

## Особенности синтеза цистеина у *Corynebacterium*, *Mycobacterium* и *Propionibacterium*

А.В. Селиверстов, В.А. Любецкий

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

Поступила в редколлегию 2.09.2004

**Аннотация**—В статье представлены результаты компьютерного анализа регуляторных областей генов, связанных с синтезом цистеина в актинобактериях. Обсуждаются особенности оперонной структуры. Предсказан новый тип регуляции транскрипции. Это иллюстрирует возможности ранее созданного алгоритма поиска близких слов, основанного на поиске клики в многодольном графе.

### 1. ВВЕДЕНИЕ

Рассматриваются бактерии: *Corynebacterium diphtheriae*, *Corynebacterium efficiens* YS-314, *Corynebacterium glutamicum* ATCC 13032, *Mycobacterium avium* subsp. paratuberculosis str. k10, *Mycobacterium bovis* subsp. bovis AF2122/97, *Mycobacterium leprae* strain TN, *Mycobacterium tuberculosis* H37Rv, *Propionibacterium acnes* KPA171202.

Один из путей синтеза цистеина из серина вовлекает два фермента [1]. Серин ацетируется по гидроксигруппе при участии *серин ацетилтрансферазы CysE* и затем взаимодействует с сероводородом с образованием цистеина. Последняя реакция катализируется *цистеин синтазой GysK*. Гены *cysK* и *cysE* часто входят в один оперон.

Гены всех рассматриваемых ферментов определены для актинобактерий по гомологии. При этом многие гены имеют по несколько гомологов. Структура оперонов предсказана по ориентации и по расстоянию между генами. Некоторые опероны включают неизвестные гены, обозначенные *X*.

### 2. CORYNEBACTERIUM

У *Corynebacterium* гены *cysK* и *cysE* могут транскрибироваться вместе, но разделены значительным промежутком. Более того, после стоп кодона первого гена *cysK* есть типичный терминатор — *gc*-богатая шпилька, за которой следует *t*-богатый участок. Поэтому вероятность транскрипции второго гена *cysE* меньше, чем первого гена *cysK*. Эти последовательности начинаются со стоп кодона *taa* гена *cysK*. Плечи шпильки терминатора подчеркнуты.

*C. diphtheriae*    taatcaactcaccggttgacacaacgcccctgaccaccgaaacataagtgggtcagggcggttttttc  
*C. efficiens*    taatcagccgcccgatcaccatccgccccttccccgtgccagttcggggaagggcggttgcattc  
*C. glutamicum* taattcttagcgaactggttaaccactcaagctctttgcttgggtgggttttttc

Заметим, что без фермента *CysE* синтез цистеина не начинается. С другой стороны, РНК полимераза может с ненулевой вероятностью продолжить транскрипцию, преодолев терминатор. Возможно, что этот механизм связан с поддержанием отношения концентраций ферментов *CysE* и *CysK*.

### 3. MYCOBACTERIUM И PROPIONIBACTERIUM

У оперонов, связанных с синтезом цистеина, в *Mycobacterium* и *Propionibacterium acnes* обнаружены возможные **лидерные пептиды** с длинным повтором цистеиновых кодонов. Соответствующие

последовательности нуклеотидов для *M. bovis* и *M. tuberculosis* совпадают. Лидерные пептиды сильно консервативны по аминокислотному составу. Они обнаружены ранее созданным алгоритмом поиска близких слов, основанным на поиске клики в многодольном графе. [2]

Непосредственно перед старт кодоном лидерного пептида есть богатые пурином области Шайна–Дальгарно.

Бактерия	Оперон	Перед Л.П.	Лидерный пептид	Стоп
<i>M. avium</i>	<i>X<sub>1</sub>cysKE</i>	tatagtggtgac	MQHRLQPRFAPSRCLVVACCCCCCR	tga
<i>M. bovis</i>	<i>cysK<sub>1</sub>E</i>	tatagtgggccc	MQQAIQLRFILPRRLAVQCCCC	tga
<i>M. leprae</i>	<i>XcysKE</i>	tatagtggaact	MHQSTQPRFVFTRRFVTVDCYCRCC	tga
<i>P. acnes</i>	<i>cysK</i>	ggtcaatcggtt	MTSAMMVCICRCCC	tga

Во всех случаях стоп кодон лидерного пептида tga, что хорошо согласуется со средними частотами стоп кодонов у всех генов этих бактерий. Это позволяет предположить, что рибосома не задерживается на стоп кодоне, но освобождает мРНК сразу после окончания трансляции лидерного пептида.

Бактерия	tga	tag	taa
<i>Mycobacterium avium</i> subsp. paratuberculosis str. k10	63	26	11
<i>Mycobacterium bovis</i> subsp. bovis AF2122/97	54	30	16
<i>Mycobacterium leprae</i>	46	30	24
<i>Mycobacterium tuberculosis</i> H37Rv	55	29	16
<i>Propionibacterium acnes</i> KPA171202	81	7	12

Ниже приведены фрагменты ДНК, окружающие стоп кодон лидерного пептида. Как расстояние от стоп кодона лидерного пептида до старт кодона гена, так и нуклеотидный состав этого участка весьма изменчивы.

Участок перед стоп кодоном почти консервативен и имеет консенсус *yg<sub>1</sub>tg<sub>2</sub>tyg<sub>3</sub>tg<sub>4</sub>*, а после стоп кодона следует слабоконсервативный участок, богатый пиримидином. Возможно, что этот участок играет роль в связывании мРНК с белком **Rho**, вызывающим терминацию транскрипции. Известно [3, 4], что у *Mycobacterium* редко встречаются **Rho**-независимые терминаторы. Поэтому естественно предположить, что участки, с которыми связывается белок **Rho**, должны часто встречаться на концах генов сразу после стоп кодонов. Ниже приведены результаты поиска слов, близких по Хэммингу к слову длины 14 *ttcctggcgtccac*, которое следует за стоп кодоном лидерного пептида в *M. bovis* и является предполагаемым местом посадки белка **Rho**. Рассматривались только слова, расположенные между генами на расстоянии от 0 до 100 нуклеотидов от 3'-конца какого-либо гена, указанного в аннотации. Для сравнения взяты *E. coli*, для которой характерны **Rho**-независимые терминаторы, и *B. subtilis*, у которой концентрация **Rho** в 50 раз ниже, чем у *E. coli* [5].

Расстояние	2	3	4	5
<i>M. bovis</i>	1	26	185	1040
<i>M. leprae</i>	3	15	96	624
<i>E. coli</i>	0	6	115	696
<i>B. subtilis</i>	0	4	35	333

Полученные данные хорошо согласуются с высказанной гипотезой.

Гомологи для генов *rho*, *nusA*, *nusB* и *nusG* у актинобактерий существуют. Ниже дано краткое описание соответствующих белков. [5, 6, 7]

**Rho** (или  $\rho$ -белок) образует незамкнутое кольцо, на которое наматывается мРНК. Место посадки обычно представляет собой с-богатый участок *rut*. Взаимодействуя с мРНК и  $\beta$  и  $\beta'$  субъединицами РНК полимеразы, **Rho** может вызывать терминацию транскрипции. **Rho** образован 6 одинаковыми субъединицами, каждая из которых имеет по два домена. N-концевой домен связывает мРНК, С-

концевой домен взаимодействует с РНК полимеразой. Гомологи в актинобактериях очень близки экспериментально изученному гену *rho*.

**NusG** связывает одновременно РНК полимеразу и белок **Rho**. Повышает эффективность **Rho**-зависимых терминаторов, ограничивая скорость транскрипции. **rho**-зависимая терминация без участия **NusG** наблюдалась *in vitro*, но только при скорости транскрипции не более 50 нуклеотидов в секунду.

**NusA** имеет по несколько доменов для связывания и с мРНК и с  $\alpha$ -субъединицей РНК полимеразы. При этом у *Mycobacterium* у белка **NusA** отсутствует С-концевой домен.

Т.к. **Rho** образован 6 одинаковыми субъединицами, можно ожидать, что область мРНК, связываемая **Rho**, длинная и содержит повтор некоторого слова, с произвольными вставками между словами.

У *Mycobacterium* есть 3-х и 4-х кратные повторы RYYYYY

*M. avium* tgATTTCCgcaaGCCCTctgacgctgtagaaATCCCCgcgctcGCCCTgccccg

*M. bovis* tgATTCCTggcgctccacagcaATTCCTcgcGCTCTTgccccg

*M. leprae* tgATTCCTgacACCTTTtaacGCTCTCagcaaatcattcacGTTCTCgccta

(первые tga — это стоп кодон лидерного пептида).

У *Propionibacterium acnes* такой повтор не найден. Вот участок, начинающийся со стоп кодона

*P. acnes* tgatttagccagcagcgcgagcatgagcggctccgtggat

Предполагаемая регуляция синтеза цистеина у рассматриваемых актинобактерий основана на **Rho**-зависимой терминации транскрипции вблизи стоп кодона лидерного пептида.

При **недостатке** цистеина участок мРНК вокруг стоп кодона лидерного пептида закрыт рибосомой длительное время, за которое РНК полимеразы успевает уйти далеко.

При **избытке** цистеина рибосома быстро завершает трансляцию лидерного пептида. Т.к. именно tga является обычным стоп-кодоном, эффективность отделения рибосомы от мРНК должна быть высокой. При этом открывается как участок мРНК с консенсусом ygytgtygytga, так и следующий за ним с-богатый участок, характерный для **Rho**-зависимого терминатора.

То, что рибосома закрывает не весь, а только часть предполагаемого участка связывания с **Rho** не мешает регуляции, т.к. и рибосома и **Rho** достаточно большие (диаметр **Rho** 150-200 Å). Поэтому он может надёжно связать мРНК только после окончания трансляции лидерного пептида.

Отметим, что известен аналогичный механизм регуляции экспрессии гена фермента, катализирующего деградацию триптофана в кишечной палочке [8]. Участок связывания мРНК с **Rho** и, возможно, с белком **NusA** в этом случае таков: cggcccttgatttgcccttctgtagccatcacc.

В статье [9] рассматривается структура мРНК на конце оперонов у *Helicobacter pylori* и выделен с-богатый участок, окруженный at-богатыми флангами, за которым следует на некотором расстоянии ещё один t-богатый участок (расстояние между серединами участков 15 нуклеотидов). Высказано предположение о роли такой структуры для **Rho**-зависимых терминаторов.

Участок после лидерных пептидов перед геном *cusK* в *Mycobacterium* и *Propionibacterium* так же как и перед геном *tna* в *E. coli* сt-богатый и соответствует (левому) с-богатому участку из *Helicobacter pylori*. При этом второй t-богатый участок отсутствует.

У *M. avium* и *M. leprae* опероны, связанные с синтезом цистеина, включают неизвестные гены. Сходство регуляторных механизмов подтверждает предположение об оперонной структуре.

**Благодарности** М.С. Гельфанду, И. М. Ищукову, А.А. Миронову за обсуждение особенностей терминации транскрипции и структуры оперонов.

Работа частично поддержана грантом МНТЦ #2766.

#### СПИСОК ЛИТЕРАТУРЫ

1. Д.Г.Кнорре, С.Д.Мызина. *Биологическая химия*. М.: Высшая школа, 2000.
2. В.А. Любецкий, А.В. Селиверстов. Некоторые алгоритмы, связанные с конечными группами. *Информационные процессы*, 2003, том 2, № 1, 39–46.

3. Takanori Washio, Junji Sasayama, Masaru Tomita. Analysis of complete genomes suggests that many prokaryotes do not rely on hairpin formation in transcription termination. *Nucleic Acids Research*, 1998, 26, 23, 5456–5463.
4. Shyam Unniraman, Ranjana Prakash, Valakunda Nagaraja. Conserved economics of Transcription termination in eubacteria. *Nucleic Acids Research*, 2002, 30, 3, 675–684.
5. J.P.Richardson. Rho-dependent termination and ATPases in transcript termination. *Biochimica et Biophysica Acta*, June 2002, 1577, 251–260.
6. J.P.Richardson. Structural Organization of Transcription Termination Factor **Rho**. *The Journal of Biological Chemistry*, January 1996, 271, 3, 1251–1254.
7. D.L. Kaplan, M. O’Donne. Rho Factor: Transcription Termination in Four Steps Dispatch. *Current Biology*, September 2003, 13, R714–R716.
8. K.V.Konan, C.Yanofsky. Rho-dependent Transcription Termination in the *tna* Operon of *Escherichia coli*: Roles of the *boxA* Sequence and the *rut* Site. *Journal of Bacteriology*, July 2000, 182, 14, 3981–3988.
9. L.Petersen, A. Krogh. *Modeling of Rho Dependent Transcription Termination Sites in the Bacterium Helicobacter pylori*.
10. R.Landick. *RNA Polymerase structure, mechanism and regulation*. September 29, 2003.
11. I.M.Ishchukov, V.A.Likhoshvai, Yu.G.Matushkin. A new algorithm for recognizing the operon structure of prokaryotes. *Proceedings of the Fourth International Conference on Bioinformatics of Genome Regulation and Structure*, Novosibirsk, July 2004, volume 1, 73–76.