



MODELING AND MOLECULAR DYNAMICS SIMULATIONS STUDY OF ENOL-CARBONATES AND THEIR DERIVATIVES *

MILJAN BIGOVIC † LUKA FILIPOVIC ‡ ZARKO ZECEVIC § AND BOZO KRSTAJIC ¶

Abstract. In view of the fact that the mechanisms of the interaction of organic molecules with the properties of the drug occur in most cases by their binding to the receptors (proteins), we wanted to examine the interactions of small organic molecules that we synthesized and certain proteins. In this paper the comparison between molecules, modeled by different software packages, and experimental results is performed. The series of new molecules are synthesized using Barbier reaction of allylation of carbonyl compounds with 4-(bromomethyl)-1,3-dioxol-2-one as a highly fictionalized allylic synthon. Their structure was determined by spectroscopic methods (NMR-, IR- and UV/VIS-spectroscopy and MS-spectrometry), while for three individual molecules analysis of x-ray diffraction (X-ray analysis) was also performed, which gave the final confirmation of the exact arrangement of all atoms in the space. The molecules are represented in standard chemical format, visualized and prepared for simulations. In order to obtain datasets that further can be used to examine and analyze interactions with well-known proteins, molecular dynamic simulations are performed. The purpose of this research was to present that using powerful computer infrastructure and appropriate software tools, an accurate molecule models can be created.

Key words: molecule modeling, molecular dynamics simulations, organic molecules, protein, interaction, parallel simulation

AMS subject classifications. 65Y05, 92C40, 92E10

1. Introduction. Synthesis of biologically active organic molecules in the laboratory is one of the most important methods in the procedure of the creation of new drugs today. In this way, the possibility of synthesizing molecules that are available from natural sources in very small quantities is opened up. Moreover, the synthesis of molecules that are not present in nature at all, but that could have far stronger and better biological effects in relation to natural ones has become possible. All methods of chemical synthesis used in the process of obtaining such molecules are of great importance.

In order to speedup the research of synthetic molecules and their interactions, as well as reduce analysis costs, molecular dynamics (MD) simulations on distributed computing resources are used. They consist of many emerging techniques with potential applications in diverse areas of modern chemistry, pharmacology, pharmaceutical chemistry and biochemistry. Over the past three decades, MD has evolved as an area of importance for understanding the atomic basis of complex phenomena such as molecular recognition, protein folding, and the transport of ions and small molecules across membranes. The application of MD simulations with experimental approaches have provided an increased understanding of protein structure-function relationships and demonstrated capacity in pharmaceutical and medical analysis and drug discovery [1]. Existence of computing model that can simulate and predict molecule interactions with well-known proteins [2] will give enormous contribution in resolving this complex problem. The discovery of a new drug takes 12-15 years, and costs between 600 and 800 million US dollars. The application of these methods will contribute to the selection of candidate molecules that could be biologically active and thus reduce the list of tested molecules. This would certainly contribute to reducing the time spent in the laboratory and the financial expenditures [3]. A variety of publications testifies about growth and successful MD research studies in various fields of life science [4][5][6][7].

Here, models of three newly synthesized molecules are presented. In order to obtain datasets that further can be used to examine and analyze interactions with well-known proteins, molecular dynamic simulations are performed. Since the MD simulations are computationally intensive, they are significantly accelerated by using dozens or hundreds of computing cores, which gave a significant benefit of using of computer models in comparison to the traditional way of synthesis of new molecules and their biological activity. In our analysis we used GROMACS [8], a software for parallel simulations in the molecular dynamics of the given ligand-receptor

*The research leading to these results has been co-funded by the European Commission under the H2020 Research Infrastructures contract no. 675121 (project VI-SEEM).

†Faculty of natural sciences, University of Montenegro, Podgorica, Montenegro (miljan@ac.me).

‡Center of information system, University of Montenegro, Podgorica, Montenegro (lukaf@ac.me).

§Faculty of electrical engineering, University of Montenegro, Podgorica, Montenegro (zarkoz@ac.me).

¶Faculty of electrical engineering, University of Montenegro, Podgorica, Montenegro (bozok@ac.me).

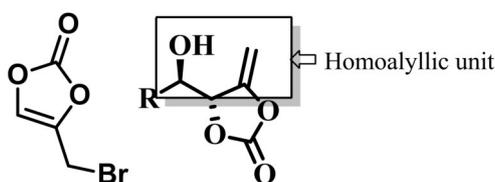


FIG. 2.1. 4-(Bromomethyl)-1,3-dioxol-2-one

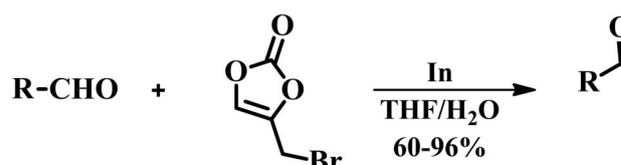


FIG. 2.2. General view of allylation reaction of carbonyl compounds (obtaining the enol-carbonates)

pair.

2. Enol Carbonates and their Derivatives as Simulation Models. The homoallylic alcoholic unit is a very interesting and very common structural unit that characterizes a large number of physiologically active molecules. In this way, important precursors are obtained, which are further used in the synthesis of a numbered natural products with pronounced antibiotic, antimicrobial, antifungal or even cytostatic effects. Some of them are: methylenolactacin [9], myxothiazol A [10], eupomatilone [11], sialic acid [12] and similarly.

We have previously reported the use of 4-(bromomethyl)-1,3-dioxol-2-one [Fig 2.1] as a highly functionalized allylic synthon [13]. We assumed that in the reaction of this reagent with aldehydes or ketones it would provide the possibility of obtaining homoallylic alcohols, obtained in the form of protected enol carbonates:

Indium mediated allylation of carbonyl compounds with this reagent in water as a solvent allows for the synthesis of α, β -hydroxy ketones, in protected or either free form. 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 [Fig 2.2].

After several decades of development, the preparation and application of organoindium reagents in organic synthesis have seen leaps and jumps. The feasibility of using these reagents in aqueous media permits the direct functionalization of water-soluble substrates, which is important both from the ecological and the economic point of view.

With obtained enol-carbonates, we have examined two sets of reactions: in first, we transferred them to free keto-diols using mercury salts, and in the second, after treatment with base, they were converted enol-carbonates into saturated cyclic carbonates¹ [Fig 2.3]:

Deprotection of enol-carbonates leads to free α, β -dihydroxyketone - a unit which is a common structural motive of many natural products and physiologically active compounds which recommended molecules for further analysis. The third direction of the development of our reaction was the creation of a new carbon-carbon bond. Namely, enol-carbonates undergo Heck reaction with aryl iodides in the presence of silver trifluoroacetate, to give the corresponding arylated products, obtained in moderate yields [14] [Fig 2.4].

¹Protocol: To a 10 mL flask equipped with a glass stopper and a magnetic stirrer, 4-(bromomethyl)-1,3-dioxol-2-one 300 (50.0 mg, 0.28 mmol), THF (0.5 mL) and water (1.0 mL), and then indium powder (32.1 mg, 0.28 mmol). The resulting suspension was intensively stirred at room temperature, whereby a white blur occurred. A carbonyl compound (0.19 mmol) was then added to the suspension. The reaction mixture was stirred for 15 minutes at room temperature, and the reaction stream was followed by thin layer chromatography (eluent: 20% acetone in petroleum ether). The reaction mixture was partitioned between dichloromethane (5 mL) and water (5 mL); the aqueous extracts were washed twice with 5 mL of dichloromethane. The combined organic extracts were dried over anhydrous magnesium sulphate, evaporated and evaporated on a rotary vacuum evaporator. The crude product was purified by column-based column chromatography, and wherever possible by subsequent recrystallization.

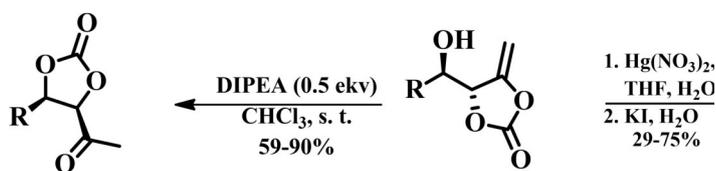


FIG. 2.3. Functional transformations of enol-carbonates

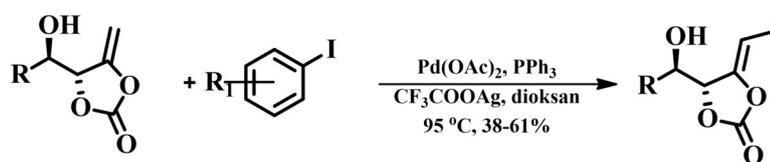


FIG. 2.4. Heck reaction with enol carbonates- method for carbon-carbon bond formation

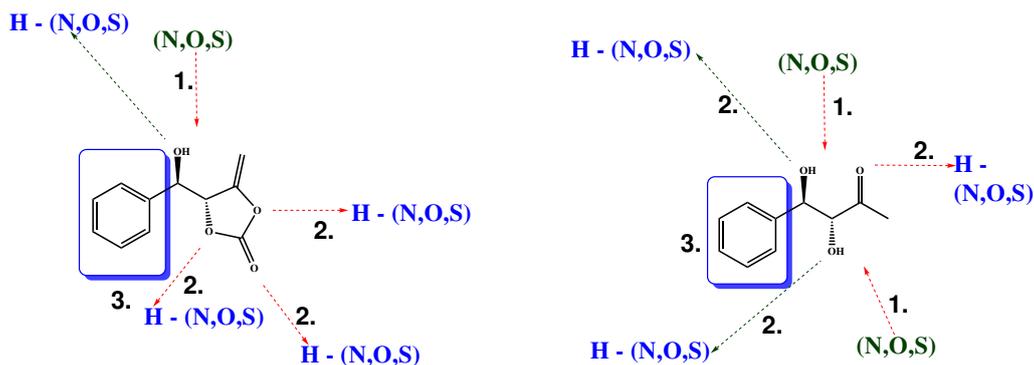


FIG. 2.5. Potential (predicted) interactions between molecules and active sites of protein

In view of the fact that all these molecules were newly synthesized and that no biological test was done with them, we thought that molecular-dynamic calculations and simulations could provide some important data on the interactions of functional groups that our molecule are having with active sites of proteins and receptors, which are, in fact, lateral strings of amino acids.

Based on theoretical knowledge, that according to the structural characteristics of the described molecules it should be assumed which regions could be involved in characteristic interactions with residues from active protein centers.

In the following examples of selected molecules from all of the listed reactions, we will show what possible interactions we expect, and which will be confirmed by molecular calculations [Fig 2.5].

In Fig 2.5, the meanings are the followings:

1. Forms hydrogen bonds with electronegative atoms from the active site of proteins (nitrogen, sulphur, oxygen).
2. Forms hydrogen bonds with hydrogen atoms with hydrogen atoms, which are in the active site of the protein, and are themselves bound directly to an electronegative atom (nitrogen, sulphur, oxygen).
3. π - π -non-polar interactions (part of the molecule that potentially builds these interactions is shown in a blue rectangle)

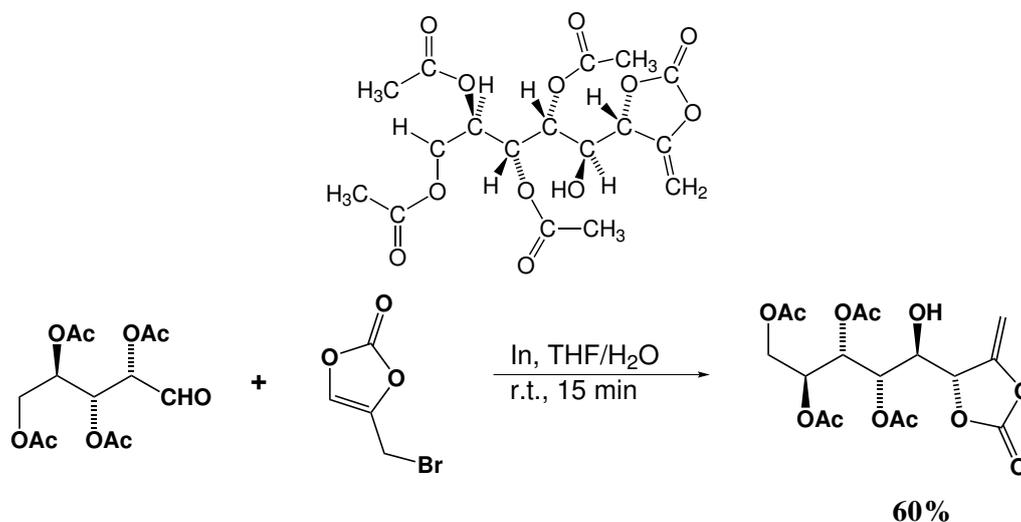


FIG. 3.1. Structural formula of Enol carbonate ((2R,3R,4R,5S)-5-hydroxy-5-((S)-5-methylene-2-oxo-1,3-dioxolan-4-yl)pentane-1,2,3,4-tetraacetate)

3. Selected Molecules for Molecular Simulations. The most interesting synthesized compounds for further analysis were:

1. (2R,3R,4R,5S)-5-hydroxy-5-((S)-5-methylene-2-oxo-1,3-dioxolan-4-yl)pentane-1,2,3,4-tetraacetate, see Fig 3.1
2. (R)-4-((R)-hydroxy(phenyl)methyl)-5-methylene-1,3-dioxolan-2-one, see Fig 3.2
3. (R)-4-((R)-(4-chlorophenyl)(hydroxy)methyl)-5-methylene-1,3-dioxolan-2-one, see Fig 3.3

Specified molecules were basis for further molecule modeling, visualization, simulations and comparison. Compounds were synthesized according to the standard experimental procedures and were 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 3.1-3.3), which was necessary for building of the compound model for further computer simulations.

Molecules are modeled and prepared for MD simulation by VMD [15], ChemSketch [16] and OpenBabel [17]. First, the molecules are sketched by using ChemSketch a molecular modeling program used to routinely draw and modify structural formula of organic and inorganic molecules. By ChemSketch, the molecules can be saved in standard formats, such as MDL molfile, which holds all information about the atoms, bonds and their types and connectivity, as well as the coordinates of the atoms in the molecule.

After geometrical optimization in OpenBabel, molecules were converted into a standard format PDB (Protein Data Bank) which is often used in public compound databases [2]. Figures 3.4-3.6 show the resulting molecules, visualized in VMD. Computer molecule models has identical structure with molecule generated in the laboratory and it can be used as an input file for molecular dynamics simulations. We conclude that there was no breakdown of links, deformation of angles or possible replacement of individual atoms. Verification of identical models was necessary as a key step for further simulations. All effects would lead to the formation of molecule with some different values of angles and distances between the atoms in relation to the one with which we want to examine the simulation.

Main difference between laboratory and computer analysis is their duration and cost. Computer simulations are much faster when it uses many computer cores and many different combinations between molecules and proteins can be analyzed in shorter time interval. Molecular simulations have great advantages in terms of synthesis and in vivo and in-vitro biological activity testing. Namely, time is shorter, there is no need expensive

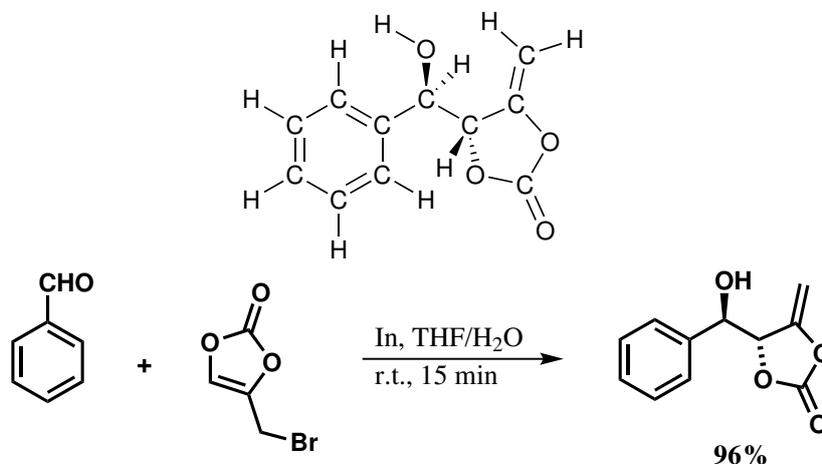


FIG. 3.2. *(R)*-4-*((R)*-hydroxy(phenyl)methyl)-5-methylene-1,3-dioxolan-2-one

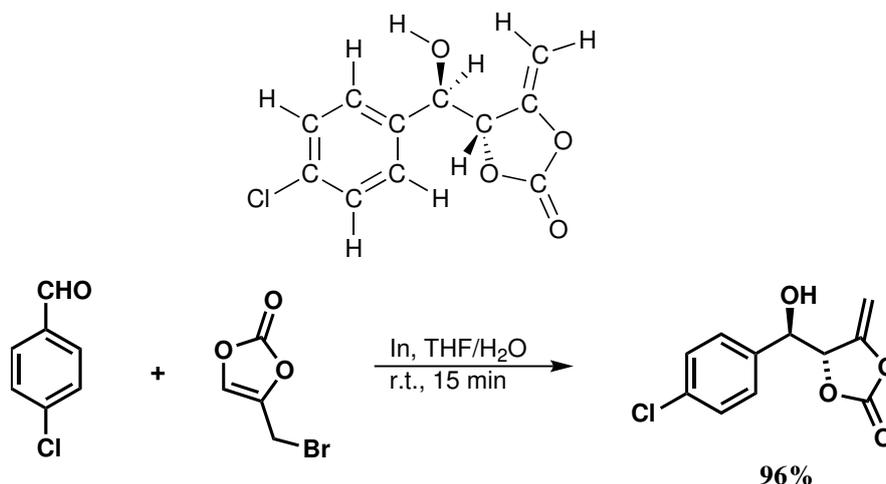


FIG. 3.3. *(R)*-4-*((R)*-(4-chlorophenyl)(hydroxy)methyl)-5-methylene-1,3-dioxolan-2-one

equipment and chemicals, and they have a very high accuracy.

4. Molecular Dynamics Simulations with Protein . We have simulated a system containing a protein (T4 lysozyme L99A/M102Q) in complex with presented ligands. The simulation results related to the second molecule are presented. The similar results are obtained for the other molecules. Gromacs in conjunction with 24 CPU cores is exploited for simulations. PRODRG server [18] is used to generate a small-ligand topology for use with GROMACS force fields family. Protein-ligand interactions are widely studied based on topologies produced by this program. Simulation are performed by following the procedures described in [19], where all necessary details about system configuration can be found.

Possible steric clashes or inappropriate geometry of the system are prevented by performing the energy minimization. The convergence of the potential energy is shown in Figure 4.1, indicating the convergence in 350 iterations.

Energy minimization ensures that the starting structure is reasonable in terms of geometry and solvent orientation. Before the real dynamics starts, the solvent and ions around the protein must be equilibrated, which is performed in two phases. The first phase is conducted under isothermal-isochoric conditions (constant

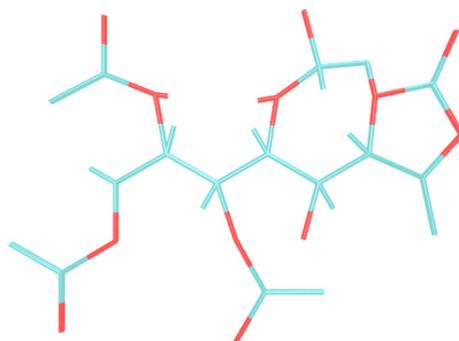


FIG. 3.4. Structural formula of Enol carbonate ((2R,3R,4R,5S)-5-hydroxy-5-((S)-5-methylene-2-oxo-1,3-dioxolan-4-yl)pentane-1,2,3,4-tetraacetate)

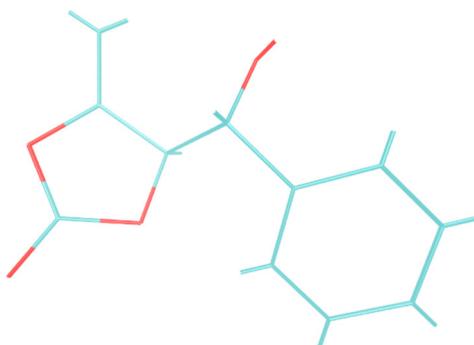


FIG. 3.5. (R)-4-((R)-hydroxy(phenyl)methyl)-5-methylene-1,3-dioxolan-2-one

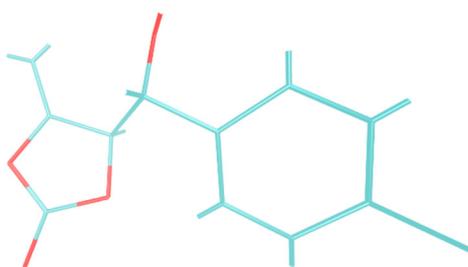
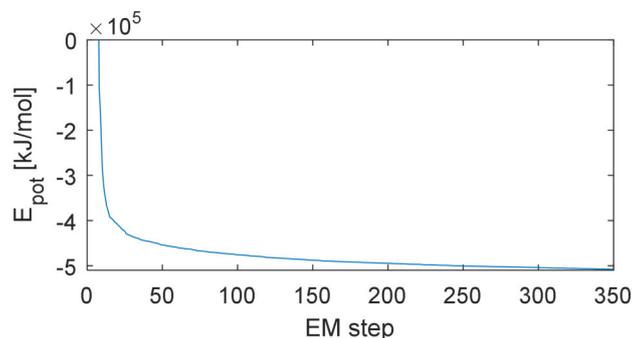
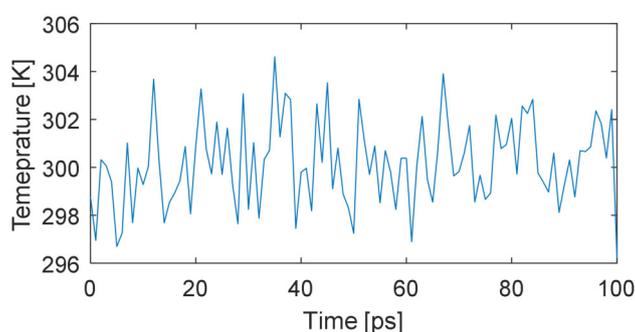
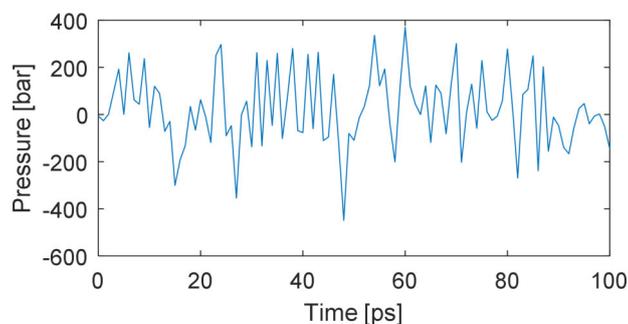


FIG. 3.6. (R)-4-((R)-(4-chlorophenyl)(hydroxy)methyl)-5-methylene-1,3-dioxolan-2-one

Number of particles, Volume, and Temperature, NVT). The timeframe for this simulation was set to 100ps. The system temperature after equilibration is shown in Figure 4.2. The 300K target value is rapidly reached, further remaining stable over time.

The second equilibration phase includes the system pressure stabilization. This phase is performed under isothermal-isochoric conditions (constant Number of particles, Pressure, and Temperature, NPT). The timeframe was also set to 100ps for this simulation. Figure 4.3 indicates the system pressure variations over the simulated timeframe, which was an expected behavior.

After equilibrating the system at the desired temperature and pressure, production MD phase is performed. For this computationally demanding simulation phase, high-performance computers are required. In this exam-

FIG. 4.1. *Energy minimization phase*FIG. 4.2. *Temperature of the system*FIG. 4.3. *Pressure of the system*

ple, 1ns simulation is performed. GROMACS has some built-in tools for MD analysis. The radius of gyration (Rg) of a protein is used as the compactness measure. Based on the results presented in Figure 4.4, 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 Root Mean Square Deviation (RMSD). The Figure 4.5 shows that the RMSD oscillates around 0.15 nm, indicating that the structure is very stable.

Figure 4.6 shows the protein that is the result of the presented simulations. It shows parts stabilized by hydrogen bonding (secondary structure) as well as specific spatial variants of the given protein. According to GROMACS-simulation results, this is the most stable structure of the given protein.

Computer-intensive calculations were performed on FINKI HPC, as one of resources of VI-SEEM project

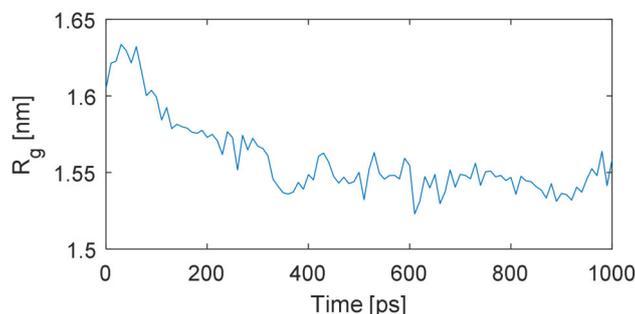


FIG. 4.4. Radius of gyration

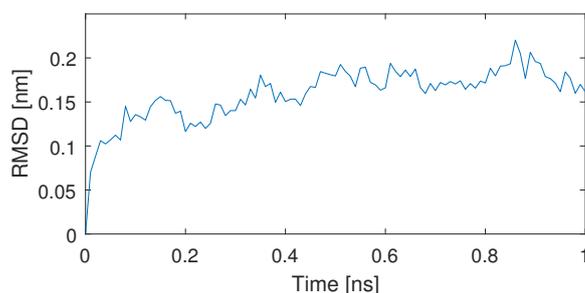


FIG. 4.5. RMSD

[20]. The parallel data processing noticeably accelerates the simulations and proves necessity of parallel processing in molecular dynamic simulations. The considered datasets and workflows provide the ability to test various combinations before going into testing in the laboratory.

5. Conclusion. By using computer simulations, which are characterized by high accuracy and precision, a valuable information usable in design new drugs can be provided. Moreover, it would be possible to investigate the interactions of such molecules that have not been made in the laboratory, and in terms of their structural or physiological properties, they are similar to those existing in nature (or synthesized artificially). On the other hand, the application of simulations contributes to a significant reduction in research time, and orientation to those groups of molecules that are highly likely to exhibit a physiological effect after interaction with the given receptor. In this way, the number of molecules that are planned to be tested is significantly reduced.

Finally, by exchanging some functional groups with some other, it would get useful information about the possible interactions (and activities) of such molecules, which would significantly contribute to the synthetic application of the tested reaction or methodology.

In this paper molecular dynamic simulations between series of new molecules with lysozyme are presented. The initial simulation results showed that there is no interaction between the synthesized molecules (enol-carbonate) and the lysosome. Before concluding that there is no interaction of an organic molecule with protein at all, several experiments with the same systems need to be conducted.

Acknowledgments. The research leading to these results has been co-funded by the European Commission under the H2020 Research Infrastructures contract no. 675121 (project VI-SEEM).

REFERENCES

- [1] NAIR, P. C., MINERS, J. O., *Molecular dynamics simulations: From structure, function relationships to drug discovery*, in *In Silico Pharmacology*, 2014, 2, 4.
- [2] *RCSB Protein Data Bank*, <http://www.rcsb.org/>

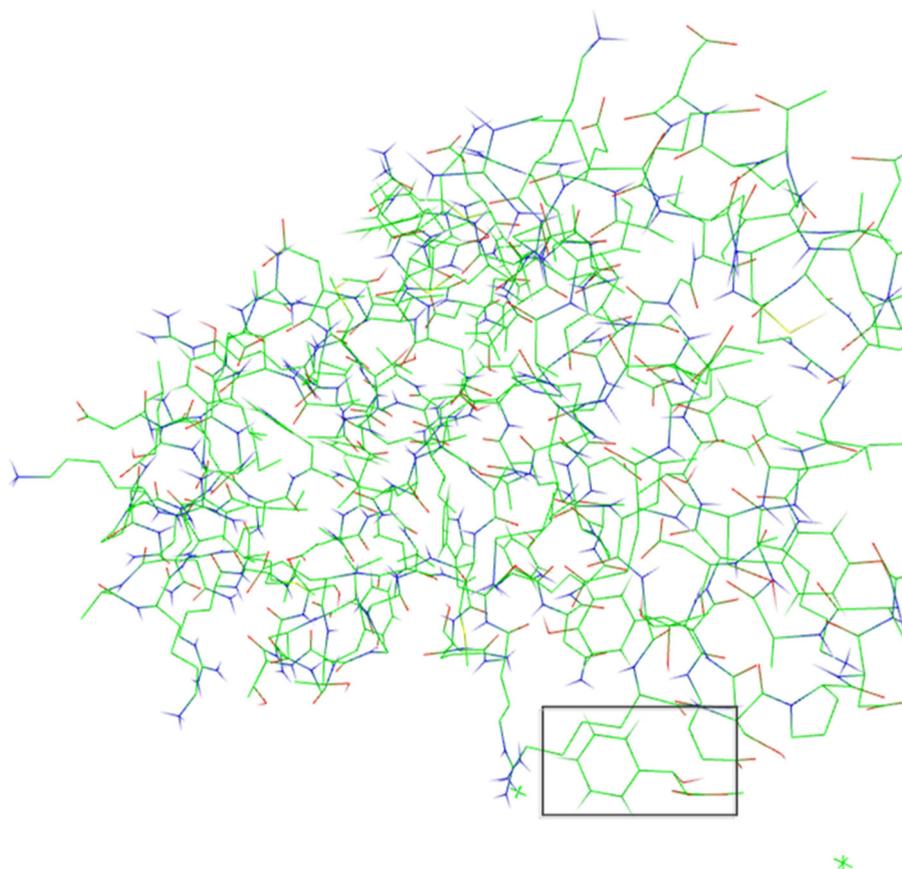


FIG. 4.6. The most stable possible conformation of protein, presented in VegaZZ

- [3] Z. CEKOVIC I SARADNICI, *Molekuli u tajnama zivota i svetu oko nas*, Zavod za udzbenike, Beograd, 2009, pp 139-149
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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., 170118.
- [9] LOH, T.P., LYE, P.L., *A concise synthesis of (-)-methylenolactocin and the formal synthesis of (-)-phaseolinic acid*, *Tetrahedron Lett*, 2001, 42, 3511.
- [10] BACKHAUS, D, *Tetrahedron Lett.*, 2000, 41, 2087.
- [11] KABALKA, G. W., VENKATAIAH, B., *Tetrahedron Lett.*, 2005, 46, 7325.
- [12] CHAPPEL, M. D., HALCOMB, R L., *Org. Lett.*, 2000, 2, 2003.
- [13] BIGOVIC, M., MASLAK, V., TOKIC-VUJOSEVIC, Z., DIVJAKOVIC, V., SAICIC, *A useful synthetic equivalent of a hydroxyacetone enolate*, *R. Org. Lett.* 2011, 13, 4720.
- [14] BIGOVIC, M.; SKARO, S.; MASLAK, V.; SAICIC, R. N., *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 Lett.* 2013, 54, 6624
- [15] HUMPHREY W., DALKE A., SCHULTEN K., *VMD: Visual molecular dynamics*, *Journal of Molecular Graphics* Volume 14, Issue 1, February 1996, Pages 33-38
- [16] *ACD/ChemSketch for Academic and Personal Use*, ACD/Labs.com. [online] Available at: <http://www.acdlabs.com/resources/freeware/chemsketch/>
- [17] *Open Babel : The Open Source Chemistry Toolbox*, <http://openbabel.org/>

- [18] A. W. SCHTTELKOPF, D. M. F. VAN AALTEN , *PRODRG: a tool for high-throughput crystallography of protein-ligand complexes*, Acta Crystallogr D60, 2004, 13551363.
- [19] *GROMACS Tutorial - Baven Lab*, <http://www.bevanlab.biochem.vt.edu/Pages/Personal/justin/gmx-tutorials/>
- [20] *H2020 project VI-SEEM*, <https://vi-seem.eu>

Edited by: Aneta Karaivanova

Received: Jan 6, 2017

Accepted: Apr 2, 2018