

УДК 575.852

АТТЕНУАТОРНАЯ РЕГУЛЯЦИЯ ОПЕРОНОВ БИОСИНТЕЗА АМИНОКИСЛОТ И АМИНОАЦИЛ-тРНК У БАКТЕРИЙ: СРАВНИТЕЛЬНЫЙ ГЕНОМНЫЙ АНАЛИЗ

© 2010 г. К. В. Лопатовская*, А. В. Селиверстов, В. А. Любецкий

Институт проблем передачи информации им. А.А. Харкевича Российской академии наук, Москва, 127994

Поступила в редакцию 13.02.2009 г.

Принята к печати 05.06.2009 г.

Проведен широкомасштабный поиск потенциальных аттенуаторных регуляторных структур у бактерий, который опирается на две оригинальные компьютерные программы – модель процесса аттенуаторной регуляции и множественное выравнивание по данному филогенетическому дереву. Эти программы доступны для свободного пользования на сайте <http://lab6.iitp.ru>. Найдены потенциально возможные способы аттенуаторной регуляции у многих бактерий из таксонов α -, β -, γ -, δ -протеобактерий, Actinobacteria, Bacteroidetes/Chlorobi, Firmicutes и Thermotogae и Cloroflexi перед генами *hisG*, *hisZ*, *hisS*, *pheA*, *pheST*, *trpEG*, *trpA*, *trpB*, *trpE*, *trpS*, *thrA*, *thrS*, *leuA*, *leuS*, *ilvB*, *ilvI*, *ilvA*, *ilvC*, *ilvD*, *ilvG*. Проведен поиск по всему разнообразию секвенированных геномов бактерий в ГенБанке NCBI. При изучении других бактерий нашим методом аттенуаторная структура не обнаружена. В некоторых случаях можно предположить, что роль РНК-триплексов при формировании эффективных антитерминатора и терминатора, а также псевдоузлов в процессе терминации весьма существенна. Предполагается, что аттенуаторная регуляция гена *lys Q* у *Lactobacillus lactis* зависит от концентрации гистидил-тРНК. Рассматриваются различные типы аттенуаторной регуляции и обсуждается ее эволюция.

Ключевые слова: экспрессия генов у бактерий, аттенуаторная регуляция, широкомасштабный поиск, алгоритм поиска аттенуаторной регуляции, алгоритм выравнивания по дереву.

ATTENUATION REGULATION OF AMINO ACID AND AMINO ACYL-tRNA BIOSYNTHETIC OPERONS IN BACTERIA: COMPARATIVE GENOMICS ANALYSIS by K. V. Lopatovskaya*, A. V. Seliverstov, V. A. Lyubetsky (Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, 127994 Russia; *e-mail: kristina@iitp.ru). We performed a large-scale search for attenuation regulation in bacteria based on two original computer programs modeling attenuation regulation and multiple alignment along a phylogenetic tree. The programs are available from <http://lab6.iitp.ru>. Candidate attenuations are predicted in many bacteria from α -, β -, γ -, δ -proteobacteria, Actinobacteria, Bacteroidetes/Chlorobi, Firmicutes and Thermotogae; in Cloroflexi – upstream genes *hisG*, *hisZ*, *hisS*, *pheA*, *pheST*, *trpEG*, *trpA*, *trpB*, *trpE*, *trpS*, *thrA*, *thrS*, *leuA*, *leuS*, *ilvB*, *ilvI*, *ilvA*, *ilvC*, *ilvD*, *ilvG*. Other bacterial taxa were not predicted to have attenuation; searches were conducted across all bacterial genomes contained in GenBank, NCBI. Evolution of attenuation is discussed.

Key words: gene expression in bacteria, attenuation regulation, large-scale searches, attenuation prediction algorithms, tree-based multiple alignment algorithms.

ВВЕДЕНИЕ И ПОСТАНОВКА ЗАДАЧИ

Эта работа представляет собой продолжение работы, опубликованной ранее [1], как в части представленного там метода поиска потенциальных аттенуаторных структур, так и в части самого широкомасштабного поиска; в частности, здесь предполагаются новые варианты аттенуаторной регуляции, исследуется роль РНК-триплексов и псевдоузлов в этом процессе. Название настоящей работы продолжает название предыдущей [1], которая включает подробный

обзор литературы. Поэтому в следующем абзаце мы лишь коснемся исторического аспекта проблемы.

Бактерии используют различные механизмы регуляции экспрессии генов на уровнях как собственно процессов транскрипции и трансляции, так и посттранскрипционных и посттрансляционных преобразований. Среди них особенно широко исследуются следующие: регуляция, основанная на белок-ДНКовых взаимодействиях (репрессия и активация инициации транскрипции гена), – этой теме посвящено большое число работ; регуляция на основе формирования Т-боксов [2–4] или рибопереключателей [5, 6]; классическая аттенуаторная регуляция,

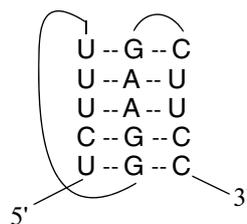
* Эл. почта: kristina@iitp.ru

описанная в работах Яновского и соавт. [7–9]. Ссылки, касающиеся этой регуляции и поясняющие ее рисунки, также приведены во введении к нашей предыдущей статье [1]. Что касается упомянутых постпреобразований, то эта тема исследуется в ряде работ, среди которых особенно отметим следующие [5, 10]. У многих протеобактерий и актинобактерий опероны *trp*, *thr*, *his*, *leu*, *ilv*, кодирующие ферменты синтеза аминокислот и аминоксил-тРНК-синтетаз, регулируются аттенуаторной структурой [11, 12]. Аттенуаторная регуляция зависит от концентрации соответствующей аминоксил-тРНК и часто распространяется на несколько оперонов из одного генома, а именно — на кодирующие ферменты синтеза самой аминокислоты и аминоксил-тРНК-синтетазы.

В этой работе принимается, что понятие *аттенуаторная регуляция* относится к структуре, которая осуществляет эту регуляцию и которая включает ген лидерного пептида с регуляторными кодонами в нем и связанные с ним альтернативные вторичные структуры мРНК. Одни из них (“терминирующие” или “секвестрирующие”) приводят к преждевременной терминации транскрипции структурного гена или к подавлению его трансляции, а другие (“анти-терминирующие” или “антисеквестрирующие”) позволяют РНК-полимеразе продолжить транскрипцию структурного гена. Эта альтернатива зависит от скорости, с которой рибосома выполняет трансляцию лидерного пептида, а, точнее, от скорости, с которой она проходит регуляторные кодоны. В свою очередь скорость зависит от концентрации регулируемого субстрата или связанного с ним вещества. Терминирующие и секвестрирующие структуры во всех наблюдаемых случаях более консервативны, чем анти-терминирующие и антисеквестрирующие структуры.

Понятие *анти-терминатор* далее относится к шпильке, разрушаемой рибосомой при ее быстром движении по гену лидерного пептида и имеющей регуляторное значение, а *терминатор* — к шпильке, которая обычно приводит к преждевременной терминации транскрипции структурного гена.

Среди типов аттенуаторной регуляции особенно важна *классическая*, которая действует на уровне транскрипции структурного гена, и ее терминаторные и анти-терминаторные вторичные структуры альтернативны *в том или ином смысле*. А именно, в понятие классической аттенуаторной регуляции входят следующие случаи. Во-первых, регуляция “по Яновскому”: терминатор и анти-терминатор непосредственно альтернативны друг другу и имеется участок, обогащенный U (полиурациловый тракт), расположенный около 3'-края терминатора. Во-вторых, “цепь спиралей”: терминатор и анти-терминатор не альтернативны друг другу, но имеется цепь обычно из четырех шпилек, в которой первая шпилька, анти-терминатор, в случае ее образования запрещает следующую (“вторую”, или “котерминатор”); тогда образуется следующая шпилька (“тре-



5' CCUGACUAGUCUUUCAGGCGAUGUGUG
CUGGAAGACAUUCAGAUCUCCAGUG 3'

Рис. 1. РНК-триплекс перед геном *hisG* у *E. coli*. “Третье плечо” триплекса равно UCUUU, спираль имеет плечи GGAAG и CUUC.

тья”, или “коанти-терминатор”), которая запрещает терминатор (“четвертую” шпильку); здесь также присутствует U-богатый участок, расположенный около 3'-края терминатора. В этом случае обычно предполагается, что стабильность котерминатора и/или коанти-терминатора поддерживается за счет образования РНК-триплексов. В-третьих, “ансамбль анти-терминаторов”: консервативный анти-терминатор отсутствует, его роль берет на себя ансамбль шпилек, каждая из которых альтернативна консервативному терминатору. Как и в предыдущих случаях, здесь имеется полиуридиловый тракт, и устойчивость некоторых шпилек может поддерживаться за счет формирования РНК-триплексов.

Напомним, что *U-богатый участок* (полиуридиловый тракт) — это участок длиной обычно около семи нуклеотидов, в основном — остатков урацила; о других использованных выше терминах можно узнать в нашей статье [13].

Отметим отдельный случай классической аттенуаторной регуляции “без U-богатого участка”, когда терминатор и анти-терминатор непосредственно альтернативны друг другу, но U-богатый участок отсутствует или содержит 2–3 остатка урацила на заметном расстоянии друг от друга. По-видимому, этот случай реально встречается, но его необходимо отличать от случая, описанного в следующем абзаце, как бы промежуточного между классической аттенуаторной регуляцией и регуляцией на уровне трансляции.

Секвестр-аттенуаторная регуляция содержит все элементы по Яновскому кроме U-богатого участка, но осуществляется на уровне трансляции: шпилька, аналогичная терминатору, перекрывает сайт связывания рибосомы.

Используя множественные выравнивания, трудно выявить ансамбль терминаторов, каждый из которых слабо консервативен; для различения вариантов регуляции без U-богатого участка, вообще говоря, также недостаточно предъявить множественное выравнивание (эти случаи ниже не показаны).

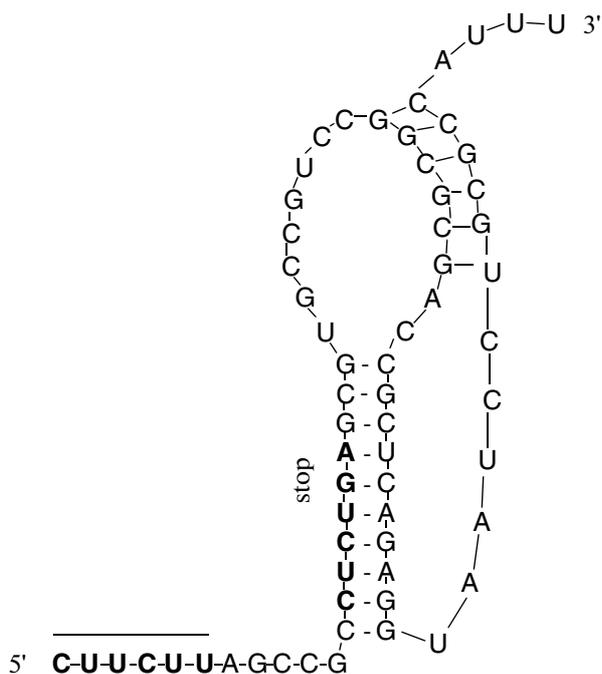


Рис. 2. LEU1-псевдоузел перед геном *leuA* у *D. shibae*. Соответствующая регуляция включает ген лидерного пептида с лейциновыми кодонами. Левое плечо спирали находится вблизи этих кодонов. Черенок отсутствует.

Цепи РНК могут образовывать *триплексы*. В них входят три участка, один из которых состоит полностью из пуриновых нуклеотидов, а два других имеют относительно пуринового участка параллельную и антипараллельную ориентации (рис. 1). Пара антипараллельных участков образует спираль РНК, в которой комплементарными считаются уотсон-криковские пары и пара GU. Таким образом, два участка триплекса образуют обычную спираль, а еще один, параллельный, связан с ней хустиновскими водородными связями. Последний участок называется *третьим плечом* триплекса, которое, как и составляющие его нуклеотиды, в записи отделяется знаком * от самой спирали (от ее нуклеотидов). В триплексе возможны триады нуклеотидов: C*GU, G*GC, G*GU, U*AU, A*AU, A*GC, C*GC и некоторые другие. Триады характеризуются рядом особенностей. Например, C*GC прочна лишь в слабокислой среде, когда происходит протонирование по атому N3 цитозина. По теме РНК-триплексов имеется много публикаций, среди которых отметим некоторые [14–19]. Если третье плечо триплекса находится на 5'-конце его спирали, то необходимо, чтобы оно располагалось на определенном расстоянии от соседнего к нему плеча спирали, достаточном для сворачивания триплекса. Мы принимали: если третье плечо содержит *l* нуклеотидов, то оно отделено не менее чем на *l* + 6 нуклеотидов от соседнего плеча. Если третье плечо находится на 3'-конце спира-

ли, то оно может быть и вблизи нее, как показано ранее [18, 19].

В качестве *неклассической аттенуаторной регуляции* в этой работе, кроме упомянутой выше секвестр-аттенуаторной, рассматриваются LEU1- и LEU-регуляции. Их структуры определяются наличием гена лидерного пептида с регуляторными кодонами в нем и формированием псевдоузлов, показанных соответственно на рис. 2 и 3¹. *Псевдоузел* в нашем случае формируется двумя спиралями, одна из которых имеет плечо во внешней петле другой. “*Левой*” называют ту из этих спиралей, 5'-край которой ближе к 5'-концу исходной последовательности, и “*правой*” — ту из них, 5'-край которой находится во внешней петле левой спирали. Собственно псевдоузел располагается во внешней петле левой спирали. Черенком псевдоузла называется ближайшая к нему спираль, в петле которой находится весь псевдоузел. LEU- и LEU1-псевдоузлы имеют некоторое структурное сходство между собой, но, в то же время, значительно отличаются по размерам, нуклеотидному составу, по положениям нуклеотидов относительно структурного гена и гена лидерного пептида. В частности, у LEU-структуры обычно имеется черенок, а у LEU1-структуры его нет.

Механизм LEU-регуляции основан на перекрывании сайта связывания рибосомы: если рибосома быстро проходит ген лидерного пептида, то черенок образуется; если рибосома “стоит” на регуляторных кодонах, то она препятствует образованию черенка (рис. 3). Подробно этот механизм обсуждается в нашей работе [2]. Возможно, возникновение LEU-элемента связано с горизонтальным переносом от предка Bifidobacteriales к предку Actinomycetales: у *Bifidobacteriales longum* сохранилась открытая рамка считывания транспозазы, в которой располагается LEU-псевдоузел, а у Actinomycetales эта рамка эволюционно трансформировалась в регуляторный элемент перед геном *leuA* [2]. Вопрос о механизме и, тем более, эволюции LEU1-регуляции менее ясен и кратко обсуждается в разделе “Результаты и обсуждение” в связи с регуляцией у α -протеобактерий.

В этой работе исследуется вопрос о том, имеется ли аттенуаторная регуляция экспрессии генов *hisG*, *hisZ*, *hisS*, *pheA*, *pheST*, *trpE*, *trpEG*, *trpA*, *trpB*, *trpS*, *thrA*, *thrS*, *leuA*, *leuS*, *ilvB*, *ilvI*, *ilvA*, *ilvC*, *ilvD*, *ilvG* во всем разнообразии секвенированных геномов бактерий из ГенБанка NCBI [20, 21]. В результате, в одних таксонах предсказано большое число новых типов аттенуаторной регуляции, в других — нашим методом ее обнаружить перед этими генами не удалось.

Перечисленные выше белоккодирующие гены отбирали по признаку того, насколько систематически в их лидерной области обнаруживается ген лидерного пептида, включающий регуляторные кодо-

¹Рисунки 3–37, подписи к ним и таблица 2 расположены на вэб-странице http://www.molecbio.com/downloads/2010/1/supp_lopat_rus.pdf

ны для аминокислот, которые биохимически связаны с последующим структурным геном.

Перечислим функции этих генов. Гены *hisG* и *hisZ* кодируют субъединицы АТР-фосфорибозилтрансферазы (ЕС2.4.2.17), которая катализирует формирование N'-5'-фосфорибозил-АТР из фосфорибозилпирофосфата. Белки HisG and HisZ формируют гетеромультимер. Ген *hisS* кодирует гистидил-тРНК-синтетазу. Ген *pheA* кодирует фермент префенатдегидратазу, участвующий в синтезе ароматических аминокислот (ЕС 4.2.1.51). Оперон *pheST* кодирует α - и β -субъединицы фенилаланил-тРНК-синтетазы; субъединицы образуют гетеротетрамер $\alpha_2\beta_2$ и связывают два иона магния. Гены *trpA* и *trpB* кодируют субъединицы триптофансинтазы, а ген *trpE* или оперон *trpEG* – субъединицы антранилатсинтазы, часто соединенные в один белок. Ген *trpS* кодирует триптофанил-тРНК-синтетазу. Ген *thrA* кодирует бифункциональную аспартаткиназу/гомосериндегидрогеназу, а *thrS* кодирует треонил-тРНК-синтетазу. Ген *leuA* кодирует 2-изопропилмалатсинтазу. У многих γ -протеобактерий этот ген входит в оперон *leuABCD*, кодирующий гены пути синтеза лейцина, у других видов ген *leuA* часто не входит в полицистронный оперон. Ген *leuS* кодирует лейцил-тРНК-синтетазу. Ген *ilvD* кодирует фермент дегидратазу дигидроксикислот, которая катализирует дегидратацию 2,3-дигидрокси-3-метилбутаноата до 3-метил-2-оксипропаноата при синтезе валина и изолейцина. Многие γ -протеобактерии содержат оперон *ilvGME-DA*, кодирующий четыре из пяти ферментов пути синтеза изолейцина и валина. Ген *ilvB* кодирует большую субъединицу ацетолактатсинтазы, часто в составе оперонов *ilvBNC* или *ilvBHC*, где *ilvN* и *ilvH* кодируют малую субъединицу ацетолактатсинтазы, а *ilvC* кодирует кетол-кислотную редуктоизомеразу. Ген *ilvA* кодирует треониндегидратазу.

Сравним метод, использованный в этой работе и изложенный ниже, с уже опубликованным методом [1]. Метод, изложенный в [1], состоит в использовании представленной там программы LLLM, которая по специальным критериям последовательно ищет в каждой лидерной области белоккодирующего структурного гена (обычно это фермент для синтеза некоторой аминокислоты или аминоксил-тРНК-синтетаза) варианты возможного расположения гена лидерного пептида. Далее, к каждому из этих вариантов она ищет возможное расположение U-богатого участка и терминаторной шпильки, а к ним – варианты возможных антитерминаторных шпилек. Затем найденные таким образом лидерные области, потенциально содержащие структуры классической аттенуаторной регуляции, выравниваются одной из стандартных программ локального выравнивания. Если получается хорошего качества локальное множественное выравнивание нуклеотидных последовательностей – участков лидерных областей генов, то оно рассматривается как предсказание классической аттенуаторной регуляции у соответствующих

видов перед соответствующими генами. Итак, упомянутый метод [1] заключается в отборе сначала отдельных лидерных областей программой LLLM, а затем в поиске хорошего локального множественного выравнивания отобранных областей, при котором элементы структуры классической аттенуаторной регуляции выравниваются между собой. Очевидно, программа LLLM имеет чисто комбинаторный характер и не описывает динамику (т.е. сам процесс) классической аттенуаторной регуляции.

В нашей работе [13] предложена модель процесса аттенуации, которая реализована быстрой компьютерной программой RNAmode1, подробно описанной нами [22] и доступной на сайте lab6.iitp.ru для свободного пользования. Эта модель применяется к отдельным лидерным последовательностям и, в этом отношении, подобна программе LLLM, но по своему содержанию радикально отличается от нее. Для настоящей работы в модель и программу RNAmode1 внесены принципиальные изменения, указанные в разделе “Метод”. Итак, метод, который использован в этой работе, состоит в применении вместо LLLM программы RNAmode1. В тех случаях, когда получался достаточно четкий (информативный) образец, применяли и поиск регуляции по образцу.

Для поиска LEU- и LEU1-регуляций в отдельных лидерных областях применяли аналогичный подход с той разницей, что вместо модели и программы, описанных ранее [13, 22] применяли их адаптированные варианты, названные соответственно LEUmodel и LEU1model, которые кратко описаны в разделе “Метод”, но пока не представлены на нашем сайте.

После применения программ RNAmode1, LEUmodel и LEU1model к отдельным лидерным областям, как и раньше, строилось множественное выравнивание отобранных участков лидерных областей, потенциально содержащих структуру аттенуаторной регуляции. Но, опять-таки, в отличие от предыдущей работы [1], здесь использовали другую программу множественного выравнивания. А именно, разработанную нами программу множественного выравнивания по филогенетическому дереву, которое может быть и политомическим [23]. Эта программа кратко описана ниже в разделе “Метод” и представлена на сайте lab6.iitp.ru для свободного пользования.

Для исключения недопредсказаний LEU- и LEU1-регуляций мы провели их поиск по образцу вдоль всего генома у нескольких бактерий, а не только в лидерных областях указанных выше генов. Этот поиск не привел к новым результатам по сравнению с теми, которые указаны в предыдущей работе [2]: гомологичный LEU-элементу фрагмент внутри гена транспозазы у *B. longum*.

МЕТОД

Бактериальные геномы находили с помощью программы BLAST [23]. Филогенетические деревья строили с помощью метода ближайших соседей по

программе CLUSTALW [24]. Наш метод состоит из первого и второго алгоритмов, излагаемых ниже и доступных для свободного пользования на сайте <http://lab6.iitp.ru>. Первый алгоритм является усовершенствованием алгоритма из нашей работы [13], второй алгоритм впервые публикуется, кроме короткой аннотации [23].

Первый алгоритм. В упомянутой работе [13] предложена модель классической аттенуаторной регуляции, которая основана на совместном описании движения рибосомы по гену лидерного пептида, движения РНК-полимеразы вплоть до конца U-участка и начала структурного гена, а также формирования вторичной структуры на участке между 3'-краем рибосомы и 5'-краем полимеразы (этот участок назван *окном*), вслед за которым идет короткий участок связывания дуплекса ДНК • мРНК в канале, образованном субъединицами полимеразы. Модель описывает вероятность $p(c)$ преждевременной термации полимеразы в зависимости от концентрации с аминоксил-тРНК, которая вычисляется на основе учета влияния концентрации c на скорость движения рибосомы по регуляторным кодонам. Для исходной, данной нуклеотидной последовательности и при данном c модель строит цепочку в некотором смысле типичных вторичных структур в каждом окне. Эта цепочка зависит от кинетики самих вторичных структур и от положения окна, которое возникает в результате движения рибосомы и полимеразы. Такую цепочку вместе с последовательными положениями рибосомы и полимеразы назовем *траекторией моделирования*.

Ключевая идея усовершенствования модели из нашей работы [13] состоит в том, что не следует выделять “паузную” шпильку, антитерминатор и терминатор с их специфическими способами действия на полимеразу, а нужно описать, как *каждая спираль* текущей вторичной структуры в окне *единообразно взаимодействует со всеми другими спиральями, с рибосомой и с полимеразой*, приводя, в зависимости от значения концентрации c , к большему или меньшему замедлению движения полимеразы или (как предельный случай замедления) к ее преждевременной термации. Эта идея соответствует нашим наблюдениям над моделью [13]: аттенуаторная регуляция сопровождается образованием ансамбля терминаторов и, чаще, образованием ансамбля антитерминаторов; антитерминатор может иметь плечи, достигающие до U-богатого участка; он может образовывать псевдоузел с терминатором и т.д. Ситуация усложняется тем, что при наличии гена лидерного пептида с регуляторными кодонами в нем U-богатый участок может включать только 1–2 остатка урацила на заметном расстоянии друг от друга (“слабый” полиурациловый тракт), или участка вообще может не быть. Иногда это означает все-таки, что имеется классическая аттенуаторная регуляция, а иногда предполагает некоторую другую регуляцию с участием гена лидерного пептида и регуляторных кодонов в нем. Ко вто-

рому случаю относятся секвестр-аттенуаторная, а также LEU- и LEU1-регуляции. Отметим, что изменение лидерной последовательности ровно в одной позиции может приводить к исчезновению аттенуаторной регуляции, которая до того была ярко выраженной [7]. Это явление воспроизводится в нашей модели.

Анализ траекторий моделирования в нашей модели позволяет приписать некоторым шпилькам роль терминатора, а другим — антитерминатора, что позволяет построить адекватное регуляции, если она имеется, локальное множественное выравнивание потенциальных регуляторных областей, используя второй алгоритм (см. ниже). Сравнение первого алгоритма с алгоритмом LLLM из предыдущей работы [1] приведено в разделе “Введение”.

Кроме того, по сравнению с предыдущей работой [13], в модель классической аттенуаторной регуляции внесены следующие изменения. Разрешено образование псевдоузлов у всех вторичных структур в каждом текущем окне. При этом энергия вторичной структуры вычисляется на основе усовершенствования метода, представленного ранее [25], и учитывается температура, типичная для среды обитания бактерии, на основе усовершенствования метода из другой работы [26]. В модели учитывается также образование РНК-триплексов, которые делают некоторые критически важные для регуляции спирали вторичной структуры более устойчивыми, понижая их энергию. Соответствующая компьютерная программа RNAmode1 доступна для свободного пользования на сайте lab6.iitp.ru. В режиме разрешенных псевдоузлов зависимость $p(c)$ качественно сохраняется для всех последовательностей из работы [13], где она наблюдалась в режиме запрещенных псевдоузлов, при этом во многих случаях эта зависимость улучшилась.

Нами проведено массовое тестирование этой модели и соответственно программы RNAmode1. А именно, сравнивали результаты, которые она выдает, с доступными результатами биоинформатического или экспериментального предсказания классической аттенуаторной регуляции у бактерий; в том числе, для случаев, когда предсказывается отсутствие такой регуляции. Результат сравнения состоит в том, что с высокой достоверностью предсказание о наличии регуляции совпадает с тем, что модель выдает возрастающую зависимость вероятности $p(c)$ от концентрации c на некотором участке значений концентрации. И наоборот: предсказание об отсутствии регуляции совпадает с тем что $p(c)$, полученная в модели, представляет собой постоянную или убывающую величину. Эти результаты сравнения доступны на вэб-странице <http://lab6.iitp.ru/mamode1> (файл Mass). Моделирование велось при одних и тех же универсальных параметрах, указанных Рубановым и Любецким [21], и также на этой вэб-странице.

Для случая LEU-регуляции разработаны модель и компьютерная программа LEUmode1, которая

позволяет определять вероятность $q(c)$ не-инициации трансляции структурного гена в зависимости от концентрации c лейцил-тРНК. Эта вероятность определяется как доля от времени жизни мРНК, в течение которой предполагаемый сайт связывания рибосомы перекрыт какой-то спиралью. Наибольший вклад в такое перекрытие вносит черенок LEU-псевдоузла, показанный на рис. 3. В случаях предсказания LEU-регуляции эта модель выдавала убывающую зависимость $q(c)$, а в случаях отсутствия такого предсказания — существенно иную, не убывающую зависимость $q(c)$.

Для случая LEU1-регуляции разработаны варианты предыдущей модели и компьютерная программа LEU1model. Программа позволяет определить вероятность $q_1(c)$ наступления регуляторного события в зависимости от концентрации c лейцил-тРНК. Эта вероятность определяется как доля от времени жизни мРНК, в течение которого существуют гипоспираль от двух фиксированных спиралей, которые наиболее консервативны и составляют LEU1-псевдоузел, показанный на рис. 2.

Второй алгоритм. Кратко опишем алгоритм и программу LmalTree локального множественного выравнивания по данному филогенетическому дереву, которую мы использовали для локального выравнивания последовательностей, потенциально содержащих соответствующие регуляторные структуры (сами последовательности отбирались на основе первого алгоритма) [22]. Как и программа RNAmodel, программа LmalTree доступна для свободного пользования на сайте <http://lab6/iitp.ru>.

Изложим вначале алгоритм для случая бинарного филогенетического дерева видов. Каждой вершине этого данного дерева видов приписывается одна последовательность распределений частот нуклеотидов (далее: “последовательность распределений”), т.е. последовательность, члены которой — векторы длины четыре. В таком векторе выписываются частоты четырех нуклеотидов в следующем порядке: А, С, G, T. Итак, в каждой позиции любой последовательности распределений находится вектор, т.е. характерное для этой позиции распределение частот нуклеотидов. Листьям дерева приписываются нуклеотидные последовательности, и они же рассматриваются как последовательности распределений, состоящие из векторов, у которых одна частота равна единице, а остальные частоты равны нулю.

Сначала предложенный алгоритм ведет обработку вершин исходного дерева от листьев к корню (“прямой ход” алгоритма), последовательно приписывая им последовательности распределений. Если два таких распределения уже приписаны двум “сыновьям” некоторой вершины, то последовательности *выравниваются*, и затем их “отцу” приписывается полусумма выровненных распределений в сыновьях. Это выравнивание строится обычным алгоритмом парного выравнивания, но функционал оценки качества выравнивания (score) вычисляется

особым образом. Вместо фиксированных величин приза/штрафа за совпадение/несовпадение нуклеотидных букв, как это делается при выравнивании обычных нуклеотидных последовательностей, здесь для каждой позиции j выравнивания, не содержащей пробела, вычисляется своя оценка a_j , например, в виде скалярного квадрата разности двух векторов $a_j = 1 - \sum_{i=1}^4 w_i(x_i - y_i)^2$, где w_i — веса разных букв, в сумме равные 1. При выравнивании вставку пробела в одну из последовательностей распределений в некоторой ее позиции отвечает нулевое распределение; в этом случае вместо a_j используется значение обычной функции штрафа, у которой величина штрафа a'_j за позицию убывает с ростом длины серии подряд идущих пробелов. Окончательная оценка парного выравнивания, максимум которой ищется, определяется как сумма величин a_j или a'_j по всем позициям выравнивания.

После получения парного выравнивания строится последовательность распределений для отца, которая в каждой из ее позиций равна Z полусумме распределений X и Y в той же позиции у выровненных сыновей, т.е. $Z = \frac{1}{2}(X + Y)$.

Если таким образом уже построена последовательность распределений для корня дерева, то выполняется “обратный ход” алгоритма, при котором сначала пробелы, вставленные в последовательности у прикорневых вершин, опускаются в низ дерева вплоть до его листьев. Затем то же самое делается с пробелами, вставленными при выравнивании на предыдущем уровне дерева (третьем, считая от корня), и так далее вплоть до уровня, расположенного непосредственно над листьями дерева.

Последовательности, вообще говоря, с многочисленными пробелами, полученные таким образом в листьях дерева, и есть *искомое множественное выравнивание* и результат работы алгоритма, при данном бинарном дереве.

Однако часто бинарное дерево видов неизвестно, а вместо него известно филогенетическое дерево, которое содержит одну или более политомических вершин (“политомическое” дерево). Тогда алгоритм разрешает все такие небинарные (политомические) вершины путем добавления в исходное небинарное дерево промежуточных вершин. Таким образом строятся все варианты бинарных деревьев, согласованных с исходно данным небинарным деревом в части отношения “ x — потомок y ”. Эта серия бинарных деревьев должна порождаться без повторений, так как число вариантов и без того велико: как известно, каждая политомическая вершина с n сыновьями допускает $\frac{(2n-3)!}{2^{n-2}(n-2)!}$ топологически различных вариантов разрешения. При нескольких по-

Потенциальная аттенуаторная регуляция у бактерий

Типы и классы	Гены						
α-протеобактерии	<i>ilvB, I</i>	<i>trpE</i>	<i>hisS</i>	<i>pheST</i>	<i>thrA</i>	<i>leuA</i>	<i>leuA</i>
β-протеобактерии	<i>ilvB</i>	<i>trpE</i>		<i>pheA</i>	<i>thrS</i>	<i>leuA</i>	<i>leuA, ilvB</i>
γ-протеобактерии	<i>ilvB, G</i>	<i>trpE</i>	<i>hisG</i>	<i>pheA, S</i>	<i>thrA</i>	<i>leuA</i>	
δ-протеобактерии	<i>ilvB</i>	<i>trpS</i>			<i>thrA, S</i>	<i>leuA</i>	
Актинобактерии	<i>ilvB, I, D</i>	<i>trpE, S, BE, BA</i>					<i>leuA</i>
Bacteroidetes/Chlorobi	<i>ilvD</i>	<i>trpE</i>	<i>hisG</i>				
Firmicutes	<i>ilvD, lysQ</i>	<i>trpB</i>	<i>hisZ</i>				
Thermotogae		<i>trpE</i>	<i>hisS</i>				
Chloroflexi	<i>ilvD</i>						

Примечание. Пустые клетки означают, что такой регуляции нет. Все столбцы, кроме последнего, содержат названия генов, подверженных классической аттенуаторной регуляции, а последний столбец – LEU1-регуляции (у протеобактерий) и LEU-регуляции (у актинобактерий).

литомиях разрешение каждой из них выполняется независимо, и тогда общее число деревьев – это произведение чисел вариантов по всем небинарным вершинам.

Множественные выравнивания, полученные при различных вариантах бинарного разрешения политомического дерева, сравниваются между собой на основе вычисления величины $(N_a + N_s)b + \sum_{i=1}^{N_s} (b+s)(l_i - 1) + N_b c$, где N_a – число абсолютно консервативных (т.е. ровно с одной буквой в столбце) позиций в выравнивании; N_s – число непрерывных абсолютно консервативных участков длиной 2 и более нуклеотидов (l_i обозначает длину i -го такого участка); N_b – число “почти” абсолютно консервативных (только с одной отличающейся буквой в столбце) позиций выравнивания; b , c и s – величины призов, являющиеся параметрами алгоритма.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Нашим методом обнаружилось потенциальные аттенуаторные структуры у видов из таксономических групп α-, β-, γ-, δ-протеобактерий, у протеобактерии *Magnetococcus* sp. MC-1, у Actinobacteria, у Bacteroidetes/Chlorobi, у Firmicutes и Thermotogae и Chloroflexi (см. сводную табл. 1). Этим методом не обнаружили аттенуаторной структуры ни у одного вида из таксонов Chlamydiae, Cyanobacteria, Mollicutes, ε-протеобактерий и Spirochaetales. Ее нет также в хлоропластах водорослей, которые имеют гены синтеза аминокислот.

Аттенуаторная регуляция наиболее часто встречается у протеобактерий и у актинобактерий, но отдельные ее случаи предсказываются у представителей других таксономических групп: у типов Firmicutes, Bacteroidetes/Chlorobi, Thermotogae и Chloroflexi. Частоты встречаемости этой регуляции разных генов значительно разнятся.

Как видно из табл. 1, наибольшее число видов и разнообразие генов биосинтеза аминокислот и аминоацил-тРНК, регулируемых именно с помощью классической аттенуаторной регуляции, наблюдается у γ-протеобактерий. Неклассическая аттенуаторная регуляция гена *leuA* наблюдается у актинобактерий и α-протеобактерий.

Далее излагаются и обсуждаются результаты отдельно по каждой аминокислоте.

Фенилаланин

Классическая аттенуаторная структура, зависящая от фенилаланил-тРНК, наблюдается исключительно у α-, β- и γ-протеобактерий и регулирует транскрипцию гена *pheA* синтеза фенилаланина и оперона *pheST*. Регуляция оперона *pheST* предсказана у небольшого числа α-протеобактерий из порядка Rhodobacterales и у протеобактерии *Magnetococcus* sp. MC-1. Она предсказана для *pheA* у рода *Bordetella* (класс β-протеобактерий) и для *pheA* и *pheS* у большинства представителей γ-протеобактерий (рис. 4 и рис. 5²). У *Psychromonas* sp. CNPT3 аттенуаторная регуляция найдена только перед геном *pheST*. В целом регуляция *pheST* наблюдается реже, чем регуляция *pheA*. По дереву видно, что эта регуляция у α-, β- и γ-протеобактерий распределена мозаично, т.е. она есть у единичных представителей больших таксономических групп. У представителей порядков Enterobacterales и Alteromonadales аттенуаторная регуляция имеется перед *pheA* и *pheST*, а у представителей порядков Aeromonadales и Vibrionales – только перед геном *pheA*. По дереву видов видно, что эта регуляция у α-, β- и γ-протеобактерий распределена мозаично, т.е. существенно неравномерно в разных таксономических группах.

²Рисунки 3–37, подписи к ним и табл. 2 расположены на вэб-странице http://www.molecbio.com/downloads/2010/1/supp_lopat_rus.pdf

На дереве регуляторных областей, построенном по соответствующему выравниванию, у γ -протеобактерий (рис. 6) структуры перед геном *pheST* образуют кладу, за единственным исключением у *Psychromonas* sp. CNPT3. Это позволяет предположить, что аттенуаторная регуляция *pheST* возникла у общего предка γ -протеобактерий и эволюционировала независимо от регуляции *pheA*. Регуляторная область перед *pheST* у *Psychromonas* sp. CNPT3 значительно отличается от таковой у других видов. Это позволяет сделать вывод, что аттенуаторная регуляция возникла здесь независимо от регуляции *pheST* у большинства других видов. Интересная особенность: эта область у *Psychromonas* sp. CNPT3 близка к регуляторной области перед *pheA* у *Alteromonas macleodii* из того же порядка Alteromonadales. Здесь можно предполагать возможность горизонтального переноса или дубликации с последующими потерями соответствующих регуляторных структур.

Треонин и изолейцин

Классическая аттенуаторная структура, связанная с треонином и изолейцином, обнаружена только у протеобактерий. У α -протеобактерий она регулирует только ген *thrA*, связанный с синтезом аминокислот, и только у *Rhodobacterales bacterium* (рис. 7a). У γ -протеобактерий она регулирует только тот же *thrA*, но зато у многих видов (рис. 8). У β -протеобактерий — только ген *thrS*, кодирующий треонил-ТРНК-синтетазу, у ряда видов (рис. 9).

Из δ -протеобактерий эта структура предсказана для *thrA* у *Mucococcus xanthus*. “Слабое” предсказание имеется и для двух паралогов *thrS* у *Bdellovibrio bacteriovorus*, и для одного из двух паралогов у *Mucococcus xanthus* и *Stigmatella aurantica* (рис. 7b). Отметим, что согласно нашей модели [13], при низкой концентрации треонина у *Mucococcus xanthus* перед *thrA* складывается антитерминатор, который мешает формироваться шпильке терминатора и при этом, из-за большой длины плеч, он располагается вблизи U-участка. При средней концентрации рибосома разрушает часть антитерминатора, но терминатор все-таки еще не может формироваться. При высокой концентрации треонина рибосома проходит до стоп-кодона гена лидерного пептида и полностью разрушает антитерминатор, в результате чего формируется короткая спираль, соответствующая классическому терминатору. Такое явление наблюдается и в других случаях моделирования. Поскольку у δ -протеобактерий найдено несколько случаев регуляции и перед неортологичными генами, можно думать, что здесь произошли горизонтальные переносы регуляции без переносов соответствующих генов (рис. 10–11). Отметим, что некоторые паралоги этих генов, не имеющие такой регуляции, также перенесены: например, *thrS* у *Anaeromyxobacter dehalogenans* и *thrA* у *Desulfobivrio desulfuricans* и *Desulfatibacillum alkenivorans*.

У δ -протеобактерий *Syntrophobacter fumaroxidans* и *Anaeromyxobacter dehalogenans* найдены пары паралогов *thrS*, которые не имеют классической аттенуаторной регуляции.

Триптофан

У актинобактерий классическая аттенуаторная структура предсказана в случае оперонов, содержащих ген *trpE*, а именно — у стрептомицетов и у трех из четырех видов коринебактерий (*C. diphtheriae*, *C. glutamicum*, *C. efficiens*) (рис. 12a). Поиск по образцу указал на эту регуляцию перед геном *trpS*. Согласно нашему методу, ее нет у *C. jeikeium*.

Регуляторные области *C. diphtheriae* выравниваются перед двумя паралогами гена *trpB*, один из которых входит в оперон *trpBA*, кодирующий триптофансинтазу, а другой — в оперон *trpBEGDC*, кодирующий антранилатсинтазу, однако качество выравнивания перед *trpBA* значительно ниже. Моделирование также указывает на отсутствие регуляции перед *trpBA*. Ее нет и перед *trpBA* у других актинобактерий. Полезно регулировать гены, кодирующие субъединицы антранилатсинтазы (первый шаг на пути синтеза триптофана), и не обязательно регулировать гены, кодирующие субъединицы триптофансинтазы (последний шаг на этом пути). Это может быть причиной значительной дивергенции аттенуаторной структуры перед *trpBA*. Вероятно, у *Corynebacterium diphtheriae* произошла серия перестроек хромосомы, при которых между регуляторной областью и самим геном *trpE* появился ген *trpB*, и затем произошла дубликация *trpB* вместе с его регуляторной областью, в результате которой паралог *trpB*, вместе с регуляторной областью, оказался позиционно сцепленным с геном *trpA*. После этого в этой регуляторной области перед *trpBA* начали накапливаться мутации.

На дереве регуляторных областей близки ветви, ведущие к *trpS* у *Streptomyces avermitilis* и к *trpE* у *Streptomyces* spp. (рис. 13). Поэтому можно предполагать, что регуляция перед *trpS* у *S. avermitilis* произошла из более древней регуляции перед *trpE*. Регуляция у видов *Nocardia farcinica* и *Saccharopolyspora erythraea* перед *trpS* имеет древнее происхождение (рис. 13).

Предсказано большое число случаев классической аттенуаторной регуляции *trpE* у α -, β -, γ -протеобактерий, у типа Bacteroidetes и у двух представителей *Thermotoga* spp. (Thermotogae) (рис. 14–17), а также у гена *trpS* у δ -протеобактерий вместе с соответствующим ансамблем спиралей (рис. 18), и для гена *trpB* у двух представителей *Bacillus* spp. (Firmicutes) (рис. 12b).

Из двух паралогов *trpE* *Vibrio fischeri* регуляция имеется только у второго из них. У γ -протеобактерий *Pseudoalteromonas tunicata* и *Alteromonadales bacterium* произошла дубликация регуляторной области гена *trpE* (рис. 16).

Все указанные выше случаи, кроме *trpS1* у δ -протеобактерий, относятся к типу регуляции по Яновскому.

Гистидин

Предсказана классическая аттенуаторная структура, связанная с гистидином, у α -протеобактерий, где она регулирует транскрипцию гена *hisS*, у γ -протеобактерий — гена *hisG*, у Firmicutes — гена *hisZ*, у Bacteroidetes — гена *hisG* и у Thermotogae — гена *hisS* (рис. 19–22). У γ -протеобактерий и у некоторых Firmicutes наблюдается регуляция на основе цепи спиралей с участием триплексов, а у Bacteroidetes, Thermotogae и у рода *Listeria* — по Яновскому.

Перед геном *hisG* у многих γ -протеобактерий (включая представителей семейств Enterobacteriales, Pasteurellales, Vibrionales, Alteromonadales, Aeromonadales) найден триплекс, который, согласно нашему методу, участвует в формировании котерминатора. Третье плечо триплекса включает много остатков урацила, что обеспечивает высокую его стабильность независимо от кислотности цитоплазмы. В большинстве случаев здесь складывается триплекс из триад Pu-Pu-Pu. Однако у *Alteromonadales bacterium* и *Pseudalteromonas haloplanktis* складывается смешанный триплекс CUGU*GAGG-CCUC, содержащий триады Pu-Pu-Pu и Pu-Pu-Pu (рис. 20). Напомним, что третье плечо триплекса везде отделяется звездочкой. Регуляторные области на рис. 20 и рис. 22 хорошо выравниваются, что указывает на их высокую консервативность у многих γ -протеобактерий и бактериоидов.

Перед геном *hisZ* у нескольких бацилл (*Bacillus cereus*, *B. thuringiensis*, *B. anthracis*, *B. weihenstephanensis*), возможно, расположен слабый цитидил-гуаниловый триплекс Pu-Pu-Pu, который участвует в формировании котерминатора. Перед геном *hisZ* у *Clostridium difficile* котерминатор складывается с участием триплекса Pu-Pu-Pu AAG*AAG-CUU (рис. 21).

Перед геном *lysQ*, кодирующим пермеазу, у *Lactococcus lactis* (Firmicutes) наблюдается классическая аттенуаторная регуляция с гистидиновыми регуляторными кодонами и цепью спиралей, в которой котерминатор поддерживается триплексом Pu-Pu-Pu AGA*AGA-UCU (рис. 23). Возможно, среди Firmicutes только у *Lactococcus lactis* эта пермеаза сменила специфичность и стала гистидиновым транспортом: здесь наблюдается регуляция, зависящая от концентрации гистидина, в то время как у других фирмикут перед ортологичным геном отсутствует даже ген лидерного пептида.

Хотя перед *hisZ* у бацилл и кластридий наблюдается цепь спиралей с триплексами, у представителей филогенетически близкого рода *Listeria* предсказывается классическая аттенуаторная регуляция по Яновскому.

У фирмикут триплексы обнаружены лишь в небольших таксономических группах. Это свидетельствует о том, что триплексы у них возникли в ходе эволюции недавно. Поскольку устойчивость некоторых триплексов зависит от кислотности цитоплазмы или от содержания в ней ионов, можно предположить, что участие триплексов в регуляции определяется сложными обратными связями, которые

согласуют ответ на концентрацию соответствующей аминокислоты с кислотностью цитоплазмы.

При моделировании показана низкая эффективность регуляции без учета триплексов, а при их учете — значительное изменение зависимости частот преждевременной терминации транскрипции от концентрации соответствующей аминокислоты-тРНК. Таким образом, участие триплексов в регуляции у γ -протеобактерий подтверждается как их консервативностью, так и при моделировании. Моделирование на лидерной области перед геном *hisG* у γ -протеобактерий также показывает, что без учета триплексов соответствующие антитерминирующие структуры РНК недостаточно стабильны для обеспечения эффективной регуляции. В случае *Clostridium difficile* и гена *hisZ* только увеличение энергии котерминатора, которое происходит в результате учета триплекса, приводит к отчетливой зависимости экспрессии структурного гена от концентрации гистидил-тРНК, что соответствует предсказанию о наличии регуляции.

Разветвленные аминокислоты

Классическая аттенуаторная структура с ансамблем антитерминаторов предсказана для гена *ilvB* большинства актинобактерий из порядка Actinomycetales, за исключением *Propionibacterium acnes*, *Tropheryma whipplei*, *Clavibacter michiganensis*, *Leifsonia xyli*, *Janibacter* и *Salinispora arenicola*. За пределами Actinomycetales такая регуляция предсказана для *Bifidobacterium adolescentis*, однако у вида *Bifidobacterium longum* из того же рода и у *Rubrobacter xylanophilus* наш метод ее не обнаруживает: у этих бактерий нет даже гена лидерного пептида. Для упомянутого большинства имеется консервативный терминатор транскрипции, но отсутствует консервативный антитерминатор, и его роль берет на себя ансамбль из нескольких спиралей и псевдоузлов (рис. 24).

Аннотировано шесть паралогов гена *ilvB* актинобактерии *Rhodococcus jostii*, из которых два (*ilvB4* и *ilvB5*) предваряются генами лидерного пептида и парой длинных альтернативных шпилек. После спирали, которая, предположительно, играет роль терминатора, U-участок имеется только перед геном *ilvB5*, но не перед *ilvB4*. Здесь можно обсуждать классическую аттенуаторную регуляцию без U-участка; не исключено, что в некоторых из этих случаев регуляция просто нет. Классическая аттенуаторная регуляция по Яновскому у еще одного паралога *ilvB1* деградировала, но некоторые ее элементы сохранились. Регуляторные области у *ilvB4* и *ilvB5* значительно отличаются между собой (рис. 24).

В роли антитерминатора *ilvB* у большинства α -протеобактерий также выступает ансамбль спиралей (рис. 25). У *Rhodobacteriales bacterium* регуляция имеется у обоих паралогов этого гена, показанных на этом рисунке.

Регуляции на уровне транскрипции *ilvB* у β -протеобактерий *Bordetella* spp., по-видимому, нет. Если принять иной иницирующий кодон структурного

гена, чем указанный в аннотации NCBI, а именно кодон, который выравнивается с иницирующими кодонами у *B. bronchiseptica* и *B. parapertussis*, то можно предсказать регуляцию на уровне трансляции, то есть секвестр-аттенуаторную регуляцию. Ее элементы показаны на рис. 26а.

У многих γ -протеобактерий оперон *leuABCD* и гены *ilvA*, *ilvC*, *ilvG* и *ilvB*, часто включенные в различные опероны, находятся под классической аттенуаторной регуляцией по Яновскому (рис. 27). Некоторые из этих случаев были приведены ранее [1].

Дерево регуляторных областей перед генами *ilv* у γ -протеобактерий показывает, что участки регуляции перед *ilvA*, *ilvC* и *ilvB*, как и перед совместно взятыми генами *ilvC* и *ilvB*, образуют четыре клады. Регуляция перед *ilvG* самая древняя и восходит к общему предку γ -протеобактерий (рис. 28).

Перед геном *ilvB* у *Stigmatella aurantiaca* (δ -протеобактерии) найдена некоторая потенциальная структура, похожая на структуру классической аттенуаторной регуляции, которая включает ген лидерного пептида, но обладает рядом особенностей (без U-участка и т.д.) (рис. 26с). То же самое имеет место у представителей порядков Desulfobacterales, Desulfuromonadales, Syntrophobacterales (δ -протеобактерии). Однако соответствующие регуляторные области не консервативны, и регуляция не подтверждается моделированием (рис. 29).

Классическая аттенуаторная регуляция предсказана для гена *ilvD* одной актинобактерии *Corynebacterium efficiens* (рис. 30а).

В регуляторной области гена *ilvD* родов *Staphylococcus* и *Listeria* (Firmicutes) наблюдается цепь спиралей: четыре консервативные спирали с образованием триплекса из триад Pu-Pu-Pu при формировании ко-антитерминатора. Правда, здесь возможен альтернативный антитерминатор, который длиннее и не требует участия триплексов (рис. 31). При моделировании регуляции *ilvD* у *Staphylococcus* и *Listeria* (рис. 32) наблюдается случай, сходный со случаем гена *hisG*, который показывает, что учет энергии триплекса в составе антитерминатора необходим. Это наблюдение, вместе с фактом консервативности, подтверждают необходимость учета триплексов при вычислении энергии вторичной структуры РНК. У *Geobacillus thermodenitrificans* наблюдается классическая регуляция по Яновскому. Такая же регуляция наблюдается у представителей типов Bacteroidetes/Chlorobi и у *Herpetosiphon aurantiacus* (Chloroflexi) (рис. 30б, с).

Для небольшого числа α - и β -протеобактерий предсказана классическая аттенуаторная регуляция гена *leuA* по Яновскому. Среди них вид *Acidovorax* sp. JS42 уникален тем, что два паралога этого гена образуют общий оперон с единой регуляцией. Из β -протеобактерий классическая аттенуаторная структура найдена у представителей таксона Burkholderiales (*Burkholderia xenovorans*, *Burkholderia phytofirmans*, *Burkholderia pseudomallei*, *Delftia acidovorans*, *Polaromonas* sp., *Comamonas testosteroni*, *Methylobium proteiphilum*, *Acidovorax* sp.) (рис. 33а, б).

У γ -протеобактерий Enterobacteriales, Pasteurellaceae, *Shewanella* spp., Vibrionaceae, Xanthomonadaceae имеется структура классической аттенуации гена *leuA* по Яновскому; здесь присутствует консервативный терминатор большой длины. Некоторые из этих видов приведены ранее [1]. Ген *leuA* у этих γ -протеобактерий, кроме Xanthomonadaceae, не имеет паралогов и входит в длинный оперон *leuABCD*, а у Xanthomonadaceae — в оперон *ilvCGM-tdcB-leuA*. Таким образом, в одном случае участок регуляции находится непосредственно перед *leuA*, а в другом случае — далеко от него, и тогда *ilv* и *leuA* регулируются совместно. Аналогичные случаи наблюдаются и у некоторых α -протеобактерий. Ген *leuA* у Pseudomonadaceae также не имеет паралогов, но не входит в состав полицистронного оперона и не имеет гена лидерного пептида (рис. 34).

Классическая аттенуаторная регуляция по Яновскому перед генами *leuA* и *ilvB* предсказана у небольшого числа δ -протеобактерий. А именно, ген *leuA* регулируется у *Desulfotalea psychrophila*, *Stigmatella aurantiaca*, *Syntrophobacter fumaroxidans*, *Syntrophus aciditrophicus* и *Plesiocystis pacifica* (рис. 33с). По дереву видов можно заключить, что у этого гена δ -протеобактерий мозаичное филогенетическое распределение регуляции.

Поиск по образцу указал на возможность существования участка классической аттенуаторной регуляции по Яновскому перед геном *leuS* у *S. avermitilis* и *S. coelicolor* (актинобактерии). Однако этот участок регуляции генов аминоксил-тРНК-синтетаз не подтверждается моделированием.

LEU-регуляция гена *leuA* у актинобактерий. Нами значительно расширен список участков LEU-регуляции, приведенный ранее [2]: ее содержит большинство актиномицетов перед этим геном (рис. 35). LEU-регуляция обнаружена у большинства актинобактерий из порядка Actinomycetales за исключением двух семейств из таксона Propionibacterineae (*Propionibacterium acnes* и *Nocardioides* sp. JS614) и семейства Cellulomonadaceae (*Tropheryma whipplei*). LEU-регуляция не обнаружена в других порядках актинобактерий: у Rubrobacterales (*Rubrobacter xylanophilus*), Coriobacteriales (*Atopobium minutum*), Biofidobacteriales (*Bifidobacterium longum* и *B. adolescentis*).

Можно предположить, что LEU-регуляция перед геном *leuA* возникла у последнего общего предка Actinomycetales и затем исчезла у общего предка таксона Propionibacterineae, хотя сам ген *leuA* хорошо сохранился в таксоне. Сам ген *leuA* у *Tropheryma whipplei* (Cellulomonadaceae, Actinomycetales) отсутствует.

Согласно данным Крастера и соавт. (Craster) [27], замена лейциновых кодонов в предполагаемом гене лидерного пептида перед геном *leuA* у *Streptomyces coelicolor* не влияет на интенсивность транскрипции. Это позволяет предположить, что весьма консервативный у Actinomycetales LEU-элемент связан с регуляцией экспрессии на уровне инициации трансляции. В пользу такой гипотезы говорит и то, что 3'-плечо черенка LEU-элемента перекрывает сайт связывания рибосомы и, таким образом, черенок пре-

пятствует инициации трансляции гена *leuA*. 5'-Плечо черенка находится внутри гена лидерного пептида вблизи лейциновых кодонов, поэтому стабильность черенка зависит от скорости трансляции этого гена.

LEU1-регуляция гена *leuA* у протеобактерий. У многих α - и некоторых β -протеобактерий (Burkholderiales) перед этим геном найден ген лидерного пептида, содержащий последовательность лейциновых кодонов, который, вместо вторичной структуры РНК, характерной для классической аттенуаторной регуляции или LEU-регуляции, сопровождается псевдоузлом, который мы назвали LEU1-псевдоузлом (рис. 2). Псевдоузел LEU1 вместе с геном лидерного пептида и регуляторными (в данном случае лейциновыми) кодонами в нем назван LEU1-регуляцией (рис. 36, 37).

Участок LEU1-регуляции перед этим геном найден у α -протеобактерий (Rhizobiales: *Agrobacterium tumefaciens*, *Aurantimonas* sp. SI85-9A1, *Brucella* spp., *Fulvamarina pelagi*, *Mesorhizobium* spp., *Rhizobium* spp., *Sinorhizobium* spp.; Rhodospirillales: *Magnetospirillum* spp.; Rhodobacterales: *Dinoroseobacter shibae*, *Jannaschia* sp. CCS1, *Loktanella vestfoldensis*, *Oceanicola* spp., *Rhodobacterales bacterium* HTCC2654, *Rhodobacter* spp., *Roseobacter denitrificans*, *Roseovarius* spp., *Sulfobacter* spp., *Alpha proteobacterium* HTCC2255) и β -протеобактерий (Burkholderiales: *Bordetella* spp., *Ralstonia eutropha*, *Ralstonia metallidurans*, *Janthinobacterium* sp., *Herminiimonas arsenicoxydans*). У Rhizobiales псевдоузел содержит дополнительную неконсервативную спираль в петле консервативного псевдоузла.

Во всех видах, содержащих несколько паралога этого гена, LEU1-псевдоузел найден не более, чем перед одним из паралогов. Для Rhizobiales такой паралог — ближайший гомолог гена *leuA* *Agrobacterium tumefaciens* (белок NP_355220.1). В случае Rhodospirillales и Rhodobacterales — это ближайший гомолог *leuA* у *Roseobacter denitrificans* (белок YP_681546.1); в случае Burkholderiales — это ближайший гомолог *leuA* у *Bordetella pertussis* Tohama I (белок NP_879030.1).

LEU1-структуры у Burkholderiales сходны со структурой из Rhizobiales и сравнительно далеки от структур у Rhodobacterales как по нуклеотидному составу спиралей, так и по расположению псевдоузла относительно стоп-кодона лидерного пептида. Во всех случаях лидерный пептид имеет длинный и совершенно не консервативный N-концевой участок.

Гомолог *leuA*, перед которым найден участок LEU1-регуляции, у α - и β -протеобактерий позиционно не связан с другими генами *leu* или *ilv*. У *Magnetospirillum* spp. (Rhodospirillales) имеется несколько паралогов *leuA*. Тот из них, который регулируется LEU1, по аминокислотной последовательности значительно дальше от гена *leuA* в кишечной палочке, чем другой паралог у *Magnetospirillum*, позиционно связанный с генами *ilv*. Таким образом, ген *leuA* у Rhizobiaceae, вероятно, является ксенологом.

Итак, LEU1-регуляция связана с геном семейства изопротилмалат-, гомоцитрат-, цитрамалат-синтаз, который сходен, но отличим от общего представителя гена *leuA* у большинства протеобакте-

рий. С другой стороны, у Rhizobiaceae имеется только один гомолог гена *leuA*, что указывает на его функциональную значимость; а все гомологи гена *leuA* у протеобактерий, предваряемые участком потенциальной LEU1-регуляцией, являются ортологами друг друга.

По крайней мере, два паралога *leuA*, кодирующих 2-изопротилмалатсинтазу у *Magnetospirillum*, имеются. Распределение аттенуаторной регуляции показывает, что один из этих паралогов унаследован от общего предка протеобактерий, а другой — LEU1-регулируемый ксенолог. Унаследованный от общего предка паралог гена *leuA*, как можно предположить, регулируется по способу классической аттенуации совместно с генами *ilv*, при этом они составляют один совместный оперон. Эта ситуация похожа на классическую аттенуацию у семейства Xanthomonadaceae из γ -протеобактерий.

LEU1- и LEU-псевдоузлы имеют некоторое структурное сходство, хотя даже на этом уровне отличаются: LEU-псевдоузел имеет черенок, а LEU1-псевдоузел не имеет его. На количественном, так сказать, уровне они отличаются еще больше: нуклеотидный состав псевдоузлов не имеет общих мотивов, число лейциновых кодонов у LEU1 меняется в пределах 4–8, а у LEU — около 3, длины плеч у левой спирали LEU-псевдоузла больше, чем у LEU1-псевдоузла. Положения стоп-кодонов внутри этих двух псевдоузлов аналогичны. У Rhodospirillales и у большинства Rhodobacterales псевдоузел расположен на небольшом расстоянии от начала гена *leuA*, а любой участок, подходящий на роль сайта связывания рибосомы перед этим геном, не комплементарен области лейциновых кодонов, как это имеет место в случае LEU-регуляции у актинобактерий. Напротив, у Rhizobiales и Burkholderiales расстояние от гена лидерного пептида до начала гена *leuA* велико и включает длинный неконсервативный участок.

Если пытаться найти у LEU1-структуры черенок, аналогичный черенку у LEU-структуры, это не удастся по следующей причине. Можно найти спираль, у которой левое плечо перекрывает участок лейциновых кодонов с длиной плеча 5–8 н. (обычно 6 н.), и правое плечо перекрывает предполагаемый сайт связывания рибосомы перед *leuA*, но у такой спирали внешняя петля будет содержать от 140 н. (у *Brucella*) до 217 н. (у *Bordetella*). Столь большая длина петли этой спирали, включающей кроме того длинный неконсервативный участок (ниже псевдоузла), сделала бы ее нестабильной, и поэтому она, по-видимому, не имеет регуляторного значения. Поэтому нет оснований предполагать, что LEU1-регуляция у этих протеобактерий функционирует на уровне трансляции гена *leuA*, как это имеет место в случае LEU-регуляции. С другой стороны, в указанных видах отсутствует и терминатор транскрипции на всем участке между лидерным пептидом и геном *leuA*.

Поэтому “слабая” гипотеза состоит в том, что спирали псевдоузла препятствуют связыванию некоторого белка, участвующего в терминации транскрипции или в разрушении мРНК *leuA*. А именно,

если лейцина мало, то рибосома редко разрушает псевдоузел, и белок не может связать мРНК. Если лейцина много, то рибосомы, транслирующие лидерный пептид, постоянно разворачивают псевдоузел, и в этой области связывается белок. Рибосома закрывает 10–12 нуклеотидов перед собой. При классической аттенуаторной регуляции расстояние между регуляторными кодонами и антитерминаторной шпилькой обычно сходно с расстоянием между регуляторными кодонами и LEU1-псевдоузлом в LEU1-регуляции. Поэтому в случае, если имеется псевдоузел, рибосома, остановившись на первых лейциновых кодонах, оказывается у начала псевдоузла, но почти не разрушает его. С другой стороны, дойдя до стоп-кодона лидерного пептида, рибосома еще не перекрывает 3'-плеча правой спирали псевдоузла и не освобождает связавшиеся там белки.

Авторы благодарны М.С. Гельфанду за ценное обсуждение результатов этой работы. Авторы благодарны студентке ФББ МГУ И. Глотовой, которая провела часть компьютерных вычислений, использованных в этой работе, и студенту Д. Колобкову, который приготовил файл Mass. Авторы глубоко благодарны В.В. Гречко за содержательное редактирование текста статьи.

СПИСОК ЛИТЕРАТУРЫ

- Vitreschak A.G., Lyubetskaya E.V., Shirshin M.A., Gelfand M.S., Lyubetsky V.A. 2004. Attenuation regulation of amino acid biosynthetic operons in proteobacteria: comparative genomics analysis. *FEMS Microbiol. Lett.* **234**, 357–370.
- Seliverstov A.V., Putzer H., Gelfand M.S., Lyubetsky V.A. 2005. Comparative analysis of RNA regulatory elements of amino acid metabolism genes in Actinobacteria. *BMC Microbiol.* **5**, 54.
- Grundy F.J., Henkin T.M. 2003. The T box and S box transcription termination control systems. *Front Biosci.* **8**, d20–31.
- Grundy F.J., Henkin T.M. 2004. Regulation of gene expression by effectors that bind to RNA. *Curr. Opin. Microbiol.* **7**, 126–131.
- Mandal M., Breaker R.R. 2004. Gene regulation by riboswitches. *Nature Rev. Mol. Cell. Biol.* **5**, 451–463.
- Vitreschak A.G., Rodionov D.A., Mironov A.A., Gelfand M.S. 2004. Riboswitches: the oldest mechanism for the regulation of gene expression? *Trends Genet.* **20**, 44–50.
- Das A., Crawford I.P., Yanofsky C. 1982. Regulation of tryptophan operon expression by attenuation in cell-free extracts of *Escherichia coli*. *J. Biol. Chem.* **15**, 8795–8798.
- Henkin T.M., Yanofsky C. 2002. Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions. *Bioessays.* **24**, 700–707.
- Yanofsky C. 2004. The different roles of tryptophan transfer RNA in regulating *trp* operon expression in *E. coli* versus *B. subtilis*. *Trends Genet.* **20**, 367–374.
- Burillo S., Luque I., Fuentes I., Contreras A. 2004. Interactions between the nitrogen signal transduction protein PII and N-Acetyl glutamate kinase in organisms that perform oxygenic photosynthesis. *J. Bacteriol.* **186**, 3346–3354.
- Heery D.M., Dunican L.K. 1993. Cloning of the *trp* gene cluster from a tryptophan-hyperproducing strain of *Corynebacterium glutamicum*: Identification of a mutation in the *trp* leader sequence. *Appl. Environ. Microbiol.* **59**, 791–799.
- Lin C., Pradkar A.S., Vining L.C. 1998. Regulation of an antranilate synthase gene in *Streptomyces venezuelae* by *trp* attenuator. *Microbiol.* **144**, 1971–1980.
- Lyubetsky V.A., Pirogov S.A., Rubanov L.I., Seliverstov A.V. 2007. Modeling classic attenuation regulation of gene expression in bacteria. *J. Bioinform. Comput. Biol.* **5**, 155–180.
- Chastain M., Tinoco I.Jr. 1992. Poly(rA) binds poly(rG) poly(rC) to form a triple helix. *Nucleic Acids Res.* **20**, 315–318.
- Кнорре Д.Г., Мызина С.Д. 2000. *Биологическая химия*. М.: Наука.
- Semerad C.L., Maher L.J. 1994. Exclusion of RNA strands from a purine motif triple helix. *Nucleic Acids Res.* **22**, 5321–5325.
- Carmona P., Molina M. 2002. Binding of oligonucleotides to a viral hairpin forming RNA triplexes with parallel G*GC triplets. *Nucleic Acids Res.* **30**, 1333–1337.
- Klinck R., Guitteta E., Liquier J., Taillandier E., Gouyetteb C., Huynh-Dinhby T. 1994. Spectroscopic evidence for an intramolecular RNA triple helix. *FEBS Lett.* **355**, 297–300.
- Holland J.A., Hoffman D.W. 1996. Structural features and stability of an RNA triple helix in solution. *Nucleic Acids Res.* **24**, 2841–2848.
- Benson D.A., Karsch-Mizrachi I., Lipman D.J., Ostell J., Wheeler D.L. 2008. GenBank. *Nucleic Acids Res.* **36**, 25–30.
- Интернет-сайт <http://www.ncbi.nlm.nih.gov>
- Rubanov L.I., Lyubetsky V.A. 2007. RNAmol Web Server: modeling classic attenuation in bacteria. *In Silico Biol.* **7**, 285–308.
- Любецкая Е.В., Горбунов К.Ю. 2008. Алгоритмы реконструкции эволюции регуляторных сигналов. *Труды 51-й научной конференции МФТИ “Современные проблемы фундаментальных и прикладных наук”*, Часть I. М.: МФТИ. 142–145.
- Craster H.L., Potter C.A., Baumberg S. 1999. End-product control of branched-chain amino acid biosynthesis genes in *Streptomyces coelicolor* A3 (2): paradoxical relationships between DNA sequence and regulatory phenotype. *Microbiol.* **145**, 2375–2384.
- Cummings L., Riley L., Black L., Souvorov A., Resenchuk S., Dondoshansky I., Tatusova T. 2002. Genomic BLAST: custom-defined virtual databases for complete and unfinished genomes. *FEMS Microbiol. Lett.* **216**, 133–138.
- Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F., Higgins D.G. 1997. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882.
- Isambert H., Siqgia E.D. 2000. Modeling RNA folding paths with pseudoknots: Application to hepatitis delta virus ribozyme. *Proc. Natl. Acad. Sci. USA.* **97**, 6515–6520.
- Tianbing Xia, John Santa Lucia, Jr, Mark E. Burkard, Ryszard Kierzek, Susan J. Schroeder, Xiaoqi Jiao, Christopher Cox, Douglas H. Turner. 1998. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry.* **37**, 14719–14735.