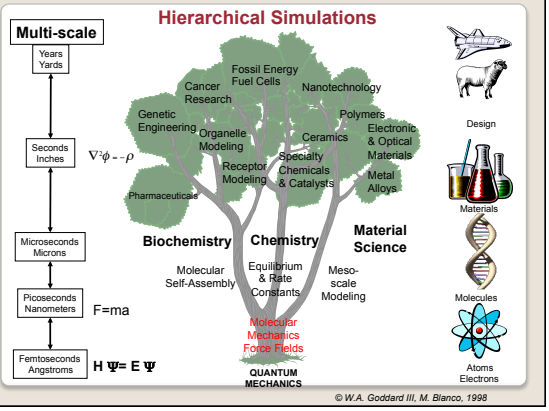
Simulations

Patrice Koehl
Department of Biological Sciences
National University of Singapore

http://www.cs.ucdavis.edu/~koehl/Teaching/BL5229
koehl@cs.ucdavis.edu

Hierarchical Simulations



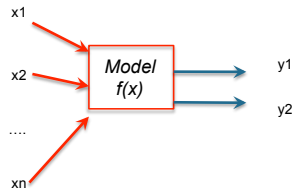
Simulations

Deterministic simulations: Molecular dynamics

Stochastic simulations: Monte Carlo

Simulations

Deterministic simulations: Molecular dynamics



What is an atom?

- Classical mechanics: a point particle
- Defined by its position (x,y,z) and its mass
- May carry an electric charge (positive or negative), usually partial (less than an electron)

Computing energy

$$U = \sum_{all\ bonds} \frac{1}{2} K_b (b - b_0)^2$$ *Bonds*

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

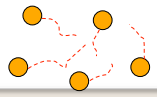
$$+ \sum_{all\ torsions} K_\phi [1 - \cos(n\phi)]$$ *Torsion angles*

$$+ \sum_{i,j\ nonbonded} \varepsilon_{ij} \left[\left(\frac{R_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}}{r_{ij}} \right)^{6} \right]$$ *Vdw interactions*

$$+ \sum_{i,j\ nonbonded} \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r_{ij}}$$ *Electrostatics*

What is a molecular dynamics simulation?

- Simulation that shows how the atoms in the system move with time
- Typically on the nanosecond timescale
- Atoms are treated like hard balls, and their motions are described by Newton's laws.



Why MD simulations?

- Link physics, chemistry and biology
- Model phenomena that cannot be observed experimentally
- Understand protein folding…
- Access to thermodynamics quantities (free energies, binding energies,…)

How do we run a MD simulation?

- **Get the initial configuration**

  From x-ray crystallography or NMR spectroscopy

- **Assign initial velocities**

  At thermal equilibrium, the expected value of the kinetic energy of the system at temperature T is:

$$\langle E_{kin} \rangle = \frac{1}{2} \sum_{i=1}^{3N} m_i v_i^2 = \frac{1}{2}(3N) k_B T$$

  This can be obtained by assigning the velocity components vi from a random Gaussian distribution with mean 0 and standard deviation ($k_B T/m_i$):

$$\langle v_i^2 \rangle = \frac{k_B T}{m_i}$$

How do we run a MD simulation?

- **For each time step:**

  Compute the force on each atom:

  $$F(X) = -\nabla E(X) = -\frac{\partial E}{\partial X}$$

  X: cartesian vector of the system

  Solve Newton's 2nd law of motion for each atom, to get new coordinates and velocities

  $$M\ddot{X} = F(X)$$

  M diagonal mass matrix

  .. means second order differentiation with respect to time

  Store coordinates

- Stop

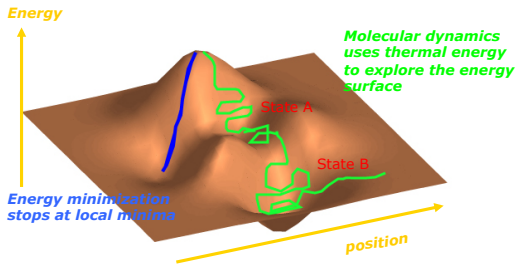
Newton's equation cannot be solved analytically:
⟶ Use stepwise numerical integration

What the integration algorithm does

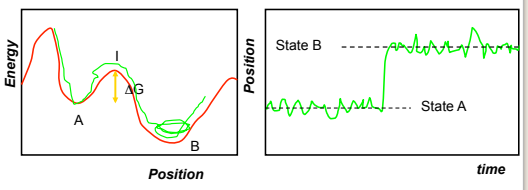
- *Advance the system by a small time step Dt during which forces are considered constant*
- *Recalculate forces and velocities*
- *Repeat the process*

If Δt is small enough, solution is a reasonable approximation

MD as a tool for minimization



Crossing energy barriers



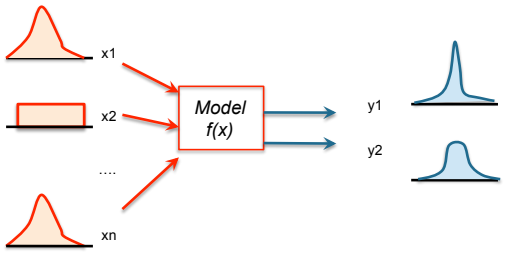
The actual transition time from A to B is very quick (a few pico seconds).

What takes time is waiting. The average waiting time for going from A to B can be expressed as:

$$\tau_{A\to B} = Ce^{\frac{\Delta G}{kT}}$$

Simulations

Stochastic simulations: Monte Carlo



Monte Carlo: random sampling

A simple example:

Evaluate numerically the one-dimensional integral:

$$I = \int_a^b f(x)dx$$

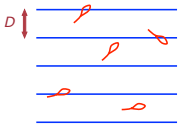
Instead of using classical quadrature, the integral can be rewritten as

$$I = (b-a)\langle f(x)\rangle$$

<f(x)> denotes the unweighted average of f(x) over [a,b], and can be determined by evaluating f(x) at a large number of x values randomly distributed over [a,b]

Monte Carlo method!

A famous example: Buffon's needle problem



The probability that a needle of length L overlaps with one of the lines, distant from each other by D, with L≤D is:

$$P = \frac{2L}{\pi D}$$

For L = D

$$P = \frac{2}{\pi}$$

Method to estimate π numerically:
"Throw" N needles on the floor, find needles that cross one of the line (say C of them). An estimate of p is:

$$\pi = 2\frac{N}{C}$$

Buffon, G. Editor's note concerning a lecture given by Mr. Le Clerc de Buffon to the Royal Academy of Sciences in Paris. Histoire de l'Acad. Roy. des Sci., pp. 43-45, 1733.
Buffon, G. "Essai d'arithmétique morale." Histoire naturelle, générale er particulière, Supplément 4, 46-123, 1777

Monte Carlo Sampling for protein structure

The probability of finding a protein (biomolecule) with a total energy E(X) is:

$$P(X) = \frac{\exp\left(-\frac{E(X)}{kT}\right)}{\int \exp\left(-\frac{E(Z)}{kT}\right) dZ}$$

→ ***Partition function***

Estimates of any average quantity of the form:

$$\langle A \rangle = \int A(X) P(X) dX$$

using uniform sampling would therefore be extremely inefficient.

→ Metropolis and coll. developed a method for directly sampling according to the actual distribution.

Metropolis et al. Equation of state calculations by fast computing machines. J. Chem. Phys. 21:1087-1092 (1953)

Monte Carlo Sampling for protein structure

Let: $P(X) = \frac{\exp\left(-\frac{E_p(X)}{kT}\right)}{\int \exp\left(-\frac{E_p(Z)}{kT}\right) dZ}$

And let $\pi(X \to Y)$ be the transition probability from state X to state Y.

Let us suppose we carry out a large number of Monte Carlo simulations, such that the number of points observed in conformation X is proportional to N(X).
The transition probability must satisfy one obvious condition: it should not destroy this equilibrium once it is reached. Metropolis proposed to realize this using the detailed balance condition:

$$P(X)\pi(X \to Y) = P(Y)\pi(Y \to X)$$

or

$$\frac{\pi(X \to Y)}{\pi(Y \to X)} = \frac{P(Y)}{P(X)} = \exp\left(-\frac{E_p(Y) - E_p(X)}{kT}\right)$$

Monte Carlo Sampling for protein structure

There are many choices for the transition probability that satisfy the balance condition. The choice of Metropolis is:

$$\pi(X \to Y) = \begin{cases} \exp\left(-\frac{E_p(Y) - E_p(X)}{kT}\right) & \text{if} \quad E_p(Y) > E_p(X) \\ 1 & \text{if} \quad E_p(Y) \le E_p(X) \end{cases}$$

The Metropolis Monte Carlo algorithm:

1. Select a conformation X at random. Compute its energy E(X)
2. Generate a new trial conformation Y. Compute its energy E(Y)
3. Accept the move from X to Y with probability:

$$P = \min\left(1, \exp\left(-\frac{E_p(Y) - E_p(X)}{kT}\right)\right)$$

Pick a random number RN, uniform in [0,1]. If RN < P, accept the move.

4. Repeat 2 and 3.

Monte Carlo Sampling for protein structure

Notes:

1. There are many types of Metropolis Monte Carlo simulations, characterized by the generation of the trial conformation.
2. The random number generator is crucial
3. Metropolis Monte Carlo simulations are used for finding thermodynamics quantities, for optimization, …
4. An extension of the Metropolis algorithm is often used for minimization: the simulated annealing technique, where the temperature is lowered as the simulation evolves, in an attempt to locate the global minimum.

Thank you!