

UDC 579.61

Identification of membrane proteins as potential drug targets in *Escherichia coli* ATCC 25922 using *in silico* approaches

I. M. Koleiev¹, S. A. Starosyla², M. V. Protopopov², G. P. Volynets², V. M. Sapelkin², L. V. Pletnova², A. R. Syniugin², N. O. Kachaput², V. I. Matiushok^{2,3}, V.G. Bdzhola², S. M. Yarmoluk²

¹ Educational and Scientific Center "Institute of Biology and Medicine",
Taras Shevchenko National University of Kyiv
64/13, Volodymyrska Str., Kyiv, Ukraine, 01601

² Institute of Molecular Biology and Genetics, NAS of Ukraine
150, Akademika Zabolotnoho Str., Kyiv, Ukraine, 03143

³ LLC "Scientific and service firm "Otava"
150, Akademika Zabolotnoho Str., Kyiv, Ukraine, 03143

Aim. The aim of this study was to identify the novel potential drug targets of *E.coli* ATCC 25922 through subtractive genomic analysis. **Methods.** The identification of non-homologous proteins to the human proteome, search for *E.coli* essential genes and estimation of drug target novelty were performed using BLAST. The unique metabolic pathways identification was done using the data and tools from KEGG (Kyoto Encyclopedia of Genes and Genomes). Prediction of the sub-cellular proteins localization was performed using a combination of the tools PSORTb, CELLO and ngLOC. The homology modeling was performed by web-server I-TASSER, the models being validated using MolProbity web-server. The binding sites were analyzed using Discovery Studio 2017 with web servers ProBis and PrankWeb. **Results.** Proteome of *Escherichia coli* ATCC 25922, which contains 4808 proteins, has been taken to form the initial set. Using the subtractive genome analysis we identified 9 membrane proteins which are essential, non-homologous to human proteome, involved in unique metabolic pathways and are not described as the drug targets. A study of the spatial structure of this proteins showed that 6 of them have binding sites for ligands. **Conclusions.** Using classical bioinformatics approaches we identified 6 molecular targets of *Escherichia coli* ATCC 25922, which can be exploited for further rational drug design in order to find novel therapeutic agents for the treatment of infection caused by *E.coli*.

Key words: *Escherichia coli*, antibiotic resistance, drug target, antibiotics, subtractive genomic analysis.

Introduction

Uncontrolled using of antibiotics led to antibiotic resistance. The spread of antibiotic resistance accelerated. For example, ceftazidime-avibactam, a drug composed of ceftazidime, cephalosporin antibiotic, and avibactam, a β -lactamase inhibitor, was approved for medical use in the United States and Europe in 2015. It is used to treat certain multidrug-resistant gram-negative infections. In December 2018 a ceftazidime-avibactam-resistant *Klebsiella pneumoniae* strain was isolated in Finland [1]. The microorganisms become resistant to antibiotics due to mutations in drug target proteins or the use of new alternative metabolic or signaling pathways to avoid harmful effect [2]. One way to pass up the pathogen resistance is search for the new drug targets among bacterial proteins. The methods of gene knockout and RNA interference are widely used in order to find novel targets. These methods have proven to be extremely effective, as the result can be immediately seen on living cells. Along with the advantages, these methods have a number of disadvantages, including high cost and time-consumption to set up the experiment. The bioinformatics methods are an alternative way to find novel drug targets. To date, they are successfully used in searching for new targets and are the fastest and cheapest [3].

The *Escherichia coli* infections are the major challenge for public health. Diarrheal illness causes high mortality worldwide, particularly in children under the age of 5 and particularly in countries of Africa and South Asia [4]. Even in the United States only one pathogenic strain of *E.coli* causes 73 000 illnesses annually [5].

Pathogenic *E. coli* can cause a broad range of human diseases that span from those of the gastrointestinal tract to the urinary tract, bloodstream, and central nervous system [6]. Over the last decade, there has been a significant increase in antibiotic resistance of *Escherichia coli*. In 2013 in Ile-Ife, Nigeria, over 130 isolates of *E.coli* were tested for the sensitivity to the most commonly used antibiotics. The results showed the resistance to the most commonly used antibiotics from 51.1 % to 94.3 % [7].

We used subtractive genomic analysis to identify novel molecular drug targets in *E.coli*. *Escherichia coli* ATCC 25922 strain was used as reference for pathogenic *E.coli* strains. The standard test bacteria ATCC 25922 are used in different studies of antibacterial compounds of plants [8], in food safety studies [9] and as the quality control strain in antibiotic resistance studies [10]. Unlike the pathogenic strains of *E.coli*, the ATCC 25922 strain has biosafety level 1, which makes *in vitro* testing much safer and cheaper. Also there are known MICs of many antibacterial compounds for the ATCC 25922 strain [11]. This is gram-negative bacterium like some of the most dangerous pathogens such as *Pseudomonas aeruginosa*, *Klebsiella pneumoniae* etc., and identification of potential drug targets could help to discover new ways of fighting against the infection caused by gram-negative bacteria.

Materials and Methods

The identification and characterization of the potential drug targets of *Escherichia coli* ATCC 2592 were performed sequentially by the following methods:

Retrieval of complete proteome

Whole proteome of *E. coli* ATCC 25922 was downloaded in FASTA format from the National Center for Biotechnology Information (NCBI) protein database [12]. This set was created by Los Alamos National Laboratory and contains 4808 proteins.

Paralog sequences identification

Protein set in FASTA format was submitted to CD-HIT Suite. Paralogous proteins were detected with cut-off 0.6 (all proteins similar more than by 60 % were rejected) [13].

Determination of protein sequences that non-homologous to the human proteome

At this step the non-paralogous sequences were submitted to BLASTp through NCBI server against *Homo sapiens* proteome with threshold expectation value $E\text{-value} = 10^{-3}$ [14]. Other parameters were set by default: BLOSUM62 matrix, non-redundant (nr) database. BLAST (Basic Local Alignment Tools) is an algorithm which compares query sequences of amino acids or nucleotides with each other or with database. As a result, the sequences which were similar to the human host were removed and only the non-homologous sequences were taken for further analysis.

Identification of essential proteins

The set of non-paralogous, non-homologous sequences was subjected to similarity against the Database of Essential Genes (DEG) [15], which contains all the essential genes currently available. Protein BLAST was performed with threshold expectation value $E\text{-value} = 10^{-5}$. Identification of essential proteins was done by comparing query sequences

of the *E. coli* ATCC 2592 strain with essential proteins of the *E. coli* MG1655 I and MG1655 II strains.

Prediction of protein localization

It is a common fact that potential drugs would bind easier if the targets are membrane proteins. To find proteins localization we used combination of programs PSORTp [16], CELLO [17] and ngLOC [18]. Essential non-homologous proteins were subsequently submitted to these programs in order to get more accurate result. All proteins, which were identified as membrane proteins in three programs, were taken for further analysis.

Metabolic pathways analysis

The metabolic pathways analysis of essential membrane proteins was done by KAAS server at KEGG (Kyoto Encyclopedia of Genes and Genomes) [19]. KAAS predicts the role of proteins involved in pathways by comparing the query with the KEGG GENES database in BLAST. The output contains the metabolic pathways where submitted proteins are involved along with KO list (KEGG Orthology).

Identification of already known drug targets

To identify the known molecular targets we used Drug Bank Database [20], a free online web-resource, which contains the information about drugs and drug targets. We used Drug Bank Database version 5.1.5 (released 2020-01-03) with the information about 13489 known and over 6348 experimental drugs. The set of essential proteins involved in metabolic pathways was submitted through BLASTp against Drug Bank Database ($E\text{-value} 10^{-3}$).

The known molecular targets were excluded from further analysis.

Homology modeling of proteins

We have conducted a search for the crystal structures or homology models for every novel potential drug target that was identified in current study. For that purpose, we used UniProt to figure out the existence of reference structures for every target. The search was performed in UniProt BLAST with default settings, and the results with the best similarity score were selected. The existence of novel drug target structure was checked in “Structure” section. Homology modeling in I-TASSER [21] server was performed for the drug targets that do not have crystal structure or homology model in SWISS-MODEL repository and there is no information about protein structure in UniProt. After that, the best models were validated using MolProbity web-server [22].

Binding sites analysis

3D structures of novel potential targets were investigated to find binding sites for small organic molecules and confirm the potential use of the proteins for further rational drug design. For this purpose, we used web servers ProBis [23] and PrankWeb [24]. Additionally, for more accurate search Discovery Studio 2017 was used. Only the targets, the pockets of which were identified by at least two programs were selected as potential drug targets.

Results and Discussion

The main task of current study was to identify membrane proteins as potential drug targets in *Escherichia coli* ATCC 25922. The membrane plays a crucial role for vital processes of bac-

terial cells. It provides selective permeability for cellular homeostasis and metabolic energy-transduction. Membrane proteins are responsible for the active transport of nutrients and wastes, bacterial respiration, the establishment of the proton motive force in association with respiratory enzymes, ATP generation and cell-cell communication in biofilms. These proteins are crucial and can be considered as potential drug targets.

Another reason for choosing membrane proteins as targets may be the growing danger of quiescent bacteria. These bacteria under stress of, for example, antibiotic therapy, can slow down their own metabolism, reduce growth rate, and modify the cell wall to avoid an immune response. Under such conditions, the bacteria can wait for a long time until the stress factor disappears, after which the infection will be a serious threat to the patient [25]. Because the metabolism of such bacteria slows down and the cell wall is modified, it is difficult for antibacterial compounds to influence the processes inside the bacterial cell. One way to solve the problem of quiescent bacteria may be to choose membrane proteins as potential drug targets.

In addition to the above, the compounds targeting membrane proteins can avoid cell efflux pumps. Efflux pumps are transport proteins, which are mostly localized in cytoplasmic membrane of bacterial cells, and play an important role in their lives, particularly in providing resistance of the microorganisms to xenobiotics. These pumps are involved in the removal of toxic substrates, including antibiotics, from cells into the environment. This feature of efflux pumps is one of the mechanisms of antibiotic resistance [26]. Therefore, focus-

ing on the membrane proteins as drug targets can help to evade this mechanism.

We used subtractive genomic approach to identify potential drug targets as described previously [27-29]. Therefore, we have selected the proteins that meet the requirements for good antibacterial targets in particular, to be essential for bacterial cells, have no human homologue and to be conservative among bacteria species [30].

Whole proteome of *E. coli* ATCC 25922 which has 4808 proteins was retrieved from NCBI database. At the beginning of analysis, we removed paralogs and duplicates using server CD-HIT with cut-off = 0.6. According to the results of CD-HIT server the quantity of proteins was reduced to 4544. The resulted set of non-paralogous sequences was submitted to BLASTp against human proteome. The proteome comparison between *Escherichia coli* ATCC 25922 and *Homo sapiens* was necessary in order to eliminate the targets that are structurally similar to human proteins, and consequently to avoid the possible side effects of the antibiotic designed against drug targets. All homologues of human proteins were excluded and the number of *Escherichia coli* proteins was reduced to 2150. The next step of the analysis was the identification of essential proteins. The obtained set of sequences was submitted to BLAST against the Database of Essential Gene with E-value = 10^{-5} . There is no information about the ATCC 25922 strain in DEG, so this step was done by comparing the query with the essential proteins of the closest *E. coli* strains MG1655 I and MG1655 II. The strains ATCC 25922 and MG 1655 of *E. coli* have identity 97.81 %. There are 609 essential genes for *E. coli* MG1655 I

and 296 essential genes for *E. coli* MG1655 II in DEG. The output set included 262 essential proteins.

The next step in the process of potential drug targets identification is the prediction of protein localization. In order to identify membrane proteins the set of 262 essential non-homologous proteins was analyzed using the combination of three programs — PSORTb, CELLO and ngLOC. The resulted set included 56 proteins, which were predicted by three programs as membrane proteins.

We analyzed the set of 56 sequences by KAAS (KEGG Automated Annotation Server), in order to identify which essential membrane proteins are involved in metabolic pathways. It was found that 20 proteins are involved in metabolic pathways.

The set of 20 essential membrane proteins involved in metabolic pathways was submitted to BLAST against Drug Bank Database to exclude the already known drug targets. The sequences which have similarities with Drug Bank Database were *1-deoxy-D-xylulose 5-phosphate reductoisomerase*, *ATP synthase F0 B subunit*, *cell division protein ZipA*, *penicillin-binding protein 2*, *aerotaxis receptor*, *methyl-accepting chemotaxis (MCP) signaling domain protein*, *methyl-accepting chemotaxis protein I*, *methyl-accepting chemotaxis protein II*. According to extensive PubMed search for one potential drug target – *flagellar motor stator protein MotA*, small-molecular inhibitors have been already published [31]. Additionally, we excluded *cytochrome O ubiquinol oxidase subunit IV* and *preprotein translocase, SecE subunit* because they are too small and could not have ligand binding sites. Therefore, the number of novel potential tar-

gets was reduced from 20 to 9. The results of subtractive analysis for *Escherichia coli* are presented in the Table 1.

To evaluate the drug ability of the found targets, their 3D structures should be examined for existence of small molecules binding sites. In accordance with UniProt information, three-dimensional structures are known for 7 targets. For four proteins the structures were obtained by X-ray crystallography, for two proteins - *LPS export ABC transporter permease LptF* and *LPS export ABC transporter permease LptG* the structures were obtained by electron microscopy and for one protein – *PTS system mannitol-specific EIICBA component* a homology model is available in SWISS-MODEL repository.

For two proteins *inner membrane protein ydcZ* and *putative protein permease FtsX* 3D the structures were not known. We built homology models of these proteins using I-TASSER server. For model constructing the server used LOMETS (Local Meta-Threading Server) to collect 10 protein templates of similar folds from PDB database. To estimate the quality of predicted models by I-TASSER we used C-score. The calculation is based on the significance of threading template alignments

and the convergence parameters of the structure assembly simulations. It lies in the range from -5 to 2, where the C-score of higher value signifies a model with a high confidence. I-TASSER builds 5 homology models with different C-score. For further analysis we have chosen the homology models with best C-score. These models were validated with MolProbity score [32].

MolProbity score combines clashscore, rotamer and Ramachandran evaluations into a single score. Summary table includes a percentile relative to structures near the same resolution. Good model must have more than 66th percentile. These models of putative *protein insertion permease FtsX inner* and *membrane protein ydcZ* have MolProbity score 2.17 and 3.2 respectively with 38th percentile. It means that the created models could not be used for molecular manipulation like docking or molecular modeling. Unfortunately, at present there are not enough protein templates to build correct models, but as soon as suitable templates appear, the models' quality will increase sharply. The identified possible molecular targets involved in metabolic pathways, which would be analyzed further, are presented in the Table 2.

Table 1. Results of Subtractive analysis for *E. coli* ATCC 25922

No	Stages of study	Protein quantity
1	Proteome retrieving	4808
2	Paralog sequences rejection (CD-HIT) (>60 % similarities)	4544
3	Human non-homologous proteins (BLAST)	2150
4	Essential proteins analysis (DEG)	262
5	Membrane proteins identification (PSORTb, CELLO, ngLOC)	56
6	Metabolic pathways analysis (KAAS)	20
7	Binding sites analysis	9
8	Potential drug targets	6

The small molecules binding sites were identified using Discovery Studio 2017 and web servers PrankWeb and ProBis. For six targets the ligand binding sites were recognized with at least two tools that we used. For *hydrogenase-1 small chain*, *hydrogenase-2 small chain* and *Ni/Fe-hydrogenase, b-type cytochrome subunit* the pockets were identified with all three tools. For *PTS system, Fru family, IIB component domain protein* the pockets were determined by

Discovery Studio 2017 and web server PrankWeb. For *LPS export ABC transporter permease LptG* and *LPS export ABC transporter permease LptF* the pockets were identified by Discovery Studio 2017 and web server ProBis. There is no identified suitable pocket for *PTS system mannitol-specific EIICBA component*. The result of this step confirms potential druggability of six proteins. The results of binding site analysis are shown in the Table 3.

Table 2. Possible drug targets involved in metabolic pathways

№	Protein	Metabolic pathways	KO number
1	PTS system, Fru family, IIB component domain protein	00051 Fructose and mannose metabolism	K02769
2	PTS system mannitol-specific EIICBA component		K02800
3	hydrogenase-2 small chain	00633 Nitrotoluene degradation	K06282
4	hydrogenase-1 small chain		
5	LPS export ABC transporter permease LptF	02010 ABC transporters	K07091
6	LPS export ABC transporter permease LptG		K11720
7	Ni/Fe-hydrogenase, b-type cytochrome subunit	02020 Two-component system	K03620

Table 3. Identified pockets of potential drug targets

Proteins	Tools			Residues
	PrankWeb	ProBis	DS 2017	
PTS system, Fru family, IIB component domain protein	+	-	+	Leu24, Leu27, Glu31, Trp33, Leu 61, Phe87, Pro91, Val94, Ser96
hydrogenase-2 small chain	+	+	+	Cys154, Pro155, Pro156, Asn157, Asn160, Ile191, Thr231, Cys235, Phe240, Trp247, Pro248, Pro254, Cys255, Tyr256, Cys258, Asn259
hydrogenase-1 small chain	+	+	+	Cys149, Pro150, Ile186, Asn228, Cys230, Trp235, Asn236, Phe241, Pro242, Ile243, Cys249, Leu250, Cys252
Ni/Fe-hydrogenase, b-type cytochrome subunit	+	+	+	Met29, Leu32, Met33, Gly36, Ile39, Gly40, Arg60, His63, Phe64, Gly67, Phe70, Met143, Ile144, Gly147, Phe148, Leu150, Tyr151, His181, His184, Arg185, Met188, Ile 191
LPS export ABC transporter permease LptF	-	+	+	Met282, Leu286, Leu305, Leu341, Leu345, Trp348
LPS export ABC transporter permease LptG	-	+	+	Leu73, Leu77, Phe295, Pro299, Leu300, Val303, Val307, Arg308, Thr311, Gly312, Ser314, Phe315

Conclusion

Using subtractive genome analysis, we have identified 6 membrane proteins as novel potential drug targets in *Escherichia coli* ATCC 25922 which are essential, non-homologous to human proteome and involved in unique metabolic pathways. Three-dimensional structural analysis shows that these targets have ligand binding sites hereby can be modulated with small-molecules and appropriately are suitable for further rational drug design. Presumably, the found drug targets will be able to speed up the future discovery of new therapeutic agents against the infections caused by *E. coli* and other gram-negative bacteria.

REFERENCES:

1. Räsänen K, Koivula I, Ilmavirta H, Puranen S, Kallonen T, Lyytikäinen O, Jalava J. Emergence of ceftazidime-avibactam-resistant *Klebsiella pneumoniae* during treatment, Finland, December 2018. *Euro Surveill.* 2019;**24**(19):1900256.
2. Munita JM, Arias CA. Mechanisms of Antibiotic Resistance. *Microbiol Spectr.* 2016;**4**(2):10.1128/microbiolspec.VMBF-0016–2015.
3. Hossain M, Chowdhury DU, Farhana J, Akbar MT, Chakraborty A, Islam S, Mannan A. Identification of potential targets in *Staphylococcus aureus* N315 using computer aided protein data analysis. *Bioinformatics.* 2013;**9**(4):187–92.
4. Croxen MA, Law RJ, Scholz R, Keeney KM, Włodarska M, Finlay BB. Recent advances in understanding enteric pathogenic *Escherichia coli*. *Clin Microbiol Rev.* 2013;**26**(4):822–80.
5. Rangel JM, Sparling PH, Crowe C, Griffin PM, Swerdlow DL. Epidemiology of *Escherichia coli* O157:H7 outbreaks, United States, 1982–2002. *Emerg Infect Dis.* 2005;**11**(4):603–9.
6. Croxen MA, Finlay BB. Molecular mechanisms of *Escherichia coli* pathogenicity. *Nat Rev Microbiol.* 2010;**8**(1):26–38.
7. Olorunmola FO, Kolawole DO, Lamikanra A. Antibiotic resistance and virulence properties in *Escherichia coli* strains from cases of urinary tract infections. *Afr J Infect Dis.* 2013;**7**(1):1–7.
8. He M, Wu T, Pan S, Xu X. Antimicrobial mechanism of flavonoids against *Escherichia coli* ATCC 25922 by model membrane study. *Appl Surf Sci.* 2014;**305**:515–21.
9. Sauer A, Moraru CI. Inactivation of *Escherichia coli* ATCC 25922 and *Escherichia coli* O157:H7 in apple juice and apple cider, using pulsed light treatment. *J Food Prot.* 2009;**72**(5):937–44.
10. Hein-Kristensen L, Franzyk H, Holch A, Gram L. Adaptive evolution of *Escherichia coli* to an α -peptide/ β -peptoid peptidomimetic induces stable resistance. *PLoS One.* 2013;**8**(9):e73620.
11. Fass RJ, Barnishan J. Minimal inhibitory concentrations of 34 antimicrobial agents for control strains *Escherichia coli* ATCC 25922 and *Pseudomonas aeruginosa* ATCC 27853. *Antimicrob Agents Chemother.* 1979;**16**(5):622–4.
12. Minogue TD, Daligault HA, Davenport KW, Bishop-Lilly KA, Broomall SM, Bruce DC, Chain PS, Chertkov O, Coyne SR, Freitas T, Frey KG, Gibbons HS, Jaissle J, Redden CL, Rosenzweig CN, Xu Y, Johnson SL. Complete Genome Assembly of *Escherichia coli* ATCC 25922, a Serotype O6 Reference Strain. *Genome Announc.* 2014;**2**(5):e00969–14.
13. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics.* 2010;**26**(5):680–2.
14. Kerfeld CA, Scott KM. Using BLAST to teach “E-value-tionary” concepts. *PLoS Biol.* 2011;**9**(2):e1001014.
15. Luo H, Lin Y, Gao F, Zhang CT, Zhang R. DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. *Nucleic Acids Res.* 2014;**42**(Database issue):D574–80.
16. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FS. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics.* 2010;**26**(13):1608–15.

17. Yu CS, Lin CJ, Hwang JK. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci.* 2004;**13**(5):1402–6.
18. King BR, Vural S, Pandey S, Barteau A, Guda C. ngLOC: software and web server for predicting protein subcellular localization in prokaryotes and eukaryotes. *BMC Res Notes.* 2012;**5**:351.
19. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 2007;**35**(Web Server issue):W182–5.
20. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N, Iynkkaran I, Liu Y, Maciejewski A, Gale N, Wilson A, Chin L, Cummings R, Le D, Pon A, Knox C, Wilson M. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 2018;**46**(D1):D1074–D1082.
21. Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.* 2015;**43**(W1):W174–81.
22. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, Verma V, Keedy DA, Hintze BJ, Chen VB, Jain S, Lewis SM, Arendall WB 3rd, Snoeyink J, Adams PD, Lovell SC, Richardson JS, Richardson DC. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* 2018;**27**(1):293–315.
23. Konc J, Janežič D. ProBiS-ligands: a web server for prediction of ligands by examination of protein binding sites. *Nucleic Acids Res.* 2014;**42**(Web Server issue):W215–20.
24. Jendele L, Krivak R, Skoda P, Novotny M, Hoksza D. PrankWeb: a web server for ligand binding site prediction and visualization. *Nucleic Acids Res.* 2019;**47**(W1):W345–W349.
25. Rittershaus ES, Baek SH, Sasseti CM. The normalcy of dormancy: common themes in microbial quiescence. *Cell Host Microbe.* 2013;**13**(6):643–51.
26. Blanco P, Hernando-Amado S, Reales-Calderon JA, Corona F, Lira F, Alcalde-Rico M, Bernardini A, Sanchez MB, Martinez JL. Bacterial Multidrug Efflux Pumps: Much More Than Antibiotic Resistance Determinants. *Microorganisms.* 2016;**4**(1):14.
27. Uddin R, Saeed K. Identification and characterization of potential drug targets by subtractive genome analyses of methicillin resistant *Staphylococcus aureus*. *Comput Biol Chem.* 2014;**48**:55–63.
28. Hasan MA, Khan MA, Sharmin T, Hasan Mazumder MH, Chowdhury AS. Identification of putative drug targets in Vancomycin-resistant *Staphylococcus aureus* (VRSA) using computer aided protein data analysis. *Gene.* 2016;**575**(1):132–43.
29. Singh Sarita GSK, Pant K. K GMK. *Definition of Potential Targets in Mycoplasma Pneumoniae Through Subtractive Genome Analysis.* J Antivir Antiretrovir 2010; **2**
30. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C. The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr.* 2002;**58**(Pt 6 No 1):899–907.
31. Rasmussen L, White EL, Pathak A, Ayala JC, Wang H, Wu JH, Benitez JA, Silva AJ. A high-throughput screening assay for inhibitors of bacterial motility identifies a novel inhibitor of the Na⁺-driven flagellar motor and virulence gene expression in *Vibrio cholerae*. *Antimicrob Agents Chemother.* 2011;**55**(9):4134–43.
32. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr.* 2010; **66**(Pt 1):12–21.

Пошук мембранних білків *Escherichia coli* ATCC 25922 як нових мішеней для розробки антибіотиків, з використанням методів *in silico*

І. М. Колєєв, С. А. Старосила, М. В. Протопопов, Г. П. Волинець, В. М. Сапелкін, Л. В. Плетньова, А. Р. Синюгін, Н. О. Качапут, В. І. Матюшок, В. Г. Бджола та С. М. Ярмолюк.

Мета. Метою дослідження був пошук нових мішеней для дизайну антибіотиків проти *Escherichia coli* ATCC 25922 методами розрахункової геноміки. **Методи.** Ідентифікація негомологічних білків до людського про-

теому, визначення життєво важливих генів *Escherichia coli* ATCC 25922 та встановлення новизни знайдених мішеней проводили алгоритмами BLAST. Використовуючи базу даних KEGG, ідентифікували унікальні метаболічні шляхи бактерії. Локалізацію білків у клітині передбачали програмами PSORTp, CELLO та ngLOC. Гомологічне моделювання проводили веб-сервером I-TASSER, валідація моделей була виконана з використанням серверу MolProbity. Аналіз сайтів зв'язування малих органічних молекул проведено веб-серверами ProBis і PrankWeb та використанням Discovery Studio 2017. **Результати.** Протеом *Escherichia coli* ATCC 25922, що складається із 4808 білків, було використано для формування початкової вибірки. Використовуючи методи розрахункової геноміки, вдалося ідентифікувати 9 мембранних білків, які є важливими для виживання бактерії, не є гомологами людських білків, залучені до унікальних метаболічних шляхів та раніше не досліджувались як терапевтична мішень. Дослідження просторової структури знайдених білків показало, що 6 з них мають сайти для зв'язування лігандів. **Висновки.** У результаті дослідження методами лінійної та структурної біоінформатики знайдено 6 потенційних мішеней *Escherichia coli* ATCC 25922, які можуть бути використані для подальшої розробки антибіотиків проти інфекцій спричинених *E. coli*.

Ключові слова: *Escherichia coli*, резистентність до антибіотиків, молекулярна мішень, антибіотики, розрахункова геноміка.

Поиск мембранных белков *Escherichia coli* ATCC 25922 как новых мишеней для разработки антибиотиков, с использованием методов *in silico*

И. М. Колеев, С. А. Старосила, М. В. Протопопов, Г. П. Волинец, В. М. Сапелкин, Л. В. Плетнёва, А. Р. Синюгин, Н. А. Качапут В. И. Матюшок, В. Г. Бджола, С. Н. Ярмолук.

Цель. Целью исследования был поиск новых мишеней для дизайна антибиотиков против *Escherichia coli* ATCC 25922 методами расчетной геномики.

Методы. Идентификация негомологичных белков к человеческому протеому, определение жизненно важных генов *Escherichia coli* ATCC 25922 и установления новизны найденных мишеней проводили методами BLAST. Используя базу данных KEGG, идентифицировали уникальные метаболитические пути бактерии. Локализацию белков в клетке определяли программами PSORTp, CELLO и ngLOC. Гомологические модели были построены веб-сервером I-TASSER, валидация моделей была проведена с использованием сервера MolProbity. Анализ сайтов связывания малых органических молекул сделан веб-серверами ProBis и PrankWeb с использованием Discovery Studio 2017. **Результаты.** Протеом *Escherichia coli* ATCC 25922, состоящий из 4808 белков, было использовано для формирования начальной выборки. Используя методы расчетной геномики, удалось идентифицировать 9 мембранных белков, которые важны для выживания бактерии, не являются гомологами человеческих белков, вовлеченные в уникальные метаболитические пути и раньше не исследовались как терапевтическая мишень. Исследование пространственной структуры найденных белков показало, что 6 из них имеют сайты для связывания лигандов. **Выводы.** В ходе исследования методами линейной и структурной биоинформатики найдено 6 потенциальных мишеней *Escherichia coli* ATCC 25922, которые могут быть использованы для дальнейшей разработки антибиотиков против инфекций, вызванных *E. coli*.

Ключевые слова: *Escherichia coli*, резистентность к антибиотикам, молекулярная мишень, антибиотики, расчетная геномика.

Received 14.07.2020