JOINT INSTITUTE FOR NUCLEAR RESEARCH
Frank Laboratory of Neutron Physics

FINAL REPORT
ON STAGE 1 OF THE INTERNATIONAL STUDENT PRACTICE

"Introductory course: MD-simulation research (from atomic fragments to molecular compound)"

Supervisor:
Prof. Kholmirzo Kholmurodov

Student(s):
1- Mostafa Mahmoud Omran Mohamed (Cairo University)
2- Khlood Ahmed Ali Abdeljawaad (Mina University)
3- Mostafa Saad Ali Ebied (South Valley University)
4- Mohamed Mohey Mahmoud Sarhan (Suez canal University)

Participation period:
01-24 March

Dubna, 2022

Table of Contents

1 Abstract

Molecular dynamics (MD) is a computer simulation tool for examining atom and molecule physical motions. This technique is largely used in chemical physics, materials science, and biophysics. This report outlines the fundamental principles of molecular dynamics that have been acquired throughout our participation in this course. We studied the fundamental equations and strategies for using force fields to simulate different molecular systems in different filed such as material science and biology.

2 Project goals

The following summarizes the goal of this course:

1.Introduction to fundamental equations, potentials, and simulation approaches

2. The description of computer code for simulation of liquid model (Lenard-Jones potential.

3. The application of a general-purpose code to the modeling of molecular ionic, polymeric, and biological systems

4. study the theoretical foundations of the hybrid MD methodology (methods for simulating classical quantum-chemistry potentials)

5. and finally, MD test simulation.

3 Introduction

3.1 The basic equations, potentials and simulation techniques

The determination of a geometrically optimal structure of a molecule means the determination the structure with minimum free energy. The equilibrium free energy of a molecular structure is calculated by the use of molecular force fields. In computational physics and molecular modelling, a molecular force field is a mathematical function that describes the dependence of the potential energy of a molecule on the coordinates of its atoms. It is specified by an analytical form of the intermolecular potential energy: U(r1,r2,…,rN) and a set of input parameters. Force fields differ in their functional form as well as their fixed parameter sets. The parameter values are obtained either from quantum mechanical calculations (ab initio or semi-empirical methods), or by fitting

to experimentally determined high resolution structures, determined by X-ray diffraction, nuclear magnetic resonance (NMR), infrared and Raman spectroscopy, and other methods.

In Molecular Dynamics Simulations, Newton's equations of motion are integrated numerically for typically 100-10000 particles, starting from an initial configuration (initial positions and velocities for all particles). The atomic force field describes physical systems as collection of atoms kept together by interatomic forces. The interaction law is specified by the potential , which represents the potential energy of N interacting atoms as a function of their positions . The force acting upon the atom is then determined:

$$F_i(r) = -\frac{\partial U(r)}{\partial r_i} \quad (1)$$

Force Field can be understood as an empirical set of energy functions. It is typically the summation of bonded and non-bonded terms or covalent and non-covalent interactions among atoms and molecules

$$U(r) = U_b + U_\theta + U_\varphi + U_\omega + U_{LJ} + U_{el} + U_{HB} + \ldots \quad (2)$$

The solution of Newton's equations of motion gives the time evolution of a set of interacting particles:

$$m_i \frac{d^2 r_i(t)}{dt^2} = \boldsymbol{F_i}(\boldsymbol{r}), \;\; i = 1,2, \ldots n \, . \quad (3)$$

To solve the second-order differential equations the initial positions and velocities of the particles must be determined. The equations are discretized and solved numerically and the MD trajectories are defined by both position and velocity vectors. Position vectors $r_i(t)$ determine the changing in time positions, while the velocity vectors $v_i(t)$ determine the kinetic energy and temperature in the system.

Kinetic energy can be given by the Maxwell-Boltzmann velocity distribution:

$$E_k = \sum_{i=1}^{N} \frac{m\, v_i^2}{2} = \frac{3}{2}\, N K_b\, T \quad (4)$$

For the integration of Newton's equations of motion (Eq. 3), we can employ algorithms such as the Verlet algorithm. It is the most basic and widely used algorithm and its actually the two-third order Taylor Series expansion of the coordinate of a particle at time (t+Δt) and (t-Δt):

$$r_i(r + \Delta t) \approx 2\, r_i(t) - r_i(t - \Delta t) + \frac{F_i(t)}{m_i}\Delta t^2 \qquad (5)$$

Some of the advantages of this algorithm are that it is self-starting, requires less computer memory, it is straightforward, and the new positions are easily acquired from the previous ones. Other examples of integration algorithms are Euler algorithm, Leap Frog algorithm and Velocity Verlet schemes, Beeman's algorithm, although the Euler algorithm is not time-reversible and suffers from a tremendous energy drift.

3.2 Electrostatics - Central Problem of MD Simulations. Periodic Boundary Conditions

New velocities $\boldsymbol{v_{i+1}}$ and positions $\boldsymbol{r_{i+1}}$ are found in the stepwise numerical integration procedure by computing the forces acting upon the atoms at each step. The force fields include long-range electrostatic and dispersion interactions and summation of order N^2 has to be performed to account for all non-bonded pairs. The complexity of the algorithm consists of calculating millions of Coulombic interactions for multibody systems at each time step. There are different techniques that can be implemented to deal with the problem. Because of limited computer memory only a finite sample of an infinite system can be represented in a computer method.

The "cut-off" method tapers the interaction potential over a predefined range of distances. It usually works well with the van Der Waals interactions, but the Coulombic forces are long-range.

The treatment of long-range forces is related to the choice of boundary conditions. The two common approaches are based on either *periodic boundary conditions* and Ewald method for lattice summations or on *spherical boundary conditions* and the reaction field method. The *Ewald summation* for Coulomb interaction is a correct approach taking into account periodicity. In the Ewald summation the electrostatic Coulomb potential is divided into two sums - in wave-number space and in real space.

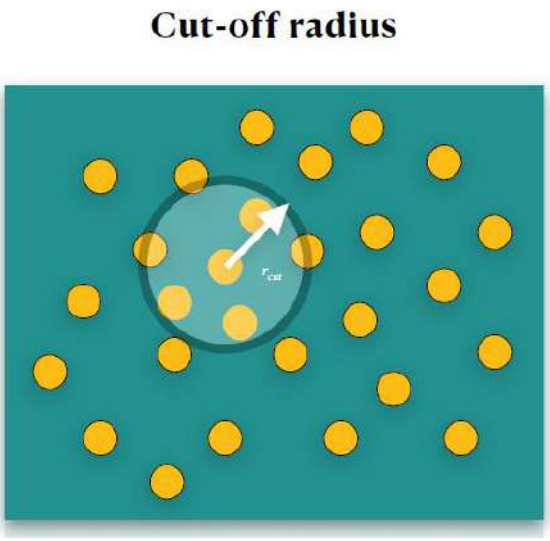
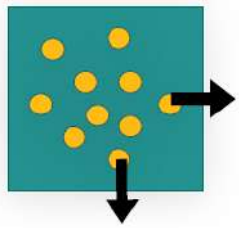


Figure 1:cut off radius

Let's imagine that the simulation of a system is within a box-shaped container. There is a high possibility that a few particles leave the box, since we examine a dynamical fluid system.



In order to overcome the issue, we generate a replica of the box that covers it from all sides. Now, whenever a particle tries to move out from the central box another will enter its place with the same speed in order to maintain a balanced system. Periodic boundary conditions enable a simulation to be performed using smaller number of particles in such a way that particles experience forces as if they are in bulk fluid.

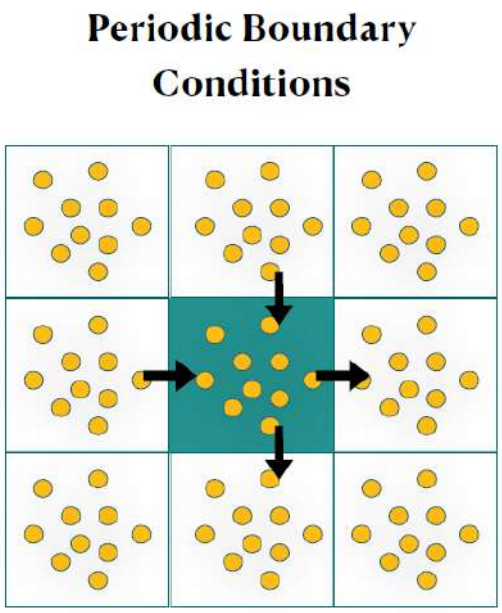


Figure 2:Periodic Boundary

3.3 Different potential energies in a molecule

Any molecule (**Fig. 3**) is characterized by the presence of a bond stretching between two atoms, an angle bending of three atoms, and a fixed torsion of four atoms. In addition to chemical bonds, there is the participation of unbound van der Waals interactions (non-bonding interactions), and if the atoms also have a charge, also electrostatic forces, and potentials (Coulomb interactions). Then the total potential energy[1]:

$$U(r) = U_b + U_\theta + U_\varphi + U_\omega + U_{LJ} + U_{el} + U_{HB} + \ldots \quad (6)$$

Using such a force field model, macromolecules are reduced to a set of atoms held together by simple harmonic forces, Coulombic interactions, and van der Waals interactions. For practical calculations, the force field must be simple enough to be evaluated quickly, but sufficiently detailed to reproduce realistic structural properties.

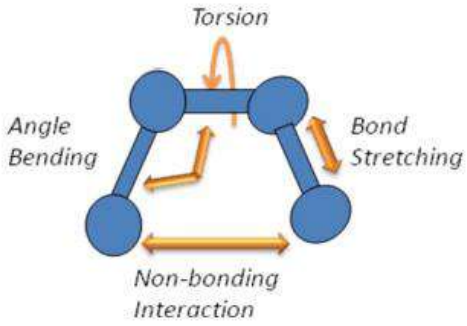


Figure 3: Different types of interactions

3.3.1 Valence Length potential U_b:

For the covalent bond terms, parametrized by bond length, bond angles and dihedral angles, the potential energy is described relative to the atoms being in their equilibrium positions, for which the energy is taken to be zero. The first term in the molecular force field describes the extension (stretching) of covalent bonds. Bond stretching is often represented by a simple harmonic function that controls the length of covalent bonds. The spring constant (K_b) can be estimated from infrared or Raman spectra.

$$U_b = \frac{1}{2}\sum_b K_b(r - b_0)^2 \tag{7}$$

3.3.2 Valence Angle Potential U_θ:

The second force field term describes the distortion of the bond angles. Distortion of bond angles is described by the energy related to bending an angle, formed by at least three atoms: A-B-C, where there is a chemical bond between A and B, and between B and C. As in the case of bond stretching, the angle bending term is expanded as a Taylor series around the equilibrium bond angle and terminated at the second order (harmonic approximation). The vibrational frequencies are in the near infrared spectrum, and the constant K_θ is measured by Raman spectra.

$$U_\theta = \frac{1}{2}\sum_\theta K_\theta(\theta - \theta_0)^2 \tag{8}$$

3.3.3 Torsion Dihedral Potential U_φ:

The third force field term describes the distortion of dihedral angles from their preferred values. If a molecule contains more than four atoms in a row, which is a given in macromolecules, the dihedral term must be included in the force field. Dihedral angles are angles of rotation of the

bonded atom pairs around the central bond. In stereochemistry, the dihedral is defined as the angle between planes through two sets of three atoms, which have two atoms in common. Changes in dihedral angles often result in major conformational changes.

$$U_{\varphi} = \frac{1}{2}\sum_{\varphi} K_{\varphi}[\cos(n\varphi - \delta) + 1] \tag{9}$$

3.3.4 Electrostatics Potential U_{el}:

The fourth term describes the electrostatic forces arising between atoms carrying a ionic charges. These interactions between positive and negative ions are called salt bridges, which play a significant role in protein structure stabilization. Since substitution of basic residues for acidic residues changes the charge from positive to negative, such changes are extremely destabilizing when they occur in the interior of the protein. They tend to be more acceptable on the protein surface where the charged residues interact with polar water molecules and charged solutes.

$$U_{el} = \sum_{i,j} \frac{q_i q_j}{r_{ij}} \tag{10}$$

3.3.5 Van der Waals Interaction Potential U_{lj}:

The fifth term of the force field describes van der Waals forces. The movement of the electrons around the atomic nucleus creates an electric dipole moment. This dipole polarises neighboring atoms, which results in a short-range attractive force between the non-bonded atoms.

$$U_{LJ} = \sum_{i,j} \left(\frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^{6}} \right) \tag{11}$$

The equilibrium geometry of a molecule (with respect to bond lengths, angles, non-overlapping van der Waals spheres, etc.) describes the coordinates of a minimum on the potential energy surface. The minimum of the potential energy function corresponds to the equilibrium geometry of the molecule. An advantage of the molecular force fields method is the speed with which calculations can be performed, enabling its application to large biomolecules. With even moderate computer power, the energies of molecules with thousands of atoms can be optimized. This facilitates the molecular modelling of proteins and nucleotide acids, which is currently done by most pharmaceutical companies[2].

3.4 The simulation of liquid model (Lenard-Jones potential)

The Lennard-Jones potential is an intermolecular pair potential. Among the intermolecular potentials, the Lennard-Jones potential is the potential that has been studied most extensively and most thoroughly. It is considered an archetype model for simple yet realistic intermolecular interactions. The commonly used expression for the Lennard-Jones potential is:

$$U_{LJ} = \sum_{i,j} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right) \tag{12}$$

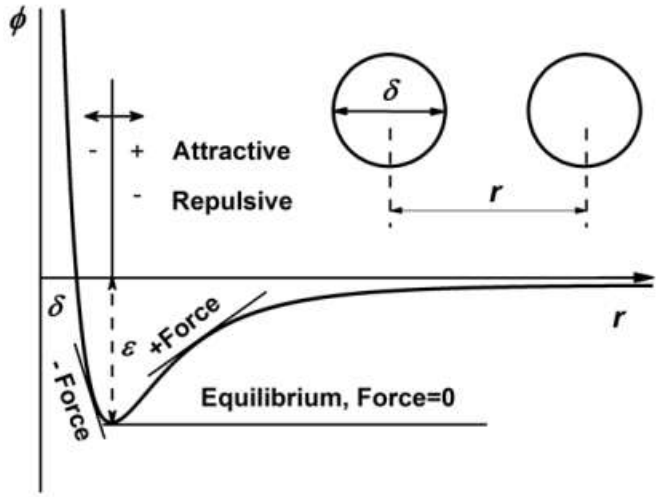


Figure 4: Lenard-Jones potential

4 MD Simulation packages

List of some common codes of multipurpose MD simulation programs which include both classical and quantum chemical methods and algorithms, is presented below:

(1) **AMBER** (www.ambermd.org) The Amber software package (Assisted Model Building with Energy Refinement) consists of a set of force fields for modeling macromolecular structures (proteins, nucleic acids and a number of other classes of molecules) and a package of quantum and molecular mechanics programs. The package is in the public domain.

(2) **CHARMM** (www.charmm.org) (Chemistry at HARvard Macromolecular mechanics) software package for molecular modeling of a wide range of systems - from small molecules to biological macromolecules, using various energy functions and models - from quantum models and force fields to molecular mechanics to full-atomic classical potentials.

(3) **DL_POLY** (www.cse.scitech.ac.uk/ccg/software/DL_POLY/) A package for modeling the molecular dynamics of complex systems with both sequential and parallel calculations. Versions available: DL_POLY_2, DL_POLY_3 and DL_POLY_4. Parallel calculations with the number of atoms up to 1 million using 1024 processors are possible. Adapted for graphics game processors, GPU (Graphical Processing Units), using the CUDA language. Freely available for research and educational purposes.

(4) **GROMACS** (www.gromacs.org) A software package for fast simulation of the dynamics of large molecular systems (from thousands to millions of particles). Designed primarily for modeling biomolecules (proteins and lipids) that have many interconnected interactions between atoms. Works in Linux environment and is free.

(5) **LAMMPS** (lammps.sandia.gov) The non-commercial package LAMMPS (Large scale Atomic Molecular Massively Parallel Simulator) uses classical molecular dynamics methods for modeling and calculating polymers, biomolecules, solids (metals, semiconductors, etc.), as well as coarse-grained mesoscopic systems at atomic, mesoscopic and continual scales.

(6) **MOE** (www.chemcomp.com) MOE (Molecular Operating Environment) is a complex of programs for modeling molecules, in particular large biomolecules. The methods of molecular mechanics and dynamics are developed in it on the basis of various force fields.

(7) **NAMD** (www.ks.uiuc.edu/Research/namd/) An object-oriented program for calculations in the field of interactive molecular dynamics, in particular for modeling large biomolecular systems that require significant resources. The program code is freely distributed for various parallel computing platforms.

5 Application of molecular dynamics

5.1 Molecular dynamics in biology

5.1.1 Application of MD simulations of future target in Drug Discovery (Coarse Grained MD simulations)

Scientific computing in biomolecular simulations has made great strides in the past decade. These achievements include advancements in the hardware, software, and force fields available to model biomolecular and polymer systems[3].

For example, quantum mechanical (QM) calculations access the microscopic scales in space and time to study small molecules and clusters, while atomic-resolution molecular dynamics (MD) classically study systems on the nanometer and nanosecond scales, such as proteins and lipid bilayers. In order to push the temporal scale of atomic-resolution molecular dynamics, algorithms that enhance conformational sampling of the system have been employed[4]–[6]. Additionally, the use of General Purpose computation on Graphics Processing Units (GPGPUs) is extending the temporal scale of QM and MD calculations. While these techniques have been successful in expanding the accessible temporal scale, the calculation of atomic-resolution potentials is not practical for larger biomolecular systems such as lipid rafts and viruses.

In order to push the limits of accessible temporal and spatial scales of biomolecular systems, coarse-grained (CG) models can be (and have been) utilized. To coarse-grain a system is to create a simplified, lower resolution model of the system. This is achieved by grouping clusters of atoms into a new, simplified CG bead (or pseudo-atoms). The level of coarse-graining depends on how many atoms are represented by a CG bead. Increasing the atom-to-bead ratio increases the degree of coarse-graining, yielding a lower resolution model. Collapsing the representation of atoms into CG beads effectively lowers the total number of degrees of freedom represented in the system. In addition, the CG beads interact with one another through more computationally efficient potentials. Together, these features yield a substantial increase in accessibility of time and space of biomolecular simulations.

Due to the fact of being a complex system (lipid bilayer) that would not be easily simulated using all atom simulations, the coarse grained simulations will be applied.

The aim of the project would be investigating the changes into the big vesicles (membrane self-organization) happened during the thermodynamic phase transitions of lipids and only in the presence of the Ab-peptide (amyloid-beta peptide), which unambiguously indicates the destructive action of the peptide. We interpreted the dramatic changes in the membrane's overall shape with parallel changes in its thickness as the amyloid-beta peptide triggered membrane disruption and a consequent reorganization of its structure. Our observations are consistent with membrane disruption and its subsequent reformation being caused by the separate peptides or small size peptide oligomers rather than the result of large fibrils. So far, understanding the mechanism of such interactions appears to be highly actual with regards to describing the onset of Alzheimer's Disease (AD). The proposed research project

related to the morphological changes of the Ab-peptide loaded into big vesicle objects (membrane) and will therefore be the direction of our further research.

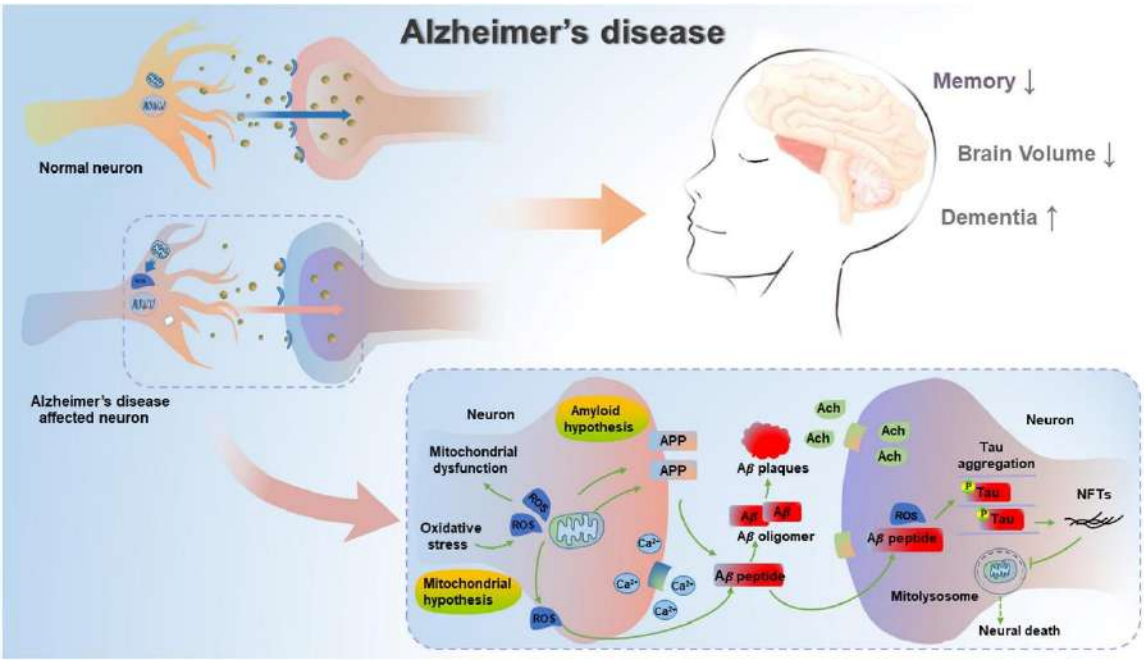


Figure 5:Alzheimer's disease mechanism.

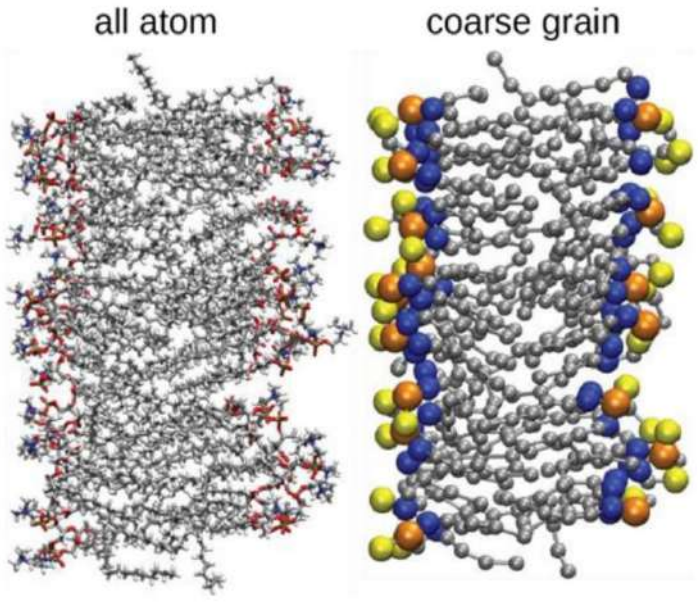


Figure 6:All atom and coarse grain lipid bilayer.

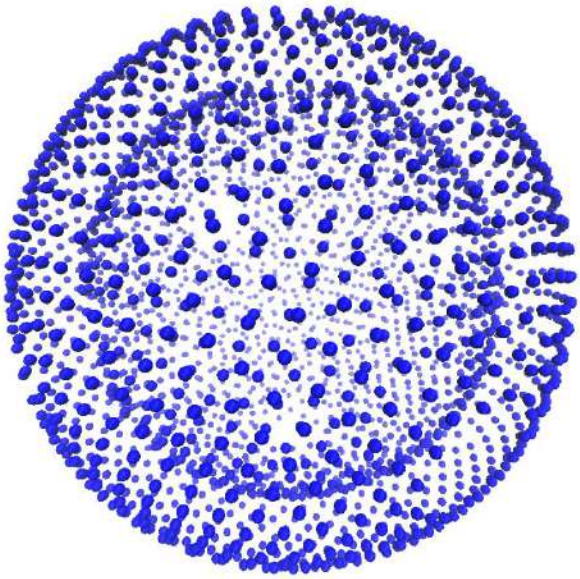


Figure 7:DPPC lipid bilayer Vesicle.

5.1.2 MD simulation in DNA

DNA is the genetic matter of all living things. Its usual common shape is double stranded helix. It consists of nucleotides connected together with covalent bonds in the same strand and with H-bonds in the opposite strand. Therefore, DNA isn't just a static structure, but it a dynamic structure with different conformational changes in space.

In this case, we used DL_POLY classic software for controlling the condition of the simulation e.g., temperature, pressure, solvation state, trajectory…. etc. In addition, visualization is done by VMD.

Two simulation conditions are performed for DNA helix:

- The simulation is done at 0 kelvin and under vacuum. The helical structure of the DNA changes with the change of temperature mainly from (0-77 °C) as DNA twisting is inversely proportional to the temperature increase. Therefore, by controlling the temperature we can control the degree of DNA helix twisting which can be beneficial in controlling the experimental conditions.

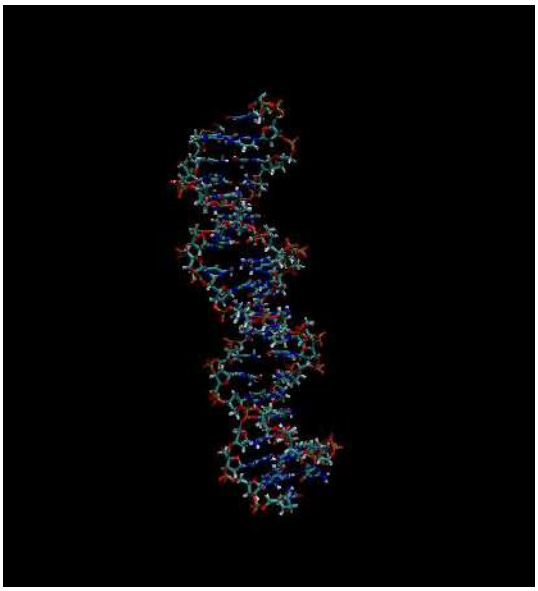


Figure 8:Zero Kelvin structure optimization of DNA

- The second condition is of DNA helix in solution of 706 (SPC) water molecules. A strand of DNA with a length of 1260 atoms is suspended in a solution of 706 (SPC) water molecules. Hexagonal prism periodic boundary conditions are used to align the DNA in the Z-direction. The Smoothed Particle Mesh Ewald method is used to calculate the electrostatic interactions. It's worth noting that the system is substantially anisotropic and has a strong overall negative charge.in terms of distribution The Dreiding force field is used for short-range forces and limitations.

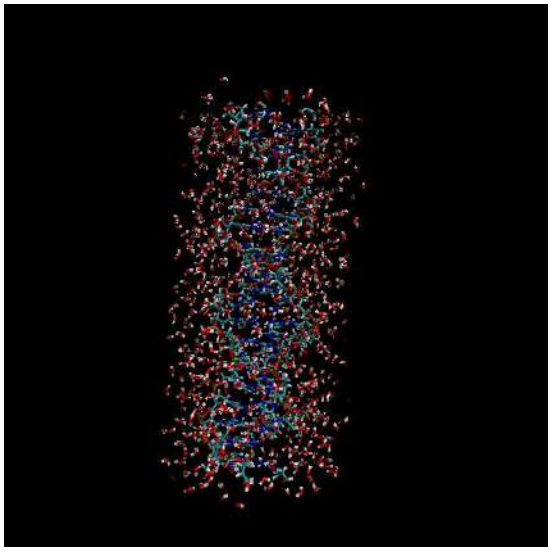


Figure 9:DNA strand in water

5.1.3 MD simulation of hen egg-white LYSOZYME

We used GROMACS, VMD for visualization and Origin for graphs. 1AKI.bdp file is used. Through GROMACS, the protein is solvated, ionized and energy minimized. Then, it undergoes NVT and NPT equilibration and finally, we performed MD simulation and analysis.

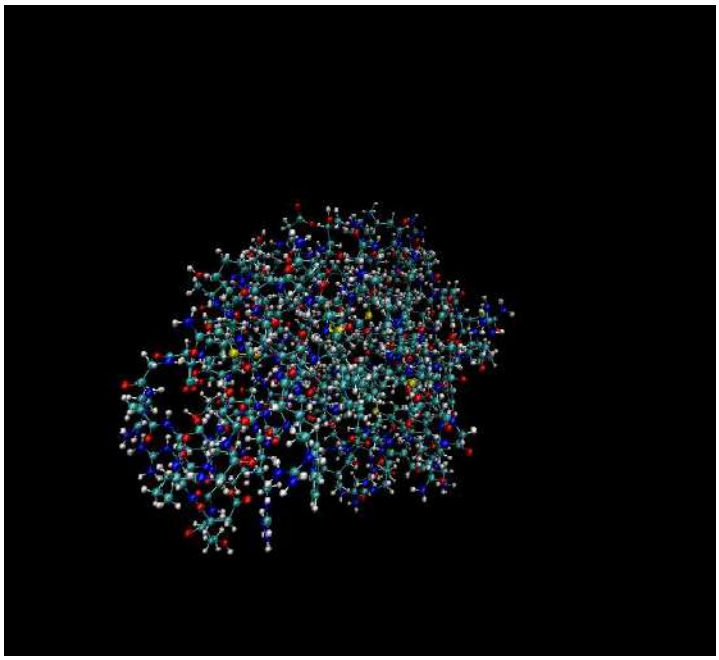


Figure 10:The structure of Lysozyme without solvent.

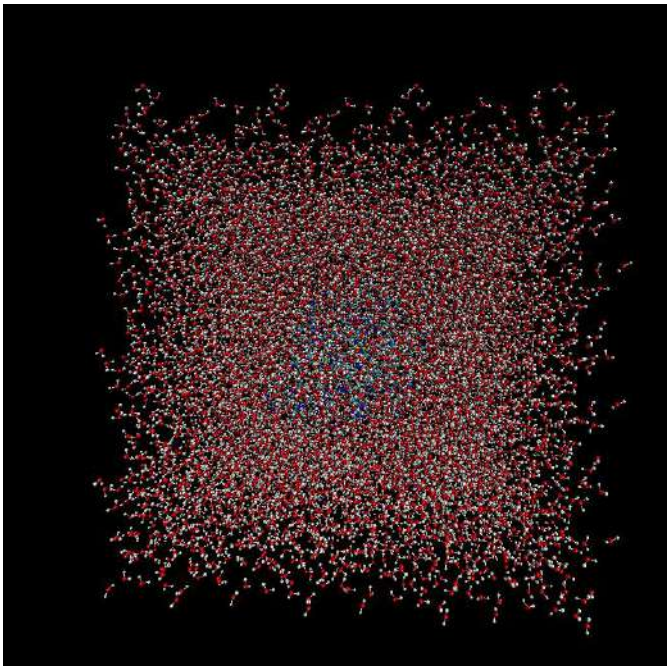


Figure 11:the structure after solvation, ion insertion and energy minimization.

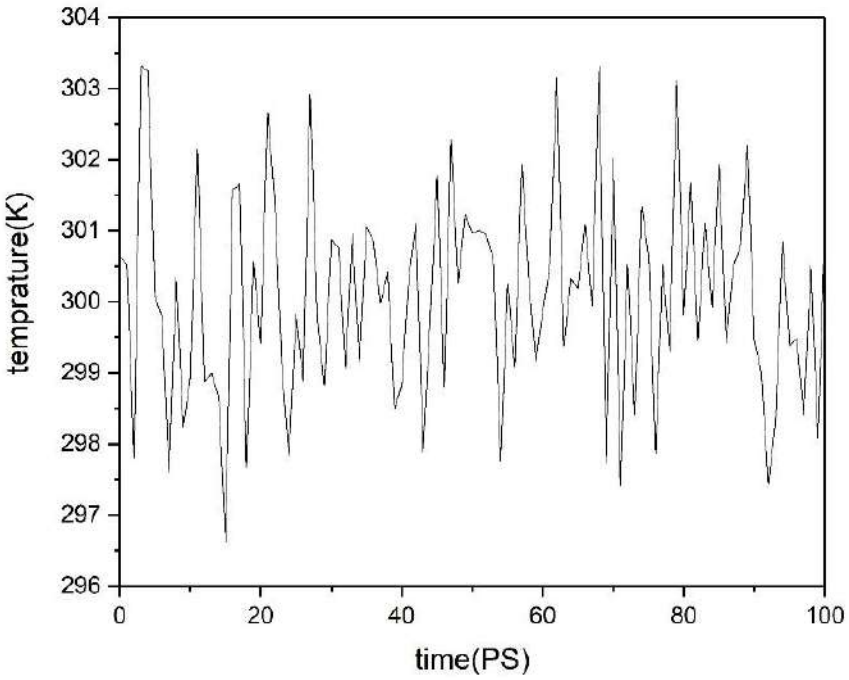


Figure 12:plotting of temperature (K) VS time (PS).

Fig. 12 shows that the system quickly reaches the target temp. (300 K) and remains stable over the remainder of the Equilibration. Only 50 PS may be enough for equilibration.

5.2 Molecular dynamics in Material Science

5.2.1 $KNaSi_2O_5$ Potassium Sodium disilicate glass:

Molecular dynamics simulation of $KNaSi_2O_5$ Potassium Sodium disilicate glass ($NaKSi_2O_5$) has been done using two body potentials using Dl_POLY classic package. Some of the two body potentials are read from the TABLE file. Electrostatics are handled by a multiple timestep Ewald sum method. Cubic periodic boundaries are in use. NVE ensemble. **Fig. 13** show the structure of $KNaSi_2O_5$ glass using VMD software. MD simulation has been performed on this structure in two different temperatures 300 K (room temperature) and 1400 K to study the effect of temperature on the strcuture. **Fig. 14** shows the total pair distribution function of $KNaSi_2O_3$ at the different temperatures which show that by increasing the temperature the PDFs peak start be wider and the intensity decreases indicating that the amorphous nature of $KNaSi_2O_3$ has been increased.

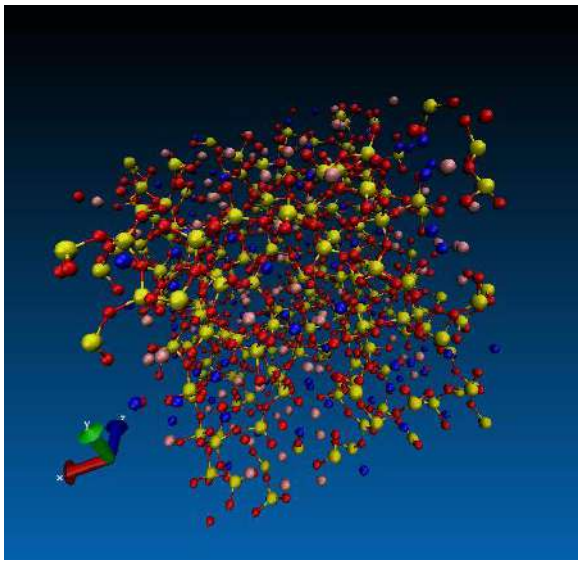


Figure 13: KNaSi2O5 glass 3D Strcture

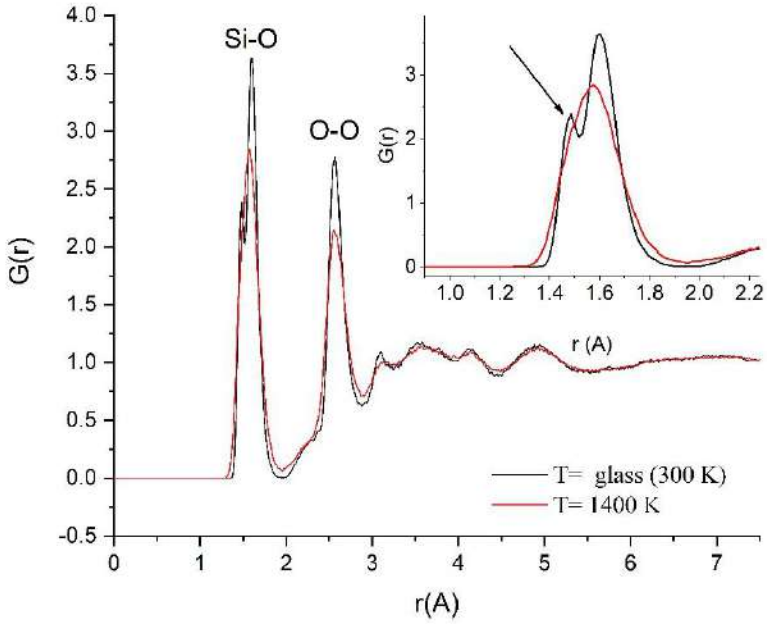


Figure 14: Total Pair Distribution Function G(r) of $KNaSi_2O_5$ glass at different temperatures.

5.2.2 MD Simulation in Catalysis

Molecular dynamics simulation has a wide variety of applications. Chemical catalysis simulation using molecular dynamics is one of the core applications of MD simulations, the nature and mechanism of liquid-liquid and solid-liquid interactions can be simulated and compared with the

practical observations. Several studies have been done for simulating the different interactions included in the multicomponent systems to get an additional understanding of the catalytic process.

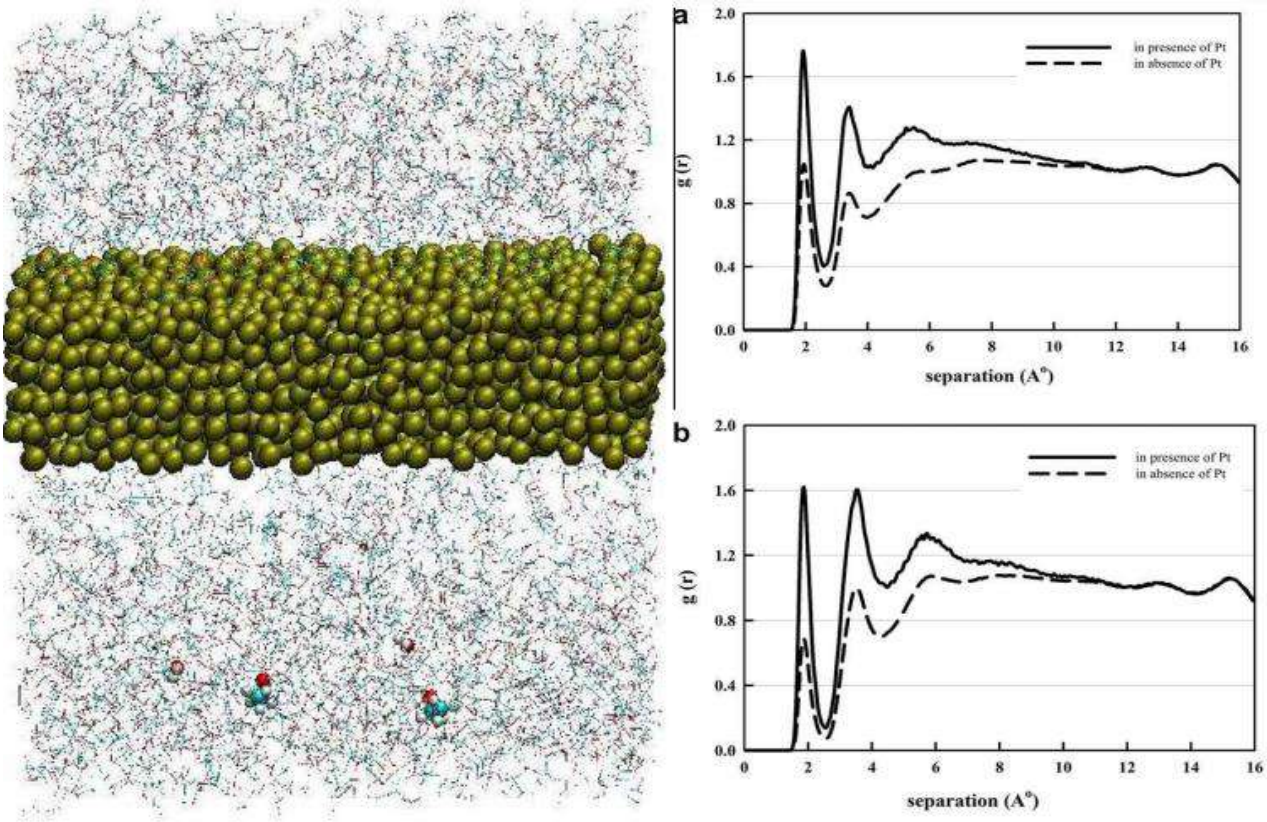


Figure 15:Snapshot of an ethanol–water–Pt (1 1 1) system.

An analysis of the molecular dynamics of ethanol solvated by water molecules in the absence and presence of the Pt (1 1 1) surface has been performed using DL_POLY , to observe the adsorption and diffusional behavior of water-ethanol mixture on the Pt (1 1 1) surface based on their affinity towards a specific type of surfaces in addition to the hydrophobicity and hydrophilicity of the metal surface, as showed in **Fig. 15**. The structure and diffusion properties of an ethanol–water system have been studied at various temperatures from 250 to 350 K. Also, the self-diffusion coefficients of a 50:50% ethanol–water system have been measured; in the absence of a Pt surface, the results have shown an excellent agreement with the experimental data (within an error of 7.4%). The enhancement of self-diffusion coefficients with the inclusion of the Pt (1 1 1) surface has been observed and estimated. Graphs of radial distribution functions (RDF) have been built; RDF correlations with the self-diffusion coefficients of both ethanol and water molecules are also illustrated as showed in **Fig. 15**.

6 Future work

In our plan for the near future, we will use the knowledge gained in this project in simulating synthetic polymer surfaces and nanoparticles for application in the field of material science and catalysis. Also, using MD as an important tool in Drug discovery.

7 Conclusion

The molecular modeling and computer design of chemical nanostructures, systems, and compounds have been studied. During the course, the following tasks were investigated: the basic equations, potentials, and simulation techniques, the simulation of a liquid model (Lenard-Jones potential), the use of selected general-purpose code for the simulation of ionic, polymeric, and biochemical molecular systems, the theory of the basics of hybrid MD approach (classical quantum-chemistry potentials simulation methods), MD test modeling.

8 Acknowledgments

We would like to thank our supervisor, Prof. Dr. Kholmirzo Kholmurodov, for allowing us to be a part of such an interesting project and for educating us about molecular dynamic modeling. His explanations and assistance aided us throughout the duration of our task. Also, we would like to express our heartfelt thanks to the Academy of Scientific Research and Technology (ASRT) and the Joint Institute for Nuclear Research (JINR) for giving such an opportunity for young researchers to broaden their knowledge of nuclear research and its applications.

9 References

[1] K. T. Kholmurodov, “Computer design for new drugs and materials: Molecular dynamics of nanoscale phenomena,” in *Computer Design for New Drugs and Materials: Molecular Dynamics of Nanoscale Phenomena*, 2017, pp. 1–231.

[2] M. Wiltgen, “Algorithms for structure comparison and analysis: homology modelling of proteins,” *Encycl. Bioinforma. Comput. Biol. Ranganathan, S., Gribskov, M., Nakai, K., Schönbach, C., Eds*, pp. 38–61, 2018.

[3] P. Sherwood, B. R. Brooks, and M. S. Sansom, “Multiscale methods for macromolecular simulations,” *Curr. Opin. Struct. Biol.*, vol. 18, no. 5, pp. 630–640, Oct. 2008, doi:

10.1016/j.sbi.2008.07.003.

[4] A. F. Voter, “A method for accelerating the molecular dynamics simulation of infrequent events,” *J. Chem. Phys.*, vol. 106, no. 11, pp. 4665–4677, Mar. 1997, doi: 10.1063/1.473503.

[5] A. F. Voter, “Parallel replica method for dynamics of infrequent events,” *Phys. Rev. B*, vol. 57, no. 22, pp. R13985–R13988, Jun. 1998, doi: 10.1103/PhysRevB.57.R13985.

[6] M. R. So/rensen and A. F. Voter, “Temperature-accelerated dynamics for simulation of infrequent events,” *J. Chem. Phys.*, vol. 112, no. 21, pp. 9599–9606, Jun. 2000, doi: 10.1063/1.481576.