International Remote Students Training

at

Joint Institute for Nuclear Research

Frank Laboratory of Neutron Physics

Final Report on the INTEREST Programme

INTRODUCTORY COURSE: MD-SIMULATION RESEARCH (FROM ATOMIC FRAGMENTS TO MOLECULAR COMPOUND)

Submitted by: Aritra Roy Prodhan

Undergraduate Student, Department of Physics, Jadavpur University

Supervisor: Professor (Dr.) Kholmirzo Kholmurodov

Frank Laboratory of Neutron Physics, Joint Institute for Nuclear Research

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to **Prof. Kholmirzo Kholmurodov** (Frank Laboratory of Neutron Physics, JINR) for his invaluable guidance, supervision, and for providing the comprehensive theoretical framework through his course, *"MD-Simulation Research: From Atomic Fragments to Molecular Compound."* His expertise and the structured insights provided in the source material were instrumental in the successful completion of this project.

I am also thankful to the **Joint Institute for Nuclear Research (JINR)** and the **INTEREST Program** organizers for providing this unique platform for international scientific collaboration and academic growth. The opportunity to work with advanced computational tools like **GROMACS** and to explore the intersection of physics, chemistry, and biology has been a transformative experience.

CONTENTS

Abstract

Here, in this report we have discussed about the methods of the Molecular Dynamics simulation, starting from its very theoretical beginning to its present computer simulation using both the classical physics and the quantum chemistry approach. A special focus has been given in the discussion of the various potential acting on the molecules, mainly the Lennard Jones potential and the Electrostatic potential. Also a discussion about various algorithms has been done in this report which are essential for the modelling of the conditions for the molecular dynamic simulation. The use of various software and packages has been also discussed here, which are used for molecular dynamic simulation, for its different application in various field of sciences.

For the project, the biophysical modelling of the trisomy and the determination of the extra chromosome using DNA pyramid and nanoparticles has been taken up. Using this method, the extra chromosome for the case of trisomy (which gives rise to Down Syndrome) can be found out, later resulting in its removal using the method of artificial trisomy rescue method.

1.Introduction: *What is Molecular Dynamics Simulation?, The steps involved in it and the fundamental equation used in it*

Molecular Dynamics Simulation is a deterministic computational approach which is used to study the behavior of classical many particle system by simulating the motion of individual atoms or molecules. This is done by numerically intergrating the Newton's equations of motion, as a result of this we get the trajectories that describe the velocity and position of the system as a function of time.

Now, looking at the various steps required for the molecular dynamics simulation:

i)Design of the initial system configuration.

ii)Description of the atom-atom interaction

iii)Control of experimental parameters and conditions

iv)Finding the equilibrium state of the system

v)Variation of the simulation parameters during the calculation time

vi)History of the system's dynamics

The fundamental equation used here is the Newton's equation of motion, i.e. Newton's 2nd law:

$$m_i \frac{d^2 \boldsymbol{r}_i(t)}{dt^2} = \boldsymbol{F}_i(\boldsymbol{r}), \quad i = 1,2,\dots,n$$

$$\{\boldsymbol{r}_i, m_i, \boldsymbol{F}_i\}$$

$$\boldsymbol{r} = \{r_1, r_2, \dots, r_n\};\ U(\boldsymbol{r})$$

$$\boldsymbol{F}_i(\boldsymbol{r}) = -\frac{\partial U(\boldsymbol{r})}{\partial \boldsymbol{r}_i}$$

Here m is the mass of the particle, r is its position and U is the total potential energy of the system.

This relation states that the total Force applied on a system is equal to the negative of the gradient of the system's potential.

A point to remember is that, here the potential U is combination of all the potentials that are acting on the system.

2.Potentials: *Types of potentials acting on the system and their mathematical equations*

There are mainly 2 types of potentials acting on a system, they are:

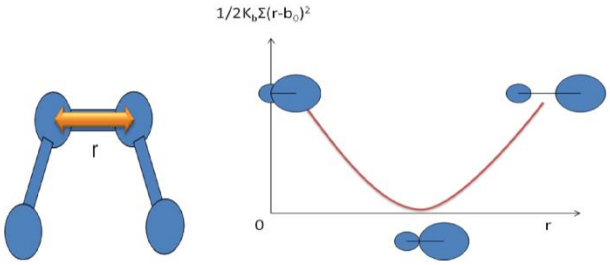
***Bonded Interactions*:**

These describe the connectivity within a molecule.

- **Bond Stretching(a.k.a. Valence length potentual):** Modelled as a harmonic oscillator

Valence length potential,

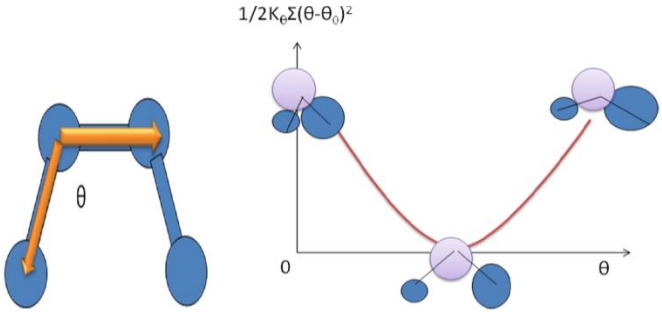
$$U_b = \frac{1}{2}\sum_b K_b\,(r - b_0)^2$$



- **Valence Angle:** Describes the vibration between 3 atoms

Valence angle potential,

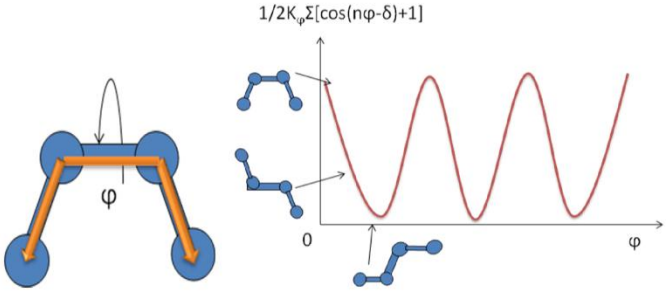
$$U_\theta = \frac{1}{2}\sum_\theta K_\theta\,(\theta - \theta_0)^2$$



- **Torsion Dihedrals:** Accounts for the rotation around a central bond, often using cosine series.

Valence angle potential,

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



Non-Bonded Interactions:

The energy between atoms or molecules that are not directly linked by covalent bonds, ionic bonds, or metallic bonds.

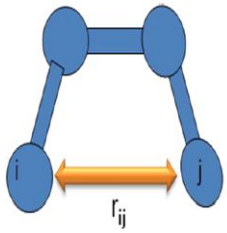
- **Lennard-Jones Potential:** Represents van der Waals forces (attraction and repulsion):
- **Electrostatic (Coulomb) Potential:** Calculated for charged particles. In periodic systems, the **Ewald Summation** method is used to handle long-range interactions efficiently.

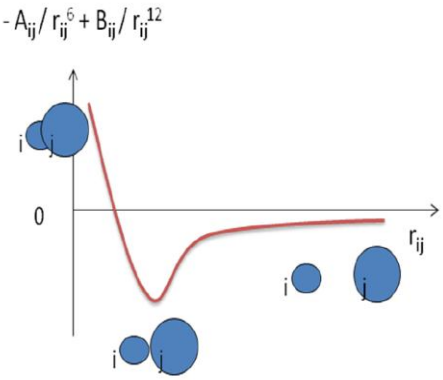
Van-der-Waals interaction potential (12-6 or Lennard-Jones (lj)):

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

Electrostatics potential,

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





3. Velocity generation:

In a Molecular Dynamics simulation, assigning initial velocities is the critical step that defines the kinetic energy of the system.

Methods used for velocity generation:

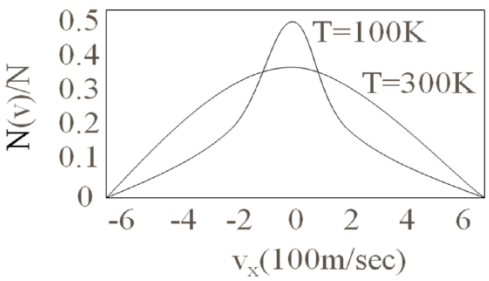
The Maxwell-Boltzmann Distribution:

According to the principles of statistical mechanics, the velocities of atoms in a system at thermal equilibrium are not uniform but follow a probability distribution.

$$f(v_x)dv_x = N(v_x)dv_x/N$$

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

$$< A(v_x) > = \int_{-\infty}^{+\infty} A(v_x)f(v_x)dv_x$$



Numerical Implementation and Random Sampling:

In the computational routines detailed in the source material (specifically the Fortran 90 implementations), the generation process involves a pseudo-random number generator. Because standard generators produce a uniform distribution, a **Box-Muller transformation** or similar algorithm is applied to convert these into a Gaussian distribution centered at zero.

4. Thermodynamic Regulation*: Thermostats and Barostats:*

Thermostats: *Maintaining Constant Temperature*

The temperature of a system is directly proportional to the average kinetic energy of its constituent particles. A thermostat acts as a "heat bath" that exchanges energy with the system to regulate thermal fluctuations.

The Berendsen Thermostat (Velocity Rescaling): This algorithm mimics a weak coupling to an external heat bath. It rescales the velocities of all particles at each time step to bring the instantaneous temperature toward the target temperature.While computationally efficient for equilibration, it does not strictly reproduce the correct canonical distribution.

$$\lambda = \left[1 + \frac{\Delta t}{\tau_T}\left(\frac{T_0}{T} - 1\right)\right]^{1/2}$$

The Nose-Hoover Thermostat: A more mathematically rigorous approach that treats the heat bath as an integral part of the system's Hamiltonian. It introduces a fictional dynamic variable (an extra degree of freedom) that acts as a friction coefficient.This method ensures that the system samples the true Canonical Ensemble.

$$\frac{d\mathbf{v}_i}{dt} = \frac{\mathbf{F}_i}{m_i} - \zeta \mathbf{v}_i$$

Barostats: *Maintaining Constant Pressure*

To maintain a constant pressure, the volume of the simulation box must be allowed to fluctuate. This is essential for studying phase transitions or systems where density changes significantly.

The Berendsen Barostat: Similar to the Berendsen thermostat, this method rescales the simulation box dimensions and the coordinates of all atoms by a factor It is governed by the system's isothermal compressibility and a pressure coupling time. If the internal pressure is higher than the target, the box expands if lower, the box contracts.

$$\mu = \left[1 - \beta \frac{\Delta t}{\tau_P}(P_0 - P)\right]^{1/3}$$

5.Softwares and packages used for Molecular Dynamics Simulation:

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.

GROMACS	This is mainly used for the simulation of various Biophysical modes like proteins, transportation across membranes etc.

6.Molecular Dynamics Simulation Project: *Trisomy to normal cell:*

This project presents the development of an Artificial Trisomy Rescue System (ATRS-1), a nanomechanical vehicle designed to identify and dock onto the extra chromosome in Trisomy 21. Using GROMACS, I have simulated coursed grain scaling of a tetrahedral DNA framework functionalized with Gold (Au) and Graphene Oxide (GO) nanoparticles, along with the 3 chromosomes. To solve structural failure during docking, a Complete Graph (K11) bond matrix was implemented. The simulation confirms that the vehicle can navigate via electrostatic steering and successfully anchor onto a methylated target site at physiological temperature (310K).

First Looking into the Biophysical modelling of the technique:

Identification of the Paternal and Maternal Chromosome in the 21st position of the cell of the Down Syndrome person, at molecular level:

In the nucleus of a cell, there consist 23 pairs of chromosomes, in these pairs one chromosome comes from the father, known as the paternal chromosome and the other one comes from the mother known as the maternal chromosome. In this pair of chromosomes, the DNA of both the chromosomes are same, but their epigenetic annotations are different.In short we can say that the paternal and maternal chromosomes have the same DNA, i.e. the same blueprint but they have different epigenetic annotations,i.e. they are the chemical tags attached to the DNA, they controls the functioning of the DNA without changing the information present in the DNA, i.e. the sequence of the DNA.

One such epigenetic annotations in the chromosomes is presenceof methylation. These are certain sequence in the 21st pair of the chromosomes 1 in a cell(irrespective of trisomy), which depends on the presence of methylation.

One such sequence is the DSCAM. Now looking from the point of view of the Trisomy case, when the methylation is high the DSCAM sequence

is silenced meaning the that there are more maternal chromosomes in the 21st position than the paternal chromosome,i.e. there is an extra copy of the maternal chromosome in the 21st position (as in the 21st position the DNA of the maternal chromosomes consist of methylation and the paternal don't). Also, when the methylation is less,that means there is less no. of maternal chromosome than the paternal, hence the is 2 copies of paternal chromosomes in the 21st position.

Using this logic we can now identify the extra chromosome present in the 21st position which is causing trisomy, and remove that using the methods as discussed in the 1st part of the paper,resulting in normalisation of the cell, i.e. from trisomic cell to normal cell.

Process for identification:

Now, for identifying the chromosome, we will be using 2 types of nanoparticles. The first type is the one, which can detect methylation and the second type is the one which detect absence of methylation. Gold nano particles detect the presence of methylation whereas Graphene oxide nano particles detect the absence of methylation.

So, now in our DNA pyramid(as discussed in the first part), we put only one anchor in it, and one gold nano particle and one graphene oxide nano particle, it the vertices of the pyramid, we can detect the extra chromosome. The gold nanoparticle will only get attached to the chromosome when the methylation is high, and it will attach only with one of the 2 chromosomes which has the methylation factor in it, as the DNA pyramid has only one anchoring spot. Also the graphene oxide nanoparticle will get attached to the chromosomes when the methylation is low, and it will attach only with the chromosome which doesn't has the methylation factor in it(will attach with only one of 2 such chromosomes, due to the presence of only one anchoring spot).

NOTE: In most of the cases of trisomy, it has been seen that, the extra chromosome is maternal, here also I have considered the same thing.

Now looking into the computational modelling of the technique:

A chromosome is made up of millions and millions of atoms, so it becomes nearly imposible for a normal computer to simulate such many no. of atoms, even it becomes very very very difficult for a super computer also, to run such simulation. So for overcoming this problem here I have taken the whole system as a 70atom course gain system, here for each chromosomes are made up of 20 atoms, and the pyramid along with the nanoparticles are made up of 11 atoms.

NOTE: Here atoms are considered as beads, for the course grained modelling.

System Modeling and Discretization:

The simulation is conducted in a **triclinic box** measuring (20*20*20)nm},to maintain the physical continuity of the environment, periodic boundary conditions are applied in all directions.

As discussed earlier I have utilized a coarse grained mapping where individual beads represent larger molecular assemblies:

- **DNA Chromosomes:** Modeled as semi-flexible chains of 20 beads each.
- **The ATRS-1 Vehicle(DNA Pyramid):** A tetrahedral assembly of 11 beads, with specific vertices functionalized as a gold nanoparticle (Au) and a graphene oxide (GO) nanoparticle.

Structural Topology: The Complete Graph Strategy:

To resolve the "shattering chassis" problem, i.e. breaking of the pyramid which takes place when the attractive docking forces exceeded the structural strength of the pyramid, we implemented a complete graph (K11) bond matrix.

- **Bonding Density:** Instead of a simple wireframe, every bead in the 11-bead pyramid is bonded to every other bead.
- **Mathematical Proof:** This results in 11(11-1)/2= 55 unique harmonic bonds.
- **Restorative Force:** We assigned an ultra-high spring constant. This ensures that even under the extreme acceleration of docking, the vehicle maintains

a rigid-body geometry, distributing the impact energy across all 55 bonds simultaneously.

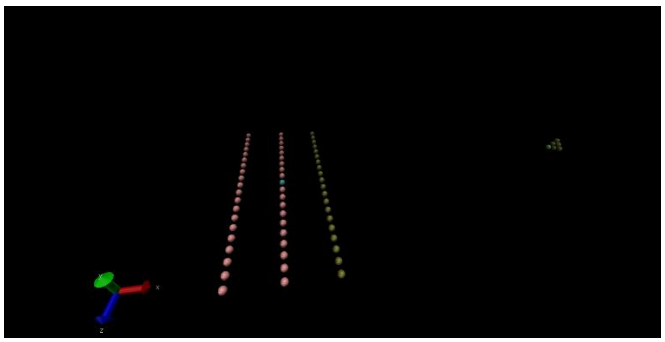
Force Field and Homing Mechanics:

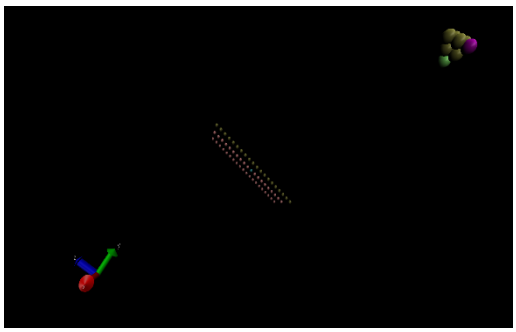
The sensing of the extra chromosome is governed by two primary non-bonded interaction types:

1. *Long-Range Electrostatic Steering:* To simulate the **aptamer-guided navigation**, we assigned a positive charge (+3e) to the gold nanoparticle and a negative charge (-3e) to the methylated target site on the extra chromosome. This creates a Coulombic attraction that pulls the vehicle from a distance of over 3 nm, overcoming random Brownian motion.
2. **Lennard-Jones Docking Potential:** Once the vehicle is within 0.5 nm of the target, a specialized Lennard-Jones (12-6) potential takes over. We increased the interaction strength (E, Epsilon) to 25 kJ/mol and reduced the collision radius to 0.25nm to simulate the "lock-and-key" fit of the aptamer sensing the methylation state.

Now, looking into the simulation:

Before reaching the docking spot:



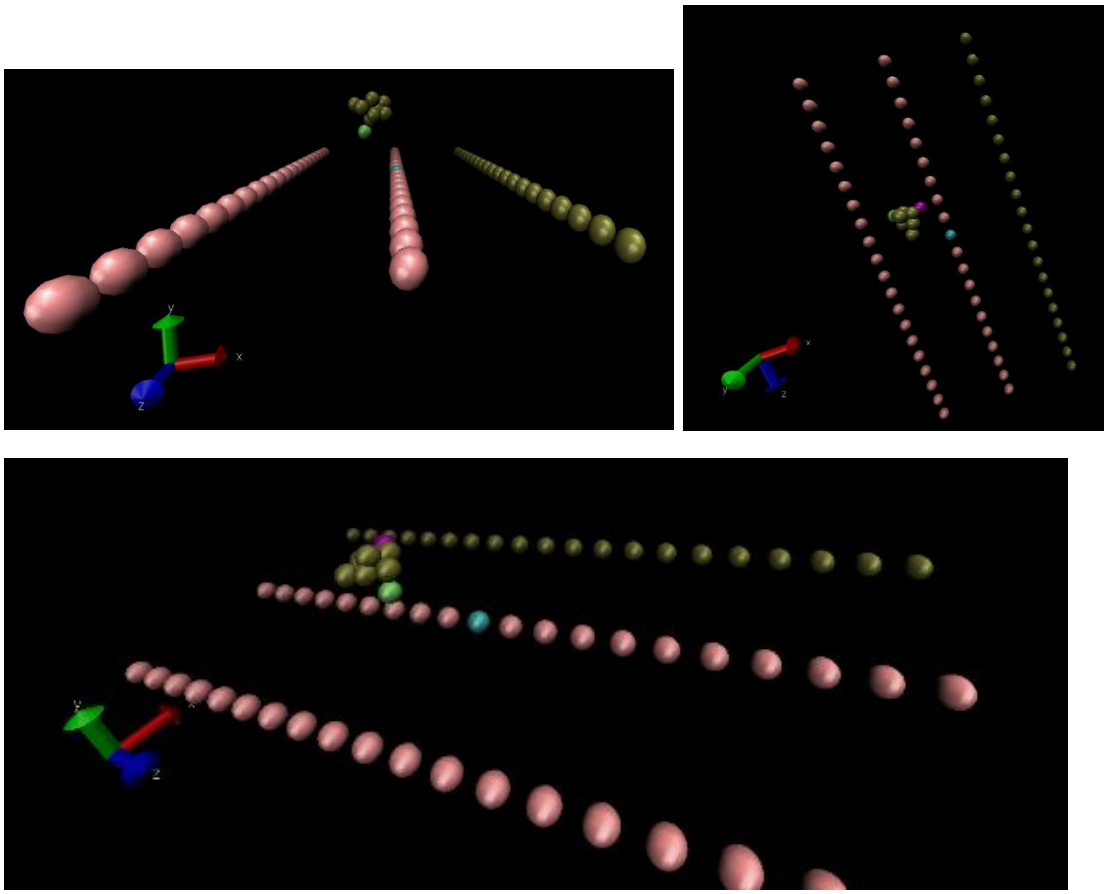


In this picture the pink beads represents the maternal chromosome and the sap green bead represent the paternal chromosome, from the picture it is clearly understood that in this case the extra chromosome is the maternal chromosome. Also the blue colour bead in the mid maternal chromosome is the position of the sequence where the docking of the gold particle (along with the whole pyramid) with the chromosome will take place.

Also at the right corner we can see the DNA pyramid, here the purple bead is the graphene oxide nanoparticle and the light green bead is the gold nanoparticle.

This picture is from the moment when the pyramid haven't yet reached the docking position.

After reaching the docking spot:



These are the pictures, when the pyramid reached the docking spots after travellinig through a distance, and docks with the extra chromosome.

7. Conclusion: *The showdown:*

Molecular Dynamics simulation is a very useful and powerful tool, for understanding all the sciences, it makes our work easier. Here I have used this tool, for simulating my idea on the biophysical modelling of converting a trisomic cell to a normal cell. Experiments related to biology requires a lot of processing and time, in both paper works domain and in the experimental domain too, molecular dynamics reduces this, and makes our work easier and faster, rather than conventional experimental techniques.

8. References:

I. Kholmirzo Kholmurodov; MD-SIMULATION IN CHEMICAL RESEARCH: FROM ATOMIC FRAGMENTS TO MOLECULAR COMPOUND

II. Musa Ozboyaci1, Daria B. Kokh1, Stefano Corni and Rebecca C. Wade; Modelling and simulation of protein surface interactions: achievements and challenges

III. IISER Pune, NPTEL; A course on Computational Physics

IV. Phapanin Charoenphol and Harry Bermudez ; Aptamer-Targeted DNA Nanostructures for Therapeutic Delivery

V. Ryotaro Hashizume , Sachiko Wakita, Hirofumi Sawada , Shin-ichiro Takebayashi , Yasuji Kitabatake , Yoshitaka Miyagawa , Yoshifumi S Hirokawa, Hiroshi Imai and Hiroki Kurahashi ; Trisomic rescue via allele-specific multiple chromosome cleavage using CRISPR-Cas9 in trisomy 21 cells