



## СТРУКТУРНОЕ РАЗНООБРАЗИЕ ТРИПТОФАНОВОГО ОПЕРОНА И ЛИДЕРНОГО ПЕПТИДА TrpL У ПРЕДСТАВИТЕЛЕЙ ENTEROBACTERIA

Бурыгин Г.Л.<sup>1,2</sup>, Крючкова Е.В.<sup>2</sup>

<sup>1</sup>Саратовский государственный аграрный университет им. Н.И. Вавилова,  
Театральная площадь, 1, 410012 г. Саратов, Россия.

<sup>2</sup>Институт биохимии и физиологии растений и микроорганизмов РАН,  
проспект Энтузиастов, 13, 410049 г. Саратов, Россия. [burygingl@gmail.com](mailto:burygingl@gmail.com)

### Резюме

Триптофановый оперон бактерий содержит гены, кодирующие ферменты биосинтеза триптофана из хоризмовой кислоты. Этот оперон является классическим примером функционирования процесса аттенюации транскрипции путём синтеза лидерного пептида TrpL. Тем не менее, в базе данных GenBank, включающей >140000 бактериальных геномов, обнаружено всего 215 генов *trpL*, из которых 151 описан для штаммов родов *Escherichia* и *Salmonella*. В данной работе проведено сравнение триптофановых оперонов представителей 7 семейств *Enterobacteriales*. Установлена консервативность структурной организации Trp-оперона энтеробактерий. Выявленные различия касались лишь присутствия в опероне отдельного гена *trpG*, либо включения кодируемого им белка как домена в состав бифункционального фермента TrpD. Анализ нуклеотидных последовательностей, предшествующих первому структурному гену оперона, позволил для всех исследованных штаммов (из них для 17 впервые) обнаружить гены лидерного пептида TrpL длиной от 14 до 36 аминокислотных остатков, содержащих от 1 (*Pantoea agglomerans*) до 5 (*Vibrio cholerae*) остатков триптофана. Пептиды TrpL большинства штаммов содержат 2 остатка триптофана в соседних положениях. Для отдельных родов и семейств энтеробактерий выявлены как консервативность структуры TrpL (внутри большинства родов и внутри семейства *Enterobacteriaceae*), так и высокая вариабельность (внутри родов *Vibrio* и *Pantoea*, внутри семейств *Erwiniaceae* и *Morganellaceae*). По результатам проведённой работы сделан вывод о некорректной работе программ аннотации полногеномных данных платформ RAST и NCBI Prokaryotic Genome Annotation Pipeline в отношении определения лидерных пептидов триптофановых оперонов бактерий и о необходимости внесения информации о лидерных пептидах и их генах в соответствующие базы данных.

**Ключевые слова:** энтеробактерии; триптофановый оперон; лидерный пептид; аттенюация транскрипции; аннотация последовательностей

**Цитирование:** Бурыгин Г.Л., Крючкова Е.В. Структурное разнообразие триптофанового оперона и лидерного пептида TrpL у представителей Enterobacteria // *Биомика*. 2019. Т.11(1). С. 101 – 106. DOI: 10.31301/2221-6197.bmcs.2019-07

## STRUCTURAL DIVERSITY OF TRYPTOPHAN OPERONS AND TrpL LEADER PEPTIDES IN ENTEROBACTERIA

Burygin G.L.<sup>1,2</sup>, Kryuchkova Ye.V.<sup>2</sup>

<sup>1</sup>Vavilov Saratov State Agrarian University, 1 Teatralnaya Ploshchad, Saratov 410012, Russia.

<sup>2</sup>Institute of Biochemistry and Physiology of Plants and Microorganisms, Russian Academy of Sciences,  
13 Prospekt Entuziastov, Saratov 410049, Russia. [burygingl@gmail.com](mailto:burygingl@gmail.com)

### Resume

The bacterial tryptophan operon contains genes that code for the enzymes of tryptophan biosynthesis from chorismic acid. Although this operon is a classic example of how transcription is attenuated in the synthesis of the TrpL leader peptide, GenBank has only 215 *trpL* genes, of which 151 have been described in *Escherichia* and *Salmonella*. In this work, we compared the tryptophan operons of 20 strains from 7 families of the order *Enterobacteriales*. The structure of the enterobacterial operon was found to be conserved. The differences we detected were due either to the presence of a separate *trpG* gene in the operon or to the inclusion of the protein encoded by this gene as part of the TrpD bifunctional enzyme as a domain. For all strains examined – and for the first for 17 strains – analysis of the nucleotide sequences located before the operon's first structural gene detected genes of the TrpL leader peptide. These peptides were 14 to 36 amino acids residues in length and contained between 1 (*Pantoea agglomerans*) and 5 (*Vibrio cholerae*) tryptophan residues. The TrpL peptides of most strains contained two tryptophan residues in neighboring positions. In individual enterobacterial genera and families, the TrpL structure was either conserved (within most genera and within the *Enterobacteriaceae*) or highly variable (within the genera *Vibrio* and *Pantoea* and within the families *Erwiniaceae* and *Morganellaceae*). Our results allow the conclusion that the programs available for whole genome data annotation (such as the RAST and NCBI Prokaryotic Genome Annotation Pipeline platform) may work incorrectly with respect to the detection of the leader peptides of bacterial tryptophan operons. Further, the results suggest the need for inclusion of information about leader peptides and their genes in the appropriate databases.

**Keywords:** enterobacteria; tryptophan operon; leader peptide; transcription attenuation; sequence annotation

**Citation:** Burygin G.L., Kryuchkova Ye.V. Structural diversity of tryptophan operons and TrpL leader peptides in Enterobacteria. *Biomcs.* 2019. V.11(1). P. 101 - 106. DOI: 10.31301/2221-6197.bmcs.2019-07

Особенностью организации генетического материала в бактериальных клетках является формирование оперонов – групп плотно сцепленных генов, транскрипция которых приводит к биосинтезу одной общей мРНК с несколькими сайтами инициации и терминации трансляции [Jacob, Monod, 1961]. В оперон обычно собраны гены, кодирующие ферменты одного биохимического пути метаболизма. Классическими примерами бактериальных оперонов являются лактозный и триптофановый опероны. Так у *Escherichia coli* в триптофановый оперон входят гены, кодирующие последовательно TrpL (лидерный пептид) и 5 полипептидов, обеспечивающих превращение хоризмовой кислоты в триптофан: TrpE (антранилатсинтаза), TrpD (антранилатфосфорибозилтрансфераза), TrpC (индол-3-глицеролфосфатсинтаза), TrpB ( $\beta$ -субъединица триптофансинтазы), TrpA ( $\alpha$ -субъединица триптофансинтазы) [Yanofsky et al., 1981].

Триптофановый оперон энтеробактерий на протяжении последних 40 лет является классическим примером негативной регуляции метаболитом транскрипции генов его биосинтеза [Yanofsky, 1981]. Помимо «стандартного» регулятора транскрипции (TrpR) генов оперона [Singleton et al., 1980], бактерии используют процесс аттенуации, т.е. прекращение транскрипции молекулой РНК-полимеразы данного участка генома в случае биосинтеза рибосомами лидерного пептида. Аттенуация может происходить

только в случае совмещения во времени процессов транскрипции и трансляции. Синтез лидерного пептида, ген которого расположен в начале оперона, приводит к возникновению специфических «шпилек» в молекуле ДНК внутри оперона между генами *trpL* и *trpE*, вследствие чего РНК-полимераза отделяется от ДНК с преждевременным прекращением транскрипции [Yanofsky, 1981]. В свою очередь синтез лидерного пептида TrpL может происходить только при избытке в клетке триптофана, а точнее тРНК<sup>Трп</sup>, связанных с триптофаном. Таким образом, бактериальная клетка не осуществляет транскрипцию генов ферментов биосинтеза этой аминокислоты в условиях её избытка.

При анализе в системе RAST (<http://rast.nmpdr.org/>) [Aziz et al., 2008] структурных генов триптофанового оперона штамма *Enterobacter cloacae* complex K7, выделенного из ризосферы топинамбура [Kryuchkova et al., 2014], было обнаружено, что при полной идентичности порядка расположения структурных генов *trpEDCBA* с *E. coli* K-12 в геноме штамма K7 не аннотирован ген лидерного пептида *trpL*. В базе данных GenBank, использующей при аннотации бактериальных геномов программную платформу NCBI Prokaryotic Genome Annotation Pipeline [Tatusova et al., 2016], для представителей рода *Enterobacter* (более 1250 геномов) пептид TrpL ранее не был описан. А из всех бактериальных геномов, представленных в GenBank

(более 140 000 геномов), TrpL аннотирован только в 215 вариантах, из них 164 (76%) для энтеробактерий, из которых 151 приходится на 2 рода: *Salmonella* и *Escherichia* (83 и 68, соответственно). Таким образом, целью данной работы стало изучения разнообразия *trp*-оперонов энтеробактерий и выявление присутствия в них генов лидерного пептида.

В данной работе проанализированы

нуклеотидные последовательности 20 геномов бактерий, относящихся к 12 родам 7 семейств порядка *Enterobacteriales* (табл. 1), на участке от гена, кодирующего фермент TrpH (5'-3' экзорибонуклеазы), предшествующего триптофановому оперону, до гена белка OmpW (белок W наружной мембраны), следующего после оперона.

Таблица 1.

Информация о геномном составе триптофановых оперонов штаммов, проанализированных в данной работе  
Table 1. Information on the gene composition of tryptophan operons of the strains that were analyzed in this work

Штамм Strain	Номер последовательности в GenBank GenBank sequence accession number	Гены триптофанового оперона, аннотированные в GenBank Annotated tryptophan operon genes	Длина нуклеотидной последовательности между фланкирующим геном и <i>trpE</i> , п.о. Length of the nucleotide sequences between the flanking genes and <i>trpE</i> (bp)
<i>Escherichia coli</i> K-12	AP012306	<i>trpLEDCBA</i>	273
<i>Enterobacter cloacae</i> ATCC 13047	CP001918	<i>trpEGDCBA</i>	279
<i>Enterobacter ludwigii</i> EN-119	CP017279	<i>trpEDCBA</i>	281
<i>Enterobacter hormaechei</i> ATCC 49162	MKEQ01000001	<i>trpEDCBA</i>	265
<i>Salmonella enterica</i> serovar Typhimurium ATCC 13311	CP009102	<i>trpEDCBA</i>	260
<i>Hafnia alvei</i> ATCC 13337	JMPK01000051	<i>trpEGDCBA</i>	452
<i>Hafnia paralvei</i> ATCC 29927	LXET01000031	<i>trpEGDCBA</i>	452
<i>Pectobacterium carotovorum</i> NCPPB 312	JQHJ01000004	<i>trpEGDCBA</i>	356
<i>Pectobacterium atrosepticum</i> SCRI1043	BX950851	<i>trpEGDCBA</i>	303
<i>Pantoea agglomerans</i> DSM3493	FYAZ01000004	<i>trpEGDCBA</i>	278
<i>Pantoea ananatis</i> LMG 2665	JFZU01000009	<i>trpEGDCBA</i>	282
<i>Erwinia amylovora</i> CFBP 1232	CAPB01000020	<i>trpEGDCBA</i>	275
<i>Photobacterium luminescens</i> ATCC 29999	FMWJ01000001	<i>trpEGDCBA</i>	861
<i>Proteus mirabilis</i> ATCC 29906	ACLE01000027	<i>trpEGDCBA</i>	318
<i>Yersinia pestis</i> NCTC5923	UAVH01000007	<i>trpLEGDCBA</i>	451
<i>Yersinia pseudotuberculosis</i> ATCC 6904	CP008943	<i>trpEGDCBA</i>	577
<i>Yersinia enterocolitica</i> ATCC 9610	JPDV01000006	<i>trpEGDCBA</i>	704
<i>Serratia marcescens</i> ATCC 13880	JMPQ01000044	<i>trpEGDCBA</i>	383
<i>Vibrio cholerae</i> ATCC 14035	JHXR01000004	<i>trpLEGDCBA</i>	335

Нуклеотидные последовательности между генами *trpH* и *trpE* (первый структурный ген оперона) были исследованы на присутствие в них открытых рамок считывания, кодирующих лидерный пептид *trpL*, с использованием ресурса ExPASy Translate (<https://web.expasy.org/translate/>). За TrpL принимался пептид, состоящий из менее 40 аминокислотных остатков, содержащий два остатка триптофана в

соседних позициях.

Структурные гены триптофанового оперона (*trpEDCBA*) были обнаружены между фланкирующими генами *trpH* и *ompW* во всех исследованных геномах. Лишь в геноме *Salmonella typhimurium* ATCC 13311 не был выявлен ген экзорибонуклеазы *trpH*, фланкирующий оперон с 5'-конца. У этого штамма гену *trpE* в аннотации генома

предшествовал ген треонилкарбомил-АМФ синтазы, который в остальных анализированных геномах располагался перед *trpH*. У большинства исследованных штаммов между *trp*-опероном и геном *ompW* аннотированы дополнительные гены гипотетических белков или белков с неясной ферментативной активностью.

Во всех рассмотренных в данной работе вариантах гены внутри оперона имели одинаковое направление транскрипции от *trpE* к *trpA*. Проанализированные триптофановые опероны можно разделить на 2 группы, в зависимости от присутствия или отсутствия в них гена *trpG* (табл. 1), кодирующего фермент TrpG (амидотрансферазный компонент антранилатсинтазы). В большинстве случаев (17 геномов) ген *trpG* присутствовал и располагался между генами *trpE* и *trpD*. Причём, во всех этих геномах (кроме *Yersinia enterocolitica* ATCC 9610) ген *trpG* имел перекрытия последовательностей с *trpE* (большинство штаммов) или с *trpD* (*Enterobacter cloacae* ATCC 13047), либо сразу с обоими соседними генами (*Erwinia amylovora* CFBP 1232). Лишь в 5 геномах (*Escherichia coli* K-12, *Enterobacter ludwigii* EN-119, *Ent. cloacae* complex K7,

*Enterobacter hormaechei* ATCC 49162 и *S. typhimurium* ATCC 13311) ген *trpG* отсутствовал. В клетках этих штаммов перенос аминогруппы с глутамина на хоризмат катализирует N-концевой домен бифункционального фермента TrpD, как это описано для *E. coli* [Yanofsky et al., 1981].

Длина нуклеотидных последовательностей от предшествующего оперону гена до начала гена *trpE* варьировала от 260 (*S. typhimurium* ATCC 13311) до 861 п.о. (штамм *Photorhabdus luminescens* ATCC 29999). В геноме штамма *Ent. cloacae* complex K7 длина этой последовательности составила 280 п.о. Из проанализированных оперонов лишь для трёх штаммов (*E. coli* K-12, *Yersinia pestis* NCTC5923 и *Vibrio cholerae* ATCC 14035) на этих участках геномов были аннотированы гены *trpL* (номера в GenBank: BAL38329, WP\_002227930 и WP\_001880954, соответственно). Перевод нуклеотидных последовательностей межгенных участков в аминокислотные последовательности с использованием ресурса ExPASy также позволил выявить во всех остальных исследованных нами геномах штаммов порядка *Enterobacteriales* гены пептидов TrpL (рис. 1).

<i>E. coli</i> K-12*	MKAI FVLK GWWRTS
<i>Ent. cloacae</i> ATCC13047	MTAHFTLHGWWRTS
<i>Ent. ludwigii</i> EN-119	MTAHFTLHGWWRTS
<i>Ent. cloacae</i> complex K7	MTAHFTLHGWWRTS
<i>Ent. hormaechei</i> ATCC49162	MTTHFALHGWWRTS
<i>S. typhimurium</i> ATCC13311	MAATFALHGWWRTS
<i>Ser. marcescens</i> ATCC13880	MNTYISLHGWWRTSLLRAV
<i>Y. pestis</i> NCTC 5923*	MKTSLISLLRWWHISLSRAM
<i>Y. pseudotuberculosis</i> ATCC6904	MKTSLISLLRWWHISLSRAM
<i>Y. enterocolitica</i> ATCC9610	MKTSLISLLRWWHISLFRVAV
<i>Erw. amylovora</i> CFBP1232	MAYLFNSITGWWRISP
<i>Pan. ananatis</i> LMG2665	MKTVITLLMGWWRNVSYGRRRCRAHLA
<i>Pan. agglomerans</i> DSM3493	MAVVITHLNGWRDVPENGRQGRAHFA
<i>H. alvei</i> ATCC13337	MTLSARWWRDSFGWAE
<i>H. paralvei</i> ATCC29927	MLSARWWRDSFWSAV
<i>Pect. atrosepticum</i> SCRI1043	MKHGWWRSSLSRV
<i>Pect. carotovorum</i> NCPPB312	MKHGWWRSSLSRV
<i>Ph. luminescens</i> ATCC29999	MTSIRWWHNFVRAVRSRT
<i>P. mirabilis</i> ATCC29906	MNINQRLIALWWWKTLWAV
<i>V. cholerae</i> ATCC14035*	MLQEFNPNHKPNFSPADAEALWWWRTWTSSWWAHVYF

Рис. 1. Аминокислотные последовательности лидерных пептидов TrpL, выявленные при анализе геномов штаммов представителей разных семейств *Enterobacteriales*. \* – лидерные пептиды, аннотированные ранее в GenBank (*E. coli* K-12 – BAL38329; *Yersinia pestis* NCTC5923 – WP\_002227930; *Vibrio cholerae* ATCC 14035 – WP\_001880954). Серым фоном отмечены остатки триптофана.

Fig. 1. Amino acid sequences of TrpL leader peptides from genomes of strains representing different families of *Enterobacteriales*. \* – leader peptides earlier annotated in GenBank (*E. coli* K-12 – BAL38329; *Yersinia pestis* NCTC5923 – WP\_002227930; *Vibrio cholerae* ATCC 14035 – WP\_001880954). Tryptophan residues are marked in gray.

Сравнение аминокислотных последовательностей лидерных пептидов TrpL двадцати штаммов энтеробактерий выявил существенные различия в их составе. Длина пептидов варьировала от 14 (*Enterobacteriaceae* и *Pectobacteriaceae*) до 36 (*V. cholerae* ATCC 14035) аминокислотных остатков. Количество остатков триптофана в лидерных пептидах также различалось, составляя один (*Pantoea agglomerans* DSM3493), два (большинство штаммов), три (*Proteus mirabilis* ATCC 29906, *Hafnia alvei* ATCC 13337 и *Hafnia paralvei* ATCC 29927) и пять (*V. cholerae* ATCC 14035).

Остатки триптофана в TrpL находятся ближе к N-концу (*Hafniaceae*, *Pectobacteriaceae* и *Ph. luminescens* ATCC 29999), в середине пептида (*Yersiniaceae*, *P. mirabilis* ATCC 29906 и 2 штамма рода *Pantoea*) и ближе к C-концу молекулы (*Enterobacteriaceae*, *Erwinia amylovora* CFBP 1232 и *V. cholerae* ATCC 14035). Исходя из функциональной значимости остатков триптофана в пептиде TrpL, можно предположить, что аттенуация выполняется наиболее строго в клетках *Pan. agglomerans* DSM3493, так как пептид синтезируется при включении единичного остатка триптофана. Для штаммов, у которых остатки триптофана расположены ближе к началу пептида, аттенуация должна быть более эффективна, так как при избытке триптофана в клетке рибосома проходит этот участок быстрее без синтеза более длинного пептида.

При сравнении пептидов TrpL штаммов внутри родов и семейств было выявлено 3 группы пептидов с идентичными структурами. Так, абсолютно идентичны аминокислотные последовательности пептидов у двух исследованных представителей рода *Pectobacterium*, при 100% идентичности и нуклеотидных последовательностей *trpL*. Также идентичны аминокислотные и нуклеотидные последовательности TrpL у штаммов *Y. pestis* NCTC 5923 и *Yersinia pseudotuberculosis* ATCC 6904. У TrpL штамма *Y. enterocolitica* ATCC 9610 имеется две аминокислотные замены относительно последовательности пептида других видов рода *Yersinia*, а TrpL штамма *Serratia marcescens* ATCC 13880 (также представитель *Yersiniaceae*) имеет ещё три дополнительных аминокислотных замены и делецию при сохранении общего сходства пептидов внутри семейства.

У штаммов *Ent. cloacae* ATCC 13047, *Ent. ludwigii* EN-119 и *Ent. cloacae* complex K7 при идентичности аминокислотных последовательностей имеются точечные различия в нуклеотидных. Так, в 6-ой и 39-ой позициях гена *trpL* у штамма *Ent. cloacae* ATCC 13047 находятся остатки тимидина вместо аденозина у двух других штаммов. При этом, TrpL *Ent. hormaechei* ATCC 49162 отличается от TrpL других штаммов рода *Enterobacter* как по

аминокислотной, так и по нуклеотидной последовательностям. Установлено, что у всех представителей семейства *Enterobacteriaceae* в TrpL имеется консервативный C-концевой участок (GWWRTS) и вариабельный N-концевой участок, в котором в позициях 5 и 7 консервативно расположены остатки фенилаланина и лейцина, соответственно.

Пептиды TrpL штаммов рода *Hafnia* (*Hafniaceae*) имеют высокое сходство между собой, различаясь отсутствием остатка треонина во 2-ом положении и заменой остатка глицина на остаток серина в 13-ом положении у *H. paralvei* ATCC 29927, в отличие от *H. alvei* ATCC 13337. Представители рода *Pantoea* (семейство *Erwiniaceae*) при сходстве своих TrpL имеют всего 16 идентичных из 26 аминокислотных остатков пептида. Другой представитель этого семейства *Erw. amylovora* CFBP 1232 имеет TrpL существенно отличающийся от *Pantoea* spp. Также у штаммов семейства *Morganellaceae* (*Ph. luminescens* ATCC 29999 и *P. mirabilis* ATCC 29906) TrpL между собой имеют большие различия, чем сходств. И совсем особое строение имеет TrpL штамм *V. cholerae* ATCC 14035, состоящий из значительно более длинной аминокислотной цепочки, содержащей сразу 5 остатков триптофана. Ранее был описан TrpL для представителя вида *Vibrio parahaemolyticus* [De Troch et al., 1997], также имеющий в своём составе 5 остатков триптофана, сходные с TrpL *V. cholerae* N- и C-концевые участки молекул, но состоящий из 41 аминокислотного остатка. Примечательно, что TrpL бактерий рода *Vibrio* содержат более 35 аминокислотных остатков и успешно аннотируются программой NCBI Prokaryotic Genome Annotation Pipeline при анализе полногеномных данных, в отличие от более коротких *trpL* других энтеробактерий.

Таким образом, несмотря на малое количество аннотированных лидерных пептидов TrpL в базах данных, гены этих пептидов легко могут быть найдены в геномах бактерий перед структурными генами биосинтеза триптофана. Программы аннотации геномов (в частности, платформы RAST и NCBI Prokaryotic Genome Annotation Pipeline) некорректно выявляют короткие кодирующие участки, такие как лидерные пептиды, состоящие из менее 30 аминокислотных остатков. В связи с этим представляется необходимой оптимизация программ аннотации геномов и внесение последовательностей лидерных пептидов бактерий и их генов в соответствующие базы данных. Это, в частности, позволит оценивать уровень экспрессии генов, кодирующих лидерные пептиды, при проведении высокопроизводительного транскриптомного профилирования.

**Литература / References**

1. Aziz R.K., Bartels D., Best A.A., DeJongh M., Disz T., Edwards R.A., Formsma K., Gerdes S., Glass E.M., Kubal M., Meyer F., Olsen G.J., Olson R., Osterman A.L., Overbeek R.A., McNeil L.K., Paarmann D., Paczian T., Parrello B., Pusch G.D. Reich C., Stevens R., Vassieva O., Vonstein V., Wilke A., Zagnitko O. The RAST server: rapid annotations using subsystems technology. *BMC Genomics*. 2008. V. 9. P. 75. doi: 10.1186/1471-2164-9-75
2. De Troch P., Dosselaere F., Keijers V., de Wilde P., Vanderleyden, J. Isolation and characterization of the *Azospirillum brasilense* trpE(G) gene, encoding anthranilate synthase. *Curr. Microbiol.* 1997. V. 34(1). P. 27-32. doi: 10.1007/s002849900139
3. Jacob F., Monod J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* 1961. V. 3(3). P. 318-356. doi: 10.1016/S0022-2836(61)80072-7
4. Kryuchkova Y.V., Burygin G.L., Gogoleva N.E., Gogolev Y.V., Chernyshova M.P., Makarov O.E., Fedorov E.E., Turkovskaya O.V. Isolation and characterization of glyphosate-degrading rhizosphere strain, *Enterobacter cloacae* K7. *Microbiol. Res.* 2014. V. 169(1). P. 99-105. doi: 10.1016/j.micres.2013.03.002
5. Singleton C.K., Roeder W.D., Bogosian G., Somerville R.L., Weith, H.L. DNA sequence of the *E. coli* trpR gene and prediction of the amino acid sequence of Trp repressor. *Nucleic Acids Res.* 1980. V. 8(7). P. 1551-1560. doi: 10.1093/nar/8.7.1551
6. Tatusova T., DiCuccio M., Badretdin A., Chetvernin V., Nawrocki E.P., Zaslavsky L., Lomsadze A., Pruitt K.D., Borodovsky M., Ostell J. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 2016. V. 44(14). P. 6614-6624. doi: 10.1093/nar/gkw569
7. Yanofsky C. Attenuation in the control of expression of bacterial operons. *Nature.* 1981. V. 289(5800). P. 751-758. doi: 10.1038/289751a0.
8. Yanofsky C., Platt T., Crawford I.P., Nichols B.P., Christie G.E., Horowitz H., VanCleemput M., Wu A.M. The complete nucleotide sequence of the tryptophan operon of *Escherichia coli*. *Nucleic Acids Res.* 1981. V. 9(24). P. 6647-6668. doi: 10.1093/nar/9.24.6647