INTERNSHIP REPORT

A report submitted for completion of

INTErnational **RE**mote **S**tudent **T**raining

By

JATIN RAWAT

Email: che21j.rawat@stuiocb.ictmumbai.edu.in

Under the supervision of
Dr. Kholmirzo Kholmurodov
Frank Laboratory for Neutron Physics
JOINT INSTITUTE FOR NUCLEAR RESEARCH

Project – Introductory course: MD-simulation research
(from atomic fragments to molecular compounds)
(Duration: 26th February 2024 to 14th April 2024)

Contents

1 Introduction

1.1 Molecular dynamics simulation

A computational approach elucidates atomic systems' equilibrium and dynamic properties. This method integrates Newton’s laws of motion to generate configurations and determine the time evolution of molecular systems, yielding microscopic details such as atomic positions and velocities. It bridges the gap between structure and function, complementing techniques like X-ray crystallography and NMR by providing additional insights. Numerical integration of Newton’s equations is utilised, with equilibrium systems yielding static properties like temperature and pressure through time-averaged measurements. Moreover, it allows for the investigation of dynamic properties such as heat transport and relaxation in systems far from equilibrium. Pioneering work on this methodology was conducted by A. Rahman (Phys. Rev. 136, A405, 1964), further expanded upon by L. Verlet (Phys. Rev. 159, 98, 1967), who introduced the Verlet algorithm and optimised calculations through the implementation of a neighbour list.

Molecular Mechanics involves using simple equations to express intra- and intermolecular forces, bypassing the use of wave functions and electron density to describe molecules. These forces, collectively termed force fields, include bond stretching, bending, torsions, electrostatic interactions, van der Waals forces, and hydrogen bonding. Harmonic oscillator formulas describe bond stretching and bond bending. Lennard-Jones equations describe van der Waals forces and hydrogen bonding, cosine functions describe torsions, and Coulomb's formula describes electrostatic interactions.

The formula represents the total energy of all forces

$$\boldsymbol{E = \sum_{bonds} \frac{kl}{2}(l - l_o)^2 + \sum_{angles} \frac{k\theta}{2}(\theta - \theta_o)^2 + \sum_{torsions} \frac{V_n}{2}(1 + cos\theta) +}$$

$$\boldsymbol{\sum_{Van\ Der\ Waals\ forces} 4 \in \left[\left(\frac{r_0}{r}\right)^{12} - \left(\frac{r_0}{r}\right)^{6}\right] + \sum_{electrostatic} \frac{q_1 q_2}{4\pi\varepsilon_o r}}$$

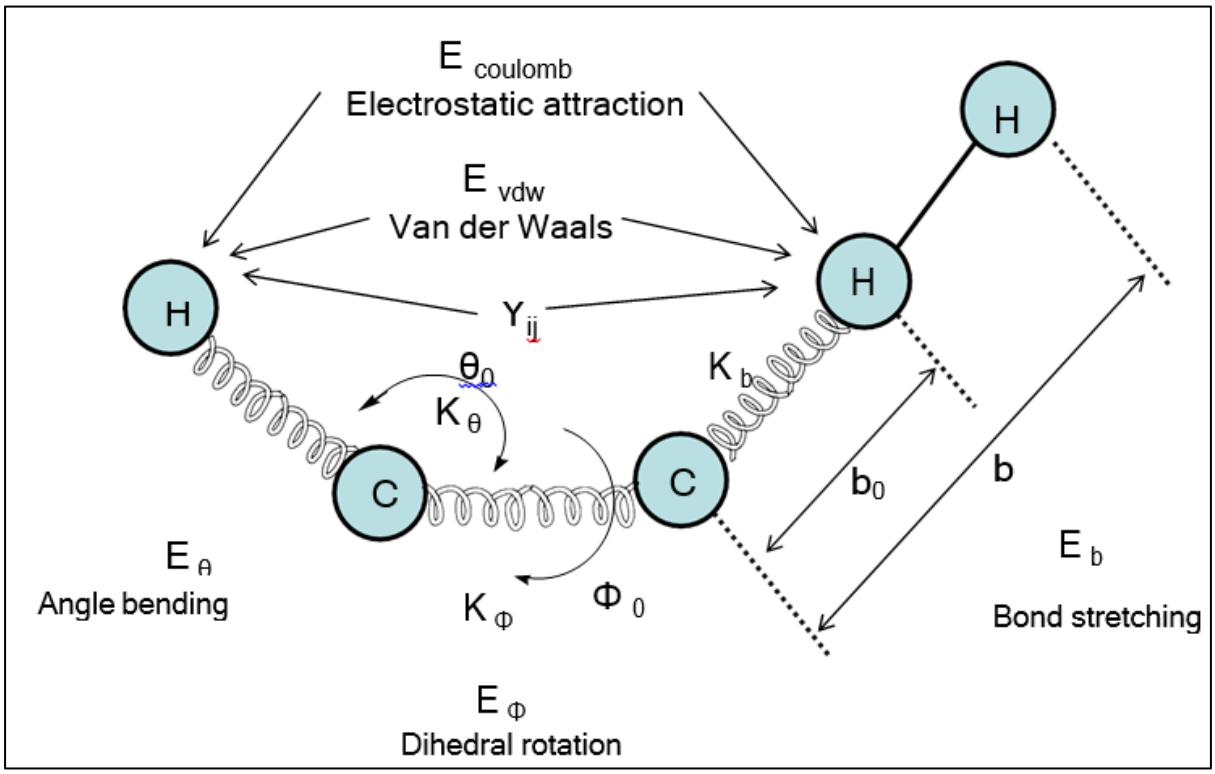

Fig1. Graphical representation of the bonded and non-bonded interaction and the corresponding energy terms.

Constants kl, and kθ are force constants, Vn is the torsional barrier, ℓℓ and θ are bond lengths and angles, r is the distance between two points, q1 and q2 are charges on two particles, ε is the well depth, and ε is the dielectric constant. Molecular Mechanics is commonly applied to calculate large molecules like polymers, proteins, and biochemical molecules. However, it may lack the ability to analyse the detailed chemical properties of these large molecules.

1.2 Historical Background

Introduced in the 1950s by Alder and Wainwright, molecular dynamics simulation was initially devised to investigate the interactions of hard spheres, representing atoms that interact solely through perfect collisions. Rahman pioneered the first simulation in 1964, focusing on liquid Argon. Subsequently, McCammon conducted the inaugural protein simulation in 1977, examining the bovine pancreatic trypsin inhibitor (BPTI). Today, the scope of molecular dynamics simulations has broadened significantly, encompassing solvated proteins, protein-DNA complexes, lipid systems, and more.

1.3 Statistical Mechanics

Molecular dynamics simulation allows us to investigate the macroscopic properties of a system by conducting microscopic simulations. The link between these microscopic simulations and macroscopic properties is established through statistical mechanics, which examines macroscopic systems from a molecular perspective. The distribution of the system within the ensemble adheres to the Boltzmann distribution.

1.4 Why Not Quantum Mechanics?

Modelling th motion of a complex molecule by solving the wave functions of the various subatomic particles would be accurate

$$\frac{-h^2}{2m}\nabla^2\psi + U(x,y,z)\Psi(x,y,z) = E\Psi(x,y,z)$$

But it would also be tough to program and use more computing power than anyone. Instead of Quantum mechanics, classical Newtonian mechanics is often utilised to model systems, offering a simplified yet less accurate representation of reality.

Numerical values derived from Quantum mechanics are incorporated into classical equations to address this limitation.

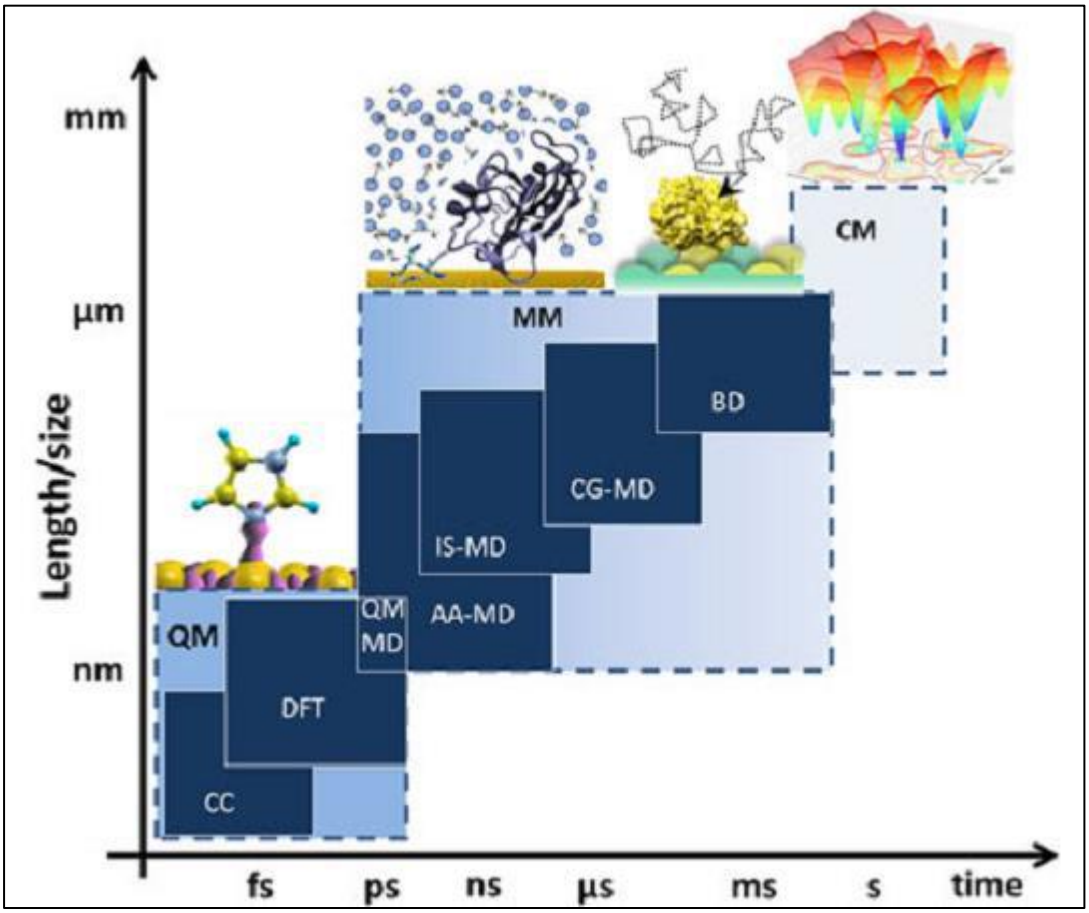

Fig 2. Typical time and length scales of different simulation techniques

In atomistic simulations, the primary objective is to accurately model, analyse, and comprehend the motion of individual atoms within a material. By understanding the collective behaviour of these atoms, insights into phenomena such as deformation, phase changes, and other transformations at the macroscopic level can be gained, establishing connections between atomic-scale dynamics and macro-scale behaviours. Extracting information from atomistic dynamics poses challenges due to factors such as vibration, spatial relocation, and connectivity, among others.

One significant advantage of Molecular Dynamics (MD) simulations is their capability to provide insights into the dynamical properties of systems. These include transport coefficients, time-dependent responses to perturbations, rheological properties, and spectral characteristics. MD simulations serve as a conduit between microscopic length and time scales and the macroscopic realm of laboratory experiments. By offering estimations of molecular interactions, MD simulations can yield precise predictions of bulk properties, with the accuracy being adjustable based on computational resources. Additionally, MD simulations can unveil hidden details behind bulk measurements, exemplified by the correlation between diffusion coefficients and velocity autocorrelation functions.

MD simulations also serve as a bridge between theory and experiment. They enable the testing of theoretical models through simulations and comparison with experimental findings, thereby facilitating the validation and refinement of models. Moreover, MD simulations allow for exploring phenomena that may be challenging or impossible to investigate in the laboratory, such as extreme temperature or pressure conditions.

While ab initio molecular dynamics aims to minimize fitting and guesswork by leveraging fundamental principles, for broader investigations or theory validation, a realistic molecular model may not always be necessary. A model encompassing essential physics can suffice for discriminating between viable and inadequate theories.

Molecular Dynamics offers a unique lens to observe the behaviour and roles of biomolecules within biological systems, serving as a crucial method for understanding their functions.

2. The basic equations and potentials

MD generates the dynamical trajectories of a system of N particles by integrating Newton's equations of motion with suitable initial and boundary conditions and proper interatomic potentials while satisfying thermodynamical (macroscopic) constraints.

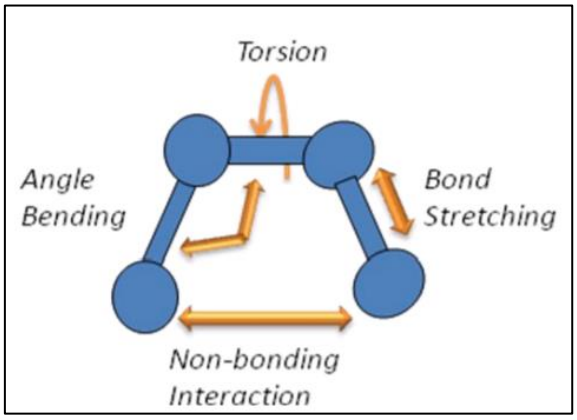

Fig.3 *Chemical bonds (bond stretching, angle bending, torsion) and non-bonding interaction*

Total energy of system $E = K + U$

$$\frac{1}{2}\sum_{j=1}^{N} mv_i$$

$$U = U(rj)$$

2.1 Equations of motion & potential

The molecular dynamics simulation is based on Newton's law of motion:

$$F_i = m_i a_i$$

For each atom in a system constituted by N atoms. Here, mi is the atom mass, its $a_i = \frac{d^r r_i}{dt^2}$acceleration, and Fi the force acting upon it due to the interactions with other atoms $r_i = const.$ A molecule initially at rest will remain at rest, and a molecule moving with a specified velocity will continue to move with that velocity until a force act on it. This

is Newton's first Law. Consider an isolated system that contains two spherical molecules,1 and 2. Hence, the total force is zero.

$$F_{total} = 0$$

Therefore, any force exerted by molecule 1 on molecule 2 must be balanced by a force exerted by 2 on 1.

$$F_{total} = F_1 + F_2 = 0$$

Hence,

$$F_1 = -F_2$$

This is Newton's third law. It's possible to determine the acceleration of each atom by knowing the force acting on each atom. Integration of the equation yields a trajectory that describes each atom's positions, velocities and acceleration as they vary with time. Once the positions and velocities of each atom are known, the system's state can be predicted at any time. The force can be written as the gradient of the potential energy:

$$Fi = -\nabla_i V$$

Combine two equations to get:

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

A trajectory is obtained by solving this differential equation.

2.2 Potential Energy

A single atom will be affected by the potential energy functions of every atom in the system: bonded neighbours and non-bonded Atoms (either other atoms in the same molecule or atoms from different molecules).

$$V(r) = E_{bonded} - E_{non-bonded}$$

2.2.1 Non-Bonded Atoms

There are two potential functions we need to be concerned about between non-bonded atoms: Van Der Waals potential and electrostatic potential.

$$E_{non-bonded} = E_{van\ der\ waals} + E_{electrostatic}$$

2.2.2 The Van Der Waals Potential

Atoms with no net electrostatic charge will still tend to attract each other at short distances as long as they don't get too close. Once the atoms are close enough to have overlapping electron clouds, they will repel each other with astounding force. One of the most widely used functions for the van der Waals potential is the Lennard-Jones. It is a compromise between accuracy and computability

$$E_{lennard\ jones} = \sum_{non\ bonded\ pairs} \left[\left(\frac{r_0}{r} \right)^{\mathbf{12}} - \left(\frac{r_0}{r} \right)^{\mathbf{6}} \right]$$

The Constants r_0 depend on the atom types and are derived from experimental data

2.2.3 The Electrostatic Potential

Opposite Charges Attract Like Charges Repel The force of the attraction is inversely proportional to the square of the distance.

$$E_{electrostatic} = \sum_{non\ bonded\ pairs} \left[\frac{\boldsymbol{q_i q_k}}{\boldsymbol{D r_{ik}}} \right]$$

2.2.4 Coulomb's Law

$$F = \frac{q_1 q_2}{4\pi\varepsilon_o r}$$

Coulomb interaction decays slowly with distance, considered prolonged range interactions. r represents the distance between two atoms having charges and ε represents the dielectric constant, a number relating the ability of a material to carry alternating current to the ability of a vacuum to carry alternating current.

2.2.5 The Non-Bonded Potential

Combination of the LJ and Electrostatic Potentials

$$E_{non-bonded} = E_{van\ der\ waals} + E_{electrostatic}$$

2.2.6 Bonded Atoms

There are three types of interaction between bonded atoms: stretching along, the bond bending between bonds and rotating around bonds

$$E_{bonded} = E_{bond-stretch} + E_{angle-bend} + E_{rotation-angle-bond}$$

2.2.7 Bond Length Potential

Both the spring constant and the ideal bond length are dependent on the atoms involved

$$E_{bonded} = \sum_{1.2 pairs} k_b (r - b_o)^2$$

2.2.8 Bond Angle potential

Describe the deviation from an ideal bond angle geometry

$$E_{bond-bend} = \sum_{angles} k_\theta (\theta - \theta_o)^2$$

k_θ represents angle bending constant, $\theta_0$represent the deviation from the ideal bond angle

2.2.9 Torsion dihedral potential

The motion associated is rotation, described by a dihedral angle around the middle bond. The potential is assumed to be periodic and expressed as a cosine function

$$E_{rotation-angle-bond} = \sum_{1.4} k_\emptyset (1 - \cos n\emptyset)^2$$

$k_\emptyset$ Represent rotation constant, n represent the periodicity of the rotational barrier and $\emptyset$ the dihedral angle.

2.2.10 Hydrogen bonding potential

$$\boldsymbol{U_{HB}} = \sum_{\boldsymbol{i,j}} \left[\frac{\boldsymbol{A}}{\boldsymbol{r_{i,j}^{12}}} - \frac{\boldsymbol{B}}{\boldsymbol{r_{i,j}^{6}}} \right]$$

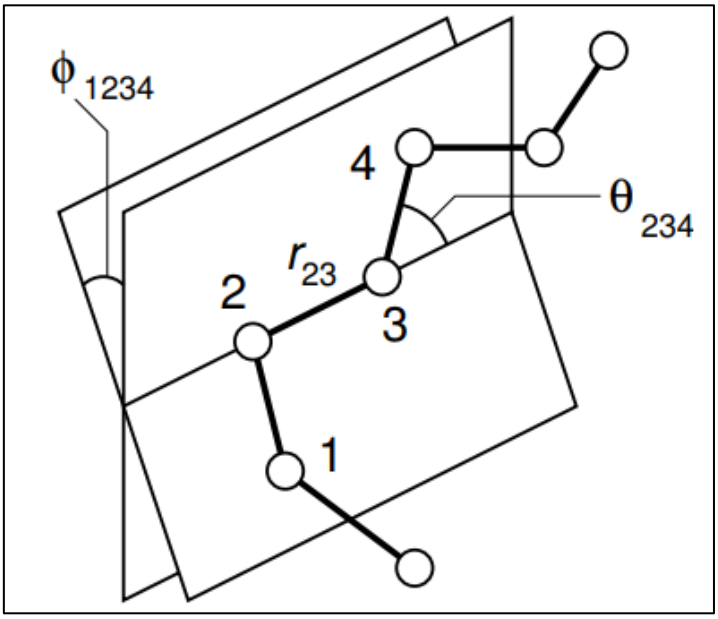

Fig4: Hydrogen Bonding Potential

3. Lennard Jones potential

The interactions at the most superficial level, occur between pairs of atoms and are responsible for providing the two principal features of an interatomic force. The first

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

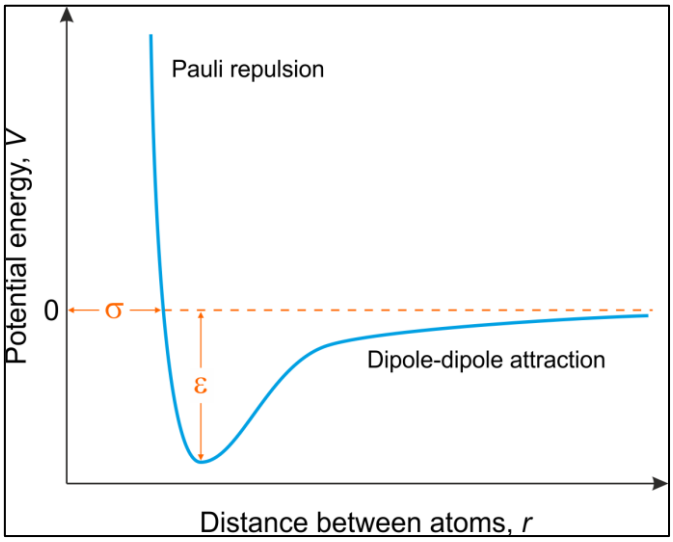

Fig.5. Graph of the Lennard- Jones potential function: Intermolecular potential energy as a function of the distance of a pair of particles.

4. Methods

4.1 Software:

i. **MD engine**-There are several software available for performing the molecular dynamic simulation of bio-molecules like GROMACS, NAMD, CHARMM, AMBER, Open Babel, VMD, UCSF Chimera, etc. We can select a software of our choice and perform the task, but always remember that different software uses different Force Fields.

ii. **Preparation and analysis-** Most MD engines, including those listed above, provide simulation preparation and analysis programs. Other options for bespoke analysis include the MD Analysis package. Basic scripting (e.g., bash) or programming skills (e.g., python) are also helpful for manipulating file and data formats and post-processing data.

iii. **Visualization:** Additional software is required to view simulation coordinate trajectories. VMD is free and compatible with the trajectory file formats produced by most MD engines. PyMOL is another standard option.

iv. **Graphical software**: Common options for plotting analysis data include gnuplot, xmgrace, qtgrace, R, and Excel.

Table 1: Application of different force fields

	Target application
DL_POLY	The multi-purpose package is used for MD modelling a wide variety of molecular systems – from simple atomic fragments to ionic structures, polymers and biochemical macromolecules.
AMBER	Simulation of peptide, protein, nucleic acid, and small organic molecules to facilitate simulations of drugs and small molecule ligands in conjunction with biomolecules, carbohydrates, lipids
CHARMM	Simulation of peptides, proteins, prosthetic groups, small molecule ligands, nucleic acids, lipids, and carbohydrates occurs in solution, crystals, and membrane environments. CHARMM also finds broad applications for inorganic materials, including applications in materials design.
NAMD	This package is designed for high-performance simulation of large biomolecular systems.

4.2 Coordinates:

i. Initial (Cartesian) coordinates for proteins can often be obtained from the Protein Data Bank (PDB) in PDB file format. These should be checked for missing or incomplete

residues, poorly resolved electron density, mutations introduced to aid crystallisation, etc. Missing residues can be built using software such as Modeller or DeepView.

ii. Software such as VMD can generate chemically plausible coordinates for biopolymers. Modeller can build structures based on homology; myriad homology modeling web servers exist. Coordinates for small molecules may be available from the Cambridge Structural Database or generated using software such as Avogadro.

iii. Coordinates not derived from experimental data should ideally be optimised using quantum chemical calculations or extensive MD simulations.

iv. Pre-equilibrated boxes of solvent coordinates are available for common solvents such as water

4.3 Parameters:

i. Force field files list parameter values and specify their assignment to atom types and bonded interactions between particular combinations of atom types. Each MD software package requires these files to be in a specific format and typically provides these files for one or more commonly used force fields.

ii. Each force field may use different terms, and even if these are the same, parameter values are not interchangeable between force fields as they are parameterised using different strategies. Thus, all molecules included in the simulation must be described using the same or a compatible force field

iii. Molecule or fragment "building blocks" provide parameters for a particular molecule; for polymers such as proteins, preparation programs provided with the MD software join amino acid building blocks together to generate the molecular "topology".

iv. If a building block is unavailable, the molecule must be parameterised. This may be done manually or using automated software such as the ATB (GROMOS),

CherryPicker, CGenFF (CHARMM) antechamber (AMBER/GAFF), or LigParGen (OPLS).

v. Similarly, each force field is compatible with particular solvent models (Table 1).

4.4 Simulation Settings:

1. MD run files contain the settings (e.g., algorithm choices) and associated parameter values for carrying out an energy minimisation or MD simulation. The file format is specific to each MD package.

2. Key settings for energy minimisation:

- **a)** Specification of energy minimisation method. Table 1 Common biomolecular force fields, their corresponding general force fields, and the most commonly used compatible water models Biomolecular force field General-purpose force field Water model CHARMM; The CHARMM force field was originally parameterised using a modified version of the TIP3P model, cTIP3P, which adds Lennard-Jones interactions between the hydrogen and oxygen atoms, but the latest version, CHARMM36, was optimised.
- **b)** Size and maximum number of minimisation steps
- **c)** Criteria for convergence.
- **d)** Frequency and properties to write to file. 3. Key settings for initialisation and heating (typically run in the NVT).
- **e)** The use of constraints, constraint algorithms, and associated parameters.
- **f)** Distances within which to explicitly calculate nonbonded interactions.
- **g)** Method for treatment of long-range electrostatic interactions and associated parameters

4.5 DL_POLY

The DL_POLY computer code, which is used to investigate a range of chemical and biochemical systems, was developed at Daresbury Laboratories, England, by a molecular modelling group led by Bill Smith. There are three input (CONFIG, CONTROL, FIELD) and three output (OUTPUT, REVCON, HISTORY) files in the DL_POLY computer code.

The initial molecular structure, CONFIG, contains 3-dimensional coordinates (x, y, z) of all atoms, set the boundary conditions, as well as the initial values of the velocities (V_x, V_y, V_z) and interatomic forces (f_x, f_y, accordance f_z). In with the CONFIG file, it is necessary to compile a FIELD, which contains information about the structure of atoms and molecules, their masses and charges, parameters, and types of interaction potentials. CONFIG and FIELD must be consistent in structure. The CONTROL file contains data on the simulation parameters (temperature, pressure, the step of integration of the equations of motion and calculation time, thermodynamic parameters, and simulation methods, etc.); each line indicates a method or algorithm for MD modelling.

DL_POLY code can be used to study the dynamic and structural properties of various molecular systems - atomic and ionic structures, polymer chains and biological macromolecules The advantage of DL_POLY software is possible to do simulations on many kinds of materials such as:

a. Simple atomic systems and mixtures e.g. Ne, Ar, Kr, etc.
b. Simple unporisable point ions e.g. NaCl, KCl, etc.
c. Polarisable point ions and molecules e.g. MgO, H_2O etc.D
d. Simple rigid molecules e.g. CCl_4, SF_6, Benzene, etc.

e. Rigid molecular ions with point charges e.g. KNO_3, $(NH_4)2SO_4$, etc.

f. Polymers with rigid bonds e.g. C_nH_{2n+2}

g. Polymers with rigid bonds and point charges e.g. proteins

h. Macromolecules and biological systems

5. Future Scope

As students of science, we understand that the macroscopic properties of elements are deeply influenced by the underlying microscopic properties and interactions among atoms and molecules. Molecular Dynamics Simulation has opened up numerous opportunities for research enthusiasts in the biological sciences, spanning from protein folding studies to exploring the physics and chemistry underlying biomolecular interactions, molecular docking for drug design, and beyond. The efficacy of MD Simulation hinges on the reliability of the model, force-field calculations, and thermodynamic property calculations, as well as the capability of software to replicate processes with high fidelity. Despite the advancement of numerous computer simulation techniques, there remains room for improvement, as simulations inherently lack complete accuracy in their results, necessitating the development of better and more precise techniques. Enhanced simulation outcomes also lay the groundwork for future studies on ion-exchange phenomena. Moreover, the quest continues for the development of more robust algorithms, with fewer computational steps and reduced computational intensity. There is also a growing need for lightweight and freely available software to broaden accessibility and accelerate progress in computational research.

I am a third-year undergraduate student at the Institute of Chemical Technology, Mumbai, Indian Oil Campus, Bhubaneswar, Odisha. My research focuses on exploring the simulation

and production of organic dyes used in dye-sensitized solar cells, aiming to investigate their production feasibility and wide-ranging applications.

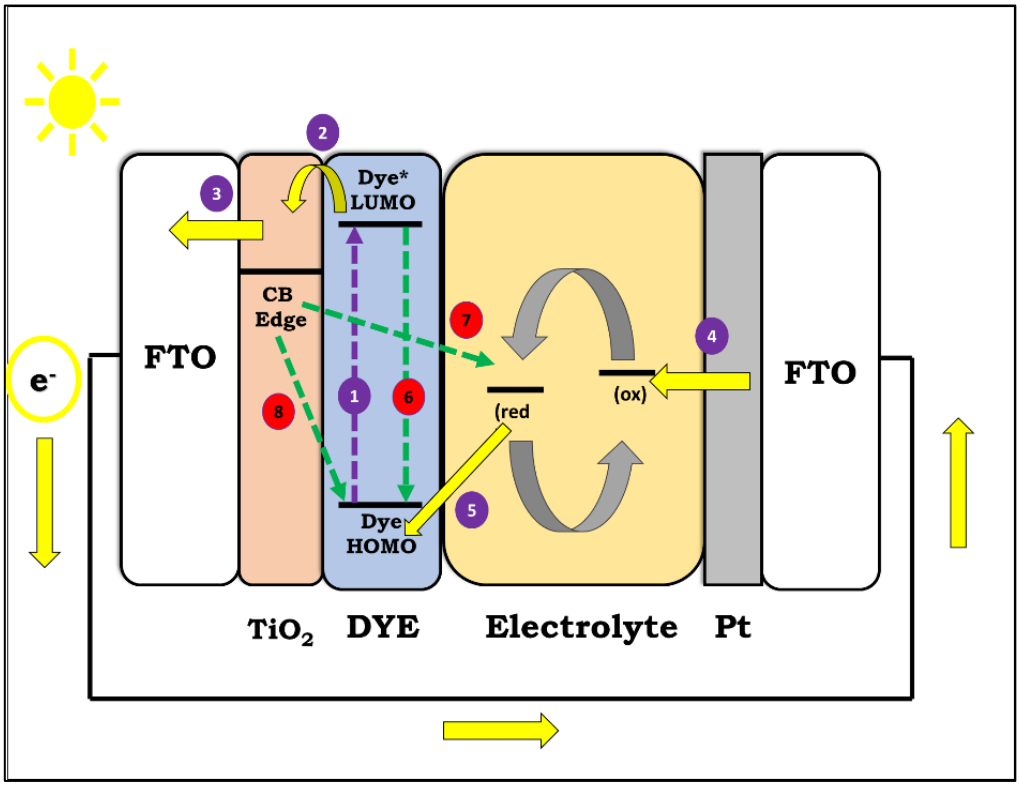

Figure 6: Working Principle of Dye-Sensitized Solar Cells (DSSCs)

Organic dyes are essential components in dye-sensitized solar cells (DSSCs), playing a crucial role in absorbing sunlight and converting it into electrical energy. Understanding their properties and behavior is vital for optimising DSSC performance and exploring new applications.

Molecular dynamics (MD) simulation offers a powerful tool for exploring various aspects of organic dyes. It can help investigate these dyes' size, shape, surface chemistry, electronic band structure, and optical properties. Additionally, MD simulation enables the study of the effects of surface functionalisation, doping, or modifications on the properties and behaviour of the dyes. Furthermore, it allows the

simulation of interactions with different functional groups or substrates, providing insights into tuning their electronic and chemical properties.

By comprehensively understanding the structure-property relationships of organic dyes through MD simulation, we can tailor them for specific applications in electronics, photonics, and optoelectronics. This knowledge is crucial for developing efficient and versatile dye-sensitized solar cells and other related technologies.

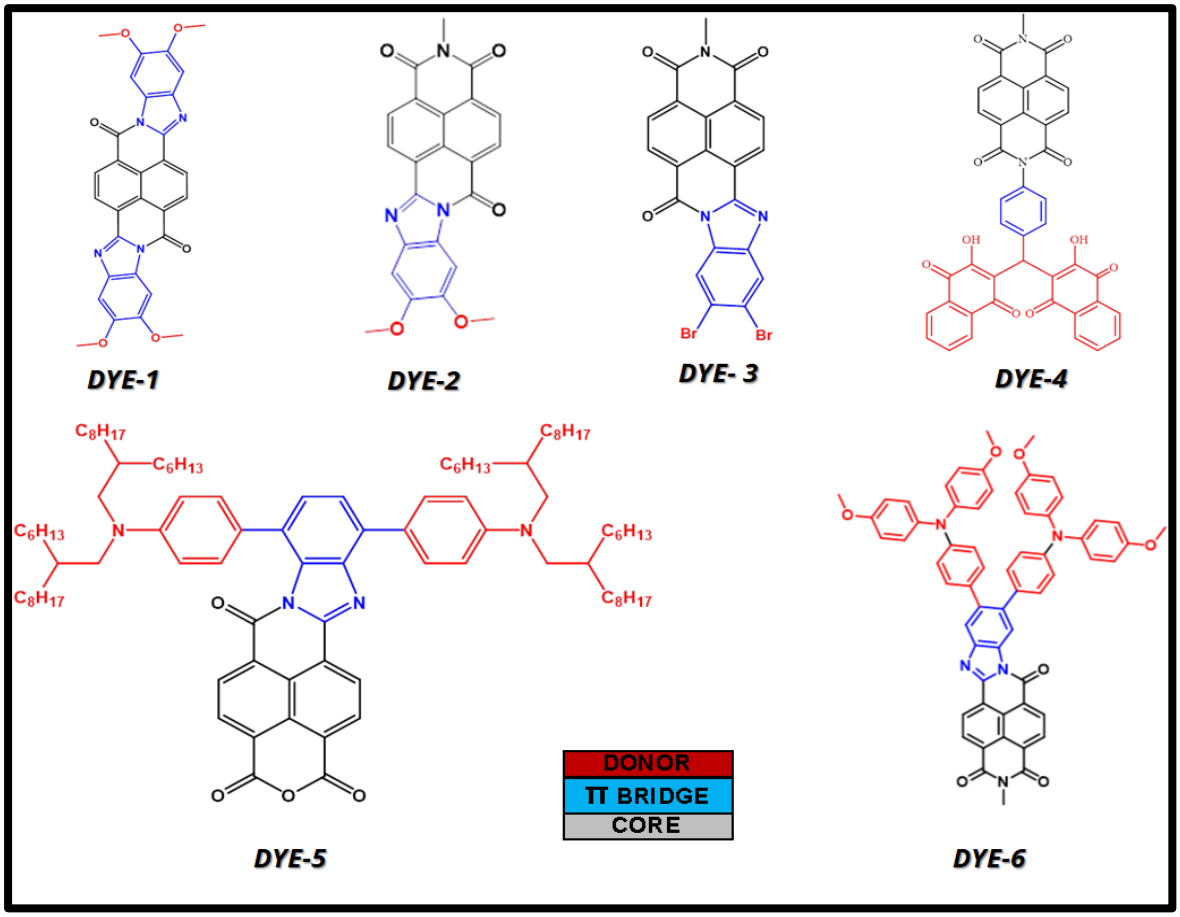

Figure 7: Structure of Naphthalene tetracarboxylic dianhydride (NTDA)

Core-Based Dyes

I am confident that the experience gained from this project in MD modelling will significantly enhance my scientific capabilities in this research area.

6. Conclusion

MD simulations have already more than 40 years of history. However, it was not until recent years that MD achieved time scales compatible with biological processes. At present, when routine simulations are approaching the microsecond scale, conformational changes or ligand binding can be effectively simulated. The improvement of the computational equipment, especially the use of GPUs, and the improvements made in the optimisation of MD algorithms, including coarse-grained ones, allow us to move from the analysis of single structures, the basis of the molecular modelling as we know it, to the analysis of conformational ensembles. Conformational ensembles are a much better representation of real macromolecules, as they account for flexibility and dynamic properties (including all thermodynamic information) and ease the match with experimental results.

Although the shift in concept is clear, and the technology is coming along, there is still a long way until biomolecular simulations, the generation of conformational ensembles, would become a routine. Tools exist that make the setup of a macromolecular system much easier, and even allow the nonexperts to enter the simulation world. However, lack of representation standards, much less optimised analysis tools, and even the difficulties in simply storing and transmitting the huge amount of trajectory data generated are still issues that remain to be solved. In any case, an MD is a valuable tool in helping one understand biology.

7. Acknowledgements

I want to thank my supervisor, Prof. Dr. Kholmirzo Kholmurodov, for allowing me to be part of an exciting project and teaching of introduction molecular dynamic simulation. All appreciation for Professor Kholmirzo Kholmurodov, who has introduced the course with perfectly organised and regular weekly lectures and is very caring about what we understand. Additionally, I would like to thank the INTEREST Team for the opportunity to do science online. Also, I would like to acknowledge the Joint Institute for Nuclear Research, Dubna, Russia.

8. References:

Books

- Tamar Schlick Molecular Modeling and Simulation: An Interdisciplinary
- Guide 2002
- Alan Hinchliffe Molecular Modelling for Beginners 2003
- D. C. Rapaport The Art of Molecular Dynamics Simulation. 2004
- Daan Frenkel, B. Smit Understanding Molecular Simulation 2001
- M. P. Allen and D. J. Tildesley, Computer Simulation of Liquids (Oxford U.
- Press, Oxford, 1987).
- D. Frenkel and B. Smit, Understanding Molecular Simulation, 2nd Ed.
- (Academic Press, New York, 2001).
- © 2007 Markus J. Buehler, CEE/MIT
- Träskelin et al., J. Appl. Phys. 93 (2002) 1826
- W. Smith, T.R. Forester and I.T. Todorov, The DL POLY Classic User
- Manual, Version 1.0, December 2010.
- C. Educational, “Kholmirzo Kholmurodov MD-SIMULATION IN
- CHEMICAL RESEARCH : Educational level : MS,” 2011.
- A. Rahman, “Correlations in the Motion of Atoms in Liquid Argon,”
- Phys. Rev., vol. 136, no. 2A, pp. A405–A411, Oct. 1964.
- L. Verlet, “Computer ‘Experiments’ on Classical Fluids. I.
- Thermodynamical Properties of Lennard-Jones Molecules,” Phys.
- Rev., vol. 159, no. 1, pp. 98–103, Jul. 1967.

Papers:

- Musa Ozboyaci, Daria B. Kokh, Stefano Corni, Rebecca C. Wade, Modeling and simulation of protein – surface interactions: achievements and challenges // Quarterly Reviews of Biophysics (2016), 49, e4, page 1 of 45.
- Khusenov M.A, Dushanov E.B, Kholmurodov Kh.T, Zaki M.M, Sweilam N.H., On Correlation Effect of the Van-der-Waals and Intramolecular Forces for the Nucleotide Chain - Metallic Nanoparticles – Carbon Nanotube Binding // The Open Biochemistry Journal, 2016, 10, 1-17.
- M. A. Khusenov, E. B. Dushanov and Kh. T. Kholmurodov, Molecular Dynamics Simulations of the Nucleotides and Metallic Nanoparticles Interaction on a Carbon anotube Matrix // Materials Transactions, Vol. 56, No. 9 (2015), pp. 1390 to 1393.
- Kholmurodov, K.T., *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. p. 1-231.

Websites

- Molecular Dynamics Tutorial at EMBnet
- http://www.ch.embnet.org/MD_tutorial/index.html
- Theoretical and Computational Biophysics Group at UIUC (home of VMD
- and NAMD)
- http://www.ks.uiuc.edu/
- http://www.weizmann.ac.il//sb/Pages/Levy/group meet/Molecular dynamics
- simulation.pdf7
- . https://www.dovepress.com/molecular-dynamics-s