

Sequence-based computational protein analysis of the MDR pump AcrAB-TolC from *Escherichia coli*

Karakozova M.^{1, 2*}, Raldugina V.¹, Nazarov P.A.¹

¹ Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia

² All-Russian Scientific Research Institute of Medicinal and Aromatic Plants, Moscow, Russia

* mvk752002@gmail.com

Key words: alignment; MDR pump; AcrAB-TolC; sequence

Motivation and Aim: Multidrug-resistance (MDR) pumps play a vital role in safeguarding bacteria against antibiotics. MDR pumps provide a foundation for non-specific protection of bacteria and are classified into six families of MDR pumps. One of the most extensively researched MDR pumps is the *Escherichia coli* AcrAB-TolC pump [1]. **Methods and Algorithms:** Reference sequences of proteins TolC, AcrA and AcrB for the *E. coli* MG1655 were found and downloaded from the NCBI database in the formats genbank and fasta using the biopython package and Erez.Esearch functions and Entrez.Efetch. To download sequences, analyze and align them the biopython package was used. Sequences were aligned between yourself using BLAST to create a custom database [2].

Results: Studying pump sequences is crucial for comprehending the development of new clinical isolates or superbugs, as efflux-related resistance plays a significant role among the total number of bacteria that become resistant due to antibiotic treatment. The genes responsible for encoding MDR pumps are seen as adaptable and are categorized as “luxury” genes, meaning they help bacteria adapt to changing environmental conditions. Examination of laboratory strains revealed that the sequences of many commonly used laboratory strains (K-12 and B) remain unchanged and show no mutations in the amino acid sequences. The conservation of amino acid sequences in the pumps mirrors the conservation seen in housekeeping genes, underscoring the significant role of these genes in bacteria, even though they are classified as nonessential genes [3].

Were analyzed 8383 TolC protein sequences, 6239 AcrA protein sequences, 3561 AcrB protein sequences. We showed that almost every amino acid is variable.

Conclusion: Having analyzed all known sequences of the AcrAB-TolC pump proteins, we found that the use of a bioinformatics approach to sequence analysis leads to incorrect results due to the quality of the database. Almost every sequence required preliminary validation, which changes the analysis pipeline.

Funding: This work was supported by grant from the Russian Science Foundation 22-15-00099 (Creation of a heterologous MDR pump system for studying bacterial resistance of pathogenic gram-negative bacteria).

References

1. Nazarov P.A., Kuznetsova A.M., Karakozova M.V. Multidrug Resistance Pumps as a Keystone of Bacterial Resistance. *Moscow Univ Biol Sci Bull.* 2022;77(4):193-200. doi 10.3103/S009639252204006X
2. Raldugina V. Sequences of TolC-containing complexes of MDR transporters of various gram-negative bacteria analysis. Diploma work, 2023
3. Karakozova M.V., Nazarov P.A. Conserved sequences of genes coding for the multidrug resistance pump AcrAB-TolC of *Escherichia coli* suggest their involvement into permanent cell “cleaning”. *Bull RSMU.* 2018;2:32-36