

УДК 577.053

РЕГУЛЯЦИЯ БИОСИНТЕЗА ПРОЛИНА У ПРОТЕОБАКТЕРИЙ

© 2007 г. А. В. Селиверстов*, В. А. Любецкий

Институт проблем передачи информации Российской академии наук, Москва, 127994

Поступила в редакцию 07.06.2006 г.

Принята к печати 28.08.2006 г.

Ключевые слова: биосинтез пролина, поиск сигнала, экспрессия генов.

REGULATION FOR PROLINE BIOSYNTHESIS IN PROTEOBACTERIA, by A. V. Seliverstov*, V. A. Lyubetsky (Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, 127994 Russia; *e-mail: slvstv@iitp.ru).

Key words: proline biosynthesis, signal finding, gene expression.

У бактерий биосинтез пролина из глутаминовой кислоты идет при последовательном участии трех ферментов: γ -глутамилкиназы, γ -глутамилфосфатредуктазы и 1-пирролин-5-карбоксилатредуктазы. Согласно данным Кришны и соавт. [1], у *Pseudomonas aeruginosa* уровень синтеза первых двух ферментов зависит от концентрации пролина, но синтез третьего фермента от нее не зависит. Гены *proA* и *proB* кодируют первые два из этих ферментов. У протеобактерий гены *proA* и *proB* обычно образуют оперон. У *Pseudomonas* гены *proA* и *proB* расположены на хромосоме далеко друг от друга и транскрибируются в противоположных направлениях. У *Bacillus subtilis* транскрипция оперона *proBA* регулируется с помощью Т-бокса [2], который отвечает за общий механизм регуляции генов, кодирующих аминокил-ТРНК-синтетазы, белки метаболизма аминокислот и аминокислотные транспортеры, и который найден в основном в грамположительных бактериях [3, 4]. В грамотрицательных бактериях такая регуляция встречается редко, поэтому представляет интерес изучить белок-ДНКовую регуляцию. У *Escherichia coli* рассматриваемые ферменты ингибируются пролином. Сообщений о белок-ДНКовой регуляции у *E. coli* нет.

В статье приводятся результаты массового поиска сайта связывания с ДНК некоторого регуляторного белка перед указанными генами синтеза пролина. Геномы бактерий взяты из ГенБанка. Поиск консервативных участков выполнен программой, основанной на оригинальном алгоритме поиска клики в многодольном графе, который описан ранее [5].

* Эл. почта: slvstv@iitp.ru

На первом этапе работы были исследованы 5'-лидерные области длиной до 200 н. у оперонов, содержащих гены *proA* или *proB*, во всех полностью секвенированных γ - и α -протеобактериях. Если лидерная область имеет большую длину, то рассматривали только первые 200 н. перед первым геном оперона. В результате массового поиска в указанных лидерных областях найдены консервативные участки, предположительно содержащие сайты регуляции соответствующих оперонов. Консервативные участки длиной 16 н. (полтора витка ДНК) и консенсусом МТАССААУНННУУУУУ обнаружены у γ -протеобактерий родов *Pseudomonas* и *Shewanella*. Перед геном *proA*, кодирующим γ -глутамилкиназу, сайт найден в ДНК *P. aeruginosa* PAO1, *P. putida* KT2440, *P. syringae* pv. *tomato* str. DC3000, *Shewanella oneidensis* MR-1, *Sh. amazonensis* SB2B, *Sh. sp.* PV-4, *Sh. frigidimarina* NCIMB 400, *Sh. baltica* OS155, *Sh. putrefaciens* CN-32, *Sh. sp.* MR-4, *Sh. sp.* MR-7, *Sh. sp.* ANA-3. Перед геном *proB*, кодирующим γ -глутамилфосфатредуктазу, сайт найден у *P. aeruginosa* PAO1 и *P. fluorescens* Pf-5. Соответствующее множество выравнивание представлено на рисунке. Отметим, что в 7 случаях из 15 найденный сайт повторяется в той же лидерной области с небольшим расстоянием между повторами. В этих семи случаях мы предполагаем кооперативное связывание регулятора. В остальных геномах перед оперонами, содержащими гены *proA* или *proB*, алгоритм не нашел консервативных участков. Алгоритм также не нашел в лидерных областях рассмотренных генов консервативных шпилек (инвертированных повторов).

Можно предположить, что найденный консервативный участок совпадает с сайтом или содержит сайт связывания белка, который регулирует

Перед геном γ -глутамилкиназы	
<i>P. aeruginosa</i>	tCTACCGACCCGGCTGTT**gсggctggggсgсgсggggctggctcgсTATAAтс
<i>P. putida</i>	gCCACCCCGCCGCTGGGC***gtACACCTGCGCGGCAGT*****TATAAтс
<i>P. syringae</i>	tCGACCATCCGCCTGTctg**cgctgtcgттсgтgccccacgacagсTATAAтс
<i>S. oneidensis</i>	tCTACCAACATCCTGCCcacttaCTACCAATTACSTAGCggattaggTATAAgg
<i>S. amazonensis</i>	aATATCCATCGGCGCCgttgtgcaaggсgggttgct*****ttgggTATAAgg
<i>S. sp. PV-4</i>	gctctaactcttttatcttaccсgCTACCAATTTGCCCTGGggattgggTATAAтс
<i>S. frigidimarina</i>	tATTCCAGTATTAGCGGctttcaCTATGAATTATCGAGTgaattgggTATAAaa
<i>S. baltica</i>	tatctcgagttaataactccttgCTACCAATTACCCAGCggattaggTATAAgg
<i>S. putrefaciens</i>	tatctcgagttaactcactcctcgCTACCAATTATCTAGCggattaggTATAAgg
<i>S. sp. MR-4</i>	tCTACCAACATCCTGCCcactaaCTACCAATTGCTCTGCggattaggTATAAgg
<i>S. sp. MR-7</i>	tATACCAACATGTAGCTcacttaCTACCAATTGCCAACggattaggTATAAgg
<i>S. sp. ANA-3</i>	tCTACCAACATCCGGCCcactaaCTACCAATTGCTCTGCggattaggTATAAgg
Перед геном γ -глутамилфосфатредуктазы	
<i>P. aeruginosa</i>	gATACCCCGCCGGCGGGссggсggggctgaagagtgсgaagcctgggctCTCAAagg
<i>P. fluorescens</i>	tCGACCATCCGCCCGTGaccgctgtcc
Перед геном гипотетического белка AAN56706	
<i>S. oneidensis</i>	cAAACCAATTCCGGAGACtt*aaaGATAACAATCGACTAAgctagacacactctcg

Множественное выравнивание лидерных областей, в которых обнаружен предполагаемый сайт связывания регуляторного белка. Прописными буквами выделены предполагаемые сайты связывания репрессора и также (–10)-боксы промотора. *P* – *Pseudomonas*; *S* – *Shewanella*.

транскрипцию в зависимости от концентрации пролина. У *Pseudomonas* и *Shewanella* сайты, регулирующие транскрипцию гена γ -глутамилкиназы, предсказаны перед участком ТАТАА, характерным для (–10)-боксы промотора.

У *Azotobacter vinelandii* перед рассмотренными оперонами такой сайт не найден, хотя эта бактерия эволюционно близка к тем, у которых сайт предсказан нами. У *P. putida* и *P. syringae* сайт обнаружен только перед геном, кодирующим γ -глутамилкиназу, а у *P. fluorescens* этот сайт обнаружен только перед геном, кодирующим γ -глутамилфосфатредуктазу; в то же время у *P. aeruginosa* сайт найден перед обоими генами *proA* и *proB*.

Найденные сайты у *P. aeruginosa* расположены перед генами *proA* и *proB*, для которых экспериментально доказано наличие регуляции посредством изменения концентрации пролина, хотя механизм такой регуляции в опыте не показан [1].

Согласно данным Кришны и соавт. [1], амплитуда изменения экспрессии генов у *P. aeruginosa* составляет лишь около 40%. И действительно, поскольку у *P. aeruginosa* предполагаемый сайт связывания регуляторного белка не повторяется, то это, возможно, объясняет столь низкую эффективность регуляции. Известно, что повтор сайта обычно увеличивает амплитуду экспрессии за счет кооперативного связывания регулятора.

Роды *Pseudomonas* и *Shewanella* достаточно далеки друг от друга, и поэтому существенно, что области между найденными сайтами перед ортологичными генами у видов из этих разных родов неконсервативны, что видно на рисунке.

На втором этапе работы предпринято исследование, направленное на то, чтобы дополнительно проверить неслучайность найденного множественного выравнивания предложенных сайтов. Обыч-

но в таких случаях ограничиваются предъявлением самого множественного выравнивания. Для этого исследовали 5'-лидерные области длиной до 200 н. или обрезанные таким образом перед всеми генами у *S. oneidensis*. В рассматриваемых областях слово, близкое к найденному консенсусу длиной 16 н., найдено только в двух случаях: перед опероном *proAB* и перед гипотетическим геном, кодирующим неизвестный белок AAN56706; и в каждом из них это слово входит с повтором. Первое из этих слов с повтором алгоритм уже нашел с самого начала, на первом этапе работы, второе из этих слов, также с повтором, включено на этом этапе в число позитивных результатов, показанных на рисунке, хотя здесь ситуация кажется менее ясной в силу следующих соображений. Белок AAN56706 включает домен, гомологичный домену из активатора транскрипции PivY у *E. coli*, и второй сайт из найденного повтора перед геном белка AAN56706 примыкает к возможному (–10)-боксы промотора. С другой стороны, неясна связь этого белка с пролином, и гомологи этого гена у других видов *Shewanella* не имеют сайта, соответствующего консенсусу, в их лидерной области (у *S. baltica* и *Shewanella* sp. PV-4) или гомологи следуют непосредственно за другим геном, т.е. фактически отсутствует подходящая лидерная область (у *S. denitrificans* и *S. frigidimarina*).

Если найденный консенсус обрезать до его более консервативной части с длиной 8 н., а именно, до слова МТАССААУ, то у *S. oneidensis* в тех же лидерных областях сайт, близкий к слову МТАССААУ, встречается только перед 80 генами. Однако и при таком, более широком, поиске (за исключением двух упомянутых выше случаев) каждый найденный сайт не повторяется два или более раз внутри лидерной области и, главное, не предше-

ствует слову, которое отличается от ТАТАА не более чем в одной позиции, и находится от этого сайта на расстоянии в 10–20 н. Таким образом, поиск по всему геному *S. oneidensis* не приводит к новому сайту, подобному тем, которые показаны на рисунке. Множественное выравнивание, показанное на рисунке, и описанное массовое обследование генома бактерии из рода *Shewanella* говорят в пользу потенциальной значимости предсказанных сайтов.

Авторы благодарны М.С. Гельфанду и А.Г. Витрецаку за ценное обсуждение и критические замечания. Авторы также благодарят рецензента за существенные исправления и замечания, позволившие улучшить текст заметки.

Работа получила финансовую поддержку Международного научно-технического центра (2766).

СПИСОК ЛИТЕРАТУРЫ

1. Krishna R.V., Beilstein P., Leisinger T. 1979. Biosynthesis of proline in *Pseudomonas aeruginosa*. Properties of gamma-glutamyl phosphate reductase and 1-pyrroline-5-carboxylate reductase. *Biochem. J.* **181**, 223–230.
2. Chopin A. 1998. Analysis of the *Bacillus subtilis* genome sequence reveals nine new T-box leaders. *Mol. Microbiol.* **29**, 661–669.
3. Grundy F.J., Henkin T.M. 2003. The T box and S box transcription termination control systems. *Frontiers Biosci.* **8**, d20–31.
4. Леонтьев Л.А., Селиверстов А.В., Любецкий В.А. 2005. Алгоритм массового поиска у бактерий вторичных структур, включающих Т-боксы. *Молекуляр. биология.* **39**, 1076–1078.
5. Селиверстов А.В., Любецкий В.А. 2006. Алгоритм поиска консервативных участков нуклеотидных последовательностей. *Информационные процессы.* **6**, 33–36.