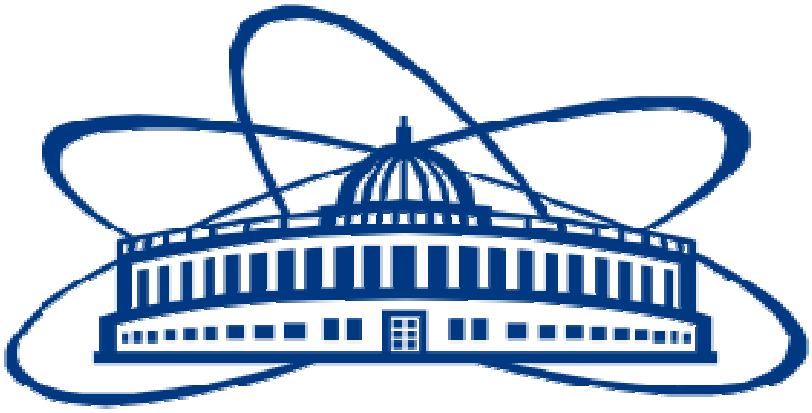


JOINT INSTITUTE OF NUCLEAR RESEARCH

DUBNA,

RUSSIAN FEDERATION

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

A tribute to the sacrifice of the mice in genetic research (*Novosibirsk' Akademgorodok, Russia***)**

Supervisor: Professor Kholmirzo Kholmurodov.

Student: Gerasimos Sapardanis, Aristotle University of Thessaloniki, Greece.

Project: Wave 6

14 February-25 March 2022.

Thessaloniki, March 2022

Contents

1 INTRODUCTION

1.1 What is Molecular Dynamics Simulation?

Molecular dynamics (MD) is a simulation methodology in which the movement of system particles is calculated in a certain period of time and the system evolution is investigated. It can be used in the studies of equilibrium and dynamic properties of a system. Usually, Newton's equations of motion are used to capture the trajectories of particles in the system and the potential energy is computed based on the force fields.

The simulation procedure is usually constructed as follows:

- Initialize the system under the criterion of zero total momentum.
- Compute the forces for each particle.
- Integrate Newton's equation of motion.
- Repeat steps 2 and 3 for a desired length of time.

The impact of molecular dynamics (MD) simulations in molecular biology and drug discovery has expanded dramatically in recent years. These simulations capture the behavior of proteins and other biomolecules in full atomic detail and at very fine temporal resolution. Major improvements in simulation speed, accuracy, and accessibility, together with the proliferation of experimental structural data, have increased the appeal of bio-molecular simulation to experimentalists—a trend particularly noticeable in, although certainly not limited to neuroscience. Simulations have proven valuable in deciphering functional mechanisms of proteins and other biomolecules, in uncovering the structural basis for disease, and in the design and optimization of small molecules, peptides, and proteins.

1.2 Historical Background

MD simulations are not new. The first MD simulations of simple gasses were performed in the late 1950s (Alder and Wainwright,1957). The first MD simulation of a protein was performed in the late 1970s

(McCammon et al., 1977), and the groundwork that enabled these simulations was among the achievements recognized by the 2013 Nobel Prize in Chemistry (Levitt and Lifson, 1969; Lifson and Warshel, 1968). MD simulations have, however, become substantially more popular and visible in recent years, particularly from the perspective of experimental molecular biologists . Simulations have begun to appear frequently in experimental structural biology papers, where they are used both to interpret experimental results and to guide experimental work.

Molecular dynamics (MD) simulation, since the late 70s, has advanced from simulating several hundreds of atoms to systems with biological relevance, including entire proteins in solution with explicit solvent representations, membrane embedded proteins, or large macromolecular complexes like nucleosomes or ribosomes. Simulation of systems having ~50,000–100,000 atoms are now routine, and simulations of approximately 500,000 atoms are common when the appropriate computer facilities are available. This remarkable improvement is in large part a consequence of the use of high performance computing (HPC), and the simplicity of the basic MD algorithm.

1.3 Computational Power

The utility of molecular dynamics simulations is still limited by two principal challenges :1)the force fields used require further refinement, and 2) high computational demands prohibit routine simulations greater than a microsecond in length, leading in many cases to an inadequate sampling of conformational states. As an example of these high computational demands, consider that a one-microsecond simulation of a relatively small system (approximately 25,000 atoms) running on 24 processors takes several months to complete.The present generation of computers takes benefit of parallelism and accelerators to speed-up the process. The most popular simulation codes (AMBER, CHARMM, GROMACS, or NAMD) have long been compatible with the messaging passing interface (MPI). When a large number of computer cores can be used simultaneously, MPI can greatly reduce the computation time. To benefit the locality of interactions, the general strategy is to distribute the system to simulate among processors. This strategy is called spatial decomposition. Only a small fragment of the system has to be simulated in each processor. The most efficient division is not based

in the list of particles, but in their position in space. Each processor deals with a region of space irrespective of which particles are present there. Communication between processors is also reduced, as only those simulating neighboring regions have to share information. As stated, the use of accelerators, mainly GPU, has become a major breakthrough in simulation codes. Originally designed to handle computer graphics, GPUs have evolved into general-purpose, fully programmable, high-performance processors and represent a major technical improvement to perform atomistic MD. Most major MD codes have already been prepared for GPUs, and even MD codes written specifically to be used on GPUs have been developed. Simulation on GPUs alone or combined with MPI is, at present, the default strategy for high-throughput MD simulations. Remarkably, while simulations have been the most popular use of HPC in life sciences, the increasing power and sophistication of GPUs is leading to a greater use of personal workstations with a comparable performance. Pure computational brute force, just making longer simulations, is not enough to extend the conformational sampling in biomolecular systems. The complex shape of the free energy landscape makes most of the simulations explore just a small region around the energy minimum closest to the initial conformation. With the availability of the present HPC systems, an obvious strategy is to perform a series of parallel simulations with several starting conformations. Although this could be efficient, it requires a specific knowledge of the system to simulate, and cannot be applied as a general strategy. This approach is particularly useful when several crystal structures are available (for instance in the case of allosterically regulated enzymes).

1.4 Statistical Mechanics

In molecular dynamics simulation we use statistical mechanics to provide a connection between the macroscopic properties of materials in thermodynamic equilibrium, and the microscopic behaviors and motions occurring inside the material. The distribution of the system inside the ensemble follows Boltzmann distribution.

1.5 Deterministic approach

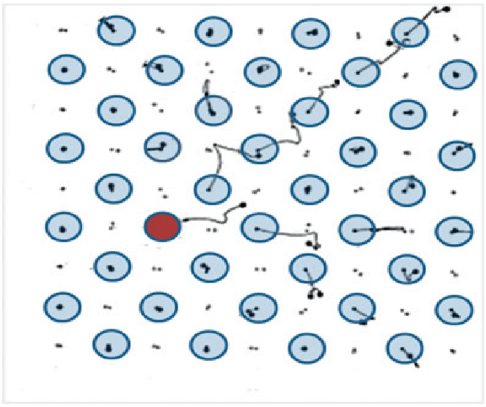
The use of quantum mechanics to describe the evolution of a complex molecule by solving the wave functions of each subatomic particle would be accurate but it would be very difficult to model it and the need of great computational power would be necessary.

Instead we use less accurate classical mechanics model to describe our system and try to alleviate the problem of accuracy by using constants in our classical equations derived from quantum mechanics.

2 Theoretical Background.

2.1 Atomic Force Field Model of Molecular Dynamics Simulation. Basic equations and potential

Molecular dynamics simulation is based on II Newton's law $\mathbf{F_i=m_ia_i}$ **(1)** , for each atom of a system of N atoms.



Here m_i is the atom's mass, a_i its acceleration, r is constant and F is the force acting upon it due to the interaction with other atoms.

If the initial positions and velocities of the particles are determined the solution of Newton's law gives the time evolution of a set of particles.

From Newton's I law we know that a particle at rest will remain at rest and a particle in motion with e certain velocity will continue to move with that velocity until an external force acts on it.

The force can be written as the gradient of the potential energy

Fi= $-\nabla i U$ (2),

where $U(r_1,\ldots,r_n)$,represents the potential energy of N interacting atoms as a function of their positions $r_i=(x_i,y_i,z_i)$.

2.1.1 What is Force Field?

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.

$$V(r)=E_{bonded}-E_{non-bonded} \quad \textbf{(3)}$$

The total potential energy can be calculated by adding the individual potential energies of the molecule caused by:

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

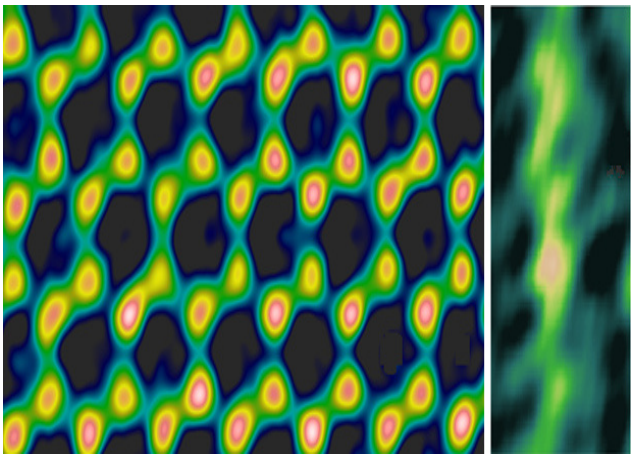
2.1.2 Non bonded atoms

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

$$U_{non-bonded}=U_{van\ der\ waals}+U_{electrostatic} \quad \textbf{(4)}$$

2.1.3 Van der Waals potential

In molecular physics, the **van der Waals force**, named after Dutch physicist is a distance-dependent interaction between atoms or molecules. Unlike ionic or covalent bonds, these attractions do not result from a chemical electronic bond. They are



comparatively weak and therefore more susceptible to disturbance. The van der Waals force quickly vanishes at longer distances between interacting molecules. 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.

$$U_{LJ} = \sum_{i,j} \left[\frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^{6}} \right] \quad (5)$$

2.1.4 The electrostatic potential

The **electric potential** (also called the *electric field potential*, potential drop, the **electrostatic potential**) is defined as the amount of work energy needed to move a unit of electric charge from a reference point to the specific point in an electric field. More precisely, it is the energy per unit charge for a test charge that is so small that the disturbance of the field under consideration is negligible.

$$U_{el} = \sum_{i,j} \frac{q_i q_j}{\varepsilon r_{ij}} \quad (6)$$

2.1.5 Bonded Atoms

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

$$U_{bonded} = U_{bond-stretch} + U_{angle-bend} + U_{torsion}$$

Where

Valence Length potential: $$U_b = \frac{1}{2} \sum_{b} K_b (r - b_0)^2 \quad (7)$$

Valence Angle Potential: $$U_\theta = \frac{1}{2} \sum_{\theta} K_\theta (\theta - \theta_0)^2 \quad (8)$$

Torsion Dihedral Potential: $$U_\varphi = \frac{1}{2} \sum_{\varphi} K_\varphi [\cos(n\varphi - \delta) + 1] \quad (9)$$

2.1.6 Hydrogen bonding potential

After giving force field potentials, the next step is velocity generation

This can be achieved by either Maxwell velocity distribution or random number generators.

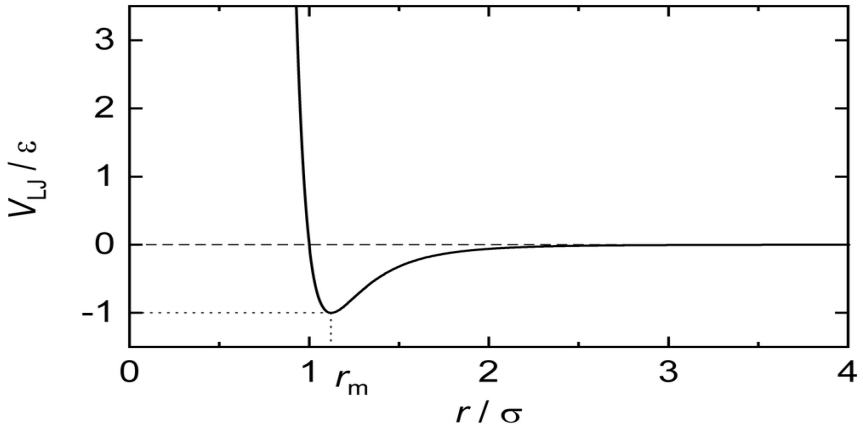
$$T(t) = \frac{1}{3Nk_B}\sum_{i=1} m_i v_i^2 \qquad v_i = \frac{dr_i}{dt}$$

2.2 Lennard Jones potential

Lennard-Jones potential is the most widely used potential and is commonly expressed by the following equation:

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

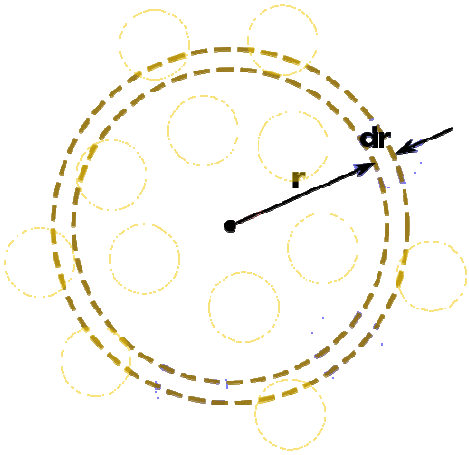
σ and ε are parameters that differ from atom to atom.



2.3 Radial Distribution Function

$$\rho g(\boldsymbol{r}) = \frac{1}{N} < \sum_{i}^{N}\sum_{j\neq i}^{N} \delta\left[\boldsymbol{r} - \boldsymbol{r}_{ij}\right] >$$

In simple terms it is a measure of the probability of finding a particle at a distance of away from a given reference particle, relative to that for an ideal gas. The general algorithm involves determining how many particles are within a distance of and away from a particle. This general theme is depicted to the right, where the red particle is our reference particle, and blue particles are those whose centers are within the circular shell, dotted in orange.



N: total number of atoms

ρ: atomic density

rij : radius vector between two centers i& j

$$g = \begin{cases} 0, for\ distance\ less\ than\ one\ atomic\ diameter \\ 1, for\ larger\ distances \end{cases}$$

2.4 Order Parameter

An **order parameter** is a measure of the degree of order across the boundaries in a phase transition system; it normally ranges between zero in one phase (usually above the critical point) and nonzero in the other. At the critical point, the order parameter susceptibility will usually diverge.

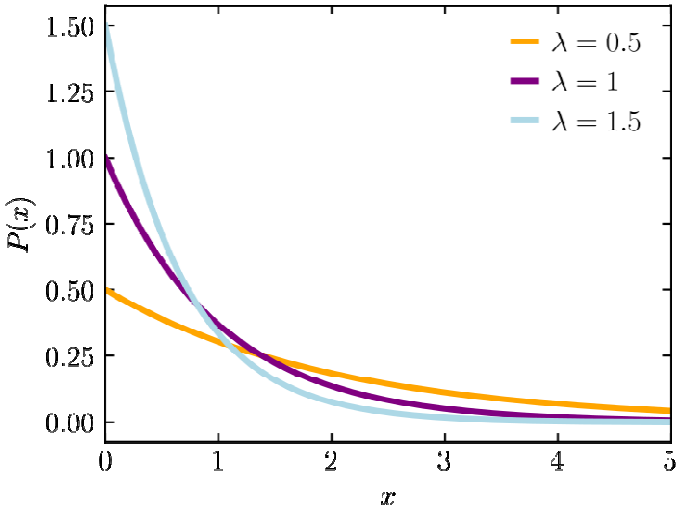
Function-parameter γ is used for distinguishing of the equilibrium states:

$$\gamma_x = \frac{1}{N}\sum \cos(4\pi x_i/a)$$
$$\gamma_y = \frac{1}{N}\sum \cos(4\pi y_i/a)$$
$$\gamma_z = \frac{1}{N}\sum \cos(4\pi z_i/a)$$
$$\gamma = \frac{1}{3}[\![\gamma_x + \gamma_y + \gamma_z]\!]$$

2.5 Boltzmann Distribution

In statistical mechanics, a **Boltzmann distribution** (also called **Gibbs distribution**) is a probability distribution or probability measure that gives the probability that a system will be in a certain state as a function of that state's energy and the temperature of the system

$$H_x(t) = \int^{+\infty} f(v_x)\, lnf(v_x)\, dv_x$$



3 MD Simulation Packages

There are many software packages available for performing bio-molecules simulations like MD-POLY, CHARMM,AMBER and the choice of use is of our favor. In the DL_POLY code there are three input and three output files.

Input files: CONFIG, CONTROL, FIELD

- CONFIG: Contains x-, y- and z- coordinates of all atoms, sets initial values of velocities (V_x, V_y, V_z) and inter-atomic forces (F_x, F_y, F_z), as well as boundary conditions.
- CONTROL: Contains data on: Temperature, Pressure, Step of Integration, Thermodynamic Parameters etc.

- FIELD: Contains information about atoms and molecules (structure, mass, charge, interaction potentials).

Below are listed some of the most common software packages.

I. **AMBER** (www.ambermd.org) The Amber software package (Assisted Model Building withEnergy 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 molecularmechanics programs. The package is in the public domain.

II. **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.

III. **DL_POLY** (www.cse.scitech.ac.uk/ccg/software/DL_POLY/) A package for modeling themolecular dynamics of complex systems with both sequential and parallel calculations.
. Freely available for research andeducational purposes.

IV. **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.

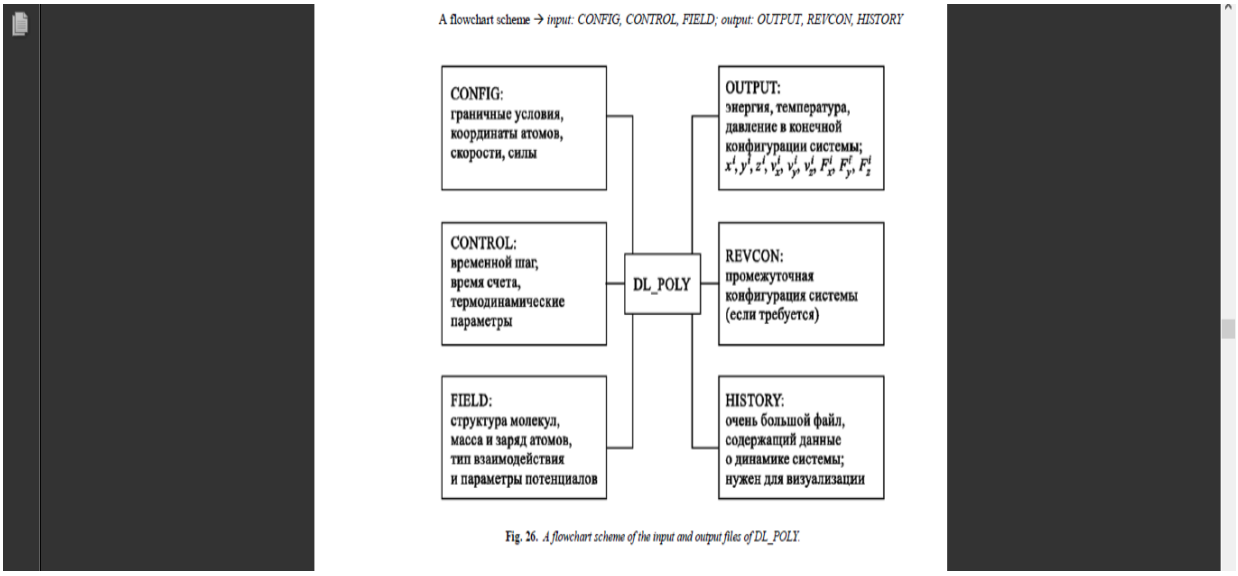
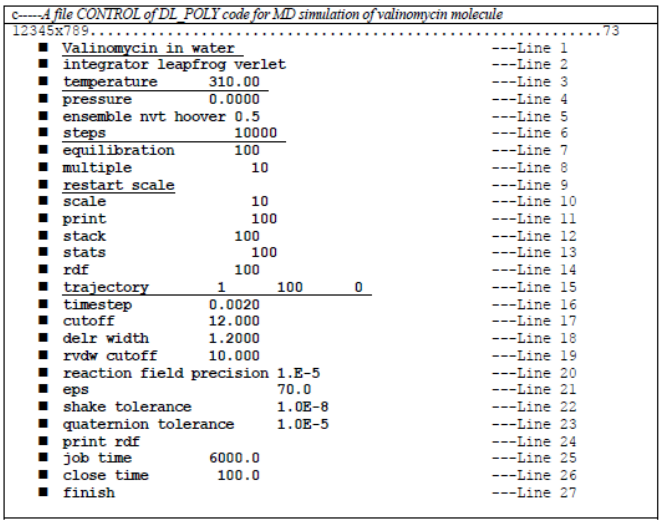


Fig. 26. *A flowchart scheme of the input and output files of DL_POLY.*

Figure 1-Flowchart scheme of the input and output files of DL-POLY



```
c-----A file CONTROL of DL_POLY code for MD simulation of valinomycin molecule
12345x789.....................................................................73
  Valinomycin in water                                          ---Line 1
  integrator leapfrog verlet                                    ---Line 2
  temperature       310.00                                      ---Line 3
  pressure          0.0000                                      ---Line 4
  ensemble nvt hoover 0.5                                       ---Line 5
  steps             10000                                       ---Line 6
  equilibration     100                                         ---Line 7
  multiple            10                                        ---Line 8
  restart scale                                                 ---Line 9
  scale               10                                        ---Line 10
  print               100                                       ---Line 11
  stack             100                                         ---Line 12
  stats               100                                       ---Line 13
  rdf               100                                         ---Line 14
  trajectory        1       100       0                       ---Line 15
  timestep          0.0020                                      ---Line 16
  cutoff            12.000                                      ---Line 17
  delr width        1.2000                                      ---Line 18
  rvdw cutoff       10.000                                      ---Line 19
  reaction field precision 1.E-5                                ---Line 20
  eps                      70.0                                 ---Line 21
  shake tolerance          1.0E-8                               ---Line 22
  quaternion tolerance     1.0E-5                               ---Line 23
  print rdf                                                     ---Line 24
  job time          6000.0                                      ---Line 25
  close time         100.0                                      ---Line 26
  finish                                                        ---Line 27
```

Figoure 2-A file control of DL-POLY for the MD simulation of valinomycin-a drug chain in a water solvent

4 MD Simulation example of amyloid-β peptide.

Amyloid-β is an important factor in Alzheimer's Disease (AD).It is a short peptide that forms soluble oligomers, filaments and fibrils and finally plaques in patients' brains.

We use VMD software for simulating the molecule and a pdb file of Zinc-binding domain of Alzheimer's disease amyloid beta-peptide in TFE-water (80-20) solution. By adjusting the graphical representation we can choose which the part of the protein to simulate.

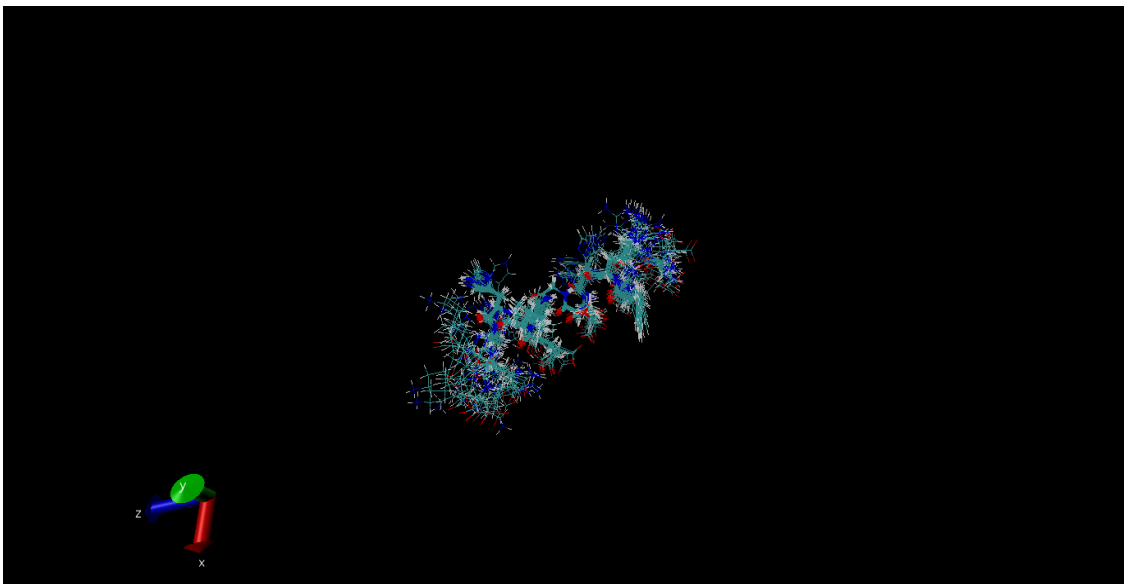


Figure 3-Protein

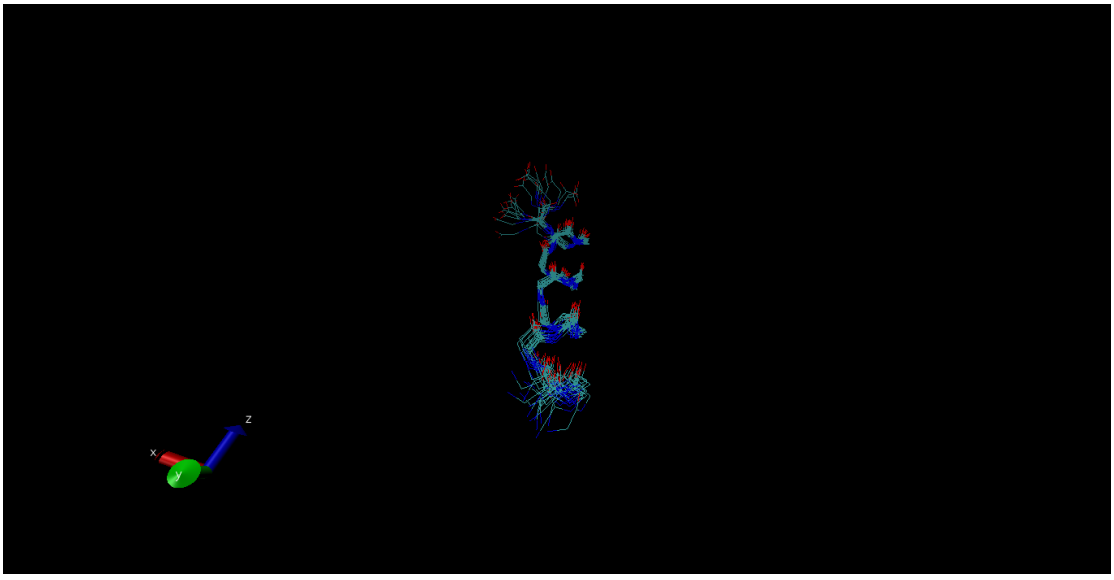


Figure 4-Backbone

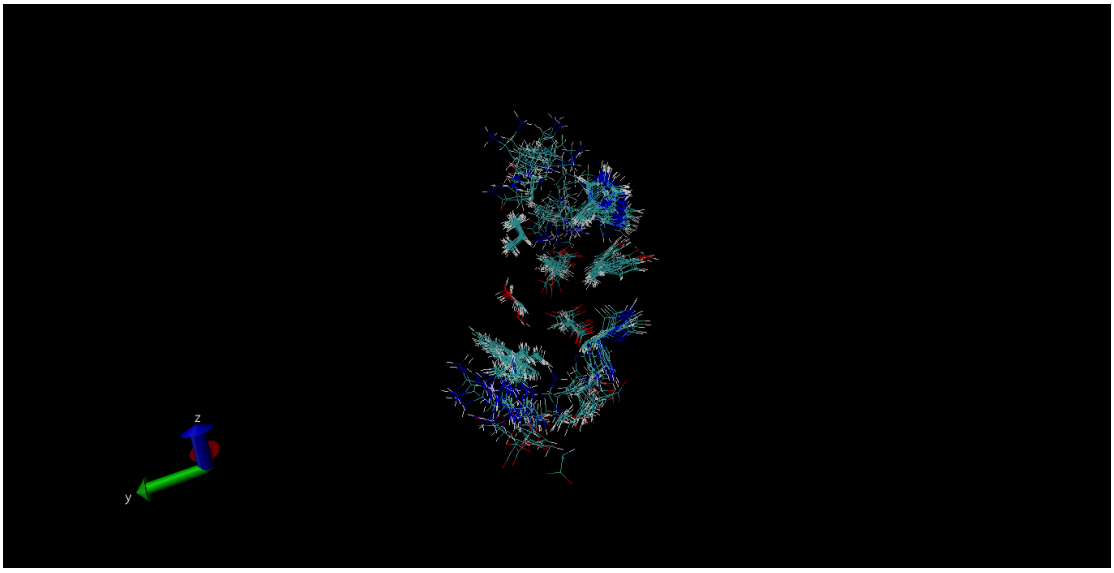


Figure 5-Sidechain

5 Conclusions

Richard Feynman, recipient of the 1965 Nobel Prize inPhysics, once famously stated: 'If we were to name the most powerful assumption of all, which leads one on and on in an attempt to understand life, it is that all things are made of atoms, and that everything that living things do can be understood in terms of the jigglings and wigglings of atoms.' Much of the biophysics of the last 50 years has been dedicated to better understanding the nature of this atomic jiggling and wiggling. The quantum-mechanical laws governing motions in the microscopic world are surprisingly foreign to those familiar with macroscopic dynamics. Motions are governed not by deterministic laws, but by probability functions; chemical bonds are formed not mechanically, but by shifting clouds of electrons that are simultaneously waves and particles. As Feynman eloquently put it, this is 'nature as she is absurd'. Understanding these absurd molecular motions is undoubtedly germane to many scientific processes such as drug discovery. For example the initial 'lock- and-key' theory of ligand binding, in which a frozen, motionless receptor was thought to accommodate a small molecule without undergoing any conformational re-arrangements, has been largely abandoned in favor of binding models that account not only for conformational changes, but also for the random jiggling of receptors and ligands.

Unfortunately, the calculations required to describe the absurd quantum-mechanical motions and chemical reactions of large molecular systems are often too complex and computationally intensive for even the best super-computers. Molecular dynamics (MD) simulations, first developed in the late 1970s, seek to overcome this limitation by using simple approximations based on Newtonian physics to simulate atomic motions, thus reducing the computational complexity. As for example ligand binding and other important macromolecular motions are microscopic events that take place in mere millionths of a second, a complete understanding of the atomistic energetics and mechanics of binding is unattainable using current experimental techniques. Molecular dynamics simulations are useful for filling in the details where experimental methods cannot. With constant improvements in both computer power and algorithm design, in the future molecular dynamics simulations are likely to play an increasingly important role in the future field of biophysics and biomedical research.

6 Future goals

As soon as I finish my undergraduate program I am planning to pursue a Master's Degree in Biophysics. My biggest interest is in the potential use of big data in oncology for a more targeted and customized treatment. Molecular Dynamics simulation has been proved an invaluable tool for cancer treatment. For example a big challenge for oncologists is non-small cell lung cancer (NCSLC) which is **the most common type of lung cancer**. **EGFR (epidermal growth factor receptor)** is a protein on healthy cells that helps them grow. A mutation in the gene EGFR can make it grow too much, which can cause cancer. A mutation in the EGFR gene is one biomarker that physicians look for in non-small cell lung cancer. Like all cancers, NSCLC begins at the cellular level and causes abnormal cells in the lungs to reproduce rapidly and out of control. NSCLCs are carcinomas, which are cancers of the cells lining the surface of the lung airways. A variety of rare mutations account for 10–20% of *EGFR* mutations in nonsmall cell lung cancer. However, due to high diversity, proper medication for patients with such mutations is impossible in daily clinic. To appropriately treat lung cancer patients harboring such rare *EGFR* mutations, a robust prediction model to predict sensitivities of rare *EGFR* mutants to existing drugs is strongly

needed. Using **molecular dynamics simulation-based model**, scientists successfully can predict diverse sensitivities of EGFR exon 20 insertion mutants to existing inhibitors. The findings suggest the usefulness of in silico simulation to overcome mutation diversity at a clinically relevant level.

7 Acknowledgements

I would like to thank Professor Kholmirzo Kholmurodov for the exceptional teaching performance and excellent collaboration during the program. I am also grateful for giving me the opportunity to participate in the Interest program which allowed me to acquire valuable scientific knowledge. At last I would like to thank JINR for this very well designed program for young scientists which lights the spark for them to dream big.

8 REFERENCES

- [1]Petrophysical Characterization and Fluids Transport in Unconventional Reservoirs
  Authors :YangNing∗BikaiJin†YuLiang‡GuanQin∗

- [2]Molecular Dynamics Simulation for All
  Authors:Scott A. Hollingsworth1,2,3,4 and Ron O. Dror

- [3]Molecular dynamics simulations: advances and applications
  Authros :Adam Hospital,[1] Josep Ramon Goñi, Modesto Orozco,and Josep L Gelpí

- [4]On the Use of Molecular Dynamics Simulations for Elucidating Fine Structural, Physico-Chemical and Thermomechanical Properties of Lignocellulosic Systems: Historical and Future Perspectives
  Authors:Krishnamurthy Prasad, Mostafa Nikzad, Shammi Sultana Nisha, Igor Sbarski.

- [5]Molecular dynamics simulation-guided drug sensitivity prediction for lung cancer with rare *EGFR* mutations(May 1, 2019 | 116 (20) 10025-10030)
  Authors: Shinnosuke Ikemura, Hiroyuki Yasuda, Shingo Matsumoto, Kenzo Soejima

- [6]Molecular dynamics simulations and drug discovery
  Authors: Jacob D Durrant and J Andrew McCammon

- [7]Kholmirzo Kholmurodov
  MD-SIMULATION IN CHEMICAL RESEARCH:
  FROM ATOMIC FRAGMENTS TO MOLECULAR COMPOUND

- [8]https://www.researchgate.net/publication/259422749_Three_decades_of_many-body_potentials_in_materials_research

- [9]Samsiq - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=115953177
- https://en.wikipedia.org/w/index.php?curid=65521054

- [10]By Newystats - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=80359746