JOINT INSTITUTE FOR NUCLEAR RESEARCH

SCIENCE BRINGS NATIONS TOGETHER

1956

3/23/2022

Report on:

MD-SIMULATION (FROM ATOMIC FRAGMENTS TO MOLECULAR COMPOUNDS)

Mohamed Bysam
INTEREST

Prof. Kholmirzo Kholmurodov
SUPERVISOR

Contents

1. Introduction

Molecular dynamics is the science of simulating the motions of a system of particles. It has been applied to systems as small as an atom and a diatomic molecule undergoing a chemical reaction, and as large as a galaxy. In all cases, the essential elements for a molecular dynamics simulation are a knowledge of the interaction potential for the particles, from which the forces can be calculated, and of the equations of motion governing the dynamics of the particles. The interaction potential may vary from the simple gravitational interaction between stars to the complex many-body forces between atoms and molecules. Classical Newtonian equations of motion are adequate for many systems, including the biomolecules of primary concern here, but for some problems (such as reactions involving tunnelling) quantum corrections are important, and for others (such as galaxy evolution) relativistic effects may have to be included.

2. Molecular dynamics simulation

Two attributes of molecular dynamics simulations have played an essential part in their explosive development and wide range of applications. Simulations provide individual particle motions as a function of time so they can be probed far more easily than experiments to answer detailed questions about the properties of a system. Further, although the potential used in a simulation is approximate, it is completely under the user's control, so that by removing or altering specific contributions their role in determining a given property can be examined. 'Computer alchemy'-changing the potential from that representing one system to another during a simulation is a powerful tool for calculating free energy differences. Equation (1) is used to express the Hamiltonian of the system as a function of nuclear variables.

$$H(q,p) = K(p) + V(q) \tag{1}$$

Where *H* is the Hamiltonian, *K(p)* is the kinetic energy as a function of momenta, that is, it contains the momenta of each atom, and *V(q)* is the potential energy as the function of generalized coordinates, that is, it contains the details of the interatomic interactions.

Molecular mechanics consider the atomic composition of a molecule such as a protein to be a collection of masses that are interacting with each other through harmonic forces. Further simplification suggests that the atoms are considered balls in molecular mechanics and the bonds between them are considered as springs.

2.1 Potential energy function

In molecular mechanics, the forces acting on the atom *i* can arise from both internal and external sources. The internal sources are basically the interaction forces acting between bonded and non-bonded atoms. The external sources can be environmental stresses including electric field, heat, and pressure, which are imposed on the system externally. To determine the forces, it is important to obtain the potential energy of the system that can be expressed as a sum of bonded, non-bonded, and cross-term interactions. The derivatives of the potential energy function are known as force fields (Equation 3)

$$E_{\text{total}} = E_{\text{bonded}} + E_{\text{non-bonded}} + E_{\text{cross-term}} \quad (2)$$

Bonded interactions are also termed as **valence interactions** and incorporate diagonal terms such as bond-stretching, angle-bending, dihedral-angle torsion, inversion, or out of plane interactions. Since in molecular mechanics it is assumed that the interactions are through harmonic forces, the potential energy associated with bond-stretching can be represented as

$$E_{\text{stretching}} = \frac{1}{2} k_b (b - b_0)^2 \quad (3)$$

where k_b is the force constant for the bond length, b_0 is the equilibrium bond length, and b is the actual bond length.

Similarly, for angle-bending, the simple harmonic expression would be

$$E_{\text{angle-bending}} = \frac{1}{2} k_\theta (\theta - \theta_0)^2 \quad (4)$$

where k_θ is the force constant for bond angles, θ_0 is the equilibrium bond angle, and θ is the actual bond angle.

The dihedral-angle torsion potential energy is represented as a cosine expression

$$E_{\text{torsion}} = \frac{1}{2} k_\varphi (1 + cos(n\varphi - \varphi_0)) \quad (5)$$

Where, k_ϕ is the force constant for the dihedral angle, ϕ_0 is the reference torsion angle, and ϕ is the actual torsion angle. All the aforementioned potential energy functions belong to Class I force-fields, which only incorporate diagonal terms and non-bonded interactions. This class of force-field is widely used to simulate complicated systems such as proteins, as one has to consider the different degrees of freedom associated with it. Some of the widely used force fields of Class I are CHARMM, AMBER, and GROMOS. The non-bonded interactions defined in this class of force-fields include electrostatic and van der Waals interactions. The electrostatic interactions can be described as an interaction between two charged particles and are expressed with a Coulombic potential function (Equation 6).

$$E_{\text{electrostatic}} = \sum_{ij} \frac{q_i q_j}{4\pi D r_{ij}} \quad (6)$$

where q is the charge of the particle, D is the dielectric constant of the material, and r_{ij} is the distance between the centres of the particles i and j.

The van der Waals interaction between non-bonded atoms can be accounted for using the Lennard–Jones potential. In molecular mechanics, the simplest potential energy function is the hard-sphere potential. This assumes that the particles would travel in a straight line until hitting another particle, which will lead to elastic scattering. This potential does not include the attractive and repulsive components, which make it inappropriate for the study of huge systems such as proteins. The Lennard–Jones potential included both attractive and repulsive components and can be expressed using Equation 7.

$$E = 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right] \quad (7)$$

where ε is the depth of the potential well, *r* is the distance between the particles, and σ is the distance at which the inter-particle potential is 0. This potential function is also called as 12–6 potential, which represents the exponent terms for repulsive and attractive components, respectively. Using Equation 7, the van der Waals interaction between non-bonded particles can be presented as

$$E_{\text{vdw}} = \sum_{ij} \varepsilon_{ij} \left[\left(\frac{R_{\min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{\min,ij}}{r_{ij}} \right)^{6} \right] \quad (8)$$

where, r_{ij} is the distance between the particles, ε_{ij} is the minimum value for van der Waals term, and $R_{\min}$ is the radius where the van der Waals term is minimum.

Some modern force-fields including CFF, MM3, and MMFF94 also include another class of force-fields, *Class II*, which incorporates the cross-term interaction between the particles. These cross-terms are required to reproduce the experimental vibrational frequencies of molecules and include the bond-bond, angle-angle, bond-angle, bond-dihedrals, angles-angles-dihedrals, and bond-bond-dihedrals. All these terms account for effects such as distortion of bond angles due to stretching of the bonds or changes in bond length due to changes in the other in an opposite direction. The purpose of these force-fields is to describe the complete potential energy surface of the molecule as accurately as possible. If we want to study the effect of external stresses on a molecule, we can look at the changes in the potential energy functions and determine how they are affected by stress.

2.2 Simulation methods (Proteins for example)

Given a potential energy function, there are several approaches to the dynamics. The most exact and detailed information is provided by molecular dynamics simulations in which Newton's equations of motion are solved for the atoms of the system and any surrounding solvent. For a simple homogeneous system, such as a box of water molecules with periodic boundary conditions, average structural and dynamic properties can be determined in simulations of only a few picoseconds. Inhomogeneous systems such as proteins require considerably longer simulations. Modern computers allow simulations of up to nanoseconds,

long enough to completely characterize the librations of small groups in a protein and to determine the dominant contributions to atomic fluctuations 1. To begin a dynamic simulation an initial set of atomic coordinates and velocities is required. The coordinates can be obtained from X-ray crystallographic or NMR structure data, or by model-building. Given a set of coordinates, a preliminary calculation serves to equilibrate the system. The structure is first refined using an iterative minimization algorithm to relieve local stresses due to overlaps of non-bonded atoms, bond length distortions, and so on. Next, atoms are assigned velocities (v) taken at random from a Maxwellian distribution for a low temperature, and a simulation is performed for a few picoseconds. This is done by finding the acceleration a_i of atom i from Newton's law

$$\frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i} \tag{9}$$

Where F_i is defined as the total force experienced by an atom i in the direction r, m_i is the mass of the atom, and r_i is the position of the atom I, and introducing it into the equation for the

position r_i at time $(t + \Delta t)$, given r_i, at time t:

$$r_i(t + \Delta t) = r_i(t) + v_i \Delta t + \frac{1}{2} a \Delta t^2 + \cdots \tag{10}$$

The equilibration is continued by alternating new velocity assignments, chosen from Maxwellian distributions for temperatures that are successively increased to some chosen value, with intervals of dynamical relaxation. The temperature T of the system is measured by the mean kinetic energy,

$$\tfrac{1}{2} \sum_{i=1}^{N} m_i \langle v_i^2 \rangle = \tfrac{3}{2} N k_B T \tag{11}$$

where N is the number of atoms in the system, $\langle v_i^2 \rangle$is the average velocity squared of the i^{th} atom and k8 is the Boltzmann constant. The equilibration period is considered finished when the temperature is stable for longer than about 10 ps, the atomic momenta obey a Maxwellian distri

2.3 MD Simulation Packages

Many readily available packages exist such as DL_POLY, AMBER, CHARMM, NAMD, etc.

In the DL_POLY code there are three input and three output files.

- Input files
    1. CONFIG (Places and velocities of atoms in x, y, and z coordinates as well as boundary conditions.)
    2. CONTROL: Contains data on: (Thermodynamic parameters such as temperature, pressure, …)
    3. FIELD: (potential energy function of atoms)
- Output files
    1. OUTPUT
    2. REVCON
    3. HISTORY

3. Future Goals

Having recently finished my master's degree in experimental biophysics, I became very interested in computational biophysics and molecular dynamic simulation. I am planning to get my PhD degree on relative computational areas such as Light-Protein interaction, protein-protein interactions, drug design, and so on…

Honestly this course gave me the starting spark.

4. References

1. Kholmirzo KHOLMURODOV (Editor), Â«Computer Design for New Drugs and Materials: Molecular Dynamics of Nanoscale PhenomenaÂ», Nova Science Publishers (N.Y.), ISBN: 978-1-53612-082-0, (2017).
2. Ashutosh Singh, Sai Kranthi Vanga, Valerie Orsat & Vijaya Raghavan (2018) Application of molecular dynamic simulation to study food proteins: A review, Critical Reviews in Food Science and Nutrition, 58:16, 2779-2789, DOI: 10.1080/10408398.2017.1341864
3. Karplus, M., Petsko, G. Molecular dynamics simulations in biology. *Nature* **347,** 631–639 (1990). https://doi.org/10.1038/347631a0