

# Supercomputing Frontiers and Innovations

2022, Vol. 9, No. 2

## Scope

- Future generation supercomputer architectures
- Exascale computing
- Parallel programming models, interfaces, languages, libraries, and tools
- Supercomputer applications and algorithms
- Novel approaches to computing targeted to solve intractable problems
- Convergence of high performance computing, machine learning and big data technologies
- Distributed operating systems and virtualization for highly scalable computing
- Management, administration, and monitoring of supercomputer systems
- Mass storage systems, protocols, and allocation
- Power consumption minimization for supercomputing systems
- Resilience, reliability, and fault tolerance for future generation highly parallel computing systems
- Scientific visualization in supercomputing environments
- Education in high performance computing and computational science

## Editorial Board

### Editors-in-Chief

- **Jack Dongarra**, University of Tennessee, Knoxville, USA
- **Vladimir Voevodin**, Moscow State University, Russia

### Editorial Director

- **Leonid Sokolinsky**, South Ural State University, Chelyabinsk, Russia

### Associate Editors

- **Pete Beckman**, Argonne National Laboratory, USA
- **Arndt Bode**, Leibniz Supercomputing Centre, Germany
- **Boris Chetverushkin**, Keldysh Institute of Applied Mathematics, RAS, Russia
- **Alok Choudhary**, Northwestern University, Evanston, USA
- **Alexei Khokhlov**, Moscow State University, Russia
- **Thomas Lippert**, Jülich Supercomputing Center, Germany

- **Satoshi Matsuoka**, Tokyo Institute of Technology, Japan
- **Mark Parsons**, EPCC, United Kingdom
- **Thomas Sterling**, CREST, Indiana University, USA
- **Mateo Valero**, Barcelona Supercomputing Center, Spain

### Subject Area Editors

- **Artur Andrzejak**, Heidelberg University, Germany
- **Rosa M. Badia**, Barcelona Supercomputing Center, Spain
- **Franck Cappello**, Argonne National Laboratory, USA
- **Barbara Chapman**, University of Houston, USA
- **Yuefan Deng**, Stony Brook University, USA
- **Ian Foster**, Argonne National Laboratory and University of Chicago, USA
- **Geoffrey Fox**, Indiana University, USA
- **William Gropp**, University of Illinois at Urbana-Champaign, USA
- **Erik Hagersten**, Uppsala University, Sweden
- **Michael Heroux**, Sandia National Laboratories, USA
- **Torsten Hoefler**, Swiss Federal Institute of Technology, Switzerland
- **Yutaka Ishikawa**, AICS RIKEN, Japan
- **David Keyes**, King Abdullah University of Science and Technology, Saudi Arabia
- **William Kramer**, University of Illinois at Urbana-Champaign, USA
- **Jesus Labarta**, Barcelona Supercomputing Center, Spain
- **Alexey Lastovetsky**, University College Dublin, Ireland
- **Yutong Lu**, National University of Defense Technology, China
- **Bob Lucas**, University of Southern California, USA
- **Thomas Ludwig**, German Climate Computing Center, Germany
- **Daniel Mallmann**, Jülich Supercomputing Centre, Germany
- **Bernd Mohr**, Jülich Supercomputing Centre, Germany
- **Onur Mutlu**, Carnegie Mellon University, USA
- **Wolfgang Nagel**, TU Dresden ZIH, Germany
- **Alexander Nemukhin**, Moscow State University, Russia
- **Edward Seidel**, National Center for Supercomputing Applications, USA
- **John Shalf**, Lawrence Berkeley National Laboratory, USA
- **Rick Stevens**, Argonne National Laboratory, USA
- **Vladimir Sulimov**, Moscow State University, Russia
- **William Tang**, Princeton University, USA
- **Michela Taufer**, University of Delaware, USA
- **Andrei Tchernykh**, CICESE Research Center, Mexico
- **Alexander Tikhonravov**, Moscow State University, Russia
- **Eugene Tyrtshnikov**, Institute of Numerical Mathematics, RAS, Russia
- **Roman Wyrzykowski**, Czestochowa University of Technology, Poland
- **Mikhail Yakobovskiy**, Keldysh Institute of Applied Mathematics, RAS, Russia

### Technical Editors

- **Yana Kraeva**, South Ural State University, Chelyabinsk, Russia
- **Mikhail Zymbler**, South Ural State University, Chelyabinsk, Russia
- **Dmitry Nikitenko**, Moscow State University, Moscow, Russia

## Contents

<b>Computational Characterization of N-acetylaspartylglutamate Synthetase: From the Protein Primary Sequence to Plausible Catalytic Mechanism</b> I.V. Polyakov, A.E. Kniga, A.V. Nemukhin .....	4
<b>Predicting the Activity of Boronate Inhibitors Against Metallo-<math>\beta</math>-lactamase Enzymes</b> E.O. Levina, M.G. Khrenova, V.G. Tsirelson .....	14
<b>Predicting Binding Free Energies for DPS Protein-DNA Complexes and Crystals Using Molecular Dynamics</b> E.V. Tereshkin, K.B. Tereshkina, Yu.F. Krupyanskii .....	33
<b>Computational Modeling of the Interaction of Molecular Oxygen with the Flavin-dependent Enzyme RutA</b> I.V. Polyakov, T.M. Domratcheva, A.M. Kulakova, A.V. Nemukhin, B.L. Grigorenko .....	46
<b>Analysis of Ion Atmosphere Around Nucleosomes Using Supercomputer MD Simulations</b> N.A. Kosarim, G.A. Armeev, M.P. Kirpichnikov, A.K. Shaytan .....	56
<b>Molecular Modeling of Penicillin Acylase Binding with a Penicillin Nucleus by High Performance Computing: Can Enzyme or its Mutants Possess <math>\beta</math>-lactamase Activity?</b> E.M. Kirilin, A.A. Bochkova, N.V. Panin, I.V. Pochinok, V. Švedas .....	68
<b>Search for Ligands Complementary to the 430-cavity of Influenza Virus Neuraminidase by Virtual Screening</b> D.K. Nilov, M. Schmidtke, V.A. Makarov, V.K. Švedas .....	79



This issue is distributed under the terms of the Creative Commons Attribution-Non Commercial 3.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.

# Computational Characterization of N-acetylaspartylglutamate Synthetase: From the Protein Primary Sequence to Plausible Catalytic Mechanism

Igor V. Polyakov<sup>1,2</sup> , Artem E. Kniga<sup>1,2</sup> , Alexander V. Nemukhin<sup>1,2</sup> 

© The Authors 2022. This paper is published with open access at SuperFri.org

The methods of supercomputer molecular modeling are applied to characterize structure and dynamics of one of the key human brain enzymes, N-acetylaspartylglutamate synthetase. The three-dimensional all-atom models of the enzyme with the reactants in the active site are constructed in several steps, starting from pilot protein structure in the *apo*-form obtained with the AlphaFold2 from the protein primary sequence. Deposition of reactant molecules into the protein cavity, construction of the reaction intermediate and relaxation of the complex are carried out with the help of large-scale classical molecular dynamics calculations. On the top of the construct, molecular dynamics simulations with the quantum mechanics/molecular mechanics interaction potentials are performed for the most promising conformations of the model system. Analysis of the latter allows us to propose plausible catalytic mechanisms of chemical reactions in the enzyme active site. The applied computational strategy opens the way towards *ab initio* enzymology using modern supercomputer simulations.

*Keywords:* molecular dynamics, quantum mechanics/molecular mechanics, QM/MM MD, GPU-accelerated algorithms, N-acetylaspartylglutamate synthetase, enzyme-substrate complexes, reaction intermediates.

## Introduction

High-performance computing plays an increasingly important role in life sciences, including simulations of chemical reactions in enzymes using advanced modeling methods based on the quantum mechanics/molecular mechanics (QM/MM) theory [1, 13, 14]. A practical goal of these simulations is to exploit the obtained information on structures and dynamics in protein systems for prediction of novel prospective drugs to fight human diseases [2]. Usually, such calculations are based upon the available experimental results, such as three-dimensional structures of macromolecules deposited in the Protein Data Bank (PDB) [6]. However, even in this case, raw crystallographic or nuclear magnetic resonance spectroscopy data undergo significant computational refinement [23] before deposition in the structure databanks.

The amount of available protein sequences and structural data enabled a number of tools to predict 3D protein structures [15] including the neural network-based models such as AlphaFold2 [12], which is claimed to be “demonstrating accuracy competitive with experimental structures in a majority of cases and greatly outperforming other methods” in the challenging 14th Critical Assessment of protein Structure Prediction (CASP14). However, very limited information if any is available, whether such structures are useful to enable accurate simulations using the QM/MM-based approaches to obtain enzymatic reaction mechanisms, locate reaction intermediates and evaluate energy profiles. Classical molecular dynamics (MD) simulations are routinely used to prepare structures for the QM/MM simulations and to refine the computationally predicted structures [10]. The downside of computational prediction of accurate protein structure and dynamics by using the all-atom classical MD is the amount of computational ef-

<sup>1</sup>Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, Moscow, Russian Federation

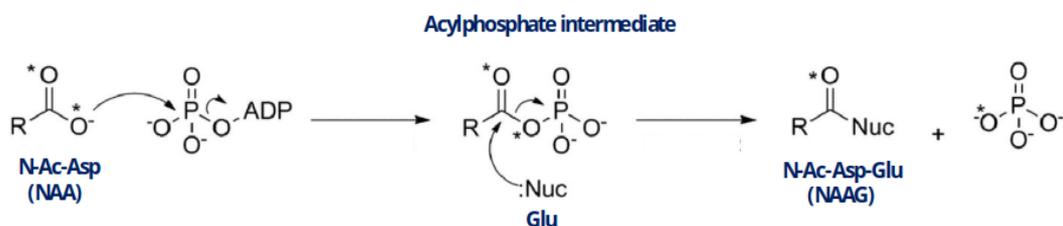
<sup>2</sup>Department of Chemistry, Lomonosov Moscow State University, Moscow, Russian Federation

fort required to obtain reliable results, because a swarm of trajectories that cover microsecond timescales are needed [10].

On the upside, the recent developments in the MD simulation software with the GPU-resident version of NAMD 3.0 [19] promise great advances for the classical MD with the GPU-heavy supercomputers. Specifically, each trajectory can be efficiently calculated on a single GPU without significant load on the CPU and interconnect. Thus, many MD trajectories can be executed at the same time with high performance and efficient hardware utilization. The follower of the classical MD, i.e. molecular dynamics simulations with QM/MM potentials (QM/MM MD), benefit greatly from such GPU systems as well, if the NAMD/Terachem software stack is used [16, 22]. Original implementation of the NAMD QM/MM script interface to Terachem used to have some pitfalls, which have been recently fixed [13]. In this work, we use the high-performance classical MD and QM/MM MD calculations in order to characterize structure and dynamics of the acylphosphate reaction intermediate in the catalytic cycle of the key human brain enzyme, N-acetylaspartylglutamate synthetase (NAAGS), responsible for the formation of the most abundant brain dipeptide N-acetylaspartylglutamate (NAAG) [4]. This dipeptide acts as a retrograde neurotransmitter selectively localized in the glutamatergic synapses; it plays an essential role in cognition and memory consolidation underlying the novel object recognition task [5].

We have previously described computer simulations aimed to predict the reaction mechanism of the related enzyme, N-acetylglutamate synthase [20] and to characterize formation of a very abundant human neuropeptide N-acetylaspartate (NAA), one of the reactants in NAAG synthesis [21]. An essential feature of the present project is that no relevant structure of NAAG synthetase enzyme is available in the PDB. Therefore, we ought to apply modern computational methods to construct a full-atom three-dimensional structure of the enzyme on the base of its primary amino acid sequence and to deposit reaction species into the enzyme active site. This is a heavy computationally demanding task, which requires the use of supercomputer facilities.

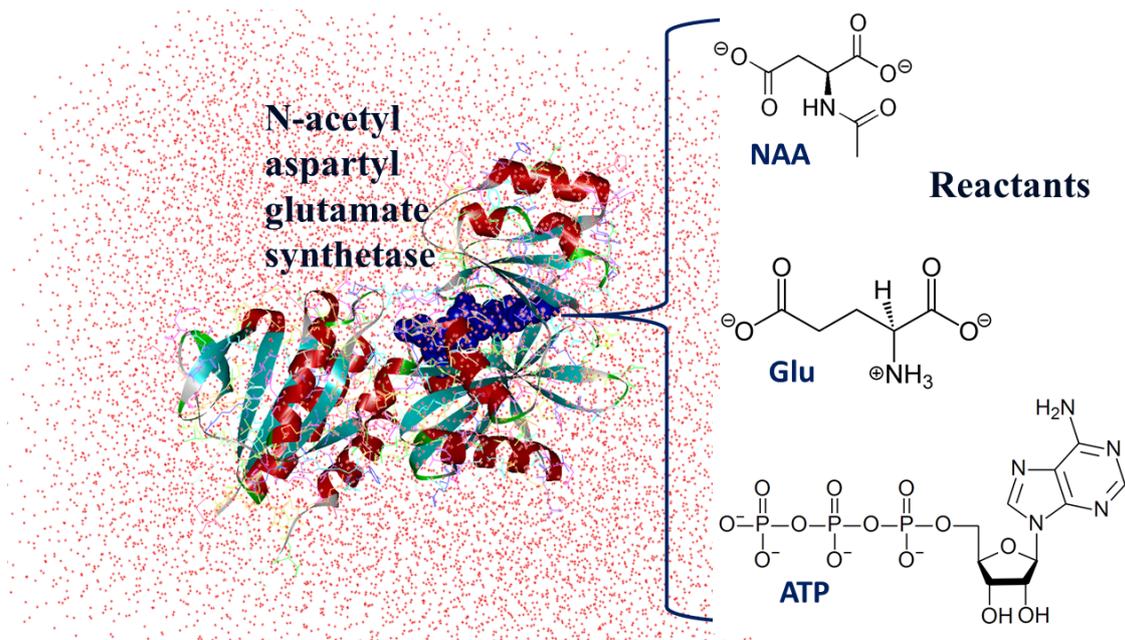
A tentative scheme of NAAG synthesis is outlined in Fig. 1; however, no attempts are known to specify the reaction mechanism. Thus, our computationally derived data present the first approach to characterize this important process.



**Figure 1.** Expected reaction mechanism of the synthesis of N-acetylaspartylglutamate (NAAG). N-acetylaspartate (NAA), glutamate (Glu) as the nucleophile (Nuc) and adenosine triphosphate are the reactants

The left panel in Fig. 2 shows the model system to be used for simulation of the reaction. Chemical formulae of three reactants are presented in the right panels.

We describe, in this work, the initial step of the construction of the acylphosphate reaction intermediate (see the central panel in Fig. 1). Rationalization of structural and dynamical features of this intermediate is a necessary stage of the entire project, because attempts to build



**Figure 2.** Computationally derived model system. Conventionally,  $\alpha$ -helices of the protein are shown in red,  $\beta$ -sheets in cyan. The reactants are schematically shown as the dark blue space-filled spheres. The red dots surrounding the protein refer to solvation water shells

*in silico* enzyme-substrate complexes without knowledge of relevant structures from PDB would most probably lead to highly uncertain starting points in modelling the reaction mechanism.

## 1. Computational Methods

The primary amino acid sequence of the NAAGS sequence coded by the RIMKLA gene (Ribosomal Modification Protein RimK Like Family Member A) was taken from the UniProtKB [3] record Q8IXN7. Multiple sequence alignment was obtained with MMSEQ2 search [24] using a local installation of ColabFold pipeline featuring AlphaFold2 model for protein structure prediction [17]. The pipeline was run with a default preset of parameters for monomeric proteins.

The obtained AlphaFold model structure did not contain any ligands or water molecules; hence, an extensive active site reconstruction was required. To assist this step, the crystal structures PDB ID: 1GSA [9] and 2DLN [8] were taken as reference structures to construct the acylphosphate intermediate, namely, to introduce two magnesium ions, ADP, acylphosphate (NAA-PO<sub>3</sub>) and Glu. We aligned and positioned ADP with the hydrogen bonds to Val192, Lys190, Gly161 main chain atoms; Gln189, Lys154, Asp199, Lys111 side chain atoms; magnesium ions and their coordination spheres with the contacts to ADP, NAA-PO<sub>3</sub>, Glu273, Asp260, Asn275 and water molecules; NAA-PO<sub>3</sub> was coordinated by Arg160, Arg201, magnesium atoms and water molecules. Unlike for other ligands, no clear reference could be found for the glutamate position; hence, several starting positions were chosen to assist the C-N bond formation during the reaction progress (Fig. 1). Water molecules were initially added to the AlphaFold model by the Dowser++ [18] software. Several water molecules were removed to avoid clashes with the ligands in the active site. The model was solvated and made charge-neutral through adding the counter-ions with VMD [11]. The model contains 39346 atoms in total.

Classical molecular dynamics simulation with NAMD [19] was chosen as an approach to refine the designed structures before running the QM/MM MD calculations. Simulations were

carried out assuming the isothermalisobaric (NPT) ensemble at  $P = 1$  atm and  $T = 300$  K using the Nos-Hoover Langevin piston pressure control and Langevin dynamics, integration step was set to 1 fs. Periodic boundary conditions along with the particle mesh Ewald method to account for the long-range electrostatic interactions were employed.

At first, harmonic restraint potential was applied to all protein backbone atoms to equilibrate the initial models. This is a standard practice [10] in the MD refinement protocols, the simulations were run for at least 100 ns each. All the constraints were released at the second step of the model refinement. At this step, we faced certain difficulties, because long trajectories ended up with the protein reorientation towards a shorter periodic cell dimension, which resulted in an artefact of MD simulations with the imposed boundary conditions, the formation of thread-like megastructures accounting for the periodicity. To overcome this difficulty, we restarted calculations from the constrained trajectories with most of the constraints released except for the CA atoms of the  $\beta$ -sheets (although the CA atoms of the  $\beta$ -sheets near the active site were constrained). In this approach, the protein conformational flexibility is greatly enhanced, as compared to initial restrained trajectories, but the periodicity artefacts can be avoided.

NAMD 3.0 was employed in order to fully harvest computational capabilities of the DGX2 supercomputer. A DGX2 node enabled us to run parallel swarms of trajectories of up to 16 simultaneously with all the GPUs and only 16 CPU cores utilized with the GPU-resident version of NAMD with CUDASOIntegrate option set to “on”. A total amount of 1.3 Tb of classical MD trajectories was produced, which accounts for 29 000 ns, whereas 1600 ns can be produced per day using all the Tesla V100 GPUs of a DGX2 node.

Four frames from the significantly different and stable classical MD trajectories were selected to start QM/MM MD simulations. The QM part contained 135 atoms described by the density function theory with the PBE0 hybrid functional, D3 dispersion corrections and 6-31G\*\* basis set with 1395 basis functions in total. The QM system included NAA-PO<sub>3</sub>, ADP (cut on the C4'-C5' bond), two magnesium atoms; Arg160, Arg201, Arg215, Glu273, Asp260, Asn275 side chains and water molecules.

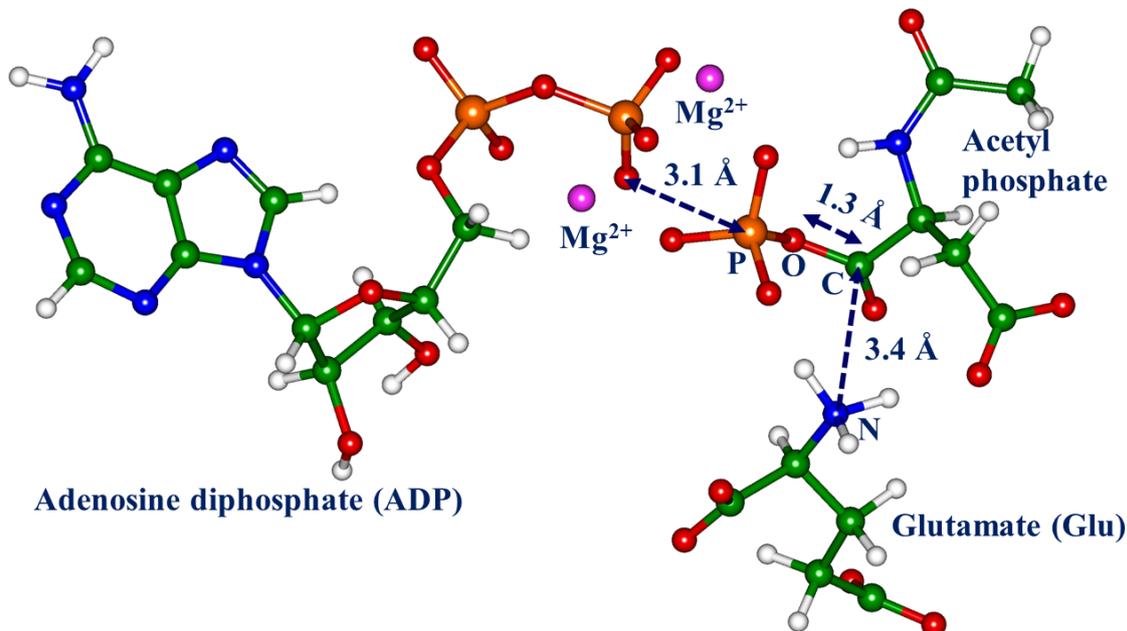
QM/MM simulations were run with the NAMD/Terachem [16, 22] software and the modified interface [13]. Combined length of computed QM/MM MD trajectories was over 50 ps, with a performance of  $\approx 1.5$  ps/day per GPU.

## 2. Analysis of the Designed Protein Structure

As described above, we designed computationally the model system in the conformation of the reaction intermediate containing phosphorylated NAA (see the central panel in Fig. 1). This structure is the best starting point for future simulations of the full energy profile, because attempts to begin construction from the enzyme-substrate complex, that is from the enzyme with the reactants (the left panel in Fig. 1), should be prohibitively expensive due to the expected huge conformation flexibility of the protein with reactants.

The model system used in simulations is composed of the protein with the embedded ligands, phosphorylated NAA, Glu, and ADP (see Fig. 1), surrounded by the shell of water molecules. The ligands are sandwiched between two  $\beta$ -strands formed by the amino acid residues #163-168, 185-191 on the one side and by five  $\beta$ -strands that include residues from the range of 197 to 277. The two loops consisting of residues #158-162 and 218-230 form a gate-like contact between these  $\beta$ -sheets.

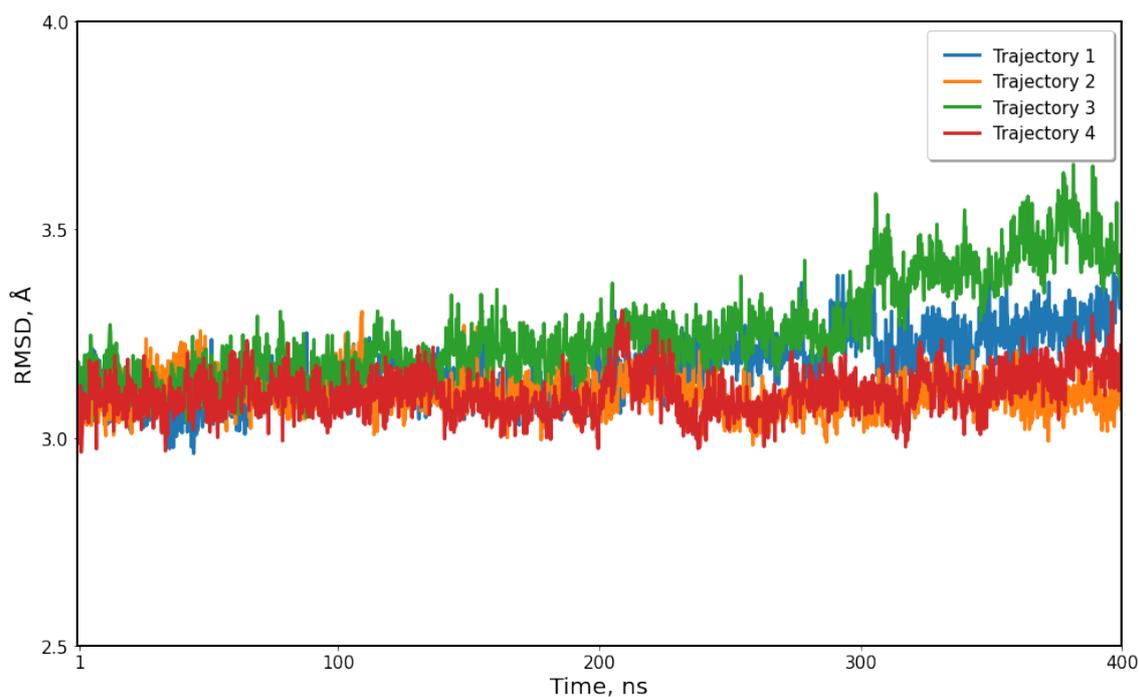
Figure 3 illustrates a typical MD frame of the system showing the atoms of the reaction intermediate. The specified distances between Glu and NAA-PO<sub>3</sub> (the C-N distance, 3.4 Å), between NAA-PO<sub>3</sub> and ADP (the distance 3.1 Å between the phosphorus atom of the  $\gamma$ -phosphate group covalently bound to NAA and the oxygen atom of the  $\beta$ -phosphate), are given only to illustrate geometry parameters, which are fluctuating along MD trajectories.



**Figure 3.** A structure of the reaction intermediate containing the acetylphosphate species with the phosphorylated NAA. The dashed dark blue arrows indicate the direction of nucleophilic attack along the N-C distance and the break of the C-O bond to form NAAGS, as well as the coordinate of the bond P-O breakage in the initial ATP molecule

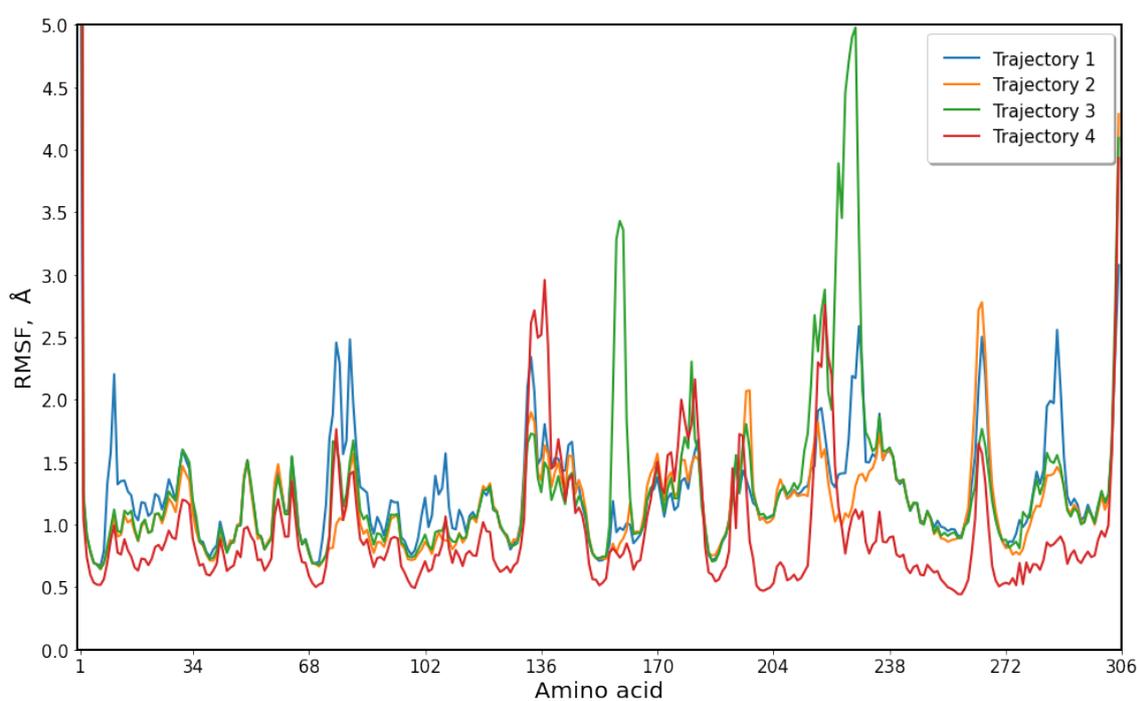
Both visual MD trajectory inspection and trajectory RMSD (root mean square deviation) analysis (Fig. 4) show that the overall protein structure remains stable during all classical MD trajectories. The RMSD from the predicted model calculated over all the CA atoms is in the range of 3-3.5 Å.

Maintaining the original protein fold is not enough for the model to be useful in further QM/MM MD calculations, which strongly depend on the conformation of the active site. While all four presented trajectories show somewhat different but stable conformations within their active sites, only trajectories 1 and 2 maintain the active site conformation during 400 ns. Trajectories 3 and 4 show degradation of the active site. This is not evident from the Fig. 4, but the RMSF (root mean square fluctuation) analysis (Fig. 5) provides additional data. Many movements observed via the RMSF analysis were found in all trajectories, and do not contribute to destabilize configurations of the active site. For example, those are large deviations near the C- and N-terminus; the regions #132-150, 170-180, which are covered by the  $\alpha$ -helices connected by loops to the  $\beta$ -strands: they are expected to be more flexible than the  $\beta$ -sheets. The residues #158-162 and 218-230 from a gate-like contact, connecting the  $\beta$ -sheet structures (Fig. 6), show the most variance (see also Fig. 5) for the trajectory 3, and this movement correlates with the active site destabilization. For the trajectory 4, no such correlation in the RMSF analysis is observed. The visual inspection of this trajectory reveals that the movement of the side chain of Arg160, which coordinates NAA-PO<sub>3</sub>, is associated with the active site conformational change. In many other trajectories, which are not shown in the manuscript, the discussed issues, namely,

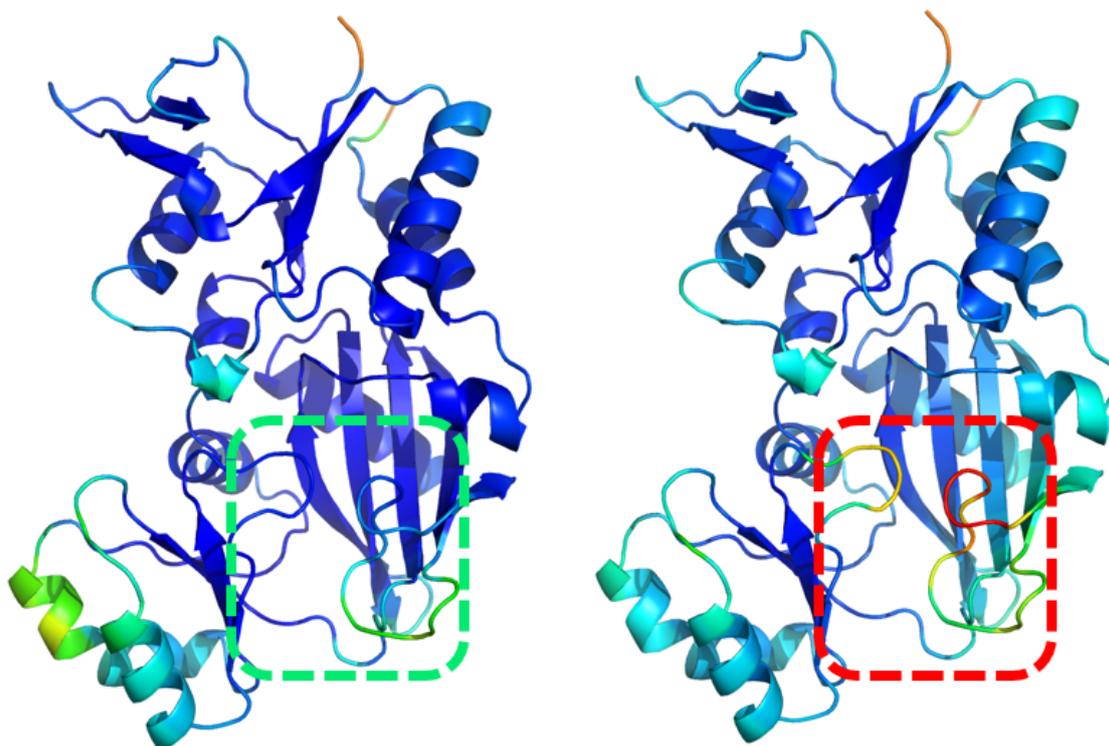


**Figure 4.** RMSD of the CA atoms compared to the original AlphaFold2 model during selected classical MD runs

lack of stable loop contact forming a gate or destabilization of Arg160 side chain contact to the NAA-PO<sub>3</sub>, are associated with the active site structure degradation or destruction. It is important to follow, whether Glu drifts away or the NAA-PO<sub>3</sub> position changes to the extent that prevents NAAG formation (Fig. 1).



**Figure 5.** RMSF during selected classical MD runs



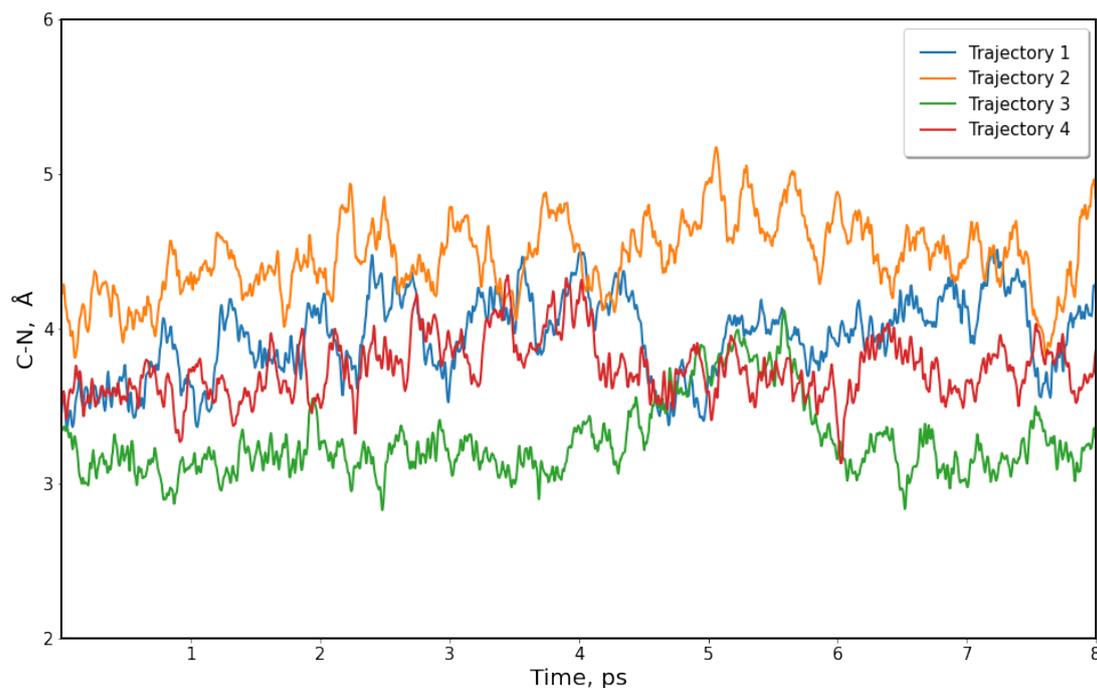
**Figure 6.** RMSF “coloring” of the 3D protein structure. Less RMSF – blue, cyan, green, yellow, red – more RMSF. Left – trajectory 4, right – trajectory 3. The green and right mark boxes highlight the region of the primary interest

### 3. Prospective Reaction Mechanism

Following results of the analysis of classical MD trajectories, we can proceed to QM/MM MD calculations. The corresponding frames were selected from those parts of classical trajectories that maintained desired conformations of the active site with the reacting species. The computed graphs in Fig. 7, as well as the data in Fig. 3, present a basis for the consideration of the reaction mechanism.

One of the key aspects of the active site structure of the model protein system is the C-N distance (Fig. 3) between the glutamate and the NAA-PO<sub>3</sub>. The short and stable value of this distance during a QM/MM MD trajectory run (Fig. 7) would indicate that the nucleophilic attack (Fig. 1) is favorable given that the H<sub>3</sub>N<sup>+</sup> group of glutamate is activated first. This can be achieved by removing a proton to a general base in the active site. Remarkably, there are three candidates for such an acceptor: two carboxyl groups of the attacking glutamate and the carboxyl group of the NAA-PO<sub>3</sub>. There are no other carboxyl groups from Glu or Asp residues nearby, which are available to accept a proton. The phosphate groups of NAA-PO<sub>3</sub> and ADP should not be favorable proton acceptors having low pK<sub>a</sub> values and being coordinated by magnesium atoms in the protein active site. Thus, we speculate, that the NAAG formation reaction is an example of the substrate assisted catalysis [7].

It is important to note that no considerable conformational changes occurred during the short QM/MM MD trajectories except for Arg160 side chain adjustment in all the calculations. The variation of the critical C-N attack distance is rather large between the different trajectories (Fig. 7), ranging from the presumably reactive 3 Å (trajectory 3) to non-reactive 5 Å (trajectory 2). Thus, we note the importance of the computational details of the refinement procedure.



**Figure 7.** The nucleophilic attack C-N distance in the active site during the QM/MM trajectories. The 8 ps windows were obtained by cutting out the first 2 ps of each trajectory

In order to obtain a single model for QM/MM MD simulation of the reaction, a user should perform an enormous amount of attempts with different active site conformations and the corresponding trajectories. It is of utmost importance that such sampling procedure is carried out in parallel with an efficient software that utilizes modern supercomputer architectures, such as the software stack we described in the Computational approaches section.

## Conclusion

In this paper, we describe a strategy to characterize *in silico* the enzyme catalysis, starting from a protein primary sequence without knowing other experimental data usually employed in such computer simulations. In particular, no relevant crystal structures are available in the Protein Data Bank for the adenosine triphosphate dependent binding of N-acetylaspartate and glutamate in the active of N-acetylaspartylglutamate synthetase (NAAGS). The primary sequence can be converted to a pilot structure of the NAAGS protein in the *apo*-form (that is, without reactants in the enzyme active site) with the help of the recently developed algorithms of AlphaFold2. The reacting species are inserted into the enzyme active site using the molecular modeling tools. The analysis of an *in silico* designed structure of NAAGS with the ligands shows that multiple manual corrections are required, which are introduced using the molecular modeling tools. The structure is refined using large-scale classical molecular dynamics simulations as well as molecular dynamics calculations with the QM/MM potentials. The performed analysis of the computationally designed complexes allows us to propose a reaction mechanism in this complicated enzyme-catalyzed chemical reaction, opening the way towards *ab initio* enzymology using modern supercomputer simulations.

## Acknowledgements

We thank Russian Science Foundation (project No. 18-13-00030) for the financial support of this work. The study was carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University.

*This paper is distributed under the terms of the Creative Commons Attribution-Non Commercial 3.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.*

## References

1. Ahmadi, S., Barrios Herrera, L., *et al.*: Multiscale modeling of enzymes: QM-cluster, QM/MM, and QM/MM/MD: A tutorial review. *International Journal of Quantum Chemistry* 118(9), e25558 (2018). <https://doi.org/10.1002/qua.25558>
2. Aminpour, M., Montemagno, C., Tuszynski, J.A.: An overview of molecular modeling for drug discovery with specific illustrative examples of applications. *Molecules* 24(9), 1693 (Apr 2019). <https://doi.org/10.3390/molecules24091693>
3. Apweiler, R.: UniProt: the Universal Protein knowledgebase. *Nucleic Acids Research* 32(90001), D115–D119 (Jan 2004). <https://doi.org/10.1093/nar/gkh131>
4. Becker, I., Lodder, J., Gieselmann, V., Eckhardt, M.: Molecular characterization of N-acetylaspartylglutamate Synthetase. *Journal of Biological Chemistry* 285(38), 29156–29164 (Sep 2010). <https://doi.org/10.1074/jbc.m110.111765>
5. Becker, I., Wang-Eckhardt, L., Lodder-Gadaczek, J., *et al.*: Mice deficient in the NAAG synthetase II gene *Rimk1a* are impaired in a novel object recognition task. *Journal of Neurochemistry* 157(6), 2008–2023 (Mar 2021). <https://doi.org/10.1111/jnc.15333>
6. Berman, H.M.: The protein data bank. *Nucleic Acids Research* 28(1), 235–242 (Jan 2000). <https://doi.org/10.1093/nar/28.1.235>
7. Dall’Acqua, W., Carter, P.: Substrate-assisted catalysis: Molecular basis and biological significance. *Protein Science* 9(1), 1–9 (Dec 2008). <https://doi.org/10.1110/ps.9.1.1>
8. Fan, C., Moews, P.C., Walsh, C.T., Knox, J.R.: Vancomycin Resistance: Structure of D-Alanine:D-Alanine Ligase at 2.3 Å Resolution. *Science* 266(5184), 439–443 (Oct 1994). <https://doi.org/10.1126/science.7939684>
9. Hara, T., Kato, H., Katsube, Y., Oda, J.: A Pseudo-Michaelis Quaternary Complex in the Reverse Reaction of a Ligase: Structure of *Escherichia coli* B Glutathione Synthetase Complexed with ADP, Glutathione, and Sulfate at 2.0 Å Resolution. *Biochemistry* 35(37), 11967–11974 (1996). <https://doi.org/10.1021/bi9605245>
10. Heo, L., Feig, M.: Experimental accuracy in protein structure refinement via molecular dynamics simulations. *Proceedings of the National Academy of Sciences* 115(52), 13276–13281 (2018). <https://doi.org/10.1073/pnas.1811364115>
11. Humphrey, W., Dalke, A., Schulten, K.: VMD: Visual molecular dynamics. *Journal of Molecular Graphics* 14(1), 33–38 (Feb 1996). [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5)

12. Jumper, J., Evans, R., Pritzel, A., *et al.*: Highly accurate protein structure prediction with AlphaFold. *Nature* 596(7873), 583–589 (Aug 2021). <https://doi.org/10.1038/s41586-021-03819-2>
13. Khrenova, M.G., Polyakov, I.V., Nemukhin, A.V.: Molecular dynamics of enzyme-substrate complexes in guanosine-binding proteins. *Khimicheskaya Fizika* 41(6), 66–72 (2022). <https://doi.org/10.31857/S0207401X22060061>
14. Khrenova, M.G., Bulavko, E.S., Mulashkin, F.D., Nemukhin, A.V.: Mechanism of Guanosine Triphosphate Hydrolysis by the Visual Proteins Arl3-RP2: Free Energy Reaction Profiles Computed with Ab Initio Type QM/MM Potentials. *Molecules* 26(13), 3998 (Jun 2021). <https://doi.org/10.3390/molecules26133998>
15. Kuhlman, B., Bradley, P.: Advances in protein structure prediction and design. *Nature Reviews Molecular Cell Biology* 20(11), 681–697 (Nov 2019). <https://doi.org/10.1038/s41580-019-0163-x>
16. Melo, M.C.R., Bernardi, R.C., Rudack, T., *et al.*: NAMD goes quantum: an integrative suite for hybrid simulations. *Nature Methods* 15(5), 351–354 (May 2018). <https://doi.org/10.1038/nmeth.4638>
17. Mirdita, M., Shtze, K., Moriwaki, Y., *et al.*: ColabFold: making protein folding accessible to all. *Nature Methods* 19(6), 679–682 (May 2022). <https://doi.org/10.1038/s41592-022-01488-1>
18. Morozenko, A., Stuchebrukhov, A.A.: Dowser++, a new method of hydrating protein structures. *Proteins: Structure, Function, and Bioinformatics* 84(10), 1347–1357 (Jul 2016). <https://doi.org/10.1002/prot.25081>
19. Phillips, J.C., Hardy, D.J., Maia, J.D.C., *et al.*: Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* 153(4), 044130 (Jul 2020). <https://doi.org/10.1063/5.0014475>
20. Polyakov, I.V., Kniga, A.E., Grigorenko, B.L., *et al.*: Computer Modeling of N-Acetylglutamate Synthase: From Primary Structure to Elemental Stages of Catalysis. *Doklady Biochemistry and Biophysics* 495(1), 334–337 (Nov 2020). <https://doi.org/10.1134/S1607672920060125>
21. Polyakov, I.V., Kniga, A.E., Grigorenko, B.L., Nemukhin, A.V.: Structure of the brain N-acetylaspartate biosynthetic enzyme NAT8l revealed by computer modeling. *ACS Chemical Neuroscience* 11(15), 2296–2302 (Jul 2020). <https://doi.org/10.1021/acscchemneuro.0c00250>
22. Seritan, S., Bannwarth, C., Fales, B.S., *et al.*: TeraChem: A graphical processing unit-accelerated electronic structure package for large-scale ab initio molecular dynamics. *WIREs Computational Molecular Science* 11(2) (Jul 2020). <https://doi.org/10.1002/wcms.1494>
23. Shabalin, I.G., Porebski, P.J., Minor, W.: Refining the macromolecular model – achieving the best agreement with the data from X-ray diffraction experiment. *Crystallography Reviews* 24(4), 236–262 (Sep 2018). <https://doi.org/10.1080/0889311x.2018.1521805>
24. Steinegger, M., Söding, J.: MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology* 35(11), 1026–1028 (Oct 2017). <https://doi.org/10.1038/nbt.3988>

# Predicting the Activity of Boronate Inhibitors Against Metallo- $\beta$ -lactamase Enzymes

*Elena O. Levina*<sup>1</sup>, *Maria G. Khrenova*<sup>1,2</sup>, *Vladimir G. Tsirelson*<sup>1,3</sup>

© The Authors 2022. This paper is published with open access at SuperFri.org

Potency of boronate inhibitors against metallo- $\beta$ -lactamases (M $\beta$ Ls) has been found to be dependent on the electrophilicity of the boron atom. It forms a covalent bond with the oxygen atom of the catalytic  $OH^-$  ion in the active site of the enzyme. The ability of the boronate inhibitor to influence the protein conformation also affects the binding potency. Molecular dynamics (MD) simulations of cyclic and non-cyclic boronate complexes with NDM-1 M $\beta$ L show their higher impact on the inhibitor efficiency compared with the electrophilicity of the boron atom. Therefore, we focus on the hardware impact on the computational speedup of the GPU-accelerated MD. Using this data, we propose a comprehensive protocol for *in silico* prediction of the activity of boronate molecules against M $\beta$ L enzymes, which includes MD simulations, combined quantum mechanics / molecular mechanics (QM/MM) computations and molecular dynamics simulations with the QM/MM potentials (QM/MM MD).

*Keywords: metallo- $\beta$ -lactamase, boronate inhibitors, MD, QM/MM MD, quantum theory of atoms in molecules (QTAIM), GPU-accelerated algorithms.*

## Introduction

Nowadays, one of the most severe threats to the healthcare system is a dramatic spread of antibiotic resistant bacteria which have flourished enormously over several decades. Resistance to  $\beta$ -lactam antibiotics via production of  $\beta$ -lactamases by infectious agents is an extremely common type of resistance mechanism [52]. Situation is exacerbated by the structural diversity of  $\beta$ -lactamases, which utilize different mechanisms of  $\beta$ -lactam hydrolysis depending on the structure of their active sites [7, 8]. Serine  $\beta$ -lactamases covalently bind antibiotic via serine residue and perform acylation/deacylation reactions, while metallo- $\beta$ -lactamases (M $\beta$ Ls) utilize  $Zn^{2+}$  cations to coordinate  $\beta$ -lactam molecule in their active site and then cleave the  $\beta$ -lactam four-membered ring by the  $OH^-$  nucleophilic attack. It significantly complicates the search for broad-spectrum inhibitors restoring  $\beta$ -lactam activity against resistant pathogens, regardless of the hydrolysis mechanism they use [40].

Recently, United States Food and Drug Administration and European Medicines Agency approved vaborbactam as a first boronate inhibitor of  $\beta$ -lactamases [28, 50]. Despite its ability to inhibit serine  $\beta$ -lactamases, this compound had no activity against M $\beta$ Ls [28]. It provoked extensive research of boronate compounds active against the entire  $\beta$ -lactamase family. The number of potent inhibitors has been proposed [9, 26, 41, 42, 58], including promising QPX7728 compound with remarkably broad spectrum of inhibition [20, 29, 53]. However, since even the first phase of its clinical trials has not yet been completed [2], there is still need in the research of other boronates effective against M $\beta$ Ls [36, 45]. In this regard, it is especially important to find protocols for *in silico* prediction of the boronate activity.

In this study, we have focused on the several non-cyclic and cyclic boronates having different activity against NDM-1 M $\beta$ L (Tab. 1). NDM-1 is an enzyme provoking a great concern due to its ability to hydrolyze last-resort carbapenem antibiotics and progressing spread over the

<sup>1</sup>A.N. Bach Institute of Biochemistry, Federal Research Centre “Fundamentals of Biotechnology” of the Russian Academy of Sciences, Moscow, Russia

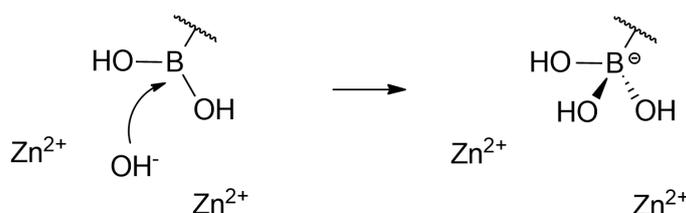
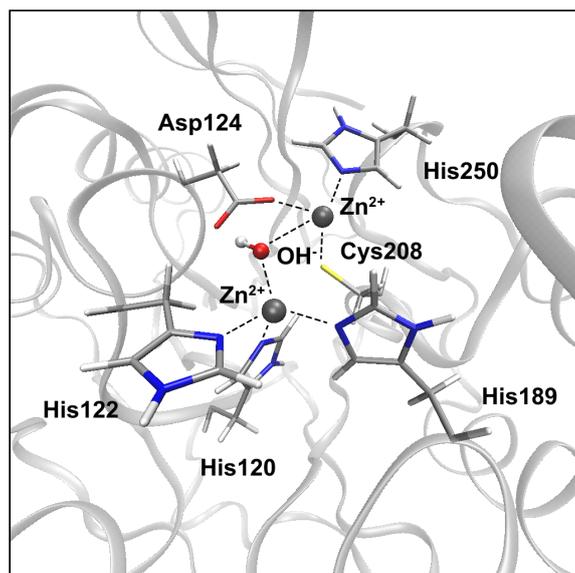
<sup>2</sup>Department of Chemistry, Lomonosov Moscow State University, Moscow, Russia

<sup>3</sup>Mendeleev University of Chemical Technology, Moscow, Russia

world [12, 28, 37, 39, 51]. It has been an ideal object for studying the features of boronate inhibition of metalloenzymes, a complex reversible process [49, 53]. Its mechanism is based on the boronate inhibitor ability to bind the catalytic  $OH^-$  ion to its boron atom and possibility of temporarily prevention  $M\beta L$  from restoring its activity by occupying the enzyme active site via  $Zn^{2+}$  coordination (Fig. 1). Two different factors should be of paramount importance in this process: the reactivity of the boronate fragment and the tendency of the inhibitor to dissociate from the enzyme. We aimed to investigate the particular reasons for the increased activity of the cyclic boronates against NDM-1 in comparison with their non-cyclic analogs and the reasons for different inhibitory activity among the cyclic boronates themselves. Our study showed that this requires taking into account the dynamic properties of the systems under consideration, hence, we were also interested in optimizing time-consuming calculations needed for their evaluation. For this purpose, we used GPU-accelerated computations and focused on the hardware impact on their speedup. Thus, our goal was to offer a comprehensive methodology for the prediction of a boronate inhibitor activity against a metallo- $\beta$ -lactamases, both in terms of the computational protocol and hardware choice.

**Table 1.** Structures of proposed boronate inhibitors of NDM-1 with different activity against the enzyme. Vaborbactam was included as an example of a compound structurally similar to cyclic boronates that lacks activity against NDM-1. Inhibitor activity is expressed as: (i)  $K_i$  value determined with imipenem as a substrate [20]; (ii)  $IC_{50}$  value measured with cephalothin for the non-cyclic inhibitor [46] or imipenem for the QPX7728 [53]; (iii) a minimum concentration of the inhibitor ( $C_{inh}$ ) needed to reduce the MIC of the biapenem to the concentration that corresponds to the PK-PD breakpoint for that  $\beta$ -lactam based on human dosing (potential target value or PTV) [20]

Type	Name	Structure	$IC_{50}$ , $\mu M$	$K_i$ , $\mu M$	$C_{inh}$ , $\mu g/L$
Non-cyclic	–		$60.7 \pm 11.8$	–	–
Cyclic	QPX7728		$0.055 \pm 0.025$	$0.032 \pm 0.014$	0.6
	1		–	–	0.6
	2		–	–	20
Vaborbactam			–	> 40	–



**Figure 1.** Metallo- $\beta$ -lactamase (NDM-1) active site and the mechanism of metallo- $\beta$ -lactamase inhibition by the boronate inhibitors. On the upper frame carbons are shown in silver, nitrogens are in blue, oxygens are in red, sulfurs are in yellow and  $Zn^{2+}$  ions are depicted as big grey spheres

## 1. Methodology

To understand origins of different activity of the boronate inhibitors against NDM-1, chosen cyclic and non-cyclic compounds (Tab. 1) were studied at various theoretical levels. We examined features of their binding with the enzyme using molecular dynamics (MD) and combined quantum mechanics / molecular mechanics (QM/MM) simulations. The dynamic changes in chemical properties of the boronate molecules were studied using a molecular dynamics simulation with the QM/MM potentials (QM/MM MD) of boronates in water solution.

The MD computations were performed for the complexes of the QPX7728 and non-cyclic boronate with NDM-1. Their crystal structures (PDB IDs: 6V1M [20] and 6Q2Y [10]) were taken as a starting point for the construction of the model systems. Both systems were solvated in the rectangular water box (distance from the protein to the cell border equaled 15 Å) and neutralized. After energy minimization and preliminary 20 ns MD simulation with fixed atomic positions of the inhibitor attacked by the  $OH^-$ ,  $Zn^{2+}$  ions and protein L3 loop (Leu65–Val73), the 500 ns simulation was carried out in the NAMD program package [43, 44]. The NPT ensemble

was chosen with  $T = 300$  K,  $p = 1$  atm and 1 fs integration step. The information about system energy, coordinates and velocities has been written every 0.05 ns. The CHARMM36 force field parameters [6, 23, 31, 32] were chosen for the ions and protein atoms, while the TIP3P parameters [25] were used for the water molecules. The inhibitors with bonded  $OH^-$  were described by the CGenFF parameters [55, 60], while their atomic charges were reparametrized via electrostatic-potential-charge derivation procedure using RED Server interface [3, 13, 56] (based on the Gaussian 09 [1] calculations). This methodology has already proven to be effective for the reparameterization of other boronate molecules [57]. The obtained trajectories were analyzed via VMD program [24] and MD Analysis Python library [16, 35].

The QM/MM simulations were performed for the complexes of cyclic inhibitors with the bacterial enzyme. First, the local minimum on the potential energy surface (PES) was located for the QPX7728 complex with NDM-1, then it was modified and reoptimized for other inhibitors. In all cases, the attainment of the PES minimum was confirmed by the analysis of the IR harmonic frequencies. The QM/MM calculations for the QPX7728-protein complex started from the structure, obtained in the preliminary 20 ns MD simulation, with reduced water shell of the model system (8710 atoms in total). The QM part of each model consisted of the inhibitor molecule attacked by the  $OH^-$ ,  $Zn^{2+}$  ions and amino acid residues bonded either with zinc cations, or with inhibitor (His120, His122, Asp124, His189, Cys208, Gly219, Asn220, His250) and several water molecules interacting with the inhibitor. These subsystems were described at the PBE0-D3/6-31G(d,p) level of theory [4, 18, 19, 21], while the MM subsystems were modeled with the AMBER force field parameters [11]. The electronic embedding scheme was applied assuming contributions of the partial charges from all MM atoms to one-electron part of QM Hamiltonian. For the QM/MM simulation itself, the NWChem program package [54] was utilized, while the subsequent quantum-topological analysis of electron density for the QM subsystems was performed in the Multiwfn program [30].

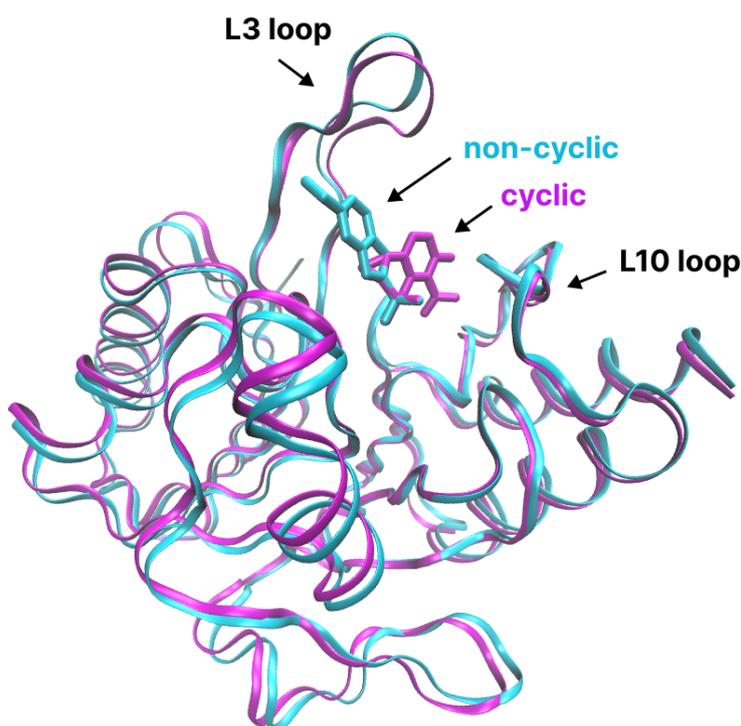
The QM/MM MD simulations of the water solutions of boronate molecules were carried out via interface [34] for the NAMD [43, 44] and TeraChem [47] software, which perform the classical MD and quantum chemical calculations, correspondingly. The inhibitor structures were obtained from the QM/MM calculations (for the non-cyclic inhibitor we used results obtained in our previous work [27]), then each system was solvated in the rectangular water box (distance from the inhibitor to the cell border equaled 15 Å) and neutralized. The preliminary 2 ns MD computation was performed before the 15 ps QM/MM MD simulation. The QM part included only the inhibitor molecule and was described at the PBE0-D3/6-31G(d,p) level of theory, while MM part contained water molecules with TIP3P parameters. Similarly, to the MD simulations, the NPT ensemble was chosen. The information about system energy, coordinates and velocities has been written every 1 fs. At the step of data analysis, first 5 ps of simulations were excluded from consideration. The atomic Fukui electrophilicity indices [38] were calculated using Hirshfeld charges [22] for 100 frames from QM/MM MD trajectory using Multiwfn code [30].

The benchmark studies were carried out for the MD simulations, since they were the most time-consuming part of all computations. The computational protocol was identical to the one used for the examination of NDM-1 conformation in its complex with the cyclic QPX7728 inhibitor. All other information, including the hardware specifications and the benchmark results, is discussed in the Section 2.3.

## 2. Results and Discussion

### 2.1. Cyclic and Non-cyclic Boronate Activity

The direct comparison of the NDM-1 crystal structures with cyclic (compound QPX7728) and non-cyclic boronate inhibitors (Tab. 1) demonstrate different location of the inhibitor molecule in the enzymatic active site (Fig. 2). The QPX7728 is located more closely to the L3 (Leu65–Val73) and L10 (Asp225–Asp254) loops, while its non-cyclic analog is directed outward the active site cavity. One can speculate that these distinctions in the orientation of the inhibitors can affect the dynamical behavior of the protein loops and, as a consequence, influence the stability of the enzyme-inhibitor complex.



**Figure 2.** Alignment of the crystal structures of the NDM-1 with the cyclic QPX7728 and non-cyclic boronate (PDB IDs: 6V1M [20] and 6Q2Y [10], correspondingly).  $IC_{50} \approx 55$  nM for the QPX7728 [53] and  $61 \mu\text{M}$  for the non-cyclic inhibitor [46]

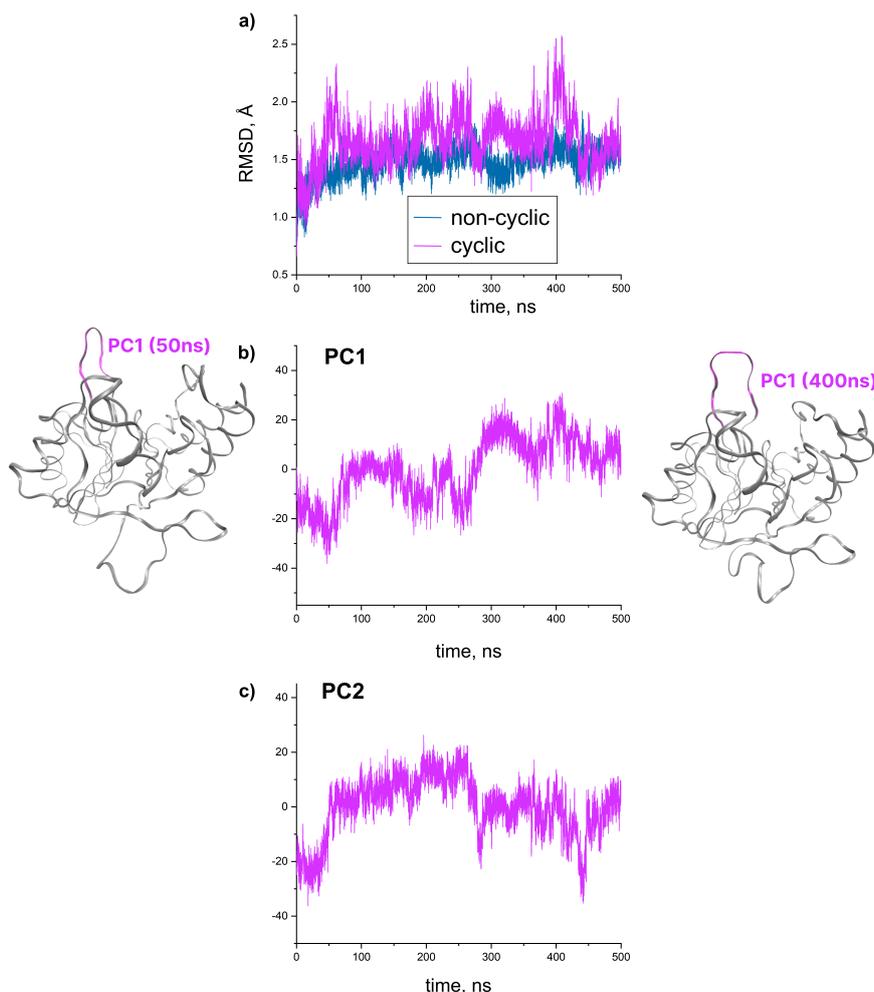
The results of the molecular dynamic simulation showed that both enzymes do not change conformation dramatically along the trajectory (Fig. 3 first panel). However, the RMSD for the complex with the non-cyclic inhibitor is substantially flatter than for the complex with the cyclic boronate. For the non-cyclic inhibitor, RMSD of the protein backbone vary within  $1 \text{ \AA}$ , while for the cyclic one the minimum and maximum values are separated by  $2 \text{ \AA}$ . This suggests that non-cyclic boronate induces higher stabilization of the flexible protein loops. To find out what structural movements are causing the RMSD bumps for the complex with the cyclic boronate, we performed a principal component analysis (PCA). It turned out that no principal component contains a large part of the overall fluctuations of the protein backbone (Tab. 2): first six components account for 50% of the total variance, while twenty PCs are needed to achieve 75%. These results confirm that overall protein has a stable conformation along the trajectory and RMSD variation is caused only by the movements in its flexible parts. Since each of the first

**Table 2.** The percentage of total variance, which is accounted for individual PCs and their cumulative variance

PC number	PC variance, %	Cumulated variance, %
1	16.0	16.0
2	12.2	28.2
3	8.6	36.8
4	6.5	43.3
5	4.7	48.0
6	3.4	51.4
7	3.0	54.4
8	2.3	56.7
9	2.3	59.0
10	2.0	61.0
11	1.7	62.7
12	1.6	64.3
13	1.6	65.9
14	1.3	67.2
15	1.1	68.3
16	1.2	69.5
17	1.0	70.5
18	0.9	71.4
19	0.9	72.3
20	0.9	73.2

two components contains more than 10% of the overall fluctuations, they were inspected more closely than others.

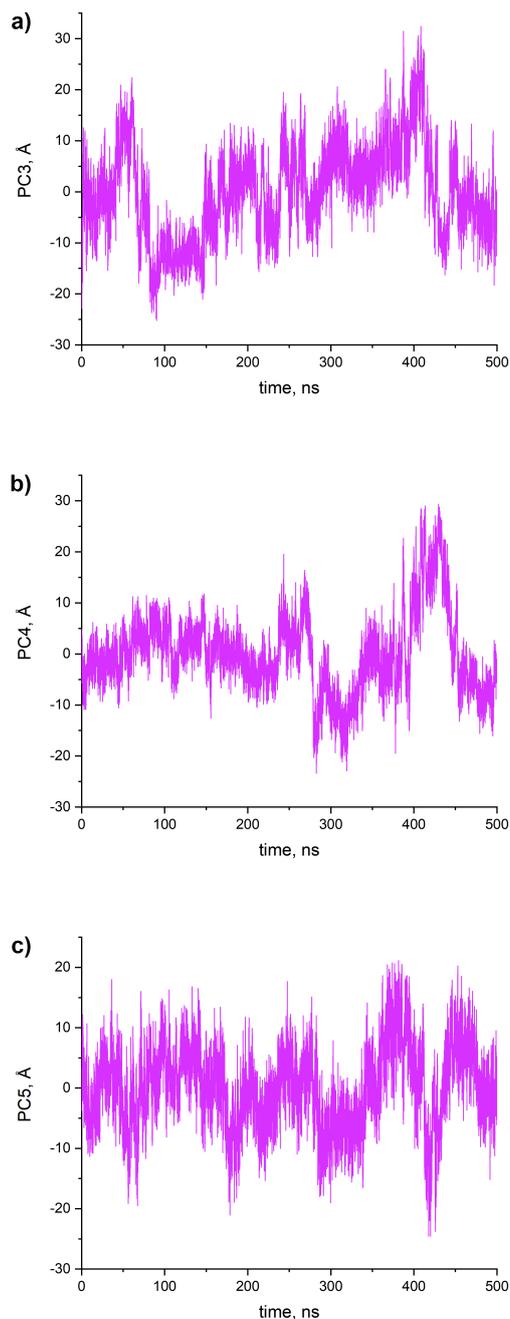
The PC1 motion is mainly connected with the L3 loop conformational changes and some motions at the bottom of the enzyme (far from the active site). At the 50 ns the L3 loop is maximally open, which corresponds to the minima in the amplitude of the fluctuations along the PC1 (Fig. 3, second panel). The pronounced RMSD peaks between 280 and 430 ns are also in the good agreement with the fluctuations along the PC1 and correspond to the loop “widening” (Fig. 3, second panel). The PC2 captures this “widening” as well, but does not reflect loops opening-closing motions. Fluctuations along the PC2 are mainly associated with N-terminus motion and are clearly visible in the 50–280 and 440–500 ns regions on the RMSD graph. Among other components that yield 5–8% of total variance (Fig. 4), only the fluctuations along the PC3 are partly connected with protein movements in the vicinity of the active site. Similar to the PC1, PC3 captures loop opening at  $\sim 50$  ns and its closing at  $\sim 100$  ns and some opening-closing movements around 400 ns. The visual inspection of the enzyme complex with non-cyclic inhibitor confirms that, in contrast to the complex with cyclic QPX7728, protein loops are less tended to change conformation. Notably that the L3 loop is almost always open along the simulation. Thus, the conducted analysis reveals that part of the RMSD variance along the trajectory for NDM-1 complex with cyclic boronate is rooted in the L3 loop movements, which is more flexible than for the structure with the non-cyclic boronate inhibitor.



**Figure 3.** The two first principal components calculated for the trajectory of NDM-1 with the cyclic QPX7728 inhibitor and RMSD for the enzymatic complexes of cyclic and non-cyclic boronates

The L3 loop, which movement is affected by the type of a boronate in the NDM-1 active site, consists of the Leu-Asp-Met-Pro-Gly-Phe-Gly-Ala-Val sequence. Thus, one can suppose that the bulky Phe70 interacts with the inhibitor when the loop is closed and serves as an obstacle to the inhibitor trying to leave the active site. To test this hypothesis, we compared the distance between the Phe70 side chain and the cyclic and non-cyclic boronates in the complexes with NDM-1. Since inhibitors have different positions in the NDM-1 active site (Fig. 2), we considered the distance between the Phe70 and both center-of-mass of the inhibitors and position of the boron atom itself, which is roughly the same for these two compounds. The obtained graphs (Fig. 5) clearly indicate that the Phe70 residue is located far from the non-cyclic boronate during simulation, except brief moments in the beginning of the trajectory and around 345 ns.

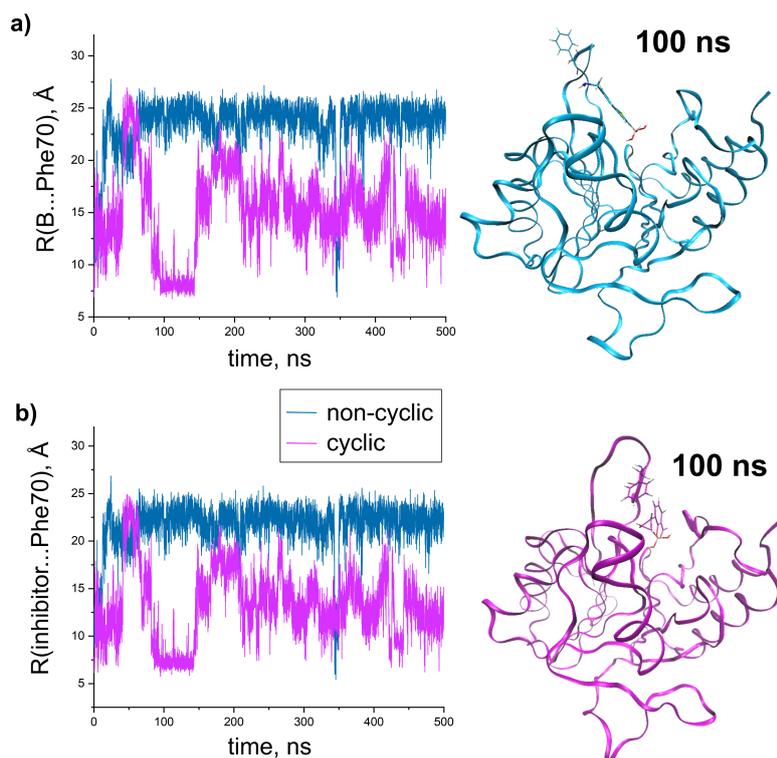
The inhibitor – Phe70 distance fluctuating around 25 Å for the complex of the non-cyclic boronate with NDM-1 testifies that the L3 loop is fixed in an open conformation. Hence, it cannot prevent the inhibitor from leaving active site. In the case of the cyclic boronate complex with NDM-1, the opening of the L3 is comparable to its opening in the case of the non-cyclic inhibitor only once per trajectory – at times close to 50 ns. All the rest of the time, it either



**Figure 4.** The principal components, which yield  $\sim 5\text{--}8\%$  of total variance (see Tab. 2), calculated for the trajectory of NDM-1 with cyclic QPX7728 inhibitor

oscillates around  $15 \text{ \AA}$  from the inhibitor (opening and closing), or it is stably closed above the active center (at the beginning of the trajectory or in the interval of 80–150 ns).

Such L3 loop behavior can definitely affect a potency of the boronate inhibitor against NDM-1. However, the inhibitor activity can be also dependent on the ability of the boronate fragment to bind catalytic  $OH^-$  ion in the enzyme active site (the boron atom electrophilicity) [27]. This raises a question whether the boron atom electrophilicity contribute to the enhanced activity of cyclic inhibitor against NDM-1 or its ability to adjust protein conformation is more important?



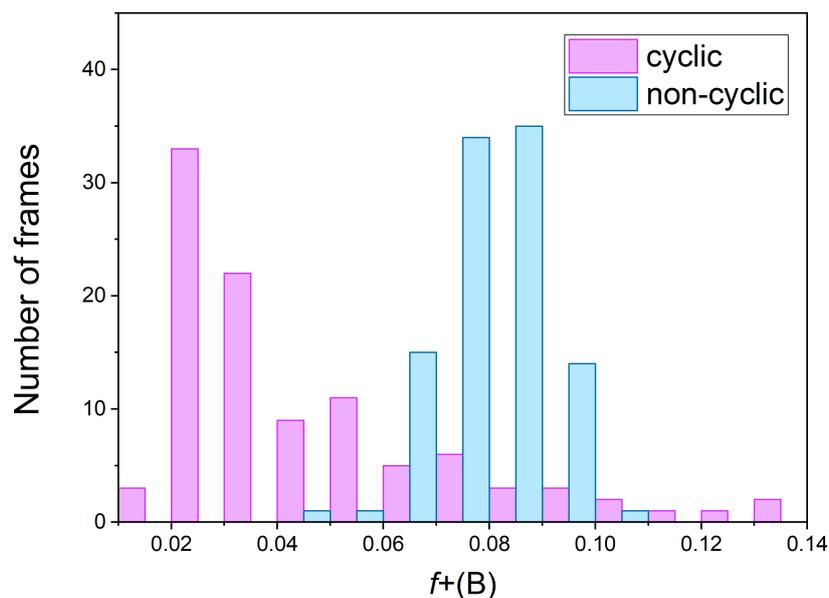
**Figure 5.** The distance between the center-of-mass of the phenyl ring in the Phe70 residue and a) the boron atom in the non-cyclic and cyclic (QPX7728) inhibitors, or b) the center of mass of the entire inhibitor. The snapshots on the right demonstrate conformations of the L3 loop and Phe70 for NDM-1 complexes with non-cyclic and cyclic inhibitors

From the computational point of view, the electrophilicity can be easily estimated via conceptual DFT approach, which defines the atomic Fukui electrophilicity indices,  $f^+$ , as a difference of the Hirshfeld atomic charges in the N and N+1 electronic states. However, this property of a boron atom is not a “static” one, and in order to argue about its greater or lesser value, it is necessary to consider numerous conformations of the inhibitors in water solution. The obtained distributions (Fig. 6) clearly demonstrate that non-cyclic boronate has more electrophilic boron atom despite its small potency as an M $\beta$ L inhibitor. Therefore, we can conclude that differences in the  $IC_{50}$  values by three orders for non-cyclic and cyclic inhibitors (Tab. 1) are caused by the dynamics of the L3 loop in the NDM-1 complex. The improved inhibitory activity of the cyclic boronate is explained by the tendency of this loop to cover NDM-1 active site, while for the non-cyclic boronate the active site stays open.

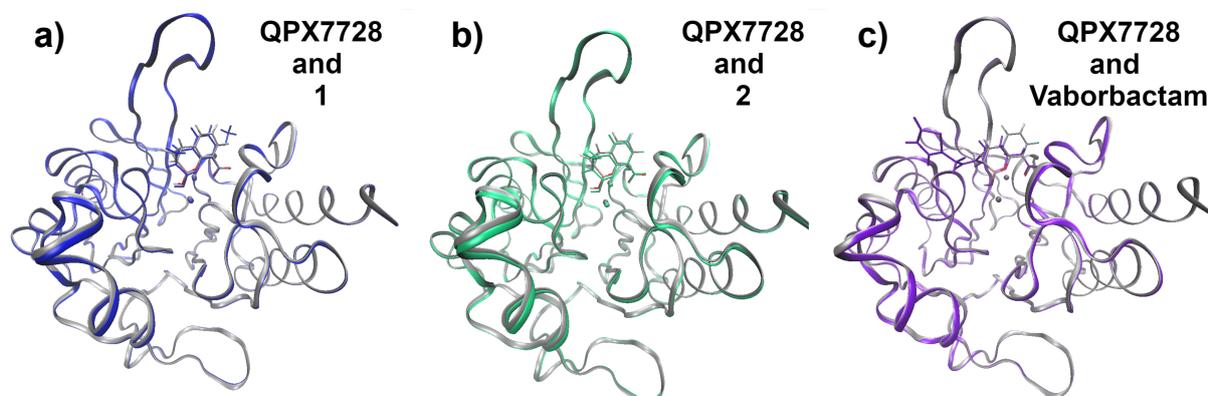
## 2.2. Origins of the Different Activity of the Cyclic Boronates

In contrast to the results of the previous section, differences in the inhibitory activity among the cyclic inhibitors (Tab. 1) cannot be explained by their influence on the enzyme conformation. It can be easily demonstrated by the superimposition of their structures, obtained in the QM/MM simulations (Fig. 7). All complexes of the QPX7728 derivatives with NDM-1 overlap perfectly with each other and no changes in the protein flexible regions are visible.

Despite the fact that cyclic inhibitors are similarly located in the active site, they differently interact with  $Zn^{2+}$  ions and amino acid residues of NDM-1 (Tab. 3). It can be revealed by the



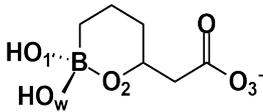
**Figure 6.** Distributions of the Fukui electrophilicity indices of the boron atom,  $f^+(B)$ , for the cyclic QPX7728 boronate and its non-cyclic analog (Tab. 1) in water solution along the QM/MM MD trajectories. The  $IC_{50}$  values are  $55 \pm 25$  nM [53]) for the cyclic inhibitor and  $60.7 \pm 11.8$   $\mu$ M [46] for the non-cyclic one



**Figure 7.** Superimposition of the QM/MM structures of NDM-1 complexes with cyclic boronates (Tab. 1)

quantum-topological analysis of electron density [5], which uses the electron-density saddle points (so-called bond critical points or BCPs) as a marker of the chemical bonding between atoms. This approach also allows one to characterize strength of interatomic interactions [14, 15, 17, 33, 48]. Applying the electron density value at BCP as a simple indicator of the interaction strength, it is easy to see, that despite dissimilarities in the bonding patterns between each of the cyclic inhibitors and protein, all considered compounds have approximately the same binding energy in the active site of the enzyme (Tab. 3). This fact indicates that only the properties of the inhibitor molecule itself can be the source of the different activity of cyclic boronates against NDM-1.

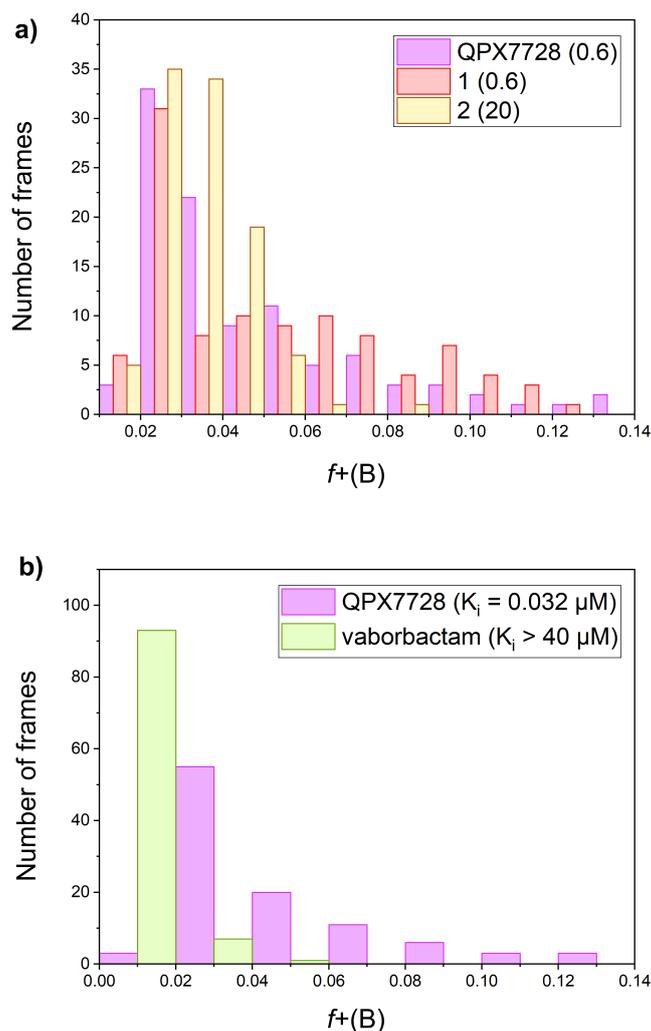
**Table 3.** Characteristics of intermolecular interactions between cyclic boronate inhibitors and the NDM-1 active site. Distances between inhibitors atoms and  $Zn^{2+}$  ions are given in parenthesis if no interatomic interaction was located from topological analysis of electron density.  $\Sigma\rho$  stands for the total electron density of the inhibitor-enzyme interactions, which included: (i) contacts between  $Zn^{2+}$  ions and a boronate molecule, (ii) conventional hydrogen bonds (H-bonds) between an inhibitor and NDM-1 amino acid residues, and (iii) weak C-H...O bonds, between an inhibitor molecule and NDM-1 active site. A schematic representation of a fragment of the boronate molecule involved in the interaction with  $Zn^{2+}$  ions is shown in the first row

	QPX7728	1	2	Vaborbactam
				
$Zn1^{2+}$				
$Zn2^{2+}$				
$R(Zn1^{2+} \dots O_1)$ , Å	– (2.7)	2.5	2.6	– (2.9)
$R(Zn1^{2+} \dots O_w)$ , Å	1.9	1.9	1.9	1.9
$R(Zn2^{2+} \dots O_1)$ , Å	– (4.1)	– (4.0)	– (4.0)	– (4.0)
$R(Zn2^{2+} \dots O_w)$ , Å	– (2.8)	2.6	– (2.8)	– (2.8)
$R(Zn2^{2+} \dots O_2)$ , Å	2.0	2.0	2.0	2.0
$R(Zn2^{2+} \dots O_3)$ , Å	2.2	2.3	2.2	2.2
$\Sigma\rho(Zn \dots inhibitor)$ , a.u.	0.22	0.26	0.25	0.24
$\Sigma\rho(H\text{-bonds})$ , a.u.	0.16	0.18	0.17	0.16
$\Sigma\rho(\text{all contacts})$ , a.u.	0.42	0.46	0.45	0.42

The obtained distributions of the  $f+(B)$  indices demonstrate that more efficient inhibitors have higher proportion of the conformations with more electrophilic boron atom, then less effective inhibitors (Fig. 8). The most dramatic difference can be seen for the QPX7728 and vaborbactam (Fig. 8b). The extremely low electrophilicity of the boron atom in the vaborbactam explains its zero activity against metallo- $\beta$ -lactamases and the presence of inhibitory potential against serine  $\beta$ -lactamases. Indeed, in contrast to serine  $\beta$ -lactamases, in metallo- $\beta$ -lactamases the boronate inhibitor does not covalently bind to the enzyme. After its reaction with  $OH^-$  inhibitor can leave the enzymatic active site, while the metallo- $\beta$ -lactamase can restore its activity. To prevent this, the metallo- $\beta$ -lactamase inhibitor must have a high electrophilicity of the boron atom in order to immediately bind its other molecule in the active site of the enzyme.

### 2.3. MD Benchmark Studies

The MD simulations of the complexes of boronate inhibitors with NDM-1 were found to be the most expensive part of our study of inhibitors activity. It was caused not by the cost of the computations themselves, but by the size of the systems under consideration and the trajectory lengths. Therefore, to optimize this part of the protocol for *in silico* prediction of inhibitors activity against metallo- $\beta$ -lactamases, we performed a benchmark study to select



**Figure 8.** Distributions of the Fukui electrophilicity indices of the boron atom,  $f^+(B)$ , for the cyclic boronates in water solution along the QM/MM MD trajectories. Data on inhibitory activity of boronate molecules against NDM-1 [20] are given in parenthesis. On the upper (a) frame it is expressed as the minimum concentration of the inhibitor ( $\mu\text{g}/\text{mL}$ ) needed to reduce the MIC of biapenem to the concentration that corresponds to the PK-PD breakpoint for that  $\beta$ -lactam based on human dosing

the optimal but widely available hardware for the MD simulation speedup. We examined the QPX7728 complex with NDM-1 in water solution, a system which contained 42938 atoms in total (see Section 1 for the computational protocol) and tested 4 GPUs (Tab. 4 and Tab. 5) with AMD Ryzen Threadripper 1950X and AMD Ryzen 9 3900X CPUs. The NAMD program package was chosen since it is a popular code for MD simulations of large biomolecular systems, which was among one of the first CUDA-accelerated applications. Main hardware and software specifications are presented in Tab. 4 and Tab. 5. Considering benchmark results for NAMD v2 (Tab. 5) one can notice that NVIDIA P102-100 GPU shows the worst results regardless of chosen CPU. The GeForce RTX 3070 and RTX 3070 Ti models demonstrate slight calculation speedup in comparison with GeForce GTX 1070 Ti GPU. However, the crucial improvement in computational speed can be achieved by the optimal hardware/software combination.

**Table 4.** Main characteristics of the considered NVIDIA GPUs

GPU model	Architecture	Release	Number of cores	Memory, GB	Theoretical performance, GFlops/s		Bandwidth, Gb/s
					SP	DP	
GeForce GTX 1070 Ti	Pascal	2017	2432	8	8186	255.8	256
GeForce RTX 3070	Pascal	2018	3200	5	10770	336.6	440
GeForce RTX 3070	Ampere	2020	5888	8	20310	317.4	448
GeForce RTX 3070 Ti	Ampere	2021	6144	8	21750	339.8	608

**Table 5.** Benchmark results for the MD simulations of the QPX7728 complex with NDM-1 obtained via NAMD v2 and 3 program [43, 44] with 1 CPU. The main hardware/software specifications are also presented

GPU	CPU	CUDA version	Motherboard	ns / day	
				NAMD v2	NAMD v3
NVIDIA GeForce GTX 1070 Ti	AMD Ryzen Threadripper 1950X	10.0	ASUS PRIME X399-A	10.5	-
NVIDIA P102-100	AMD Ryzen Threadripper 1950X	10.0	ASUS PRIME X399-A	6.4	-
	AMD Ryzen 9 3900X	9.1	ASRock X570 Pro4	7.7	57.8
NVIDIA GeForce RTX 3070	AMD Ryzen 9 3900X	9.1	ASRock X570 Pro4	14.9	81.1
NVIDIA GeForce RTX 3070 Ti	AMD Ryzen 9 3900X	9.1	ASRock X570 Pro4	15.1	84.1

Choosing between 2<sup>nd</sup> and 2<sup>rd</sup> version of the NAMD program one should note, that NAMD v2 uses GPU offloading for the force calculations and perform numerical integration on the CPUs. Despite the fact that GPU calculations constitute  $\sim 99\%$  of the overall performance, numerical integration can significantly slow down a simulation. The GPU-resident algorithm of NAMD v3 overcome this limitation and results in approximately 5x faster comput-

ing for the Ampere GPUs (Tab. 5). Even a full 12-core CPU load (AMD Ryzen 9 3900X) in the NAMD v2 calculations results only 54.7 ns/day performance (with NVIDIA GeForce RTX 3070 GPU), which is still lesser than for NAMD v3 calculations on GPUs with Ampere architecture. Therefore, the NAMD v3 simulations on modern Ampere GPUs are recommended to achieve maximal acceleration in classical MD simulations.

## Conclusion

In this work, we propose a methodology for *in silico* prediction of the activity of boronate inhibitors against metallo- $\beta$ -lactamase (M $\beta$ L) enzymes. These compounds reversibly bind  $Zn^{2+}$  ions in the M $\beta$ L active site, capturing catalytic  $OH^-$  ion from the M $\beta$ L. Our study of the NDM-1 inhibition by cyclic and non-cyclic boronates showed that, both enzyme conformation and inhibitors chemical properties can influence molecules potential as an M $\beta$ L inhibitor. Therefore, combination of various computer simulation methods (classical molecular dynamics, quantum mechanics/molecular mechanics and quantum-classical molecular dynamics) is required to study a potency of these compounds. To speed up some of these computations, we recommend GPU-accelerated calculations and propose detailed protocols for the hardware choice.

Since boronate inhibitors reversibly bind to the M $\beta$ L, the first step of their activity estimation should include an assessment of their ability to change the conformation of the protein flexible parts near its active site. Classical molecular dynamics (MD) simulations followed by principal component analysis showed that the non-cyclic boronate inhibitor fixes a flexible L3 loop in its open conformation, stipulating its exit from the enzyme. To speed up these expensive calculations one can chose NAMD v3 program package with Ampere GPUs, which demonstrated significant efficiency in this task.

For comprehensive analysis of inhibitor-protein binding, it is desirable to evaluate strength of M $\beta$ L-inhibitor interactions. It can be achieved by the quantum mechanics/molecular mechanics (QM/MM) calculations followed by the quantum-topological analysis of electron density in the QM part. However, in some cases this step can be omitted, since the protein conformation and properties of the inhibitor molecule were found to have more influence on the inhibitors activity against M $\beta$ L.

The electrophilicity of the boron atom, which binds catalytic  $OH^-$  ion in the enzyme active site, was found to be another aspect capable of strongly influencing the inhibitory activity against M $\beta$ L. To evaluate this property, one need to perform the quantum-classical molecular dynamics (QM/MM MD) modeling of aqueous solution of the inhibitor, followed by the calculation of the atomic Fukui electrophilicity index,  $f^+$ , for its boron atom along the obtained trajectory. These  $f^+$  distributions allowed us to interpret different activities of the cyclic boronates against NDM-1, clearly demonstrating that more potent inhibitors are better able to bind  $OH^-$  ion in the NDM-1 active site.

## Acknowledgements

This work was supported by the Russian Science Foundation (project No. 18-74-10056). We acknowledge the use of supercomputer resources of the Joint Supercomputer Center of the Russian Academy of Sciences and the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University [59].

*This paper is distributed under the terms of the Creative Commons Attribution-Non Commercial 3.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.*

## References

1. Gaussian09, Gaussian, Inc., Wallingford CT, 2016, <https://gaussian.com>
2. P1 Single and Multiple Ascending Dose (SAD/MAD) Study of IV QPX7728 Alone and Combined With QPX2014 in NHV, <https://clinicaltrials.gov/ct2/show/NCT04380207>
3. R.E.D. Server, <https://upjv.q4md-forcefieldtools.org/REDServer-Development/>
4. Adamo, C., Barone, V.: Toward reliable density functional methods without adjustable parameters: The PBE0 model. *The Journal of Chemical Physics* 110(13), 6158–6170 (1999). <https://doi.org/10.1063/1.478522>
5. Bader, R.F.W.: *Atoms in Molecules. Quantum Theory.* International series of monographs on chemistry, Clarendon Press (1990)
6. Best, R.B., Zhu, X., Shim, J., *et al.*: Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone  $\phi$ ,  $\psi$  and side-chain  $\chi_1$  and  $\chi_2$  dihedral angle. *Journal of Chemical Theory and Computation* 8(9), 3257–3273 (2012). <https://doi.org/10.1021/ct300400x>
7. Bonomo, R.A.:  $\beta$ -Lactamases: a focus on current challenges. *Cold Spring Harbor Perspectives in Medicine* 7(1), a025239 (2017). <https://doi.org/10.1101/cshperspect.a025239>
8. Bush, K.: Past and present perspectives on  $\beta$ -lactamases. *Antimicrobial Agents and Chemotherapy* 62(10), e01076–18 (2018). <https://doi.org/10.1128/AAC.01076-18>
9. Cahill, S., Cain, R., Wang, D.Y., *et al.*: Cyclic boronates inhibit all classes of  $\beta$ -lactamases. *Antimicrobial Agents and Chemotherapy* 61(4), e02260–16 (2017). <https://doi.org/10.1128/AAC.02260-16>
10. Cendron, L., Quotadamo, A., Maso, L., *et al.*: X-ray Crystallography Deciphers the Activity of Broad-Spectrum Boronic Acid  $\beta$ -Lactamase Inhibitors. *ACS Medicinal Chemistry Letters* 10(4), 650–655 (2019). <https://doi.org/10.1021/acsmchemlett.8b00607>
11. Cornell, W.D., Cieplak, P., Bayly, C.I., *et al.*: A second generation force field for the simulation of proteins, nucleic acids, and organic molecules *J. Am. Chem. Soc.* 1995, 117, 5179–5197. *Journal of the American Chemical Society* 118(9), 2309–2309 (1996). <https://doi.org/10.1021/ja955032e>
12. Dortet, L., Poirel, L., Nordmann, P.: Worldwide Dissemination of the NDM-Type Carbapenemases in Gram-Negative Bacteria. *BioMed Research International* 2014, 1–12 (2014). <https://doi.org/10.1155/2014/249856>
13. Dupradeau, F.Y., Pigache, A., Zaffran, T., *et al.*: The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building. *Physical Chemistry Chemical Physics* 12(28), 7821–7839 (2010). <https://doi.org/10.1039/C0CP00111B>

14. Espinosa, E., Molins, E.: Retrieving interaction potentials from the topology of the electron density distribution: The case of hydrogen bonds. *The Journal of Chemical Physics* 113(14), 5686–5694 (2000). <https://doi.org/10.1063/1.1290612>
15. Espinosa, E., Molins, E., Lecomte, C.: Hydrogen bond strengths revealed by topological analyses of experimentally observed electron densities. *Chemical Physics Letters* 285(3-4), 170–173 (1998). [https://doi.org/10.1016/S0009-2614\(98\)00036-0](https://doi.org/10.1016/S0009-2614(98)00036-0)
16. Gowers, R.J., Linke, M., Barnoud, J., *et al.*: MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. *Proceedings of the 15th Python in science conference* pp. 98–105 (2016). <https://doi.org/10.25080/Majora-629e541a-00e>
17. Grabowski, S.J.: Intramolecular Hydrogen Bond Energy and Its Decomposition-O-H · · O Interactions. *Crystals* 11(1), 22935–22952 (2020). <https://doi.org/10.3390/cryst11010005>
18. Grimme, S., Antony, J., Ehrlich, S., Krieg, H.: A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* 132(15), 154104 (2010). <https://doi.org/10.1063/1.3382344>
19. Hariharan, P.C., Pople, J.A.: The influence of polarization functions on molecular orbital hydrogenation energies. *Theoretica Chimica Acta* 28(3), 213–222 (1973). <https://doi.org/10.1007/BF00533485>
20. Hecker, S.J., Reddy, K.R., Lomovskaya, O., *et al.*: Discovery of Cyclic Boronic Acid QPX7728, an Ultrabroad-Spectrum Inhibitor of Serine and Metallo- $\beta$ -lactamases. *Journal of Medicinal Chemistry* 63(14), 7491–7507 (2020). <https://doi.org/10.1021/acs.jmedchem.9b01976>
21. Hehre, W.J., Ditchfield, R., Pople, J.A.: Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *The Journal of Chemical Physics* 56(5), 2257–2261 (1972). <https://doi.org/10.1063/1.1677527>
22. Hirshfeld, F.: Bonded-atom fragments for describing molecular charge densities. *The Journal of Physical Chemistry A* 44(2), 129–138 (1977). <https://doi.org/10.1007/BF00549096>
23. Huang, J., Rauscher, S., Nawrocki, G., *et al.*: CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* 14(1), 71–73 (2017). <https://doi.org/10.1038/nmeth.4067>
24. Humphrey, W., Dalke, A., Schulten, K.: VMD: Visual molecular dynamics. *Journal of Molecular Graphics* 14(1), 33–38 (1996). [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5)
25. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., *et al.*: Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* 79(2), 926–935 (1983). <https://doi.org/10.1063/1.445869>

26. Krajnc, A., Brem, J., Hinchliffe, P., *et al.*: Bicyclic boronate VNRX-5133 inhibits metallo- and serine- $\beta$ -lactamases. *Journal of Medicinal Chemistry* 62(18), 8544–8556 (2019). <https://doi.org/10.1021/acs.jmedchem.9b00911>
27. Krivitskaya, A.V., Khrenova, M.G.: Boronic Acids as Prospective Inhibitors of Metallo- $\beta$ -Lactamases: Efficient Chemical Reaction in the Enzymatic Active Site Revealed by Molecular Modeling. *Molecules* 26(7), 2026 (2021). <https://doi.org/10.3390/molecules26072026>
28. Lence, E., González-Bello, C.: Bicyclic Boronate  $\beta$ -Lactamase Inhibitors: The Present Hope against Deadly Bacterial Pathogens. *Advanced Therapeutics* 4(5), 2000246 (2021). <https://doi.org/10.1002/adtp.202000246>
29. Lomovskaya, O., Tsivkovski, R., Nelson, K., *et al.*: Spectrum of beta-lactamase inhibition by the cyclic boronate QPX7728, an ultrabroad-spectrum beta-lactamase inhibitor of serine and metallo-beta-lactamases: enhancement of activity of multiple antibiotics against isogenic strains expressing single beta-lactamases. *Antimicrobial Agents and Chemotherapy* 64(6), e00212–20 (2020). <https://doi.org/10.1128/AAC.00212-20>
30. Lu, T., Chen, F.: Multiwfn: A multifunctional wavefunction analyzer. *Journal of Computational Chemistry* 33(5), 580–592 (2012). <https://doi.org/10.1002/jcc.22885>
31. MacKerell, A.D., Bashford, D., Bellott, M., *et al.*: All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *The Journal of Physical Chemistry B* 102(18), 3586–3616 (1998). <https://doi.org/10.1021/jp973084f>
32. MacKerell, A.D., Feig, M., Brooks, C.L.: Improved Treatment of the Protein Backbone in Empirical Force Fields. *Journal of the American Chemical Society* 126(3), 698–699 (2004). <https://doi.org/10.1021/ja036959e>
33. Mata, I., Alkorta, I., Espinosa, E., Molins, E.: Relationships between interaction energy, intermolecular distance and electron density properties in hydrogen bonded complexes under external electric fields. *Chemical Physics Letters* 507(1-3), 185–189 (2011). <https://doi.org/10.1016/j.cpllett.2011.03.055>
34. Melo, M.C.R., Bernardi, R.C., Rudack, T., *et al.*: NAMD goes quantum: an integrative suite for hybrid simulations. *Nature Methods* 15(5), 351–354 (2018). <https://doi.org/10.1038/nmeth.4638>
35. Michaud-Agrawal, N., Denning, E.J., Woolf, T.B., Beckstein, O.: MDAAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* 32(10), 2319–2327 (2011). <https://doi.org/10.1002/jcc.21787>
36. Mojica, M.F., Rossi, M.A., Vila, A.J., Bonomo, R.A.: The urgent need for metallo- $\beta$ -lactamase inhibitors: an unattended global threat. *The Lancet Infectious Diseases* 22(1), e28–e34 (2022). [https://doi.org/10.1016/S1473-3099\(20\)30868-9](https://doi.org/10.1016/S1473-3099(20)30868-9)
37. Novais, A., Ferraz, R.V., Viana, M., *et al.*: NDM-1 Introduction in Portugal through a ST11 KL105 *Klebsiella pneumoniae* Widespread in Europe. *Antibiotics* 11(1), 92 (2022). <https://doi.org/10.3390/antibiotics11010092>

38. Oláh, J., Van Alsenoy, C., Sannigrahi, A.B.: Condensed Fukui Functions Derived from Stockholder Charges: Assessment of Their Performance as Local Reactivity Descriptors. *The Journal of Physical Chemistry A* 106(15), 3885–3890 (2002). <https://doi.org/10.1021/jp014039h>
39. Osei Sekyere, J.: Current State of Resistance to Antibiotics of Last-Resort in South Africa: A Review from a Public Health Perspective. *Frontiers in Public Health* 4, 209 (2016). <https://doi.org/10.3389/fpubh.2016.00209>
40. Palacios, A.R., Rossi, M.A., Mahler, G.S., Vila, A.J.: Metallo- $\beta$ -lactamase inhibitors inspired on snapshots from the catalytic mechanism. *Biomolecules* 10(6), a025239 (2020). <https://doi.org/10.3390/biom10060854>
41. Parkova, A., Lucic, A., Krajnc, A., *et al.*: Broad spectrum  $\beta$ -lactamase inhibition by a thioether substituted bicyclic boronate. *ACS Infectious Diseases* 6(6), 1398–1404 (2019). <https://doi.org/10.1021/acsinfecdis.9b00330>
42. Parkova, A., Lucic, A., Krajnc, A., *et al.*: A 2018-2019 patent review of metallo beta-lactamase inhibitors. *ACS Infectious Diseases* 30(7), 541–555 (2020). <https://doi.org/10.1080/13543776.2020.1767070>
43. Phillips, J.C., Hardy, D.J., Maia, J.D.C., *et al.*: Scalable molecular dynamics on CPU and GPU architectures with NAMD. *The Journal of Chemical Physics* 153(4), 044130 (2020). <https://doi.org/10.1063/5.0014475>
44. Phillips, L.C., Braun, R., Wang, W., *et al.*: Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* 26(16), 1781–1802 (2005). <https://doi.org/10.1002/jcc.20289>
45. Sabet, M., Tarazi, Z., Griffith, D.C.: *In vivo* activity of QPX7728, an ultrabroad-spectrum beta-lactamase inhibitor, in combination with beta-lactams against carbapenem-resistant *Klebsiella pneumoniae*. *Antimicrobial Agents and Chemotherapy* 64(11), e01267–20 (2020). <https://doi.org/10.1128/AAC.01267-20>
46. Santucci, M., Spyraakis, F., Cross, S., *et al.*: Computational and biological profile of boronic acids for the detection of bacterial serine- and metallo- $\beta$ -lactamases. *Scientific reports* 7(17716), 1–15 (2017). <https://doi.org/10.1038/s41598-017-17399-7>
47. Seritan, S., Bannwarth, C., Fales, B.S., *et al.*: TeraChem: A graphical processing unit-accelerated electronic structure package for largescale ab initio molecular dynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 11(2), e1494 (2021). <https://doi.org/10.1002/wcms.1494>
48. Shahi, A., Arunan, E.: Hydrogen bonding, halogen bonding and lithium bonding: an atoms in molecules and natural bond orbital perspective towards conservation of total bond order, inter- and intra-molecular bonding. *Physical Chemistry Chemical Physics* 16(4), 22935–22952 (2014). <https://doi.org/10.1039/C4CP02585G>
49. Shi, C., Chen, J., Kang, X., *et al.*: Approaches for the discovery of metallo- $\beta$ -lactamase inhibitors: A review. *Chemical Biology & Drug Design* 94(2), 1427–1440 (2019). <https://doi.org/10.1111/cbdd.13526>

50. Tehrani, K.H., Martin, N.I.:  $\beta$ -lactam/ $\beta$ -lactamase inhibitor combinations: an update. *Med-ChemComm* 9(9), 1439–1456 (2018). <https://doi.org/10.1039/c8md00342d>
51. Thapa, S., Adhikari, N., Shah, A.K., *et al.*: Detection of NDM-1 and VIM genes in carbapenem-resistant *Klebsiella pneumoniae* isolates from a tertiary health-care center in Kathmandu, Nepal. *Chemotherapy* 66(5-6), 199–209 (2021). <https://doi.org/10.1159/000518256>
52. Tooke, C.L., Hinchliffe, P., Bragginton, E.C., *et al.*: Lactamases and  $\beta$ -Lactamase Inhibitors in the 21st Century. *Journal of Molecular Biology* 431(18), 3472–3500 (2019). <https://doi.org/10.1016/j.jmb.2019.04.002>
53. Tsivkovski, R., Totrov, M., Lomovskaya, O.: Biochemical characterization of QPX7728, a new ultrabroad-spectrum beta-lactamase inhibitor of serine and metallo-beta-lactamases. *Antimicrobial Agents and Chemotherapy* 64(6), e00130–20 (2020). <https://doi.org/10.1128/AAC.00130-20>
54. Valiev, M., Bylaska, J., Govind, N., *et al.*: NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Computer Physics Communications* 181(9), 1477–1489 (2010). <https://doi.org/10.1016/j.cpc.2010.04.018>
55. Vanommeslaeghe, K., Hatcher, E., Acharya, C., *et al.*: CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* 31(4), 671–690 (2009). <https://doi.org/10.1002/jcc.21367>
56. Vanqualef, E., Simon, S., Marquant, G., *et al.*: R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Research* 39, W511–W517 (2011). <https://doi.org/10.1093/nar/gkr288>
57. Vázquez-Ucha, J.C., Rodríguez, D., Lasarte-Monterrubbio, C., *et al.*: 6-Halopyridylmethylidene Penicillin-Based Sulfones Efficiently Inactivate the Natural Resistance of *Pseudomonas aeruginosa* to  $\beta$ -Lactam Antibiotics. *Journal of Medicinal Chemistry* 64(9), 6310–6328 (2021). <https://doi.org/10.1021/acs.jmedchem.1c00369>
58. Venkata, K.C.N., Ellebrecht, M., Tripathi, S.K.: Efforts towards the inhibitor design for New Delhi metallo-beta-lactamase (NDM-1). *European Journal of Medicinal Chemistry* 225(5), 113747 (2021). <https://doi.org/10.1016/j.ejmech.2021.113747>
59. Voevodin, V.V., Antonov, A.S., Nikitenko, D.A., *et al.*: Supercomputer Lomonosov-2: Large Scale, Deep Monitoring and Fine Analytics for the User Community. *Supercomputing Frontiers and Innovations* 6(2), 4–11 (2019). <https://doi.org/10.14529/jsfi190201>
60. Yu, W., He, X., Vanommeslaeghe, K., MacKerell, A.D.: Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *Journal of Computational Chemistry* 33(31), 2451–2468 (2012). <https://doi.org/10.1002/jcc.23067>

# Predicting Binding Free Energies for DPS Protein-DNA Complexes and Crystals Using Molecular Dynamics

*Eduard V. Tereshkin*<sup>1</sup> , *Ksenia B. Tereshkina*<sup>1</sup> , *Yurii F. Krupyanski*<sup>1</sup> 

© The Authors 2022. This paper is published with open access at SuperFri.org

The interaction between deoxyribonucleic acid (DNA) and deoxyribonucleic acid-binding protein from starved cells (DPS) in bacterial cells leads to intracellular crystallization of the genetic material of bacteria, which contributes to the survival of bacteria under stress factors, including antibacterial agents. Molecular modeling can help explain the molecular mechanisms of DNA binding to this protein. In this paper, we report a supercomputer simulation of the molecular dynamics of several types DNA-DPS complexes and crystals ranging from DPS+DNA dimer to DNA in periodic crystal channels of *Escherichia coli* DPS protein using a coarse-grained Martini force field. By modeling DNA of 24 base pairs, comparable in size to the diameter of the DPS protein, we use the slow-growth thermodynamic integration method to find binding protein-DNA free energy and discuss the contribution of ions and the length of trajectories sufficient for this type of simulations. The results obtained are important for further research in the field of simulation of biological DNA-protein crystals and the study of the molecular mechanisms of DNA interaction with the DPS protein.

*Keywords:* molecular dynamics, slow-growth thermodynamic integration method, DPS protein, DNA stabilization, DNA-DPS binding free energy.

## Introduction

The formation of intracellular crystals of bacterial DNA in response to stress bears much interest both in biophysics and pharmaceuticals, as well as in industry. Despite the confirmation of the formation of the crystals by DPS proteins and their homologues, the search for DNA positions in such crystals remains of immediate importance. Bacteria have numerous strategies that allow them to survive unfavorable environmental conditions [21]. The preservation of a part of the bacterial colony under stressful conditions is the key to the survival of the population in the long term [39]. Primarily, survival depends on the cell's ability to retain its DNA. Unlike eukaryotic cells, bacterial cells do not have a system of histone proteins and unified mechanisms for chromosome condensation. However, prokaryotic DNA is also capable of compacting, undergoing hierarchical packaging [41]. The chromosomal DNA of bacteria is compactly folded into the so-called nucleoid, consisting of DNA, RNA and proteins, which differs sharply from the rest of the cytoplasm. Bacteria use a number of nucleoid-associated proteins (NAPs) that influence the lower-level organization of the nucleoid by bending, stiffening, bunching, wrapping, or bridging the DNA [2, 18, 33, 41]. Bacterial chromatin can change its shape and composition depending on the growth phase of bacterial population and the state of the cells. During prolonged starvation, DNA, together with bacterial NAPs, can form various condensed structures [17, 23]. One of the most notable strategies is the crystallization of the bacterial nucleoid with the help of DPS proteins (DNA-binding Proteins from Starved cells) and their homologues [42]. DPS was first isolated from cells of *Escherichia coli* (*E. coli*) under stress of starvation [1]. Studies of bacteria and archaea revealed a wide distribution of proteins of the DPS family and proteins homologous to them among the representatives of these domains. The crystal structure of *E. coli* DPS was first obtained in 1998 by X-ray diffraction analysis with a resolution of 1.6 Å [9]. DPS forms a

<sup>1</sup>N. N. Semenov Federal Research Center for Chemical Physics, Russian Academy of Sciences, Moscow, Russian Federation

dodecamer with 23 (tetrahedral) point group symmetry. It has a spherical shape with a cavity inside ( $\sim 9$  nm outer diameter,  $\sim 5$  nm inner diameter) and pores at the three-folds.

During the logarithmic phase of growth, DPS has been shown to be a minor component of the nucleoid ( $\sim 6$  thousand molecules per cell). In the stationary phase, its synthesis sharply increases, and DPS becomes the predominant nucleoid-associated protein ( $\sim 180$  thousand molecules per cell [14]. DPS can protect bacteria against multiple stresses during stationary phase of grow [27]. The conventional point of view that DPS binds DNA without sequence or structural specificity, but there is controversial data [3]. The structure of this protein and probable interactions with DNA was studied *in vitro* [6, 7, 9, 16, 26, 34]. *In silico* studies have shown how the DPS protein can form crystals and co-crystals with DNA under various conditions [36–38]. The long N-terminal regions and the surface of the protein are rich in lysine residues. Like the tails of eukaryotic histone proteins [12, 32], the N-terminals of DPS provide stability and conformational mobility of DNA, and also determine the binding of DNA to DPS protein complexes. However, the amino acid sequences and structures of bacterial NAPs and eukaryotic histones differ greatly. The spherical shape of the molecule allows the formation of crystals with different unit cell parameters and space groups, all of which have extended channels inside.

Presumably, during *in vivo* crystallization, DNA can be located inside the channels of DPS crystals and on the surface of DPS molecules. Unfortunately, it has not yet been possible to resolve the structure of DPS crystals in which the DNA coordinates would be obtained and the location of the DNA would be shown. Therefore, in this work, using computer simulation, we tried to obtain putative molecular structures of DPS clusters and crystals with DNA and make an attempt to evaluate the advantageous position of DNA in one or another of them. The aim of this work was to find favorable DNA positions relative to the DPS protein using molecular dynamics modeling by suitable protocols.

## 1. Methods

### 1.1. Building of the Molecular Models

Each subunit of *E. coli* DPS protein contains 167 amino acid residues. It consists of four long  $\alpha$ -helices parallel to each other and one small  $\alpha$ -helix perpendicular to them [5]. The N-terminal regions of the molecule consist of ordered and disordered parts. The ordered part of each N-terminus (amino acid residues numbered 14 till 25) can be easily deciphered by X-ray diffraction analysis [9, 15]. The three-dimensional structure of the disordered part of the N-terminus is not defined in any structure of the RCSB Protein Data Bank. As shown in our work [38], this is due to the extreme mobility of these regions of the molecule. N-termini and C-termini provide connection of DPS monomers in dodecamers, binding of dodecamers between themselves and with DNA. We have modelled a spatial structure of the DPS protein and its 2D and 3D crystals on the basis of the crystal structure 6GCM.pdb. Because the file lacks the positions of the flexible amino acid residues from the unordered part of all the N-termini, we added them using UCSF Chimera program package [30]. Each N-terminus was modeled separately rather than copying the same spatial structure for all 12 fragments [38].

Double-stranded linear DNA fragments, containing 24 base pairs, 5'-AAGTCGACCCTAGAGGATCTTTGT-3', were used to build the DNA-DPS complexes. Three-dimensional models of DPS crystals, DNA-DPS clusters and co-crystals were constructed

in the UCSF Chimera program package. The systems were placed in a periodic box. We modeled systems both with counterions only and with ions at concentrations corresponding to physiological concentrations in the *E. coli* cells [35]. It should be noted that the set of “ions” in the MARTINI force field is poor and consist not of usual ions, but of charged particles that imitate hydrated ions. Therefore, here we are only talking about concentrations, not about the exact ionic composition. Ion names are given according to *martini.v2.0.ions.itp* topology file. Counterions, namely 48 sodium ions per protein molecule and 46 sodium ions per one DNA molecule were added to the systems to maintain their electrical neutrality. Then the rest of the ions were added to the same system to obtain the ion concentration of *E. coli* cells. Since DPS crystals are large, the systems were studied in the coarse-grained approximation. Models were built and verified using the MARTINI 2.1\_DNA coarse-grained force field [20, 40]. We used *dssp* program for protein structure assignment. The Python scripts *martinize.py* [13] and *martinize-dna.py* were used to convert atomistic protein and DNA molecules, respectively, to MARTINI coarse grain structures and topologies. The topology files contained standard constraints used in the MARTINI force field to maintain the secondary structure elements such as  $\alpha$ -helices of protein. Elastic Network in Dynamics (elnedyn22 modification) was used for DNA [29, 40].

Due to the use of large particles when describing water molecules (one “molecule” of water in MARTINI corresponds to four all-atom water molecules), in the MARTINI force field, even with correctly selected molecular dynamics protocols, artifact freezing of water in the system is possible [10]. Depending on the simulation conditions, in the temperature range of 280K–300K [19], nucleation and freezing of the entire volume of water in the simulation box are observed. For our systems, we observed this process at 300K. To avoid this, we replaced at least 10% of the non-polarizable water beads (W) by anti-freeze [19] water beads (WF). The difference between “antifreeze” water and “ordinary” water is as follows. Non-bonded interactions of MARTINI particles are described by shifted Lennard-Jones 12-6 potential energy function (1)

$$U_{LJ}(r) = 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r} \right)^{12} - \left( \frac{\sigma_{ij}}{r} \right)^6 \right], \quad (1)$$

where  $\sigma_{ij}$  represents the closest distance of approach between two particles  $i$  and  $j$ ,  $\epsilon_{ij}$  is the strength of their interaction. For most interacting pairs, including pairs with ordinary water beads,  $\sigma_{ij} = 0.47$  nm. For water and antifreeze water interactions,  $\sigma_{W-WF} = 0.57$  nm, that makes it possible to disturb the lattice packing of the uniformly sized solvent particles. To avoid phase separating of antifreeze and other water particles,  $\epsilon_{W-WF} = 5.6$  kJ/mol vs  $\epsilon_{W-W} = \epsilon_{WF-WF} = 5.0$  kJ/mol. Antifreeze water beads are not distinguishable for particles of other types, because of the same values of  $\sigma$  and  $\epsilon$  as for ordinary water.

6GCM crystals contain an extensive network of channels. Three types of mutually orthogonal channels were found inside the crystal, surrounded (if we consider the cut in a plane perpendicular to the channel) by four DPS molecules: fairly wide channels of the first type (denoted as x) and narrow channels of the second (y) and the third types (z). Also in the crystal there are short channels surrounded by three DPS molecules (denoted as 3). It is assumed that the DNA in the channels is arranged randomly, filling the channels with bends from one channel to another. In this work, we studied a crystal in which short DNA fragments are packed into channels of all four types. DNA bends were not considered.

We examined both DNA in periodic crystal channels and DPS clusters. Clusters were obtained by cutting out 3 or 4 DPS molecules from the crystal, surrounding a channel. Table 1 informs on the composition of the various DNA and DNA-DPS systems. The DPS dodecamer contains 4344 coarse-grained particles, the double-stranded DNA molecule (24 base pairs) is built from 310 particles.

**Table 1.** The number of coarse-grained particles in simulated systems. In the *Name* column, 2D are two-dimensional periodic crystals, 3D are three-dimensional periodic crystals. Designations in brackets: 3, x, y, z are the names of the channels (see in the text), c/i – only counterions are added to the system. The number of molecules for DNA and DPS, and the number of coarse-grained particles for ions and water are shown. W – water, WF – antifreeze water

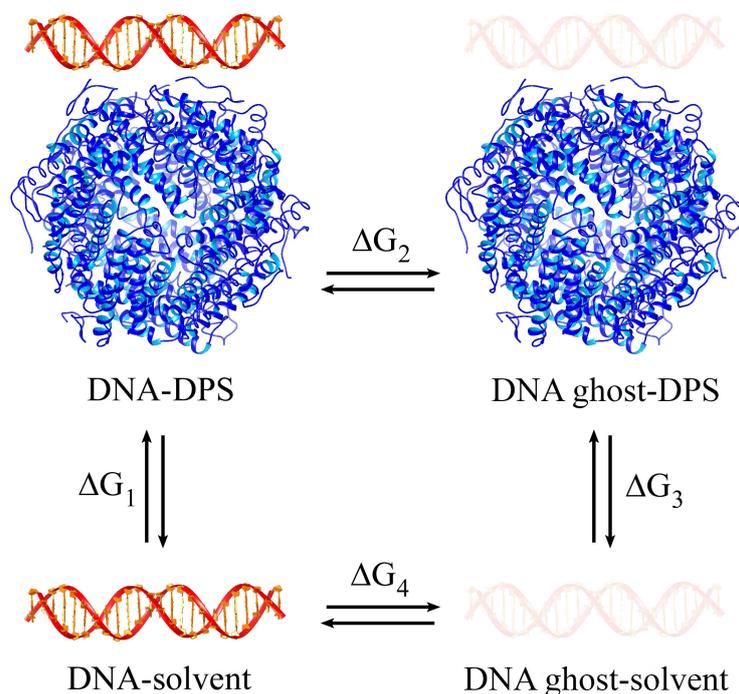
Name	DPS	DNA	NA <sup>+</sup>	CL <sup>-</sup>	CA <sup>2+</sup>	W	WF
DNA	-	1	46	-	-	13212	1468
DNA (c/i)	-	1	252	208	1	12838	1427
1DPS	1	1	816	740	4	34942	3882
1DPS (c/i)	1	1	94	-	-	36252	4028
2DPS (c/i)	2	1	142	-	-	94860	10540
3DPS (c/i)	3	1	190	-	-	97920	10880
4DPS (x)	4	1	2750	2544	16	124046	13782
4DPS (x, c/i)	4	1	238	-	-	128610	14290
4DPS (y, c/i)	4	1	238	-	-	115380	12820
2D	4	1	814	584	4	34352	3817
3D (3)	8	1	599	175	3	19227	2136
3D (x)	8	1	599	175	3	19240	2134
3D (x, c/i)	8	1	430	-	-	19517	2168
3D (y)	8	1	599	175	3	19282	2142
3D (z)	8	1	599	175	3	19280	2138

## 1.2. Molecular Dynamics Simulations

Molecular modeling of every system included two stages: classical dynamics simulation and the search for free energy by the slow-growth thermodynamics integration (TI) method. The first stage was to obtain stable DNA conformations at the protein surface and in solution. We followed our earlier established simulation protocols. Energy minimization using steepest descent algorithm followed by relaxation at constant volume (100 ps) and constant pressure (100 ps). Then we ran simulations in the NPT ensemble at 300 K and 1 bar controlled by means of a velocity Langevin thermostat [8] with a time constant of 1 ps and a Parrinello-Rahman barostat [28] with a time constant of 4 ps, respectively. The barostat provided isotropic pressure coupling for all systems, with the exception of two-dimensional periodic crystals, for which the pressure regulation was semi-isotropic. The isothermal compressibility of water was  $4.5 * 10^{-5} \text{ bar}^{-1}$ . The fastest degrees of freedom were removed by the parallel linear constraint solver. A cutoff radii of 1.2 nm was used for Coulombic and van der Waals interactions. The

integration step was 10 fs, providing trajectories up to 100000000 steps. The dielectric constant of the medium was equal to 15 for implicit screening. The simulations were carried out with the Gromacs 5.1 package [11].

Then we performed free energy simulations using TI. We were interested in the free energy differences between DNA in DNA-DPS clusters and crystals and DNA in solution (2). The binding free energy,  $\Delta G = -\Delta G_1$ , is a work required to transfer a DNA molecule from the solution (water+ions) into protein-DNA complex and was calculated according the thermodynamic cycle shown in Fig. 1. This cycle can be expanded by other contributions [24, 25]. In Fig. 1 DNA (24 base pairs) is shown in red. The DPS molecule is shown in blue; instead of a single DPS molecule, we also considered DPS clusters and crystals. In Fig. 1,  $\Delta G = -\Delta G_2$  is the work required to remove all the internal nonbonded (Coulombic and van der Waals) interactions of DNA in DPS cluster or crystal with surroundings. This was achieved by the gradual transformation of DNA atoms from fully interacting with the environment to completely non-interacting, which DNA state is designated in the figure as “DNA ghost”. Non-interacting atoms do not keep unbonded interactions, but completely keep their masses and intramolecular bonded interactions. In this case, we used a series of TI simulations. The work required to transfer the completely non-interacting DNA from DNA-protein complex to solution,  $\Delta G = -\Delta G_3$ . The term is effectively zero due to volume equality and lack of interaction with the surroundings.  $\Delta G = -\Delta G_4$  is the work required to turn DNA into a non-interacting “ghost” DNA in solution. To obtain it, we also used a series of TI simulations.



**Figure 1.** Thermodynamic cycle for the determination of binding free energies for DNA-DPS cluster. DPS protein is shown in blue, DNA is red. “DNA ghost” is the DNA that is completely non-interacting with surroundings

$$\Delta\Delta G = \Delta G_1 = \Delta G_4 - \Delta G_3 - \Delta G_2 \quad (2)$$

We chose the slow-growth thermodynamic integration method because it allows us to obtain DNA binding energies inside DPS crystals without destroying them. It is known that the Hamiltonian of the system in the slow-growth method becomes dependent on the coupling parameter,  $\lambda$ :  $H=H(p,q;\lambda)$ , where  $p$  and  $q$  are momenta and coordinates of the particles, respectively. In our simulations, Coulomb and Lennard-Jones terms were controlled by  $\lambda$  independently. The number of coupling parameter values varied to achieve computational efficiency. This will be described in more detail below. To ensure convergence when particles grow out of nothing, the regular Lennard-Jones and Coulomb potentials were modified with soft-core potentials [31]. The soft-core parameter was 0.5, the soft-core  $\lambda$  power was 1, the power of the radial term in the soft-core equation was 6 as recommended for the MARTINI force field. For each value of the coupling parameter, energy minimization and equilibrations were performed according to the protocol described above. The production run was simulated up to 30 ns. The frequency for writing  $dH/d\lambda$  was 0.1 ps. Bennetts Acceptance Ratio (BAR) method [4] was used for calculating values of  $\Delta G$  for transformations from two states using the *bar* module of Gromacs.

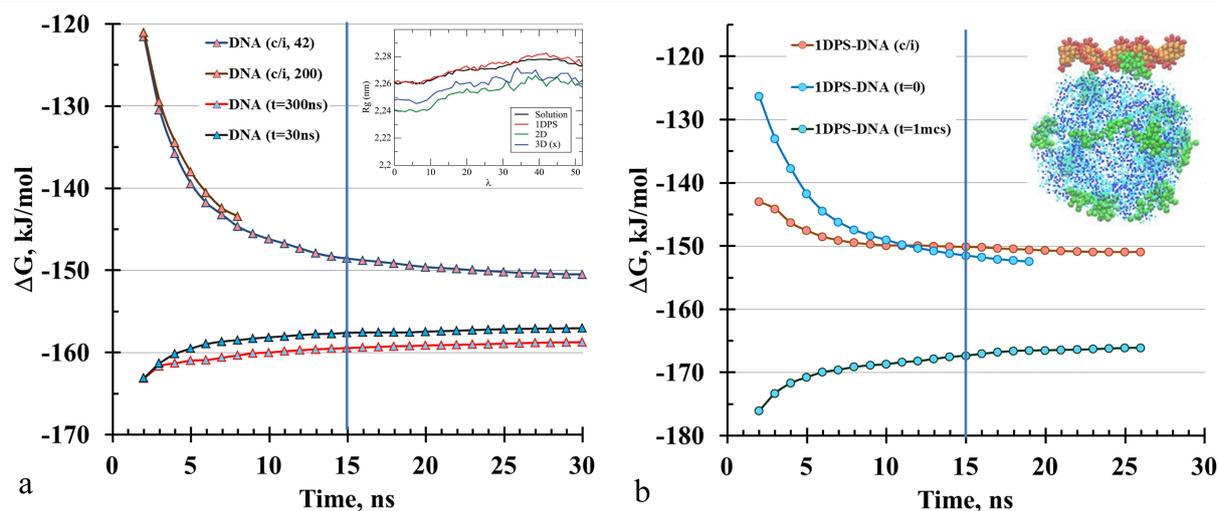
## 2. Results and Discussion

### 2.1. Free DNA and 1DPS-DNA Clusters

The systems of DNA in solution and DNA bound to single DPS molecule were considered to develop a protocol for calculating the Gibbs free energy. Using the coupling parameter, we gradually increased from zero the van der Waals interactions between DNA and the environment at the electrostatic interactions turned off. When the van der Waals interactions became fully turned on, we gradually increased the electrostatic interactions. This correspond to  $\Delta G = -\Delta G_4$  and  $\Delta G = -\Delta G_2$  calculation. Unless otherwise noted, simulations were made for of DNA configuration in solution at 30 ns and DNA with protein at 0.2  $\mu$ s of the classical trajectory.

In Fig. 2a one can see the  $\Delta G = -\Delta G_4/N$  depending on the simulation time of the TI trajectory.  $N = 24$  is the number of base pairs of DNA. For computational efficiency, we tried to find the minimum number of well-chosen values of the coupling parameter  $N(\lambda)$  by testing between 42 and 200  $\lambda$  values, and trajectory lengths sufficient to estimate the energies. We recalculated the values of the Gibbs energy every 1 ns starting from the trajectory time of 2 ns to track the output of the function to a plateau, the points are shown in the figure by triangles (DNA-solution) or circles (DNA-DPS). Curves with pink triangles (circles) correspond to simulations with counterions, with blue triangles to simulations with the *E. coli* cell concentration of ions (*E. coli* ions) [35].

DNA in solution and a DNA cluster with one DPS molecule with counterions was considered as the simplest model. Choosing  $\Delta\lambda=0.01$ , we simulated 200 trajectories with different values of  $\lambda$  (blue curve in Fig. 2a). Reducing  $N(\lambda)$  to 42 made it possible to increase the simulation performance by a factor of 5, while a change in the free energy value of only 1.5% was observed (brown curve in Fig. 2a). This value  $-2.2$  kJ/mol is less than the statistical error ( $\pm 3.1$  kJ/mol). It provides sufficient overlap of the phase space. A further decrease in  $N(\lambda)$  is impractical, as it leads to results that violate the second law of thermodynamics because of undersampling regardless of the length of the trajectory. Unfortunately, it is difficult to judge the quality of the obtained data due to the impossibility of obtaining the values of the free energy of solvation of this DNA by experimental methods. However, the obtained values are in good agreement with other data on the solvation of nucleic bases [22].



**Figure 2.** The change in the Gibbs energy for the transition of DNA from completely non-interacting to fully interacting with the solution (a) and with one DPS molecule in the solution (b). The insets show DNA radius of gyration (a) in different systems and the structure of a DNA cluster with one DPS molecule (b). DNA molecule is shown in red, DPS molecule is blue, mobile N-termini of DPS (1-20 terminal residues) are green

Simulations of DNA at *E. coli* ion concentration in solution and with DPS were chosen to demonstrate how the time of the original classical trajectory may affect free energy calculations. As starting points for TI, the conformations of the system were taken at two points of the classical trajectory, namely 30 ns (Fig. 2a black line) and 300 ns (Fig. 2a red line). The difference in free energy values turned out to be about 5% ( $-8.25$  kJ/mol). As it can be seen from the comparison of the blue and red curves, a solution with *E. coli* ions seems to be more energetically favorable for DNA than a solution with counterions. Inset in Fig. 2a shows that the radius of DNA gyration practically does not change from  $\lambda$ , regardless of the considered system (free DNA or DNA-protein).

Figure 2b shows that the value of the Gibbs energy ( $-\Delta G_2$ ) of a DNA cluster with one DPS molecule (see the structure in the inset) is affected by both the ionic composition and the adjustment of the ionic and protein environment. A change in the conformation of the protein may also be important, but verification of this was beyond the scope of this work. The starting point of preliminary dynamics simulation of DNA-DPS with *E. coli* ions is clearly insufficient to obtain acceptable value of the free energy, because the ionic environment does not have time to accurately rearrange (blue curve in Fig. 2b). The simulation with counterions after 100 ns dynamics (brown curve) gives a similar value. While the preliminary simulation of the system with *E. coli* ions for  $1 \mu\text{s}$  (black curve) brings lower values of the Gibbs energy. Thus, to obtain Gibbs energies, it is desirable to carry out a preliminary simulation of the dynamics of DNA in solution for at least 30 ns. As for DNA-protein complexes, a preliminary simulation of at least 100 ns is required. What about TI, it is desirable to simulate trajectories of at least 15 ns with  $N(\lambda)$  about 40 for DNA in solution and closer to 30 ns with  $N(\lambda)$  about 50 for DNA in DNA-DPS complexes.

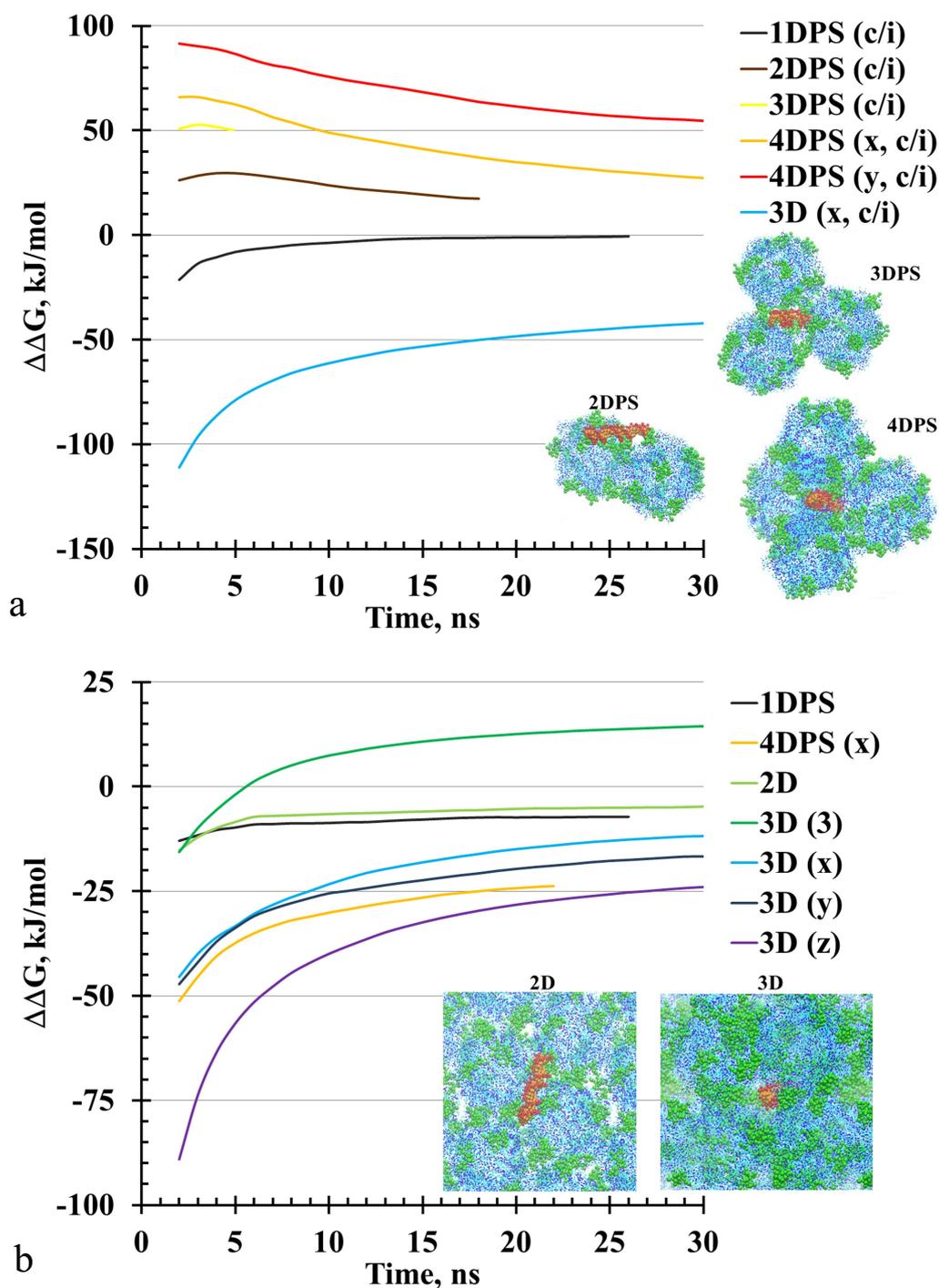
## 2.2. DNA-DPS Clusters and Crystals

Figure 3 shows the difference in Gibbs free energies according to equation (1), i.e. free energies of DNA-protein binding depending on the simulation time of TI. Average values per one base pair of DNA are given. Figure 3a shows curves for the systems with counterions, while Fig. 3b is for the *E. coli* ions. The structure of the DNA cluster with 1DPS is shown above in inset of Fig. 2b. The structure of DNA clusters with 2DPS, 3DPS, and 4DPS can be seen in Fig. 3a. Clusters of DNA-4DPS have three different shapes (denoted x, y, z) due to different positions of DPS molecules relative to each other. This is described in details in 1.1. DNA adsorbed on 2D-crystal of DPS and embedded in 3D-crystal are shown in Fig. 3b. In a crystal, DNA can be located in several directions, which corresponds to clusters with 3 DPS and 4 DPS, which is also detailed in 1.1. The composition of the systems is given in Tab. 1.

In Fig. 3a, the black curve corresponds to the free energy of DNA binding to one DPS molecule and becomes close to zero already after 15 ns of simulation and approaches the value  $\Delta\Delta G = -0.63 \pm 2.12$  kJ/mol in 26 ns. When clusters are enlarged (by adding DPS molecules), we obtain even positive values of  $\Delta\Delta G$  up to  $\sim 50$  kJ/mol per base pair, i.e., such a process is energetically unfavorable (brown, yellow, orange and red curves in Fig. 3a). However, as it can be seen from the blue curve, the presence of DNA inside the DPS crystal is favorable with  $\Delta\Delta G = -44.12 \pm 6.07$  kJ/mol. As it was shown in [38], DNA stabilizes DPS crystals due to the formation of additional interactions in the system (N-terminus of one DPS molecule)-DNA-(another DPS molecule). The black curve in Fig. 3b (DNA cluster with 1DPS) lies in the negative region. That is, in the presence of ions, the formation of such a cluster becomes advantageous with  $\Delta\Delta G = -7.21 \pm 1.32$  kJ/mol. Slightly less favorable is the adsorption of DNA on a 2D-crystals of DPS (light green curve in Fig. 3b),  $\Delta\Delta G = -4.77 \pm 1.36$  kJ/mol. It should be noted that during the formation of a cluster with one DPS molecule, only one N-terminus is involved in DNA binding. While DNA adsorption on the crystal surface, several N-termini are involved (green DPS regions in the insets). Channels bordered by three DPS molecules (green curve in Fig. 3b), as well as clusters of three DPS molecules, are unfavorable for DNA ( $\Delta\Delta G$  15kJ/mol). Presumably, these channels cannot serve as the main “warehouses” of DNA, but possibly can serve for DNA as transition regions of crystals between other channels. The presence of ions favors the growth of clusters (orange line in Fig. 3b). Binding free energy of DNA with 4 molecules of DPS is  $\Delta\Delta G = -23.76 \pm 4.39$  kJ/mol. The binding free energies of DNA in the main (blue line in Fig. 3b) and additional (blue and purple lines) channels of DPS 6GCM protein crystals are negative and equal, respectively,  $\Delta\Delta G = -11.82 \pm 5.2$  kJ/mol,  $-16.7 \pm 4.29$  kJ/mol, and  $-23.96 \pm 4.63$  kJ/mol. Nevertheless, these values are lower by absolute value than for DNA in the presence of counterions given above. Thus, the formation of DNA-DPS clusters is more favorable and possible in the presence of ions, while the retention of DNA in a bound state is more favorable at low ionic strength.

## Conclusions

In this work, we have studied the binding free energies of DNA with the DPS protein in different clusters and channels of 6PCM.pdb crystals of *Escherichia coli* DPS protein. In order to obtain acceptable values of free energy when simulating in GROMACS 5.1 with the MARTINI force field, two-stage simulations were carried out. At the first stage, we simulated the dynamics on trajectories up to  $0.2 \mu\text{s}$  (up to  $1 \mu\text{s}$  for a DNA cluster with one DPS molecule). It was



**Figure 3.** Binding Gibbs free energies of DNA (per base pair) in DNA-DPS complexes and crystals for simulation with counterions (a) and normal concentration of ions in *E. coli* cells (b). The insets show the structure of DNA-DPS clusters and crystals. DNA molecules are shown in red, DPS molecules are blue, mobile N-termini are green

shown, that the trajectories for simulation of DNA dynamics  $\sim 30$  ns are quite enough, while for DNA-DPS complexes should be not less than  $0.1 \mu\text{s}$  and longer depending on the system. Even for already formed DNA-DPS complexes, it is important to wait for the complete adjustment of the ionic environment, which requires rather long trajectories for such systems. Subsequent simulations by slow-growth thermodynamic integration method can be performed for DNA at

~40 points of the binding parameter. For computer-efficient calculation of the free energies of binding of DNA in DNA-DPS complexes (clusters and crystals), it seems optimal to choose ~50 points of the binding parameter. It has been shown that a solution with an ion concentration corresponding to that in the *E. coli* cells may be more favorable for DNA binding into DNA-protein clusters, while a solution with low ionic strength may be more optimal for co-crystals.

## Acknowledgements

The computations were carried out on MVS-10P at Joint Supercomputer Center of the Russian Academy of Sciences (JSCC RAS). This work was supported within frameworks of the state task for N. N. Semenov Federal Research Center for Chemical Physics RAS FFZE-2022-0011 (State Registration No. 122040400089-6).

*This paper is distributed under the terms of the Creative Commons Attribution-Non Commercial 3.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.*

## References

1. Almiron, M., Link, A.J., Furlong, D., Kolter, R.: A novel DNA-binding protein with regulatory and protective roles in starved escherichia coli. *Genes & Development* 6(12b), 2646–2654 (dec 1992). <https://doi.org/10.1101/gad.6.12b.2646>
2. Amemiya, H.M., Schroeder, J., Freddolino, P.L.: Nucleoid-associated proteins shape chromatin structure and transcriptional regulation across the bacterial kingdom. *Transcription* 12(4), 182–218 (aug 2021). <https://doi.org/10.1080/21541264.2021.1973865>
3. Antipov, S.S., Tutukina, M.N., Preobrazhenskaya, E.V., *et al.*: The nucleoid protein Dps binds genomic DNA of Escherichia coli in a non-random manner. *PLOS ONE* 12(8), e0182800 (aug 2017). <https://doi.org/10.1371/journal.pone.0182800>
4. Bennett, C.H.: Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics* 22(2), 245–268 (oct 1976). [https://doi.org/10.1016/0021-9991\(76\)90078-4](https://doi.org/10.1016/0021-9991(76)90078-4)
5. Chiancone, E., Ceci, P.: The multifaceted capacity of dps proteins to combat bacterial stress conditions: Detoxification of iron and hydrogen peroxide and DNA binding. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1800(8), 798–805 (aug 2010). <https://doi.org/10.1016/j.bbagen.2010.01.013>
6. Dadinova, L.A., Chesnokov, Y.M., Kamyshinsky, R.A., *et al.*: Protective Dps-DNA co-crystallization in stressed cells: an in vitro structural study by small-angle X-ray scattering and cryo-electron tomography. *FEBS Letters* 593(12), 1360–1371 (2019). <https://doi.org/10.1002/1873-3468.13439>
7. Frenkiel-Krispin, D., Minsky, A.: Nucleoid organization and the maintenance of DNA integrity in *E. coli*, *B. subtilis* and *D. radiodurans*. *Journal of Structural Biology* 156(2), 311–319 (nov 2006). <https://doi.org/10.1016/j.jsb.2006.05.014>

8. Goga, N., Rzepiela, A.J., de Vries, A.H., *et al.*: Efficient algorithms for langevin and DPD dynamics. *Journal of Chemical Theory and Computation* 8(10), 3637–3649 (jun 2012). <https://doi.org/10.1021/ct3000876>
9. Grant, R.A., Filman, D.J., Finkel, S.E., *et al.*: The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nature Structural Biology* 5(4), 294–303 (apr 1998). <https://doi.org/10.1038/nsb0498-294>
10. Hadley, K.R., McCabe, C.: Coarse-grained molecular models of water: a review. *Molecular Simulation* 38(8-9), 671–681 (jul 2012). <https://doi.org/10.1080/08927022.2012.671942>
11. Hess, B., Kutzner, C., van der Spoel, D., Lindahl, E.: GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation* 4(3), 435–447 (feb 2008). <https://doi.org/10.1021/ct700301q>
12. Huertas, J., Cojocar, V.: Breaths, twists, and turns of atomistic nucleosomes. *Journal of Molecular Biology* 433(6), 166744 (mar 2021). <https://doi.org/10.1016/j.jmb.2020.166744>
13. de Jong, D.H., Singh, G., Bennett, W.F.D., *et al.*: Improved Parameters for the Martini Coarse-Grained Protein Force Field. *Journal of chemical theory and computation* 9(1), 687–697 (2013). <https://doi.org/10.1021/ct300646g>
14. Karas, V.O., Westerlaken, I., Meyer, A.S.: The DNA-binding protein from starved cells (dps) utilizes dual functions to defend cells against multiple stresses. *Journal of Bacteriology* 197(19), 3206–3215 (oct 2015). <https://doi.org/10.1128/jb.00475-15>
15. Kovalenko, V., Popov, A., Santoni, G., *et al.*: Multi-crystal data collection using synchrotron radiation as exemplified with low-symmetry crystals of Dps. *Acta Crystallographica Section F Structural Biology Communications* 76(11), 568–576 (oct 2020). <https://doi.org/10.1107/s2053230x20012571>
16. Krupyanski, Y.F., Loiko, N.G., Sinityn, D.O., *et al.*: Biocrystallization in bacterial and fungal cells and spores. *Crystallography Reports* 63(4), 594–599 (jul 2018). <https://doi.org/10.1134/s1063774518040144>
17. Loiko, N., Danilova, Y., Moiseenko, A., *et al.*: Morphological peculiarities of the DNA-protein complexes in starved Escherichia coli cells. *PLOS ONE* 15(10), e0231562 (oct 2020). <https://doi.org/10.1371/journal.pone.0231562>
18. Luijsterburg, M.S., White, M.F., van Driel, R., Dame, R.T.: The major architects of chromatin: Architectural proteins in bacteria, archaea and eukaryotes. *Critical Reviews in Biochemistry and Molecular Biology* 43(6), 393–418 (jan 2008). <https://doi.org/10.1080/10409230802528488>
19. Marrink, S.J., Risselada, H.J., Yefimov, S., *et al.*: The MARTINI force field: coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry B* 111(27), 7812–7824 (2007). <https://doi.org/10.1021/jp071097f>

20. Marrink, S.J., Tieleman, D.P.: Perspective on the Martini model. *Chemical Society Reviews* 42(16), 6801 (2013). <https://doi.org/10.1039/c3cs60093a>
21. Martinez, J.L.: The role of natural environments in the evolution of resistance traits in pathogenic bacteria. *Proceedings of the Royal Society B: Biological Sciences* 276(1667), 2521–2530 (apr 2009). <https://doi.org/10.1098/rspb.2009.0320>
22. Miller, J.L., Kollman, P.A.: Solvation free energies of the nucleic acid bases. *The Journal of Physical Chemistry* 100(20), 8587–8594 (jan 1996). <https://doi.org/10.1021/jp9605358>
23. Minsky, A., Shimoni, E., Frenkiel-Krispin, D.: Stress, order and survival. *Nature Reviews Molecular Cell Biology* 3(1), 50–60 (jan 2002). <https://doi.org/10.1038/nrm700>
24. Mobley, D.L., Chodera, J.D., Dill, K.A.: On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. *The Journal of Chemical Physics* 125(8), 084902 (aug 2006). <https://doi.org/10.1063/1.2221683>
25. Mobley, D.L., Chodera, J.D., Dill, K.A.: Confine-and-release method: Obtaining correct binding free energies in the presence of protein conformational change. *Journal of Chemical Theory and Computation* 3(4), 1231–1235 (may 2007). <https://doi.org/10.1021/ct700032n>
26. Moiseenko, A., Loiko, N., Tereshkina, K., *et al.*: Projection structures reveal the position of the DNA within DNA-Dps Co-crystals. *Biochemical and Biophysical Research Communications* 517(3), 463–469 (sep 2019). <https://doi.org/10.1016/j.bbrc.2019.07.103>
27. Nair, S., Finkel, S.E.: Dps protects cells against multiple stresses during stationary phase. *Journal of Bacteriology* 186(13), 4192–4198 (jul 2004). <https://doi.org/10.1128/jb.186.13.4192-4198.2004>
28. Parrinello, M., Rahman, A.: Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* 52(12), 7182–7190 (dec 1981). <https://doi.org/10.1063/1.328693>
29. Periole, X., Cavalli, M., Marrink, S.J., Ceruso, M.A.: Combining an elastic network with a coarse-grained molecular force field: Structure, dynamics, and intermolecular recognition. *Journal of chemical theory and computation* 5(9), 2531–2543 (2013). <https://doi.org/10.1021/ct300646g>
30. Pettersen, E.F., Goddard, T.D., Huang, C.C., *et al.*: UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* 25(13), 1605–1612 (2004). <https://doi.org/10.1002/jcc.20084>
31. Pham, T.T., Shirts, M.R.: Optimal pairwise and non-pairwise alchemical pathways for free energy calculations of molecular transformation in solution phase. *The Journal of Chemical Physics* 136(12), 124120 (mar 2012). <https://doi.org/10.1063/1.3697833>
32. Shaytan, A.K., Armeev, G.A., Goncarencu, A., *et al.*: Coupling between histone conformations and DNA geometry in nucleosomes on a microsecond timescale: Atomistic insights into nucleosome functions. *Journal of Molecular Biology* 428(1), 221–237 (jan 2016). <https://doi.org/10.1016/j.jmb.2015.12.004>

33. Shen, B.A., Landick, R.: Transcription of bacterial chromatin. *Journal of Molecular Biology* 431(20), 4040–4066 (sep 2019). <https://doi.org/10.1016/j.jmb.2019.05.041>
34. Sinitsyn, D.O., Loiko, N.G., Gularyan, S.K., *et al.*: Biocrystallization of bacterial nucleoid under stress. *Russian Journal of Physical Chemistry B* 11(5), 833–838 (sep 2017). <https://doi.org/10.1134/s1990793117050128>
35. Szatmári, D., Sárkány, P., Kocsis, B., *et al.*: Intracellular ion concentrations and cation-dependent remodelling of bacterial MreB assemblies. *Scientific Reports* 10(1) (jul 2020). <https://doi.org/10.1038/s41598-020-68960-w>
36. Tereshkin, E.V., Tereshkina, K.B., Kovalenko, V.V., *et al.*: Structure of DPS protein complexes with DNA. *Russian Journal of Physical Chemistry B* 13(5), 769–777 (sep 2019). <https://doi.org/10.1134/s199079311905021x>
37. Tereshkin, E.V., Tereshkina, K.B., Krupyanskii, Y.F.: Molecular dynamics of DNA-binding protein and its 2D-crystals. *Journal of Physics: Conference Series* 2056(1), 012016 (oct 2021). <https://doi.org/10.1088/1742-6596/2056/1/012016>
38. Tereshkin, E., Tereshkina, K., Loiko, N., *et al.*: Interaction of deoxyribonucleic acid with deoxyribonucleic acid-binding protein from starved cells: cluster formation and crystal growing as a model of initial stages of nucleoid biocrystallization. *Journal of Biomolecular Structure and Dynamics* 37(10), 2600–2607 (nov 2018). <https://doi.org/10.1080/07391102.2018.1492458>
39. Tkachenko, A.G.: Stress responses of bacterial cells as mechanism of development of antibiotic tolerance (review). *Applied Biochemistry and Microbiology* 54(2), 108–127 (mar 2018). <https://doi.org/10.1134/s0003683818020114>
40. Uusitalo, J.J., Ingólfsson, H.I., Akhshi, P., *et al.*: Martini coarse-grained force field: Extension to DNA. *Journal of chemical theory and computation* 11(8), 3932–3945 (2015). <https://doi.org/10.1021/acs.jctc.5b00286>
41. Verma, S.C., Qian, Z., Adhya, S.L.: Architecture of the escherichia coli nucleoid. *PLOS Genetics* 15(12), e1008456 (dec 2019). <https://doi.org/10.1371/journal.pgen.1008456>
42. Wolf, S.G., Frenkiel, D., Arad, T., *et al.*: DNA protection by stress-induced biocrystallization. *Nature* 400(6739), 83–85 (jul 1999). <https://doi.org/10.1038/21918>

# Computational Modeling of the Interaction of Molecular Oxygen with the Flavin-dependent Enzyme RutA

Igor V. Polyakov<sup>1,2</sup> , Tatiana M. Domratcheva<sup>1</sup> , Anna M. Kulakova<sup>1</sup> ,  
Alexander V. Nemukhin<sup>1,2</sup> , Bella L. Grigorenko<sup>1,2</sup> 

© The Authors 2022. This paper is published with open access at SuperFri.org

Supercomputer molecular modeling methods are applied to characterize structure and dynamics of the flavin-dependent enzyme RutA in the complex with molecular oxygen. Following construction of a model protein system, molecular dynamics (MD) simulations were carried out using either classical force field interaction potentials or the quantum mechanics/molecular mechanics (QM/MM) potentials. Several oxygen-binding pockets in the protein cavities were located in these simulations. The QM/MM-based MD calculations rely on the interface between the quantum chemistry package TeraChem and the MD package NAMD. The results show a stable localization of the oxygen molecule in the enzyme active site. Static QM/MM calculations carried out with two different packages, NWChem and TURBOMOLE, allowed us to establish the structure of the RutA-O<sub>2</sub> complex. Biochemical perspectives of the hallmark reaction of incorporating oxygen into organic compounds emerged from these simulations are formulated.

*Keywords:* computational modeling, molecular dynamics, quantum mechanics/molecular mechanics, protein-oxygen interaction, flavin-dependent enzymes.

## Introduction

Modeling vitally important reactions of biomolecules with molecular oxygen is one of the most demanding subjects in life sciences. In particular, it is necessary to account for activation of the initially chemically inert molecule O<sub>2</sub> in the triplet spin state in order to obtain chemically reactive species. Computer simulations of elementary steps of the oxidation reactions require application of various modeling tools, including construction of all-atom three-dimensional model systems, preliminary relaxation of geometry parameters with the help of large-scale classical molecular dynamics (MD) simulations, and application of the quantum-based methods to locate stationary points on the potential energy surfaces of different spin multiplicity to estimate transition energies between electronic states, and to scan energy profiles along properly selected reaction coordinates. Application of the quantum mechanics/molecular mechanics (QM/MM) theory is an essential step in these simulations, which require high-performance supercomputer facilities [21].

An additional requirement in studies of the oxygen-binding proteins is to characterize the accommodation of O<sub>2</sub> in the protein cavities, because the localization of oxygen in these sites is a nontrivial experimental task. Only few structures in the Protein Data Bank [7] report complexes of proteins with molecular oxygen, e.g., [5, 10, 18]. It should be also pointed out that the resulting experimental electron density is produced using computational refinement [24] before deposition to the structure databanks. Thus, computer simulations are indispensable for revealing structures of the oxygen-binding proteins. In this work, we demonstrate the achievements and the problems of the required computational procedures when considering interaction of O<sub>2</sub> with the flavin-dependent enzyme RutA from *Escherichia coli* as an important example.

Flavin-binding monooxygenase RutA, catalyzing critical steps in degradation of uracil [16], has been characterized by performing O<sub>2</sub> pressurized crystallography experiments and a specific

<sup>1</sup>Department of Chemistry, Lomonosov Moscow State University, Moscow, Russian Federation

<sup>2</sup>Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, Moscow, Russian Federation

O<sub>2</sub> binding site has been suggested on the basis of these experiments [18]. RutA and other monooxygenases employ flavin cofactors derived from riboflavin (vitamin B2) in the form of flavin mononucleotide (FMN) or flavin-adenine dinucleotide that share the reactive isoalloxazine ring and adopt various redox states of the system [26]. In general, the interaction of O<sub>2</sub> with flavin cofactors in flavoproteins is of a particular importance for biochemistry [1, 17].

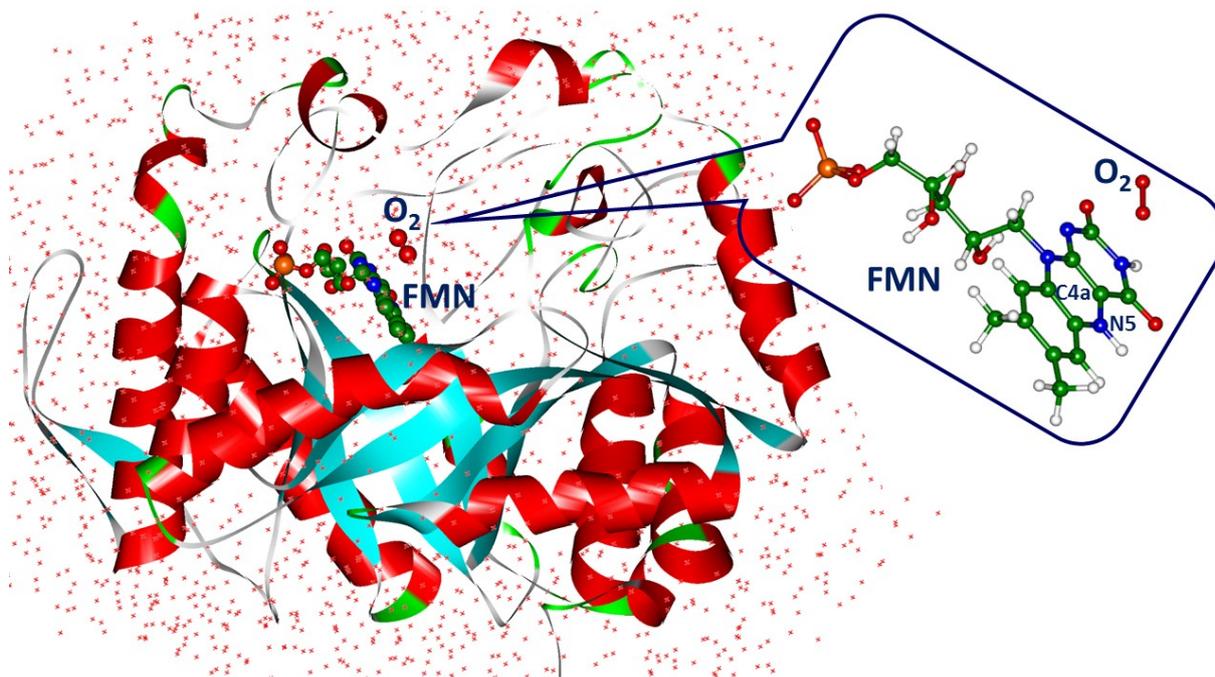
## 1. Molecular Model Setup

We explore the recently published crystal structure PDB ID 6SGG [18] as an initial guess of atomic coordinates of heavy atoms in the construction of the model system. According to the reported data, the protein structure contains the non-covalently bound FMN cofactor and the oxygen molecule near the isoalloxazine ring of FMN. In our simulations, hydrogen atoms were added assuming the conventional protonation states of the polar residues Arg, Lys (positively charged), Glu, Asp (negatively charged), at neutral pH. The histidine residues were assumed in the neutral state. The loop Glu292 - Ala310, which is poorly defined in the crystal structure, was manually refined. Topology files for the protein for MD simulations were constructed using the psfgen toolkit of NAMD [22]. The force field parameters correspond to the CHARMM36 library [8]. Parameters of the FMN molecule in the reduced form (with the total charge -3 e) were taken from Ref. [3] and those for the molecular oxygen were taken from Ref. [29]. Water molecules resolved in the crystal structure were kept in the model system, while solvation water shells were built using the visual molecular dynamics (VMD) plugins [13], the TIP3P water model was employed. The model system composed of 42174 atoms in total.

The constructed model system is illustrated in the main frame in Fig. 1 showing the secondary structure of the RutA protein ( $\alpha$ -helices and  $\beta$ -sheets), the FMN molecule without hydrogen atoms, and the oxygen molecule. The inset in Fig. 1 shows the all-atom model of FMN and the O<sub>2</sub> molecule near the isoalloxazine ring. The atoms of the latter, N5 and C4a, are considered as competing sites for flavin oxidation, according to the current dispute [18].

Classical MD simulations were carried out using the NAMD 3.0 software package [22]. The isothermal-isobaric (NPT) ensemble at  $P = 1$  atm and  $T = 300$  K using the Nosé-Hoover Langevin piston pressure control and the Langevin dynamics were employed. Integration step was set to 1 fs. Periodic boundary conditions and the particle mesh Ewald algorithm to account for the long-range electrostatic interactions were applied, whereas the non-bonded interaction cut-off parameter was set to 12 Å. The solvation water box was equilibrated in a short 1 ns calculation while keeping the protein heavy atoms fixed. The use of NAMD allowed us to efficiently utilize the DGX-2 supercomputer. The node configuration included Dual Intel Xeon Platinum 8168, 1.5 TB RAM, 16 NVIDIA Tesla V100 GPUs (NVSwitch interconnect) and a 30 TB NVMe RAID0 SSD storage. We ran a swarm of 8 trajectories simultaneously on a single DGX-2 node, while only 8 CPU cores were utilized. This is possible with the new GPU-resident version of NAMD 3.0 with the CUDASOAintegrate option set to “on”. Our benchmark showed that at maximum, 16 parallel trajectories can be executed to exploit all the V100 GPUs of a DGX-2 node, which results in a performance of over 1000 ns/day for model systems considered in this work.

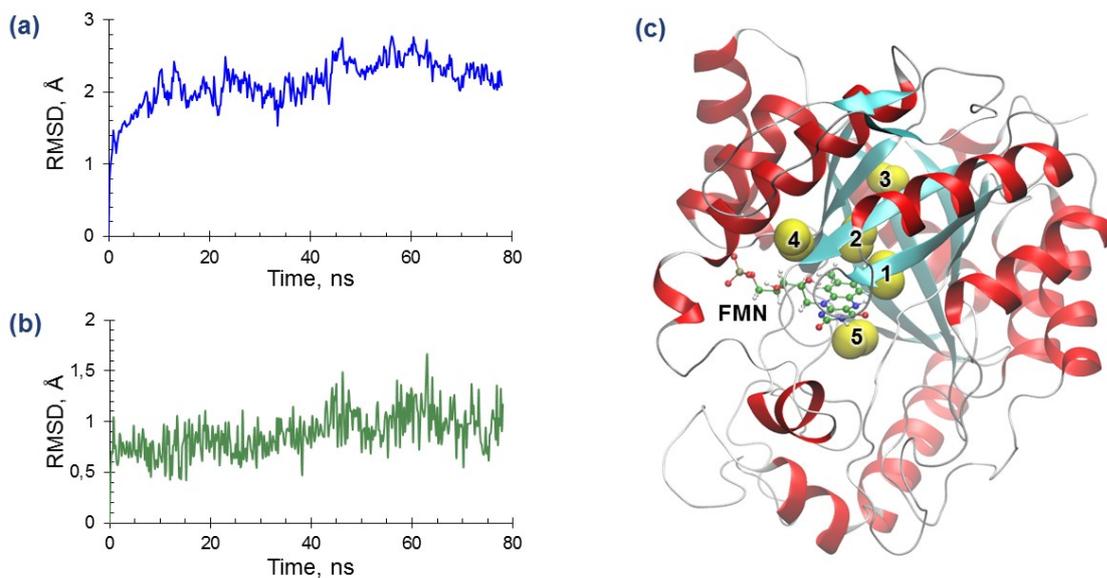
Panels in Fig. 2 illustrate the main conclusions, which can be drawn from classical MD simulations. Panel (a) reports that the overall protein structure remains stable throughout the classical MD trajectories, the graph shows the root mean square deviation (RMSD) calculated



**Figure 1.** Computationally derived model system of the RutA protein with the FMN cofactor and the oxygen molecule. Conventionally,  $\alpha$ -helices of the protein are shown in red,  $\beta$ -sheets in cyan. The red dots surrounding the protein refer to solvation water shells. The inset in the right part shows the FMN- $O_2$  species with the hydrogen atoms. Carbon atoms are colored green, oxygen – red, nitrogen – blue, phosphorus – orange, hydrogen – white. C4a and N5 atoms form a covalent bond

over the CA atoms of the amino acid residues in a typical run. Panel (b) reports that the non-covalently bound FMN cofactor remains in a single stable conformation as well.

The most important finding refers to the location of the  $O_2$  molecule. The calculations show (see Fig. 2) that the molecule is localized in different protein pockets in MD runs. Pocket-1 roughly corresponds to the arrangement identified in the crystal structure PDB ID 6SGG. Here, the oxygen molecule fluctuates in the region between the flavin moiety, polar residues Asn134, Thr105 and a hydrophobic cluster of Leu65, Met67 and Phe25 sidechains. The retention time of oxygen in any pocket varies from hundreds of picoseconds to tens of nanoseconds. We notice that along the MD trajectories, the oxygen species tends to move from pocket-1 to pocket-2 or pocket-5. In pocket-5, the  $O_2$  molecule is surrounded by water molecules and resides closer to the C4a atom of the flavin molecule (see Fig. 1). Pocket-2 is mostly hydrophobic; it is formed by the sidechains of Asn134, Phe224, Ala206, Leu65 and flavin. Pocket-3 consists of the sidechains Phe222, Phe224, Tyr257, Ile204, Phe63; it is close to the protein outer surface, showing one of the obvious ways connecting the pocket-1 (identified in the crystal structure) to the bulk,  $1 \rightarrow 2 \rightarrow 3$ . Pocket-4 is formed by the FMN phosphate group, the Tyr160, Cys205 sidechains and nearby main chain of  $\beta$ -strands. From the practical point of view, classical MD simulations with the CHARMM36 force field described here are feasible for conformational sampling in this complicated biomolecular system. However, these simulations lack to describe an important feature of this system, that is to model a partial electron transfer from the reduced FMN moiety to the  $O_2$  species [5]. Therefore, we consider MD simulations for this complex with the QM/MM potentials, that is the QM/MM MD approach.



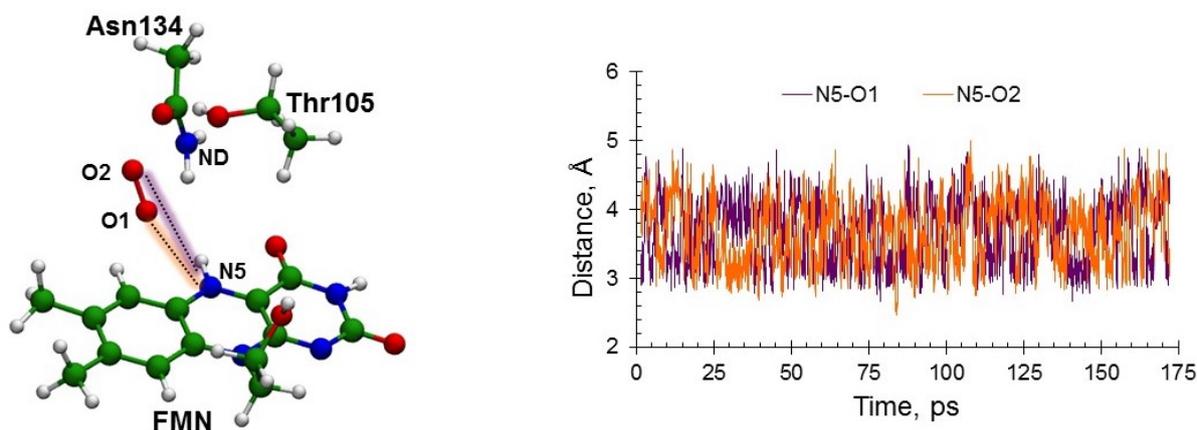
**Figure 2.** The results obtained in classical MD simulations: (a) RMSD over the CA atoms during a classical MD trajectory; (b) RMSD over the heavy atoms of FMN (in both cases, a, and b, the alignment was performed for the protein backbone in all the frames relative to the crystal structure PDB ID 6SGG); (c) the located pockets in protein cavities for the  $O_2$  molecule shown by pairs of large yellow balls

## 2. QM/MM MD Simulations

The QM/MM MD trajectories were initiated from the frames of classical MD in the vicinity of pocket-1. We assigned to the quantum subsystem the oxygen molecule and the atoms from the molecular groups from the isoalloxazine ring of FMN, Asn134 and Thr105 side chains (see Fig. 3). The QM part consisted of 56 atoms in total, and it was treated at the density functional theory (DFT) level with the range-separated  $\omega$ B97X functional [9] and the D3 dispersion correction [12]. The 6-31G\*\* basis set was employed for carbon and hydrogen atoms and 6-31+G\*\* for nitrogen and oxygen atoms. The QM subsystem was treated in the triplet state due to the ground electronic state of  $O_2$  using the unrestricted DFT approach. The remaining atoms of the model system were assigned to the MM part described by the CHARMM36 [8] force field. Presently, MD simulations with the QM/MM potentials benefit greatly from the GPU systems because of the use of the software stack of TeraChem [23] and NAMD [22] with the appropriate interfaces. The originally implemented interface [19] had some pitfalls, which were recently fixed [14].

The present simulations were carried out on 8 GPUs in parallel resulting in a combined 172 ps trajectory. The measured performance was  $\approx 2.5$  ps per day on a single GPU or 40 ps per single DGX-2 node.

The important results of these simulations are as follows. First, we observe a significant electron density transfer to the  $O_2$  molecule from FMN. For instance, the amount of the Mulliken charge on the  $O_2$  moiety was up to 0.4 e, when O1-N5 or O2-N5 distances were less than 3 Å. Second, the graph of distances between the oxygen atoms in  $O_2$  and the N5 atom in the isoalloxazine ring of FMN shown in the right part in Fig. 3 demonstrates that the oxygen molecule stably resides in the protein. Third, this graph predicts a N5-O1/O2 distance of 3.5 Å,



**Figure 3.** Left – quantum subsystem in the QM/MM MD calculations. Right – graph of the distances between the oxygen atoms in  $O_2$  and the N5 atom in the isoalloxazine ring of FMN along the QM/MM MD trajectories

which is significantly longer than the distance 2.09 Å observed in the parent crystal structure PDB ID 6SGG [18]. Hence, we observe a clear discrepancy in the experimentally determined and computed O-N5 distances.

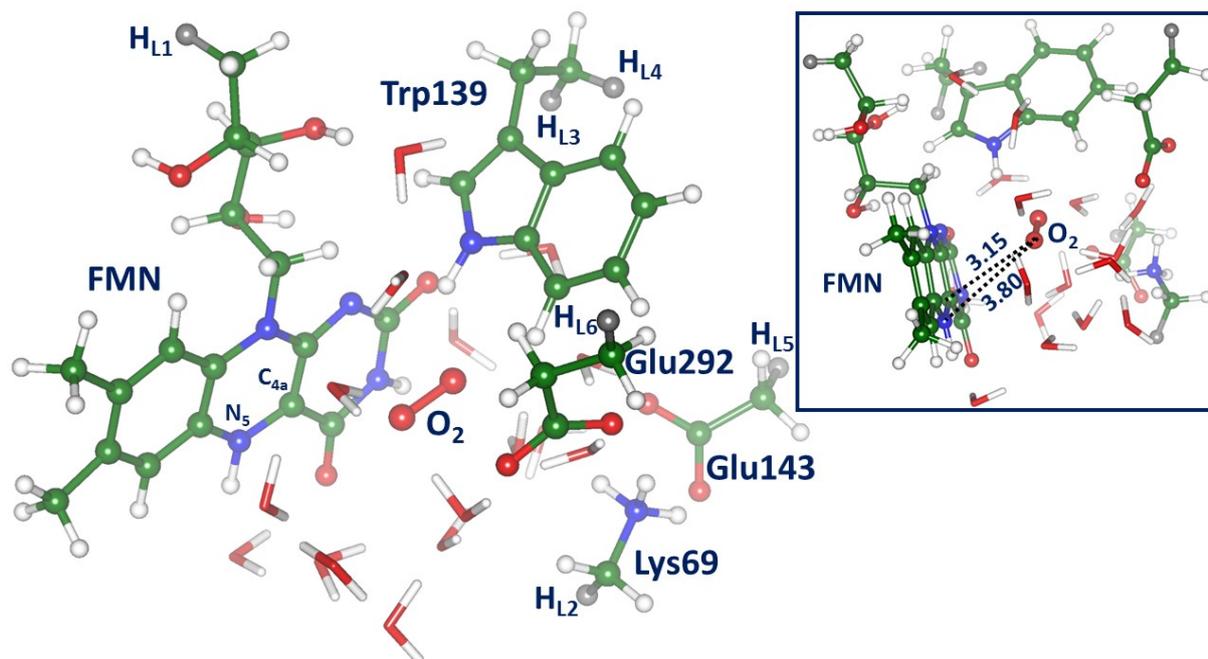
### 3. QM/MM Calculations

In this subsection, we report the results of static QM/MM calculations carried out to locate the minimum energy points on the potential energy surface of the RutA- $O_2$  complexes using two different quantum chemistry software packages. We note that QM/MM simulations are less demanding in computational efforts as compared to the QM/MM MD approach. For instance, a micro iterative structure optimization in QM/MM requires approximately  $10^2$  energy-gradient calculations, whereas a mere 10 ps QM/MM MD trajectory requires  $10^4$  such calculations. Nevertheless, QM/MM simulations, which rely on *ab initio*-type quantum chemistry approaches in QM, are still quite expensive for biomolecular systems composed of thousands of atoms as in our case.

First, we describe the use of a modern computer based on a consumer-grade 16-core Ryzen 5950X CPU. We applied the software stack of Tcl ChemShell [25] with the efficient DL-FIND optimizer [15] and TURBOMOLE quantum chemistry software package [6]. ChemShell was compiled on Ubuntu Linux 18.04 using gfortran/gcc version 7.5, OpenMPI v. 2.1.1. Preliminary testing showed that this strategy is very efficient for QM/MM calculations.

In the TURBOMOLE/ChemShell calculations, the same QM-MM partitioning was assumed for the protein with the FMN and  $O_2$  cofactors as in the previously described QM/MM MD simulations using TeraChem/NAMD. The MM part was treated with the CHARMM36 force field, but no cut-off or periodical boundary conditions were imposed unlike the case of MD calculations. The protein solvation box was trimmed to a water monolayer; therefore, the system contained 11445 atoms in total. No protein residues, ligands or water molecules were frozen during energy minimization. The typical step of the QM energy and gradient evaluation in the field of MM charges took  $\approx 1.5$ –2 minutes to calculate with 16 OpenMP threads using TURBOMOLE, while DL-FIND optimizer iteration (including the MM energy and gradient evaluation) takes 0.3 seconds with 16 MPI processes. Thus, a typical calculation takes around

6 hours of wall time to reach convergence. Using all 16 cores allows for significant speedup of 10 times compared to a single core for both the TURBOMOLE and DL-FIND. If the starting conformation for the minimum energy search is taken from a MD frame with the oxygen molecule in pocket-1 (see Fig. 2), then the QM/MM minimum corresponds to the structure shown in the left side in Fig. 3 with the following distances: 3.05 Å for ND(Asn134)-O<sub>2</sub>, 1.20 Å for O1-O2, 3.09 Å for O1-N5(FMN). The latter value is the most important result of this set of calculations. Consistently with the results of QM/MM MD simulations described above, this shows that the distance between any of the oxygen atom from O<sub>2</sub> to the N5 of FMN cannot be as low as 2.09 Å as reported in the crystal structure. The second approach to characterize the RutA-O<sub>2</sub> minimum energy structure using the QM/MM method was to apply the NWChem software package [4] implemented at the Lomonosov-2 supercomputer [27]. That NWChem 6.6 local installation was compiled with the ifort and icc (Intel Parallel Studio 2019) while using the MKL2019 and openmpi-4.0.1 libraries. Calculations were carried out utilizing 6–12 nodes per task (84 to 168 MPI processes). Few novel options were introduced as compared to the QM/MM calculations with TURBOMOLE/ChemShell. We considerably enlarged the QM-subsystem by including the almost entire FMN molecule, the oxygen molecule, fragments of the side chains of Lys 69, Trp139, Glu143, Glu292 and 15 water molecules (see the main frame in Fig. 4). We took the initial conformation for the structure optimization from the MD frame corresponding to the location of oxygen in pocket-5 of the complex (see Fig. 2). Now the selected groups in QM completely cover the interacting groups if the oxygen molecule is placed at the center of the QM-part. All other atoms of the model system are assigned to the MM part.



**Figure 4.** Molecular groups assigned to the quantum subsystem. The grey balls refer to the hydrogen link atoms (HL<sub>#</sub>) introduced to saturate the broken bonds C-O (HL<sub>1</sub>), C-C (HL<sub>2</sub>, HL<sub>5</sub>, HL<sub>6</sub>, HL<sub>4</sub>) and C-N (HL<sub>3</sub>) upon QM-MM partitioning. The inset in the right side shows the same subsystem in another perspective to emphasize that the oxygen molecule is located above the plane of the isoalloxazine ring of FMN. The distances from the oxygen atom of O<sub>2</sub> to the C4a and N5 atoms of FMN are given in Å

Finally, we used other approximations in the QM/MM scheme. On the base of the previous experience in modeling properties of flavin-containing photoactive proteins [20] we use here the density functional theory (DFT) approach with the PBE0 functional [2], the 6-31G\* basis set, and the D3 dispersion correction [12] to compute energies and energy gradients in QM upon optimization geometry parameters of the model system. As before, to model the system in the triplet electronic state, the unrestricted DFT approach was used. The AMBER99 force field parameters [28] are employed to estimate energy and energy gradients in MM.

As shown in the very recent study [11], the treatment of the QM-MM boundary as well as of the embedding scheme is still an open question despite years of successful application of QM/MM codes in modern software packages. It seems that the only way to report results of QM/MM application that can be reproduced in subsequent simulations is to provide as many technical details as possible. Therefore, we explicitly point out that the hydrogen link atom scheme is applied here to saturate the dangling chemical bonds. These atoms and their positions are shown using the dark grey color in Fig. 4. Distances between the hydrogen link atoms and the partner atoms assigned to the QM part (here, all of them are the carbon atoms) are the same as in the  $-\text{CH}_3$  groups.

The electrostatic embedding scheme is responsible for the interaction of the electron density in QM and the partial charges on MM atoms. To provide the necessary details, we note that NWChem performs a multi-region optimization procedure, i.e. at each optimization cycle, the QM region is optimized for M1 steps (10 in our calculations), whereas the MM region is optimized for M2 steps (300 in our case). If one applies the “density esplit” option in NWChem, this means that the electron density of the QM region is approximated with the electric charges obtained in the end of the QM part of the optimization cycle and used throughout the MM part. The “density static” option corresponds to the numerical integration to obtain the electron density of the QM region in the end of the QM part of each optimization cycle to use it during the MM part of that optimization cycle. Finally, we note that each link atom contacts the partner MM atom from the broken chemical bond. We employed in our calculations the NWChem option “mm charges exclude linkbond”, which means that all the MM point charges are taken into consideration except the very ones located on the covalent QM/MM boundary. These MM point charges are set to zero. As mentioned above, these computational details are necessary to make the results reproducible.

The inset in Fig. 4 shows that the located equilibrium geometry configuration corresponds to the location of the oxygen molecule near the isoalloxazine ring of FMN. The computed distance between the closest to the ring oxygen atom and the C4a atom of flavin is 3.15 Å, whereas the O-N5 distance is 3.80 Å. The corresponding distances in the crystal structure PDB ID 6SGG are 3.11 Å and 2.09 Å.

To sum up, we conclude that a variety of computational approaches applied in this work show that the complex of the flavin-dependent protein RutA is able to bind the oxygen molecule in protein cavities, thus by forming the active site suitable for oxidation reactions. One of the found oxygen-binding pockets is very close to the site identified by the O<sub>2</sub> pressurized crystallography experiments [18]. However, our calculations could not reproduce a rather short O(oxygen)-N5(FMN) distance 2.09 Å in the structure of RutA-O<sub>2</sub> complex reported by the results of crystallography experiments. The experimentally obtained value for the distance seems to be too short for a hydrogen bond, if N5 is protonated in the reduced form of FMN, and too long for a covalent bond, if one assumes a covalent binding of oxygen to the deprotonated nitrogen in FMN.

As location of a small molecule O<sub>2</sub> in protein cavities remains a challenging experimental task, we suggest that in order to obtain reliable results, experimental data should be combined with the results of appropriate computational approaches, including the results of computationally demanding quantum-based methods. This issue becomes even more critical, if the precise location of the oxygen molecule in the structure of the active site is considered as a crucial point defining possible reaction mechanisms of a flavin-containing monooxygenase [1, 17, 26].

## Conclusion

We demonstrate the application of high-performance computational methods to model the puzzling biochemical reactions of molecular oxygen with proteins, taking the interaction of oxygen with the flavin-containing monooxygenase RutA as a representative example. The presence of the oxygen molecule in model systems considerably complicates studies of structure and dynamics of large protein macromolecules, especially taking into account that localization of oxygen in protein sites is a nontrivial experimental task. Application of a wide spectrum of computational approaches including classical and QM/MM-based molecular dynamics as well as QM/MM optimization allowed us to characterize the RutA-O<sub>2</sub> complex and to dispute the recently reported crystallography structure of this complex. We show how the computational strategies can be optimized to apply several modern software packages in molecular modeling (NAMD, NWChem, TeraChem, ChemShell) using their options, which are often hidden for a user.

## Acknowledgements

We thank the Russian Science Foundation (project No. 22-13-00012) for the financial support of this work. The study was carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University [27].

*This paper is distributed under the terms of the Creative Commons Attribution-Non Commercial 3.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.*

## References

1. Adak, S., Begley, T.P.: RutA-Catalyzed Oxidative Cleavage of the Uracil Amide Involves Formation of a Flavin-N5-oxide. *Biochemistry* 56(29), 3708–3709 (2017). <https://doi.org/10.1021/acs.biochem.7b00493>
2. Adamo, C., Barone, V.: Toward reliable density functional methods without adjustable parameters: The PBE0 model. *The Journal of Chemical Physics* 110(13), 6158–6170 (1999). <https://doi.org/10.1063/1.478522>
3. Aleksandrov, A.: A molecular mechanics model for flavins. *Journal of Computational Chemistry* 40(32), 2834–2842 (2019). <https://doi.org/10.1002/jcc.26061>
4. Aprà, E., Bylaska, E.J., de Jong, W.A., *et al.*: NWChem: Past, present, and future. *The Journal of Chemical Physics* 152(18), 184102 (2020). <https://doi.org/10.1063/5.0004997>

5. Auhim, H.S., Grigorenko, B.L., Harris, T.K., *et al.*: Stalling chromophore synthesis of the fluorescent protein venus reveals the molecular basis of the final oxidation step. *Chem. Sci.* 12, 7735–7745 (2021). <https://doi.org/10.1039/D0SC06693A>
6. Balasubramani, S.G., Chen, G.P., Coriani, S., *et al.*: Turbomole: Modular program suite for ab initio quantum-chemical and condensed-matter simulations. *The Journal of Chemical Physics* 152(18), 184107 (2020). <https://doi.org/10.1063/5.0004635>
7. Berman, H.M., Westbrook, J., Feng, Z., *et al.*: The Protein Data Bank. *Nucleic Acids Research* 28(1), 235–242 (2000). <https://doi.org/10.1093/nar/28.1.235>
8. Best, R.B., Zhu, X., Shim, J., *et al.*: Optimization of the additive charmm all-atom protein force field targeting improved sampling of the backbone  $\phi$ ,  $\psi$  and side-chain  $\xi_1$  and  $\xi_2$  dihedral angles. *Journal of Chemical Theory and Computation* 8(9), 3257–3273 (2012). <https://doi.org/10.1021/ct300400x>
9. Chai, J.D., Head-Gordon, M.: Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.* 10, 6615–6620 (2008). <https://doi.org/10.1039/B810189B>
10. Colloc'h, N., Gabison, L., Monard, G., *et al.*: Oxygen Pressurized X-Ray Crystallography: Probing the Dioxygen Binding Site in Cofactorless Urate Oxidase and Implications for Its Catalytic Mechanism. *Biophysical Journal* 95(5), 2415–2422 (2008). <https://doi.org/10.1529/biophysj.107.122184>
11. Giudetti, G., Polyakov, I., Grigorenko, B.L., *et al.*: How Reproducible are QM/MM Simulations? Lessons from Computational Studies of the Covalent Inhibition of the SARS-CoV-2 Main Protease by Carmofur. *ChemRxiv* (2022). <https://doi.org/10.26434/chemrxiv-2022-4mmdc>
12. Grimme, S., Antony, J., Ehrlich, S., Krieg, H.: A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* 132(15), 154104 (2010). <https://doi.org/10.1063/1.3382344>
13. Humphrey, W., Dalke, A., Schulten, K.: VMD: Visual molecular dynamics. *Journal of Molecular Graphics* 14(1), 33–38 (1996). [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5)
14. Khrenova, M.G., Polyakov, I.V., Nemukhin, A.V.: Molecular dynamics of enzyme-substrate complexes in guanosine-binding proteins. *Khimicheskaya Fizika* 41(6), 66–72 (2022). <https://doi.org/10.31857/S0207401X22060061>
15. Kstner, J., Carr, J.M., Keal, T.W., *et al.*: DL-FIND: An Open-Source Geometry Optimizer for Atomistic Simulations. *The Journal of Physical Chemistry A* 113(43), 11856–11865 (2009). <https://doi.org/10.1021/jp9028968>
16. Loh, K.D., Gyaneshwar, P., Papadimitriou, E.M., *et al.*: A previously undescribed pathway for pyrimidine catabolism. *Proceedings of the National Academy of Sciences* 103(13), 5114–5119 (2006). <https://doi.org/10.1073/pnas.0600521103>
17. Massey, V.: Activation of molecular oxygen by flavins and flavoproteins. *Journal of Biological Chemistry* 269(36), 22459–22462 (1994). [https://doi.org/10.1016/S0021-9258\(17\)31664-2](https://doi.org/10.1016/S0021-9258(17)31664-2)

18. Matthews, A., Saleem-Batcha, R., Sanders, J.N., *et al.*: Aminoperoxide adducts expand the catalytic repertoire of flavin monooxygenases. *Nature Chemical Biology* 16(5), 556–563 (2020). <https://doi.org/10.1038/s41589-020-0476-2>
19. Melo, M.C.R., Bernardi, R.C., Rudack, T., *et al.*: NAMD goes quantum: an integrative suite for hybrid simulations. *Nature Methods* 15(5), 351–354 (2018). <https://doi.org/10.1038/nmeth.4638>
20. Nemukhin, A.V., Grigorenko, B.L., Khrenova, M.G., Krylov, A.I.: Computational challenges in modeling of representative bioimaging proteins: GFP-like proteins, flavoproteins, and phytochromes. *The Journal of Physical Chemistry B* 123(29), 6133–6149 (2019). <https://doi.org/10.1021/acs.jpcc.9b00591>
21. Nemukhin, A.V., Grigorenko, B.L., Polyakov, I.V., Lushchekina, S.V.: Computational modeling of the SARS-CoV-2 main protease inhibition by the covalent binding of prospective drug molecules. *Supercomputing Frontiers and Innovations* 7(3), 25–32 (2020). <https://doi.org/10.14529/jsfi200303>
22. Phillips, J.C., Hardy, D.J., Maia, J.D.C., *et al.*: Scalable molecular dynamics on CPU and GPU architectures with NAMD. *The Journal of Chemical Physics* 153(4), 044130 (2020). <https://doi.org/10.1063/5.0014475>
23. Seritan, S., Bannwarth, C., Fales, B.S., *et al.*: Terachem: A graphical processing unit-accelerated electronic structure package for large-scale ab initio molecular dynamics. *WIREs Computational Molecular Science* 11(2), e1494 (2021). <https://doi.org/10.1002/wcms.1494>
24. Shabalin, I.G., Porebski, P.J., Minor, W.: Refining the macromolecular model – achieving the best agreement with the data from X-ray diffraction experiment. *Crystallography Reviews* 24(4), 236–262 (2018). <https://doi.org/10.1080/0889311X.2018.1521805>
25. Sherwood, P., de Vries, A.H., Guest, M.F., *et al.*: QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis. *Journal of Molecular Structure: THEOCHEM* 632(1), 1–28 (2003). [https://doi.org/10.1016/S0166-1280\(03\)00285-9](https://doi.org/10.1016/S0166-1280(03)00285-9)
26. Teufel, R.: Flavin-catalyzed redox tailoring reactions in natural product biosynthesis. *Archives of Biochemistry and Biophysics* 632, 20–27 (2017). <https://doi.org/10.1016/j.abb.2017.06.008>
27. Voevodin, V.V., Antonov, A.S., Nikitenko, D.A., *et al.*: Supercomputer Lomonosov-2: Large scale, deep monitoring and fine analytics for the user community. *Supercomputing Frontiers and Innovations* 6(2), 4–11 (2019). <https://doi.org/10.14529/jsfi190201>
28. Wang, J., Cieplak, P., Kollman, P.A.: How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry* 21(12), 1049–1074 (2000). [https://doi.org/10.1002/1096-987X\(200009\)21:12<1049::AID-JCC3>3.0.CO;2-F](https://doi.org/10.1002/1096-987X(200009)21:12<1049::AID-JCC3>3.0.CO;2-F)
29. Wang, S., Hou, K., Heinz, H.: Accurate and compatible force fields for molecular oxygen, nitrogen, and hydrogen to simulate gases, electrolytes, and heterogeneous interfaces. *Journal of Chemical Theory and Computation* 17(8), 5198–5213 (2021). <https://doi.org/10.1021/acs.jctc.0c01132>

# Analysis of Ion Atmosphere Around Nucleosomes Using Supercomputer MD Simulations

Nikita A. Kosarim<sup>1</sup>, Grigoriy A. Armeev<sup>1</sup> , Mikhail P. Kirpichnikov<sup>1,2</sup>, Alexey K. Shaytan<sup>1,3</sup> 

© The Authors 2022. This paper is published with open access at SuperFri.org

The nucleosome is the basic unit of eukaryotic DNA compaction. It consists of about 147 base pairs wrapped around an octamer of histone proteins. Nucleosomal dynamics provides the availability of packaged DNA for various factors that carry out the vital processes associated with chromatin. It is not completely known how the structure and dynamics of the nucleosome depends on the ionic environment. The current researches do not give an unambiguous answer and often contradict each other. In this paper, we demonstrate supercomputer molecular dynamics simulations of nucleosome models surrounded by monovalent sodium and potassium cations. Analyzing the trajectories, we have shown the details of the distribution of sodium and potassium ions around the linker DNA, nucleosomal DNA at the sites of nucleosomal opening, and histone residues involved in the process of nucleosomal breathing. We have demonstrated the mobility of DNA linkers and the process of nucleosomal unwrapping in various ionic environments, and also assessed the probable mechanisms of the dependence of nucleosome unwrapping on the type of ions in the system. Our study is intended to emphasize the importance of understanding the role of the ionic environment in the functioning of chromatin.

*Keywords: molecular modeling, molecular dynamics simulations, nucleosomes, protein-DNA interactions, monovalent cations, sodium, potassium.*

## Introduction

The length of DNA in a human cell is about 2 meters, which is 200,000 times the diameter of the nucleus [37]. To fit DNA in the nucleus, it must be packed into a compact structure – chromatin, the elementary unit of which is the nucleosome. The nucleosome is formed by a histone octamer (two H2A-H2B dimers in complex with a tetramer H3-H4/H3-H4), which is wrapped by DNA about 147 base pairs long in a 1.67-turn of a left-handed superhelix (Fig. 1). This structure is also called nucleosome core particle (NCP) [26, 38]. Neighboring nucleosomes are connected by the so-called linker DNA with a length of 10 to 90 bp depending on the type of organism, cell type, stage of the cell cycle, and location in the cell nucleus [2, 21].

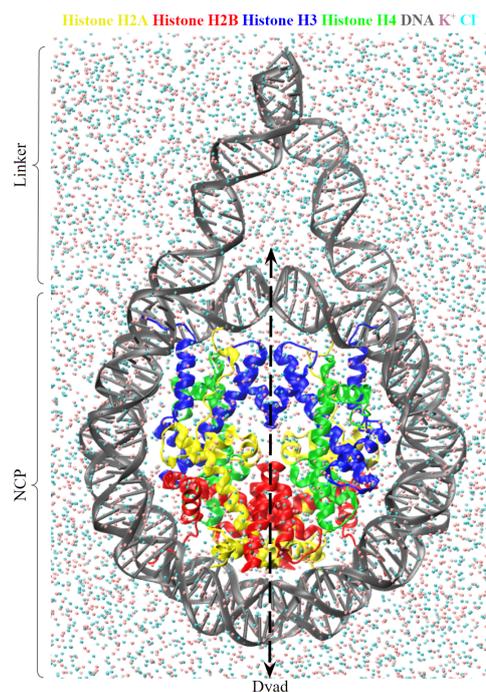
The nucleosome is dynamic [4]. Its dynamic plays a key role in the vital processes that take place in the nucleus. An important part of nucleosomal dynamics is nucleosome breathing, which is the unwrapping of up to twenty base pairs from the NCP. The process of nucleosomal breathing is driven by dynamic interactions between neighboring DNA strands, as well as interactions between DNA and histones. The main meaning of nucleosomal breathing is to regulate the availability of DNA for factors that carry out chromatin-associated processes.

The nucleosome is a charged structure [17], and most of the contacts between DNA and histones are electrostatic. This causes the dependence of nucleosome structure and dynamics on the ionic environment. In living systems, the main monovalent ions are Na<sup>+</sup> and K<sup>+</sup>. While sodium prevails in the extracellular environment, potassium dominates in the intracellular environment, being an essential element of the DNA environment. One of the hypotheses for such distribution of ions in eukaryotic cells, suggests a more efficient implementation of chromatin-associated

<sup>1</sup>Department of Biology, Lomonosov Moscow State University, Moscow, Russia

<sup>2</sup>Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia

<sup>3</sup>Bioinformatics Lab, Faculty of Computer Science, HSE University, Moscow, Russia



**Figure 1.** Nucleosome model in simulated system

processes in the potassium environment compared to sodium [21]. Nevertheless, due to the similarity of  $\text{Na}^+$  and  $\text{K}^+$  ions, most of the experimental work related to the study of nucleosomes is carried out in media containing sodium ions. Some works demonstrated the identity of experimental results obtained in potassium and sodium solutions [7, 22–24]. However, today, more and more researchers are declaring significant differences between these ionic environments. The key difference between sodium and potassium is the diameter of the ions: sodium has a smaller diameter than potassium. At the same time, the diameter of the hydration shell of sodium is larger than that of potassium. This difference can affect the features of ions interactions with the nucleosome and their distribution around it. As a result, structure and dynamics of the nucleosome, which determine the efficiency of various processes associated with chromatin, can be affected.

For example, the question of the comparative affinity of sodium and potassium ions for DNA has not yet been unambiguously resolved. Some computational work suggests a higher affinity for sodium [40]. In turn, the results of experimental work are contradictory. Earlier studies suggest a greater affinity for potassium [47]. More recent studies, including works made with the ion counting method, indicate that the affinity of sodium and potassium ions for DNA does not differ or differs slightly [6, 23, 25].

Molecular dynamics and experiments show that sodium ions pack DNA [47] and polynucleosomal chains [2] more effectively than potassium ions. Thus, sodium ions can suppress the natural dynamics of chromatin. However, at the level of mononucleosomes the opposite effect is observed: the nucleosome acquires a more compact conformation in the potassium environment [28].

The type of ions in the medium can also influence the contacts between DNA and histones. Binding of DNA and histones is primarily due to electrostatic interactions between the negatively charged phosphate groups of DNA and the positively charged side chains of arginine and lysine amino acids through ion pairs [27]. It was shown that in the presence of mono- and

divalent cations, these interactions weaken, and their geometry changes due to the polarization of the phosphate groups charges towards ions [13]. Thus, the main role of cations in interactions between DNA and histones is the regulation of these interactions, in particular, the modulation of the dynamics of nucleosomal DNA opening and closing by weakening intermolecular contacts.

One of the powerful tools for studying the structure and dynamics of nucleosomes is the molecular dynamics (MD) method, which makes it possible to interpret and supplement the results of experimental studies. MD simulations have been applied to study the structure and dynamics of nucleosomal DNA [39], nucleosomal unfolding [3, 12, 18, 44], functions and dynamics of histone tails [9, 11, 15, 32, 36, 41], the role of histone post-translational modifications [14, 35], DNA-DNA and DNA-protein interactions [46], counterion distribution around nucleosomes [30], role of DNA sequence in nucleosome dynamics [42], etc. However, a detailed analysis of the ionic environment of the nucleosome and the dependence of the nucleosomal structure and dynamics on it has not been previously carried out. Due to the complexity and computation time, models often require various approximations such as removing parts of the system (e.g. histone tails), using implicit solvent models, or representing the system as a coarse-grained model to achieve significant simulation times. At the moment, the duration of the longest trajectory of an all-atom nucleosome model with intact histone tails, obtained using supercomputing, is about 15  $\mu$ s [5]. However, many important dynamic processes of the nucleosome occur in the range from microseconds to milliseconds [16], which requires an increase in speed and efficiency of computational methods.

In this work, motivated by the ambiguity of the available evidence about the effect of the monovalent cations type on the nucleosome structure and dynamics, we aimed to use all-atom supercomputer molecular dynamics simulations to model nucleosome dynamics in different ionic environments. Simulating the time trajectory of such a large all-atom model as the nucleosome with ions requires considerable computing power. Thus, the most efficient way is to carry out MD calculations on supercomputers, which make it possible to achieve significant computational speeds and trajectory lengths. Using the Lomonosov-2 supercomputer, we obtained eight trajectories for nucleosome models (including 20 bp linker DNA segment) in solution with  $\text{Na}^+$  and  $\text{K}^+$  ionic concentrations of 150 mM and 3 M. In our work, we analyzed the details of the distribution of sodium and potassium ions around DNA of the nucleosome, and assessed the possible effect of ions on the process of nucleosomal breathing.

## 1. Methods

### 1.1. Simulation Preparation

The work was carried out using the molecular dynamics simulations method. The systems were based on the initial X-ray structure 1KX5 from the PDB database [10]. The 1KX5 structure was modified by adding 20 bp DNA-linkers with identical sequence and truncating histone tails. Sites of truncation were chosen in order to remove histone tails' flexible parts and avoid removing parts near the globular core that make stable contacts with DNA and histones. The final modified structure was placed in rhombododecahedral cell with periodic boundary conditions and minimal distance from nucleosome to cell wall 2 or 3 nm depending on the system. Then water molecules (TIP3P model) and  $\text{K}^+$  or  $\text{Na}^+$  and  $\text{Cl}^-$  ions were added to reach the desired concentration for the whole cell volume (150 mM or 3 M) (Fig. 1). In order to avoid possible fraying of base pairs at the ends of DNA, an additional harmonic potential (with a force constant of 1000 kJ

$\text{mol}^{-1} \text{ nm}^{-2}$ ) was applied to the distance between glycosidic nitrogen atoms of the terminal base pairs. Simulations were performed with the GROMACS 2020.2 software package [1]. An AMBER ff14SB force field [29] with parmbsc1 DNA parameter correction [20] and CUFIX ion parameter correction [45] was used. The prepared systems were energy minimized and step-by-step equilibrated with a gradual relaxation of harmonic positional restraints imposed on all atoms except hydrogens. Minimization was carried out using the steepest descent gradient method with positional restraints of  $500 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ . Then the five stages of equilibration were performed: 1. 100 ps with positional restraints of  $500 \text{ kJ mol}^{-1} \text{ nm}^{-2}$  with a step of 0.5 fs; 2. 200 ps with positional restraints of  $50 \text{ kJ mol}^{-1} \text{ nm}^{-2}$  with a step of 2 fs (then the step was kept); 3. 200 ps with positional restraints of  $5 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ ; 4. 200 ps with positional restraints of  $0.5 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ ; 5. 200 ps free simulation. The systems were modeled in an NPT ensemble with a temperature of 300 K using the velocity rescale scheme [8] and with a 1 bar pressure coupling using Parrinello-Rahman barostat [34]. The molecular dynamics simulation step was 2 fs, and frames in the molecular dynamics trajectory simulation were saved every 1 ns. Detailed simulated systems description and production run protocol files may be found at [https://github.com/intbio/Kosarim\\_et\\_al\\_2022](https://github.com/intbio/Kosarim_et_al_2022).

Simulations were performed on the Lomonosov-2 supercomputer [43] using 1 computing node, having 14 CPU cores and one NVidia Tesla K40 GPU. Average speed of simulation was 6 ns per day. For calculation of non-bonded interactions GPU was used.

## 1.2. Trajectory Analysis

The resulting trajectories were analyzed using Gromacs [1], MDAnalysis [31] and VMD [19] programs, as well as algorithms developed by our scientific group using Python 3.

The unwrapping of nucleosomal DNA during nucleosome breathing was measured as the length of a DNA fragment in which all base pair centers were more than 7 angstroms further from the centers of any base pair in the initial structure.

Radial distribution functions (RDF) of the of sodium and potassium ions were plotted around two regions of the nucleosome DNA, namely, around any atom of DNA linkers (74-92 bp, “linker”) and nucleosomal DNA in the region of the exit from the nucleosome (54-73 bp, “core”). The functions were normalized in such a way as to converge to 1 at large values of the distance.

Survival probability (SP) reflects the probability of finding ions within a given radius from DNA within a certain time. In our case, we considered the probability of detecting sodium and potassium ions at concentrations of 150 mM and 3 M within a radius of 5 angstroms from any DNA atom around DNA linkers (74-92 bp, “linker”) and nucleosomal DNA (54-73 bp, “core”).

For a comparative analysis of the spatial distribution of sodium and potassium ions around DNA, the occupancy of ions near the nucleosomal DNA in the region of the DNA exit from the nucleosome (base pair number 54-73) was plotted. The occupancy values are obtained as follows: the space is divided by a cubic lattice in a certain cell size (in this case, 1 angstrom); during the analysis the lattice nodes are assigned the values 0 or 1 depending on whether one or more of the considered atoms intersect with them; then the results are averaged along all the trajectory. For this analysis, we selected pairs of systems that demonstrate the greatest opening of the DNA in nucleosome (K150/Na150 and K3000a/Na3000a).

## 2. Results and Discussion

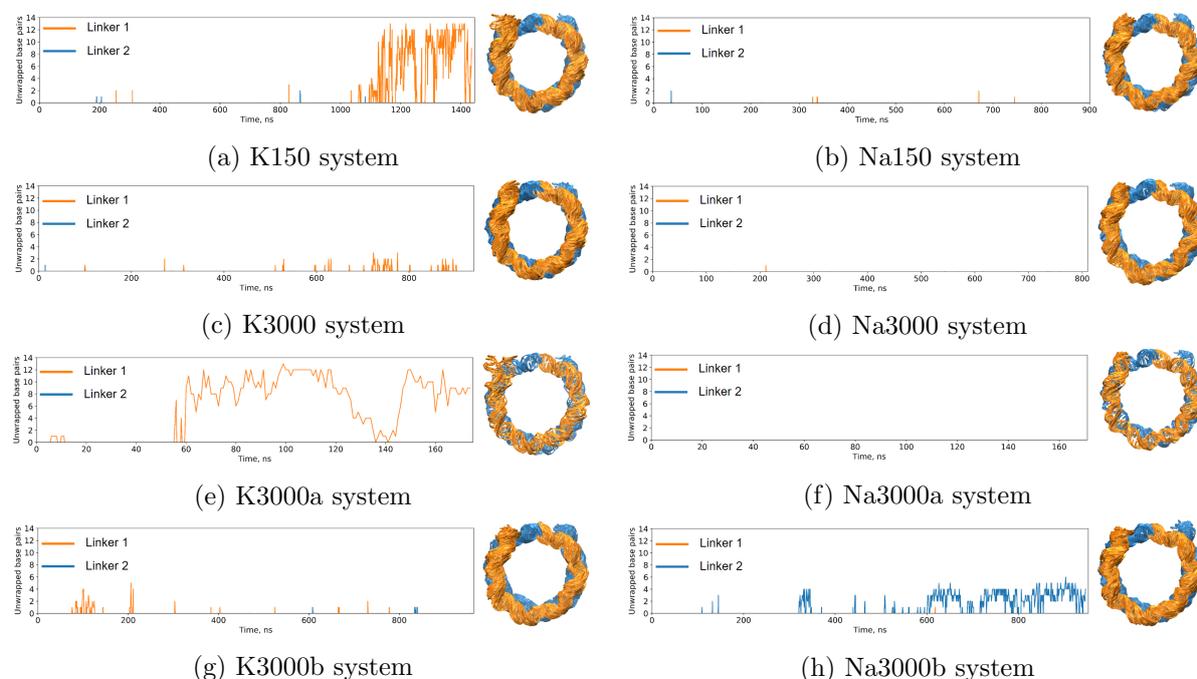
The study of the influence of the monovalent cations type on the process of nucleosome breathing was carried out on eight systems, mainly differing in the type of cations, their concentration and simulation time (Tab. 1).

**Table 1.** Simulated systems

System	K150	K3000	K3000a	K3000b	Na150	Na3000	Na3000a	Na3000b
Atom number	319435	302890	397389	302890	323520	302890	397389	302890
Water molecules	99490	89184	119461	89184	99490	89184	119461	89184
Cation type	K <sup>+</sup>	K <sup>+</sup>	K <sup>+</sup>	K <sup>+</sup>	Na <sup>+</sup>	Na <sup>+</sup>	Na <sup>+</sup>	Na <sup>+</sup>
Ions number, cation/Cl	577/271	5723/5417	7557/7251	5723/5417	577/271	5723/5417	7557/7251	5723/5417
Cation concentration, mM	150	3000	3000	3000	150	3000	3000	3000
Distance to the box, nm	2	2	3	2	2	2	3	2
Time, $\mu$ s	1.37	0.92	0.94	0.16	0.88	0.80	0.93	0.16

### 2.1. DNA Unwrapping

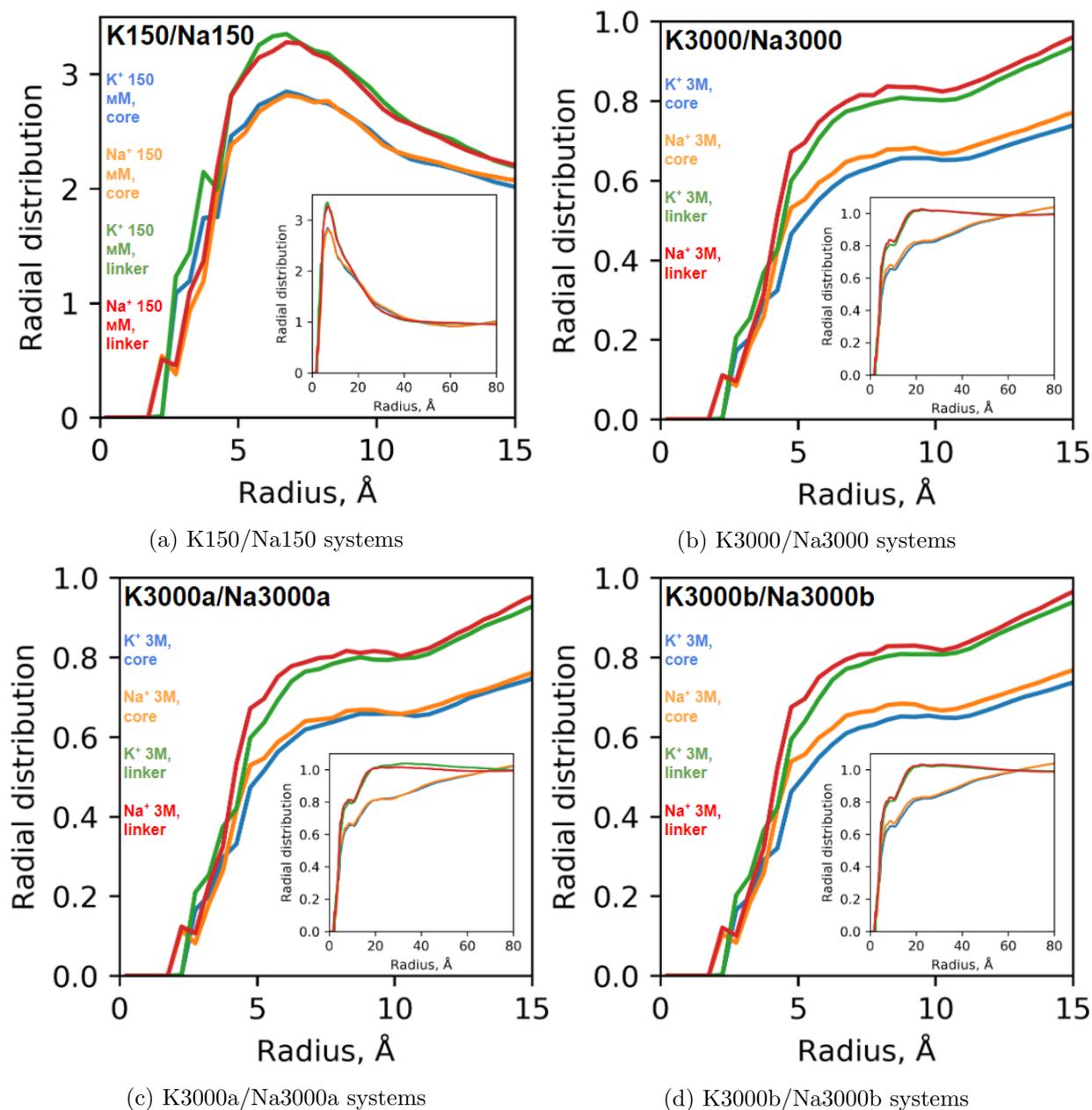
DNA unwrapping was observed in trajectories at timescales of greater than 1 microsecond for ion concentrations of 150 mM and at timescales of hundreds of nanoseconds for 3M concentrations (Fig. 2). The obtained results clearly demonstrate the asymmetric nature of nucleosome breathing in our simulations, that was experimentally shown by Ngo and colleagues [33]. Despite the fact that the nucleosome unwrapping in the obtained graphs is observed mainly in the potassium environment, this, unfortunately, cannot serve as unambiguous evidence of the influence of the ions type due to the probabilistic nature of nucleosome breathing and the short duration of simulations.



**Figure 2.** Profiles of the extent of DNA unwrapping during MD simulations and corresponding overlays of MD snapshots spaced 50 ns apart

## 2.2. Radial Distribution of Ions

According to the radial distribution function (RDF) plots (Fig. 3), several important observations can be made: firstly, the plots of the ionic distribution around DNA linkers are higher than around nucleosomal DNA, which indicates a greater probability of potassium and sodium ions binding to the linker DNA than to the nucleosomal DNA; secondly, the plots for sodium and potassium ions do not differ significantly, which indicates a similar probability of detecting potassium and sodium ions near both linker and nucleosomal DNA.



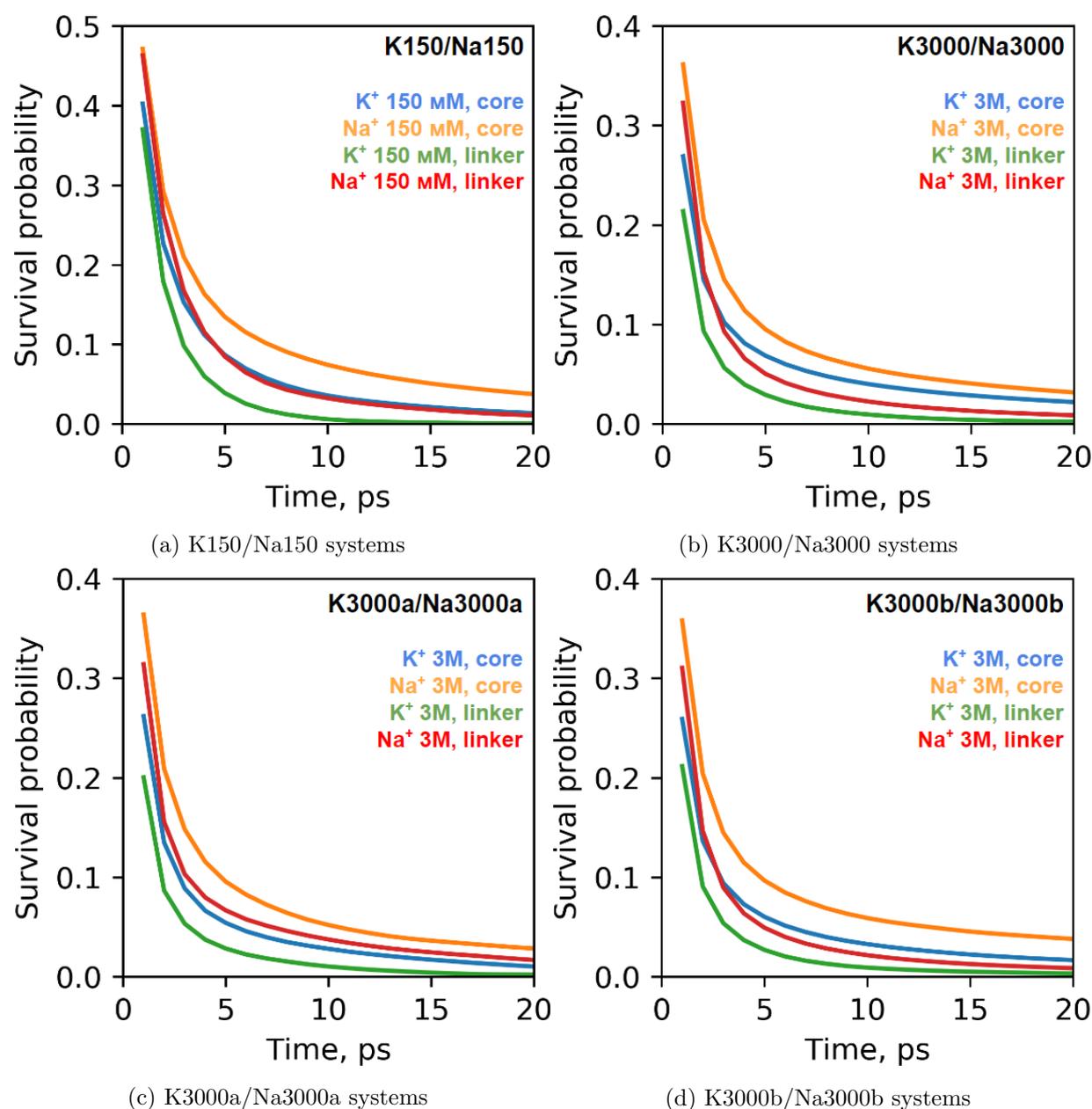
**Figure 3.** Radial distribution functions of  $\text{Na}^+$  and  $\text{K}^+$  ions around linker (“linker”) and nucleosomal (“core”) DNA

The first observation can be explained due to the fact that some of the ion-binding sites on nucleosomal DNA can be occupied by the amino acid residues of the adjacent histones or by

adjacent DNA gyres. The second observation may be a consequence of the fact that both types of ions have similar interaction patterns with the DNA.

### 2.3. Survival Probability of Ions

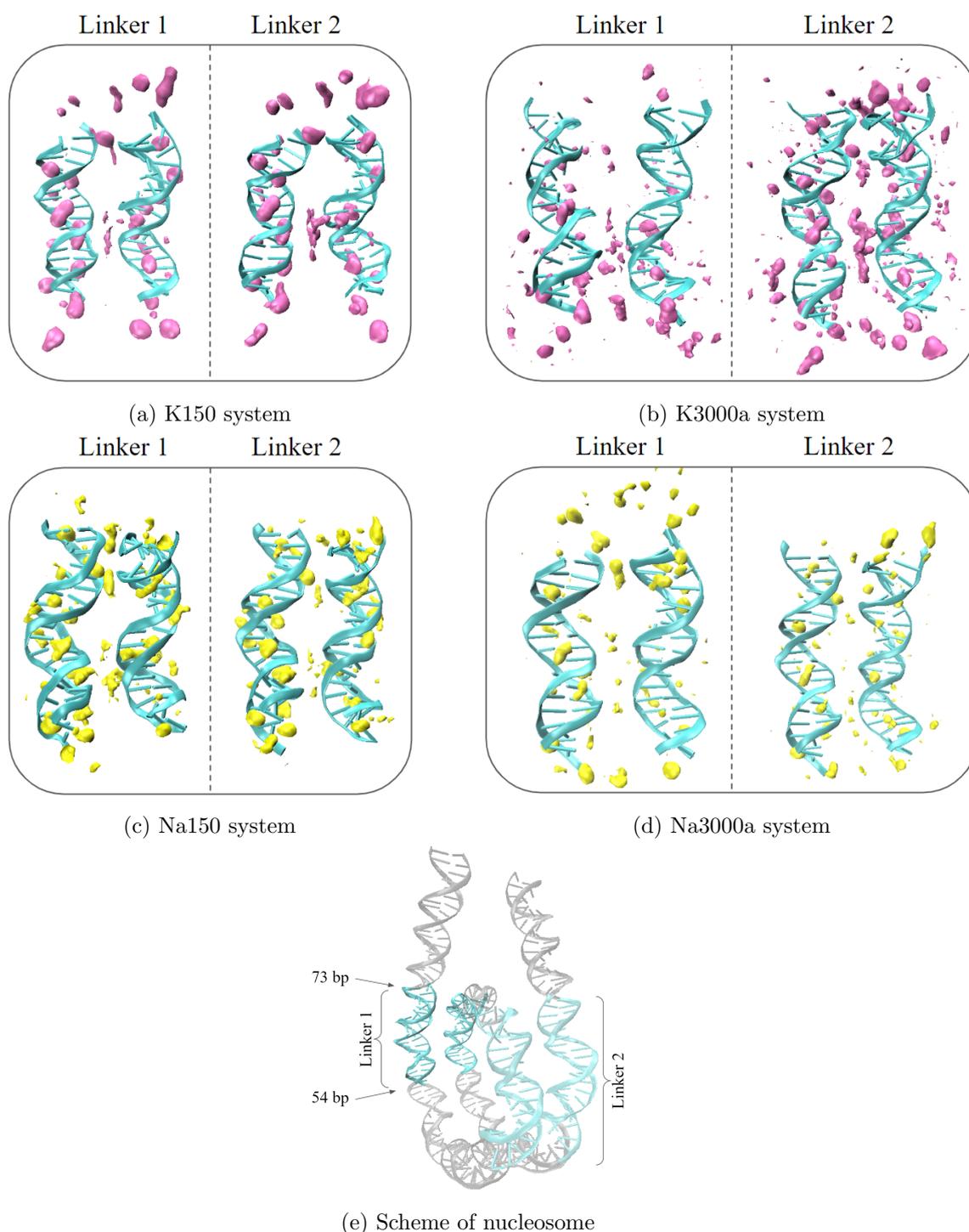
The calculated survival probability (SP) plots demonstrate higher values of the SP for sodium ions compared to potassium ions near the linker and nucleosomal DNA, both at a concentration of 150 mM and at 3 M (Fig. 4). This indicates a longer interaction of sodium ions with DNA. We would also like to note the increased SP of both types of ions near nucleosomal DNA compared to the SP near linker DNA. This observation indicates a relatively longer residence time of both types of ions near nucleosomal DNA. This can be explained by the interactions of ions with histones or the neighboring gyres of the nucleosomal DNA, preventing their movement.



**Figure 4.** Survival probability of Na<sup>+</sup> and K<sup>+</sup> ions in 5 Å around linker ("linker") and nucleosomal ("core") DNA in simulation Na pairs

## 2.4. Ionic Occupancy

In systems with a sodium environment, where the nucleosomal DNA is predominantly wrapped on both sides during the entire simulation, the occupancy of sodium ions does not significantly differ (Fig. 5). At the same time, in systems with potassium ions, and especially in the K3000a system, there is a clear difference in the ion occupancy in the space between adjacent DNA gyres.



**Figure 5.** Occupancy of  $\text{Na}^+$  and  $\text{K}^+$  ions around nucleosomal DNA at entry/exit regions and corresponding scheme of nucleosome

The occupancy is increased on the non-unwrapping side of the nucleosome (“Linker 2”), where the DNA segment of the upper and lower DNA gyre reside closer together throughout the entire trajectory. At the same time, occupancy on the unwrapping side (“Linker 1”) is lower. This may indicate both that the approach of DNA segments leads to the accumulation of ions in the intermolecular space, and the opposite – ions, accumulating between and around the adjacent DNA segments, prevent their repulsion and, in this case, the nucleosome breathing.

## Conclusion

As a result of this work, we have made important observations concerning the distribution of monovalent sodium and potassium cations around the nucleosomal DNA, as well as their influence on the structure and dynamics of the nucleosome linker region. In particular, plots of nucleosome unwrapping, radial and temporal distribution functions, and populations of both types of ions near linker and nucleosomal DNA were obtained. We showed that the concentration and mobility of both types of ions near the nucleosomal DNA is lower than around the linker DNA, which can be explained by the interactions of DNA with histones and interactions between neighboring DNA gyres. The noted longer residence time of sodium ions near the DNA presumably indicates their greater affinity for DNA compared to potassium ions at least within the force field model used in our simulations. Among other things, it has been demonstrated that ions tend to accumulate in the space between the nucleosomal DNA gyres, probably contributing to their attraction and thus preventing nucleosome breathing. The problem of monovalent cations type effect on nucleosome requires further examination due to its ambiguity. Nevertheless, the use of potassium, which is predominant *in vivo*, may be preferred in computer simulations.

## Acknowledgements

This research was funded by the Ministry of Science and Higher Education of the Russian Federation, grant No. 075-15- 2021-1354. This research has been non-profit supported by the Interdisciplinary Scientific and Educational Schools of Moscow University “Molecular Technologies of the Living Systems and Synthetic Biology”.

*This paper is distributed under the terms of the Creative Commons Attribution-Non Commercial 3.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.*

## References

1. Abraham, M.J., Murtola, T., Schulz, R., *et al.*: Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1-2, 19–25 (2015). <https://doi.org/10.1016/j.softx.2015.06.001>
2. Allahverdi, A., Chen, Q., Korolev, N., Nordenskiöld, L.: Chromatin compaction under mixed salt conditions: Opposite effects of sodium and potassium ions on nucleosome array folding. *Scientific Reports* 5 (2015). <https://doi.org/10.1038/srep08512>
3. Armeev, G.A., Shaitan, K.V., Shaytan, A.K.: Nucleosome structure relaxation during dna unwrapping: Molecular dynamics simulation study. *Moscow University Biological Sciences Bulletin* 71, 141–144 (2016). <https://doi.org/10.3103/S0096392516030020>

4. Armeev, G.A., Gribkova, A.K., Pospelova, I., *et al.*: Linking chromatin composition and structural dynamics at the nucleosome level. *Current Opinion in Structural Biology* 56, 46–55 (2019). <https://doi.org/10.1016/j.sbi.2018.11.006>
5. Armeev, G.A., Kniazeva, A.S., Komarova, G.A., *et al.*: Histone dynamics mediate DNA unwrapping and sliding in nucleosomes. *Nature Communications* 12 (2021). <https://doi.org/10.1038/s41467-021-22636-9>
6. Bai, Y., Greenfeld, M., Travers, K.J., *et al.*: Quantitative and comprehensive decomposition of the ion atmosphere around nucleic acids. *Journal of the American Chemical Society* 129, 14981–14988 (2007). <https://doi.org/10.1021/ja075020g>
7. Braunlin, W.H., Nordenskiöld, L.: A potassium-39 NMR study of potassium binding to double-helical DNA (1984)
8. Bussi, G., Donadio, D., Parrinello, M.: Canonical sampling through velocity rescaling. *Journal of Chemical Physics* 126 (2007). <https://doi.org/10.1063/1.2408420>
9. Chakraborty, K., Kang, M., Loverde, S.M.: Molecular mechanism for the role of the H2A and H2B histone tails in nucleosome repositioning. *Journal of Physical Chemistry B* 122, 11827–11840 (2018). <https://doi.org/10.1021/acs.jpcc.8b07881>
10. Davey, C.A., Sargent, D.F., Luger, K., *et al.*: Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9 Å Resolution. *Journal of Molecular Biology* 319(5), 1097–1113 (2002). [https://doi.org/https://doi.org/10.1016/S0022-2836\(02\)00386-8](https://doi.org/https://doi.org/10.1016/S0022-2836(02)00386-8)
11. Erler, J., Zhang, R., Petridis, L., *et al.*: The role of histone tails in the nucleosome: A computational study. *Biophysical Journal* 107, 2911–2922 (2014). <https://doi.org/10.1016/j.bpj.2014.10.065>
12. Ettig, R., Kepper, N., Stehr, R., *et al.*: Dissecting DNA-histone interactions in the nucleosome by molecular dynamics simulations of DNA unwrapping. *Biophysical Journal* 101, 1999–2008 (2011). <https://doi.org/10.1016/j.bpj.2011.07.057>
13. Farca, A.A., Bende, A.: The influence of monovalent and divalent metal cations on the stability of the DNA-protein interaction in the nucleosome core particle. *Advances in Quantum Chemistry* 81, 269–290 (2020). <https://doi.org/10.1016/bs.aiq.2020.05.001>
14. Fu, I., Geacintov, N.E., Broyde, S.: Molecular dynamics simulations reveal how H3K56 acetylation impacts nucleosome structure to promote DNA exposure for lesion sensing. *DNA Repair* 107, 103201 (2021). <https://doi.org/10.1016/j.dnarep.2021.103201>
15. Furukawa, A., Wakamori, M., Arimura, Y., *et al.*: Characteristic H3 N-tail dynamics in the nucleosome core particle, nucleosome, and chromatosome. *iScience* 25(3) (2022). <https://doi.org/10.1016/j.isci.2022.103937>
16. Gansen, A., Felekyan, S., Khnemuth, R., *et al.*: High precision FRET studies reveal reversible transitions in nucleosomes between microseconds and minutes. *Nature Communications* 9 (2018). <https://doi.org/10.1038/s41467-018-06758-1>
17. Gebala, M., Johnson, S.L., Narlikar, G.J., Herschlag, D.: Ion counting demonstrates a high electrostatic field generated by the nucleosome. *eLife* 8, e44993 (2019). <https://doi.org/10.7554/eLife.44993>

18. Huertas, J., Cojocaru, V.: Breaths, twists, and turns of atomistic nucleosomes. *Journal of Molecular Biology* 433 (2021). <https://doi.org/10.1016/j.jmb.2020.166744>
19. Humphrey, W., Dalke, A., Schulten, K.: VMD: Visual Molecular Dynamics (1996)
20. Ivani, I., Dans, P.D., Noy, A., *et al.*: Parmbsc1: A refined force field for DNA simulations. *Nature Methods* 13, 55–58 (2015). <https://doi.org/10.1038/nmeth.3658>
21. Korolev, N.: How potassium came to be the dominant biological cation: of metabolism, chemiosmosis, and cation selectivity since the beginnings of life. *BioEssays* 43 (2021). <https://doi.org/10.1002/bies.202000108>
22. Korolev, N., Lyubartsev, A.P., Rupprecht, A., Nordenskiöld, L.: Competitive binding of  $Mg^{2+}$ ,  $Ca^{2+}$ ,  $Na^+$ , and  $K^+$  ions to DNA in oriented DNA fibers: Experimental and Monte Carlo simulation results. *Biophysical Journal* 77, 2736–2749 (1999). [https://doi.org/10.1016/S0006-3495\(99\)77107-9](https://doi.org/10.1016/S0006-3495(99)77107-9)
23. Korolev, N., Lyubartsev, A.P., Rupprecht, A., Nordenskiöld, L.: Experimental and Monte Carlo simulation studies on the competitive binding of  $Li^+$ ,  $Na^+$ , and  $K^+$  ions to DNA in oriented DNA fibers. *Journal of Physical Chemistry B* 103, 9008–9019 (1999). <https://doi.org/10.1021/jp9913517>
24. Korolev, N., Nordenskiöld, L.: Influence of alkali cation nature on structural transitions and reactions of biopolyelectrolytes. *Biomacromolecules* 1, 648–655 (2000). <https://doi.org/10.1021/bm000042f>
25. Kuznetsov, I., Gorshkov, V., Ivanov, V., *et al.*: Ion-exchange properties of immobilized DNA (1984)
26. Luger, K., Der, A.W.M., Richmond, R.K., *et al.*: Crystal structure of the nucleosome core particle at 2.8 Å resolution (1997)
27. Luscombe, N.M., Laskowski, R.A., Thornton, J.M.: Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level (2001)
28. Lyubitelev, A.V., Studitsky, V.M., Feofanov, A.V., Kirpichnikov, M.P.: Effect of sodium and potassium ions on conformation of linker parts of nucleosomes. *Moscow University Biological Sciences Bulletin* 72, 146–150 (2017). <https://doi.org/10.3103/S0096392517030075>
29. Maier, J.A., Martinez, C., Kasavajhala, K., *et al.*: ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB (2015)
30. Materese, C.K., Savelyev, A., Papoian, G.A.: Counterion atmosphere and hydration patterns near a nucleosome core particle. *Journal of the American Chemical Society* 131, 15005–15013 (2009). <https://doi.org/10.1021/ja905376q>
31. Michaud-Agrawal, N., Denning, E.J., Woolf, T.B., Beckstein, O.: MDAAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* 32, 2319–2327 (2011). <https://doi.org/10.1002/jcc.21787>
32. Morrison, E.A., Bowerman, S., Sylvers, K.L., *et al.*: The conformation of the histone H3 tail inhibits association of the BPTF PHD finger with the nucleosome . <https://doi.org/10.7554/eLife.31481.001>

33. Ngo, T.T., Zhang, Q., Zhou, R., *et al.*: Asymmetric unwrapping of nucleosomes under tension directed by DNA local flexibility. *Cell* 160, 1135–1144 (2015). <https://doi.org/10.1016/j.cell.2015.02.001>
34. Parrinello, M., Rahman, A.: Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* 52, 7182–7190 (1981). <https://doi.org/10.1063/1.328693>
35. Peng, Y., Espiritu, D., Li, S., *et al.*: Deciphering the effects of post-translational modifications and mutations on histone tail dynamics and interactions. *Biophysical Journal* 121, 361a (2022). <https://doi.org/10.1016/j.bpj.2021.11.931>
36. Peng, Y., Li, S., Landsman, D., Panchenko, A.R.: Histone tails as signaling antennas of chromatin. *Current Opinion in Structural Biology* 67, 153–160 (2021). <https://doi.org/10.1016/j.sbi.2020.10.018>
37. Razin, S.V., Ulianov, S.V.: Gene functioning and storage within a folded genome. *Cellular & Molecular Biology Letters* 22, 18 (2017). <https://doi.org/10.1186/s11658-017-0050-4>
38. Richmond, T.J., Davey, C.A.: The structure of DNA in the nucleosome core (2003)
39. Roccatano, D., Barthel, A., Zacharias, M.: Structural flexibility of the nucleosome core particle at atomic resolution studied by molecular dynamics simulation. *Biopolymers* 85, 407–421 (2007). <https://doi.org/10.1002/bip.20690>
40. Savelyev, A., Papoian, G.A.: Electrostatic, steric, and hydration interactions favor Na<sup>+</sup> condensation around DNA compared with K<sup>+</sup>. *Journal of the American Chemical Society* 128, 14506–14518 (2006). <https://doi.org/10.1021/ja0629460>
41. Shaytan, A.K., Armeev, G.A., Goncarencu, A., *et al.*: Coupling between histone conformations and DNA geometry in nucleosomes on a microsecond timescale: Atomistic insights into nucleosome functions. *Journal of Molecular Biology* 428, 221–237 (2016). <https://doi.org/10.1016/j.jmb.2015.12.004>
42. Sun, R., Li, Z., Bishop, T.C.: TMB library of nucleosome simulations. *Journal of Chemical Information and Modeling* (2019). <https://doi.org/10.1021/acs.jcim.9b00252>
43. Voevodin, V.V., Antonov, A.S., Nikitenko, D.A., *et al.*: Supercomputer Lomonosov-2: Large scale, deep monitoring and fine analytics for the user community. *Supercomputing Frontiers and Innovations* 6, 4–11 (2019). <https://doi.org/10.14529/jsfi190201>
44. Winogradoff, D., Aksimentiev, A.: Molecular mechanism of spontaneous nucleosome unraveling. *Journal of Molecular Biology* 431, 323–335 (2019). <https://doi.org/10.1016/j.jmb.2018.11.013>
45. Yoo, J., Aksimentiev, A.: New tricks for old dogs: Improving the accuracy of biomolecular force fields by pair-specific corrections to non-bonded interactions. *Physical Chemistry Chemical Physics* 20, 8432–8449 (2018). <https://doi.org/10.1039/c7cp08185e>
46. Yoo, J., Winogradoff, D., Aksimentiev, A.: Molecular dynamics simulations of DNADNA and DNAProtein interactions. *Current Opinion in Structural Biology* 64, 88–96 (2020). <https://doi.org/10.1016/j.sbi.2020.06.007>
47. Zinchenko, A.A., Yoshikawa, K.: Na<sup>+</sup> shows a markedly higher potential than K<sup>+</sup> in DNA compaction in a crowded environment. *Biophysical Journal* 88, 4118–4123 (2005). <https://doi.org/10.1529/biophysj.104.057323>

# Molecular Modeling of Penicillin Acylase Binding with a Penicillin Nucleus by High Performance Computing: Can Enzyme or its Mutants Possess $\beta$ -lactamase Activity?

*Evgeny M. Kirilin*<sup>1,2</sup> , *Anna A. Bochkova*<sup>1</sup>, *Nikolay V. Panin*<sup>1</sup>,  
*Igor V. Pochinok*<sup>2</sup>, *Vytas Švedas*<sup>2,3</sup> 

© The Authors 2022. This paper is published with open access at SuperFri.org

High-performance computing has been used for molecular modeling of penicillin acylase interaction with a penicillin nucleus 6-aminopenicillanic acid (6-APA) to assess whether the wild-type enzyme or its mutants could possess  $\beta$ -lactamase activity. Applying parallel hybrid GPU/CPU computing technologies for metadynamics calculations with the PLUMED library in conjunction with AMBER software suite it has been shown that trace amounts of wild-type penicillin acylase-6-APA complexes leading to a  $\beta$ -lactamase reaction can be formed. Higher  $\beta$ -lactamase activity can be observed in enzyme mutants by introducing charged residue in the substrate binding pocket and its proper positioning with respect to a catalytic nucleophile, including stabilization of the tetrahedral intermediate in the oxyanion hole. Thus, it has been shown that the certain mutations facilitate the orientation of the substrate required for the manifestation of  $\beta$ -lactamase activity in the penicillin acylase active center.

*Keywords:* moonlighting protein, penicillin acylase engineering,  $\beta$ -lactam antibiotics resistance,  $\beta$ -lactamase design, metadynamics.

## Introduction

In the recent years, we are witnessing increasing evidences that a “one protein-one function” paradigm is no longer valid and seems to be rather an exception to the rule. Well documented examples of an ability of the same active center to catalyze different chemical conversions as well as a presence in a protein/enzyme molecule of several active sites responsible for various functional properties or catalytic activities have opened new intriguing period in enzymology. The terms of enzyme promiscuity and protein moonlighting are being filled with new experimental data and theoretical substantiation. This recently observed phenomenon is of not only fundamental, but also of significant practical interest, since the presence of novel activities, in addition to the traditionally known and the attributed canonical one, significantly expands our understanding of the role of enzymes in the functioning of living systems, as well as options for their applications [7, 10, 13]. In this context, it is interesting to take a look at the enzymatic transformations of  $\beta$ -lactam antibiotics.

Two types of enzymes that convert the same substrate penicillin are the most known: penicillin acylases capable of hydrolyzing the more stable amide bond formed by the side chain acyl group with the antibiotic nucleus [20] and  $\beta$ -lactamases which cleave the less stable amide bond of the  $\beta$ -lactam ring in the antibiotic nucleus, thus irreversibly inactivating it [15]. Despite the canonical catalytic activities of these two enzyme families are quite different, there are some experimental facts to suppose that penicillin acylases may have minor  $\beta$ -lactamase activity and thus possess moonlighting properties. It is based on the experimental observation that at the penicillin acylase-catalyzed industrial conversion of penicillin to produce a penicillin nucleus 6-aminopenicillanic acid there is some accumulation of the  $\beta$ -lactam hydrolysis products along

<sup>1</sup>Lomonosov Moscow State University, Belozersky Institute of Physicochemical Biology, Moscow, Russia

<sup>2</sup>Lomonosov Moscow State University, Research Computing Center, Moscow, Russia

<sup>3</sup>Lomonosov Moscow State University, Faculty of Bioengineering and Bioinformatics, Moscow, Russia

with a formation of the target product. In this regard, the question of whether a side reaction of hydrolysis of the  $\beta$ -lactam ring can occur in the active center of penicillin acylase and whether mutations in the enzyme structure can increase such a minor nonspecific activity is of interest. The answer to this question may also help to understand whether, in the course of evolution, with random mutations, penicillin acylase can become a resistance factor to the action of  $\beta$ -lactam antibiotics. Molecular modeling can be used for evaluation of potential moonlighting activity of an enzyme, especially when this activity is much lower compared to the canonical one and therefore is rather complicated to characterize it experimentally what is expected with the  $\beta$ -lactamase activity of penicillin acylases.

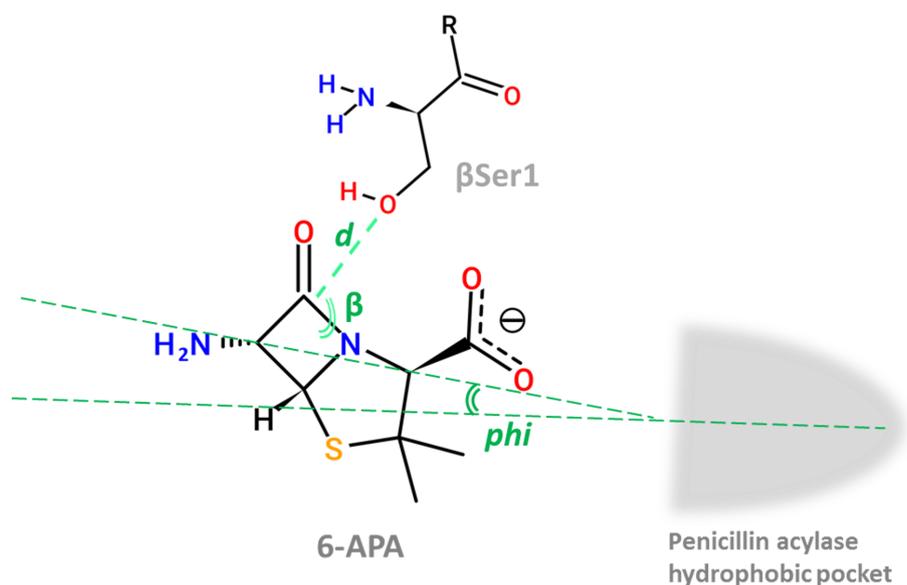
The ability of an enzyme to convert one or another substrate is primarily determined by its recognition and binding in the active center of the enzyme followed by a proper orientation toward the catalytic amino acid residues and conversion to the products. The search over the entire phase space which describes all possible conformational and orientational states of the substrate in the active site is a resource-intensive task, therefore usually specialized methods are used which are limited to either the search and evaluation of the final most favorable bound state (molecular docking [6], MMPBSA calculation [11], thermodynamic integration [22], free energy perturbation [26] methods) or partial analysis of the substrate delivery paths (implementations of steered molecular dynamics [8] and umbrella sampling [12]). One of the advanced implementations of the search for the substrate phase space is funnel metadynamics [14] which allows not only to analyze substrate binding in active centers of different configuration but also explore alternative bindings as well as physically accessible routes of substrate delivery with subsequent estimation of the binding energies for these states. The method is based on the advanced sampling technology – metadynamics implemented in the PLUMED software package which works with a variety of molecular modeling programs [2, 23]. To study the possible major and alternative prereaction states of the penicillin acylase and  $\beta$ -lactamase substrate (6-APA), it is important to take into account not only the configuration of the active site, but also the orientational possibilities of the substrate itself and therefore the redefinition of the so-called collective variables – generalized coordinates of ligand binding in the active site – is needed. Possibilities of recent incorporation of metadynamics to high performance molecular dynamics package AMBER [4] utilizing GPU allow efficient sampling of states without the need to limit the diffusion of the substrate in the solvent when approaching the active site [5]. The protocol used in this article makes it possible, using parallel hybrid GPU/CPU computing technologies, to calculate both a significant amount of the phase space of the enzyme-substrate interaction for both the wild-type enzyme and selected mutant forms with the expected activity.

## 1. Material and Methods

The classical approach to assessing the energy of the chemical stages of enzymatic transformations is the choice of a reaction coordinate and QM/MM simulation to obtain an energy profile, however these calculations are time as well as computational resource consuming and it is practically impossible to perform such studies for a wide range of enzyme mutants and their potential substrates. As a practical alternative the criteria of theoretical chemistry based on the analysis of the so-called reactive conformational states (“Near-to-Attack Conformations”, NAC) can be used to estimate the energy landscape of enzymatic reactions which began to be actively used along with the methods of molecular dynamics [3, 19]. The use of this approach makes it possible not to perform sophisticated QM/MM calculation but to focus on the analysis

of the classical molecular dynamics trajectory in the characterization of native enzymes and the selection of their mutant forms with new properties.

In metadynamics, the free energy profile is reconstructed by introducing an additional bias potential, which makes it easier for the system to overcome potential barriers and move to other regions of the phase space [23]. The bias potential acts through selected collective variables which are functions of the atomic coordinates. Accordingly, NAC-criteria might serve as prospective collective variables. To assess the possibility of enzymatic hydrolysis of amides in metadynamics (to classify each particular conformation as capable or not capable of entering into a subsequent chemical transformation) it is possible to propose two criteria (Fig. 1) – the angle of the nucleophilic attack on the plane of the hydrolysable amide bond ( $\beta$ ) and the distance between a nucleophile atom and an accepting amide carbon atom ( $d$ ). A conformation is considered reactive if the angle  $\beta$  lies in the range of 70–110° and the distance  $d$  is less than or equal to the sum of the van der Waals radii of the nucleophile and electrophile. In the case of hydrolysis of 6-APA by penicillin acylase the nucleophile is the oxygen atom of the hydroxyl group of the  $\beta$ Ser1 residue, the electrophile is the carbon atom of the amide bond in the  $\beta$ -lactam group and the threshold value for  $d$  is 3.5Å. It is justified to use the angle of rotation of the substrate relative to the axis (the line connecting bW154 and bY376 residues, the residues are labeled with small letter depicting the chain, if available, followed by one-letter code of original amino acid residue and its position; one-letter code of new residue is added to the end if mutation is considered) passing through the center of the penicillin acylase hydrophobic pocket as the third collective variable  $\phi$  (the orientation of the substrate with the amide part directed into the hydrophobic pocket corresponds to  $\beta=0^\circ$ ) which in addition to the positional search with respect to the variable  $d$  will expand the search for various substrate orientations.



**Figure 1.** Collective variables ( $d$ ;  $\beta$ ;  $\phi$ ) used to perform metadynamics in the penicillin acylase-6-APA complex.  $d$  corresponds to the distance between a nucleophile atom and an accepting amide carbon atom,  $\beta$  corresponds to the angle of the nucleophilic attack on the plane of the hydrolysable amide bond and  $\phi$ ( $\phi$ ) is the angle of rotation of the substrate relative to the axis passing through the center of the penicillin acylase hydrophobic pocket

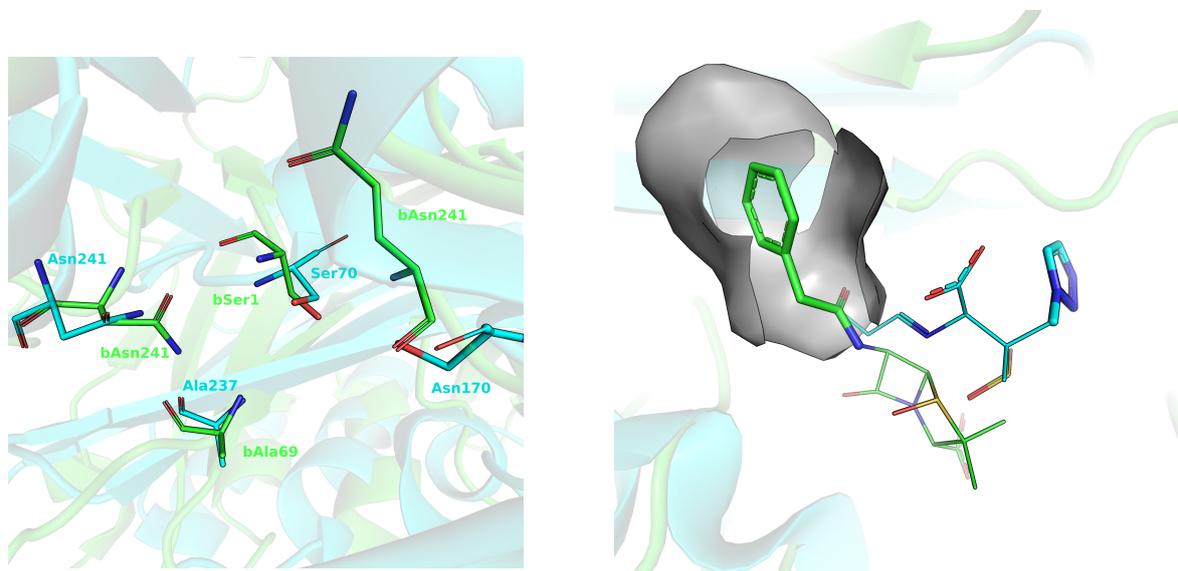
To calculate the energy landscape of various enzyme-substrate complexes and determine the reaction states the metadynamics has been performed using the phase space of three collective variables ( $d;\beta;\phi$ ) while imposing the wall potential to restrict nucleophilic attack distance  $d$  and prevent the substrate to leave active site area ( $> 10\text{\AA}$ ).

For molecular dynamics calculations the AMBER18 package was used [4]; for metadynamics calculations – the open-source, community-developed PLUMED library [2] version 2.6 [23] in conjunction with AMBER18. Recently developed PLUMED incorporation to AMBERs pmemd program to resolve performance issues within hybrid molecular dynamics/metadynamics runs utilizing GPU units has helped to gain free energy landscape convergence within valuable timing period [5]. As a result, the computational performance of the AMBER18+PLUMED bundle performed approximately 4 times faster than the NAMD 2.13 [17]+PLUMED bundle under similar conditions (the systems under study of the order of 80,000 atoms and more). The enzyme-6-APA systems (native or mutant forms of the enzyme) were preliminarily prepared by removing low molecular substances from the initial structure (RCSB 1gm9); protons were added to the protein according to pH 7.5 using the pdb2pqr30 utility [9]; geometry of 6-APK was optimized in package GAMESS [21] using hybrid density functional B3LYP with the 6-31+G\* basis set followed by deriving the point atomic charges using RESP method at the RED server [24]. A system was created for enzyme-substrate complexes that included the TIP3P solvent in which the distance from the edge of periodic box to any atom of solute was no less than  $15\text{\AA}$ . Cl<sup>-</sup> and Na<sup>+</sup> ions (up to 0.1M concentration) were added to the system to create conditions close to physiological ones. The total charge of the system was adjusted to zero by adding corresponding ions. Energy minimization was carried out for each system (2500 steps) followed by heating to 300 K for 60 ps, lifting of limitations from heavy atoms for 290 ps and relaxation for 5 ns in the NPT ensemble with a gradual adjustment of the density to the constant value of  $1.04\text{ g/cm}^3$ . Standard simulation in solution was performed with the AMBER18 package at 300 K with the integration step of 2 fs under conditions of the NVT ensemble. The system temperature was controlled using a Langevin thermostat [16]. The height of the Gaussian hills in metadynamics was set to the initial value of 0.6 kcal/mol; the width was set 0.01, 0.0075 and 0.1 1 for  $d$ ,  $\beta$  and  $\phi$ , respectively. The well-tempered variant of metadynamics was used by setting BIAS-FACTOR value to 13 [1]. New potentials were added every 300 steps at the integration step of 2 fs. The temperature of simulation was set to 300 K. Plumed library restraining potential UPPER\_WALLS was applied to distance collective variable at  $10\text{\AA}$  with KAPPA value 2500 to limit the phase space accessible during simulation. Ten parallel calculations with shared metadynamics potential (multiple-walkers technique) were used to accelerate calculations [18]. The convergence of metadynamics was estimated from the time dependence of the difference in free energy values at each minima point of ( $d;\beta;\phi$ ) energy landscape.

## 2. Results and Discussion

When considering the catalytic mechanisms of penicillin acylase and  $\beta$ -lactamase it should be noted, first, that the nucleophile of penicillin acylase is formed by N-terminal serine residue while  $\beta$ -lactamase uses a non-terminal serine residue as a nucleophilic reagent. Pairwise structural superposition of enzymes by key catalytic residues including serine residues as well as residues involved in the stabilization of the oxyanion bAla69 and Ala237, bAsn241 and Asn170, bGln23 and Ser130 of penicillin acylase and  $\beta$ -lactamase, respectively, demonstrates spatial matching of the catalytic cores but a low similarity index for the entire protein structure (Fig. 2a). The

enzymes do not share common protein fold and belong to different superfamilies while sharing analogous substrate amide group recognition. Enzyme substrates from superimposed crystallographic structures are rotated relative to each other by 180 degrees: the acyl part of the substrate is located in the hydrophobic pocket of penicillin acylase contrary to bulk water orientation in  $\beta$ -lactamase (Fig. 2b).



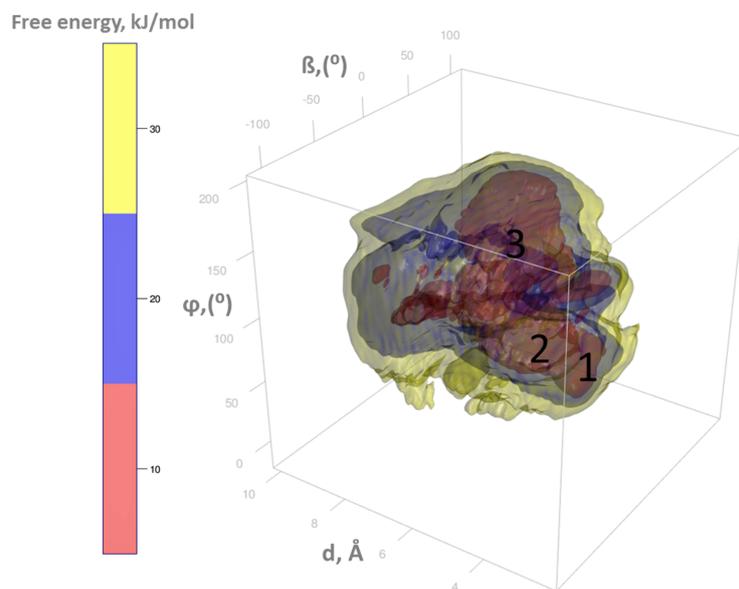
(a) bSer1-Ser70, bAla69-Ala237, bAsn241-Asn170, bGln23-Ser130 pairs of residues of penicillin acylase (1gm9, green sticks) and  $\beta$ -lactamase (4a5r, cyan sticks), respectively, were used for superposition (RMSD = 1.2Å)

(b) Substrates orientation in the superimposed crystallographic structures of penicillin acylase (1gm9, green) and  $\beta$ -lactamase (4a5r, cyan). The acyl groups of the substrates are shown as bold sticks; the surface represents penicillin acylase hydrophobic pocket location

**Figure 2.** Pairwise structural superposition of penicillin acylase and  $\beta$ -lactamase by active site amino acid residues

The study of the orientational and conformational states of 6-APA in the active site of penicillin acylase have been carried out by means of metadynamics in the selected collective variables ( $d; \beta; \phi$ ) followed by reconstruction of the free energy landscape. The review of the landscape obtained for wild-type penicillin acylase reveals lack of stabilized regions corresponding to the required orientation of 6-APA to support  $\beta$ -lactamase activity (Fig. 3). For the hydrolysis of the  $\beta$ -lactam ring to proceed, it is necessary the attack angle  $\beta$  be around  $90^\circ$  while the nucleophilic attack distance  $d$  be less than  $3.5\text{\AA}$ : in the case of wild-type penicillin acylase low-energy states conferring such parameters are absent. The shortest distance from catalytic bSer1 to amide bond carbon  $d$  corresponded to 2 regions – region 1 and region 2 in the Fig. 3. In region 1, the distance  $d$  varies from  $3.5$  to  $5\text{\AA}$ , the attack angle  $\beta$  do not exceed  $25^\circ$ . In region 2, the distance varies from  $2.5$  to  $3.5\text{\AA}$  but attack angle  $\beta$  do not exceed  $50^\circ$  turning the substrates amide group to the bulk water. This orientation is supported by the formation of a hydrogen bond between the carboxyl group of 6-APA and the OH group of bSer386 however no states are observed in which the stabilization of the transition state in the oxyanion hole is possible. Even though it is possible to isolate individual frames of the molecular dynamics trajectory that corresponds to the correct orientation of the substrate ( $d < 3.5\text{\AA}$ ,  $\beta \approx 90^\circ$ ) as well as formation of hydrogen bonds between 6-APA carbonyl oxygen and the bAla69 main chain hydrogen and bAsn241 side

chain hydrogen it can only explain the trace activity of the enzyme in purified preparations (purification by affinity chromatography, unpublished data).

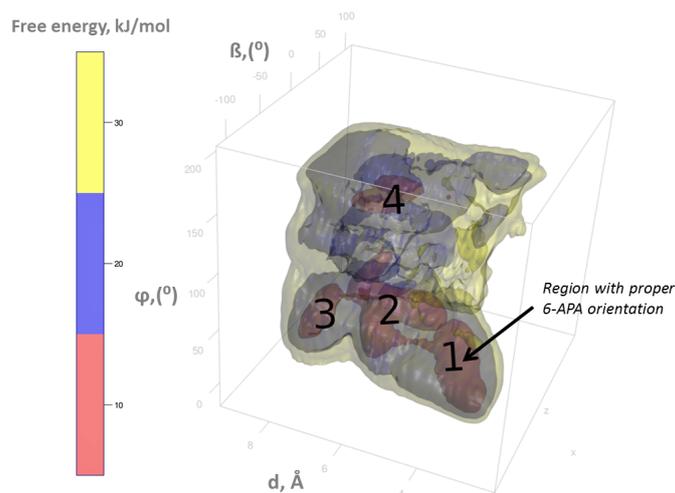


**Figure 3.** Free Energy Landscape of the 6-APA/(wild type)penicillin acylase binding is reconstructed using well-tempered metadynamics as a function of nucleophilic attack distance  $d$ , the attack angle  $\beta$  and the angle of rotation of the substrate  $\phi$ . The landscape is rendered using isometric surfaces with a step of 10 kJ/mol. Areas corresponding to major minima are labeled by numbers

Mutant forms of penicillin acylase capable of efficiently catalyzing alternative reactions could evolve in the course of natural selection in a variety of ways. Thus for  $\beta$ -lactamase activity it would be reasonable to introduce a new nucleophilic serine residue next to the amide bond of the  $\beta$ -lactam ring, however, then it would be necessary also to move the oxyanion hole due to additional mutations. Such an interference could affect the maturation of the enzyme. The shortest one for  $\beta$ -lactamase reaction seems to be introduction of the charged residue into the substrate binding pocket thus moving the  $\beta$ -lactam amide bond toward bSer1 nucleophilic attack. Given the trace presence of the corresponding orientations of 6-APA in the active site of the wild-type penicillin acylase an evolutionary path leading to further stabilization of these states would be an option. In order to identify corresponding penicillin acylase mutants with  $\beta$ -lactamase activity, it is necessary to select substitutions that allow to bind and orient the 6-APA in the enzymes active site properly directing the substrate carboxyl group with respect to the catalytic nucleophile residue, and then, if necessary, select additional mutations. In order to check adequacy of this hypothesis we have modelled the introduction of a positively charged amino acid residue within the substrate binding pocket.

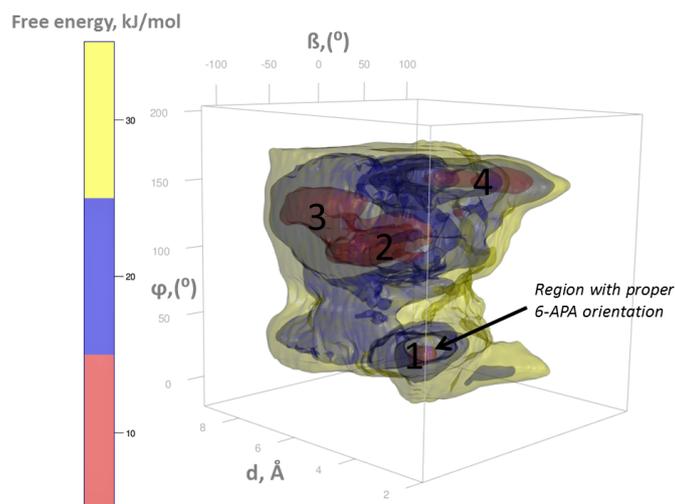
With the introduction of bV56R substitution stabilized states appear on the free energy landscape which correspond to the substrate orientation required for  $\beta$ -lactam hydrolysis to occur (Fig. 4, the region 1).

The nucleophilic attack distance  $d$  in these states reaches the limit of  $2.75\text{\AA}$ , the attack angle  $\beta$  is  $60\text{--}90^\circ$ , the substrate is oriented in the required manner relative to the hydrophobic pocket ( $\phi = 40\text{--}50^\circ$ ).



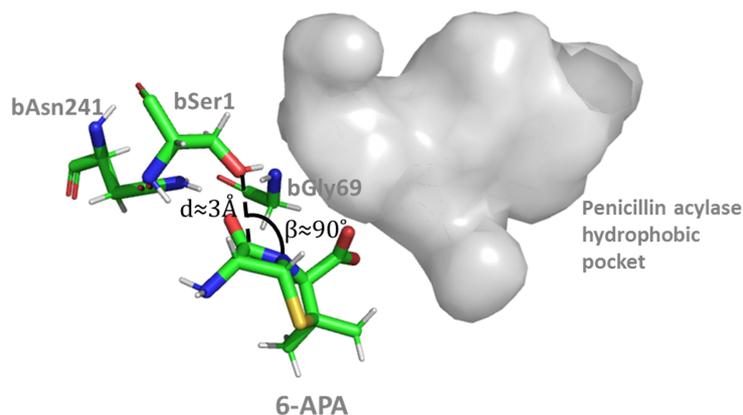
**Figure 4.** Free Energy Landscape of the 6-APA/(bV56R)penicillin acylase binding is reconstructed using well-tempered metadynamics as a function of nucleophilic attack distance  $d$ , the attack angle  $\beta$  and the angle of rotation of the substrate  $\phi$ . The landscape is rendered using isometric surfaces with a step of 10 kJ/mol. Areas corresponding to major minima are labeled by numbers

We also proposed to additionally introduce the substitution of bA69 for a less bulky glycine residue (bA69G) so that the substrate molecule could more effectively approach the catalytic bSer1 from solution. The double mutant bV56R+bA69G retains a region with the 6-APA orientation required for the reaction to proceed (indicated by the arrow in Fig. 5) in which the substrate is oriented with the carboxyl group into the hydrophobic pocket: the collective variables vary within the appropriate limits ( $d = 2.75\text{--}3.5\text{\AA}$ ,  $\beta = 60\text{--}90$ ,  $\phi = 40\text{--}60$ ). For double



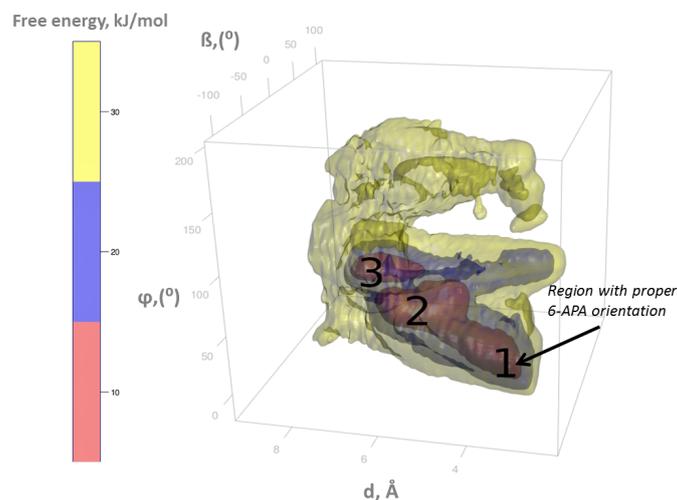
**Figure 5.** Free Energy Landscape of the 6-APA/(bV56R+bA69G)Penicillin Acylase binding is reconstructed using well-tempered metadynamics as a function of nucleophilic attack distance  $d$ , the attack angle  $\beta$  and the angle of rotation of the substrate  $\phi$ . The landscape is rendered using isometric surfaces with a step of 10 kJ/mol. Areas corresponding to major minima are labeled by numbers

penicillin acylase mutant 6-APA orientation completely fulfill stabilization of amide bond within oxyanion hole directing the carbonyl oxygen towards bAla69 and bAsn241 residues (Fig. 6).



**Figure 6.** Orientation of 6-APA in the active site of the penicillin acylase double mutant bA69G+bV56R which completely fulfills  $\beta$ -lactam hydrolysis transition state stabilization by oxyanion hole formed by bAla69 and bAsn241 residues

When analyzing the states corresponding to different energy minima on the free energy landscape, a large number of conformations were revealed, in which the carboxyl group of 6-APA forms a hydrogen bond with the aR145 residue. In order to reduce the proportion of such states in favor of the more preferable orientation for the alternative  $\beta$ -lactamase reaction (when the carboxyl group is directed to the bV56 residue), it is also reasonable to propose aR145G substitution. For the bA69G+bV56R+aR145G mutant the free energy landscape has also been constructed to support the major minima of  $\beta$ -lactamase-like substrate orientation (Fig. 7, indicated by the arrow).



**Figure 7.** Free Energy Landscape of the 6-APA/(bV56R+bA69G+aR145G)penicillin acylase binding is reconstructed using well-tempered metadynamics as a function of nucleophilic attack distance  $d$ , the attack angle  $\beta$  and the angle of rotation of the substrate  $\phi$ . The landscape is rendered using isometric surfaces with a step of 10 kJ/mol. Areas corresponding to major minima are labeled by numbers

## Conclusion

High-performance computing has been used for molecular modeling of penicillin acylase interaction with a penicillin nucleus 6-aminopenicillanic acid (6-APA) to assess whether the wild-type enzyme or its mutants could possess  $\beta$ -lactamase activity. Despite the trace amounts of wild-type penicillin acylase-6-APA complexes leading to a  $\beta$ -lactamase reaction have been shown to be formed, the manifestation of  $\beta$ -lactamase activity by a wild-type penicillin acylase is unlikely. However  $\beta$ -lactamase activity can be observed due to penicillin acylase mutations when charged residue is introduced in the substrate binding pocket leading to its proper positioning with respect to a bSer1 nucleophilic attack, including stabilization of the tetrahedral intermediate in the oxyanion hole. It has been shown that the mutations bV56R, bA69G and aR145G can facilitate the orientation of the substrate required for the manifestation of  $\beta$ -lactamase activity in the penicillin acylase active center.

## Acknowledgements

This work was supported by the Russian Foundation for Basic Research (grant № 20-04-01119). The research was carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University [25].

*This paper is distributed under the terms of the Creative Commons Attribution-Non Commercial 3.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.*

## References

1. Barducci, A., Bussi, G., Parrinello, M.: Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Physical Review Letters* 100(2), 020603 (Jan 2008). <https://doi.org/10.1103/PhysRevLett.100.020603>
2. Bonomi, M., Bussi, G., Camilloni, C., *et al.*: Promoting transparency and reproducibility in enhanced molecular simulations. *Nature Methods* 16(8), 670–673 (Aug 2019). <https://doi.org/10.1038/s41592-019-0506-8>
3. Bruice, T.C.: A View at the Millennium: the Efficiency of Enzymatic Catalysis. *Accounts of Chemical Research* 35(3), 139–148 (Mar 2002). <https://doi.org/10.1021/ar0001665>
4. Case, D.A., Aktulga, H.M., Belfon, K., *et al.*: Amber 2018. University of California Press (2018), <http://ambermd.org/doc12/Amber18.pdf>
5. Drobot, V.V., Kirilin, E.M., Kopylov, K.E., Švedas, V.K.: PLUMED Plugin Integration into High Performance Pmemd Program for Enhanced Molecular Dynamics Simulations. *Supercomputing Frontiers and Innovations* 8(4), 94–99 (2021). <https://doi.org/10.14529/jsfi210408>
6. Gohlke, H., Klebe, G.: Approaches to the Description and Prediction of the Binding Affinity of Small-Molecule Ligands to Macromolecular Receptors. *Angewandte Chemie International Edition* 41(15), 2644–2676 (2002). [https://doi.org/10.1002/1521-3773\(20020802\)41:15<2644::AID-ANIE2644>3.0.CO;2-0](https://doi.org/10.1002/1521-3773(20020802)41:15<2644::AID-ANIE2644>3.0.CO;2-0)

7. Hult, K., Berglund, P.: Enzyme promiscuity: mechanism and applications. *Trends in Biotechnology* 25(5), 231–238 (May 2007). <https://doi.org/10.1016/j.tibtech.2007.03.002>
8. Izrailev, S., Stepaniants, S., Israilewitz, B., *et al.*: Steered Molecular Dynamics. In: Deuffhard, P., Hermans, J., Leimkuhler, B., *et al.* (eds.) *Computational Molecular Dynamics: Challenges, Methods, Ideas*. pp. 39–65. *Lecture Notes in Computational Science and Engineering*, Springer, Berlin, Heidelberg (1999). [https://doi.org/10.1007/978-3-642-58360-5\\_2](https://doi.org/10.1007/978-3-642-58360-5_2)
9. Jurrus, E., Engel, D., Star, K., *et al.*: Improvements to the APBS biomolecular solvation software suite. *Protein Science* 27(1), 112–128 (2018). <https://doi.org/10.1002/pro.3280>
10. Khersonsky, O., Tawfik, D.S.: Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annual Review of Biochemistry* 79, 471–505 (2010). <https://doi.org/10.1146/annurev-biochem-030409-143718>
11. Kollman, P.A., Massova, I., Reyes, C., *et al.*: Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Accounts of Chemical Research* 33(12), 889–897 (Dec 2000). <https://doi.org/10.1021/ar000033j>
12. Kumar, S., Rosenberg, J.M., Bouzida, D., *et al.*: THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry* 13(8), 1011–1021 (1992). <https://doi.org/10.1002/jcc.540130812>
13. Leveson-Gower, R.B., Mayer, C., Roelfes, G.: The importance of catalytic promiscuity for enzyme design and evolution. *Nature Reviews Chemistry* 3(12), 687–705 (Dec 2019). <https://doi.org/10.1038/s41570-019-0143-x>
14. Limongelli, V., Bonomi, M., Parrinello, M.: Funnel metadynamics as accurate binding free-energy method. *Proceedings of the National Academy of Sciences* 110(16), 6358–6363 (Apr 2013). <https://doi.org/10.1073/pnas.1303186110>
15. Livermore, D.M.: Beta-lactamase-mediated resistance and opportunities for its control. *Journal of Antimicrobial Chemotherapy* 41(suppl\_4), 25–41 (Jun 1998). [https://doi.org/10.1093/jac/41.suppl\\_4.25](https://doi.org/10.1093/jac/41.suppl_4.25)
16. Paterlini, M.G., Ferguson, D.M.: Constant temperature simulations using the Langevin equation with velocity Verlet integration. *Chemical Physics* 236(1), 243–252 (Sep 1998). [https://doi.org/10.1016/S0301-0104\(98\)00214-6](https://doi.org/10.1016/S0301-0104(98)00214-6)
17. Phillips, J.C., Hardy, D.J., Maia, J.D.C., *et al.*: Scalable molecular dynamics on CPU and GPU architectures with NAMD. *The Journal of Chemical Physics* 153(4), 044130 (Jul 2020). <https://doi.org/10.1063/5.0014475>
18. Raiteri, P., Laio, A., Gervasio, F.L., *et al.*: Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics. *The Journal of Physical Chemistry B* 110(8), 3533–3539 (Mar 2006). <https://doi.org/10.1021/jp054359r>

19. Sadiq, S.K., Coveney, P.V.: Computing the Role of Near Attack Conformations in an Enzyme-Catalyzed Nucleophilic Bimolecular Reaction. *Journal of Chemical Theory and Computation* 11(1), 316–324 (Jan 2015). <https://doi.org/10.1021/ct5008845>
20. Sakaguchi, K., Murao, S.: A Preliminary Report on a New Enzyme, Penicillin-amidase. *Journal of the agricultural chemical society of Japan* 23(9), 411–411 (1950). <https://doi.org/10.1271/nogeikagaku1924.23.411>
21. Schmidt, M.W., Baldrige, K.K., Boatz, J.A., *et al.*: General atomic and molecular electronic structure system. *Journal of Computational Chemistry* 14(11), 1347–1363 (1993). <https://doi.org/10.1002/jcc.540141112>
22. Straatsma, T.P., McCammon, J.A.: Multiconfiguration thermodynamic integration. *The Journal of Chemical Physics* 95(2), 1175–1188 (Jul 1991). <https://doi.org/10.1063/1.461148>
23. Tribello, G.A., Bonomi, M., Branduardi, D., *et al.*: PLUMED 2: New feathers for an old bird. *Computer Physics Communications* 185(2), 604–613 (Feb 2014). <https://doi.org/10.1016/j.cpc.2013.09.018>
24. Vanquelef, E., Simon, S., Marquant, G., *et al.*: R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Research* 39(suppl.2), W511–W517 (Jul 2011). <https://doi.org/10.1093/nar/gkr288>
25. Voevodin, V.V., Antonov, A.S., Nikitenko, D.A., *et al.*: Supercomputer Lomonosov-2: Large Scale, Deep Monitoring and Fine Analytics for the User Community. *Supercomputing Frontiers and Innovations* 6(2), 4–11 (Jun 2019). <https://doi.org/10.14529/jsfi190201>
26. Zwanzig, R.W.: High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *The Journal of Chemical Physics* 22(8), 1420–1426 (Aug 1954). <https://doi.org/10.1063/1.1740409>

# Search for Ligands Complementary to the 430-cavity of Influenza Virus Neuraminidase by Virtual Screening

*Dmitry K. Nilov*<sup>1,2</sup>, *Michaela Schmidtke*<sup>3</sup>, *Vadim A. Makarov*<sup>4</sup>,  
*Vytas K. Švedas*<sup>2,5</sup>

© The Authors 2022. This paper is published with open access at SuperFri.org

An anthrapyrazole derivative STK663786 has been identified as a selective ligand of the so-called 430-cavity of influenza virus neuraminidase at virtual screening of a library of low-molecular-weight compounds. It is able to form favorable contacts with hydrophobic residues as well as cation- $\pi$  interaction and hydrogen bonds with the polar Arg371 residue. The experimentally determined EC<sub>50</sub> values have been found to be 19 and 30  $\mu$ M for viruses H1N1 and H3N2, respectively. Complementarity of STK663786 to the 430-cavity adjacent to the sialic acid binding subsite in the active center of neuraminidase makes this compound a valuable structural fragment at construction of bifunctional inhibitors of the enzyme.

*Keywords: influenza, neuraminidase, 430-cavity, inhibitor, anthrapyrazole.*

## Introduction

Seasonal flu affects nearly 10% of the world's population every year, and the pandemic influenza virus strains pose serious danger worldwide [1–3]. Two glycoproteins on the surface of the viral envelope, hemagglutinin and neuraminidase, are responsible for infectivity. Hemagglutinin binds to terminal sialic acid residues of epithelial receptors, and then virus enters the cell by endocytosis. The neuraminidase enzyme, on the contrary, cleaves sialic acid residues, which promotes the release of newly formed viral particles from the cell surface [4, 5]. The widely used anti-influenza drugs zanamivir and oseltamivir, being structurally similar to the sialic acid residue in natural enzyme substrates, competitively inhibit neuraminidase activity [6, 7]. However, virus strains resistant to these drugs are quickly emerging due to mutations in the sialic acid binding subsite [8, 9].

Another putative binding site of neuraminidase inhibitors, the so-called 430-cavity, is formed by a series of hydrophobic residues [10–13]. The interface between the sialic acid binding subsite and 430-cavity is formed by three arginine residues (Arg118, Arg292, Arg371), with the Arg371 residue making a decisive contribution to the positioning of the carboxyl group of the substrate or competitive inhibitors due to the formation of two hydrogen bonds [14]. In the present study, we have performed virtual screening to identify ligands complementary to the 430-cavity, which can interact with both the hydrophobic residues of the cavity and the guanidine group of Arg371.

## 1. Results and Discussion

The molecular model of N1 neuraminidase was constructed based on the 3b7e crystal structure [15] using the Amber 12 package [16]. Hydrogen atoms were added to the protein structure, and then it was solvated by a layer of TIP3P water. The energy minimization of the obtained system included 2500 steps of the steepest descent algorithm followed by 2500 steps of conjugate

<sup>1</sup>Lomonosov Moscow State University, Belozersky Institute of Physicochemical Biology, Moscow, Russia

<sup>2</sup>Lomonosov Moscow State University, Research Computing Center, Moscow, Russia

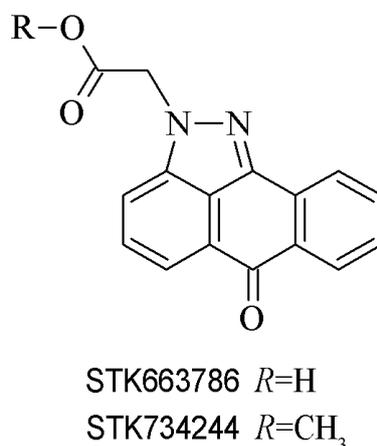
<sup>3</sup>Department of Virology and Antiviral Therapy, Jena University Hospital, Jena, Germany

<sup>4</sup>Research Center of Biotechnology of the Russian Academy of Sciences, Moscow, Russia

<sup>5</sup>Lomonosov Moscow State University, Faculty of Bioengineering and Bioinformatics, Moscow, Russia

gradient algorithm, and the heavy atoms of the protein were being fixed by positional restraints. The *ff99SB* force field was applied to describe the protein by molecular mechanics [17]. The optimized N1 model was used in docking experiments described below.

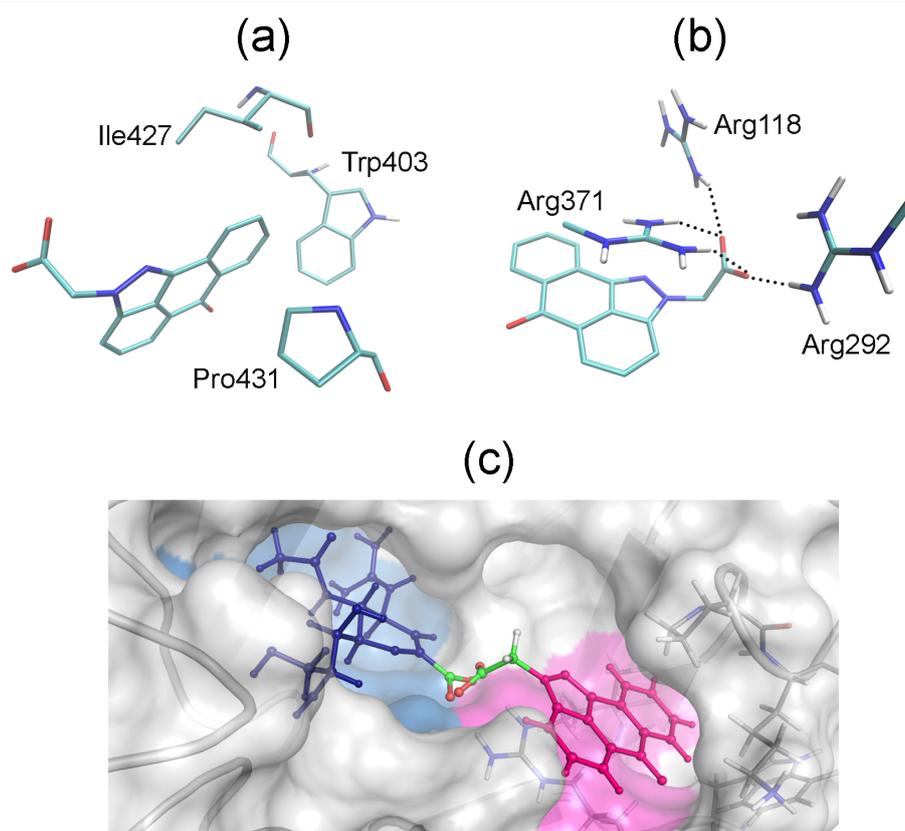
Virtual screening was carried out using a subset of low-molecular-weight compounds retrieved from the Vitas-M commercial library (<https://vitasmlab.biz>). Molecules containing a carboxyl group and obeying the rule of three (molecular weight < 300,  $\log P \leq 3$ , hydrogen bond donors  $\leq 3$ , hydrogen bond acceptors  $\leq 3$ , rotatable bonds  $\leq 3$ ) [18] were retrieved by a substructure search in ACD/ChemFolder (<http://www.acdlabs.com>): in total, 3734 compounds. 3D structures of the compounds were generated with the CORINA software [19]. Each compound was docked into the N1 active site using Lead Finder 1.1.15 [20, 21], and then the modeled positions were subjected to structural filtration [22] to select molecules capable of forming hydrophobic contacts with 430-cavity residues as well as hydrogen bonds with the Arg371 guanidinium group. For this purpose, the vsFilt software [23] integrated into a high-throughput virtual screening platform of the Lomonosov Moscow State University supercomputer was used. Visual inspection of the selected molecules allowed us to identify the anthrapyrazole derivative STK663786 (Fig. 1;  $\Delta G^{calc} = -5.9$  kcal/mol) as a promising ligand of the 430-cavity.



**Figure 1.** Chemical structure of anthrapyrazole derivatives

In the obtained model, the anthrapyrazole scaffold of STK663786 forms favorable hydrophobic contacts with the Trp403, Ile427, and Pro431 residues (Fig. 2a), while the carboxyl group is involved in a network of hydrogen bonds with the arginine triad residues Arg118, Arg292, and Arg371 (Fig. 2b). In addition, effective binding of the STK663786 scaffold is facilitated by a cation- $\pi$  interaction with the guanidine group of Arg371. The positioning of the ligand in the 430-cavity makes it possible to elongate its structure towards the adjacent sialic acid binding subsite. The carboxyl groups of zanamivir and STK663786 occupy nearly the same position (Fig. 2c), and therefore we believe that prototypes of bifunctional inhibitors can be constructed by combining two scaffolds, zanamivir (or its structural analogue) and STK663786, into a single chimeric molecule. An appropriate ester linker or isosteric analogues of the carboxyl linker (sulfo, phosphono, etc.) capable of interacting with the arginine triad may be of interest.

A preliminary *in vitro* study of STK663786 has confirmed its inhibitory effect against influenza viruses. In cytopathic effect (CPE) assay [25], the  $EC_{50}$  values were found to be 19 and 30  $\mu M$  for viruses H1N1 (8178/09) and H3N2 (HK/68), respectively. We have also tested a methyl ester of STK663786, compound STK734244 (Fig. 1). It showed an inhibitory effect ( $EC_{50}$  equal to 53 and 85  $\mu M$  for viruses H1N1 and H3N2, respectively), though less pronounced compared



**Figure 2.** Binding mode of STK663786 to the N1 neuraminidase: (a) Interactions with hydrophobic residues of the 430-cavity, for clarity, non-polar hydrogen atoms are omitted; (b) Interactions with the arginine triad, guanidinium groups of arginine residues are shown; (c) Comparison of zanamivir and STK663786 positions in the neuraminidase active center: the sialic acid binding subsite and zanamivir molecule are colored blue, the 430-cavity and STK663786 molecule are colored magenta. Carboxyl substituents forming hydrogen bonds with Arg371 are shown in green. The figure was prepared using VMD and PyMOL [24]

to STK663786. Apparently, this is due to the absence of a negative charge on the esterified carboxyl group and the weakening of the interaction with the positively charged arginine triad. Thus, virtual screening made it possible to identify the ligand STK663786, complementary to the 430-cavity, which by itself inhibits neuraminidase, disrupting the interaction between the substrate and the arginine triad, and can be used as a structural fragment of a chimeric molecule of bifunctional inhibitors capable of occupying both the cavity-430 and sialic acid binding subsite in the neuraminidase active center.

## Acknowledgements

This work was supported by the Russian Science Foundation (grant No. 21-71-30003). The research was carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University.

*This paper is distributed under the terms of the Creative Commons Attribution-Non Commercial 3.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.*

## References

1. Zambon, M.C.: Epidemiology and pathogenesis of influenza. *J. Antimicrob. Chemother.* 44, 3–9 (1999). [https://www.doi.org/10.1093/jac/44.suppl\\_2.3](https://www.doi.org/10.1093/jac/44.suppl_2.3)
2. Layne, S.P., Monto, A.S., Taubenberger, J.K.: Pandemic influenza: an inconvenient mutation. *Science*. 323(5921), 1560–1561 (2009). <https://www.doi.org/10.1126/science.323.5921.1560>
3. Hussain, M., Galvin, H.D., Haw, T.Y., *et al.*: Drug resistance in influenza A virus: the epidemiology and management. *Infect. Drug Resist.* 10, 121–134 (2017). <https://www.doi.org/10.2147/IDR.S105473>
4. Bouvier, N.M., Palese, P.: The biology of influenza viruses. (2008) *Vaccine*. 26, D49–D53 (2008). <https://www.doi.org/10.1016/j.vaccine.2008.07.039>
5. Herold, S., Becker, C., Ridge, K.M., Budinger, G.R.: Influenza virus-induced lung injury: pathogenesis and implications for treatment. *Eur. Respir. J.* 45(5), 1463–1478 (2015). <https://www.doi.org/10.1183/09031936.00186214>
6. von Itzstein, M., Wu, W.Y., Kok, G.B., *et al.*: Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature*. 363(6428), 418–423 (1993). <https://www.doi.org/10.1038/363418a0>
7. Lew, W., Chen, X., Kim, C.U.: Discovery and development of GS 4104 (oseltamivir): an orally active influenza neuraminidase inhibitor. *Curr. Med. Chem.* 7(6), 663–672 (2000). <https://www.doi.org/10.2174/0929867003374886>
8. Thorlund, K., Awad, T., Boivin, G., Thabane, L.: Systematic review of influenza resistance to the neuraminidase inhibitors. *BMC Infect. Dis.* 11, 134 (2011). <https://www.doi.org/10.1186/1471-2334-11-134>
9. Hurt A.C., Besselaar T.G., Daniels R.S., *et al.* Global update on the susceptibility of human influenza viruses to neuraminidase inhibitors, 2014-2015. *Antiviral Res.* 132, 178–185 (2016). <https://www.doi.org/10.1016/j.antiviral.2016.06.001>
10. Amaro, R.E., Minh, D.D., Cheng, L.S., *et al.*: Remarkable loop flexibility in avian influenza N1 and its implications for antiviral drug design. *J. Am. Chem. Soc.* 129(25), 7764–7765 (2007). <https://www.doi.org/10.1021/ja0723535>
11. Landon, M.R., Amaro, R.E., Baron, R., *et al.*: Novel druggable hot spots in avian influenza neuraminidase H5N1 revealed by computational solvent mapping of a reduced and representative receptor ensemble. *Chem. Biol. Drug Des.* 71(2), 106–116 (2008). <https://www.doi.org/10.1111/j.1747-0285.2007.00614.x>
12. Cheng, L.S., Amaro, R.E., Xu, D., *et al.*: Ensemble-based virtual screening reveals potential novel antiviral compounds for avian influenza neuraminidase. *J. Med. Chem.* 51(13), 3878–3894 (2008). <https://www.doi.org/10.1021/jm8001197>
13. Feng, E., Shin, W.J., Zhu, X., *et al.*: Structure-based design and synthesis of C-1- and C-4-modified analogs of zanamivir as neuraminidase inhibitors. *J. Med. Chem.* 56(3), 671–684 (2013). <https://www.doi.org/10.1021/jm3009713>

14. Taylor, N.R., von Itzstein, M.: Molecular modeling studies on ligand binding to sialidase from influenza virus and the mechanism of catalysis. *J. Med. Chem.* 37(5), 616624 (1994). <https://www.doi.org/10.1021/jm00031a011>
15. Xu, X., Zhu, X., Dwek, R.A., *et al.*: Structural characterization of the 1918 influenza virus H1N1 neuraminidase. *J. Virol.* 82(21), 10493–10501 (2008). <https://www.doi.org/10.1128/JVI.00959-08>
16. Case, D.A., Darden, T.A., Cheatham III, T.E., *et al.*: AMBER 12. University of California, San Francisco (2012). <http://ambermd.org/doc12/Amber12.pdf>
17. Hornak, V., Abel, R., Okur, A., *et al.*: Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins.* 65(3), 712–725 (2006). <https://www.doi.org/10.1002/prot.21123>
18. Congreve, M., Carr, R., Murray, C., Jhoti H.: A ‘rule of three’ for fragment-based lead discovery? *Drug Discov. Today.* 8(19), 876–877 (2003). [https://www.doi.org/10.1016/s1359-6446\(03\)02831-9](https://www.doi.org/10.1016/s1359-6446(03)02831-9)
19. Schwab, C.H.: Conformations and 3D pharmacophore searching. *Drug Discov. Today Technol.* 7(4), e245–e253 (2010). <https://www.doi.org/10.1016/j.ddtec.2010.10.003>
20. Stroganov, O.V., Novikov F.N., Stroylov, V.S., *et al.*: Lead finder: an approach to improve accuracy of protein-ligand docking, binding energy estimation, and virtual screening. *J. Chem. Inf. Model.* 48(12), 2371–2385 (2008). <https://www.doi.org/10.1021/ci800166p>
21. Novikov, F.N., Stroylov, V.S., Zeifman, A.A., *et al.*: Lead Finder docking and virtual screening evaluation with Astex and DUD test sets. *J. Comput. Aided Mol. Des.* 26(6), 725–735 (2012). <https://www.doi.org/10.1007/s10822-012-9549-y>
22. Novikov, F.N., Stroylov, V.S., Stroganov, O.V., Chilov, G.G.: Improving performance of docking-based virtual screening by structural filtration. *J. Mol. Model.* 16(7), 1223–1230 (2010). <https://www.doi.org/10.1007/s00894-009-0633-8>
23. Gushchina, I.V., Polenova, A.M., Suplatov, D.A., *et al.*: vsFilt: a tool to improve virtual screening by structural filtration of docking poses. *J. Chem. Inf. Model.* 60(8), 3692–3696 (2020). <https://www.doi.org/10.1021/acs.jcim.0c00303>
24. Martinez, X., Krone, M., Alharbi, N., *et al.*: Molecular graphics: bridging structural biologists and computer scientists. *Structure.* 27(11), 1617–1623 (2019). <https://www.doi.org/10.1016/j.str.2019.09.001>
25. Schmidtke, M., Schnittler, U., Jahn, B., *et al.*: A rapid assay for evaluation of antiviral activity against coxsackie virus B3, influenza virus A, and herpes simplex virus type 1. *J. Virol. Methods.* 95(1-2), 133–143 (2001). [https://www.doi.org/10.1016/s0166-0934\(01\)00305-6](https://www.doi.org/10.1016/s0166-0934(01)00305-6)