

JOINT INSTITUTE FOR NUCLEAR RESEARCH

Frank Laboratory of Neutron Physics

**FINAL REPORT ON THE**

**INTEREST PROGRAMME**

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

**Supervisor:**

Prof. Kholmirzo Kholmurodov

**Student:**

Uliana Alatyrtseva, Russia  
Far Eastern Federal University

**Participation period:**

May 24 – July 02, Wave 4

Dubna, 2021

**Abstract**

In the report, I describe the basic concepts of molecular dynamics that I have mastered during the course. I have studied the basic equations and methods of applying force fields to modeling molecular systems. I analyzed the specifics of using different force fields for different tasks of molecular dynamics. The methods of molecular dynamics are based not only on the laws of classical physics, but also on the laws of quantum physics, for example, molecular orbitals (MO), density functional theory (DFT). The direction of my research is the study of knotted proteins. Previous studies have already created molecular dynamics models for knotted proteins. Now, I use a soliton model for modeling knotted proteins, it has high accuracy and efficiency. During the course, I received the necessary knowledge and skills that will allow me to combine the use of two approaches to obtain more complete and accurate research results.

**Introduction**

A fundamental appreciation for how biological macromolecules work requires knowledge of structure and dynamics. Molecular dynamics simulations provide powerful tools for the exploration of the conformational energy landscape accessible to these molecules, and the rapid increase in computational power coupled with improvements in methodology makes this an exciting time for the application of simulation to structural biology.

Field of application of the molecular dynamics:

* Structure of small molecules
* Structure of macromolecules-proteins and nucleic acids
* Study of intermolecular interactions
* Study of the dynamic behavior of molecules
* Study of reactivity
* Molecular design

**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. For the energy minimization of macromolecules, adequate molecular force fields such as AMBER, CHARMM, DY\_POLY, GROMOS have been developed. In these approaches, the coordinates of all atoms of the macromolecule are treated as free variables [1]. The basic functional form of a molecular force field includes terms for covalent bonds and terms for long-range forces (non-bonded interactions) inside the molecule:

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.

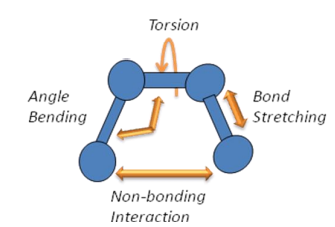
For the covalent bond terms, parametrized by bond length, bond angles and dihedral angles, the potential energy is described relative to the atoms being in their equilibrium positions, for which the energy is taken to be zero. The first term in the molecular force field describes the extension (stretching)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

Figure 1 - Different types of interactions [1]

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 Kθ is measured by Raman spectra.

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.

Non-bonded interactions

In common molecular force fields the non-bonded interactions include electrostatic forces and van der Waal forces.

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.

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.

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. [3]

**The computer code description for 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. For simulation of liquid model is using FORTRAN.

The use of FORTRAN as the main algorithmic language in the MD method is primarily motivated by the very essence of the MD modeling technology or method. In a computer implementation of MD, we have to deal with a huge set of similar fragments or loops that are executed repeatedly in the same way. For example, calculations of the same quantities (potentials and forces of interaction of atoms, values of velocities and coordinates, etc.) are achieved on the basis of-promotion of the identical cycles in the program, only moving from one an atomic index to another. The calculation of such computational cycles and the overall MD of the program, it turns out, is most quickly performed in the FORTRAN language, which, moreover, is convenient and easy to write, as well as compile and load under the standard Linux operating system [2].

**The use of selected general-purpose code for the simulation of ionic, polymeric and biochemical molecular systems.**

In today computational chemistry and nanotechnological application a lot of potentials have developed, optimized and adapted in general-purpose packages like DL\_POLY, AMBER, CHARMM, NAMD, etc.. The peculiarity of these programs is that they cover a wide range of molecular systems-from simple atomic structures to ionic systems, polymer and biochemical macromolecules.

Table 1 – Application of different force fields

|  |  |
| --- | --- |
|  | **Target application** |
| DL\_POLY | Multi-purpose package is used for MD modeling of a wide variety of molecular systems – from simple atomic fragments to ionic structures, polymers and biochemical macromolecules |
| AMBER | Simulation of peptide, protein, nucleic acid, 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, as they occur in solution, crystals, and membrane environments. CHARMM also finds broad applications for inorganic materials with applications in materials design. |
| NAMD | This package is designed for high-performance simulation of large biomolecular systems |

At the top of hierarchical complexity for simulation proteins one finds full atomistic representations combined with realistic force fields (e.g. AMBER and the GROMOS are popular choices among researchers), which are explored with classical MD simulations. Apart from providing realistic energetics this approach allows one to simulate folding in explicit water. Its major advantage is the possibility to directly compare simulation data with data from in vitro experiments. A fundamental problem of classical MD is the accuracy of the force fields. Another (less important) drawback is the need to consider very small time steps to integrate the equations of motion. This constraint imposes severe limitations on the total amount of simulation time and renders their systematic application to protein folding (and other dynamical processes involving large scale conformational changes) non-trivial. For this reason, smart sampling methods and sophisticated distributed computing schemes have been developed to conduct classical MD of protein folding, and novel algorithms and machine architectures have been created to execute MD simulations orders of magnitude faster than was previously possible. A paradigmatic example of the latter is the ANTON machine developed by DE Shaw Research [4].

**The theory of the basics of hybrid MD approach (classical quantum-chemistry potentials simulation methods)**

Quantum chemical methods of molecular dynamics (ab initio quantum chemistry or qMD) are based on the Schrodinger equation. In contrast to ab initio or first principles calculations, the conventional methods of MM (molecular mechanics) and MD (molecular dynamics) are based on classical concepts. Methods of molecular mechanics use the approach of traditional chemistry, when in which the molecules are represented as a set of balls and rods, with each ball representing an atom, and each rod representing a bond between them. Particles are considered here as material points interacting through force fields, which, in turn, are determined by interaction potentials. Depending on the type of bonds, the interaction potentials are selected, as well as the energy and parameters corresponding to certain local configurations of the atoms.

In this approach, molecular mechanics treats the potential energy as the sum of the terms, describing the stretching, bending, and torsion of bonds, as well as, in the case of the van der Waals interaction, the overlap and electrostatic interaction between unbound atoms.

But the general goal of molecular mechanics, as well as calculations-from the first principles - is to find stable configurations (local minima) of potential energy for systems of many particles. The solution of the quantum Schrodinger equation for a system of many particles is not possible - the required calculation time increases exponentially with an increase in the number of particles.

Therefore, the essence of the hybrid approach in MD modeling, involving more accurate quantum chemical approximations, will consist in finding reasonable approximations and possibilities separation of variables that simplify the calculation scheme without involving experimental data. The most common ab initio calculation methods, ―from the first principles, are self-consistent field models, molecular orbitals( MO), density functional theory (DFT) [2]

**Application of molecular dynamics**

The topic of my research is the knotted proteins, in particular protein MJ0366 in the form of trefoil knot. Knots in proteins were identified in the year 1994 but their topological arrangement, mechanism and biological significance are still indistinguishable. Knots are usually defined according to the number of crossings in a projection on a plane. To date around 2% of protein structures in PDB were reported to possess as knotted structural arrangements. At present various types of KNOT's were identified and reported, which includes 31, 41, 52, 61, 62 and 63 types with respect to the number of crossing. Most of the knot proteins adopt 31 topology whereas 61 topological arrangements are scarce. It is reported that DNA structure also form knotted conformation, leads a vital role in biological functions from transcription and replications.

Knots play an important role in mechanical and thermal stability of proteins. The theoretical and experimental studies of transcarbamylase protein indicate that the knotted proteins are more stable than unknotted structures. It is also believed that knotted proteins involved in disease related functional disorders such as Parkinson disease, neurological disorders etc, nevertheless they are also used as drug targets for many diseases [5].

Researchers have already modeled the dynamics of natively-knotted protein, MJ0366, based on a realistic force field. They used the AMBER ff99SB force field in implicit solvent, within the Generalized Born formalism implemented in GROMACS 4.5.2. In such an approach, the Born radii are calculated according to the Onufriev-Bashford-Case algorithm. The hydrophobic tendency of non-polar residues is taken into account through an interaction term proportional to the solvent-accessible-surface-area (SASA). The solvent-exposed surface of the different atoms is calculated from the Born-radii, according to the approximation developed by Schaefer, Bartelsand Karplus.

They investigated the folding pathways of a protein with complex native topology. By using the dominant-reaction pathway scheme they collected about 30 successful folding trajectories for the 82-amino acid long trefoil-knotted protein. Despite the dissimilarity of the initial unfolded configuration, these trajectories reach the natively-knotted state through a remarkably similar succession of steps.

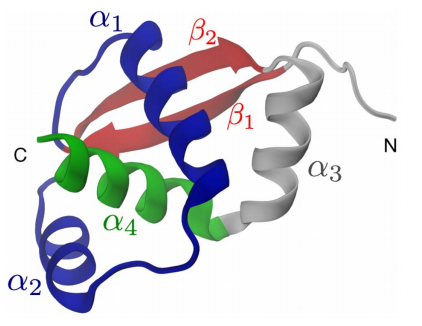
In particular it is found that knotting occurs essentially through a threading mechanism, involving the passage of the C-terminal through an open region created by the formation of the native b-sheet at an earlier stage. The dominance of the knotting by threading mechanism is not observed in MJ0366 folding simulations using simplified, native-centric models. This points to a previously underappreciated role of concerted amino acid interactions, including non-native ones, in aiding the appropriate order of contact formation to achieve knotting [6].

Figure 2 - Crystal structure of protein MJ0366, PDB code: 2EFV, modeling using VMD [6]

I also made model of protein M0366 using VMD by myself.

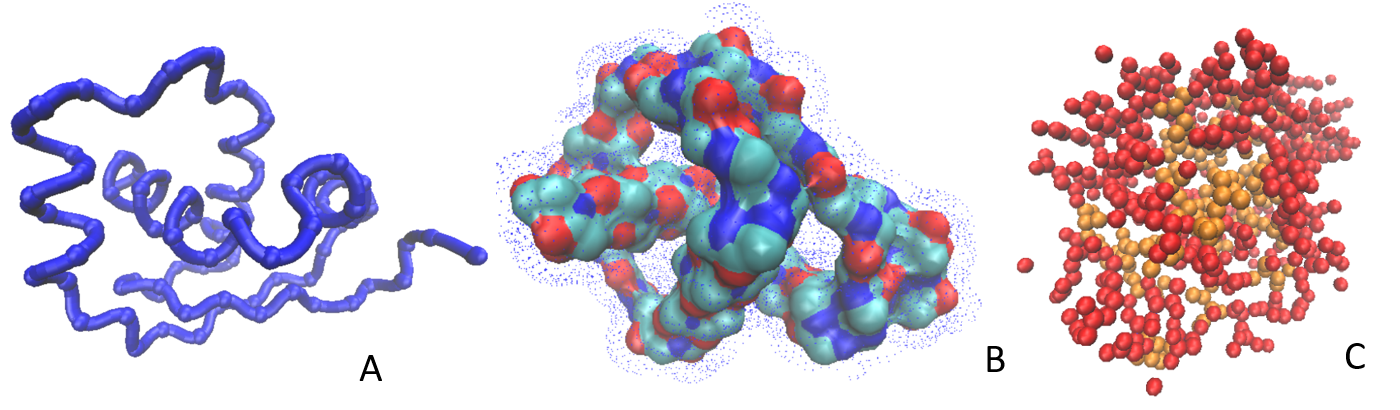
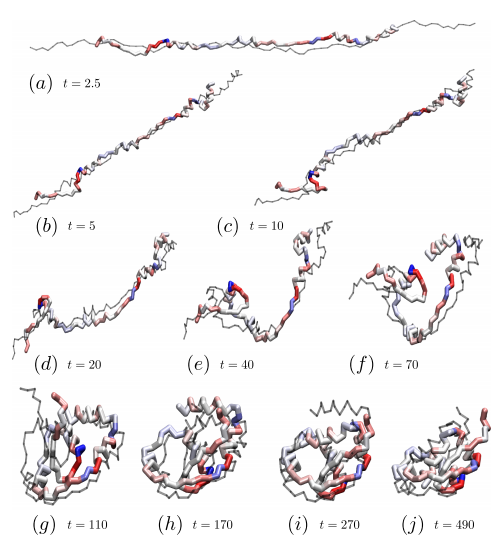
Another researchers have used a Coarse-grained model: Elastic Folder Model (EFM). The EFM employs a minimalistic representation of the protein based on the sole Cα atoms, connected to their first neighbors by means of stiff bonds. The only non-bonded interaction to which the centroids are subject is a short-ranged excluded volume, enforcing steric hindrance and preventing the chain from crossing itself. This tube-like model of the protein is then provided with bending and torsion potentials, whose reference angles are parametrized on a target structure. The latter could be, in principle, the PDB crystal structure. By construction, the reference structure represents the global minimum of the EFM; hence, the latter, being based on the folded (and knotted) conformation, falls in the category of structure-based models. It is important to stress, however, that the interactions in the system do not favor the formation of native contacts: the effective energy is minimized only by the onset of the target local arrangement of the residues. The reference angles represent the sole input parameters introduced in the model. The strength of the bending and torsion potentials is in fact determined by a Monte Carlo (MC) search, in which the set of parameters maximizing the successful collapse in the reference state is obtained. As the result, researchers have got an optimal knotting process of protein MJ0366 as obtained from a simulation employing the refined force field of the EFM on the figure 3 [7].

Figure 3 - Models of protein M0366 using VMD: A) Protein backbone, Cα atoms, B) Protein surface in solvent C) Hydrophobic residues (orange) inside the molecule

Figure 4 - Optimal knotting process of protein MJ0366 [7]

I with my colleagues make my molecular modeling and molecular dynamics based on the soliton model of protein. This approach based on the concept of gauge invariance to scrutinize the extrinsic geometry of strings in three dimensional space. General principles of symmetry are being used in combination with Wilsonian universality and derived an essentially unique Landau-Ginzburg energy that describes the dynamics of a generic string-like conﬁguration in the far infrared. The energy supports topological solitons, that pertain to an anomaly in the manner how a string is framed around its inﬂection points. The solitons can operate as modular building blocks from which folded proteins are composed. The crystallographic protein structures is described by multi-solitons with experimental precision, and investigated in the non-equilibrium dynamics of proteins under varying temperature. Using model of solitons it is possible to simulate the folding process of a protein at in vivo speed and with close to pico-scale accuracy using a standard laptop computer [8].

It was developed a global technique that is rooted on topological concepts, to analyze and describe the formation of topological structures in proteins, in particular aspects of knottiness and self-entanglement. For this, the entire Cα backbone of a protein have been modeled in terms of a single topological multi-soliton entity; the multi-soliton describes the minimum of a mechanical free energy. As a case study, we have investigated the folding of the protein MJ0366 in the form of trefoil, instigated by variable ambient temperature, using powerful state-of-art Monte Carlo techniques of non-equilibrium thermodynamics. We have found that the multi-soliton describes the formation of the trefoil knot very accurately. As a results, we made soliton model of the protein MJ0366 with accuracy less than 0.6 Å (RMSD distance between the two structures). On the figure 5 shown overlay comparison of PDB structure MJ0366 (green) and its multi-soliton structure (blue). Also we are able to describe the folding pathways and make predictions on the physical origin of knot formation.

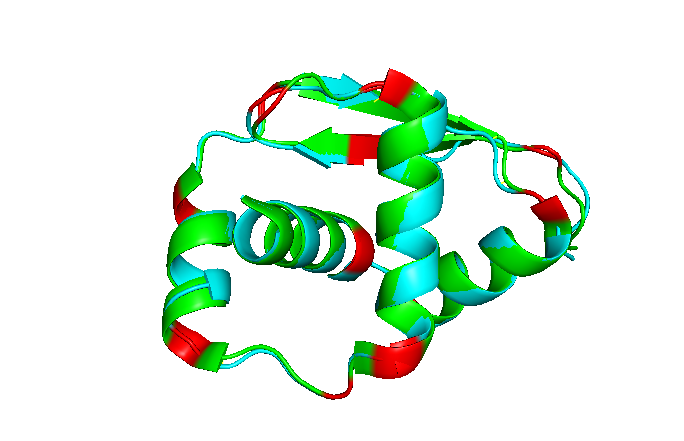
Our results demonstrate the value of developing global approaches to protein folding and dynamics; global approaches are highly accurate, and even though they may lack in atomic level details they appear to correctly capture the global, topological aspects of self-entanglement during protein folding and dynamics.

Figure 5 – The overlay comparison of PDB structure 2J6B and its multi-soliton structure.

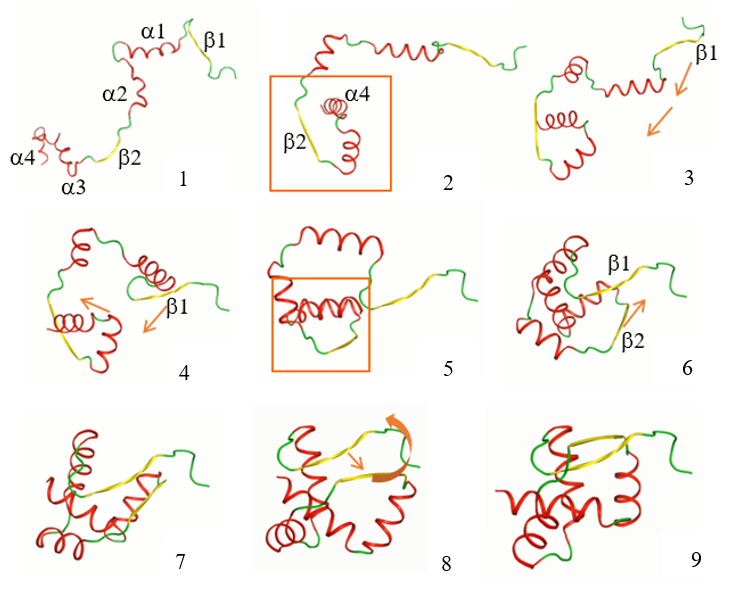
My colleagues using he Glauber algorithm for modeling dynamics. In the simulations they adiabatically decrease and increase the value of the temperature factor β, to simulate heating and cooling. The results of protein dynamic under the cooling is shown on the figure 6.

Figure 6 - Protein dynamic (folding) under the cooling conditions

This modeling correspond with previous studies of folding protein MJ0366 [7].

**Prospects**

In the future, I plan to continue modeling proteins based on the soliton model of other proteins in the form of a different knots, but I also will use the knowledge gained in this course in the application of molecular dynamics methods.

**References**

1. Kholmirzo KHOLMURODOV (Editor), «Computer Design for New Drugs and Materials: Molecular Dynamics of Nanoscale Phenomena», Nova Science Publishers (N.Y.), ISBN: 978-1-53612-082-0, (2017).
2. Kholmirzo KHOLMURODOV (Editor), «Computational Materials and Biological Sciences», Editor: Kholmirzo T. Kholmurodov (Leading Scientist, Frank Laboratory of Neutron Physics, Joint Institute of Nuclear Research, Dubna, Moscow Region, Russia), Nova Science Publishers (N.Y.), ISBN: 978-1-63482-541-2, (2015).
3. Marco Wiltgen, Algorithms for Structure Comparison and Analysis: Homology Modelling of Proteins, Graz General Hospital and University Clinics, Graz, Austria 2019 Elsevier Inc
4. Faísca, P. F. N. Knotted Proteins: A Tangled Tale of Structural Biology. Comput. Struct. Biotechnol. J. 13, 459–468. doi:10.1016/j.csbj.2015.08.003, (2015).
5. Thiruselvam, V., Kumarevel, T., Karthe, P., Kuramitsu, S., Yokoyama, S., & Ponnuswamy, M. N., Crystal structure analysis of a hypothetical protein (MJ0366) from Methanocaldococcus jannaschii revealed a novel topological arrangement of the knot fold. Biochemical and Biophysical Research Communications, 482(2), 264–269, 2017
6. Beccara S, Sˇkrbic´ T, Covino R, Micheletti C, Faccioli, P Folding Pathways of a Knotted Protein with a Realistic Atomistic Force Field. PLoS Comput Biol 9(3),2013
7. Najafi S, Potestio R. Folding of small knotted proteins: Insights from a mean field coarse-grained model. J Chem Phys. 2015 Dec 28;143(24):243121. doi: 10.1063/1.4934541. PMID: 26723606.
8. Niemi, A.J. Gauge fields, strings, solitons, anomalies, and the speed of life. *Theor Math Phys* 181**,**1235–1262 (2014). https://doi.org/10.1007/s11232-014-0210-x