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)

Student:

Roman Moryachkov, Russia

Kirensky Institute of Physics

Supervisor:

Prof. Kholmirzo Kholmurodov

Participation period:

September 27 – November 05, Wave 5

Dubna, 2021

Contents

Summary

In the course basic approaches to molecular dynamics (MD) simulations are presented and corresponding calculations implemented into the demonstrated software are shown. Computer simulations are becoming a very useful instrument in chemistry, biology and other branches of science. To model large biomolecules in their native environment one needs corresponding computational resources and applying methods. In our work we applied molecular simulations to reconstruct the spatial structure of the DNA aptamer molecule on the basis of X-ray scattering data.

Introduction

Modern world raises new challenges for the scientific community. To develop new materials with desirable properties and characteristics, investigate any biomolecule and create a new drug, we need to use new approaches to accelerate, facilitate and simplify their design and development, validation and interpretation of our experimental data. For this purpose the computational methods are applied and become very useful and convenient tools for scientists.

New compact and cheap agents for therapy and diagnostics are wanted with the modern development of personalized medicine. Compact (15-100 bases) synthetic aptamers - single-stranded oligonucleotides based on DNA or RNA - can be such therapeutics agents. Their primary sequence determines the future three-dimensional structure. The unique shape of the molecule and the special distribution of charges on its surface, capable of electrostatic and van der Waals interactions and hydrogen bonds, determines the ability of aptamers to specifically bind with specific types of proteins, for example, proteins presented in the membranes of tumor cells. It makes them synthetic analogs of antibodies [1,2].

The sensitivity of diagnostic test systems and the effectiveness of targeted therapy are determined by the energy of interaction between the recognizing and/or affecting element and the target. Currently, there is no universal algorithm for studying recognition mechanisms in biological systems due to the variety of types of interactions: conformational, induced, electrostatic, dispersion interactions, hydrogen bonds, etc. Traditionally, highly specific antibodies are used as recognition elements in biosensors. However, at present, DNA and RNA oligonucleotides are becoming more and more preferable, since they have a large combinatorial variety. This variety allows them to bind to molecular targets of various nature: from small molecules and ions to proteins and cells. Determining the mechanisms of interaction between molecular recognition elements based on oligonucleotides with targets of various nature is an important fundamental problem. Its solution will increase the accuracy

of diagnostics and the effectiveness of targeted therapy. The main task is to use an integrated approach to study the features of molecular interactions, binding sites, kinetics, energies, and the type of bonds between targets of different nature with molecular recognition elements based on oligonucleotides.

Molecular dynamics (MD) is a computer simulation method used in the theoretical study of biological molecules, such as proteins and nucleic acid, to analyze the physical movements of the constituent atoms and molecules. In the computer simulation, these atoms and molecules interact over time and give a sense of the dynamic evolution of the system. MD simulation mimics the changes in the structures of biological molecules over a given period of time, giving us atomic insights about the change in structure. This data helps us understand biological functions. These simulations give us detailed information about the fluctuations and conformational changes of the proteins and nucleic acids under study. These methods are applied to thoroughly study the organization and dynamics of biological molecules, their complexes, and the conformational changes of proteins and nucleic acids. Many mysteries, on the femtosecond scale, have been revealed through the study of these conformational changes. These methods are applied in chemical physics, materials science, and biophysics. MD simulations are often used in computational biology to study protein-protein interaction, protein-ligand docking, the effects of mutation on interactions, protein folding, and flexibility of the biological molecules.

SAXS is a powerful technique for structure determination of biomolecules in solution as well as their higher-order assemblies, if the initial constituents are known [3-5].The theoretical basis of structural analysis by small-angle scattering on particles was first presented by Guinier [6] and applied to nucleic acids by Timasheff in 1961 [7]. Small-angle scattering intensity from a particle drops exponentially with an angle, and the index of power is associated with a generalized particle size and its radius of gyration. The magnitude of gyration radius depends on the distribution of the scattering center in space, particle size, and shape. Moreover, the form of the scattered radiation intensity curve with an increasing angle is significantly different for the three basic conformations of biological macromolecules: a ball, stick, and disk. SAXS has several advantages over the other structural techniques including minimal sample preparation time, analyzing individual macromolecules or their complexes in solutions at physiological conditions or at any required temperature, pH, and buffer, and rapid data collection and processing [5].

Overview

The goal of Molecular Dynamic Simulation is to predict the behavior of atoms in a biological system and how they move as a time-dependent function thereby providing the ultimate details concerning the atoms based on algorithms of physics that govern the interatomic interactions [6].
MD simulation requires the step-by-step numerical solution of the classical equations of motion [9-11].

The Potential Energy Function is described by an equation:

$$U(\vec{R}) = \underbrace{\sum_{bonds} k_i^{bond}(r_i - r_0)^2}_{U_{bond}} + \underbrace{\sum_{angles} k_i^{angle}(\theta_i - \theta_0)^2}_{U_{angle}} +$$
$$\underbrace{\sum_{dihedrals} k_i^{dihe}[1 + \cos(n_i\phi_i + \delta_i)]}_{U_{dihedral}} +$$
$$\underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}}_{U_{nonbond}}$$

U_{bond} = oscillations about the equilibrium bond length
U_{angle} = oscillations of 3 atoms about an equilibrium bond angle
$U_{dihedral}$ = torsional rotation of 4 atoms about a central bond
$U_{nonbond}$ = non-bonded energy terms (electrostatics and Lenard-Jones)

In molecular dynamics a molecule is described as a series of charged points (atoms) linked by springs (bonds). To describe the time evolution of bond lengths, bond angles and torsions, also the non-bonding van der Waals and elecrostatic interactions between atoms, one uses a force field. The force field is a collection of equations and associated constants designed to reproduce molecular geometry and selected properties of tested structures.

In the context of chemistry and molecular modelling, a force field is a computational method that is used to estimate the forces between atoms within molecules and also between molecules. More precisely, the force field refers to the functional form and parameter sets used to calculate the potential energy of a system of atoms or coarse-grained particles in molecular mechanics, molecular dynamics, or Monte Carlo simulations. The parameters for a chosen energy function may be derived from experiments in physics and chemistry, calculations in quantum mechanics, or both. Force fields are interatomic potentials and utilize the same concept as force fields in classical physics, with the difference that the force field parameters in chemistry describe the energy landscape, from which the acting forces on every particle are derived as a gradient of the potential energy with respect to the particle coordinates.

Implementation of classical potential energy functions:
1. Theoretical functional forms are derived for the potential energy V(r).
2. Definition of atom types that differ by their atomic number and chemical environment, e.g. the carbons in C=O or C-C are of different types.
3. Parameters are determined so as to reproduce the interactions between the various atom types by fitting procedures:
- experimental enthalpies (CHARMM)
- experimental free energies (GROMOS, AMBER)

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

(1) **AMBER** (www.ambermd.org) The Amber software package (Assisted Model Building with Energy Refinement) consists of a set of force fields for modeling macromolecular structures (proteins, nucleic acids and a number of other classes of molecules) and a package of quantum and molecular mechanics programs. The package is in the public domain.
(2) **CHARMM** (www.charmm.org) (Chemistry at HARvard Macromolecular mechanics) software package for molecular modeling of a wide range of systems - from small molecules to biological macromolecules, using various energy functions and models - from quantum models and force fields to molecular mechanics to full-atomic classical potentials.
(3) **DL_POLY** (www.cse.scitech.ac.uk/ccg/software/DL_POLY/) A package for modeling the molecular dynamics of complex systems with both sequential and parallel calculations. Versions available: DL_POLY_2, DL_POLY_3 and DL_POLY_4. Parallel calculations with the number of atoms up to 1 million using 1024 processors are possible. Adapted for graphics game processors, GPU (Graphical Processing Units), using the CUDA language. Freely available for research and educational purposes.
(4) **GROMACS** (www.gromacs.org) A software package for fast simulation of the dynamics of large molecular systems (from thousands to millions of particles). Designed primarily for modeling biomolecules (proteins and lipids) that have many interconnected interactions between atoms. Works in Linux environment and is free.
(5) **LAMMPS** (lammps.sandia.gov) The non-commercial package LAMMPS (Large scale Atomic / Molecular Massively Parallel Simulator) uses classical molecular dynamics methods for modeling and calculating polymers, biomolecules, solids (metals, semiconductors, etc.), as well as coarse-grained mesoscopic systems at atomic, mesoscopic and continual scales.
(6) **MOE** (www.chemcomp.com) MOE (Molecular Operating Environment) is a complex of programs for modeling molecules, in particular large biomolecules. The methods of molecular mechanics and dynamics are developed in it on the basis of various force fields.
(7) **NAMD** (www.ks.uiuc.edu/Research/namd/) An object-oriented program for calculations in the field of interactive molecular dynamics, in particular for modeling large biomolecular systems that require significant resources. The program code is freely distributed for various parallel computing platforms.
(8) **HyperChem** (www.hyper.com) The HyperChem software package (latest version 8.0) includes programs that implement quantum-chemical calculation methods "from first principles" and semi-empirical methods, as well as modeling methods in molecular mechanics and molecular dynamics.
(9) **GAMESS** (www.msg.ameslab.gov/GAMESS/) GAMESS (General Atomic and Molecular Electronic Structure System) is a non-commercial quantum-chemical package that allows calculating molecular wave functions by the self-consistent field method in the RHF, UHF, ROHF, GVB and MCSCF approximations. The main features of the package are: taking into account the electron correlation energy based on perturbation theory, configuration interaction, coupled clusters and the density functional; automatic geometry optimization, search for transient states using analytical gradients; calculation of molecular properties, in particular, dipole moment, electrostatic potential, electron and spin density.
(10) **Gaussian** (www.gaussian.com) Commercial electronic structure modeling package (latest version Gaussian 03) is used for research in chemistry and biochemistry, physics and other known and developing fields related to chemical processes. The Gaussian package, based on first principles methods, allows predictions of energies, molecular structures and vibrational frequencies of molecular systems, along with many other properties of molecules. Methods for accounting for correlation energy

are widely implemented: energy calculation and optimization with analytical gradients for perturbation theory methods, coupled clusters, configuration interaction, density functional, multiconfiguration self-consistent field method are possible.

(11) **VASP** (www.cms.mpi.univie.ac.at/vasp/) Using the VASP (Vienna Ab initio SimulationPackage) package, quantum-mechanical calculations `` from first principles " are carried out in the field of molecular dynamics using pseudopotentials, the method of calculating the electron band PAW structure and plane wave basis.

(12) **MOPAC** (www.openmopac.net) A package of semiempirical programs is used to calculate the electronic structure of the main and excited states of atoms, molecules and solids. MOPAC implements semiempirical methods RM1, PM6, MNDO, AM1 and PM3. When studying the electronic structure of macromolecules (proteins, DNA, polymers and solids), it allows calculating large (up to 15,000 atoms) biomolecules (including enzymes, DNA, etc.) based on the use of localized molecular orbitals.

Molecular structure restoration based on Small-angle X-ray scattering data

The molecular structure restoration modeling of RE31 was carried out with the Avogadro program [12] according to its primary sequence and small-angle scattering experimental data (Fig. 1-II). The primary sequence of RE31 was used to predict the secondary structure by using the Mfold web server [13] on the basis of free energy minimization techniques. The aptamer was divided into three structural parts and their location in space was varied relative to each other in accordance withexperimentaldata.ThefirststructuralpartisaG-quadruplex with the following sequence GGTTGGTGTGGTTGG. It is a well-known 15-nt thrombin binding aptamer (15TBA) formed by two tetrads of four guanines in the classical conformation anti-syn-anti-syn and a potassium ion located in the center [14,15]. The X-ray data of the thrombin complex with 15TBA aptamer has been described previously by the Tulinsky group [16]. The G-quadruplex consists of two stacked planar G-quartets(G-stem) that are associated with three lateral loops. Each G-quartet is connected by eight hydrogen bonds. According to the X-ray data, 15TBA binds to thrombin by embracing the protruding region of exosite I through their TT loops [17,18]. The second structural part of RE31 is a double-stranded part. This is a right-handed B-type helix. The third structural part consists of four non-complementary nucleotides T–G and A–Gs (Fig. 1-III).

The molecular modeling was carried out to construct the initial geometry of aptamer molecules in space. The resulting atomic structure was compared with an experimental form obtained from small-angle scattering. At this stage, determining the linear dimensions of molecules in all three projections was the main goal, while it was difficult to achieve a perfect match in the fine details. As DNA is a chiral molecule, different variants of aptamer local twisting are possible. The largest difficulty arises in the molecular design of non-complementary nucleotides, as they have a large number of possible positions. At first, the model with the basic nitrogen atoms turned in different directions was built. However, this structure is not compliant with the experimental volume. In order to make the model fit with the experimental figure, nucleotides making up the double chain were twisted, similar to a double helix. Thus, we achieved the maximum coincidence with the experimental shape. Therefore, in this step, local twisting of nucleotides was carried out to obtain the best possible match with features of the experimental figure (Fig. 1-IV).

The modeled aptamer molecule should match linear dimensions obtained according to the analysis of the SAXS curve. For example, the linear dimensions from SAXS (height 44 Å; width 26,

23, and 25 Å in Fig. 1-III) match well the linear dimensions of the constructed 3D model (44 Å and 23, 20, and 22 Å, respectively) in Fig. 1-III. Furthermore, even on the existing linear dimensions, it is possible to model a large number of conformations. At the second stage of theoretical modeling it is necessary to ensure that the constructed theoretical model takes into account the peculiarities of the "relief" of the experimentally obtained shape (this shape may have a complex topology). Therefore, in the "fine tuning" step, we achieved (by minor rotations around sigma bonds) the most complete fitting of the theoretical model into the experimental volume. After changing the theoretical geometry of the molecule in accordance with the shape of the electron density from SAXS, we plotted the SAXS curve from the theoretical molecule and compared it with the experimental curve; thus, we checked the consistency of the model with the experiment.

Figure 2 shows the RE31 structure obtained by three different methods. The top image (Fig. 2A) shows the aptamer binding with thrombin revealed by XRD [19]. Figure 2B shows the aptamer structure obtained by XRD, SAXS and MSR. Comparing aptamer conformations obtained by SAXS in solution with XRD from a crystal form with thrombin and MSR showed a good match (Fig. 2C). In both cases, the main difference consists of the quadruplex part of the aptamer, as it acts as a thrombin-binding site. Previously, it was shown that RE31 forms a specific complex with thrombin with an apparent dissociation constant of 0.16 nM [20]. The differences between the aptamer structure before and after addition of the protein can be interpreted as an induced fit model when the aptamer changes its shape upon the binding. Figure 4D shows the RE31 structure constructed fromXRD analysis and MSR. The bonding between terminal nucleotides (G1 and C31) is flexible. The duplex part and non-complementary region between the quadruplex and duplex are not static, but rather flexible.

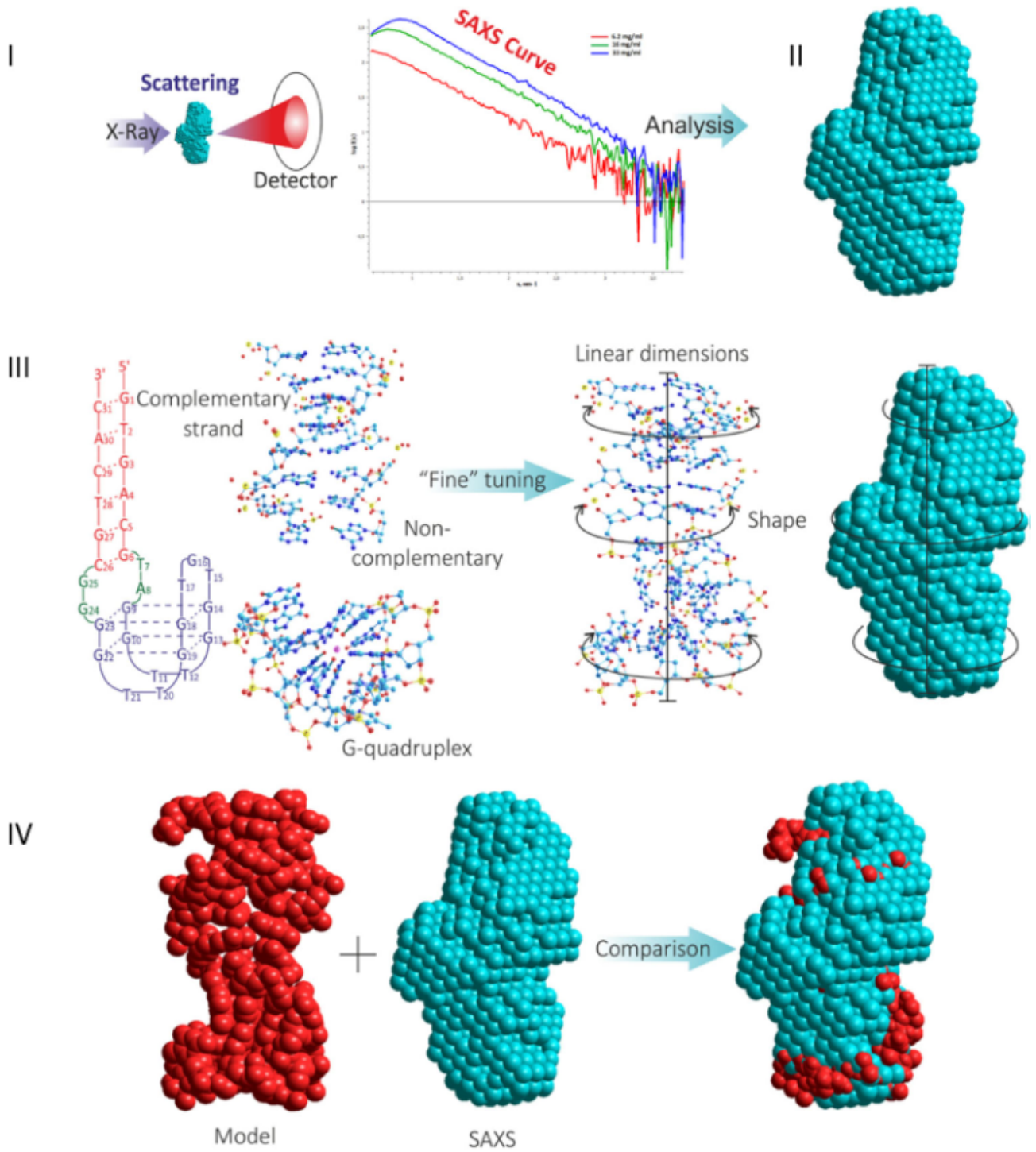

Fig. 1 General scheme of finding 3D structure of an aptamer with SAXS and molecular structure restoration (MSR). (I) acquiring SAXS experimental data of an aptamer in solution, (II) building a spatial distribution of the molecule electron density using SAXS results, (III) constructing a 3D model of the aptamer from its nucleotide primary sequence and secondary structure, and (IV) comparing and refining the modeled 3D structures with the experimental SAXS model (red color, from molecular modeling; blue color, from SAXS data). The modeled aptamer molecule should match linear dimensions obtained according to the analysis of the SAXS curve.

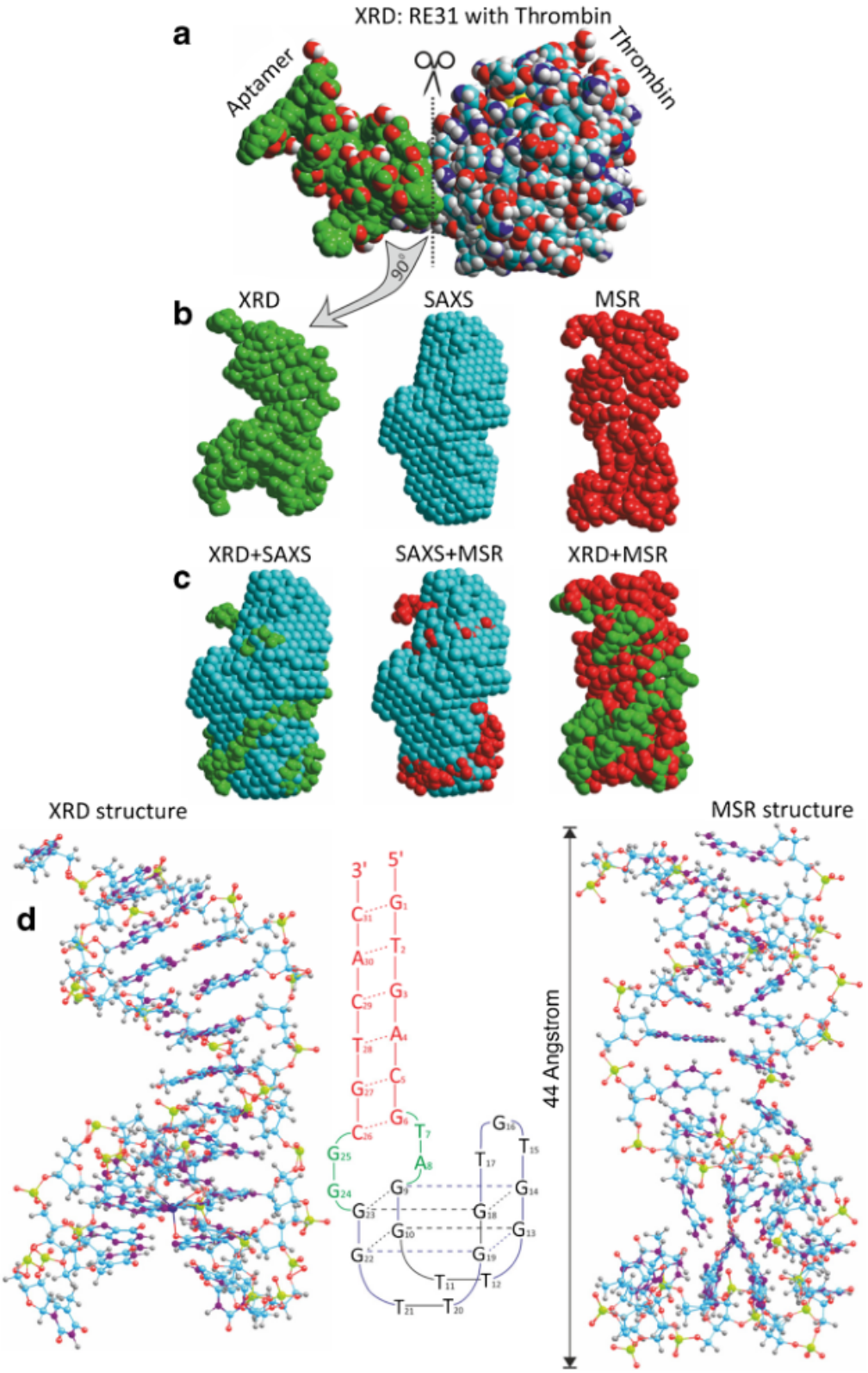

Fig. 2 Comparing RE31 aptamer structures from different methods. (A) RE31–thrombin complex by XRD. (B) Aptamer structures obtained by XRD in green color, SAXS in turquoise color, MSR in red color. (C) Comparison of the structures obtained by different methods. (D) RE31 structures from XRD analysis and MSR. Between them there is RE31 secondary structure divided into the three colored parts; red, a B-form duplex; green, a non-complementary region; blue, a G-quadruplex.

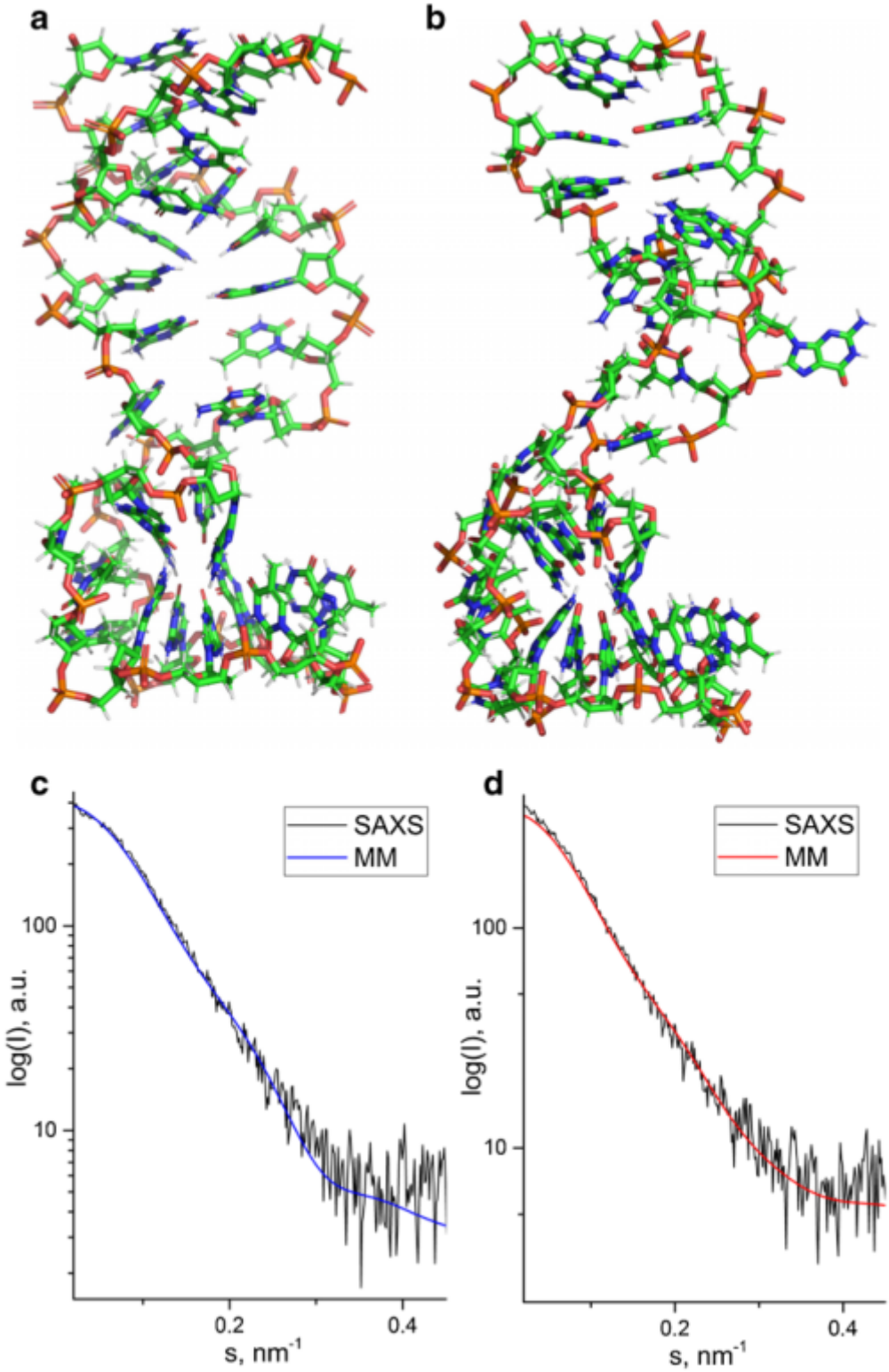

Fig. 3 Reconstructed atomic structures and SAXS profiles of the RE31 aptamer at different temperatures. Atomic structures of RE31 at 25 °C (a) and 40 °C (b). Experimental SAXS and MSR simulated scattering curves at 25 °C (c) and 40 °C (d).

An atomic model in the PDB format can be converted to the SAXS profile and compared with the experimental curve using the CRYSOL program [21].The program uses spherical harmonic approximation to calculate the scattering pattern from atomic structure factors and Gaussian Sphere Approximation for each atom's position in the molecule considering the hydration layer enveloping it. Figure 5 represents the correspondence between the measured SAXS profile and the scattering profiles

from XRD and MSR. A discrepancy χ 2 between experimental SAXS curve and evaluated one from the XRD model by the program CRYSOL is 5.851; χ^2 between SAXS and the MSR model is 5.279.

Conclusions

The purpose of this work is 3D molecular structure restoration of the aptamer from SAXS experiments obtained in solution. The SAXS method gives the shape of a molecule, but not the atomic structure of the aptamer. To solve this problem we used Avogadro software. The building of a molecule uses bond lengths and angles between atoms. In the first approximation, it is reasonable tousethemethods of molecular dynamics and/or semi-empirical methods of quantum chemistry, because they allow one to obtain the geometry of the molecule in a reasonable time. Therefore, the PM6 method was used as an auxiliary method for quickly obtaining the bond lengths and angles inside the aptamer molecule. Consequently, neither pH nor other solvent parameters were taken into account (because our purpose is to get a theoretical 3D model according to SAXS experiments).

The SAXS-based approach finds 3D structure of aptamers in solution without the crystallization required for X-ray diffraction, which may not provide reliable information about the structure in solution of such flexible molecules as nucleic acids. The success of the proposed approach is highly dependent on the knowledge about primary DNA structure. In general, the work described here creates the starting structure for an actual structure determination using molecular mechanics, but without the energy minimization or water molecules. It may further be applied for finding 3D structures of aptamers, DNAzymes, and ribozymes, and could supply new opportunities for developing functional nucleic acids.

Future prospects

It is usually impossible to uniquely construct a three-dimensional spatial model of a molecule from the primary DNA sequence. It is necessary to know at least its form, for which it is already possible to select a single location of nucleotides and to clarify the tertiary structure of the aptamer. In my work I would like to apply these computational methods, especially the MD simulations and Q-MD simulations to model our DNA aptamer biomolecules in solution. One of the complex tasks we faced is to model the oligomer state of the DNA aptamers constructing a G-quadruplex domain from several strands of nucleic acid (Fig. 4) [15]. I hope these methods will expand my contribution to the development of new theoretical and experimental methods.

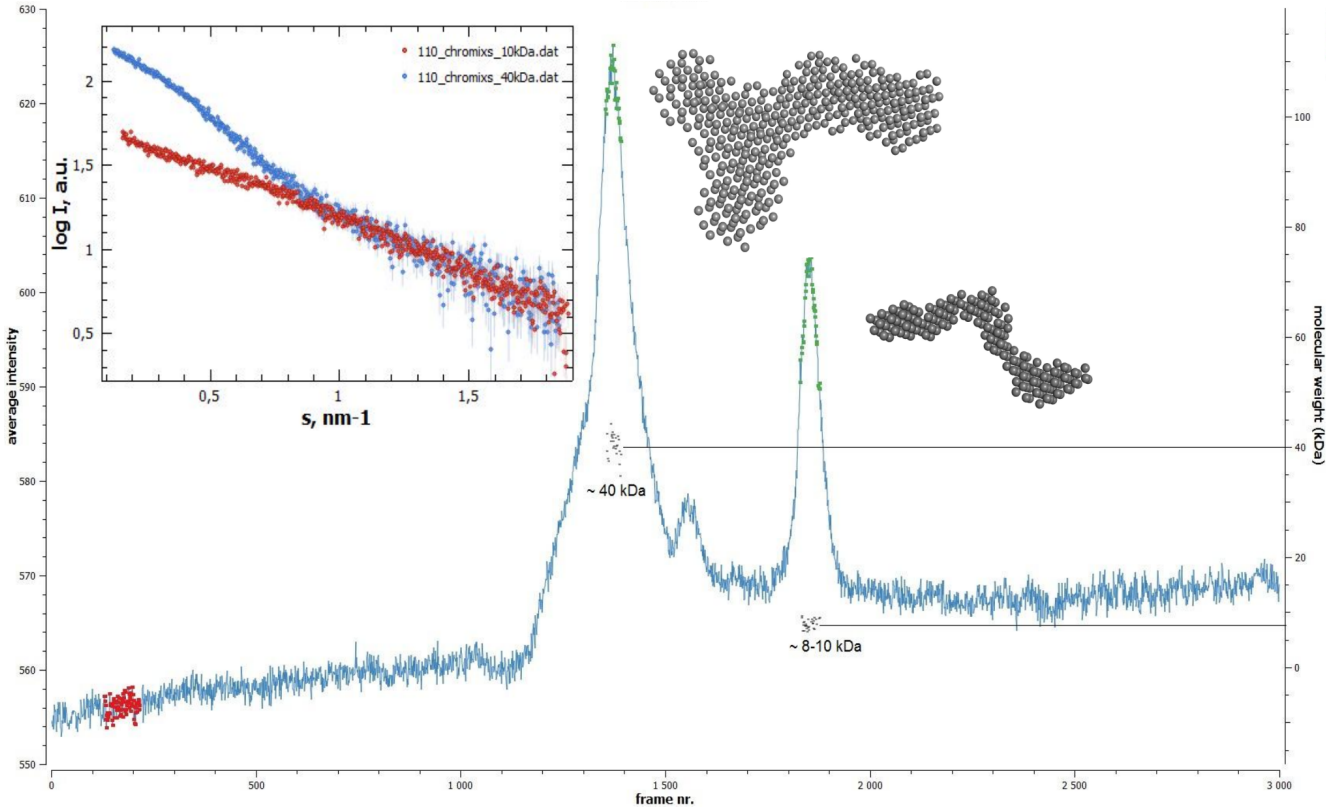

Fig. 4 SEC-SAXS - Size-Exclusion Chromatography + SAXS data for the DNA aptamer 110 to interleukin-6.

Acknowledgements

I would like to thank the JINR Program INTEREST for allowing me to participate in that science project. I express my gratitude to my supervisor Prof. Kholmirzo Kholmurodov for this course, introducing the molecular dynamics approaches, providing information in this field and productive communication.

References

1. Meyer C., Hahn U., Rentmeister A. Cell-Specific Aptamers as Emerging Therapeutics // Journal of Nucleic Acids. – 2011. – Vol. 904750. – P. 1-18.
2. Patel D.J., Suri A.K. Structure, recognition and discrimination in RNA aptamer complexes with cofactors, amino acids, drugs and aminoglycoside antibiotics // J. Biotechnol. – 2000. – Vol. 74. – P. 39-60.
3. Hammel M. Validation of macromolecular flexibility in solution by small-angle x-ray scattering (SAXS). Eur Biophys J. 2012;41(10): 789–99.
4. Rambo RP, Tainer JA. Super-resolution in solution x-ray scattering and its applications to structural systems biology. Annu Rev Biophys. 2013;42:415–41.
5. Jeffries CM, Graewert MA, Blanchet CE, Langley DB, Whitten AE, Svergun DI. Preparing monodisperse macromolecular samples for successful biological small-angle x-ray and neutron-scattering experiments. Nat Protoc. 2016;11(11):2122–53.
6. Guinier A. L'esprit de la recherche aux U. S. A. Atomes. 1947;2(20):378–82.
7. Timasheff SN, Witz J, Luzzati V. The structure of high molecular weight ribonucleic acid in solution. A smallangle x-ray scattering study. Biophys J. 1961;1:525–37.

8. Hollingsworth, S. A., and Dror, R. O. (2018) Molecular dynamics simulation for all. Neuron 99 (6), 1129-1143
9. 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).
10. 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).
11. Kholmirzo KHOLMURODOV (Editor), «Models in Bioscience and Materials Research: Molecular Dynamics and Related Techniques», Nova Science Publishers (N.Y.), ISBN: 978-1-62808-052-0 (2013).
12. Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J Cheminform. 2012;4(1):17.
13. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 2003;31(13):3406–15.
14. Ikebukuro K, Okumura Y, Sumikura K, Karube I. A novel method of screening thrombin-inhibiting DNA aptamers using an evolution-mimicking algorithm. Nucleic Acids Res. 2005;33(12):e108.
15. Macaya RF, WaldronJA, Beutel BA, Gao H, Joesten ME, Yang M, et al. Structural and functional characterization of potent antithrombotic oligonucleotides possessing both quadruplex and duplex motifs. Biochemistry. 1995;34(13):4478–92.
16. Padmanabhan K, Tulinsky A. An ambiguous structure of a DNA 15-mer thrombin complex. Acta Crystallogr D Biol Crystallogr. 1996;52(Pt 2):272–82
17. RussoKraussI,MerlinoA,GiancolaC,RandazzoA,MazzarellaL, Sica F. Thrombin-aptamer recognition: a revealed ambiguity. Nucleic Acids Res. 2011;39(17):7858–67.
18. Russo Krauss I, Merlino A, Randazzo A, Novellino E, Mazzarella L, Sica F. High-resolution structures of two complexes between thrombin and thrombin-binding aptamer shed light on the role of cations in the aptamer inhibitory activity. Nucleic Acids Res. 2012;40(16):8119–28
19. Werner A, Konarev PV, Svergun DI, Hahn U. Characterization of a fluorophore binding RNA aptamer by fluorescence correlation spectroscopy and small angle x-ray scattering. Anal Biochem. 2009;389(1):52–62.
20. Spiridonova VA, Barinova KV, Glinkina KA, Melnichuk AV, Gainutdynov AA, Safenkova IV, et al. A family of DNA aptamers with varied duplex region length that forms complexes with thrombin and prothrombin.FEBSLett.2015;589(16):2043–9.
21. Svergun D, Barberato C, Koch MHJ. CRYSOL - a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. J Appl Crystallogr. 1995;28(6):768–73.