

30-я КОНФЕРЕНЦИЯ МОЛОДЫХ УЧЕНЫХ
И СПЕЦИАЛИСТОВ ИППИ РАН

Информационные технологии и системы ИТиС'07

СБОРНИК ТРУДОВ КОНФЕРЕНЦИИ

г. Звенигород, 18-21 сентября 2007 г.

ББК 32.81
И 741

Информационные Технологии и Системы (ИТиС'07)
Москва: ИППИ РАН, 2007. – 370с.

Издание содержит труды 30-й конференции молодых ученых и специалистов Института проблем передачи информации им. А.А.Харкевича Российской академии наук (ИППИ РАН) «Информационные технологии и системы» (ИТиС'07). Конференция явилась продолжением серии традиционных конференций, организованных Советом молодых ученых и специалистов ИППИ РАН в предшествующие годы.

Основная цель Конференции ИТиС'07 – 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.

Все включенные в данный сборник работы опубликованы в том виде, в котором они были представлены авторами, среди которых молодые ученые и специалисты ИППИ РАН, ИМБ РАН, ИМГ РАН, ИОГен РАН, ИПУ РАН, МГТУ им. Н.Э. Баумана, МГУ им. М.В. Ломоносова, МИФИ (ГУ), МТУСИ, МФТИ (ГУ), НИИ биомедицинской химии РАМН, УФСИН России по Псковской области, ФГУП ГосНИИГенетика и Центрального НИИ эпидемиологии Роспотребнадзора.

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 Целевой программы Президиума РАН «Поддержка молодых ученых 2007 г.».

ISBN 978-5-7834-0193-0

© Институт проблем передачи информации им. А.А. Харкевича
Российской академии наук, 2007

Псевдоузлы и РНК-триплексы в модели классической аттенуаторной регуляции

И. Глотова¹, Л. Рубанов², А. Селиверстов², В. Любецкий^{1,2}

¹Московский государственный университет им. М.В. Ломоносова,

²Институт проблем передачи информации РАН

igbox@mail.ru, lyubetsk@iitp.ru

Аннотация

Нами расширена модель классической аттенуаторной регуляции, предложенная в [1], и также ее компьютерная реализация, доступная на сервере <http://lab6.iitp.ru/rnamodel>, за счет разрешения микросостояний с псевдоузлами и специального учета РНК-триплексов.

1. Введение

Геномы организмов содержат как гены, так и для многих генов и групп генов регуляторные элементы. Роль последних состоит в том, чтобы включать и выключать (синонимы: регулировать, экспрессировать) работу групп генов (оперонов). Известно некоторое число типов регуляции экспрессии генов. Наиболее изученным типом является белок-ДНКовая регуляция. Ее механизм состоит в том, что некоторая белковая молекула, называемая фактором транскрипции, связывается или не связывается с определенным участком, называемым сайтом фактора транскрипции, что и служит сигналом к экспрессии гена или оперона. Сигнал может быть позитивным (тогда происходит активация гена) или негативным (тогда происходит репрессия гена). Другой тип регуляции называется классической аттенуаторной регуляцией и основан на согласованном движении рибосомы и РНК-полимеразы по области, которую можно назвать сайтом этой регуляции. Геном имеет фиксированное направление, которое обозначается «от 5'-конца к 3'-концу». В области между 3'-концом рибосомы и 5'-началом полимеразы, которая называется «окном», складывается та или иная вторичная структура, которая принимает существенное участие в классической аттенуаторной регуляции: рибосома разрушает эту структуру, а она, в свою очередь, задерживает полимеразу, а иногда и обрывает связь полимеразы с комплексом ДНК/мРНК, тем самым прекращая

процесс транскрипции. Модель классической аттенуаторной регуляции была предложена в [1]. В этой модели не были разрешены микросостояния с псевдоузлами, а при вычислении энергии антитерминаторов не учитывался ряд сопутствующих явлений, как, например, образование РНК-триплексов, которые приводят к понижению энергии антитерминатора и, следовательно, к его большей устойчивости. В работе рассмотрены такие две возможности: в пункте 2 - микросостояний с псевдоузлами и в пункте 3 - понижение энергии у некоторых шпилек. Для этого были разработаны новые специальные программные средства на языке С с Web-интерфейсом, представленные на сервере <http://lab6.iitp.ru/rnamodel>, и проведен обширный счет, некоторые результаты которого указаны ниже, в пункте 4. Пункт 5 содержит предварительные выводы из результатов счета. Авторы предполагают написать развернутую статью по этому материалу.

2. Модель классической аттенуаторной регуляции с псевдоузлами у микро-состояний

Ниже излагаются только отличия текущей модели от той, которая приведена в [1].

Для определения вклада потенциальных псевдо-узлов в свободную энергию спирали мы использовали подход, описанный в [2]. Пусть две гипо-спирали γ и γ' образуют псевдоузел; примем следующую нумерацию участков псевдоузла: $5' - \bar{n}_1 s_2 \bar{n}_2 s_3 \bar{n}_1 s_1 \bar{n}_2 - 3'$, где \bar{n}_1, \bar{n}_1 - плечи гипоспиралей γ и \bar{n}_2, \bar{n}_2 - плечи гипоспиралей γ' . Весь псевдоузел заключен между точками А и D'. Энергия стекинга вычисляется так, что края гипоспиралей около 5' и 3' концов псевдоузла являются свободным краем спирали, а противоположные концы гипоспиралей обращены к выпячиваниям, внешним, одно- или двусторонним в зависимости от того, примыкают ли они к другим гипоспиралам (добавочным или

же совпадающим со второй спиралью псевдоузла). Только этот вклад зависит от температуры. Как обычно, в качестве G_{helix} используется безразмерное частное от деления энергии одного моля на RT .

Для элементарного псевдоузла с длинами n_1 и n_2 гипоспиралей и длинами s_1 , s_2 и s_3 промежутков между гипоспиральями энергия петли (или, тоже самое, энтропия) равна

$$G_{loop}(\gamma, \gamma') = -\ln \left[\alpha^2 \frac{\exp(-A_1 f_1^2 - A_2 f_2^2)}{D^{3/2}} \times \frac{\exp(2A_3 f_1 f_2) - \exp(-2A_3 f_1 f_2)}{4A_3 f_1 f_2} \right],$$

где $\alpha = 0.0068$, $D = s_1 s_2 + s_1 s_3 + s_2 s_3$,

$$f_{1,2} = \sqrt{16 \sin^2(\pi n_{1,2} / 11) + 25(n_{1,2} / 11)^2},$$

$$A_1 = \frac{3(s_1 + s_3)}{5D}, \quad A_2 = \frac{3(s_2 + s_3)}{5D}, \quad A_3 = \frac{3s_3}{5D}.$$

Если на участках s_1 , s_2 и s_3 имеются вставки других гипоспиралей, то от них при вычислении энергии петли берутся только два нуклеотида А и D. Если вставлен псевдоузел, то от него берутся только два крайних нуклеотида А и D'. Таким образом длины s_1 , s_2 и s_3 должны удовлетворять условию

$$\begin{aligned} 3 \leq n_1 \leq s_1 + s_3 + 2 \\ 3 \leq n_2 \leq s_2 + s_3 + 2 \end{aligned} \quad (*)$$

при котором гипоспираль выбираются максимально продолженными: насколько позволяет это условие, размер текущего окна и совместимость с другими гипоспиральями из рассматриваемого микросостояния.

При расчете замедления полимеразы под влиянием сложившейся вторичной структуры принято, что псевдоузел не оказывает на нее влияния, т.е. его вклад со всеми вложенными в него шпильками и узлами в «силу» F в формулах (20-21) из [1] считается нулевым, и в (21) при расчете τ_i считается, что узел представлен только двумя крайними нуклеотидами А и D'.

3. Модель классической аттенуаторной регуляции с возможностью формирoвания РНК-триплексов

В процессе укладки вторичной структуры мРНК в окне наряду с образованием типичных шпилек возможны взаимодействия, которые

уменьшают энергии некоторых шпилек, например, формирoвание РНК-триплексов, как результат возникно-вения так называемых хугстинских взаимодействий между сравнительно коротким участком цепи мРНК и близко расположенным черенком соответствующей спирали. Образование РНК-триплекса приводит к понижению энергии спирали. Если эта спираль является антитерминатором, его устой-чивость повышается, что существенно влияет на взаимодействие антитерминатора и терминатора. Понижение энергии для стандартного мРНК-триплекса может быть заключено в пределах примерно от -2 до -12 ккал/моль.

Поэтому для последовательностей мРНК бактерий, у которых в ранее предложенной нами модели зависимость вероятности $p(c)$ терминации от концентрации аминоацил-тРНК, вероятно, была не вполне адекватной (например, не имела моно-тонный график), мы имитировали образование мРНК-триплекса. Для этого энергия антитер-минатора в соответствующей структуре мРНК [3] избирательно понижалась на величину δ , так что

$$\Delta G_{helix}' = \Delta G_{helix} + \delta,$$

где ΔG_{helix} - свободная энергия спирали без добавки δ , где δ менялось от -2 до -12 ккал/моль.

4. Особенности программной реализации и результаты счета

С самого начала программной реализации модели классической аттенуаторной регуляции в программе существенно использовались допущения, которые прямо противоречат нынешнему расширению модели. Наиболее существенными из этих допущений были следующие три: (1) при нахождении с помощью отношений на спиральях, предложенных в [4], окрестности текущего макро-состояния для моделирования очередного перехода не допускалось само построение диаграмм-кандидатов, в которых имеются псевдоузлы; (2) рассматривались только микросостояния с максимально продолженными (в пределах теку-щего) окна гипоспиральями; и, наконец, (3) в любом микросостоянии плечи гипоспиралей не пересе-каются, т.е. каждый нуклеотид может участвовать не более чем в одной гипоспирали.

Допущение (1), препятствующее появлению псевдоузла, недостаточно было просто отменить, т.е. убрать из программы некоторое условие, при выполнении которого запрещенные диаграммы отбрасываются. Напомним, что с помощью набора раз вычисленных бинарных некоммутативных отношений [4] между всеми

спиралями заданной последовательности алгоритм не строит всевозможные соседние диаграммы, отбрасывая из них недопустимые (что заняло бы слишком много времени), а конструирует преимущественно допустимые (непустые и не содержащие псевдоузлов) диаграммы, за очень редкими исключениями. Мы дополнили систему отношений L-R-X еще одним отношением Y, которое описывает допустимые фрагменты диаграммы макросостояния вида $(i)_j$ для всевозможных пар номеров гипоспиралей $i \neq j$ (как легко видеть, именно отсутствие фрагментов такого вида не позволяет построить диаграмму с элементарным псевдоузлом).

От использования в программе допущения (2), сильно сокращающего перебор, также пришлось отказаться, поскольку приоритет имеют условия (*), так что возможны псевдоузлы, у которых одна или обе образующие его гипоспираль в принципе могут быть продолжены в данном микросостоянии, но только с нарушением условий (*). В итоге проверка допустимости построенной конфигурации осуществляется в программе с помощью оригинальной перекрестно-рекурсивной структуры, состоящей из двух рекурсивных подпрограмм: одна проверяет допустимость гипоспираль, которая является одним из двух «черенков» псевдоузла, другая - допустимость гипоспираль, являющейся черенком обычной шпильки. Рекурсия обусловлена продолжением проверки допустимости дальше в сторону петли черенка (или петель псевдоузла), а перекрестный вызов – тем, что в каждой из петель может встретиться как новая гипоспираль, так и новый псевдоузел.

В результате описанных изменений модель получила способность перехода в макросостояния с псевдоузлами. Это привело к небольшому замедлению программы по сравнению с режимом, при котором образование псевдоузлов не допускается, по причине большего объема перебора, но в целом не повлияло на ее пригодность для оперативного моделирования, в том числе непосредственно на Web-сервере.

Что касается допущения (3), противоречащего структуре РНК-триплекса, то оно настолько глубоко используется в нынешней программной реализации (в том числе и неявно), что отказ от него фактически означал бы разработку программы заново. Поэтому мы решили предварительно проверить эффективность такого добавления, прибегнув к имитации оказываемого триплексом влияния. В случае успеха это будет реализовано в автоматическом режиме в будущих версиях программы.

Ниже описываются результаты, полученные с помощью усовершенствованной модели, и

обсуждается эффект от учета возможностей образования псевдоузлов или РНК-триплексов.

4.1 Моделирование с разрешенными псевдоузлами

Прежде всего, в режиме разрешенных псевдоузлов зависимость $p(c)$ качественно сохранилась для всех последовательностей из [1], где она наблюдалась в режиме запрещенных псевдоузлов, табл. 1. Зависимость $p(c)$ приобрела монотонный характер в случае двух новых последовательностей, для которых она не являлась монотонной в режиме запрещенных псевдоузлов, табл. 1, рис. 1 и рис. 2.

Для большинства последовательностей псевдо-узлы, которые охватывают участок спирали анти-терминатора и таким образом мешают его образованию, разрушаются в конце траектории моделирования.

4.2 Моделирование с имитацией образования мРНК-триплексов

При имитации образования мРНК триплексов, которые включают черенок анти-терминатора, для последовательности, соответствующей гену *Sv_trpE*, для которой без этого зависимость $p(c)$ не имела монотонного характера, были получены следующие результаты. При добавлении к энергии анти-терминатора, рис. 3, добавки $\delta = -7$ ккал/моль, значение $p(0)$ значительно уменьшилось и получилась стандартная зависимость $p(c)$, табл.2, рис.4. При этом на интервале концентраций [0.02, 0.15] наблюдается строго монотонный рост $p(c)$. Можно предположить, что регуляция транскрипции у *S. venezuelae* происходит как раз на этом интервале значений концентрации, рис.4.

Таблица 1. Вероятности терминации $p(c)$ в интервале значений с концентрации от 0 до 0.5 в режиме разрешенных псевдоузлов. Здесь $p(c)_{kp}$ - вероятность терминации, полученная в режиме разрешенных псевдоузлов, $p(c)_0$ - вероятность терминации, полученная в режиме запрещенных псевдоузлов. Число псевдоузлов, возникших на одной траектории, и затем усредненное по всем траекториям, указано в строке "Число псевдоузлов". L75_trp и L132_trp – мутантные формы гена Ec_trp.

с		0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
I. Последовательности со стандартной зависимостью $p(c)$ от с в режиме запрещенных узлов												
Ec_trp	$p(c)_{kp}$	0.19	0.3	0.45	0.73	0.61	0.68	0.61	0.7	0.67	0.69	0.72
	Число псевдоузлов	20	11	8	3	3	2	1	1	1	0.4	0.4
Vc_trp	$p(c)_{kp}$	0.04	0.1	0.34	0.51	0.62	0.75	0.71	0.68	0.67	0.76	0.72
	Число псевдоузлов	6	7	7	7	5	3	3	4	3	2	1
Rle_trp	$p(c)_{kp}$	0.16	0.26	0.3	0.42	0.51	0.58	0.5	0.6	0.52	0.57	0.63
	Число псевдоузлов	158	76	56	30	26	16	15	8	7	4	3
Rp_trp	$p(c)_{kp}$	0	0.13	0.22	0.32	0.35	0.36	0.35	0.42	0.33	0.38	0.42
	Число псевдоузлов	5	6	8	10	12	12	11	11	13	12	13
Sm_trp	$p(c)_{kp}$	0.03	0.09	0.16	0.37	0.31	0.37	0.4	0.36	0.47	0.45	0.34
	Число псевдоузлов	17	14	10	10	9	6	4	4	3	3	2
Ec_his	$p(c)_{kp}$	0.22	0.29	0.43	0.52	0.49	0.54	0.56	0.59	0.53	0.51	0.55
	Число псевдоузлов	4	3	9	9	10	9	8	5	6	7	5
L75_trp	$p(c)_{kp}$	0.71	0.71	0.75	0.71	0.67	0.63	0.67	0.7	0.7	0.73	0.71
	Число псевдоузлов	6	4	3	2	1	1	0.5	0.3	0.2	0.4	0.1
L132_trp	$p(c)_{kp}$	0	0.08	0.16	0.16	0.14	0.15	0.15	0.22	0.24	0.18	0.23
	Число псевдоузлов	17	10	8	4	3	2	1	1	0.5	0.2	0.1
II. Последовательности с нестандартной зависимостью $p(c)$ от с в режиме запрещенных псевдоузлов												
Sc_trpE	$p(c)_0$	0.33	0.32	0.39	0.32	0.41	0.34	0.31	0.40	0.38	0.32	0.35
	$p(c)_{kp}$	0.27	0.29	0.31	0.37	0.36	0.47	0.29	0.4	0.4	0.37	0.43
	Число псевдоузлов	7	8	8	7	7	6	5	6	6	4	5
Ce_trpEG	$p(c)_0$	0.15	0.06	0.19	0.17	0.24	0.22	0.24	0.34	0.28	0.23	0.35
	$p(c)_{kp}$	0.12	0.13	0.2	0.22	0.26	0.18	0.26	0.27	0.24	0.31	0.3
	Число псевдоузлов	33	34	55	57	58	49	53	49	49	51	49

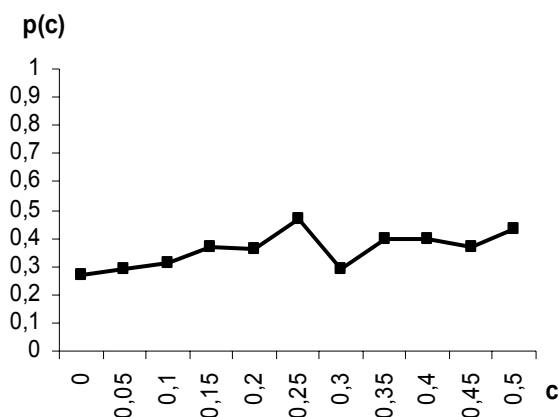


Рис. 1 Вероятность терминции $p(c)$ для гена *Sc_trpE* в интервале концентраций от 0 до 0.5 в режиме разрешенных псевдоузлов.

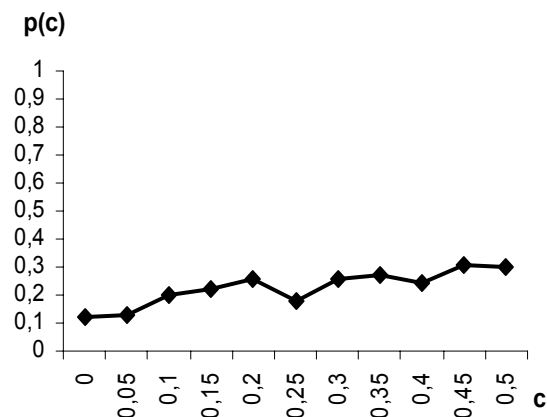


Рис. 2 Вероятность терминции $p(c)$ для гена *Ce_trpEG* в интервале концентраций от 0 до 0.5 в режиме разрешенных псевдоузлов.

G=-123.002 -79 -82 82 79 -105 105

atgttcgcgcaactcgatccagaactgggtggtggaccgctcaccGGCGGCCcactTGATCgcgcggtacacgGATCAcacgcacaGGCCGCC
cGAGGggcggCCTTtctcg

Рис. 3 Предполагаемый мРНК-триплекс в последовательности мРНК *S. venezuelae* (ген *Sv_trpE*). Показана структура мРНК в момент антитерминации и положение антитерминатора (синий цвет), у которого энергия уменьшилась в связи с образованием РНК-триплекса, что привело к уменьшению вероятности терминции. Участок, предположительно связывающийся с левой половиной черенка антитерминатора, подчеркнут.

Таблица 2. Вероятность терминции $p(c)$ для *Sv_trpE* в интервале концентраций *trp*-тРНК от 0 до 0.5 без механизма образования РНК-триплекса и при его наличии.

c	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
$\delta = 0$	0.52	0.63	0.75	0.60	0.70	0.68	0.68	0.61	0.69	0.74	0.67
$\delta = -7$	0.06	0.11	0.24	0.42	0.49	0.58	0.63	0.59	0.72	0.63	0.70

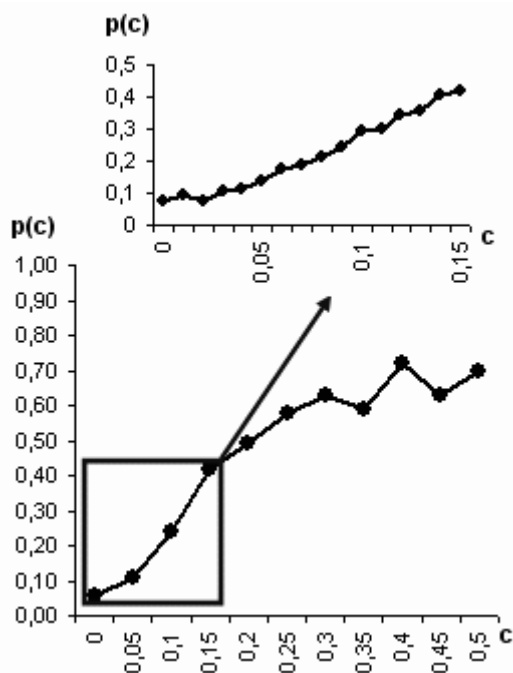


Рис. 4 Вероятность терминции $p(c)$ для Sv_trpE в интервале концентраций trp -тРНК от 0 до 0.5 при $\delta = -7$ ккал/моль. На верхнем графике указано изменение $p(c)$ на участке графика $[0, 0.15]$.

5. Выводы

Модель классической аттенуаторной регуляции экспрессии генов у бактерий расширена за счет разрешения псевдоузлов при образовании микро- и макросостояний, и также

за счет имитации РНК-триплексов. Это расширение компьютерно реализовано на языке С с Web-интерфейсом и представлено на сервере <http://lab6.iitp.ru/rnamodel>. При счете на биологических последовательностях результаты, представленные в нашей предыдущей публикации [1], качественно не изменились. А для ряда других последовательностей вероятность терминции в зависимости от концентрации amino-ацил-тРНК приобрела стандартный характер.

Литература

- [1] V.A. Lyubetsky, S.A. Pirogov, L.I. Rubanov, A.V. Seliverstov, "Modeling classic attenuation regulation of gene expression in bacteria", *Journal of Bioinformatics and Computational Biology*, vol.5, no.1, 2007, pp. 155-180.
- [2] Herve Isambert and Eric D. Siggia, "Modeling RNA folding paths with Pseudoknots: Application to hepatic delta virus ribozyme", *PNAS*, vol.97, no.12, 2007, pp. 6515-6520.
- [3] Alexander V. Seliverstov, Harald Putzer, Mikhail S. Gelfand and Vassily A. Lyubetsky, "Comparative analysis of RNA regulatory elements of amino acid metabolism genes in Actinobacteria", *BMC Microbiology*, vol.5, no.54, 1-14.
- [4] Lev Rubanov and Vassily Lyubetsky, "RNAmodel Web Server: Modeling Classic Attenuation in Bacteria", *In Silico Biology*, 2007 (in print)