

# Компьютерное предсказание регуляторных сайтов мРНК. Анализ регуляции экспрессии генов биосинтеза аминокислот и генов, кодирующих тРНК-синтетазы в грам-положительных бактериях

А.Г.Витрещак

Институт проблем передачи информации РАН, 101447, Москва, Большой Каретный переулок, 19, Россия  
e-mail: veter@imb.imb.ac.ru Тел. (095) 135-20-41.

Поступила в редколлегию 17.05.2002

**Аннотация**—Сравнительный анализ является мощным орудием предсказания различных вторичных структур РНК. Он использовался для предсказания как структурных, так и регуляторных РНК. В настоящей работе анализируются регуляторные вторичные структуры мРНК (Т-бокс) в Грам-положительных бактериях. Экспериментально известно около десятка Т-бокс сигналов в *Bacillus subtilis*, *Bacillus. stearothermophilus*, *Lactococcus lactis* и *Staphylococcus aureus*. Используя выборку известных сигналов был составлен паттерн, и было проанализировано около сотни геномов при помощи программы RNAPattern, написанной для этих целей. Вторичная структура РНК задавалась как набор параметров, включающих в себя длины спиралей и петель, расстояния между спиральями и консервативные позиции нуклеотидов. В ряде геномов Грам-положительных бактерий были найдены Т-бокс сигналы перед новыми генами, включая предполагаемые аминокислотные транспортеры.

## 1. ВВЕДЕНИЕ

Сравнительный анализ является мощным орудием предсказания различных вторичных структур РНК. Он использовался для предсказания как структурных, так и регуляторных РНК [1],[2]. Различные подходы применялись для предсказания регуляции генов по последовательности РНК [3],[4],[5]. В настоящей работе анализируются регуляторные вторичные структуры мРНК (Т-бокс) в Грам-положительных бактериях. Экспериментально известно около десятка Т-бокс сигналов в *Bacillus subtilis*, *Bacillus. stearothermophilus*, *Lactococcus lactis* и *Staphylococcus aureus*. Используя выборку известных сигналов был сделан паттерн, и было проанализировано около сотни геномов при помощи программы RNAPattern, написанной для этих целей. В ряде геномов были найдены новые Т-бокс сигналы (обсуждается ниже). Предсказание регуляторных вторичных структур мРНК (Т-бокс) представляет интерес с точки зрения описания регуляции экспрессии генов метаболизма аминокислот в различных бактериях. Кроме того, по вторичной структуре мРНК Т-бокс сигнала имеется возможность предсказания аминокислоты индуцирующей регуляцию, что является полезным для выяснения функции генов, кодирующих неизвестные белки. Например, анализ регуляторной вторичной структуры РНК позволяет предсказывать по анти-антикодону Т-бокс сигнала специфичность транспортера по аминокислоте.

## 2. ПОИСК Т-ВОХ СИГНАЛОВ

Механизм регуляции Т-бокс, экспериментально показанный для аминокислотных генов (биосинтетических, аминоацил-тРНК синтетаз и некоторых транспортеров) в Грам-положительных бактериях, является одной из разновидностей аттенюации [6]. Вторичная структура РНК (Т-бокс), расположенная

в лидерной области оперона, может находиться в антитерминирующем или терминирующем состоянии. Незагруженная аминокислотой тРНК соединяется с этой структурой при помощи антикодонной петли и акцепторного конца и стабилизирует антитерминирующее состояние вторичной структуры. Загруженная аминокислотой тРНК не взаимодействует с вторичной структурой и РНК укладывается в терминирующее состояние. Такой механизм регуляции хорошо изучен, и для ряда генов в нескольких Грам-положительных бактериях были экспериментально показаны T-box регуляторные сигналы [7],[8].

Однако предсказание этих сигналов в массовом порядке сделано не было. Неясно, насколько филогенетически распространены T-box сигналы и существуют ли новые гены, относящиеся к биосинтезу или транспорту аминокислот, регулируемые по этому механизму. Для поиска новых T-box сигналов был составлен паттерн на основании выборки известных структур и при помощи программы RNAr-*attem* были найдены новые вторичные структуры T-box. В паттерн были включены антитерминатор, терминатор и наиболее консервативный участок — T-box (рис. 1). В программе вторичная структура РНК задавалась как набор параметров, учитывающих длины спиралей, размеры однонитевых участков и расположение консервативных нуклеотидов. Анализировалось около сотни геномов. Рассмотрим результаты на примере хорошо изученной бактерии *Bacillus subtilis*. У этой бактерии было найдено программой 19 потенциальных T-box сигналов, 8 из них входило в обучающую выборку. 11 экспериментально известных T-box сигналов не входило в обучающую выборку, но эти сигналы были найдены программой. При этом шум отсутствовал, что показывает высокую избирательность паттерна. Длина генома *B. subtilis* составляет  $5 \times 10^6$  нуклеотидов, длина сигнала —  $< 100$  нуклеотид). Тест программы на *B. subtilis* показал применимость поиска T-box сигналов в других бактериях. Кроме того, исходя из сравнительного подхода, нахождение сигналов перед ортологичными генами повышает надежность предсказания.

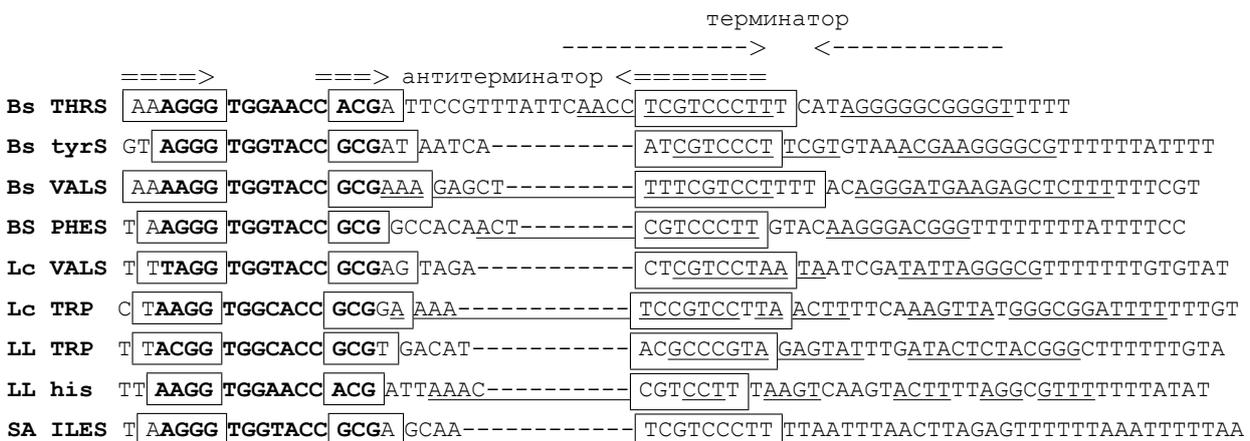


Рис. 1. Выборка известных T-box сайтов. В первой колонке указаны аббревиатуры геномов (BS — *Bacillus subtilis*, LC — *Lactobacillus casei*, LL — *Lactococcus lactis*, SA — *Staphylococcus aureus*), во второй колонке — названия генов, в третьей — участок последовательности мРНК лидерной некодирующей области. Жирным шрифтом показан 14-нуклеотидный консервативный участок (T-box), рамкой и подчёркиванием обозначена вторичная структура РНК антитерминатора и терминатора соответственно. Стрелками указаны комплементарные участки мРНК.

### 3. АНАЛИЗ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

В таблице 1 приведены найденные T-box сигналы перед аминокислотными генами. Подавляющее число T-box регуляторных сигналов было найдено в Грам-положительных бактериях, принадлежащих группе *Bacillus/Clostridium*. Кроме этих бактерий несколько T-box сигналов были найдены в *Chlo-*

*roflexus aurantiacus* (зелёные несерные бактерии) и *Deinococcus radiodurans* (*Thermus/Deinococcus*). В других бактериях не было найдено ни одного регуляторного сайта. Предполагается, что область распространения рассматриваемого регуляторного механизма ограничивается Грам-положительными бактериями или, возможно, подобный механизм регуляции существует у других бактерий, но регуляторный сигнал (вторичная структура РНК и консервативный участок T-box) сильно отличается от сигнала, обнаруженного у Грам-положительных бактерий.

Рассмотрим предсказание регуляции аминокислотных биосинтетических генов и аминоксил-тРНК синтетаз. Найдено 40 генов, регулируемых T-box сигналами в различных бактериях. Регуляция 17 из найденных генов ранее не была экспериментально известна. Наибольшее количество регулируемых генов обнаружено у *Bacillus anthracis* и *Bacillus cereus*. На основании анализа цепочек ортоголических генов в различных бактериях были построены предполагаемые опероны, регулируемые T-box сигналом. Анализ расположения регуляторных сайтов перед аминокислотными оперонами дал интересные результаты. Впервые для регуляторных сигналов T-box было показано сохранение сигнала при различных оперонных перестройках.

бактерии	кол-во T-бок сигналов в геноме
<b>Грам-положительные:</b>	
<i>Bacillus subtilis</i>	19
<i>Bacillus stearothermophilus</i>	14
<i>Bacillus halodurans</i>	18
<i>Bacillus anthracis</i>	42
<i>Bacillus cereus</i>	37
<i>Staphylococcus aureus</i>	12
<i>Enterococcus faecalis</i>	18
<i>Lactococcus lactis</i>	11
<i>Clostridium acetobutylicum</i>	23
<i>Clostridium difficile</i>	22
<i>Streptococcus pyogenes</i>	8
<i>Streptococcus pneumoniae</i>	8
<i>Streptococcus mutans</i>	10
<i>Streptococcus equi</i>	7
<i>Desulfotobacterium hafnience</i>	14
<b>Другие группы:</b>	
<i>Chloroflexus aurantiacus</i>	1
<i>Deinococcus radiodurans</i>	2

Табл. 1. Результаты предсказания T-бок сигналов в различных геномах.

Кроме аминокислотных биосинтетических генов и генов, кодирующих аминоксил-тРНК синтетазы, была предсказана регуляция генов, кодирующих аминокислотные транспортеры. В отличие от аминокислотных биосинтетических генов и генов, кодирующих аминоксил-тРНК-синтетазы, регуляция аминокислотных транспортеров у Грам-положительных бактерий практически не изучена. Недавно [9] экспериментально показан T-бок сигнал перед генами *yczA-ycbK* и *yhaG* в *B. subtilis*. Предсказание регуляции генов, кодирующих аминокислотные транспортеры, интересно тем, что по виду T-бок сигнала (участок анти-антикодон) можно предсказывать специфичность транспортеров по аминокислоте у транспортеров с неизвестной функцией. В настоящей работе были найдены T-бок

сигналы перед 16 генами, кодирующими потенциальные транспортеры с неизвестной функцией. По анти-антикодонам T-box структур была предсказана специфичность предполагаемых транспортеров.

В случае аминокислотных биосинтетических генов и аминоксил-тРНК синтетаз предсказанная регуляция проверяется через функцию гена. Например ген *pah*, кодирующий фенилаланин-4-гидроксилазу (реакция L-фенилаланин  $\rightarrow$  L-тирозин) регулируется тирозиновым T-box сигналом. Уровень экспрессии гена *pah* зависит от концентрации продукта реакции (тирозина), что разумно с биологической точки зрения. В случае же аминокислотных транспортеров предсказываемую регуляцию (T-box) можно подтвердить и через другие механизмы регуляции. Известно, что механизмы регуляции одних и тех же генов в разных бактериях могут отличаться друг от друга. Ниже приводятся несколько случаев пересечения предсказываемой регуляции (T-box) с другими механизмами регуляции.

1. Известно из экспериментальных данных, что ген *yhaG* в *B. subtilis* кодирует триптофановый транспортер и регулируется TRAP репрессором. TRAP репрессия индуцируется высокой концентрацией триптофана внутри бактерии. В настоящей работе были найдены триптофановые T-box сигналы перед геном *yhaG* в *C. acetobutylicum* и *C. difficile*, что указывает на индуцированную триптофаном репрессию этого гена. Таким образом, механизмы регуляции гена *yhaG* в этих бактериях различаются, но принцип регуляции аналогичен. Следует отметить, что ген *yhaG* не имеет ортологов в других бактериях.

*Bacillus subtilis* (TRAP) *yhaG*  
*Clostridium acetobutylicum* (Trp T-box) *yhaG*  
*Clostridium difficile* (Trp T-box) *yhaG*

2. Еще один случай пересечения различных механизмов регуляции можно показать на примере оперона *yqiX-yqiY-yqiZ*, который кодирует гены потенциальной ABC системы транспорта. Ранее перед этим опероном в *C. acetobutylicum* был предсказан сайт для аргининового репрессора *ArgR* [10]. В настоящей работе в *Clostridium difficile* перед опероном *yqiX-yqiY-yqiZ* был найден аргининовый T-box.

*Clostridium acetobutylicum* (ArgR box) *yqiX-yqiY-yqiZ*  
*Clostridium difficile* (Arg T-box) *yqiX-yqiY-yqiZ*

3. Оперон *yusCBA* содержит гены потенциальной ABC системы транспорта. Показано [11], что в *Bacillus subtilis* этот оперон регулируется при помощи механизма S-box. S-box сайты были найдены перед генами, кодирующими ферменты биосинтеза метионина и цистеина, а также перед генами, кодирующими потенциальные метиониновые транспортеры. В настоящей работе в *Enterococcus faecalis* перед опероном *yusCBA* был найден метиониновый T-box. При разности механизмов регуляция *yusCBA* оперона похожа. При избытке метионина экспрессия оперона *yusCBA* репрессируется.

*Bacillus subtilis* (S-box) *yusCBA*  
*Clostridium acetobutylicum* (S-box) *yusCBA*  
*Clostridium difficile* (Met T-box) *yusCBA*

#### 4. ВЫВОДЫ

Предсказаны регуляторные вторичные структуры мРНК (T-box сигналы) перед генами, кодирующими аминоксил-тРНК синтетазы и ферменты биосинтеза аминокислот в Грам-положительных бактериях. Семейство регулируемых T-box сигналами генов было значительно расширено за счет предсказания регуляции у новых генов. Кроме того, предсказана специфичность по аминокислоте у ряда транспортеров с неизвестной функцией. На примере T-box сигналов показано пересечение предсказываемой регуляции с другими известными механизмами регуляции. Кроме того, на примере T-box сигналов показаны оперонные перестройки в различных бактериях при сохранении регуляции, осуществляемой вторичной структурой РНК.

## СПИСОК ЛИТЕРАТУРЫ

1. Waterman M.S. *Mathematical Methods for DNA sequences*. — Boca Raton: CRC Press, 1989.
2. Gelfand M.S., Mironov A.A., Jomantas J., Kozlov Yu.I. and Perumov D.A. A conserved RNA structure element involved in the regulation of bacterial riboflavin biosynthesis genes. *Trends Genet.*, 1999, vol. 15, pp. 439-442.
3. Dandekar T. and Sibbald P.R. Trans-splicing of pre-mRNA is predicted to occur in a wide range of organisms including vertebrates. *Nucleic Acids Res.*, 1990, vol. 18, pp. 4719-4725.
4. Knoop U., Kloska S. and Brennicke A. On the identification of group II introns in nucleotide sequence data. *J. Mol. Biol.*, 1994, vol. 242, pp. 389-396.
5. Vitreschak A., Bansal A.K. and Gelfand M.S. Conserved RNA structures regulate initiation of translation of *Escherichia coli* and *Haemophilus influenzae* ribosomal protein operons. *1st Conf. BGRS-98*, 1998, vol. 1, p. 229.
6. Landick R., Yanofsky C. et al. *Escherichia coli* and *Salmonella*. *Cellular and Molecular Biology*. Neidhardt, F.C., Ed., Washington DC: ASM Press, 1996. vol. 1, Ch. 81, pp. 1263-1286.
7. Henkin T.M. tRNA-directed transcription antitermination. *Mol. Microbiol.*, 1994, vol. 13, no. 3, pp. 381-387.
8. Steiner K., Malke H. relA-Independent amino acid starvation response network of *Streptococcus pyogenes*. *J. Bacteriol.*, 2001, vol. 183, no. 24, pp. 7354-7364.
9. Sarsero J.P., Merino E., Yanofsky C. A *Bacillus subtilis* operon containing genes of unknown function senses tRNA<sup>Trp</sup> charging and regulates expression of the genes of tryptophan biosynthesis. *Proc. Natl. Acad. Sci. U S A*, 2000, vol. 97, no. 6, pp. 2656-2661.
10. Makarova K.S., Mironov A.A., Gelfand M.S. Conservation of the binding site for the arginine repressor in all bacterial lineages. *Genome Biol.*, 2001, vol. 2, no. 4:RESEARCH 0013.
11. Grundy F.J., Henkin T.M. The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in gram-positive bacteria. *Mol. Microbiol.*, 1998, vol. 30, no. 4, pp. 737-749.

*Статью представил к публикации член редколлегии В.А. Любецкий*