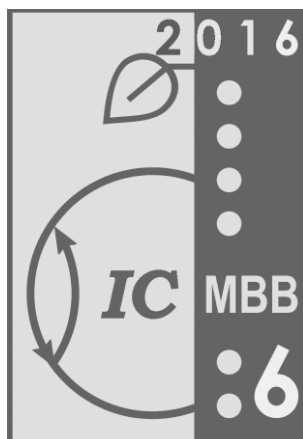


**РОССИЙСКАЯ АКАДЕМИЯ НАУК  
«ФЕДЕРАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР  
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ  
ИМЕНИ М.В.КЕЛДЫША»  
ИНСТИТУТ МАТЕМАТИЧЕСКИХ ПРОБЛЕМ БИОЛОГИИ**



## **МАТЕМАТИЧЕСКАЯ БИОЛОГИЯ И БИОИНФОРМАТИКА**

**ДОКЛАДЫ VI МЕЖДУНАРОДНОЙ КОНФЕРЕНЦИИ**

Пушино, 16-21 октября 2016 г.

**Под редакцией д.ф.-м.н. В.Д.Лахно**

**Пушино 2016**

**Математическая биология и биоинформатика:** Доклады VI Международной конференции.

16-21 октября 2016 г., г. Пущино

В сборнике представлены доклады VI-й Международной конференции «Математическая биология и биоинформатика», проводимой Институтом математических проблем биологии РАН при участии Научного совета по математической биологии и биоинформатике РАН. Конференция проводится при финансовой поддержке Российского фонда фундаментальных исследований (проект 16-07-20497).

**Состав Оргкомитета**

Лахно В.Д., д.ф.-м.н., – председатель  
Устинин М.Н., д.ф.-м.н. - зам.председателя  
Махортых С.А., к.ф.-м.н. – ученый секретарь  
Александров А.А., д.б.н.  
Арсеньев А.С., д.х.н.  
Васин А.А., д.ф.-м.н.  
Ильин В.А., д.ф.-м.н.  
Козлов Н.Н., д.ф.-м.н.  
Коротков Е.В., д.б.н.  
Марков А.В., д.б.н.  
Миронов А.А., д.ф.-м.н.  
Назипова Н.Н., к.ф.-м.н.  
Ризниченко Г.Ю., д.ф.-м.н.  
Романюха А.А., д.ф.-м.н.  
Самсонова М.Г., к.б.н.  
Туманян В.Г., д.ф.-м.н.  
Шабанов Б.М., к.т.н.  
Шайтан К.В., д.ф.-м.н.

**Редакционная группа:**

Махортых С.А., к.ф.-м.н.  
Назипова Н.Н., к.ф.-м.н.  
Якушева А.А.  
Махортых Е.С.

**Состав Программного комитета**

Журавлёв Ю.И., д.ф.-м.н., академик РАН –  
председатель  
Колчанов Н.А., д.б.н., академик РАН -  
зам.председателя  
Рудаков К.В., д.ф.-м.н., член-корреспондент РАН -  
зам.председателя  
Арчаков А.И., д.б.н., академик РАН  
Бердышев В.И., д.ф.-м.н., академик РАН  
Дегерменджи А.Г., д.ф.-м.н., академик РАН  
Евтушенко Ю.Г., д.ф.-м.н., академик РАН  
Жижченко А.Б., д.ф.-м.н., академик РАН  
Левин В.К., д.ф.-м.н., академик РАН  
Матвеев С.В., д.ф.-м.н., член-корреспондент РАН  
Рубин А.Б., д.ф.-м.н., член-корреспондент РАН  
Федотов А.М., д.ф.-м.н., член-корреспондент РАН  
Фрисман Е.Я., д.б.н., член-корреспондент РАН  
Четверушкин Б.Н., д.ф.-м.н., академик РАН  
Шокин Ю.И., д.ф.-м.н., академик РАН

# Отбор структур стебель-петля в промоторных областях транскрипционных факторов с помощью эволюционного алгоритма

Заикин А.В.<sup>1</sup>, Попцова М.С.<sup>1</sup>

Московский Государственный Университет им. М.В.Ломоносова, физический факультет<sup>1</sup>

[maria.poptsova@gmail.com](mailto:maria.poptsova@gmail.com)

## Selection of stem-loops in the promoter regions of transcription factors by an evolutionary algorithm

Zaikin A.V.<sup>1</sup>, Poptsova M.S.<sup>1</sup>

Lomonosov Moscow State University, Physics Department<sup>1</sup>

DNA secondary structures play an important role in various processes of genome functioning from bacteria to human. A stem-loop structure is an elementary structural motif that can function as terminator, attenuator, promoter and recognition elements. Earlier we annotated bacterial genomes with stem-loop structures and found that promoter regions of transcription factors (TF) are enriched with stem-loops. Here we posed a question whether stem-loops can play a role in TF autoregulation. To test this hypothesis we applied an evolutionary algorithm selecting statistically significant stem-loop structures in the TF promoter regions. For the directed selection we used the Pearson's chi-squared test – the greater the test value, the better the stem-loop parameters. For autoregulated TFs the evolutionary algorithm found the set of parameters so that the selected type of stem-loops was found in the promoter regions of the TFs with the feedback loop, and other TFs did not have any stem-loop structures. These results support the hypothesis of a possible regulatory role of stem-loops in the TF promoter regions.

## 1 Вступление

### 1.1 Вторичные структуры ДНК

Со времени открытия Уотсоном и Криком ДНК-структуры в виде двойной спирали [1], так называемой канонической В-формы, или В-ДНК, были получены экспериментальные доказательства существования структур, отличных от классической В-формы [2]. К таким структурам относятся структуры стебель-петля (известные также как шпильки или крестообразные структуры), триплексы, квадруплексы, А-ДНК, Н-ДНК, Z-ДНК и другие.

Важная роль вторичных структур ДНК была показана для разных геномных процессов, от бактерий до человека. Было показано, что структуры стебель-петля могут функционировать как терминаторы, аттенуаторы, промотеры и рекогниционные элементы [3-5]. В настоящей работе мы рассматриваем отдельный тип вторичных структур – структуры стебель-петля. Данные структуры могут образовываться в местах палиндромных участков последовательностей ДНК или РНК и образовывать шпильку в одинарной последовательности или крестообразную структуру в двойной спирали.

Ранее нами были проаннотированы геномы бактерий палиндромными структурами (аннотация доступна на сайте [www.dnapunctuation.org](http://www.dnapunctuation.org)).

Распределение этих структур по геному показало, что в геноме бактерий гены транскрипционных факторов (ТФ) и их окрестности статистически значимо обогащены структурами стебель-петля межгенных областей, прилегающих непосредственно к гену, выявили наличие шпилек как в терминаторных, так и в промоторных областях. Хорошо известно, что у бактерий шпильки часто служат терминаторами транскрипции, однако роль шпилек в промоторной области еще плохо изучена.

### 1.2 Транскрипционные факторы

Транскрипционные факторы – это гены, кодирующие белки-регуляторы множества генов, кодирующих разнофункциональные белки. ТФ обычно связываются с промоторными последовательностями ДНК, в областях до 50 нуклеотидов отстоящих от сайта. Однако, остается открытым вопрос, как осуществляется регуляция самих транскрипционных факторов.

Существуют экспериментальные данные, подтверждающие влияние вторичных структур на уровень экспрессии кластера генов [6]. В данной работе было показано, что шпилька между генами *arsC* и *arsD* сильно влияет на экспрессию кластера, но механизмы действия пока еще плохо изучены. Аналогично, регуляция оперона *dra-nupC-pdr* у бактерии *Bacillus subtilis* происходит с посредством

шпильки, расположенной в промоторной области оперона [7].

## 2 Эволюционный алгоритм

В настоящей работе мы выделили подкласс авторегуляторных ТФ, чья регуляция осуществляется структурами стебель-петля. Для этого был произведен анализ транскрипционной регуляторной сети бактерии *Bacillus subtilis*, из которой были вычлнены элементы с положительной и отрицательной обратной связью, в которых участвуют ТФ. Также были определены ТФ, которые потенциально регулируются структурами стебель-петля.

Для этого сначала с посредством эволюционного алгоритма был определен статистически значимый набор параметров структуры стебель-петля в промоторных областях ТФ. В эволюционном алгоритме был использован критерий согласия Пирсона для направленной селекции (чем больше значение критерия, тем лучше набор параметров шпильки). Критерий согласия Пирсона считался для двух вариантов выборов: ТФ противопоставлялись всем остальным генам *B. subtilis* и все авторегуляторные ТФ (ТФ с петлей обратной связи) – всем остальным ТФ.

В случае выборки второго типа с помощью эволюционного алгоритма был найден такой набор параметров, что характеризующие им шпильки были найдены в промоторных областях генов ТФ с петлей обратной связи, а во всех остальных ТФ структуры стебель-петля отсутствовали в принципе. Этот результат говорит в пользу возможной регуляторной роли шпилек, находящихся в промоторных областях ТФ.

Для выборок первого типа было обнаружено 27 генов, имеющих шпильку с определенным набором параметром в промоторной области. В этом случае гены можно было разбить на четыре основных группы: имеющие идентичную шпильку с идентичными координатами (12 генов), имеющие идентичную шпильку с идентичными координатами (4 гена), имеющие идентичную шпильку с идентичными координатами (тоже 4 гена) и все остальные (7 генов).

В первом случае (12 генов) было обнаружено, что вся исследуемая промоторная область (100 нуклеотидов) почти абсолютно идентична у этих 12 генов, хотя гены имеют разные функции (резистентность к мышьяку, регуляция уровня фруктозы и т.д.) и находятся в разных штаммах *B. subtilis*. Вторая и третья группы (по 4 гена) имеют аналогичное свойство (почти абсолютная идентичность исследуемой промоторной области). Четвертая же группа не имеет каких-то явно выраженных общих свойств.

Полученные результаты свидетельствуют о важной роли структур стебель-петля в авторегуляторных циклах ТФ. Требуется дальнейшее исследование генетической сети и усовершенствования методов эволюционного алгоритма для поиска статистически значимых структур в промоторных областях ТФ.

## Список литературы

- [1] Watson, J.D. and F.H. Crick, Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 1953. **171**(4356): p. 737-8.
- [2] Smith, G.R., Meeting DNA palindromes head-to-head. *Genes Dev*, 2008. **22**(19): p. 2612-20.
- [3] Wilson, K.S. and P.H. von Hippel, Transcription termination at intrinsic terminators: the role of the RNA hairpin. *Proc Natl Acad Sci USA*, 1995. **92**(19): p. 8793-7.
- [4] Wong, T.N., T.R. Sosnick, and T. Pan, Folding of noncoding RNAs during transcription facilitated by pausing-induced nonnative structures. *Proc Natl Acad Sci USA*, 2007. **104**(46): p. 17995-8000.
- [5] Haasnoot, P.C., et al., A conserved hairpin structure in Alfamovirus and Bromovirus subgenomic promoters is required for efficient RNA synthesis in vitro. *Rna*, 2000. **6**(5): p. 708-16.
- [6] Yu, X., et al., Transcriptional and posttranscriptional regulation of *Bacillus* sp. CDB3 arsenic-resistance operon *arsI*. *PeerJ*, 2015. **3**: p. e1230.
- [7] Zeng, X. and H.H. Saxild, Identification and characterization of a DeoR-specific operator sequence essential for induction of *dra-nupC-pdp* operon expression in *Bacillus subtilis*. *J Bacteriol*, 1999. **181**(6): p. 1719-27.