Molecular Dynamics, Monte Carlo and Docking

Lecture 21

Introduction to Bioinformatics

MNW2
Overview

Today's subject: Molecular Dynamics, Monte Carlo, Docking

- 1. Protein motion
- 2. Molecular dynamics simulations (MD)
  - Equations of motion
  - Verlet algorithm
  - Energetics and force field
  - Trajectories and ensemble properties
- 3. Monte-Carlo simulations (MC)
  - The idea
  - Conformational search with MC
  - Metropolis criterion
- 4. Docking
  - Surface complementarity
  - How to find the correct surface match
1.1 Protein motion: Time and displacement

[Diagram showing a 3D bar graph with logarithmic scales for time and displacement. The categories include:
- Loops
- Termini
- Sidechains
- Fluctuations
- Subunits
- Domains
- Helices
- (Un)folding
- Diss./Ass.
- Transitions]
Allowed phi-psi angles

Red areas are preferred, yellow areas are allowed, and white is avoided.
1.2 Protein motion: Ramachandran plot
1.3 Protein motion: Configuration space
2.1 History of Molecular Dynamic and Monte Carlo Simulations

- Gelatine balls (Morrell and Hildebrand, 1936)
- Invention of MC simulations (Metropolis et al., 1953)
- MC of Lennard-Jones spheres (Wood and Parker, 1957)
- MD of hard spheres (Alder and Wainwright, 1957, 1959)
- MD of liquid argon (Rahman, 1964)
- MD of liquid water (Stillinger and Rahman, 1974)
- MD of protein BPTI (McCcammon, et al, 1977)
- Non-equilibrium methods (1970ies)
- Stochastic dynamics methods (1970ies)
- Quantum-mechanical effects (1980ies)

Today (2000ies):
- Large proteins or complexes in water or membrane
- Simulated time up to micro-seconds
2.2 Newton Equations of motion

A system has coordinates $q$ and impulses $p\ (= mv)$

\[
q = (q_1, q_2, ..., q_N) \\
p = (p_1, p_2, ..., p_N)
\]

The energy can be split into kinetic energy ($K$) and potential energy ($V$)

\[
K(p) = \frac{1}{2}mv^2 = \frac{p^2}{2m} \\
V(q) \Rightarrow \text{depends on interaction}
\]

Here are two examples for potentials:
Coulomb and Lennard-Jones potential

\[
V_C(q) = \frac{z_i z_j}{4\pi \varepsilon_0 r_{qi} r_{qj}} \\
V_{LJ}(q) = 4\varepsilon((\sigma/r)^{12} - (\sigma/r)^6)
\]
2.3a Hamiltonian equations of motion

Hamiltonian equations (one degree of freedom):

\[ \frac{dq}{dt} = \frac{\partial \mathcal{H}}{\partial p}, \quad \frac{dp}{dt} = -\frac{\partial \mathcal{H}}{\partial q}. \]

\( \mathcal{H} \) — Hamiltonian function, *Hamiltonian,*

\( q, p \) — *Canonical variables: generalized coordinate and momentum.*

\( \mathcal{H} \) — Hamiltonian function, *Hamiltonian,*

\( q, p \) — *Canonical variables: generalized coordinate (\( q \)) and momentum [impulses] (\( p \)).
2.3 The Hamiltonian

The Hamiltonian describes the total energy of the system

\[ H(q, p) = K(p) + V(q) \]

An elegant formulation of the Hamiltonian is given by the momenta

\[ \dot{q}_k = \delta H / \delta p_k \]
\[ \dot{p}_k = -\delta H / \delta q_k \]

The Hamiltonian is a fundamental measure for dynamic systems. It is more general than the classical description \( F = ma \), because the forces are often not known explicitly in complicated systems; however, the forces can be extracted from the Hamiltonian (see below).
2.4 Finite difference method: Verlet algorithm

1. Predict \( q, v \) and \( a \) at time \( t + dt \) using current values
2. Evaluate forces \( F = -dH/dr \) and hence accelerations \( a = F/m \) from the new positions.
3. Correct the predicted \( q, v, \) and \( a \) using the new accelerations
4. Return to step 1.

One commonly used algorithm for the above scheme is the Verlet algorithm:
It is derived from a Taylor expansion about time \( t \):

\[
\begin{align*}
  r(t + \delta t) &= r(t) + \delta t v(t) + (1/2) \delta t^2 a(t) + ... \\
  r(t - \delta t) &= r(t) - \delta t v(t) + (1/2) \delta t^2 a(t) - ...
\end{align*}
\]

We can add both equations and obtain:

\[
  r(t + \delta t) = 2r(t) - r(t - \delta t) + \delta t^2 a(t)
\]

\[
  v(t) = \frac{(r(t + \Delta t) - r(t - \Delta t))}{2\Delta t}
\]

\[
  r(t + \Delta t) = r(t) + v(t)\Delta t + (1/2)a(t)\Delta t^2
\]
2.5 The Hamilton principle

"The time derivative of the integral over the energy of a system is a minimum."

\[ \delta \int (p\dot{q} - H(q, \dot{q})) \, dt = 0 \]

This means that dynamic systems move along the energetically lowest trajectory.
Molecular Dynamics

Knowledge of the atomic forces and masses can be used to solve the position of each atom along a series of extremely small time steps (on the order of femtoseconds $= 10^{-15}$ seconds). The resulting series of snapshots of structural changes over time is called a trajectory. The use of this method to compute trajectories can be more easily seen when Newton's equation is expressed in the following form:

$$- \frac{dE}{dr_i} = F_i$$

$$- \frac{dE}{dr_i} = m_i \frac{d^2r_i}{dt^2}$$

The "leapfrog" method is a common numerical approach to calculating trajectories based on Newton's equation. The steps can be summarized as follows:

1. **solve for $a_i$ at $t$ using:**
   $$- \frac{dE}{dr_i} = F_i = m_i \ a_i(t)$$

2. **update $v_i$ at $t + \Delta t/2$ using:**
   $$v_i(t + \Delta t/2) = v_i(t - \Delta t/2) + a_i(t) \ \Delta t$$

3. **update $r_i$ at $t + \Delta t$ using:**
   $$r_i(t + \Delta t) = r_i(t) + v_i(t + \Delta t/2) \ \Delta t$$
2.6 Newton dynamics is conservative

Newton equations of motion obey conservation rules:

- Conservation of energy
- Conservation of linear momentum
- Conservation of angular momentum
- Time reversibility
Force field

The potential energy of a system can be expressed as a sum of valence (or bond), crossterm, and nonbond interactions:

\[ E_{\text{total}} = E_{\text{valence}} + E_{\text{crosstern}} + E_{\text{nonbond}} \]

The energy of valence interactions comprises bond stretching \((E_{\text{bond}})\), valence angle bending \((E_{\text{angle}})\), dihedral angle torsion \((E_{\text{torsion}})\), and inversion (also called out-of-plane interactions) \((E_{\text{inversion}} \text{ or } E_{\text{oop}})\) terms, which are part of nearly all forcefields for covalent systems. A Urey-Bradley term \((E_{\text{UB}})\) may be used to account for interactions between atom pairs involved in 1-3 configurations (i.e., atoms bound to a common atom):

\[ E_{\text{valence}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{torsion}} + E_{\text{oop}} + E_{\text{UB}} \]

Modern (second-generation) forcefields include cross terms to account for such factors as bond or angle distortions caused by nearby atoms. Crossterms can include the following terms: stretch-stretch, stretch-bend-stretch, bend-bend, torsion-stretch, torsion-bend-bend, bend-torsion-bend, stretch-torsion-stretch.

The energy of interactions between nonbonded atoms is accounted for by van der Waals \((E_{\text{vdW}})\), electrostatic \((E_{\text{Coulomb}})\), and (in some older forcefields) hydrogen bond \((E_{\text{hbond}})\) terms:

\[ E_{\text{nonbond}} = E_{\text{vdW}} + E_{\text{Coulomb}} + E_{\text{hbond}} \]
\[ V(R) = \sum_b D_b \left[ 1 - \exp(-a(b - b_0)) \right]^2 + \sum_\theta H_\theta (\theta - \theta_0)^2 + \sum_\phi H_\phi \left[ 1 + s \cos(n\phi) \right] \\
+ \sum_\chi \sum_b H_\chi \chi^2 + \sum_\chi \sum_{b'} F_{bb'} (b - b_0)(b' - b_0') + \sum_\theta \sum_{\theta'} F_{\theta\theta'} (\theta - \theta_0)(\theta' - \theta'_0) \\
+ \sum_{b_0} \sum_{\theta} F_{b\theta} (b - b_0)(\theta - \theta_0) + \sum_{\theta_0} \sum_{\theta'} F_{\theta\theta'} (\theta - \theta_0)(\theta' - \theta'_0) \cos \phi \\
+ \sum_{\chi_{\chi}} \sum_{\chi'} F_{\chi\chi'} \chi \chi' + \sum_{i} \sum_{j > i} \left[ \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{r_{ij}} \right] \]
2.7 A (simple) Lennard-Jones force field

\[ f = \frac{a}{r^6} - \frac{b}{r^{12}} \]
Van der Waals forces
2.8 A (complicated) protein force field

1-4 interactions
covalent bonds
bond angles
improper dihedrals
dihedrals
Lennard-Jones interactions

At each time step:
  • Solve Newton equations for each atom

At longer intervals:
  • Update neighbour list
  • Adjust temperature (velocities)
  • Adjust pressure or volume

Difficult: Coulomb interactions

\[ F = \frac{kq_1q_2}{r^2} \]
2.9 A characteristic setup of MD

Molecular topology -> Forcefield

Molecular coordinates

Box of water
Periodic boundary conditions
Counter ions
Coupling to thermal bath

Run time 200 ns (100 000 000 steps)
Timestep: 2 fs (10^(-15) s)
Coupling to thermal bath

Run time 200 ns
2.10 A trajectory

**Figure:** Snapshots of ubiquitin pulling with constant velocity at three different time steps.
2.11 Ensemble properties

Probability of occurrence ($Z$: partition function)

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

Averages along a dynamic trajectory

$$\langle A(q,p)_T \rangle = \frac{1}{T} \int A(q(t), p(t)) \, dt$$

Ergodic hypothesis

$$\lim_{T \to \infty} \langle A(q,p)_T \rangle = \langle A(q,p) \rangle_Z$$
3.1 The idea behind Monte Carlo simulations

Developed by von Neumann, Ulam and Metropolis at Los Alamos in 1947

Solve problems by stochastic sampling experiment:
- Generate random number
- Perform arithmetic operations
- Interpret results

Very well suited for computers
Often fast approximation to complicated problems
Precision improves with number of trials
Markov chain of events: the outcome of a trial is only dependent on the previous step (no history)
3.2 Example: Calculation of $\pi = 3.14...$

Calculate $\pi$ by MC simulation:
Generate random $x, y$ coordinates between 0 and 1
If $\sqrt{x^2 + y^2} \leq 1$, point within circle
else point outside circle.

Demonstration with program 'mcddemo'
3.3 Example: Results
3.4 Monte Carlo simulation of molecule conformations

Randomised moves with constraints

Distance Geometry Rules

<table>
<thead>
<tr>
<th>Type</th>
<th>d / A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–2</td>
<td>0.02</td>
</tr>
<tr>
<td>1–3</td>
<td>0.05</td>
</tr>
<tr>
<td>1–4</td>
<td>0.1–0.4</td>
</tr>
<tr>
<td>Ring</td>
<td>0.1</td>
</tr>
<tr>
<td>Omega</td>
<td>0.1</td>
</tr>
<tr>
<td>Phi / Psi</td>
<td>0.2–0.4</td>
</tr>
<tr>
<td>Sec. str.</td>
<td>0.5</td>
</tr>
<tr>
<td>Hydrogen bonds</td>
<td>0.5</td>
</tr>
<tr>
<td>Hydrophobic</td>
<td>0.5–1</td>
</tr>
</tbody>
</table>
3.5 The Metropolis criterion

The aim of the Metropolis criterion is to select conformations (generated by an MC method) to create a realistic ensemble.

1. Generate a random number $p(r)$ between 0 and 1
2. Calculate the energy of the conformation
3. Select a temperature (for example 300 K)
4. $p(c) = e^{-E/kT}$
5. If $p(c) > p(r)$ accept conformation else reject conformation

The Metropolis criterion generates a so-called ergodic ensemble, which has the same energy distribution as a 'natural' ensemble.
4.1 Surface complementarity

Protein complexes are often highly specific
Specificity is based on molecular recognition

Characteristic is the complex binding constant:
\[ A + B \rightarrow AB \quad K = \frac{c(AB)}{c(A)c(b)} \]

High-affinity complexes have \[ K \sim 10^9 \]
Binding energy is \(-RT \ln K\)

Surface complementarity by means of evolution:
receptor-ligand
antibody-antigen
enzyme-substrate
4.2 How to find the correct surface match (1)

Systematic search (2 times 3D space)

- \( \bar{u}_{l,m,n} = \begin{cases} 
1 & \text{on the surface of the molecule} \\
\rho & \text{inside the molecule} \\
0 & \text{outside the molecule,} 
\end{cases} \)

- \( \bar{v}_{l,m,n} = \begin{cases} 
1 & \text{on the surface of the molecule} \\
\delta & \text{inside the molecule} \\
0 & \text{outside the molecule,} 
\end{cases} \)
4.2 How to find the correct surface match (2)

Correlation function

\[
\bar{v}_{\alpha,\beta,\gamma} = \sum_{l=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} a_{i,m,n} * b_{l+\alpha,m+\beta,n+\gamma}
\]

A plot of the correlation function indicates favourable docking positions
antibody HyHEL-63 (cyan) complexed with Hen Egg White Lysozyme

The X-ray structure of the antibody HyHEL-63 (cyan) uncomplexed and complexed with Hen Egg White Lysozyme (yellow) has shown that there are small but significant, local conformational changes in the antibody paratope on binding. The structure also reveals that most of the charged epitope residues face the antibody. Details are in Li YL, Li HM, Smith-Gill SJ and Mariuzza RA (2000) The conformations of the X-ray structure Three-dimensional structures of the free and antigen-bound Fab from monoclonal antilysozyme antibody HyHEL-63. Biochemistry 39: 6296-6309.

Salt links and electrostatic interactions provide much of the free energy of binding. Most of the charged residues face in interface in the X-ray structure. The importance of the salt link between Lys97 of HEL and Asp27 of the antibody heavy chain is revealed by molecular dynamics simulations. After 1NSec of MD simulation at 100°C the overall conformation of the complex has changed, but the salt link persists. Details are described in Sinha N and Smith-Gill SJ (2002) Electrostatics in protein binding and function. Current Protein & Peptide Science 3: 601-614.