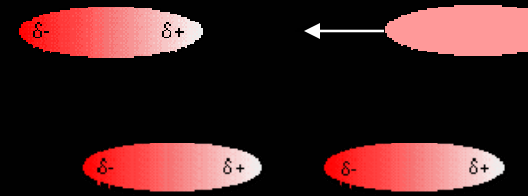
Model using the Lennard-Jones potential for two atoms at distance r

$$U(r) = 4\epsilon \left[\left(\frac{A}{r}\right)^{12} - \left(\frac{B}{r}\right)^{6} \right]$$

- r^6 term is attractive energy

- r^{12} term repulsive energy (electron cloud overlap)

- A, B and ε describe shape of potential (from quantum mechanics)

- Energy decreases rapidly with distance

$$- \frac{A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}}$$

Repulsion regime

van der Waals attraction regime

r_{ij}

Optimum energy
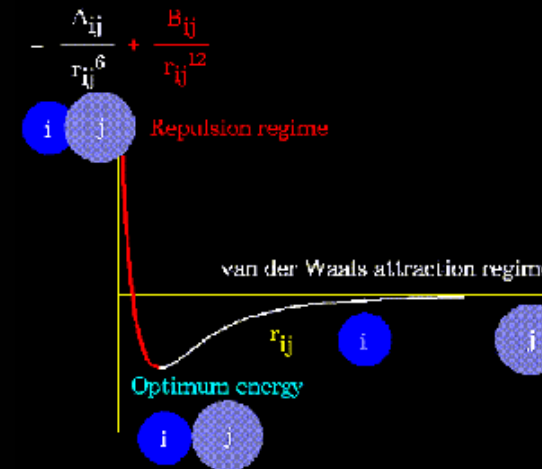
Potential Energy Function: Electrostatics

Coulomb's law models electrostatic forces with charges treated as point charges centered on atoms.

For two charges q_1 and q_2 separated by distance r

$$U(r) = \frac{q_1 q_2}{4\pi\epsilon_0 r}$$

$1/4\pi\epsilon_0$ is the electrostatic constant.

Electrostatic energy decreases slowly and affects atoms over a significant distance.

Potential Energy Function

$$
U(\mathbf{r}^N) = \sum_{bonds} \frac{K_i}{2}(r_i - r_{i0})^2 + \sum_{angles} \frac{\Theta_i}{2}(\theta_i - \theta_{i0})^2
$$

$$
+ \sum_{torsions} \frac{V_n}{2}(1 + \cos(n\omega - \gamma))
$$

$$
+ \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\epsilon_{ij} \left[\left(\frac{A_{ij}}{r_{ij}}\right)^{12} - \left(\frac{B_{ij}}{r_{ij}}\right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right)
$$

Bond stretching and bending DOF are very stiff so energy and conformational changes for a molecule are mainly due to torsional and non-bonded energy.

Energy Term	Scale (kcal/mol)
Bond stretching	100
Angle Bending	10
Torsion	1
Hydrogen Bond	2
Electrostatic interaction	0.5
Van der Waals	0.1

Force Fields

A force field refers to the functional form and parameters used to describe the potential energy of a system of atoms. Parameters are derived using small organic molecules from

- Experiments (bond lengths, spring constants)

- Quantum mechanical calculations (atom partial charges)

Parameters depend on

- Atom type

- Bonding patterns

Parameters include values for
- atomic mass
- partial charge for individual atoms
- equilibrium bond lengths and angles

Different force fields are included in some MD software packages
 Gromacs has: AMBER, GROMOS, OPLS, ENCAD

Force Fields: Amber

AMBER has 40 atom types.

- 13 types of carbon: sp3 carbon, carbonyl sp2 carbon, aromatic sp2 carbon, …

- Bond stretch and angle parameters for each valid combination of atom types

- 1 to 3 torsion parameters for many combinations of atoms and there are 30 improper torsions

- Van der Waals parameters for each atom type (combined for each pairwise interaction)

Bond Parameters

| bond	K_r^b	r_{eq}^c	bond	K_r^b	r_{eq}^c	bond	K_r^b	r_{eq}^c	bond	K_r^b	r_{eq}^c
C−CA	469.0	1.409	CA−HA	367.0	1.080	CM−HA	367.0	1.080	CT−S	227.0	1.810
C−CB	447.0	1.419	CA−N2	481.0	1.340	CM−N*	448.0	1.365	CT−SH	237.0	1.810
C−CM	410.0	1.444	CA−NA	427.0	1.381	CN−NA	428.0	1.380	CV−H4	367.0	1.080
C−CT	317.0	1.522	CA−NC	483.0	1.339	CQ−H5	367.0	1.080	CV−NB	410.0	1.394
C−N	490.0	1.335	CB−CB	520.0	1.370	CQ−NC	502.0	1.324	CW−H4	367.0	1.080
C−N*	424.0	1.383	CB−CN	447.0	1.419	CR−H5	367.0	1.080	CW−NA	427.0	1.381
C−NA	418.0	1.388	CB−N*	436.0	1.374	CR−NA	477.0	1.343	H−N	434.0	1.010
C−NC	457.0	1.358	CB−NB	414.0	1.391	CR−NB	488.0	1.335	H−N*	434.0	1.010
C−O	570.0	1.229	CB−NC	461.0	1.354	CT−CT	310.0	1.526	H−N2	434.0	1.010
C−O2	656.0	1.250	CC−CT	317.0	1.504	CT−F	367.0	1.380	H−N3	434.0	1.010
C−OH	450.0	1.364	CC−CV	512.0	1.375	CT−H1	340.0	1.090	H−NA	434.0	1.010
C*−CB	388.0	1.459	CC−CW	518.0	1.371	CT−H2	340.0	1.090	HO−OH	553.0	0.960
C*−CT	317.0	1.495	CC−NA	422.0	1.385	CT−H3	340.0	1.090	HO−OS	553.0	0.960
C*−CW	546.0	1.352	CC−NB	410.0	1.394	CT−HC	340.0	1.090	HS−SH	274.0	1.336
C*−HC	367.0	1.080	CK−H5	367.0	1.080	CT−HP	340.0	1.090	O2−P	525.0	1.480
CA−CA	469.0	1.400	CK−N*	440.0	1.371	CT−N	337.0	1.449	OH−P	230.0	1.610
CA−CB	469.0	1.404	CK−NB	529.0	1.304	CT−N*	337.0	1.475	OS−P	230.0	1.610
CA−CM	427.0	1.433	CM−CM	549.0	1.350	CT−N2	337.0	1.463	OW−HW	553.0	0.9572
CA−CN	469.0	1.400	CM−CT	317.0	1.510	CT−N3	367.0	1.471	S−S	166.0	2.038
CA−CT	317.0	1.510	CM−H4	367.0	1.080	CT−OH	320.0	1.410			
CA−H4	367.0	1.080	CM−H5	367.0	1.080	CT−OS	320.0	1.410			

MD Simulation (Big Picture)

The key ingredients in an MD simulation are

1. Initial conditions

   - Initial structure: atoms, coordinates, topology

   - Initial atom velocities (temperature)

   - Add water, ions

2. Boundary conditions

   - How to simulate finite system

3. Energy Minimization

4. Computing forces given an interaction potential

   - Efficiently compute atom-atom forces

5. Numerical integration of equations of motion

   - Efficiently and accurately integrate equations of motion

6. Extracting useful data from simulation
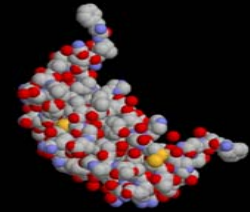
Initial Structure

Initial atom types and coordinates from protein X-ray or NMR crystallographic structures
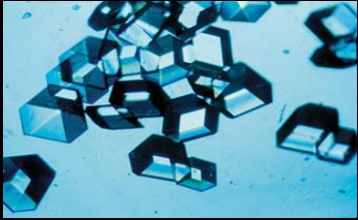
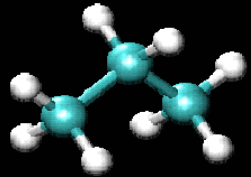
Insulin crystals (1mm)
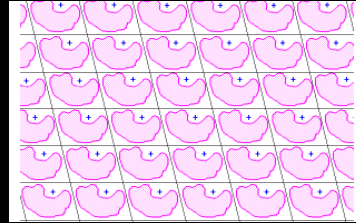
2D section of crystal lattice

PDB 1B9E

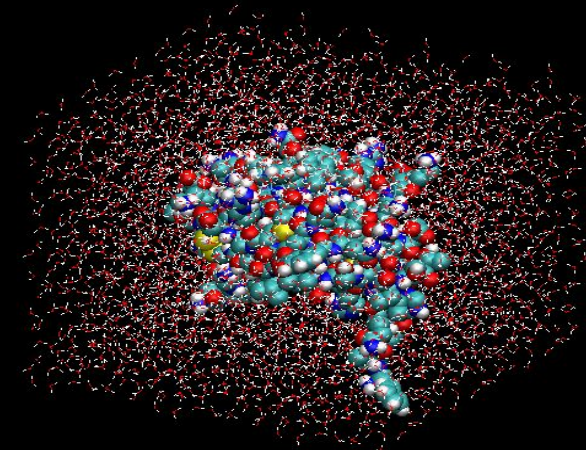Compute atom connectivity for force field computations

Propane C_3H_8

10 bonds,   18 valence angles,   18 torsional terms, and 27 non-bonded terms

Total = 27 terms

Add hydrogen atoms, create a box around the protein, and fill it with water.

Initial Velocities

Initial velocities for each atom can be assigned via statistical mechanics

- At temperature T, the average velocity of an atom is $<v_i^2> = k_B*T / m$

- Average kinetic energy of an atom is $<1/2\ m\ |v|^2> = 3/2\ k_BT$

- To scale velocities to a target temperature $T^\#$

  equate original velocities $<v_i^2> = k_BT / m$

  with scaled velocities $<(sv_i)^2> = s^2 <v_i^2> k_BT^\# / m$

  to get $s = sqrt(T^\# / T)$

We can assign random initial velocities using the algorithm:

1. For each atom i and direction j assign velocity v_{ij} a uniform random number from [-1, 1]

2. Remove mass center motion (subtract average velocity from from each atom)

3. Compute $<v_{ij}^2>$ and multiply v_{ij} by sqrt($k_BT^\#$ / $<mv_{ij}^2>$)

Boundary Conditions

In biological systems, proteins are immersed in a sea of water molecules so want to place in a box and fill it with water.

A box of water interfacing to a vacuum will have significant surface tension. How can these surface effects be eliminated?

Surround simulation box with virtual images of itself

- Effectively creates an infinite system
- Atoms interact with atoms in virtual images
- Atoms leaving one side appear on other
- Ignore atoms at large distances from each other
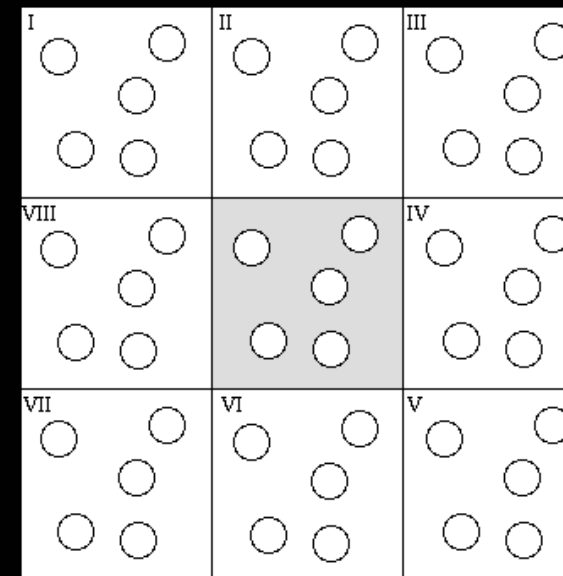
Energy Minimization

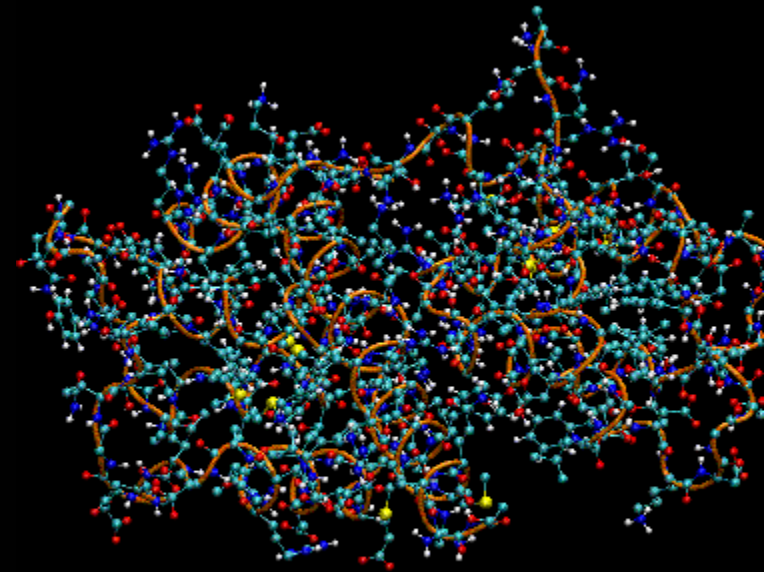Before running an MD simulation we need to perform an energy minimization to eliminate excessively large forces from

- Starting configuration

- Adding water

Steepest descent

- Robust and easy to implement

- Move atom coordinates in the direction –grad U

    $$\mathbf{r}_{new} = \mathbf{r}_{old} + h*\mathbf{F}$$

- Compute **F** from new configuration and repeat

- Slow convergence

The conjugate gradient method converges faster by using gradient information from previous steps.

Computing Forces

Computing non-bonded forces in an MD simulation is expensive

- Need to compute forces between all atoms

- Ignore certain forces for atoms at large distances from each other (cutoff)

- Electrostatic force is long-range so can't use small cutoff

Speed up non-bonded force calculation by using neighbor lists

- For each atom, store list of atoms within cutoff range

- Update lists as atoms move closer and further away
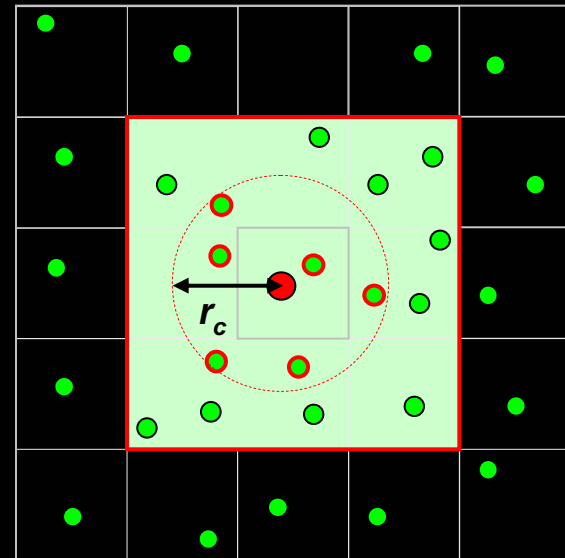
Integration Algorithms

An MD simulation solves Newton's equations of motion for a system of N interacting atoms with mass m_i and position $\mathbf{r}_i$

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = \mathbf{F}_i(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) \quad i = 1, 2, \ldots, N$$

$$\mathbf{F}_i(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = -\nabla_{\mathbf{r}_i} U(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$$

Numerical integration of the equations of motion will advance the system by a small time step dt

- Forces on the system are considered approximately constant during dt
- Coordinates as a function of time represent a trajectory of the system
- Energy and momentum should be conserved

Integration Algorithms

For MD, we want a fast and accurate integrator with:

1. High accuracy

2. Large range of stability (wrt time step)

3. Reversibility (wrt time)

4. Speed (computation time per step)

5. Low memory requirement

6. Conserve energy

7. Simple form and easy to implement

In MD, forces are rapidly changing, (e.g., van der Waals forces) so stability is very important.

Integration Algorithms

Euler's Method: $x(t + dt) = x(t) + x'(t)dt$

- Low accuracy

- Unstable integrating nonlinear terms of potential

2nd Order Runge-Kutta Method:

$$k1 = f(t_i, x_i)dt \qquad k2 = f(t_i + t, x_i + k1)dt$$

$$y_i+1 = y_i + a*k1 + b*k2$$

- Good accuracy and stability

- Expensive: need multiple evaluations of forces

- Poor energy conservation

- Not time reversible

Integration Algorithms

Verlet Method:

$$r(t + dt) = r(t) + v(t)dt + \tfrac{1}{2}\, a(t)dt^2$$

$$v(t + dt/2) = v(t) + \tfrac{1}{2}\, a(t)dt$$

$$v(t + dt) = v(t + dt/2) + \tfrac{1}{2}\, a(t+dt)dt$$

- Good accuracy and stability

- Time reversible

- Conserves energy

- Single force evaluation per time step

Verlet belongs to a class of integrators called symplectic integrators that conserve energy, are time reversible and conserve phase space area.

Analyzing Results

What are the results of an MD simulation?

- Trajectory: atom positions, velocities

- System potential and kinetic energy

- Energy between groups of atoms

- Temperature, pressure, density, …

What to look at

- Animations of trajectory

- Geometric measures

  - Root Mean-Squared Deviation (RMSD): measure structural similarity

  $$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(v_{ix} - w_{ix})^2 + (v_{iy} - w_{iy})^2 + (v_{iz} - w_{iz})^2}$$

  - Radius of gyration: useful estimate of the size of a molecule

  $$\frac{1}{N}\sum_{k=1}^{N}(\mathbf{r}_k - \mathbf{r}_{mean})^2$$

- Energy, temperature, pressure plots

MD Algorithm

The steps in an MD simulation are:

1. Input initial conditions

    - Interaction potential U as function of atom positions
    - Positions **r** of all atoms
    - Velocities **v** of all atoms

2. Compute forces on all atoms **F** = -grad U

    - Compute forces between non-bonded atoms
    - Compute forces between bonded atoms

3. Update atoms positions

    - Compute acceleration using **a** = **F** / m
    - Solve for velocity using d**v** / dt = **a**
    - Solve for position using d**x** / dt = **v**

$$a_{new} = F / m$$

$$r_{new} = r_{old} + v_{old} + \tfrac{1}{2} a_{old}dt^2$$

$$v_{1/2} = v_{old} + \tfrac{1}{2} a_{old}dt$$

$$v_{new} = v_{1/2} + \tfrac{1}{2} a_{new}dt$$

4. Output trajectory, energy

5. Go to 2 until computer explodes or universe ends

6. Analyze results

Example: Argon

Argon is an inert element that interacts via van der Waals energy only

$$V(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \qquad \mathbf{F}(r) = -\nabla V(r) = -\frac{d}{dr}V(r)\hat{\mathbf{r}} = 4\epsilon \left(12\frac{\sigma^{12}}{r^{13}} - 6\frac{\sigma^{6}}{r^{7}} \right) \hat{\mathbf{r}}$$
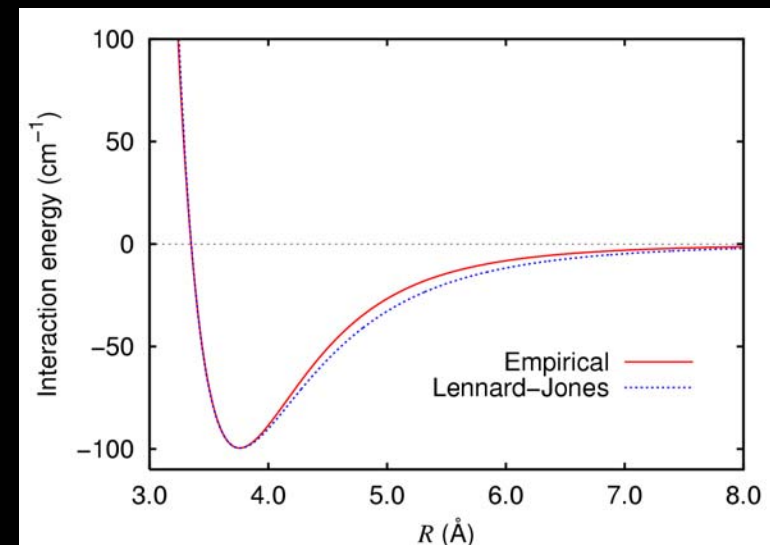
The appropriate parameter values for this system are

- $\epsilon = 1.03235 \times 10^{-2}$ eV
- $\sigma = 3.405$ Å
- m = 39.948 amu

For MD use the following basic units

- eV be energy unit
- Å length unit
- amu be mass unit
- fs be time unit

Forces will then be expressed in units eV / Å.

References

T. Schlick, "Molecular Modeling and Simulation", Springer, 2003.

A. R. Leach, "Molecular Modeling. Principles and Applications, Prentice Hall, 2001.

Molecular Dynamics Tutorial at EMBnet
   http://www.ch.embnet.org/MD_tutorial/index.html