Molecular Dynamics

Another Brief Introduction

What is MD?

- **Molecular dynamics is concerned with time dependent processes in molecular systems.**
- **Each dynamic process (i.e., motion) has a characteristic time-scale, amplitude and energy range.**
- **Can be applied to chemical, biochemical as well as solid-state systems**

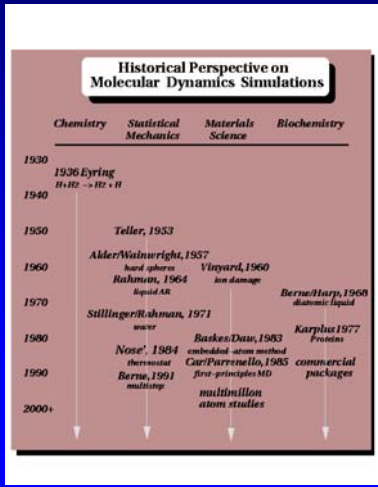
Time scale in biosystems

Type of Motion	Example	Functionality Examples	Time and Amplitude Scales
Local	Atomic fluctuation Side chain motion	Ligand docking flexibility Temporal diffusion pathways	fs - ps (10^{-15} - 10^{-12} s) less than 1 A
Medium Scale	Loop motion Terminal-arm motion Rigid-body motion (helices)	Active site conformation adaptation Binding specificity	ns - micro s (10^{-9} - 10^{-6} s) 1 - 5 A
Large Scale	Domain motion Subunit motion	Hinge bending motion Allosteric transitions	micro s - ms (10^{-6} - 10^{-3} s) 5 - 10 A
Global	Heix-coil transition Folding/unfolding Subunit association	Hormone activation Protein functionality	ms - h (10^{-3} - 10^{4} s) more than 5 A

What we can compute?

- How the structure of our system changes in time (conformational changes, phase transitions, etc…)
- Time dependent properties (diffusion)
- Thermodynamic properties (energy, heat capacity, pressure, etc….)

Historical Perspective



Some Remarks on Statistics

Our system is described by a classical Hamiltonian, a function of the coordinates q and momenta p.

$$H = H(q,p) = K(p) + V(q) = p^2/2m + V(q)$$

Each state of the system is thus characterized by the set (q,p); i.e., each state of the system is a point in the space defined by both coordinates and momenta. This space is called Phase Space.

Boltzmann distribution

To compute the thermodynamics averages over a canonic ensemble (N,V,T=constant) it is necessary to know the probability for finding the system at each and every point (=state) in phase space.

$$P(\mathbf{q},\mathbf{p}) = e^{-H(\mathbf{q},\mathbf{p})/kT} / Z$$

Z = Canonical Partition Function

Property average

If this P is known, an average of any of the system's dynamical variable *A(q,p)* can be calculated as **ensemble average**.

$$\langle A(\mathbf{q},\mathbf{p})\rangle_Z = \int_V d\mathbf{q} \int_{-\infty}^{\infty} d\mathbf{p} P(\mathbf{q},\mathbf{p}) A(\mathbf{q},\mathbf{p})$$

Of course, it is necessary to know the Boltzmann probability for each and every state of the system simultaneously.

Not an easy task at all!

Averages along a Trajectory

- An alternative averaging approach is to calculate averages along the motion of a single point through phase space. Such an average is called a **dynamic average**.

- The two averaging approaches are reciprocal:

  Thermodynamic average The average over **all** points in phase space at a **single** time.

  Dynamic average The average over a **single** points in phase space at **all** times.

Equation of motion

- The motion of a single point through time is obtained by integrating the equation of motion of the system.
- Starting from the point $\{\mathbf{q}(0), \mathbf{p}(0)\}$ the integration yields a **trajectory**

$$\{\mathbf{q}(0), \mathbf{p}(0)\} \rightarrow \{\mathbf{q}(t), \mathbf{p}(t)\}$$

- Averages of any dynamical variable $A(q,p)$ can now be calculated along this trajectory:

$$\langle A(\mathbf{q},\mathbf{p})\rangle_T = \frac{1}{T}\int_0^T A(\mathbf{q}(t),\mathbf{p}(t))dt$$

where T is the length of the simulation

Ergodic Hypothesis

For an infinitely long trajectory the ensemble average and the dynamics average are interchangeable

$$\lim_{T\to\infty}\langle A(\mathbf{q},\mathbf{p})\rangle_T = \langle A(\mathbf{q},\mathbf{p})\rangle_Z$$

Ergodic Hypothesis

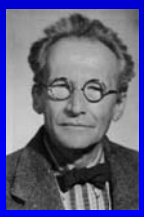
Two assumptions:

✓ The system is at a **stationary state** (e.g., at equilibrium).
✓ An infinitely long dynamic trajectory adequately covers all of phase space, so that the trajectory passes through all possible system states.

The application of "long" but *finite molecular dynamics simulations assumes that the system is "ergodic" and that the finite simulation is "long enough".*

Equations of Motion

- Time-dependent Schrödinger's Equation (quantum)
- Newton's Equation (classical)
- Langevin's Equation (stochastic)



1887-1961

1643-1727

1872-1946

Newton's Equation of Motion

- The Newton's second law of motion

$$F_i = m_i a_i = m\ddot{q}_i = -\nabla_i V(\mathbf{q})$$

The force F_i is given as the derivative of the potential V
(as computed by any Molecular Mechanics approach!!)

Hamilton equation

- Another, more basic, formulation of this equation can be given in terms of the system's Hamiltonian

$$H(\mathbf{q},\mathbf{p}) = \sum_i \frac{1}{2m_i} P_i^2 + V(\mathbf{q})$$

The combination of Newton and Hamilton equation

$$\dot{q}_i = \frac{\partial H}{\partial p_i}$$

$$\dot{p}_i = -\frac{\partial H}{\partial q_i}$$

Properties of Newton's equation

Some important properties of Newton's equation of motion are:

- ✓**Conservation of energy**
- ✓**Conservation of linear momentum**
- ✓**Conservation of angular momentum**
- ✓**Time reversibility**

These properties are used to test whether the numerical solution of the equation (i.e., the molecular dynamics simulation) is stable and reliable.

Integration of Newton's equation

Solving Newton's equation of motion requires a numerical solution of the differential equation. This procedure is called **numeric integration**. Numeric integration involves several components:

Integration Algorithms

- Verlet integrator
- Leap-Frog integrator
- Velocity Verlet integrator

Initial conditions

- Initial coordinates
- Initial velocities

Mesaures for the stability of integrations

Finite Difference Methods.

- They use the information available at time t to predict the system's coordinates and velocities at a time $t + dt$
- based on a Taylor expansion of the position at time $t + dt$

$$\mathbf{r}(t+dt) = \mathbf{r}(t) + \mathbf{v}(t)\,dt + \mathbf{a}(t)\,dt^2/2 + \ldots$$

or in terms of the variables at step n and $n+1$.

$$\mathbf{r}_{n+1} = \mathbf{r}_n + \mathbf{v}_n\,dt + \mathbf{F}_n/2\mathbf{m}\,dt^2 + O(dt^3)$$

- The different integration algorithms vary in the way they implement this basic expansion.

Verlet integrator

Two Taylor expansions, one forward and one backward

$$\mathbf{r}_{n+1} = \mathbf{r}_n + \mathbf{v}_n\, dt + \mathbf{F}_n/2\mathbf{m}\, dt^2 + \ldots$$
$$\mathbf{r}_{n-1} = \mathbf{r}_n - \mathbf{v}_n\, dt + \mathbf{F}_n/2\mathbf{m}\, dt^2 - \ldots$$

These two expansions are than added to give the basic Verlet integration formalism:

$$\mathbf{r}_{n+1} = 2\mathbf{r}_n - \mathbf{r}_{n-1} + \mathbf{F}_n/\mathbf{m}\, dt^2 + O(dt^4$$

Verlet integrator

Advantages:

- Integration does not require the velocities, only position information is taken into account.
- Only a single force evaluation per integration cycle.
- This formulation is naturally reversible in time (a property of the equation of motion).

Disadvantages:

- Introduces rather large numerical errors. An $O(dt^2)$ terms is added to an $O(dt^0)$ term.
- The velocities, which are required for energy evaluation are calculated in an approximate manner only through the equation: $\mathbf{v}_n = (\mathbf{r}_{n+1} - \mathbf{r}_{n-1})/2\, dt$.

Leap Frog integrator

- Velocities are evaluated at the mid-point of the position evaluation and vice versa.

$$\mathbf{v}(t+dt/2) = \mathbf{v}(t-dt/2) + \mathbf{a}(t)\,dt$$
$$\mathbf{r}(t+dt) = \mathbf{r}(t) + \mathbf{v}(t+dt/2)\,dt$$

Each integration cycle involves three steps:

- Calculate $\mathbf{a}(t)\,dt$ (based on $\mathbf{r}(t)$)
- Calculate $\mathbf{v}(t+dt/2)$
- Calculate $\mathbf{r}(t+dt)$

The *instanteneous* velocity at time t is then calculated as

$$\mathbf{v}(t) = (\mathbf{v}(t+dt/2) + \mathbf{v}(t-dt/2)) / 2$$

Leap Frog integrator

Advantages:

- Improved evaluation of velocities.
  - Direct evaluation of velocities gives a useful handle for controling the temperature in the simulation.
  - Reduces the numerical error problem of the Verlet algorithm. Here $O(dt^1)$ terms are added to $O(dt^0)$ terms.

Disadvantages:

- The velocities at time t are still approximate.
- Computationally a little more expensive than Verlet.

Initial Coordinates

- Initial coordinates are usually taken **from experiments** (x-ray, NMR).
- Since experimentally determined structures do not contain coordinates for hydrogen atoms, the first step is to **add hydrogen atoms** to the model.
- **Initial minimization** is essential to remove bad interactions from the initial structure.

Initial Velocities

- Since the only information regarding the veloocities is the desired **temperature** of the simulation, initial velocities (v_x, v_y,v_z) are randomly assigned according to a Maxwellian distribution:

$$P(v_i)dv = (m/2pik_BT)^{1/2} \exp[-mv^2/2k_BT]$$

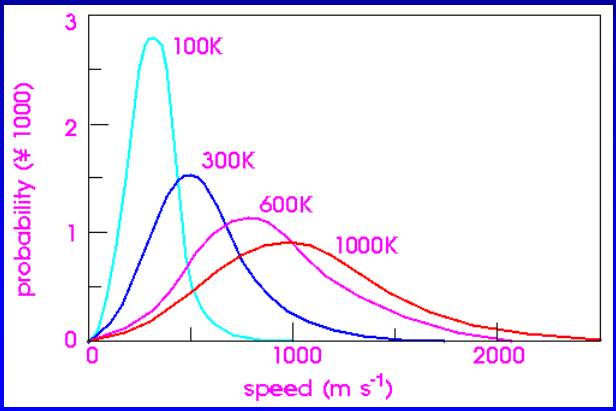
- The **instantaneous temperature** (necessary for heating the system) is defined based on *equipartition theorem* as

$$T(t) = (1/k_B(3N-n)) \sum_i (m_i v_i^2)$$

where *N* is the number of atoms, and *n* is the number of constrained degrees of freedom.

Velocities and Temperature

Maxwell distribution of velocity of water at various temperatures



Stability of integration

Any numerical integration is susceptible to errors and may become unstable. To ensure stability of the simulations the properties of Newton's equation of motion, representing an (N, V, E) ensemble, must be preserved

- *Conservation of energy*
- *Conservation of Linear Momentum*
- *Conservation of Angular Momentum*
- *Time reversibility*

Stability of integration

Conservation of energy

Newton's equation preserves energy. Assuming the average energy fluctuations during a simulation at energy E are dE, a stable simulation will have:

$$log_{10}\,(dE/E) < -4$$

Conservation of Linear Momentum

Total linear momentum, $\mathbf{P} = \sum_i \mathbf{P}_i$, must be conserved

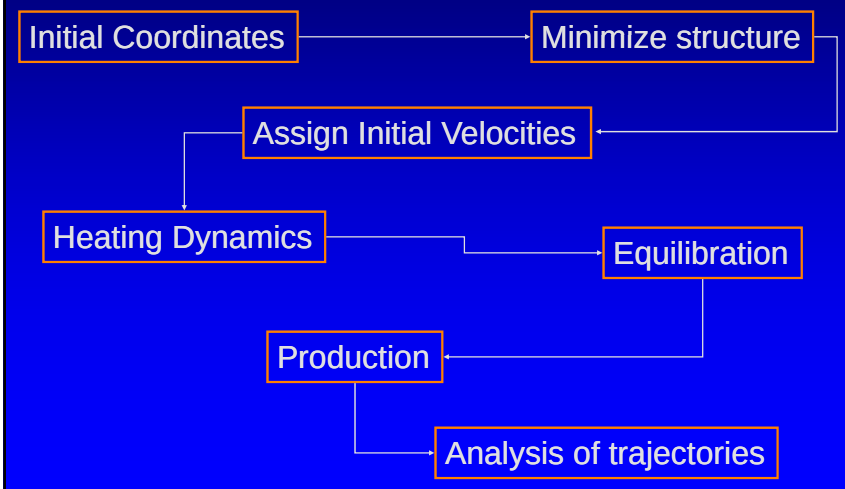
Conservation of Angular Momentum

Total angular momentum, $\mathbf{L} = \sum \mathbf{q}_i x \mathbf{P}_i$, must be conserved

Time reversibility

Newton's equation is reversible in time. Numerical errors, however, hamper this reversibility introducing chaotic effects. Nontheless, for short periods of time a stable integration should exhibit apparent temporal reversibility.

MD in practice



MD begin

Initial Coordinates → Minimize structure

Cartesian, internal coordinates
Or different file formats

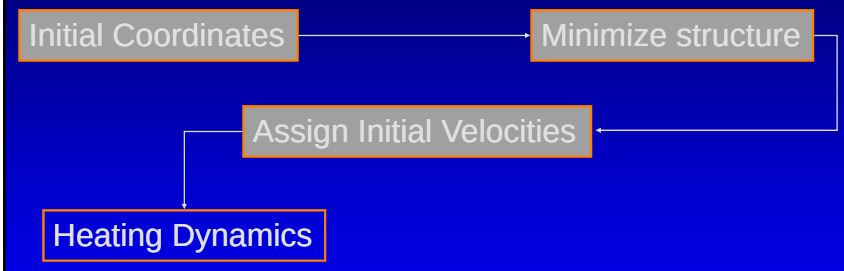
Methods: Newton-Raphson (block diagonal), steepest descent, conjugate gradient, others.

MD Initial Velocities

Initial Coordinates → Minimize structure → Assign Initial Velocities

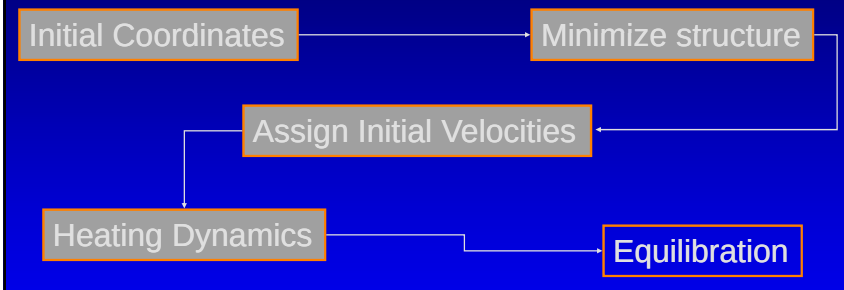
Initial velocities at a low temperature are assigned to each atom and Newton's equations are integrated to propagate the system in time.

MD Heating



new velocities are assigned at a slightly higher temperature and the simulation is allowed to continue. This is repeated until the desired temperature is reached

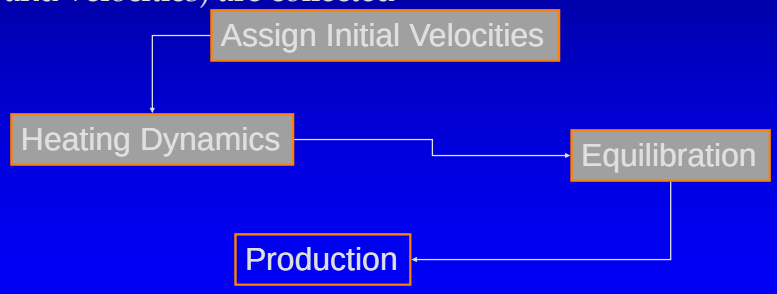
MD Equilibration



Once the desired temperature is reached, the simulation of the system continues and several properties are monitored (the structure, the pressure, the temperature and the energy).

MD Production

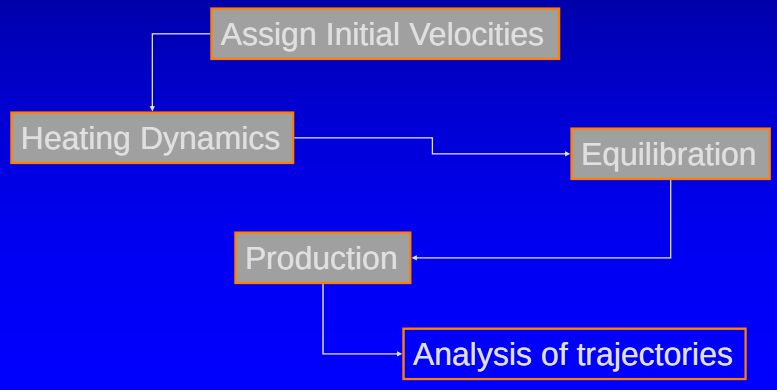
The "production" phase is runned for the time length desired (from several hundred ps to ns). The trajectories (coordinates and velocities) are collected



MD in practice

Using saved trajectories, thermodinamical properties (average values) are computed

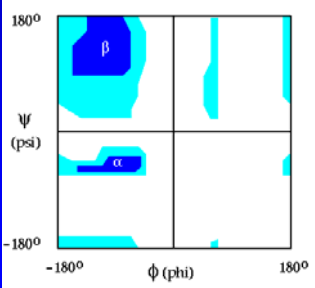
Structural (conformational) changes are observed



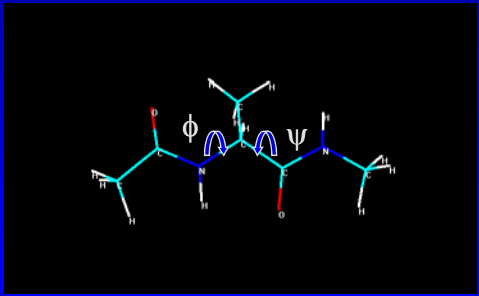
Gas-phase simulation

Amber Force Field

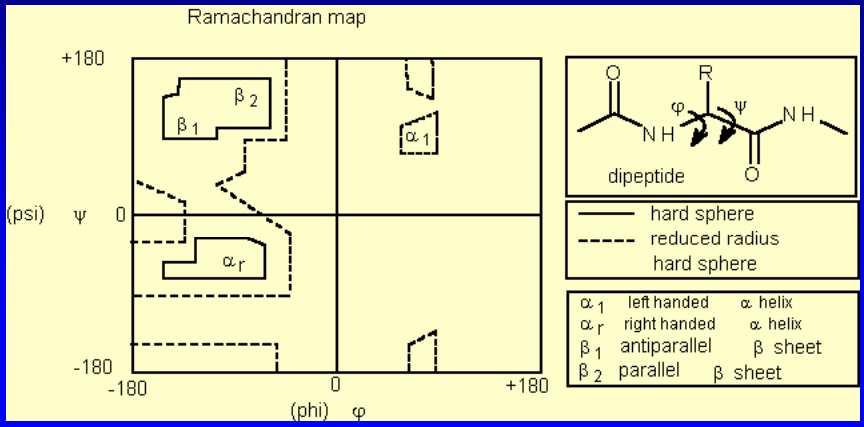
Variables to monitor
E_{tot}, T, ϕ, ψ



Alanine dipeptide

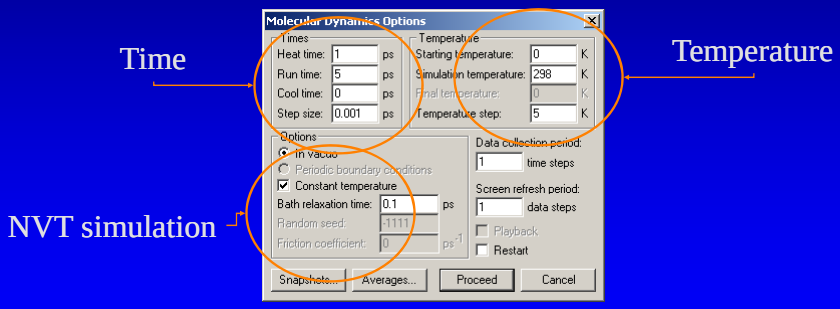


Ramachandran map



Running

Simulation parameters



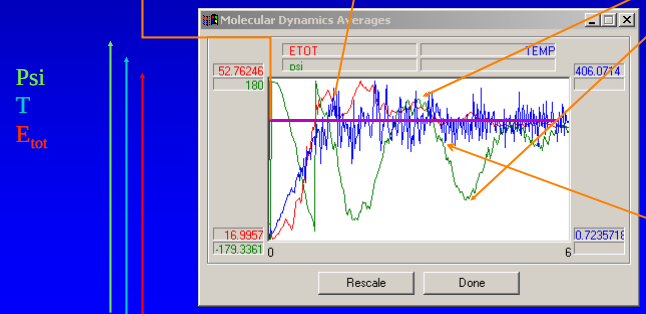
Starting value ϕ=180°, ψ=180° (β-sheet)

Averages

E is conserved in average, oscillations depend on finite integration (same for T)

End of the heating (after 1ps)

β sheet



Psi
T
E_{tot}

α helix

Time (ps)

Solvent in MD simulations

Models for solvent

Implicit model

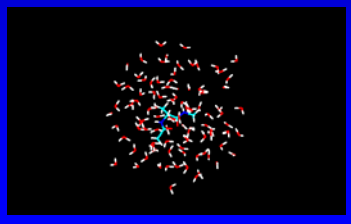
Screening of the electrostatic term

$$V_{elec} = \frac{q_i q_j}{\varepsilon_{eff} r_{ij}}$$

effective dielectric constant (ε_{eff})

Explicit model

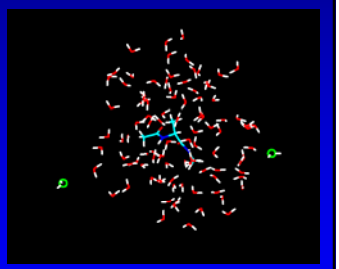
Just add water …..



Explicit model

112 water molecules

Shake

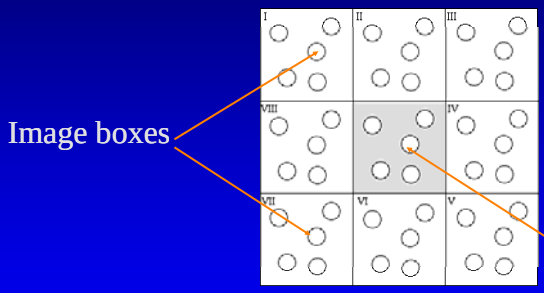


OO distance 21 Å

explosion

OO distance 32 Å

Periodic Boundary Conditions

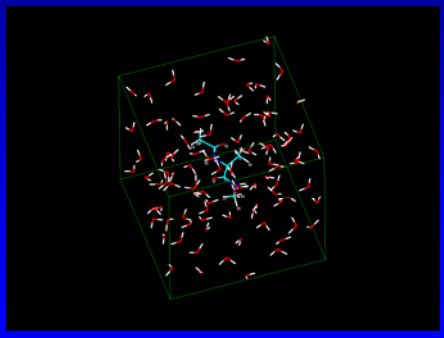


Primary box is replicated in xyz directions
Each particle in the primary box interacts with « image » particles throught a potential

Running

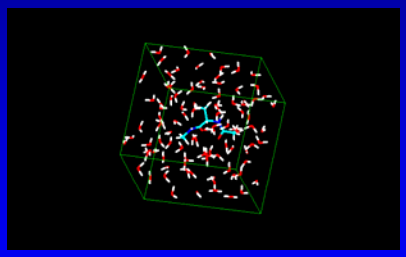
Initial configuration

Box 15 Å of side
112 water molecules
Heat time 1 ps
Run time 5 ps
Step siz 1 fs
Starting T 0 K
Simulation T 298 K
T step 5 K

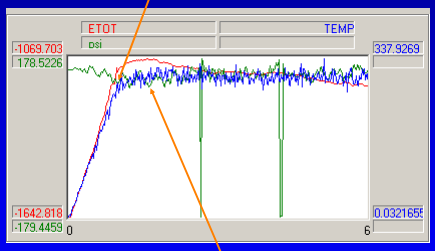


Results

Final configuration



End of the heating
(after 1ps)



β sheet