

УДК 577.1

## КОМПЬЮТЕРНЫЙ АНАЛИЗ РЕГУЛЯЦИИ МЕТАБОЛИЗМА ГЛИЦЕРОЛ-3-ФОСФАТА В ГЕНОМАХ ПРОТЕОБАКТЕРИЙ

© 2003 г. Л. В. Данилова\*, М. С. Гельфанд<sup>1</sup>, В. А. Любецкий, О. Н. Лайкова<sup>1</sup>

Институт проблем передачи информации, Москва, 101447

<sup>1</sup> Федеральное государственное унитарное предприятие "ГосНИИгенетика", Москва, 113545

Поступила в редакцию 01.04.2003 г.

Проведен сравнительный компьютерный анализ потенциальных GlpR-регулонов, отвечающих за метаболизм глицерола и глицерол-3-фосфата (ГЗФ), в геномах  $\alpha$ -,  $\beta$ - и  $\gamma$ -протеобактерий. Идентифицированы новые потенциальные палиндромные сигналы связывания GlpR в  $\gamma$ -протеобактериях с консенсусом TGTTCGATAACGAACA для семейства Enterobacteriaceae, wTTTTCGTATACGAAAAA для семейства Pseudomonadaceae, AATGCTCGATCGAGCATT для семейства Vibrionaceae, а также сигналы в  $\alpha$ - и  $\beta$ -протеобактериях, состоящие из 3–4 прямых повторов слова TTTTCGTT через 3–4 пары нуклеотидов.

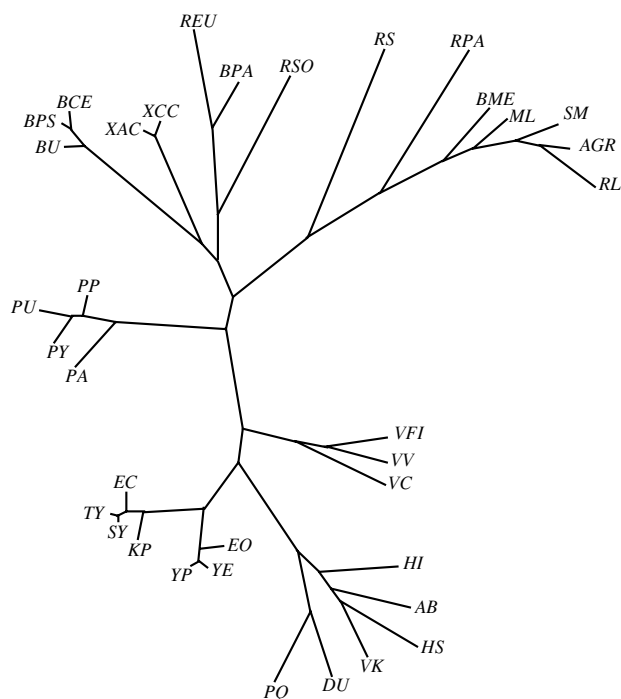
**Ключевые слова:** GlpR, тандемные повторы, компьютерный анализ, оперонная структура,  $\alpha$ -,  $\beta$ -,  $\gamma$ -протеобактерии.

GlpR, принадлежащий к семейству регуляторов DeoR, контролирует экспрессию генов метаболизма глицерола и глицерол-3-фосфата (ГЗФ). Хорошо изучен GlpR-регулон в *Escherichia coli* [1–3] и частично – в *Pseudomonas aeruginosa* [4].

Глицерол поступает извне в цитоплазму путем облегченной диффузии, обеспечиваемой продуктом гена *glpF*, а ГЗФ активно транспортируется продуктом гена *glpT*. Внутриклеточный глицерол фосфорилируется глицеролкиназой (*glpK*), давая ГЗФ. ГЗФ затем может быть превращен в дигидроксиацетонфосфат под действием одной из двух имеющихся у *E. coli* ГЗФ-дегидрогеназ – аэробной (*glpD*) или анаэробной (*glpA*). Кроме того, к GlpR регулону *E. coli* относится ген *glpQ*, кодирующий периплазматическую глицерофосфодиэстеразу, гидролизующую глицерофосфодиэфиры с высвобождением ГЗФ, гены *glpB* и *glpC*, кодирующие дополнительные структурные компоненты анаэробной ГЗФ-дегидрогеназы, а также гены *glpE*, *glpG* и *glpX*, функции которых неясны. Названные гены собраны в три локуса на хромосоме *E. coli*: *glpTQ/glpABC*, *glpEGR/glpD* и *glpFKX* (косая линия разделяет опероны, ориентированные в разные стороны). GlpR имеет наибольшее сродство к регуляторной области *glpD*. Истинным индуктором GlpR-регулона является ГЗФ.

Целью данной работы было определение состава GlpR-регулонов и сигналов связывания GlpR в геномах протеобактерий.

\* Эл. почта: dlv2k@mail.ru



**Рис. 1.** Филогенетическое дерево гомологов репрессора GlpR в  $\alpha$ -,  $\beta$ -,  $\gamma$ -протеобактериях. Сокращения: EC – *E. coli*, TY – *S. typhi*, SY – *S. typhimurium*, KP – *K. pneumoniae*, YP – *Y. pestis*, YE – *Y. enterocolitica*, EO – *E. carotovora*, VC – *V. cholerae*, VV – *V. vulnificus*, VFI – *V. fischeri*, HI – *H. influenzae*, DU – *H. ducrey*, HS – *Haemophilus somnus*, VK – *P. multocida*, PQ – *P. haemolytica*, AB – *A. actinomycetemcomitans*, PA – *P. aeruginosa*, PP – *P. putida*, PU – *P. fluorescens*, PY – *P. syringae*, BU – *B. fungorum*, BPS – *B. pseudomallei*, BCE – *B. cepacia*, XAC – *Xanthomonas axonopodis*, XCC – *Xanthomonas campestris*, BPA – *B. paraptusis*, REU – *R. eutropha*, RSO – *R. solanacearum*, RL – *R. leguminosarum*, AGR – *A. tumefaciens*, SM – *S. meliloti*, ML – *M. loti*, BME – *B. melitensis*, RPA – *R. palustris*, RS – *R. sphaeroides*.

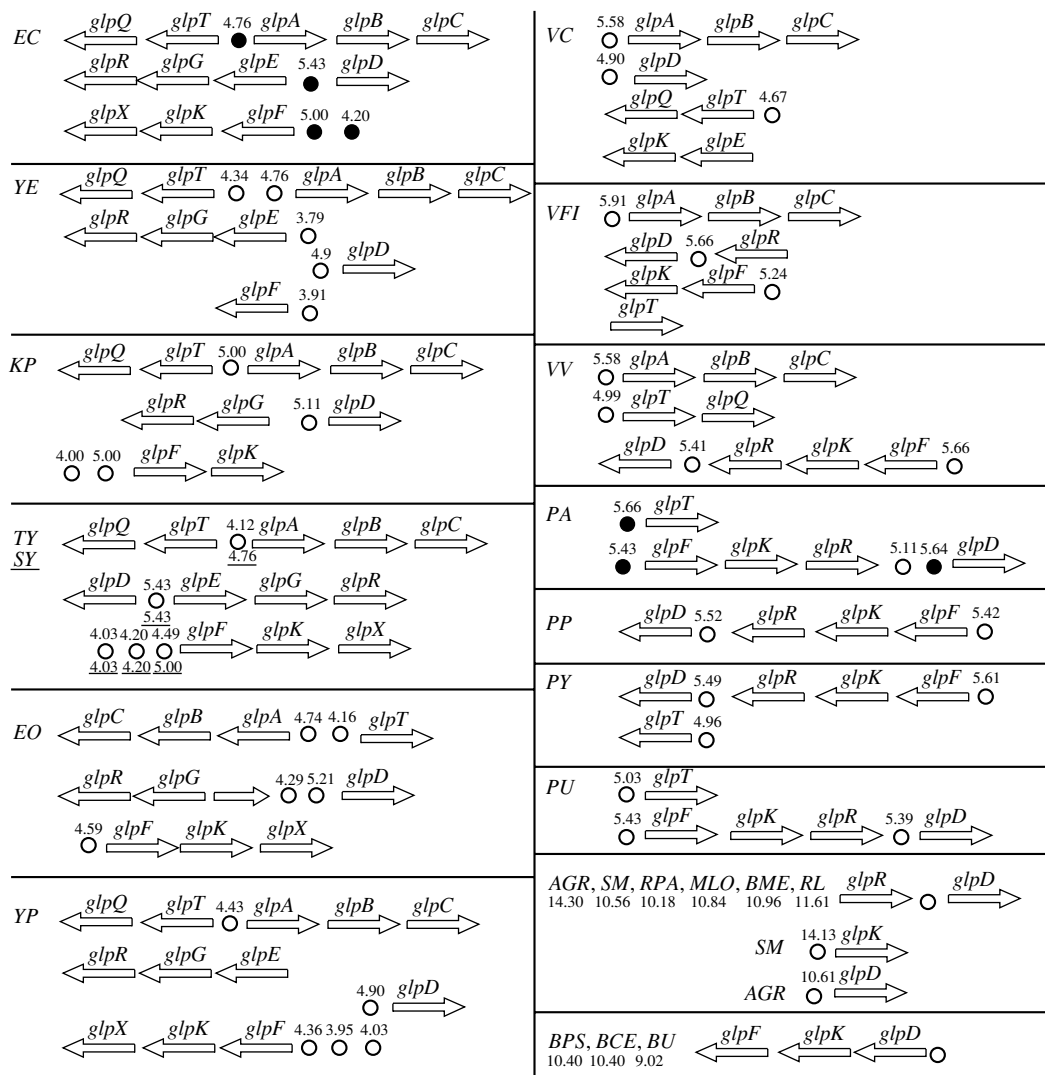


Рис. 2. Оперонная структура GlpR-регулонов в α-, β-, γ-протеобактериях. Закрашенными кружочками отмечены известные сайты, а незакрашенными – предсказанные, с их весами. Сокращения как на рис. 1.

## МАТЕРИАЛЫ И МЕТОДЫ

**Исследуемые геномы:** *Escherichia coli* [5], *Salmonella typhi* [6], *S. typhimurium* [7], *Klebsiella pneumoniae* [8], *Erwinia carotovora*, *Yersinia pestis* [9], *Y. enterocolitica* [9], *Vibrio cholerae* [10], *V. vulnificus* [9], *V. fischeri*, *Pasteurella multocida* [11], *P. haemolytica*, *Haemophilus influenzae* [12], *H. ducrey*, *Actinobacillus actinomycetemcomitans* [16], *Pseudomonas aeruginosa* [13], *P. fluorescens* [14], *P. putida* [9], *P. syringae* [15], *Burkholderia fungorum*, *B. pseudomallei*, *B. cepacia*, *Bordetella parapertussis*, *Ralstonia eutropha*, *R. solanacearum* [9], *Mesorhizobium loti* [9], *Sinorhizobium meliloti* [9], *Rhizobium leguminosarum*, *Agrobacterium tumefaciens* [9], *Brucella melitensis* [9], *Rhodopseudomonas palustris*. Близкие гомологи GlpR обнаружены во многих геномах, построено дерево (рис. 1), и все эти геномы исследованы.

Выравнивание последовательностей белков проводили при помощи программы ClustalW [17], а построение дерева генов – при помощи программы proml пакета Phylip [18].

Обозначения генов, данные в статье, соответствуют их ортологам в *E. coli*.

Для поиска сайтов и построения обучающей выборки и матриц позиционных весов использовали программы GenomeExplorer [19], SignalX [19] и IRSA [20].

Матрицу позиционных весов определяли по формуле:

$$W(b, k) = \sum_{i=A, T, C, G} 0.25 \log[(N(b, k) + 0.5)/(N(i, k) + 0.5)],$$

где  $N(b, k)$  – количество появлений нуклеотида  $b$  в позиции  $k$ . Вес потенциального сигнала определяется как сумма позиционных весов составляющих его нуклеотидов. Основание логарифма выбирается таким образом, что веса случайных олигонуклеотидов распределены по нормальному закону с нулевым средним и дисперсией, равной единице.

Все матрицы позиционных весов применялись к геномам для поиска новых сайтов в областях –400...+50 п.н. относительно начала гена.

В тех случаях, когда оперонная структура данного фрагмента ДНК неизвестна, гены относили к одному потенциальному оперону, если они имели одинаковое направления считывания, а расстояние между ними не превышало 100 п.н.. За начало такого потенциального оперона принимали ген, перед которым найден GlpR-сайт (см. рис. 2).

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Во всех исследованных геномах были определены ортологи генов GlpR-регулона. Области перед этими генами составили обучающую выборку, к которой были применены программы выделения сигнала. Основные таксономические группы, соответствующие ветвям дерева белков GlpR, рассмотрены отдельно.

### $\gamma$ -Протеобактерии, семейство Vibrionaceae.

В этом семействе из трех геномов: *V. cholerae*, *V. vulnificus*, *V. fischeri* – удалось выделить палиндромный сигнал длиной 18 п.н. с консенсусом ААТГСТСГАТСГАГСАТТ (рис. 3б). Найденные сайты и матрица позиционных весов представлены в табл. 1б и табл. 2б соответственно. При сканировании геномов с использованием этой матрицы новых потенциальных сайтов обнаружено не было.

**$\gamma$ -Протеобактерии, семейство Pseudomonadaceae.** Далее рассмотрены четыре генома из семейства Pseudomonadaceae и в регуляторных областях найден палиндромный сигнал wTTTTCG-TATACGAAAaw длиной 18 п.н. (рис. 3в), включающий ранее предсказанные сайты у *P. aeruginosa* [4] (табл. 1в); построена позиционная матрица (табл. 2в). С помощью этой матрицы с порогом 4.9 были найдены новые потенциальные сайты связывания GlpR, указанные в табл. 1в.

**$\alpha$ -,  $\beta$ -Протеобактерии.** При исследовании геномов  $\alpha$ -протеобактерий найдены 3–4 tandemных повтора ТТТСГТТ (рис. 3з), идущих друг за другом через 3–4 нуклеотида (табл. 1з), и построена позиционная матрица (табл. 2з). При дальнейшем исследовании в *A. tumefaciens* обнаружено два ортолога гена *glpD*, которые имеют в регуляторных областях tandemные повторы. Аналогичные повторы обнаружены в геномах бактерий рода

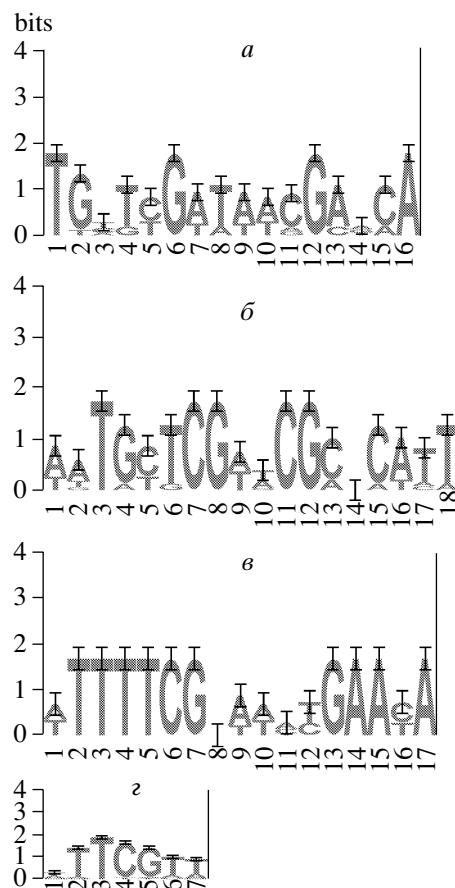


Рис. 3. Матрицы позиционных весов сигнала GlpR представлены в графическом виде, построенном с помощью описанной процедуры [23], для организмов: а – *E. coli*, *E. carotovora*, *Y. enterocolitica*, *K. pneumoniae*; б – *V. cholerae*, *V. vulnificus*, *V. fischeri*; в – *P. aeruginosa*, *P. syringae*, *P. fluorescens*, *P. putida*; з – *B. melitensis*, *M. loti*, *A. tumefaciens*, *S. meliloti*, *R. palustris*.

*Burkholderia* из группы  $\beta$ -протеобактерии (см. табл. 1з).

### $\gamma$ -протеобактерии, семейство Enterobacteriaceae.

Сначала рассматривали регуляторные области из четырех геномов – *E. coli*, *E. carotovora*, *Y. enterocolitica*, *K. pneumoniae* – и по ним составлена обучающая выборка, включающая уже известные сайты *E. coli* (табл. 1а) с консенсусом TGTTCGATAACGAACA (рис. 3а). По этой выборке построена матрица позиционных весов (табл. 2а) для поиска палиндромных сайтов длиной 16 п.н.. С помощью этой матрицы исследованы геномы *Y. pestis*, *S. typhimurium*, *S. typhi* с порогом 4.1. Найденные сайты представлены в табл. 1а.

Снижение порога при исследовании геномов из семейства Enterobacteriaceae приводит к сильному “перепредсказанию”, но при этом обнаруживаются относительно слабые сайты перед генами, входящими в ГЗФ регулон (табл. 1з). В то же время даже при низком пороге не во всех об-

**Таблица 1.** Сайты перед генами, входящими в ГЗФ-регулон в геномах указанных организмов

Геном	Ген	Вес сайта	Сайт
а – семейство Enterobacteriaceae			
<i>E. coli</i>	<i>glpD</i>	5.41	<b>TGTTTCGATAaCGAACA</b>
»	<i>glpF</i>	4.99	<b>TGcTCGtTAaCGAtaA</b>
»	<i>glpT</i>	4.76	<b>TGTTtGATtTCGcgCA</b>
<i>E. carotovora</i>	<i>glpD</i>	5.20	<b>TGcTCGAaAaCGAACA</b>
»	<i>glpT</i>	4.72	<b>TGTTtGATAaAGAgCA</b>
»	<i>glpF</i>	4.59	<b>TtcTCGtTtTCGctCA</b>
<i>K. pneumoniae</i>	<i>glpD</i>	5.10	<b>TGAgCGATATCGAgCA</b>
»	<i>glpT</i>	5.00	<b>TGTTtGATtTCGAgCA</b>
»	<i>glpF</i>	4.99	<b>TGcTCGtTAaCGAtaA</b>
<i>Y. enterocolitica</i>	<i>glpD</i>	4.89	<b>TGAgCGAaAaCGAACA</b>
»	<i>glpT</i>	4.74	<b>cGcTCGtTATgGAACA</b>
<i>E. coli</i>	<i>glpF</i>	4.20	<b>gGcgCGATAaCGctCA</b>
<i>E. carotovora</i>	<i>glpD</i>	4.29	<b>TGTTtGtTtTCGAttA</b>
»	<i>glpA</i>	4.16	<b>TGTTTcATtaCGAACg</b>
<i>S. typhi</i>	<i>glpD</i>	5.43	<b>TGTTTCGATAaCGAACA</b>
»	<i>glpF</i>	4.49	<b>TGcTCGtTAgCGAtaA</b>
»	<i>glpF</i>	4.20	<b>gGcgCGATAaCGctCA</b>
»	<i>glpT</i>	4.12	<b>TGTTtGATtTCGcgCg</b>
<i>S. typhimurium</i>	<i>glpD</i>	5.43	<b>TGTTTCGATAaCGAACA</b>
»	<i>glpF</i>	5.00	<b>TGcTCGtTAaCGAtaA</b>
»	<i>glpT</i>	4.76	<b>TGTTtGATtTCGcgCA</b>
»	<i>glpF</i>	4.20	<b>gGcgCGATAaCGctCA</b>
<i>Y. enterocolitica</i>	<i>glpA</i>	4.34	<b>TGTTTcATAaCGAgCg</b>
<i>Y. pestis</i>	<i>glpD</i>	4.90	<b>TGTTTCGtTtTCGctCA</b>
»	<i>glpA</i>	4.43	<b>TGTTtctTATCaAtCA</b>
»	<i>glpF</i>	4.36	<b>cGcTCGtTAaCGAtaA</b>
б – семейство Vibrionaceae			
<i>V. cholerae</i>	<i>glpA</i>	5.57	<b>AATGCTCGtTCGcGCtTT</b>
»	<i>glpD</i>	4.92	<b>AATaTTCGAgCGctCATT</b>
»	<i>glpT</i>	4.56	<b>AtTGCTCGtTCGcctTT</b>
<i>V. fischeri</i>	<i>glpA</i>	5.91	<b>AATGCgCGAaCGAGCATT</b>
»	<i>glpD</i>	5.66	<b>AATGtTCGtTCGctCATT</b>
»	<i>glpF</i>	5.24	<b>tgTGCTCGAaCGctCATT</b>
<i>V. vulnificus</i>	<i>glpF</i>	5.69	<b>tATGCTCGAaCGcGCATT</b>
»	<i>glpA</i>	5.66	<b>AATGtTCGAaCGctCATT</b>
»	<i>glpD</i>	5.36	<b>AATGCTCGtTCGAaCAaa</b>
»	<i>glpT</i>	5.02	<b>ttTGCTCGtTCGcaCAcT</b>
в – семейство Pseudomonadaceae			
<i>P. aeruginosa</i>	<i>glpD</i>	5.64	<b>ATTTTCGAtTtCGAAcAA</b>
»	<i>glpF</i>	5.43	<b>TTTTTCGAaAcTGAAcAA</b>
<i>P. fluorescens</i>	<i>glpF</i>	5.43	<b>TTTTTCGAATctGAAtAA</b>
»	<i>glpD</i>	5.39	<b>ATTTTCGcAaAtGAAcAT</b>
<i>P. putida</i>	<i>glpD</i>	5.52	<b>ATTTTCGcAaACGAAcAT</b>

Таблица 1. Окончание

Геном	Ген	Вес сайта	Сайт
»	<i>glpF</i>	5.42	<b>TTTTTCGTt Tc t GAAt AA</b>
<i>P. syringae</i>	<i>glpF</i>	5.61	<b>TTTTTCGTt TACGAAt AT</b>
»	<i>glpD</i>	5.49	<b>ATTTTCGg Aa At GAAc AT</b>
<i>P. aeruginosa</i>	<i>glpT</i>	5.66	TTTTTCa Tt TACGAAAA
»	<i>glpD</i>	5.11	ATg TTCGTt Tc a GAAAA
<i>P. fluorescens</i>	<i>glpT</i>	5.03	ATTTTCGg t a ACGAAAc T
<i>P. syringae</i>	<i>glpT</i>	4.96	TTTTTC t g t a At GAAAAT
г – α-, β-протеобактерии			
<i>A. tumefaciens</i>	<i>glpD</i>	14.30	<b>g TTCGTTt t a t TTTCt TT t gac a TTCGTT t tgt TTTCGc T</b>
»	<i>glpD</i>	10.61	<b>TTTCGTTt t g a c a TTCGTTt t g t CTTCGAA</b>
<i>B. melitensis</i>	<i>glpD</i>	10.96	<b>TTTCGTTt t g a t TTTCa TTt gc TTTCGTa</b>
<i>M. loti</i>	<i>glpD</i>	10.84	<b>TTTCGTTt t g a c a TTCGTTat g a g TTCGa a</b>
<i>R. leguminosarum</i>	<i>glpD</i>	11.61	<b>a TTCGTTt t g a c a TTCGTa t t c c TTTCGTT</b>
<i>R. palustris</i>	<i>glpD</i>	10.18	<b>TTTCGTTt t t g g TTt GTg c t t a TTCGTT</b>
<i>S. meliloti</i>	<i>glpK</i>	14.13	<b>TTTCGTTt t g a c a TTCGTTt t t c Ta TCTa t t gaa g TCGTT</b>
»	<i>glpD</i>	10.56	<b>a TTCGTTt t g a c a TTCGa a at a t TTTCGc T</b>
<i>B. pseudomallei</i>	<i>glpD</i>	10.40	TTTCGa T t a t g TTCGTT a aaTTTCGa a
<i>B. cepacia</i>	<i>glpD</i>	10.40	TTTCGa T t c c g TTCGTT a aaTTTCGa a
<i>B. fungorum</i>	<i>glpD</i>	9.02	TTTCGa a t a t g TTCa TT a ag TTCGa a
д – потенциальные сайты в геномах Enterobacteriaceae при сильном “перепредсказании”			
<i>K. pneumoniae</i>	<i>glpF</i>	4.00	GGc g CGAa Aa CGc t CA
<i>S. typhi</i>	<i>glpF</i>	4.03	T t c a CGt a Aa CGc g CA
<i>S. typhimurium</i>	<i>glpF</i>	4.03	T t c a CGt a Aa CGc g CA
<i>Y. enterocolitica</i>	<i>glpF</i>	3.91	AGc T t GATAa Ca At a A
»	<i>glpE</i>	3.79	T t a g Ca ATATg GAACA
<i>Y. pestis</i>	<i>glpF</i>	4.03	T a c g CGAa Aa CGc t CA
»	<i>glpF</i>	3.95	T t c TCGt Tt TCGc t Cg

Примечание. Обучающая выборка выделена жирным шрифтом.

ластях футпринта, показанных в статьях [1–3], удается обнаружить потенциальные сайты.

Поскольку данные о трехмерной структуре регуляторов семейства DeoR отсутствуют, нет оснований полагать, что эти регуляторы во всех случаях образуют димеры, связывающиеся с палиндромными сайтами. Более того, принимая во внимание, что в некоторых других группах организмов нам удалось выделить сигнал в виде тандемного повтора, можно было предположить, что структура сайта в семействе Enterobacteriaceae тоже представляет собой тандемный повтор. При выравнивании областей перед геном *glpD* в организмах из семейства Enterobacteriaceae – *E. coli*, *S. typhimurium*, *K. pneumoniae*, *Y. enterocolitica* – можно усмотреть как палиндромную симметрию, так и повторы (табл. 3). В то же время построить хорошую матрицу позиционных весов для этих регуляторных областей не удается и в предпо-

жении наличия палиндромной симметрии, и для симметрии типа тандемного повтора.

На рис. 2 представлены примеры оперонов в некоторых из рассмотренных геномов, регулируемых репрессором GlpR и имеющих соответствующий сайт. Закрашенными кружками отмечены известные сайты, а незакрашенными – предсказанные, с их весами.

Итак, нами идентифицированы предположительные сигналы связывания GlpR у Vibrionaceae и Pseudomonadaceae (палиндром) и α-, β-протеобактерий (тандемный повтор). Ситуация в Enterobacteriaceae остается открытой. Выравнивание регуляторных областей и их анализ с помощью программ выделения сигналов позволяет предположить существование как палиндромного сигнала, так и сигнала в виде фазированного прямого повтора. Нельзя исключать того, что мономеры GlpR в этих бактериях могут связываться с регу-

**Таблица 2.** Матрица позиционных весов сигнала GlpR в указанных организмах

а – <i>E. coli</i> , <i>E. carotovora</i> , <i>Y. enterocolitica</i> , <i>K. pneumoniae</i>				б – <i>V. cholerae</i> , <i>V. vulnificus</i> , <i>V. fischeri</i>				в – <i>P. aeruginosa</i> , <i>P. syringae</i> , <i>P. fluorescens</i> , <i>P. putida</i>			
а	с	g	t	а	с	g	t	а	с	g	t
-0.16	-0.16	-0.16	0.48	0.35	-0.24	-0.24	0.13	0.25	-0.25	-0.25	0.25
-0.23	-0.23	0.38	0.09	0.29	-0.30	-0.02	0.03	-0.15	-0.15	-0.15	0.46
0.06	0.13	-0.34	0.15	0.05	-0.22	-0.22	0.39	0.03	-0.31	0.11	0.18
-0.24	-0.24	0.12	0.36	-0.06	-0.25	0.37	-0.06	-0.15	-0.15	-0.15	0.46
-0.24	0.36	-0.24	0.12	0.00	0.15	-0.18	0.04	-0.15	-0.15	-0.15	0.46
-0.25	-0.06	0.37	-0.06	-0.26	-0.26	0.24	0.27	-0.15	0.46	-0.15	-0.15
0.28	-0.26	-0.26	0.25	-0.16	0.47	-0.16	-0.16	0.18	-0.24	0.31	-0.24
0.15	-0.25	-0.25	0.35	-0.16	-0.16	0.47	-0.16	0.01	-0.09	0.01	0.07
0.35	-0.25	-0.25	0.15	0.22	-0.11	-0.30	0.20	0.29	-0.24	-0.24	0.20
0.25	-0.26	-0.26	0.28	0.20	-0.30	-0.11	0.22	0.20	-0.24	-0.24	0.29
-0.06	0.37	-0.06	-0.25	-0.16	0.47	-0.16	-0.16	0.07	0.01	-0.09	0.01
0.12	-0.24	0.36	-0.24	-0.16	-0.16	0.47	-0.16	-0.24	0.31	-0.24	0.18
0.36	0.12	-0.24	-0.24	0.27	0.24	-0.26	-0.26	-0.15	-0.15	0.46	-0.15
0.15	-0.34	0.13	0.06	0.04	-0.18	0.15	0.00	0.46	-0.15	-0.15	-0.15
0.09	0.38	-0.23	-0.23	-0.06	0.37	-0.25	-0.06	0.46	-0.15	-0.15	-0.15
0.48	-0.16	-0.16	-0.16	0.39	-0.22	-0.22	0.05	0.18	0.11	-0.31	0.03
				0.03	-0.02	-0.30	0.29	0.46	-0.15	-0.15	-0.15
				0.13	-0.24	-0.24	0.35	0.25	-0.25	-0.25	0.25
г – <i>B. melitensis</i> , <i>M. loti</i> , <i>A. tumefaciens</i> , <i>S. meliloti</i> , <i>R. palustris</i>											
0.19	-0.33	-0.10	0.24								
-0.11	-0.40	-0.11	0.63								
-0.26	-0.26	-0.26	0.79								
-0.33	0.71	-0.33	-0.04								
-0.11	-0.40	0.63	-0.11								
0.03	-0.06	-0.49	0.51								
0.19	-0.49	-0.19	0.49								

**Таблица 3.** Выравнивание областей перед геном *glpD* в *E. coli*, *S. typhimurium*, *K. pneumoniae*, *Y. enterocolitica*

Геном	Ген	Вес сайта	Сайт
<i>E. coli</i>	<i>glpD</i>	10.0	aat atgt t <b>cgat aacgaacatt</b> t at <b>gagctt</b> t <b>aacgaaagt</b> gaat
<i>S. typhimurium</i>	<i>glpD</i>	10.0	at t atgt t <b>cgat aacgaacatt</b> t t <b>gaactt</b> t <b>aacgaaagt</b> gca a
<i>K. pneumoniae</i>	<i>glpD</i>	9.8	at agt <b>gagcgat at cgagcatt</b> t at <b>gagctt</b> t <b>aacgaaagt</b> gt ga
<i>Y. enterocolitica</i>	<i>glpD</i>	8.9	at cgt <b>gagcgaaaacgaacatt</b> a <b>aagagctgtt</b> t <b>cgaaactt</b> t gg

Примечание. Жирным шрифтом показаны палиндромные области, а подчеркнуты повторы. Вес сайта соответствует палиндромному сайту.

ляторными областями в произвольных направлениях, образуя кооперативные комплексы (ср. регуляторы NarL [21] и Fur [22]). Ситуация может проясниться после появления экспериментальных данных, секвенирования новых геномов из этой группы или анализа других регуляторов из семейства DeoR.

Мы благодарны В.Ю. Макееву, А.А. Миронову и Д.А. Родионову за полезные обсуждения, а также К.Ю. Горбунову за помощь с алгоритмом.

Эта работа получила финансовую поддержку от Медицинского института Ховарда Хьюза (грант 55000309).

## СПИСОК ЛИТЕРАТУРЫ

1. Weissenborn D.L., Wittekindt N., Larson T.J. 1992. Structure and regulation of the *glpFK* operon encoding glycerol diffusion facilitator and glycerol kinase of *Escherichia coli* K-12. *J. Biol. Chem.* **267**, 6122–6131.
2. Larson T.J., Cantwell J.S., van Loo-Bhattacharya A.T. 1992. Interaction at a distance between multiple operators controls the adjacent, divergently transcribed *glpTQ-glpABC* operons of *Escherichia coli* K-12. *J. Biol. Chem.* **267**, 6114–6121.
3. Yang B., Larson T.J. 1996. Action at a distance for negative control of transcription of the *glpD* gene encoding sn-glycerol 3-phosphate dehydrogenase of *Escherichia coli* K-12. *J. Bacteriol.* **178**, 7090–7098.
4. Schweizer H.P., Po C. 1996. Regulation of glycerol metabolism in *Pseudomonas aeruginosa*: characterization of *glpR* repressor gene. *J. Bacteriol.* **178**, 5215–5221.
5. Blattner F.R., Plunkett G. 3rd, Bloch C.A. et al. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science.* **277**, 1453–1474.
6. Parkhill J., Dougan G., James K.D. et al. 2000. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar typhi CT18. *Nature.* **413**, 848–852.
7. McClelland M., Sanderson K.E., Spieth J. et al. 2001. Complete genome sequence of *Salmonella enterica* serovar typhimurium LT2. *Nature.* **413**, 852–856.
8. <http://genome.wustl.edu>
9. <http://www.ncbi.nlm.nih.gov/Genbank/index.html>
10. Parkhill J., Wren B.W., Thomson N.R. et al. 2001. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature.* **413**, 523–527.
11. Heidelberg J.F., Eisen J.A., Nelson W.C. et al. 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature.* **406**, 477–483.
12. May B.J., Zhang Q., Li L.L., Paustian M.L., Whitam T.S., Kapur V. 2001. Complete genomic sequence of *Pasteurella multocida* Pm70. *Proc. Natl. Acad. Sci. USA.* **98**, 3460–3465.
13. Fleischmann R.D., Adams M.D., White O. et al. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science.* **269**, 496–512.
14. [http://spider.jgi-psf.org/JGI\\_microbial/html/](http://spider.jgi-psf.org/JGI_microbial/html/)
15. <http://www.tigr.org>
16. <http://www.genome.ou.edu/act.html>
17. Thompson J.D., Higgins D.G., Gibson T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positionspecific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.
18. Lim A., Zhang L. 1999. WebPHYLP: a web interface to PHYLIP. *Bioinformatics.* **15**, 1068–1069.
19. Миронов А.А., Винокурова Н.П., Гельфанд М.С. 2000. Программное обеспечение анализа бактериальных геномов. *Молекуляр. биология.* **34**, 253–262.
20. Данилова Л.В., Горбунов К.Ю., Гельфанд М.С., Любецкий В.А. 2001. Алгоритм выделения регуляторных сигналов в последовательностях ДНК. *Молекуляр. биология.* **35**, 987–995.
21. Gennis R.B., Stewart V. 1996. *Escherichia coli* and *Salmonella*. Cellular and Molecular biology. ASM Press. Washington DC, 2822 p.
22. Panina E.M., Mironov A.A., Gelfand M.S. 2001. Comparative analysis of FUR regulons in gamma-proteobacteria. *Nucleic Acids Res.* **29**, 5195–5206.
23. <http://www.bio.cam.ac.uk/seqlogo/logo.cgi>

## Computer-assisted Analysis of Regulation of the Glycerol-3-Phosphate Metabolism in Genomes of Proteobacteria

L. V. Danilova<sup>1</sup>, M. S. Gelfand<sup>2</sup>, V. A. Lyubetsky<sup>1</sup>, and O. N. Laikova<sup>2</sup>

<sup>1</sup>*Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, 101447 Russia;*  
E-mail: [dlv2k@mail.ru](mailto:dlv2k@mail.ru)

<sup>2</sup>*State Research Center GosNIIGenetika, Moscow, 113545 Russia*

**Abstract**—Comparative computer-assisted analysis was used to study putative GlpR-regulons responsible for metabolism of glycerol and glycerol-3-phosphate in genomes of  $\alpha$ -,  $\beta$ -, and  $\gamma$ -proteobacteria. New palindromic GlpR-binding signals were identified in  $\gamma$ -proteobacteria; consensus sequences being TGTTCGATAACGAACA for Enterobacteriaceae, wTTTTTCGTATACGAAAaw for Pseudomonadaceae, and AATGCTCGATCGAGCATT for Vibrionaceae. The signals in  $\alpha$ - and  $\beta$ -proteobacteria were also identified: they contained 3–4 direct TTTTCGTT repeats separated by 3–4 nucleotide pairs.