

## Вычисление эффективности регуляции биосинтеза триптофана у бактерий на основе модели классической аттенюации.

В.А. Любецкий, А.В. Селиверстов

Институт проблем передачи информации РАН, 127994, Россия,  
Москва, Большой Каретный переулок, 19, e-mail: lyubetsk@iitp.ru, slvstv@iitp.ru

Поступила в редколлегия 27.03.2006

**Аннотация**—Для ранее опубликованной авторами модели классической аттенюаторной регуляции транскрипции оперонов, кодирующих ферменты биосинтеза триптофана, и гена, кодирующего триптофанил-тРНК синтетазу, приводятся результаты счёта при новых значениях двух параметров, один из которых относится к вычислению энергии петли, а другой к вычислению скорости перехода между микросостояниями. И также для отчасти нового набора бактерий.

### 1. ВВЕДЕНИЕ

Под эффективностью регуляции биосинтеза триптофана у бактерий понимается вероятность  $p(c)$  терминации транскрипции, вычисленная на основе нашей модели классической аттенюации, в зависимости от концентрации  $c$  триптофанил-тРНК. И, следовательно, от концентрации триптофана в клетке или от концентрации триптофанил-тРНК синтетазы. Модель описана в статье [1], но сейчас выбраны значения параметров  $C = 5$  и  $\kappa = 10^3 c^{-1}$  и симметричная формула для константы медленного перехода между макросостояниями. Это привело к более естественным результатам для альфа- и гамма-протеобактерий и актинобактерий.

### 2. МАТЕРИАЛЫ И МЕТОДЫ

Геномы бактерий получены из базы данных ГенБанка (NCBI). В качестве набора последовательностей нами были взяты 5'-нетранслируемые области перед оперонами, включающими ген *trpE* антранилат синтазы, у актинобактерий *Corynebacterium diphtheriae* и *Corynebacterium glutamicum*, у альфа-протеобактерий *Agrobacterium tumefaciens*, *Bradyrhizobium japonicum*, *Rhodospseudomonas palustris*, *Rhizobium leguminosarum*, *Sinorhizobium meliloti* и у гамма-протеобактерий *Escherichia coli* и *Vibrio cholerae*. И также для гена *trpS*, кодирующего триптофанил-тРНК синтетазу, у актинобактерии *Streptomyces avermitilis*. Эти аттенюаторы описаны в статьях [2] и [3].

Отметим некоторые отличия нового подхода вычисления кинетики вторичной структуры РНК. Свободная энергия стекинга  $G_{helix}$  вычисляется как и раньше. Энергия петель микросостояния  $\omega$ , измеряемая в единицах  $RT$ , принимается равной сумме:

$$G_{loop} = \sum_i (1.77 \ln(\ell_i + 1) + B + \frac{C}{\ell_i + 1}),$$

где индекс  $i$  пробегает все гипоспиральи из микросостояния  $\omega$ . Эта формула хорошо согласуется с экспериментальными данными [4] для энергий всех петель при всех  $\ell_i > 2$ , если положить  $B = 6.5$  для концевых петель,  $B = 0$  для двусторонних выпячиваний,  $B = 4$  для односторонних выпячиваний,  $C = 5$  для всех случаев. Случаи  $\ell_i \leq 2$  рассматриваются нами отдельно в

соответствии с таблицами из [4]. А именно, принимаются значения энергий петель: для двустороннего выпячивания 0.8 (при  $\ell_i = 2$ ), для одностороннего выпячивания 6.2 (при  $\ell_i = 1$ ) и 4.5 (при  $\ell_i = 2$ ). Концевые петли с такими длинами в нашей модели исключаются. Переходы между микросостояниями делятся на быстрые и медленные. В модели предполагается, что на множестве  $\Omega$  микросостояний  $\omega$ , между любыми двумя из которых возможен быстрый переход, устанавливается стационарное распределение вероятностей Больцмана-Гиббса:

$$p(\omega) = \frac{\exp(-G_{loop}(\omega) - G_{helix}(\omega))}{\sum_{\omega \in \Omega} \exp(-G_{loop}(\omega) - G_{helix}(\omega))}.$$

Такое множество микросостояний объединяется в одно макросостояние. Для медленных переходов между микросостояниями, константы скоростей переходов между микросостояниями в модели вычисляются по формуле:

$$K(\omega \rightarrow \omega') = \kappa \cdot \exp\left[\frac{(G_{loop}(\omega) + G_{helix}(\omega)) - (G_{loop}(\omega') + G_{helix}(\omega'))}{2}\right]$$

Моделирование с использованием этой формулы позволило точнее, чем при использовании аналогичных формул, ранее предложенных А.А. Мироновым, смоделировать аттенуацию экспрессии генов антранилат синтазы. Напомним, что в том варианте предлагалось различать случаи распада и образования гипоспирали. Был проведён счёт для разных вариантов. В частности, вычисление для *E. coli* по ранее предложенным формулам для  $K$ . В то время как рассматриваемая здесь симметричная формула для  $K$  позволила получить хороший результат. Отметим, что в любом из вариантов выполнен принцип детального равновесия:

$$\frac{K(\omega \rightarrow \omega')}{K(\omega' \rightarrow \omega)} = \exp[G(\omega) - G(\omega')],$$

где  $G(\omega) = G_{loop}(\omega) + G_{helix}(\omega)$  — энергия, приписываемая микросостоянию  $\omega$ . В приведенных ниже расчетах принималось значение константы  $\kappa = 1000c^{-1}$ .

### 3. РЕЗУЛЬТАТЫ И ДИСКУССИЯ

Неизвестные заранее параметры модели подобраны так, что для *E. coli* и *C. diphtheriae*, у которых аттенуаторная регуляция подтверждена экспериментально, при малой концентрации наблюдается рост частоты терминации с увеличением концентрации триптофанил-тРНК, а при больших концентрациях — насыщение и выход на плато. Было исследовано поведение функции  $p(c)$  при различных значениях параметров и выбраны те, при которых у этих оперонов наиболее выражен эффект регуляции транскрипции. Для параметров замедления трансляции шпилькой РНК, описанных в [2], приняты значения  $\delta = 25$ ,  $L_1 = 14.5$ ,  $p_0 = 0.167$ ,  $r_0 = 2$ . Влияние "размеров" рибосомы и полимеразы имеет место, но заметно слабее и одинаковое для всех рассмотренных организмов и генов. Это позволило выбрать для них общие значения  $s_0 = 12$ ,  $s_1 = 5$ . С другой стороны, для гена *trpE* у *S. avermitilis*, *S. coelicolor* и *S. venezuelae* не удалось подобрать такие значения параметров, при которых аттенуация проявляется столь же заметно. В этих случаях расчётная вероятность терминации оказывается высокой даже при малых значениях концентраций триптофана, что объясняется высокой устойчивостью шпильки терминатора. И сравнительно медленно растёт с увеличением концентрации. В случае гена *trpS* у *S. avermitilis* вычисленное изменение вероятности терминации также мало при различных значениях параметров, но монотонно возрастает от 0.06 до 0.30. Получены также разумные зависимости для вероятности  $p(c)$  для триптофанового оперона у актинобактерии *C. glutamicum*, у гамма-протеобактерии *V. cholerae* и у альфа-протеобактерий.

Результаты вычислений приведены в таблице. Для *C. glutamicum* вычисленная вероятность терминации возрастает при росте концентрации триптофана в два раза, но остаётся очень низкой. В то же время для некоторых альфа-протеобактерий предсказанная вероятность терминации изменяется весьма значительно: для *R. palustris* в 48 раз, для *S. meliloti* в 7.6 раза. Также значительный рост вероятности терминации предсказан для *V. cholerae*: в 16.6 раза. Значение константы замыкания, равное  $\kappa = 1000c^{-1}$ , является близким к оптимальному. При меньшем её значении перестройка вторичной структуры почти не происходит, большинство событий в процессе моделирования — это сдвиг рибосомы или полимеразы. Напротив, при увеличении  $\kappa$  число смен вторичной структуры в текущем окне быстро растёт, а зависимость вероятности терминации от концентрации триптофана становится менее выраженной. При значении  $\kappa = 10^6c^{-1}$  вероятность терминации для случая *E. coli* близка к 0.7 независимо от концентрации  $c$ . Напротив, для *V. cholerae* результат моделирования слабо зависит от выбора  $\kappa$ , хорошие результаты получаются и при  $\kappa = 10^6c^{-1}$ .

Вид	Концентрация $c$										
	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
<i>C. diphtheriae</i>	0.34	0.34	0.39	0.46	0.50	0.54	0.53	0.53	0.53	0.52	0.54
<i>C. glutamicum</i>	0.05	0.06	0.08	0.10	0.10	0.09	0.09	0.09	0.10	0.10	0.10
<i>S. avermitilis, trpS</i>	0.06	0.13	0.21	0.26	0.28	0.29	0.30	0.30	0.32	0.32	0.30
<i>A. tumefaciens</i>	0.49	0.50	0.62	0.70	0.74	0.78	0.77	0.78	0.82	0.80	0.79
<i>B. japonicum</i>	0.19	0.20	0.24	0.26	0.28	0.26	0.26	0.27	0.26	0.26	0.26
<i>R. leguminosarum</i>	0.23	0.30	0.42	0.55	0.60	0.65	0.67	0.70	0.71	0.71	0.71
<i>R. palustris</i>	0.01	0.22	0.40	0.48	0.56	0.59	0.60	0.60	0.63	0.61	0.62
<i>S. meliloti</i>	0.07	0.11	0.23	0.37	0.43	0.49	0.48	0.51	0.50	0.53	0.51
<i>E. coli</i>	0.34	0.46	0.54	0.68	0.70	0.70	0.71	0.73	0.75	0.75	0.74
<i>V. cholerae</i>	0.05	0.16	0.39	0.57	0.70	0.74	0.77	0.77	0.80	0.79	0.81

**Таблица.** Указаны вероятности терминации  $p(c)$  в зависимости от концентрации  $c$  триптофан-ил-тРНК для различных видов бактерий.

Модельный счёт на регуляторных областях перед триптофановыми оперонами у *E. coli*, *C. diphtheriae*, *V. cholerae* и у нескольких альфа-протеобактерий приводит к результатам, качественно согласным с экспериментальными данными.

Работа частично поддержана грантом МНТЦ (2766).

### СПИСОК ЛИТЕРАТУРЫ

1. В.А. Любецкий, К.Ю. Горбунов, С.А. Пирогов, Л.И. Рубанов, А.В. Селиверстов. Алгоритм и результаты счёта для модели регуляции экспрессии генов у бактерий на основе формирования вторичных структур РНК. *Информационные процессы*, 2005, 5, 337-366. В.А. Любецкий, А.В. Селиверстов. Некоторые алгоритмы, связанные с конечными группами. *Информационные процессы*, 2003, 3(1), стр. 39-46.
2. A.V. Seliverstov, H. Putzer, M.S. Gelfand, V.A. Lyubetsky. Comparative analysis of RNA regulatory elements of amino acid metabolism genes in Actinobacteria. *BMC Microbiology*, 2005, 5: 54.
3. A.G. Vitreschak, E.V. Lyubetskaya, M.A. Shirshin, M.S. Gelfand, V.A. Lyubetsky. Attenuation regulation of amino acid biosynthetic operons in proteobacteria: comparative genomics analysis. *FEMS Microbiology Letters*, 2004, 234, 357-370.
4. D.H. Mathews, M.D. Disney, J.L. Childs, S.J. Schroeder, M. Zuker, D.H. Turner. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *PNAS*, 2004, 101, 7287-7292.