Molecular dynamic simulations as a verification of the three-dimensional structure of synthesized molecule

Miljan Bigović
Faculty of natural sciences
University of Montenegro
Podgorica, Montenegro
miljan@ac.me

Žarko Zečević
Faculty of electrical engineering
University of Montenegro
Podgorica, Montenegro
zarkoz@ac.me

Luka Filipović
Center of information system
University of Montenegro
Podgorica, Montenegro
lukaf@ac.me

Božo Krstajić
Faculty of electrical engineering
University of Montenegro
Podgorica, Montenegro
bozok@ac.me

Abstract - **Paper presents comparison between molecular models made using software packages and experimental results and performed simulations in molecular dynamics area with created molecule. Molecule was synthetized using organic synthesis strategy, performed structural analysis and determined precise layout of all the atoms in space. It is converted into standard chemical format, visualized and prepared for simulations. Study of the interaction of the selected molecule with receptor proteins proved that the arrangement of all the atoms is identical with the obtained experimental values. With our work, we intend to show that it is possible, by using powerful computer infrastructure and related applications, to create good models of molecules, which can be used in simulations that give the result valid characteristics of subject molecules interacting with various organic compounds.**

Keywords - molecule modeling, molecular dynamics, organic molecule, protein, parallel computing simulations

I. INTRODUCTION (*HEADING 1*)

Distributed computing simulations in molecular dynamics area are very popular and used in last decade [1]. They are based on research of various types of non-covalent intermolecular interactions (Van der Waals interactions, hydrogen bonds, creating ion bridges) and covalent interactions between small organic molecules synthesized in the laboratory and the receptor sites in the body (cell membrane proteins as a target-centers the effect of drugs and biologically active chemical substances). These interactions are foundation for modern chemistry, pharmacology, pharmaceutical chemistry and biochemistry. However, mentioned interactions are mostly studied on isolated tissues and cell cultures (*in vitro*) or in experimental animals and volunteers (*in vivo*). Existence of computing model which can simulate and predict molecule interactions with well-known proteins [2] will gave enormous contribution in resolving this complex problem. Computer simulations can reduce costs of development of new cures, which takes 12-15 years and costs between 600 and 800 million US dollars and accelerate process of compounds selection that are considered as part of new medicine [3]. Mixture of molecular dynamics and distributed computing is very relevant in modern research of potential cures. A variety of publications testify about growth and successful research studies in various fields of life science [6-9].

In this paper, we focused on test of interactions between our synthetized molecule and proteins of the cell membrane. Our intention was to show possibility of creation of good molecule models and simulations using powerful computing infrastructure. Whole simulation process can be accelerated using dozens or hundreds of computing cores, which gives a significant advantage of using of computer models in comparison to the traditional way of synthesis of new molecules, and subsequent analysis of their biological activity. In our simulations, we used NAMD (NAnoscale Molecular Dynamics, [4]) and GROMACS (GROningen MAchine for Chemical Simulations, [5]), applications for parallel simulations in the molecular dynamics of the given ligand-receptor pair. Simulations in molecular dynamics area are very computing intensive and require large processing power.

This research was realized within H2020 project „Virtual Research Environment (VRE) in Southeast Europe and the Eastern Mediterranean (SEEM) – VI-SEEM" [10], which gathers partners form 16 countries. Research institutions shares computing resources (more than 300000 CPU cores and 11 PB of storage) for simulations in life sciences, cultural heritage, climate modeling and weather forecasting.

Computing resources and services [16], expertise and knowledge sharing form project partners dealing with similar issues were valuable for research presented in this paper.

II. PREPARATION OF MOLECULE MODEL

We synthetized 17 new and previously unknown molecules, based on the existence of enol carbonate structural units [11]. Specified unit is a common structural motive of many natural products and physiologically active compounds [12,13], which makes molecules suitable for further analysis. Synthesis of enol carbonates has been performed been performed by allylation of aldehydes under very mild reaction conditions in the aqueous medium and with very good yields the expected product, Figure 1.
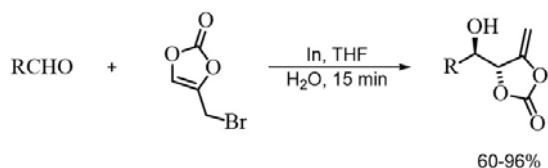

Fig. 1. Reaction of obtaining Enol carbonate

Complete spectroscopic documentation for each synthetized compound is available [11]. The presence of fragment with three atoms of oxygen is certainly interesting, especially in drug design industry, because it provides possibilities for further functionalization of derivatives and in organic synthesis [14]. One of synthetized compounds was (2R,3R,4R,5S)-5-hydroxy-5-((S)-5-methylene-2-oxo-1,3-dioxolan-4-yl) pentane-1,2,3,4-tetraacetate, which structure is shown on Figure 2 and which was basis for further molecule visualization and simulations.
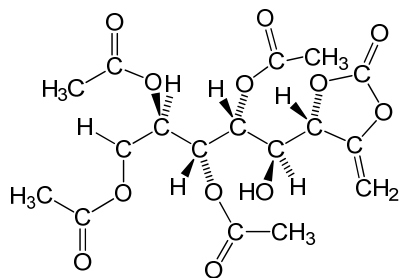

Fig. 2. Structural formula of Enol carbonate ((2R,3R,4R,5S)-5-hydroxy-5-((S)-5-methylene-2-oxo-1,3-dioxolan-4-yl) pentane-1,2,3,4-tetraacetate)

This compound was synthesized according to the standard experimental procedures and was obtained as white needle crystals. The crystals were carefully purified by the method of recrystallization from a mixture of organic solvents in order to obtain a representative monocrystal. This high quality monocrystal was used for X-ray structural analysis, which belongs to a powerful, accurate and representative technique of analysis of organic molecules. As a result, diffraction analysis provides the exact spatial model of atoms and bonds in a given molecule (Figure 3), which was necessary for building of model of compound for further computer simulations.
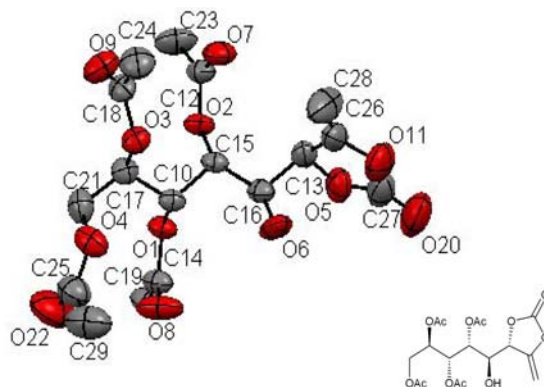

Fig. 3. ORTEP diagram of enol carbonate compound

Molecule modeling and preparation for computer simulation can be performed using VegaZZ [17], ChemSketch [18] and OpenBabel [15]. Enol carbonate, (2R, 3R, 4R, 5S) - 5-hydroxy-5-((S)-5-methylene-2-oxo-1,3-dioxolan-4-yl) pentane-1,2,3,4-tetraacetate was first sketched using a software package ChemSketch, which is routinely used to draw the structural formula of organic and inorganic molecules. It has the ability to save drawn molecules in their standard formats such as MDL molfile, which contains all the information about the atoms, bonds and their types, as well as the coordinates of the atoms in the molecule. Figure 4 shows the structure of the observed molecules in ChemSketch, while Figure 5 presents data files molecules in the program VegaZZ, preserved in MDL molfile format.
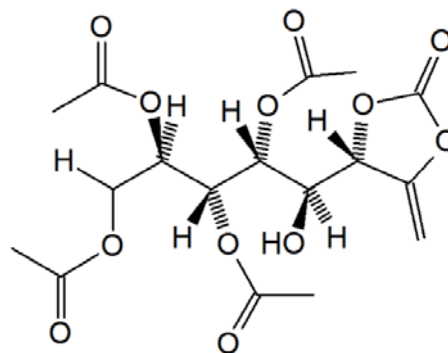

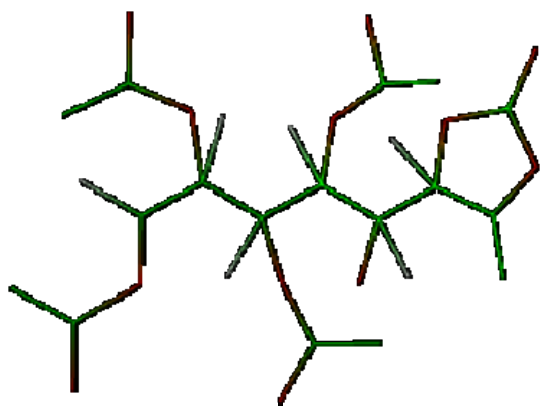Fig. 4. Model of enol carbonate from ChemSketch

Fig. 5. Enol carbonate - VegaZZ (MOL format)

Figure 5 presents molecule model in planar structure and it's necessary to optimize their geometry in order to prepare the molecule for further simulations with the appropriate protein. This operation is possible using the open source program OpenBabel. In addition to the geometrical optimization, OpenBabel can convert a molecule through standard chemical formats. Figure 6 shows the observed molecules after geometrical optimization and converted into a standard format PDB (Protein Data Bank) which is often used in public compound databases [2].
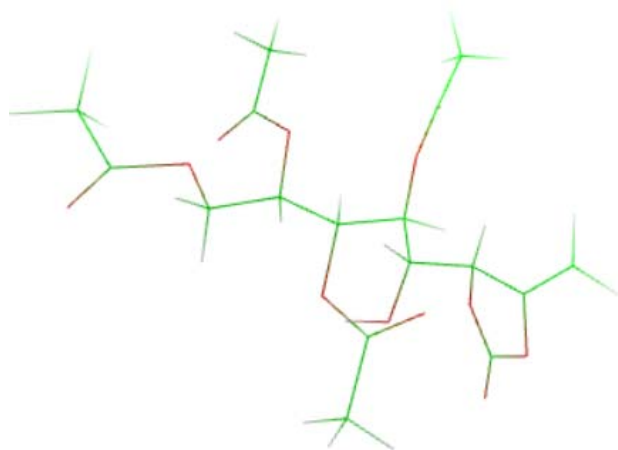

Fig. 6. Preview of enol carbonate from VegaZZ (PDB format)

Model of our molecule on Figure 6 proves identical structure of molecule can be generated using software tools as the structure of molecule generated in the laboratory. We conclude that model has correct structure and it can be used as input file for molecule dynamics computer simulations, same as molecule simulations in real world. Main difference between mentioned simulations is their duration – computer simulations are much faster when it uses many computer cores and can analyze, in shorter time interval, many different combinations between molecules and proteins.

III. MD SIMULATIONS WITH PROTEIN

System containing a protein (T4 lysozyme L99A/M102Q) in complex with a ligand was simulated. All simulations were performed by Gromacs using 24 computing cores. To generate small-ligand topology for use with GROMACS family of force fields we used PRODRG server [19]. Topologies produced by this program have been used widely in studies of protein-ligand interactions.

Energy minimization should be performed to be ensured that the system has no steric clashes or inappropriate geometry. Figure 7 shows the convergence of potential energy. It can be determined that potential energy converges in 350 iterations.
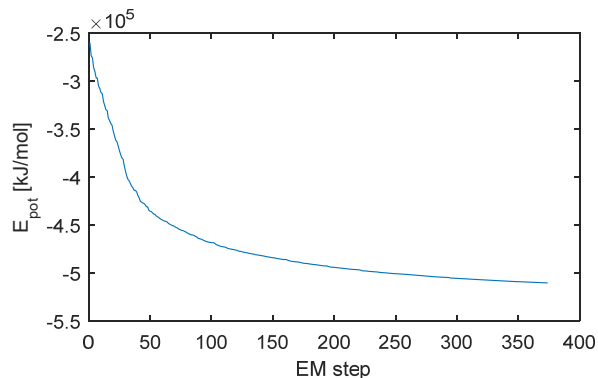

Fig. 7. Energy minimization phase

Energy minimization ensures that the starting structure is reasonable in terms of geometry and solvent orientation. Before beginning real dynamics, the solvent and ions around the protein must be equilibrated. Equilibration is performed in two phases. The first phase is conducted under "isothermal-isochoric" conditions (constant Number of particles, Volume, and Temperature, NVT). The timeframe for this simulation was set to 100ps. Figure 8 shows temperature of the system after equilibration. Temperature quickly reaches the target value of 300K and remains stable over the remainder of time.
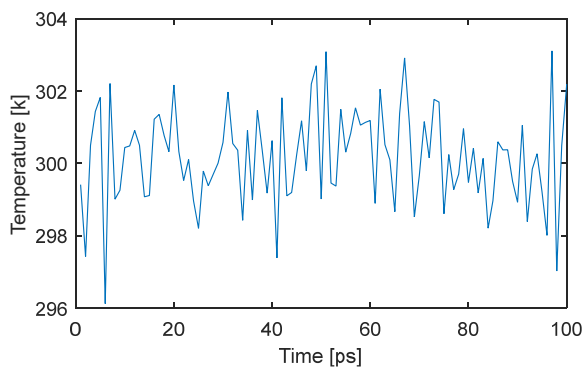

Fig. 8. Temperature of the system

In the second phase of equilibration, the pressure of the system should be stabilized. This phase is performed under "isothermal-isochoric" conditions (constant Number of particles, Pressure, and Temperature, NPT). For this simulation, the timeframe also has set to 100ps. Figures 9-11 shows pressure of the system varies over the simulated timeframe, but this behavior is not unexpected.
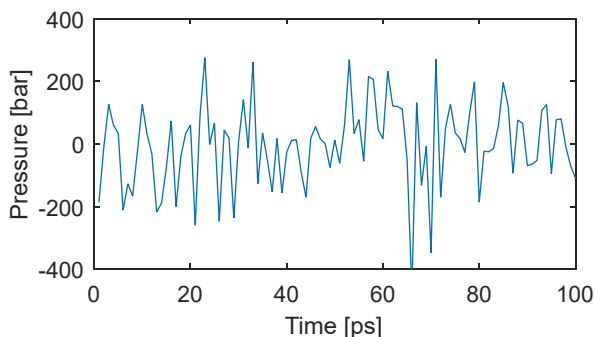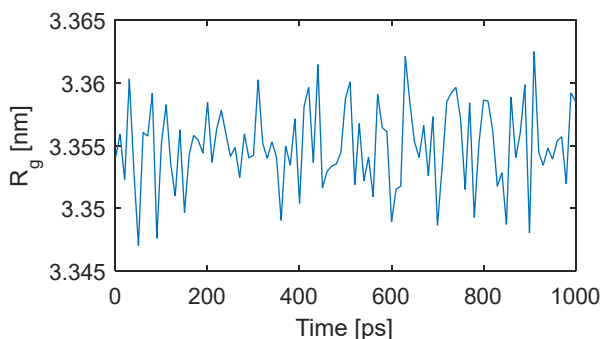

Fig. 9. Pressure of the system
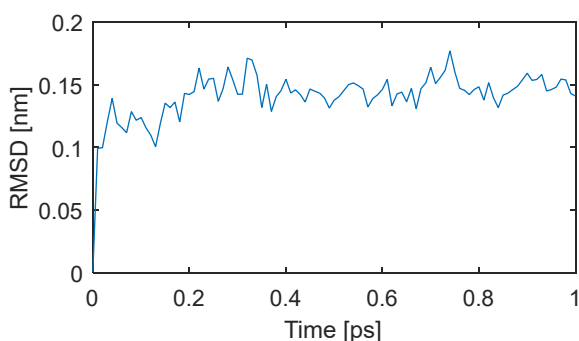

Fig. 10. Pressure of the system

Fig. 11. Pressure of the system

After equilibrating the system at the desired temperature and pressure, production MD phase is carried. This simulation phase is computationally the most demanding, so high performance computers are recommended here. In this example, 1ns simulation is performed. GROMACS has some built in tools for MD analysis. The radius of gyration (Rg) of

a protein is a measure of its compactness. From Figure 10 it can be observed that Rg has reasonably invariant values, which means that the protein remains in its folded form over the timeframe of 1ns at 300 K. Another indicator of the system compactness is RMSD. The Figure 11 shows that the RMSD oscillates around 0.15 nm, indicating that the structure is very stable.
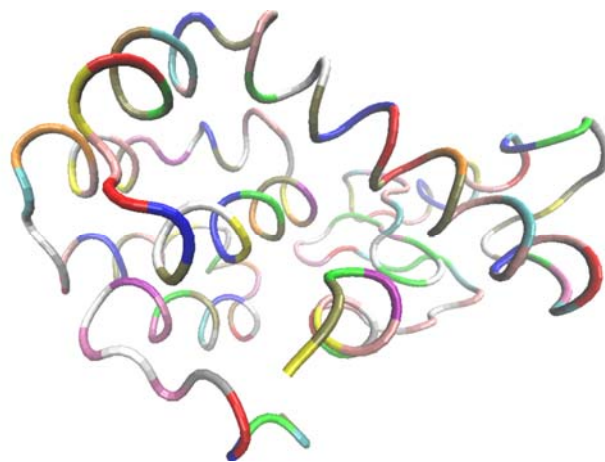

Fig. 12. The most stable possible conformations of proteins, presented in VMD

Figure 12 shows the most stable and thermodynamic most favorable conformation of the tested proteins.

Computer-intensive calculations were performed on 24 core FINKI HPC. By using parallel data processing, simulations are noticeably speeded up, where the efficiency was more than 92%. We received results quicker and had possibility to try many combinations before recommendations for reaction testing in the laboratory.

IV. CONCLUSION

The existence of a computer model, that could predict the presence of non-covalent and covalent interaction with the well-known and well-defined structural proteins, would give a huge contribution from at least two reasons:

1. We should study the interaction of organic molecules, which previously did not synthesized in the laboratory (which significantly reduces the time and research, but also the price of the process), with appropriate protein (i.e. receptor);

2. It would be available additional data on possible new structural units of the molecule, which would be responsible for additional or possibly even more stronger interactions with receptors (proteins) of cell membrane. In that manner, the same pharmacological properties of the test molecule would be much improved.

Further, using the knowledge and tools of Organic synthesis, these and similar molecules could be synthesized in the laboratory in an amount that would be sufficient for pre-clinical and clinical tests. Understanding these interactions and their detailed description and creation of molecules that are

structurally similar to investigated represents the future of this field computational chemistry and molecular modeling.

REFERENCES

[1] J. Donngara, I. Foster, G. Fox, W. Gropp, K. Kennedy, L. Torczon and A. White, "Sourcebook of parallel computing", Morgan Kaufman publishers, ISBN 1-55860-871-0

[2] RCSB Protein Data Bank: http://www.rcsb.org/

[3] Ž. Čeković i saradnici "*Molekuli u tajnama života i svetu oko nas*", 1. Izdanje, Zavod za udžbenike, Beograd, 2009, strane 139-149

[4] M. Nelson, W. Humphrey, A. Gursoy, A. Dalke, L. Kalhe, R. D. Skeel, K. Schulten, ,,A Parallel, Object-Oriented Molecular Dynamics Program", Theoretical Biophysics Group University of Illinois and Beckman Institute 405 North Matthews Urbana, IL 61801 June 10, 1996.

[5] *D.* Van Der Spoel, E. Lindahl,B. Hess, G. Groenhof, A.E. Mark, H.J. Berendsen*.* *"GROMACS: fast, flexible, and free". J Comput Chem. Vol.*16, 2005., 1701–18.

[6] E. Athanasiadis, Z. Cournia, G. Spyrou, ,,ChemBioServer: a web-based pipeline for filtering, clustering and visualization of chemical compounds used in drug discovery", .*Bioinformatics* Vol. 28, No. 22, November 2012, pp. 3002-3003.

[7] E. Lionta, G. Spyrou, D. K. Vassilatis, Z. Cournia, ,, Structure-based virtual screening for drug discovery: principles, applications and recent advances. " , *Curr. Top. Med. Chem.*, Vol.14, No. 16, Oct. 2014, pp. 1923-1938.

[8] M. Mansha, U. U. Kumari, Z. Cournia, N. Ullah, ,, Pyrazole-based potent inhibitors of GGT1: Synthesis, biological evaluation, and molecular docking studies.", *Eur. J. Med. Chem.*, Vol. 124, Nov. 2016, pp. 666-676.

[9] Y. Wang, P. Gkeka, J. E. Fuchs, K. R. Liedl, Z. Cournia, ,, DPPC-cholesterol phase diagram using coarse-grained Molecular Dynamics simulations.", *Biochim. Biophys. Acta*, Vol. 11, Nov. 2016, pp. 2846-2857.

[10] H2020 project VI-SEEM, https://vi-seem.eu

[11] M. Bigovic, V. Maslak, Z. Tokic-Vujosevic, V. Divjakovic, R. N..Saicic, ,,A useful synthetic equivalent of a hydroxyacetone enolate", *Organic Letters,* vol. 13, July 2011, pp. 4720-4723.

[12] S.S.V. Ramasastry, H. Zhang, F. Tanaka, C.F. Barbas, ,,Direct Catalytic Asymmetric Synthesis of anti-1,2-Amino Alcohols and syn-1,2-Diols through Organocatalytic anti-Mannich and syn-Aldol Reactions", *J. Am. Chem. Soc*. Vol. 129, Jul. 2007, pp. 288–289.

[13] M. M. Market, M. Mulzer, B. Schetter, R. Mahrwald, ,,Amine-Catalyzed Direct Aldol Addition", *J. Am. Chem. Soc*. Vol. 129, Nov. 2007, pp. 7258–7259. 3) J. Li, S. Luo, J. P. Cheng, ,, Chiral Primary−Tertiary Diamine Catalysts Derived From Natural Amino Acids for syn-Aldol Reactions of Hydroxy Ketones", *J. Org. Chem.* Vol. 74, Avg. 2009, pp. 1747-1750.

[14] M. Bigovic, S. Skaro, V. Maslak, R. N. Saicic, ,,Expanding the scope of the indium-promoted allylation reaction: 4-(bromomethyl)-1.3-dioxol-2-one as a synthetic equivalent of a 3-arylhydroxyacetone enolate, *Tetrahedron Letters,* No. 54, September 2013, pp. 6624-6626.

[15] Open Babel: The Open Source Chemistry Toolbox, http://openbabel.org/wiki/Main_Page.

[16] E. Atanassov, A. Karaivanova, T. Gurov, "Services And Infrastructure For Virtual Research Environments - For Use By Science And Business", International Scientific Journal Industry 4.0, Issue 2, 2016, pp. 110-113

[17] A. Pedretti, L. Villa, G. Vistoli, "VEGA - an open platform to develop chemo-bio-informatics applications, using plug-in architecture and script programming", J.C.A.M.D., Vol. 18, 167-173 (2004)

[18] ACD/ChemSketch for Academic and Personal Use :: ACD/Labs.com. [online] Available at: http://www.acdlabs.com/resources/freeware/chemsketch/

[19] A. W. Schüttelkopf and D. M. F. van Aalten (2004). "PRODRG: a tool for high-throughput crystallography of protein-ligand complexes", Acta Crystallogr D60, 1355–1363.