02601 Project Proposal

Group: GoMad

Wen Li, Xingyu Hu, Ziyan Zhang

October 2024

1 Scientific Problem

How can we perform the MD simulation in the water environment?

2 Scientific article

Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. BMC biology, 9, 71. https://doi.org/10.1186/1741-7007-9-71

3 Project background

The "jiggling" and "wiggling" of atoms are the foundation of life. Molecular dynamics (MD) simulations enable us to capture atomic motion and gain deeper insights into the workings of biological molecules. These simulations are particularly significant in pharmacology, helping to better understand molecular recognition and drug binding during drug discovery. Different from crystallography, MD simulations computationally predict protein motions,which provide greater efficiency and produce results that are highly agreed with experimental data. To accurately replicate the behavior of real molecules in motion, various factors must be considered in the force field, including interactions between bonded and non-bonded atoms, chemical bonds and atomic angles, energy differences, van der Waals interactions, and electrostatic interactions. Once the forces acting on each atom in the system are calculated, their positions will be simulated according to Newton's laws of motion.

4 Approach

In this project, we aim to use Go and R to build a simple classical MD simulation web app. Our approach is to perform classical MD simulation to calculate the physical movement of atoms of one protein in the water environment. Firstly, initialize the atoms, and secondly, calculate the force which can be obtained from the potential energy(related to position). Thirdly, calculate the movement according to force. Finally, repeat step 2 and step 3 until the protein is stable in the water environment.

In classical MD settings, each atom is treated as a "classical" object which interacts with the other atoms in the system. The interaction is represented through a classical potential and the time evolution of the system is determined by the second Newton's law $F = ma$.

1. Data collection: Single protein's structure file (PDB format) from RCSB PDB, ensure the file only contains the pure protein information.
2. Read data: Parse PDB files into Go through Go package "pdb". Topology database of atoms from Gromacs (e.g., ffbonded.itp, ffnonbonded.itp)
3. Classical force fields setup: Since the electronic configuration is frozen in the ground state, we can simplify the form of the energy of interaction in mamny different ways:
   (a) A charge is assigned to each atom, and it is equal to the average charge of that atom

(b) Interactions between bonded atoms reproduce chemical and geometrical properties of the molecule, in particular regarding bond distances, angles and strength

The properties of the chemical bonds depend on the electronic configuration of the atoms in the molecule, and not only on the kind of atoms. Atoms oscillate around their equilibrium average distances, thus chemical bonds are treated as harmonic oscillators. The spring constant depend on the atoms involved and on their electronic configurations

$$U_r = k_r(r - r_0)^2$$

Two consecutive bonds form an angel. And we assume a harmonic approximation: the angle can oscillate around the equilibrium configuration with a well-defined angular strength

$$U_\theta = k_\theta(\theta - \theta_0)^2$$

And for consecutive atoms define three bonds and two planes. The two planes describe an angle which is called dihedral angle. Due to steric and electrostatic interaction, not all the angular position are allowed for the dihedrals. Several minima of the can exist for a single dihedral angle, corresponding to different possible configuration (isomers of rotation or rotamers). Dihedral angles allow biological macromoleculea to be flexible and assume many different configurations.

$$U_\phi = \sum c_k \cos(\phi - \phi_{0,k})$$

(c) Non-bonded atoms interact exclusively through Coulomb and Lennard-Jones potentials

Atoms that are not connect by stable chemical bonds are considered non-bonded, and they interact through long range potentials

i. Coulomb interaction: the charge represent the average charge on the atom due to the distribution of the electrons in that particular molecule

$$U_c = \frac{q_i q_j}{4\pi\epsilon_0 r}$$

ii. Lennard-Jones potential: It is the sum of an attractive part (quantum effect) and a repulsive part which represent the effect of electronic clouds in close proximity

$$U_{LJ} = \left(\frac{A}{r^{12}} - \frac{B}{r^6}\right)$$

iii. Hydrogen bonds: an empirical potential that take explicitly into account of the possibility of creation of hydrogen bonds

$$U_{HB} = \left(\frac{C}{r^{12}} - \frac{D}{r^{10}}\right)$$

Then, the total energy of the system can be

$$\begin{aligned} U = &\sum_{bonds} k_r(r - r_0)^2 + \sum_{angle} k_\theta(\theta - \theta_0)^2 + \\ &\sum_{dihedrals} C_k \cos(\phi - \phi_{0,k}) + \sum_{impropers} k_\alpha(\alpha - \alpha_0)^2 + \\ &\sum_{i<j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \sum_{i<j} \left(\frac{A}{r^{12}} - \frac{B}{r^6}\right) + \sum_{h.b.} \left(\frac{C}{r^{12}} - \frac{D}{r^{10}}\right) \end{aligned}$$

4. Energy minimization: Each of the particle in the system will feel a certain force, which is the gradient of the potential energy given by the interaction with all the other particles

$$U_i = U_i(x, y, z)$$

In this case, the force is given by the opposite of the gradient of the potential energy

$$F = -\nabla U_i = \left(-\frac{\partial U_i}{\partial x}, -\frac{\partial U_i}{\partial y}, -\frac{\partial U_i}{\partial z} \right)$$

Use the second Newton's law we can obtain the acceleration of each atom of the system if we know the force.

5. Uniformly acceleration motion: By integrating the equation of velocity in the time variable we can obtain the equation for the position

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{1}{2}a(\Delta t)^2$$

Since position and velocity are vectors, we can write the equations for each of the components

$$\begin{aligned}
v_x(t + \Delta t) &= v_x(t) + v_x \Delta t \\
v_y(t + \Delta t) &= v_y(t) + v_y \Delta t \\
v_z(t + \Delta t) &= v_z(t) + v_z \Delta t \\
x(t + \Delta t) &= x(t) + v_x(t)\Delta t + \frac{1}{2}a_x(\Delta t)^2 \\
y(t + \Delta t) &= y(t) + v_y(t)\Delta t + \frac{1}{2}a_y(\Delta t)^2 \\
z(t + \Delta t) &= z(t) + v_z(t)\Delta t + \frac{1}{2}a_z(\Delta t)^2
\end{aligned}$$

6. Periodic Boundary Conditions: An isolate box of water would evaporate very fast, losing contacts with the protein. To solve this issue periodic boundary conditions are introduced. The system is repeated identical infinite times in the 3-dimensional space.

7. Trajectory analysis: The result of a MD simulation is a trajectory. We can follow each atom movement for a limited amount of time (ns to μs). If we have reached equilibrium the time averages corresponds to the phase average.

   (a) Root Mean Squared Deviations (RMSD): It is a measure of the average displacement of a group of atoms from their initial/reference positions over the course of a molecular dynamics (MD) simulation trajectory. It is calculated as a function of time/frame and provides insight into the overall structural deviation during the simulation. The RMSD is expected to increase initially and then typically plateaus once equilibrium is reached, oscillating around a relatively constant value (of a few Å).

$$RMSD(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_i(t) - r_i(0))^2}$$

(b) Root Mean Squared Fluctuations (RMSF): It quantifies the mobility of individual atoms or residues by measuring their timeaveraged deviation from a reference position (usually the initial structure or average structure) over the entire MD trajectory. It is calculated as a function of the residue index along the protein sequence. Higher RMSF values correspond to more flexible regions, while lower values indicate rigid or well-packed regions.

$$RMSF(i) = \sqrt{\frac{1}{T}\sum_{t=0}^{T}(r_i(t) - r_{i,AVE})^2}$$

5 External resources

The input data about the protein(such as atom types and positions of every atom) can be obtained from RCSB PDB. The key parameters like the bond force constant in energy calculation can be obtained from the files(such as ffbonded.itp) in the Gromacs.