JOINT INSTITUTE FOR NUCLEAR RESEARCH
Frank Laboratory of Neutron Physics

Introductory Course: MD-Simulation Research (From Atomic Fragments To Molecular Compound

Supervisor:
Prof. Kholmirzo Kholmurodov
Frank Laboratory of Neutron Physics, JINR

Student:
Zeyad Mansour B. Radwan
Cairo University

Participation date: 02 March – 19 April, Wave 14

Dubna, 2026

Table of Contents

Table of Figures

I. Abstract

Molecular dynamics act as the bridge between the laboratory and the microscopic world in the femtoseconds scales up to microsecond scales. It's used in life sciences, material sciences, and many other fields to study how ensembles of molecules interact with one another, and it's of great importance in biophysics when studying how proteins or other biological molecules behave in cell compartments. MD has an extensive mathematical models that needs to be reviewed thoroughly before starting a simulation, because it the parameters and inputs included in the simulation must resemble what is seen in real life. This report outlines the principles of MD and how it works in essence.

II. Introduction

Molecular Dynamics (MD) simulations are a major type of computer simulations that are used to study molecular assemblies and their behavior, the other type being Monte Carlo simulations. They are used to simulate the physical movements of atoms and molecules over time, and are generally done by numerically solving newton's law of motions; thus simulating how particles move and interact. They are used to simulate microscopic systems like proteins, liquids, nanoparticles, and materials at the femtoseconds resolution. Therefore, computer simulations like molecular dynamics act as a bridge between theory and experimentation; we may test a theory by conducting a simulation using the same model and comparing it with the experimental results. This approach has become a cornerstone in fields of computational chemistry, biophysics, materials science, and drug discovery.

In MD simulations, the atoms are treated as classical particles and bonds between them are treated as springs, all interacting through a predefined potential energy function. This potential field is what is known as a **force field.** These force fields such as **AMBER**, **CHARMM**, **GROMOS**, or **OPLS**; describe bonded interactions (including bonds, angles, etc) as well as non-bonded interactions (van der Waals and electrostatic forces). The resulting trajectories provide detailed insights into the conformational changes and binding interactions under various environmental conditions.

Molecular dynamics simulations can be sub-categorized based on the level of structural resolution, treatment of the environment, and the underlying equations of motion and potential. At the highest level of detail, all-atom molecular dynamics (**AA-MD**) explicitly represents every atom in the system, while coarse-grained molecular dynamics (**CG-MD**) reduces complexity by grouping atoms into interaction sites. The inclusion of a solvent system in the MD simulation is also of imperative importance, and as such, simulations may employ **explicit** or **implicit** solvent models depending on how the surrounding environment is treated. Explicit solvent models like those used for water (e.g. **TIP3P, TIP4P, TIP5P, and SPC**) represent the solvent molecules explicitly (as any other molecule in the system). Implicit solvent models (e.g. **Generalized Born**) treat the solvent system as a continuum of dielectric medium to reduce computational cost. In terms of accuracy, explicit solvents are more accurate because local interactions like hydrogen bonding are better described by this model, while in terms of efficiency, implicit solvents are far more computationally reasonable. Other MD variants such as **Brownian dynamics** modify the equations of motion to include **stochastic forces**, enabling the study of diffusive processes. At larger spatial and temporal scales, continuum mechanics approaches (**CM**) replace atomistic descriptions altogether, although these are no longer considered molecular dynamics in the strict sense. It should be noted the difference between molecular mechanics and molecular dynamics resolves around how the molecular system is studied. Molecular mechanics is focused on finding the most stable static conformation (minimum energy) of the system that is independent of time. Molecular dynamics are focused on how molecules interact based on their movement and binding, which is time dependent.

Usually, simulating material/compounds of hetero-atoms like gold, metals, metal oxides, or others require a far more robust description of the system as it heavily depends on electronic structure of these atoms. In which case, the system is simulated using a different approach to the casual MD simulations utilizing quantum mechanics (QM) in what is known as **qMD**. In this approach, a highly accurate description of the system and its behavior is provided, albiet the computational demand of this approach is very expensive. Common electronic-structure-based simulations like Density Function Theory (DFT) and qMD simulations are so computationally demanding that it's commonly done on merely 30-100 atoms and at very small time steps ranging from femto to pico-seconds, depending on the available computational resources. The fundamental idea behind DFT, as described by Kohn-Hohenberg theorm, it that for non-degenerate systems, the **ground state electron density, n(r),** alone determines the ground-state behavior of the system and as such n(r) minimizes the energy of the system. This can be summarized in the **figure 1**.

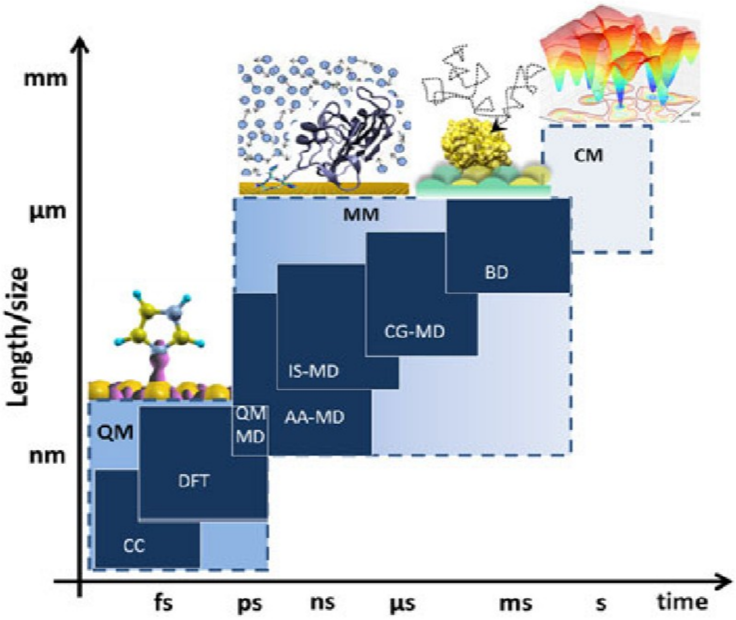


Figure 1: Time and length scales of different simulation techniques: Quantum Mechanics (QM); includes Coupled cluster (CC) and DFT method. Molecular mechanics (MM), includes all-atom molecular dynamics (AA-MD), Implicit solvent MD (IS-MD), coarse grained MD (CG-MD), and Brownian dynamics (BD-MD). Finally, simulating larger systems for longer timescales are done using continuum mechanics (CM).

Quantum molecular dynamics can be used with classical molecular dynamics in what is known as hybrid qMD/MD. In this approach, exponential potential terms are used to describe hetero-atoms present in the system, while the rest of the atoms are described using classical mechanics.
In general, Molecular dynamics provide us with a comprehensive framework of any molecular system behavior we're studying, provided that we use proper definitions for the specific system being simulated.

In this course, the theoretical background that addresses how molecular dynamics are performed were discussed in detail, and the following manuscript is the summarized version of what was studied.

III. Theoretical Background For interactions

As discussed, molecular dynamics simulations are primarily based on a classical mechanics framework, where atoms are treated as point particles whose motion is governed by Newton's equations of motion. This framework enables the description of atomic motion, including translation, vibration, and conformational changes over time. To initialize an MD simulation, the following quantities must be defined for each atom i: position ri(t), velocity vi(t), mass (mi), acceleration ai(t). The forces acting on each atom are obtained from the system's potential energy fnction and the trajectories are generated by integrating the equations of motion over discrete time steps.

A typical molecular dynamics simulation follows an iterative **predictor-corrector algorithm**, the most basic of which are as follows:

---

1. **Initiation**: initial atomic coordinates r_i, velocities vi, and system parameters are assigned. Velocities are often initialized according to Maxwell-Boltzmann distribution at a given temperature as shown in **figure 2**. Time (t) is set to zero, and the step is indexed from n=0 with time step Δt is also set for simulation progression.

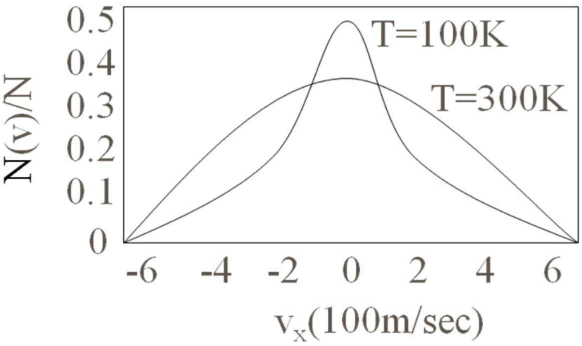


Figure 2: Maxwell-Boltzmann velocity distribution depending on temperature

The velocities of the atoms/molecules in the system defines the temperature of said system by the following relationships:

$$T(t)=\frac{1}{3NK_B}\sum_{i=1}^{n} m_i v_i^2 , v_i=\frac{dr_i}{dt} \tag{1}$$

And in turn, the velocity for each components in 3D space of the system (x,y, and z) are calculated as a probability (f(v)) in a distribution**,** which is defined by:

$$f(v_x)dv_x=N(v_x)\frac{dv_x}{N} \tag{2}$$

$$f(v_x)dv_x=N(v_x)\frac{dv_x}{N}=\sqrt{\left(\frac{m}{(2\pi kT)}\right)}.e^{-\left(\frac{mv_x^2}{2kT}\right)}dv_x \tag{3}$$

And this ultimately describes the maxwell velocity distribution shown in fig 2, and it tells us how likely a specific velocity is at the set temperature T.
The average velocities of these atoms are thus defined by:

$$\langle A(v_x)\rangle=\int_{-\infty}^{\infty} A(v_x)f(v_x)dv_x \tag{4}$$

2. Predictor step: Atom positions are predicted using velocities at t+Δt using new current values. A **Verlet-type** scheme for this is:

$$r_i(t+\Delta t)=r_i(t)+\Delta t\, v(t+\frac{1}{2}\Delta t) \tag{5}$$

3. Force estimation: the forces experienced by each atom are computed on the updated positions. This can be calculated by Newton’s second law of motion:

$$F_i(r)=m_i\frac{d^2 r_i(t)}{dt^2}, i=1,2,\ldots,n$$

$$r=\{r_1,r_2,\ldots,r_n\}; U(r), \text{where } U \text{ is the potential energy}$$

$$F_i(r)=-\frac{\partial U(r)}{\partial r_i}$$

$$\text{at a specific time } t; F_i(r_i(t))=m_i\frac{d^2 r_i(t)}{dt^2}$$

Which can be modified to include friction/dampening effects (so that the particle energy doesn’t keep increasing freely) to simulate interaction with environment (solvent drag) , plus a random thermal component . This modified system is called **thermostatted system** or **Langevin MD**, and is calculated as follows for the force: (6)

$$F_i(r_i(t))=m_i(\frac{d^2 r_i(t)}{dt^2})+\gamma_i m_i(\frac{dr_i(t)}{dt})-R_i(t)$$

Where γ is the friction coefficient in ps^{-1} , and $R_i(t)$ is the stochastic force with distribution around 0 and variance that’s proportional to $\gamma k_B T$. This helps maintain a constant temperature. This is also known s the fluctuation-dissipation theorem. Other temperature control methods are discussed later like **Nose-Hoover** method that relies on deterministic scaling.

4. Corrector step: Adjust velocities using both current and updated accelerations by following a general equation:

$$v_i(t+\Delta t)=v_i(t)+\frac{1}{2}[a_i(t)+a_i(t+\Delta t)]\Delta t \tag{7}$$

The previous correction for velocity is **NOT** thermostatted. A thermostatted velocity correction like the one used in **Nose-Hoover** is described by knowing the **adaptive friction coefficient σ**, defined by:

$$\sigma(t+\frac{1}{2}\Delta t)=\sigma(t-\frac{1}{2}\Delta t)+(\frac{\Delta t}{\tau_T^2})(\frac{T}{T_{ext}}-1)$$

Where τ is the termostat relaxation time , and in essence, the whole rightside expression simulates an exchange between an external heat bath with temperature T_{ext} and how quickly/strongly it is coupled with the system .

$$v(t+\frac{1}{2}\Delta t)=v(t-\frac{1}{2}\Delta t)+\Delta t[(\frac{F(r_i(t))}{m})-\sigma(t)v(t)]$$

Note: The simulation for velocities and adaptive friction is done at half steps, whereas positions are defined at integer time steps. (8)

Thus the average velocity and friction coefficient are then defined by,

$$\sigma(t)=\frac{1}{2}\left\llbracket \sigma(t-\frac{1}{2}\Delta t)+\sigma(t+\frac{1}{2}\Delta t)\right\rrbracket$$

$$v(t)=\frac{1}{2}\left\llbracket v(t-\frac{1}{2}\Delta t)+v(t+\frac{1}{2}\Delta t)\right\rrbracket$$

With respect that updated position $r_i(t)$ is equation **5** (Verelt scheme).

5. Apply Boundary condition and constraints: The simulated system is confined by a boundaries like one would see with a container. The boundary is often made **periodic**, in which case an initial system is treated as a unit cell surrounded by identical repeating copies. In this scenario, when an atom exits one side of the simulation container or unit cell, it will re-enter from the opposite side, thusly emulating an infinite system in a sense (figure 3). This as aspect is known as **periodic boundary conditions (PBC).** Constraints are also applied to ensure system closely resembles what is observed in a real experiment. It should be noted that in a PBC simulation, electrostatic forces are simulated in specific manner as to ensure that the electrostatic forces reaches far. This will be defined in **Ewal Potential.**

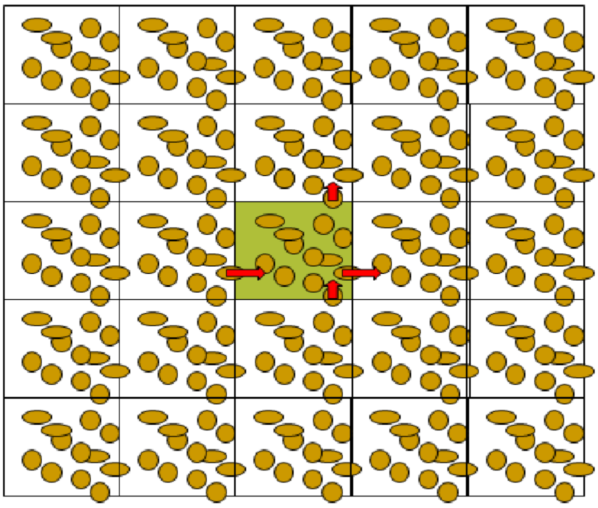


Figure 3: Periodic Boundary Condition (PBC).

6. Computing physical properties like structural metrics (RMSD, RMSF, etc).

7. Time progression by updating simulation time t to t+ Δt while indexing step from n to n+1

8. Iteration from step 2 till 7 until the simulation reaches an desired time.

The calculation of forces and potential will be explained in more details below.

A. Equations of Motion and force field potential

As mentioned, the motion of atoms and their trajectories are calculated using classical Newtonian mechanics, where the force *F* done by each atom *i*, with respect to its position *r* is calculated by:

$$F_i(r)=m_i\frac{d^2r_i(t)}{dt^2}=-\frac{\partial U(r)}{\partial r_i},i=1,2,\dots,n \qquad (9)$$

Where U is the potential energy with respect to the atom’s position. The total potential energy is estimated by:

$$U(r)=U_b+U_\theta+U_\phi+U_\omega+U_{Lj}+U_{el}+U_{HB}+\dots \qquad (10)$$

In computational chemistry and nanotechnology applications, a lot of potentials have been developed depending on the purpose of the simulation or what is being simulated. For example, force fields like AMBER is designed to simulate biological molecules, and has a simplified force field potential defined by:

$$U(r)=U_b+U_\theta+U_{phi}+U_{Lj}+U_{el}+U_{HB} \qquad (11)$$

Where **U_b** is the **valence length potential**, **U** is the **angle potential**, **U_ϕ** is the **torsion dihedral potential**, **U_{lj}** is the v**an der Waals interaction potential** following **Lennard-Jones** relation, **U_{el}** is the **electrostatic potential**, and **U_{HB}** is the **hydrogen bonding potential**. Additional terms are incorporated when simulating more fastidious systems that relies on electronic structures of the atoms involved.

An expanded form of this total potential or **force field** is shown below:

$$U(r)=\frac{1}{2}\sum_b K_b(r-b_0)^2+\frac{1}{2}\sum_\theta K_\theta(\theta-\theta_0)^2+\frac{1}{2}\sum_\phi K_\phi[\cos(n\phi-\delta)+1]+\sum_{Lj}[\frac{A}{r_{ij}^{12}}-\frac{B}{r_{ij}^{6}}]+\dots$$
$$\dots+\sum_{i,j}\frac{q_iq_j}{\varepsilon r_{ij}}+\sum_{i,j}[\frac{A'}{r_{ij}^{12}}-\frac{B'}{r_{ij}^{10}}] \qquad (12)$$

We can see from equation 12 that interaction between atoms can be split in general into bonding and non-bonding interaction (figure 4).

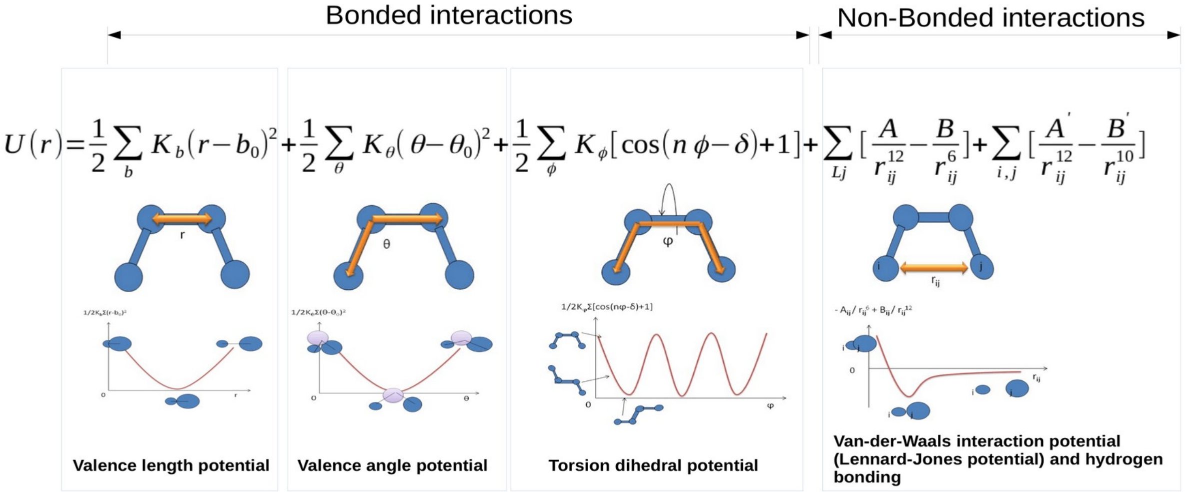


Figure 4: Simplified force field potential equation. Can be summarized into bonded and non-bonding interactions.

We will discuss each type of interaction and their components below.

C. Bonded Interactions (1-2 Angestroms)

In classical mechanics, atoms are treated as classical point particles with bonds between them as springs. These interactions dictate the internal potential energy from the atoms that are covalently linked to one another. They also dictate internal flexibility, vibrational frequencies, and stable conformations of molecules. As outlined in equation 12, the bonded potential is normally divided into three primary terms; **valence bond length**, **valence angle**, and **torsional dihedral potentials** (figure 5).

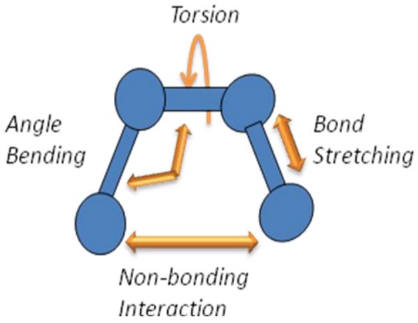


Figure 5: Chemical bonds between atoms.

1. Valence Bond Length (Bond stretching):

Covalent bond stretching between two atoms is simulated in classical terms as a harmonic oscillator model, or simply a spring following Hooke’s Law:

$$U_b(r)=\frac{1}{2}\sum_b K_b(r-b_0)^2 \tag{13}$$

Where K_b represents bond force constant or spring stiffness in Hooke’s description of a spring connecting two bodies (atoms). The variable r is the instantaneous distance between the atoms, while b_0 represents the ideal bond length in equilibrium. The potential can be described in graph 6. Any deviation from b0 results in a quadratic increase in potential energy, forcing the atom back toward their equilibrium distance.

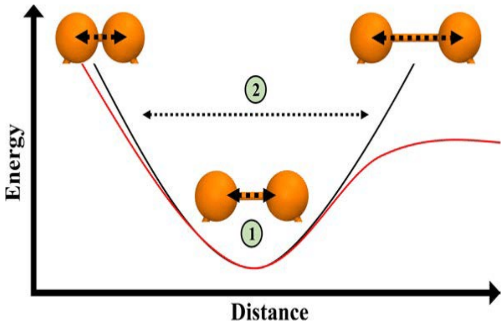


Figure 6: Bonded interaction between two covalently bonded atoms stretching. A harmonic potential is used to keep the distance between the two atoms at a reference value (bond length at equilibrium). [1]

Constraint algorithms are also used to ensure that vibration or twisting of atoms are within known parameters as to not have the simulation act erratically. Algorithms like **SHAKE** are often employed to constrain bond vibration.

2. Valence Angle Potential (Angle Bending):

The angle potential restricts the geometry of three bonded atoms to ensure the simulated molecules maintain their correct chemical geometry (such as 104.5 degrees bend of water molecule). It's also described harmonically following the equation:

$$U_\theta(r)=\frac{1}{2}\sum_\theta K_\theta(\theta-\theta_0)^2 \tag{14}$$

In this expression $\mathbf{K_\theta}$ is the angle bending force constant, $\boldsymbol{\theta}$ is the calculated instantaneous angle between the atoms, and $\boldsymbol{\theta_0}$ is the reference equilibrium angle (known angle between the two atoms at an equilibrium). This can be summarized in **figure 7**.

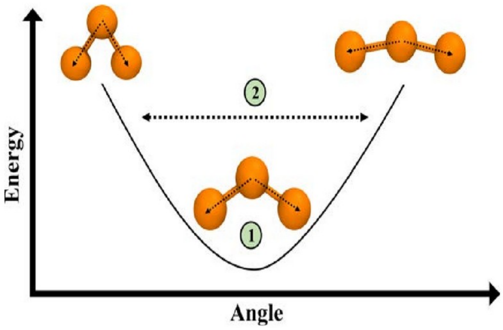


Figure 7: Bending Potential. A harmonic potential is used to keep the angle between the three atoms at its reference value. [1]

3. Valence dihedral torsion potential:

The torsional potential models the energy required to rotate a sequence of four covalently bonded atoms around a central bond. This is not a harmonic spring, but rather a periodic function because rotating bond by 360 degrees gets the molecule back to its original spatial arrangement and potential. It can be expressed as a cosine series:

$$U_\phi(r)=\frac{1}{2}\sum_\phi K_\phi[\cos(n\phi-\delta)+1] \tag{15}$$

Where $\mathbf{K_\phi}$ determines the energy barrier height to rotation (how hard it is for the bond to rotate), **n** is the multiplicity that indicates how many minimum energy valleys exist in a single 360 rotation, $\boldsymbol{\phi}$ is the current torsion/dihedral angle, and $\boldsymbol{\delta}$ is the phase shift that sets the angle where the potential energy is at its minimum (at equilibrium).

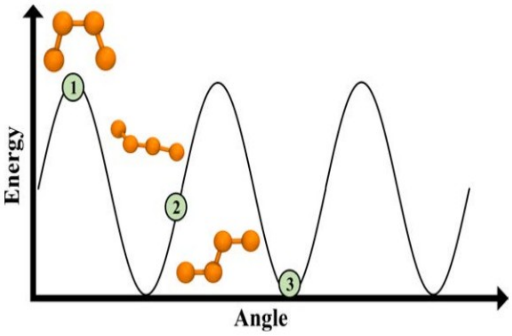


Figure 8: Proper dihedral (torsional) potential.

The torsional potential (**figure 8**) has an important role in forcing the molecule to settle into its known conformation. For example, it is used to impose the cis-trans isomerism o a molecule, the planarity of aromatic compounds, or the chirality of a stero-specific center. The potential is usually described by a dihedral angle potential. A proper dihedral angle is represented periodically by four consecutively boned atoms, while an improper dihedral angle is represented harmonically through a similar equation to angle bending potential, and is based on three consecutively bonded atoms connected to a fourth “central” atom.

D. Non-bonded Interactions

As seen before, bonded interactions describe the internal structure of a molecule, while non-bonded interactions describe the forces that act between atoms that are not directly connected through a covalent bond. Such interactions are crucial for modeling intermolecular forces, phase behaviors, and the 3D folding of large macromolecules. Non-bonded interactions represent most of the computational needs for molecular dynamics because of how intricate they are. The following are major types of non-bonded interactions.

1. Electrostatic (Long-ranged)

Electrostatic interactions are formed by unqual distribution of charge across molecules, and are universally described by Coulomb’s law, appearing in the potential function in equation 12 as:

$$U_{el}(r)=\sum_{ij}\frac{q_i q_j}{\varepsilon r_{ij}} \tag{16}$$

Where q_i and q_j are point charges of the interacting atoms, r_{ij} is the distance between them, and ε is the dielectric constant of the medium. Because electrostatic forces decay slowly when the interacting atoms are parted away from each other (1/r), they are considered a **long-range interactions** and often require specialized algorithms like he **Ewald Summation** which is specialized for calculating electrostatic interactions across periodic boundaries conditions (PBC).

The Ewald Summation of the electrostatic Coulomb potential is described by the following equation:

$$\frac{1}{4\pi\varepsilon_0}\sum_{n<j}^{N}\frac{q_i q_j}{r_{ij}} \tag{17}$$

And is divided into two sums, in the **wave-number space**:

$$\frac{1}{2V_0\,\varepsilon_0}\sum_{k\neq 0}^{\infty}\frac{e^{-k^2/4\alpha^2}}{k^2}\left|\sum_{j}^{N} q_i e^{(-ikr_j)}\right|^2 \tag{18}$$

In which the calculates the long-range interactions (background electric field), and maps the charges onto a grid and uses **Fourier transform** for this aim (as seen with the e^{-ikr} part). k represents wave vectors. There are multiple Fourier transform algorithm made for this purpose, including the well known algorithm Fastest Fourier Transform in the West (FFTW).

And in the **real space:**

$$\frac{1}{4\pi\varepsilon_0}\sum_{n<j}^{N}\frac{q_i q_j}{r_{ij}}erfc(\alpha r_{nj}) \tag{19}$$

Where erfc(αr_{nj}) is the complementary error function and α is a tuning parameter, and in essence works as a switch. When the interacting atoms are at their closest distance, this term is at its maximum value, and as the distance increases, the value also rapidly drops to zero. Thus, it will ignore the infinite copies in PBC.

Finally, the correction sums for the intermolecular (chemical bonds) interactions is calculated by:

$$-\frac{1}{4\pi\varepsilon_0}\sum_{molecules}\sum_{l\leq m}^{M} q_l q_m \{\delta_{lm}\frac{\alpha}{\sqrt{\pi}}+\frac{erfc(\alpha r_{mj})}{r_{lm}^{1-\delta_{lm}}}\} \quad (20)$$

Ewald summation is the most computationally intensive operation and often requires a dedicated hardware to complete it's calculations.

2. Hydrogen bond

Hydrogen bonds are strong and highly directional dipole-dipole type of interactions. They can be generally calculated by combining electrostatic and van der waals terms. Hydrogen bond can be estimated using quantum mechanics far more precisely, but for general applications, It can be estimated following the equation:

$$U_{HB}=\sum_{i,j}[\frac{A^{'}}{r_{ij}^{12}}-\frac{B^{'}}{r_{ij}^{10}}] \quad (21)$$

A' and B' are parameters that dictates repulsion and attraction for hydrogen bonding, respectively. Basically, hydrogen bond is simulated similarly to Van der Waals but with parameters for binding/repulsion strength changed at specific distances.

3. Van der Waals

Van der Waals interactions are probably one of the most important types of interactions in computational chemistry. They are experienced by all atoms in the simulation and are most famously estimated through **Lennard-Jones potential,** which has the following equation:

$$U_{LJ}(r_{ij})=4\,\varepsilon_{ij}[(\frac{\sigma}{r_{ij}})^{12}-(\frac{\sigma}{r_{ij}})^{6}] \quad (22)$$

Where $i\,j$ and $i\,j$ represent depth of potential well, as seen in the**figure 9.**

'

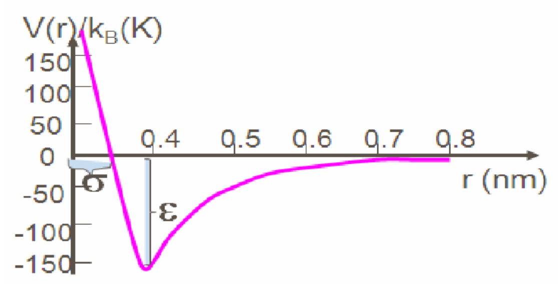


Figure 9: Lennard-Jones Potential energy dependence on the distance between two atoms.

Estimating interaction parameters of $_{ij}$ and $_{ij}$ between two **different** species based on known parameters of **pure** species (ii, and jj) can cause many issues in performing a simulation, and thus **Lorentz-Bertholt** mixing rules are used to estimate the these two parameters for the mixed species, and are calculated by:

$$\sigma_{ij}=\frac{\sigma_{ii}+\sigma_{jj}}{2},\ \varepsilon_{ij}=[\varepsilon_{ii}*\varepsilon_{jj}]^{1/2} \quad (23)$$

4. Metal oxides and Buckingham potential

The standard force fields like AMBER uses Lennard-jones potential for Van-der-Waals forces, but the term $1/r^{12}$ that specifies repulsion does not work properly because metal ions are large and more or less polarizable than normally studied atoms in biology (C,N,S,O, H, etc.). For this reason, a new potential field is given to metal ions following the equation:

$$U=A\exp\left(\frac{-r}{\rho}\right)-\frac{C}{r^6} \tag{24}$$

Where A, ρ, and C are parameters of the model tuned to the specific type of each atom pair.

E. Quantum Molecular Dynamics (qMD) description

Classical molecular dynamics are highly efficient, but they completely ignore electrons and electronic structures of atoms in the simulations, making it impossible to simulation chemical reactions (bond breaking/formation), electron transfer, or the coordination between metals and non-metal atoms. Quantum molecular dynamics (qMD) solves this by abandoning classical force fields in favor of quantum mechanics to calculate forces. Other types qMD like Ab initio quantum chemistry, and density function theory involve solution of Schrodinger equations.

Example, hydrogen molecule H_2; two proton (a & b) and two electrons (i=1.2); r_{12} &r_{ab} ; r is interatomic distance; the general formula for solving the potential field between them is through the following formulas.

$$E = E_a + E_b + U$$

$$U(r) = E(r) - E_a - E_b$$

$$E \sim \frac{\int \dots \int \psi^* \hat{\mathrm{H}} \psi \, dr_{a1} \dots dr_{b2} dr_{12}}{\int \dots \int \psi^* \psi \, dr_{a1} \dots dr_{b2} dr_{12}} \tag{25}$$

$$E \sim \frac{\langle \psi^* \mid \hat{\mathrm{H}} \mid \psi \rangle}{\langle \psi^* \mid \psi \rangle}$$

$$U(r) = \frac{\langle \psi^*(r) \mid \hat{\mathrm{H}} \mid \psi(r) \rangle}{\langle \psi^*(r) \mid \psi(r) \rangle} - E_a - E_b$$

IV. Monitoring of equilibrium states in MD simulations

MD Simulation results in a lot of data in the form of trajectories, and before any data can be extracted from the molecular dynamic simulation, the system must reach a state of thermodynamic equilibrium. This means the system has relaxed from the artificial starting configuration and its properties are fluctuating around a stable mean. Equilibrium monitoring is done via several metrics:

A. Structural metrics:

Such as **Root Mean Square Deviation (RMSD)** to track how much the structure's atomic coordinates deviate from the initial starting structure over time. A flattening RMSD curve indicate that the molecule has settled into a stable conformational state. The following equation describes **mass-weighted** RMSD is calculated:

$$RMSD=\sqrt{\frac{1}{M}\sum_{i}^{N} m_i \| r_{i,1} - r_{i,2} \|^2} \tag{26}$$

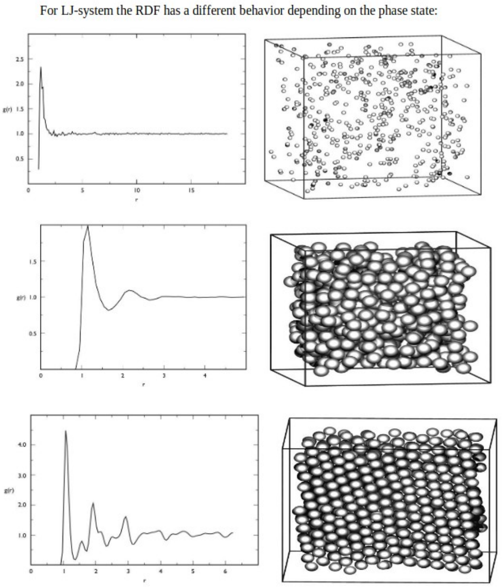
Massive heavy atoms like carbon or iron moving for example 2 A is far more significant than a light atom moving the same distance.

Where: $r_{i,1}$ is the 3D vector of atom i in starting reference frame, $r_{i,2}$ is the position of the exact same atom i at the current time step, m is mass of atom i, M is the total mass of all atoms combined.

Another parameter is the **Radial distribution function (RDF)**, which measures the local structure and packing of atoms. Basically, it shows the likelihood of finding another atom around a specific atom/molecule. It's particularly great when discussing solvation shells, where peaks show hydration shell formation. Also, in phase states, a solid crystal will have very sharp RDF peaks that go on for long distance, and a liquid will have few peaks that smooth out quickly into a flat line, while a gas will basically be a flat line because all atoms are distributed randomly (for an idealized condition) (**figure 10**). Radial distribution function can be calculated following the equation:

$$\rho g(r) = \frac{1}{N} \langle \sum_{i}^{N} \sum_{j \neq i}^{N} \delta[r - r_{ij}] \rangle \tag{27}$$

Where N is the total number of atoms, **rho** is the atomic density (N/V), r_{ij} is the radius vector between two centers I & j, g(r) is the radial distribution function, **delta** represents the **Dirac delta function,** < … > is the time/configuration average; for distance less than one atomic diameter g(r)=0, and for larger distances g(r) =1.



For LJ-system the RDF has a different behavior depending on the phase state:

Figure 10: RDF for different material phases

In addition, the function gamma is also used to distinguish equilibrium states:

$$\begin{gathered}
\gamma_x = \frac{1}{N} \sum \cos\left(\frac{4\pi x_i}{\alpha}\right) \\
\gamma_y = \frac{1}{N} \sum \cos\left(\frac{4\pi y_i}{\alpha}\right) \\
\gamma_z = \frac{1}{N} \sum \cos\left(\frac{4\pi z_i}{\alpha}\right) \\
\gamma = \frac{1}{3}[\gamma_x + \gamma_y + \gamma_z]
\end{gathered} \tag{28}$$

Finally, the H-function or Boltzmann distribution is used for monitoring the equilibrium:

$$H_x(t) = \int_{-\infty}^{+\infty} f(v_x) \ln f(v_x) dv_x \tag{29}$$

B. Thermodynamic properties:

The total energy, potential energy, kinetic energy, temperature, and volume/density can all be plotted over time. The system can be considered to be in equilibrium when these values plateau and only exhibit normal statistical fluctuations (**figure 11**).

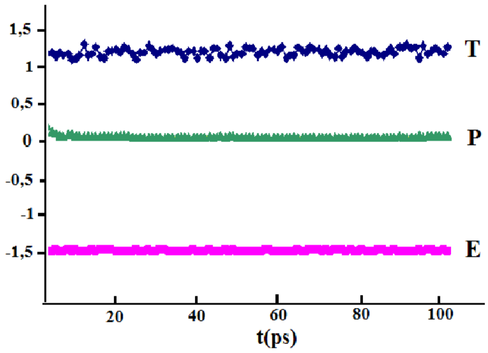


Figure 11: T,E,P measurement and used as a indicator of equilibrium.

VI. Simulation Ensembles (thermostats and barostats)

The simulated system is usually modeled to an external hot-bath as shown in the algorithm scheme above, as if that was not realized, the system temperature and pressure would be no adequately controlled. Usually, this is known as **micro-canonical ensemble (NVE)** which is characterized by constant number of atoms, volume, and total energy. It's usually good enough for many purposes, but does not match studied conditions in experiments. In actual experiments, factors like temperature, pH, pressure, external electric field, solvent, etc is extremly important factor in studying behaviors of molecules. **NVT** ensembles (constant number of atoms, volume, and temperature), **NpT** ensembles (constant number of atoms, pressure, and temperature), and finally constant-pH ensembles help address this issue.

A. Temperature Coupling

Temperature is connected to the average kinetic energy of the simulated particles and thus their velocities. The equipartition theorem gives the relation between these quantities as follows:

$$KE=\frac{1}{2}\sum_{i=1}^{N} m_i v_i^2=\frac{1}{2}\sum_{i=1}^{N} N_f k_B T \qquad (30)$$

Where N_f is the number of degrees of freedom, m_i iis the masses of particles, and k_B is the Boltzmann's constant. We can derive the equation of temperature by:

$$T=\sum_{i=1}^{N}\frac{m_i v_i^2}{N_f k_B T} \qquad (31)$$

The temperature can be kept constant by constantly rescaling velocities. And this is basically how the **Berendsen thermostat** works. Using the reference temperature $T_{0.\,,}$the system is kept at constant T by scaling velocities at a regular time interval $\tau_{T.}$ The Berendsen algorithm was briefly mentioned in the algorithmic scheme, but it can be summarized in the equations below:

$$\frac{dT}{dt}=\frac{T_0-T}{\tau_i}$$

$$\begin{gathered}
\sigma = \left\llbracket 1+\frac{\Delta t}{\tau_T}\left(\frac{T}{T_{ext}}-1\right)\right\rrbracket^{1/2} \\
\boldsymbol{v}\left(t+\frac{1}{2}\Delta t\right) = \left\llbracket \boldsymbol{v}\left(t-\frac{1}{2}\Delta t\right)+\Delta t\frac{\boldsymbol{f}(t)}{m}\right\rrbracket \sigma \\
\boldsymbol{v}(t) = \frac{1}{2}\left\llbracket \boldsymbol{v}\left(t-\frac{1}{2}\Delta t\right)+\boldsymbol{v}\left(t+\frac{1}{2}\Delta t\right)\right\rrbracket \\
\boldsymbol{r}(t+\Delta t) = \boldsymbol{r}(t)+\Delta t\,\boldsymbol{v}\left(t+\frac{1}{2}\Delta t\right)
\end{gathered} \tag{32}$$

The Berendsen thermostat is simple and robust for most applications. However it does not correctly reproduce the canonical NVT ensemble. For this reason, the is preferable to use it only during equilibration phase. For the production runs, instead, it is preferred to employ the **Nose-Hoover thermostat** as it is consistent with the canonical ensemble**.** The basic idea behind is coupling the system to an external heat bath with mass Q>0 and use a frictional coefficient μ to control velocities of particles. Below is the summarized algorithm used in Nose-Hoover thermostat:

$$\begin{gathered}
\sigma\left(t+\frac{1}{2}\Delta t\right) = \sigma\left(t-\frac{1}{2}\Delta t\right)+\frac{\Delta t}{\tau_T^2}\left(\frac{T}{T_{ext}}-1\right) \\
\sigma(t) = \frac{1}{2}\left\llbracket \sigma\left(t-\frac{1}{2}\Delta t\right)+\sigma\left(t+\frac{1}{2}\Delta t\right)\right\rrbracket \\
\boldsymbol{v}\left(t+\frac{1}{2}\Delta t\right) = \boldsymbol{v}\left(t-\frac{1}{2}\Delta t\right)+\Delta t\left\llbracket \frac{\boldsymbol{f}(t)}{m}-\sigma(t)\boldsymbol{v}(t)\right\rrbracket \\
\boldsymbol{v}(t) = \frac{1}{2}\left\llbracket \boldsymbol{v}\left(t-\frac{1}{2}\Delta t\right)+\boldsymbol{v}\left(t+\frac{1}{2}\Delta t\right)\right\rrbracket \\
\boldsymbol{r}(t+\Delta t) = \boldsymbol{r}(t)+\Delta t\,\boldsymbol{v}\left(t+\frac{1}{2}\Delta t\right)
\end{gathered} \tag{33}$$

Equations 5 → 8 explain the variables in both algorithms.

B. Pressure Coupling

The pressure in MD simulation can be maintained through the same way as with the Berendsen approach in that I rescales the pressure to match target value P_0:

$$\frac{dP}{dt}=\frac{P_0-P}{\tau_P} \tag{34}$$

Pressure is calculated with the theorem:

$$P=\frac{Nk_B\langle T\rangle}{V}+\frac{1}{3V}\langle\sum_{i-1}^{N}\vec{v}_i.\vec{F}_i\rangle \tag{35}$$

Barostats that modify equations of motions like Andersens or Parrinello-Rahman allow the volume of pressure to fluctuate around constant value.

VII. MD Simulation Packages and Force Fields

There are different packages and force field algorithms made for specific applications. For life sciences that relies heavily on large number of atoms that are mostly carbon, oxygen, hydrogen, and nitrogen; packages like AMBER are usually preferred. While dynamics that relies on potential energy from metal atom coordination with other atoms require more robust definitions of the system due to presence of hetero-atoms. Packages like DL_POLY address this issue. Variations in potential energy calculation and energy/pressure coupling are incorporated in several other potential fields like CHARMM and GROMOS.

A. AMBER

Assisted Model Building with Energy Refinement (AMBER) refers to a family of highly popular force field and comprehensive suite of molecular simulation programs. Heavily optimized for bimolecular systems, particularly DNA, RNA, and carbohydrates. AMBER force fields are generally an “all-atom” models where every single atom including the hydrogen is explicitly calculated. It’s highly parameterized for standard biological elements (C,H,N,O,P, and S), and thus incredibly accurate for folding dynamics and drug-protein binding. It’s less suited for complex solid-state inorganic materials.

B. DL_POLY

Developed by Daresbury Laboratory in the UK, it is highly versatile and general purpose MD package designed to primarily address the needs of material science and solid-state chemistry. Unlike AMBER, DL_POLY excels at handling complex systems containing hetero-atoms, metals, and ions. As mentioned, it incorporates specialized potential expressions for metal-oxides like Bukingham potential, or Morse potential for bond breaking, and many-body potentials like Sutton-Chen or EAM. This helps it accurately model metal surfaces and ionic diffusion, which is especially important for battery electrolyte research.

C. CHARMM

Chemistry at HARvard Macromolecular Mechanics (CHARMM) is another versatile simulation package with force fields addressing modeling of proteins, nucleic acids, and especially suitable for lipid bilayer and cell membranes. The CHARMM force field is famous for it parametrization processes that often realies heavily on quantum mechanical data to define its bonded and non-bonded terms. It also includes the Urey-Bradely term alongside the standard harmonic angle potentials to accurately model the vibration coupling between bonds and angles.

D. GROMOS

GROningen Molecular Simulation (GROMOS) is developed by university of Groningen in the Netherlands and it’s usually paired with simulation engines like GROMACs. GROMOS is unique in that it uses the united-atom approach, where rather than simulating every single hydrogne atom, GROMOS often lumps non-polar hydrogen atoms like those in CH_2 or CH_3 groups together with their parent carbon atom into a single slightly larger atom. This significantly reduces the total number of particles in the simulation and vastly improves calculation speed and cost while maintaining thermodynamic accuracy for large macromolecules/complexes.

E. OPLS

Optimized Potential for Liquid Simulation (OPLS) is developed by the Jorgensen group, and was originally specifically parameterized to reproduce the macroscopic properties of organic liquids (such as densities, heat of vaporization, and free energies of solvation). Other force fields focus heavily on gas-phase quantum mechanics for their parameters, while OPLS is unique in that it’s tuned to replicate experimental liquid-phase behaviors. OPLS-AA (all-atoms) is one of the most reliable force fields for simulating small organic molecules with mixed solvent environments, as well as complex liquid electrolytes.

VIII. Applications in future projects

Title: Aptamer design for small organic molecules biosensing

Aptamers are single stranded oligonucleotides that have specific 2D and 3D conformations that allows it to bind specifically to small organic molecules with high specificity (f**igure 12**). This unique capability makes aptamers an attractive area of research for biosensing and genetic riboswitches within the cell.

Aptamer sequences are mostly elucidated through a process known as Systematic Evolution of Ligands by Exponential Enrichment (SELEX), which is an in vitro process used to identify high-affinity and specific nucleic acid (DNA or RNA) to a target molecule. SELEX is expensive and time consuming for most labs to perform, thus, a computational approach is needed to help design aptamer sequences more effectively.

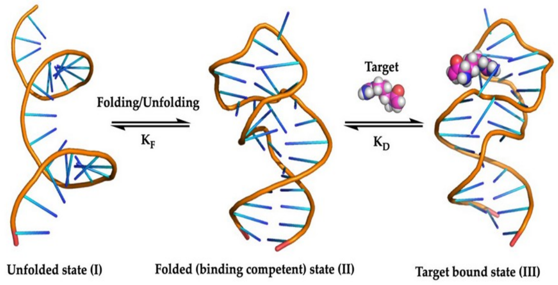


Figure 12: Aptamer binding with small organic molecules. [2]

This project will revolve around predicting and simulating binding of the designed oligonucleotides using packages like CHARMM or AMBER, as well as their function on top of a metal surface using packages like DL_POLY or qMD/MD hybrid approach. Solvation models like TIP3P or TIP4P will also be incorporated due its significance in the binding process. A thermostatted, barostated system with periodic boundaries will be applied, with attention to applying cutoff values around the aptamer structure and incorporating salts to the simulation like sodium and magnesium chloride. Ph adjustment is also important for ionization, thus the simulation will be kept at physiological pH. Figure 13 shows similar project.

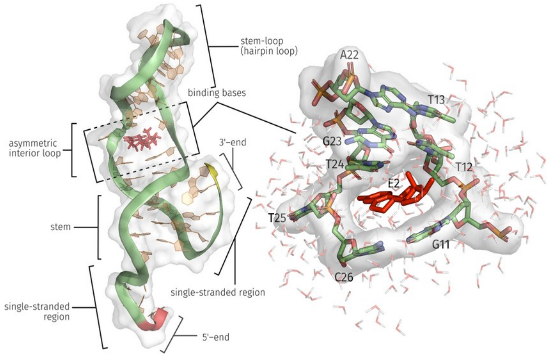


Figure 13: MD of 17Beta-estradiol with an Aptamer using GROMACs. [3]

This approach will help develop new biosensors for detection of fungal toxins, heavy metals, or even proteins in the environment or inside the cell itself.

IX. Acknowledgment

I want to express my sincere gratitude to Prof. Kholmirzo Kholmurodov for his valuable time and effort in teaching and guiding us throughout the program. I'm particularly grateful for his insights and thorough explanations of the theoretical aspects of molecular dynamics and keenness in engaging in discussions to help us understand better. I would also like to thank JINR for providing this valuable opportunity to me, which will be a substantial catalyst in my career.

X. References

1. Lolicato, Fabio. (2020). Exploring the Plasma Membrane via Computer Simulations. 10.13140/RG.2.2.16521.88168.

2. Elskens, J. P., Elskens, J. M., & Madder, A. (2020). Chemical Modification of Aptamers for Increased Binding Affinity in Diagnostic Applications: Current Status and Future Prospects. *International Journal of Molecular Sciences*, *21*(12), 4522. https://doi.org/10.3390/ijms21124522

3.Eisold, A., & Labudde, D. (2018). Detailed Analysis of 17β-Estradiol-Aptamer Interactions: A Molecular Dynamics Simulation Study. *Molecules*, *23*(7), 1690. https://doi.org/10.3390/molecules23071690

4. Mathematical modeling for MD and background were reviewed from Prof. Kholmurodov's literature:

K. Kholmurodov, Computer Design for New Drugs and Materials: Molecular Dynamics of Nanoscale Phenomena, New York: Nova Science Publishers, 2017.

K. T. Kholmurodov, ed., Computational Materials and Biological Sciences. NewYork: Nova Science Publishers, 2015.

K. T. Kholmurodov, ed., Models in Bioscience and Materials Research: Molecular Dynamics and Related Techniques. New York: Nova Science Publishers, 2013.

K. T. Kholmurodov, ed., Molecular Simulation in Material and Biological Research. New York: Nova Science Publishers, 2009