



КОНФЕРЕНЦИЯ МОЛОДЫХ УЧЕНЫХ И СПЕЦИАЛИСТОВ

**Информационные  
ТЕХНОЛОГИИ И СИСТЕМЫ**  
**ИТиС'09**

СБОРНИК ТРУДОВ КОНФЕРЕНЦИИ

**ISBN 978-5-901158-11-1**

пос. д/о Бекасово,  
15 – 18 декабря 2009 г.

Информационные технологии и системы (ИТиС'09):  
сборник трудов конференции. – М.: ИППИ РАН, 2009. – 463 с.

Издание содержит труды ежегодной конференции молодых ученых и специалистов «Информационные технологии и системы» (ИТиС'09), традиционно организуемой Советом молодых ученых и специалистов ИППИ РАН.

Основная цель Конференции ИТиС'09 – дать возможность молодым ученым и специалистам различных подразделений ИППИ РАН, а также студентам, аспирантам и молодым ученым других институтов РАН, отраслевых институтов, университетов и вузов, познакомиться с коллегами и обменяться научными достижениями по основным для ИППИ РАН направлениям научной деятельности: теория передачи и защиты информации; математическая теория информации и управления, многокомпонентные случайные системы; информационно-коммуникационные технологии и их применение в сложных системах и сетях; информационные процессы в живых системах и биоинформатика; компьютерная лингвистика и моделирование естественного языка; высшая математика.

Все включенные в данный сборник работы прошли рецензирование и опубликованы в том виде, в котором они были представлены авторами, среди которых молодые ученые и специалисты ИППИ РАН, ИПУ РАН, ИККИ РАН, ИПМ РАН, ИПТМУ РАН, ИМБ РАН, ИОГен РАН, МГУ им. М.В. Ломоносова, МФТИ (ГУ), МАИ, НГУ, СПбГУ ИТМО, ЛГТУ, МПГУ, МТУСИ, РХТУ и др.

Труды Конференции могут представлять интерес для ученых, студентов и аспирантов, специализирующихся в областях науки, связанных с перечисленными выше научными направлениями.

Конференция проведена при финансовой поддержке ИППИ РАН, Отделения нанотехнологий и информационных технологий РАН, Целевой программы Президиума РАН «Поддержка молодых ученых 2009 г.» и РФФИ (грант 09-07-06810моб\_г)



**ISBN 978-5-901158-11-1**

© Коллектив авторов, 2009

© Учреждение Российской академии наук Институт проблем передачи информации им. А.А. Харкевича РАН, 2009

## Транскрипция генов синтеза пролина у бактерий родов *Marinobacter*, *Pseudomonas* и *Shewanella* регулируется белком семейства *tetR*

К.В. Лопатовская, О.А. Зверков, А.В. Селиверстов, В.А. Любецкий

ИППИ РАН

lyubetsk@iitp.ru

### Аннотация

Найден консервативный мотив перед генами *proA* и *proB* синтеза пролина у большинства видов  $\gamma$ -протеобактерий (родов *Pseudomonas*, *Marinobacter*, *Shewanella*) и у некоторых других видов. С помощью описанного ниже алгоритма сигнал отделен от шума и затем с помощью другой нашей программы проведено сравнение его филогенетического профиля с профилями всех бактериальных белков. Отсюда предположено, что соответствующим транскрипционным фактором является белок из семейства *tetR*, ортологичный белку NP\_249058 из *Pseudomonas aeruginosa* PAO1.

### 1. Введение

В работах [1, 2] экспериментально показана зависимость (от концентрации пролина) транскрипции генов, кодирующих два фермента (гаммаглутамилкиназу и гаммаглутамилфосфат-редуктазу) синтеза пролина, у *Pseudomonas aeruginosa*. Сначала анализ небольшого числа геномов позволил выявить консервативный мотив перед генами *pro* у некоторых представителей родов *Pseudomonas* и *Shewanella*, [3]. В настоящей работе расширен список видов, у которых обнаружен этот мотив, рассмотрен филогенетический профиль предсказанного сигнала, сайта связывания белка, перед генами *proA* и *proB* у  $\gamma$ -протеобактерий и в результате предположительно определен сам белок.

### 2. Методы

Отбор белков с филогенетическим профилем близким к профилю сигнала выполнен программой, написанной О. Зверковым по алгоритму из [4]. Для тех из этих белков, которые оказались представленными в *Pseudomonas aeruginosa* PAO1, проведена аннотация с помощью базы данных Pfam с целью определить, какой из них может выступать в качестве регулятора. Затем для этого белка (или белков) программой BLASTP 2.2.21 в базе данных NCBI

определены высокого качества гомологи и так полученный филогенетический профиль сравнивался с филогенетическими профилями исходного белка и снова сигнала. Близость этих профилей рассматривалась как подтверждение выделенного белка. Ниже описан примененный нами алгоритм выделения сигнала из шума. Для этого сначала каждому сайту из найденного мотива (множества сайтов) приписывался штраф. Затем в мотиве выделялся кластер, который и рассматривался в качестве сигнала. Простейший способ такой кластеризации состоял в том, чтобы удалить из мотива сайты со значением штрафа выше некоторого порога.

Штраф определялся следующим образом. Дано множество  $S$  «образцов» — известных сайтов одинаковой длины  $N$ , которые в нашем случае были составлены по сайтам из *P. aeruginosa* и близких к ней видов, в которых найден мотив. В этих сайтах мы уверены в силу экспериментального подтверждения сайта в *P. aeruginosa*. Затем дано множество  $C$  «кандидатов» — более длинных ( $\geq N$ ) сайтов, которые вошли в найденный мотив. Большая длина объясняется тем, что множественное выравнивание при удалении от *P. aeruginosa* «расплывается» и предполагаемый сайт на выравнивании берется с окрестностью; по крайней мере, такая картина наблюдалась в наших данных. По множеству образцов составлялась частотная матрица нуклеотидов, которая для наших целей представлялась в виде центра масс, см. ниже. А именно, каждое слово  $w$  длины  $N$  записывается точкой  $c(w)$  из  $(4^*N)$ -мерного пространства. Фактически эта точка принадлежит поверхности единичного гиперкуба с ребром от 0 до 1, точнее даже ее специальной части, но это не существенно для дальнейшего: любой букве в слове  $w$  соответствует в  $c(w)$  последовательная четверка координат, в которой три нуля и одна единица соответственно на 1-м месте для буквы А, на 2-м месте для буквы С, на 3-м месте для G, на 4-м месте для Т. Например, А соответствует четверке координат  $\{1;0;0;0\}$ , а С — четверке  $\{0;1;0;0\}$  и т.д.

Вычисляем центр масс  $c^0$  множества образцов как по координатное среднее арифметическое всех

точек  $c(s)$ , где  $s$  пробегает  $S$ . Затем для каждого слова  $s$  из  $S$  вычисляем его *штраф* - расстояние  $l(s, c^0)$  как сумму модулей разностей соответствующих координат. Вычисляем расстояние от центра масс  $c^0$  до каждого под слова длины  $N$  из каждого кандидата  $c$  из  $C$ . В качестве *штрафа* для самого кандидата  $c$  берем минимум расстояний от  $c^0$  до всех входящих в  $c$  под слов длины  $N$ .

Таким образом, каждый образец и каждый кандидат получает численную характеристику - штраф, который указан ниже.

### 3. Результаты

**Результаты для  $\gamma$ -протеобактерий.** Поиск привел к предсказанию сайтов связывания репрессора вблизи промотора перед генами *proA* или *proB* у большинства видов родов *Pseudomonas*, *Marinobacter* и *Shewanella*. Слабые сайты обнаружены также у *Oceanospirillum*, *Aeromonas hydrophila* и *Beggiatoa*. Штрафы сайтов приведены в таблице 1. В результате массового поиска в лидерных областях найдены консервативные сайты связывания белка с ДНК длиной 16 н (полтора витка ДНК) с консенсусом MTACCAAYNNYRSY. Перед некоторыми генами определяются пары сайтов, которые могут кооперативно связывать этот белок. Соответственно указаны одно или два значения штрафов.

Сопоставление филогенетических профилей предсказанных нами сайтов и всех белков у  $\gamma$ -протеобактерий выделило белок, который является гипотетическим регулятором транскрипции из семейства tetR. Это - белок NP\_249058 из *P. aeruginosa* PAO1, который содержит на N-конце гипотетический ДНК-связывающий домен, характерный для транскрипционных регуляторов семейства tetR (Pfam номер PF00440, для участка от 11-й до 57-й аминокислоты с уровнем отличия от консенсуса  $E=1.5 \cdot 10^{-13}$ ). Трехмерная структура домена гомологичного к этому определена в [5].

В таблице 2 приведены виды, имеющие белки высоко гомологичные белку NP\_249058 из *P. aeruginosa*, и для них указано значение E-характеристики (которое здесь не превосходит  $10^{-25}$ ). Данные получены той же программой BLASTP 2.2.21 на основе аминокислотных последовательностей из базы GenBank, NCBI.

**Результаты по другим бактериям.** Похожие сайты были найдены у небольшого числа представителей  $\alpha$ -,  $\beta$ -,  $\delta$ -протеобактерий и у типа Bacteroidetes. Они перечислены в таблице 3. Эти

сайты несколько отличаются от консенсуса, что может быть связано с удаленностью соответствующих видов. Но они относительно хорошо выравниваются между собой внутри каждой таксономической группы. Эти сайты предсказаны перед участком ТАТАА, характерным для (-10)-бокса промотора. В большинстве случаев эти сайты повторяются в той же лидерной области с небольшим расстоянием между повторами.

**Таблица 1. Штрафы потенциальных сайтов связывания регуляторного белка перед генами *proA* и *proB* у  $\gamma$ -протеобактерий**

Вид	Штрафы сайта
ген <i>proA</i>	
<i>Pseudomonas aeruginosa</i>	14.25
<i>Pseudomonas mendocina</i>	13.25; 19.5
<i>Pseudomonas syringae</i>	11.5
<i>Pseudomonas putida</i>	12.75; 22
<i>Pseudomonas entomophila</i>	12.5; 21.5
<i>Pseudomonas stutzeri</i>	14
<i>Marinobacter aquaeolei</i>	16.75; 24
<i>Marinobacter algicola</i>	15; 17.5
<i>Oceanospirillum</i> sp. MED92	18.5
<i>Marine <math>\gamma</math>-proteobacteria</i>	21.5; 23
ген <i>proB</i>	
<i>Shewanella amazonensis</i>	15.67
<i>Shewanella frigidimarina</i>	15.67; 11.25
<i>Shewanella oneidensis</i>	8.67; 5.75
<i>Shewanella</i> sp. MR-4	8.67; 6.5
<i>Shewanella</i> sp. MR-7	11; 6.5
<i>Shewanella</i> sp. ANA-3	8.33; 6.5
<i>Shewanella loihica</i>	8.25
<i>Shewanella baltica</i>	5
<i>Shewanella putrefaciens</i>	6.75
<i>Shewanella benthica</i>	11.25
<i>Shewanella sediminis</i>	10.25
<i>Shewanella woodyi</i>	14
<i>Shewanella piezotolerans</i>	14.5
<i>Shewanella pealeana</i>	14.5
<i>Shewanella halifaxensis</i>	16
<i>Aeromonas hydrophila</i>	19.75
<i>Beggiatoa</i> sp. PS	15.67; 15.75

<i>Pseudomonas aeruginosa</i>	22.25
<i>Pseudomonas fluorescens</i>	16.75

<i>Hahella chejuensis</i>	4·10 <sup>-25</sup>
<i>Vibrio shilonii</i>	4·10 <sup>-25</sup>
<i>Colwellia psychrerythraea</i>	6·10 <sup>-25</sup>

**Таблица 2. Виды, имеющие белки высоко гомологичные белку NP\_249058**

Вид	Expect value
<i>Pseudomonas aeruginosa</i>	4·10 <sup>-123</sup>
<i>Pseudomonas mendocina</i>	3·10 <sup>-88</sup>
<i>Pseudomonas fluorescens</i>	1·10 <sup>-87</sup>
<i>Pseudomonas putida</i>	2·10 <sup>-83</sup>
<i>Pseudomonas entomophila</i>	1·10 <sup>-81</sup>
<i>Pseudomonas stutzeri</i>	1·10 <sup>-77</sup>
<i>Pseudomonas syringae</i>	2·10 <sup>-74</sup>
<i>Oceanobacter sp. RED65</i>	6·10 <sup>-63</sup>
<i>Marinobacter aquaeolei</i>	1·10 <sup>-52</sup>
<i>Marinobacter algicola</i>	1·10 <sup>-51</sup>
<i>Hahella chejuensis</i>	1·10 <sup>-51</sup>
<i>Acinetobacter baumannii</i>	2·10 <sup>-46</sup>
<i>Acinetobacter sp. ADP1</i>	4·10 <sup>-46</sup>
<i>Marinobacter sp. ELB17</i>	8·10 <sup>-46</sup>
<i>Acinetobacter sp. ATCC 27244</i>	2·10 <sup>-45</sup>
<i>Marinomonas sp. MED121</i>	9·10 <sup>-45</sup>
<i>Alcanivorax borkumensis</i>	1·10 <sup>-43</sup>
<i>Shewanella woodyi</i>	6·10 <sup>-33</sup>
<i>Shewanella frigidimarina</i>	7·10 <sup>-33</sup>
<i>Shewanella pealeana</i>	2·10 <sup>-32</sup>
<i>Shewanella piezotolerans</i>	2·10 <sup>-32</sup>
<i>Shewanella sediminis</i>	2·10 <sup>-32</sup>
<i>Shewanella halifaxensis</i>	3·10 <sup>-32</sup>
<i>Shewanella oneidensis</i>	8·10 <sup>-32</sup>
<i>Shewanella benthica</i>	9·10 <sup>-32</sup>
<i>Vibrio cholerae</i>	1·10 <sup>-31</sup>
<i>Shewanella sp. MR-7</i>	1·10 <sup>-31</sup>
<i>Moritella sp. PE36</i>	2·10 <sup>-31</sup>
<i>Vibrio vulnificus</i>	2·10 <sup>-31</sup>
<i>Shewanella sp. ANA-3</i>	3·10 <sup>-31</sup>
<i>Shewanella sp. W3-18-1</i>	4·10 <sup>-31</sup>
<i>Pseudoalteromonas tunicata D2</i>	7·10 <sup>-31</sup>
<i>Vibrio parahaemolyticus</i>	7·10 <sup>-31</sup>
<i>Vibrionales bacterium SWAT-3</i>	9·10 <sup>-31</sup>
<i>Aeromonas hydrophila</i>	1·10 <sup>-30</sup>
<i>Aeromonas salmonicida</i>	2·10 <sup>-30</sup>
<i>Shewanella amazonensis</i>	2·10 <sup>-30</sup>
<i>Vibrio sp. MED222</i>	2·10 <sup>-30</sup>
<i>Shewanella baltica</i>	2·10 <sup>-30</sup>
<i>Vibrio campbellii</i>	3·10 <sup>-30</sup>
<i>Vibrio splendidus</i>	3·10 <sup>-30</sup>
<i>Vibrio alginolyticus</i>	3·10 <sup>-30</sup>
<i>Vibrio shilonii</i>	3·10 <sup>-30</sup>
<i>Shewanella sediminis</i>	6·10 <sup>-30</sup>
<i>Vibrio harveyi</i>	7·10 <sup>-30</sup>
<i>Vibrio splendidus</i>	9·10 <sup>-30</sup>
<i>Shewanella loihica</i>	1·10 <sup>-29</sup>
<i>Vibrio fischeri</i>	2·10 <sup>-29</sup>
<i>Shewanella putrefaciens</i>	2·10 <sup>-29</sup>
<i>Vibrio salmonicida</i>	4·10 <sup>-29</sup>
<i>Marinobacter algicola</i>	1·10 <sup>-28</sup>
<i>Pseudoalteromonas haloplanktis</i>	3·10 <sup>-28</sup>
<i>Vibrio angustum</i>	5·10 <sup>-28</sup>
<i>Marinobacter sp. ELB17</i>	9·10 <sup>-28</sup>
<i>Photobacterium sp. SKA34</i>	2·10 <sup>-27</sup>
<i>Shewanella denitrificans</i>	2·10 <sup>-27</sup>
<i>Alteromonadales bacterium TW-7</i>	2·10 <sup>-26</sup>
<i>Marinobacter aquaeolei</i>	3·10 <sup>-26</sup>
<i>Photobacterium profundum</i>	3·10 <sup>-26</sup>
<i>Reinekea sp. MED297</i>	6·10 <sup>-26</sup>

**Таблица 3. Штрафы потенциальных сайтов связывания регуляторного белка перед генами *proA* и *proB* у  $\alpha$ -,  $\beta$ -,  $\delta$ -протеобактерий и у типа Bacteroidetes**

Вид	Штраф сайта
ген <i>proA</i>	
<b><math>\alpha</math>-протеобактерии:</b>	
<i>Magnetospirillum magnetotacticum</i>	19.33
<b><math>\beta</math>-протеобактерии:</b>	
<i>Nitrosomonas eutropha</i>	20; 26.29
<i>Nitrosomonas europaea</i>	25.14
<i>Dechloromonas aromatica</i>	22.29
<i>Kingella oralis</i>	22.57
<i>Janthinobacterium sp. Marseille</i>	19.78; 20.57
<b><math>\delta</math>-протеобактерии:</b>	
<i>Desulfuromonas acetoxidans</i>	26.57
<i>Geobacter metallireducens</i>	20; 20.86
<i>Geobacter uraniireducens</i>	23.11; 22.86
<i>Desulfobalobium retbaense</i>	25.14
<i>Pelobacter carbinolicus</i>	26
<i>Haliangium ochraceum</i>	21.78; 20.57
<i>Anaeromyxobacter dehalogenans</i>	17.11; 21.57
<i>Anaeromyxobacter sp. Fw109-5</i>	17.11; 21.43
<b>Bacteroidetes:</b>	
<i>Gramella forsetii</i>	21
<i>Chlorobium phaeobacteroides</i>	23.5; 18.75
<i>Prevotella copri</i>	23.43
<i>Bacteroides intestinalis</i>	21
<i>Bacteroides eggerthii</i>	22.75; 19.43
<i>Bacteroides thetaiotaomicron</i>	22
<i>Bacteroides pectinophilus</i>	22.75; 22.29
<i>Pedobacter sp. BAL39</i>	22.75; 22.57
<i>Pedobacter heparinus</i>	25; 21.71
ген <i>proB</i>	
<b><math>\alpha</math>-протеобактерии:</b>	
<i>Beijerinckia indica</i>	16.25
<i>Agrobacterium radiobacter</i>	23
<i>Agrobacterium tumefaciens</i>	23.5
<i>Rhizobium leguminosarum</i>	21.5
<i>Rhizobium etli</i>	21.75; 20
<i>Methylocella silvestris</i>	21.5; 19
<i>Nitrobacter winogradskyi</i>	19.75
<i>Candidatus Pelagibacter</i>	25.25
<b><math>\beta</math>-протеобактерии:</b>	
<i>Methylobacillus flagellatus</i>	20.5
<i>Nitrospira multiformis</i>	22.25
<i>Chromobacterium violaceum</i>	14.25
<i>Polaromonas naphthalenivorans</i>	18
<b><math>\delta</math>-протеобактерии:</b>	
<i>Desulfonatrosospira thiodismutans</i>	16
<i>Syntrophus aciditrophicus</i>	16.75
<b>Bacteroidetes:</b>	
<i>Bacteroides coprophilus</i>	17
<i>Bacteroides uniformis</i>	16.5
<i>Candidatus Azobacteroides pseudotrichonymphae</i>	14.5
<i>Psychroflexus torquis</i>	16.5
<i>Flavobacteriales bacterium</i>	20.75
<i>Sphingobacterium spiritivorum</i>	19.5
<i>Pelodictyon phaeoclathratiforme</i>	15.5

#### 4. Обсуждение

Итак, у  $\gamma$ -протеобактерий филогенетический профиль белков гомологичных белку NP\_249058 близок к профилю сайтов связывания регуляторов транскрипции перед *pro* генами. Гомологи этого белка присутствуют у видов из родов *Pseudomonas*, *Marinobacter* и *Shewanella*, для которых предсказана регуляция. Найденный белок имеет паралог NP\_253746 в том же виде *P. aeruginosa*. Близкие гомологи обоих паралогов обнаружены нами у *Vibrio* и у некоторых других видов. Можно думать, что эти паралоги и гомологи у видов, в которых не обнаружен сигнал (синтез пролина не регулируется на уровне транскрипции), имеют альтернативную специфичность хотя и принадлежат тому же семейству tetR.

Можно думать, что описанная здесь регуляция генов *pro* осуществляется белком семейства tetR, который впервые возник у *Pseudomonas* в результате дупликации. В дальнейшем регуляция распространялась между видами с помощью горизонтальных переносов. Это согласуется с мозаичным характером филогенетического профиля белка NP\_249058.

#### 5. Список литературы

- [1] R.V. Krishna, P. Beilstein, T. Leisinger, "Biosynthesis of proline in *Pseudomonas aeruginosa*. Partial purification and characterization of  $\gamma$ -glutamyl kinase", *Biochem. J.* 1979, 181, pp. 215-222.
- [2] R.V. Krishna, P. Beilstein, T. Leisinger, "Biosynthesis of proline in *Pseudomonas aeruginosa*. Properties of gamma-glutamyl phosphate reductase and 1-pyrroline-5-carboxylate reductase", *Biochem. J.*, 1979, 181, pp. 223-230.
- [3] А.В. Селиверстов, В.А. Любецкий, "Регуляция биосинтеза пролина у протеобактерий", *Молекулярная биология*, 2007, 41(3), с. 572-574.
- [4] Л.А. Леонтьев, В.А. Любецкий, "Алгоритм определения белка, согласованного с заданным филогенетическим профилем", *Информационные процессы*, 2006, 6(1), с. 24-32.
- [5] C. Kisker, W. Hinrichs, K. Tovar, W. Hillen, W. Saenger, "The complex formed between Tet repressor and tetracycline-Mg<sup>2+</sup> reveals mechanism of antibiotic resistance", *Journal of Molecular Biology*. 1995, 247(2), с. 260-280.