

*Лопатовская К.В.*¹, *Селиверстов А.В.*¹, *Любецкий В.А.*¹

¹ Институт проблем передачи информации им. А.А. Харкевича РАН

Об одном подходе к поиску регуляции в дискретных последовательностях

Постановка задачи и описание метода. В [1] формулируются понятия множественного выравнивания и сигнала, составленного из сайтов в данных последовательностях. Часто сигнал характеризуется дополнительными требованиями на искомые сайты. Например, можно требовать, чтобы сигнал допускал определенную комбинацию спиралей, определенным образом расположенных относительно друг друга. Спиралью называется пара отрезков («плеч»), которые могут быть побуквенно спарены в соответствии с правилом G с C и A с T. В этом случае говорят о сигнале, основанном на вторичной структуре; при этом сами исходные последовательности называют первичными структурами. Поиск такого сигнала, т.е. в конечном счете построение множественного выравнивания, о котором говорится в [1], представляет большую трудность. Здесь в качестве сигнала рассматривается классическая аттенуаторная регуляция, описанная, например, в [2]. Нами предложен метод, основанный на усовершенствовании модели из [2]. А именно, теперь эта модель позволяет учитывать РНК-триплексы, псевдоузлы разных типов и также преждевременную терминацию транскрипции, которая может происходить не только на полиурациле, но и в любом месте по ходу движения РНК-полимеразы, если только тормозящее действие соответствующей спирали неблагоприятно сочетается с ослаблением связи полимеразы с ДНК. Эта модель применялась для отбора лидерных областей аминокислотных генов, потенциально содержащих данную регуляцию, с последующим выравниванием сайтов регуляции. Таким образом, отбор лидерных областей происходил принципиально иначе, чем в работе [3], где использовался описанный там алгоритм LLLM.

Такой метод был применен самым широким образом к данным, содержащимся в базе NCBI. Полученные результаты собраны в файле, помещенном по адресу <http://lab6.iitp.ru> (пункт 3а). Там же выложены выравнивания полученных сайтов для нескольких случаев, которыми мы ограничимся в этом докладе. А именно, будет рассмотрена потенциальная классическая аттенуаторная регуляция у гамма-

протеобактерий, фирмикутов, бактериоидов и термотоговов для генов *hisG*, *hisZ*, *hisS*, *ilvD*, *lysQ*, *trpE*, *trpB*. Небольшая часть из этих сайтов уже была указана в [3].

Результат применения метода. Для гена *hisG* по всему спектру, кроме семейств *Pseudomonadaceae* и *Xanthomonadaceae*, гамма-протеобактерий (на выравнивании по указанному адресу находятся одновременно 64 лидерных области этого гена) наблюдаются после гена лидерного пептида 4 консервативных спирали. Первая из них играет роль антитерминатора, антитерминатор (спираль 1) далеко не пересекается с терминатором (спираль 4), а передает свое действие через ко-антитерминатор (спираль 2, которая складывается при наличии антитерминатора). Если антитерминатор не может сложиться из-за соответствующего положения рибосомы на мРНК гена лидерного пептида, то складывается ко-терминатор (спираль 3, которая не позволяет сложиться спирали 2, но позволяет сложиться терминатору). Как показывает наша модель, образование триплексов на антитерминаторе и отчасти на ко-терминаторе существенно для процесса регуляции.

Аналогичная картина регуляции для генов *ilvD*, *lysQ*, *hisZ* и *trpB* наблюдается на меньшем выравнивании у ряда фирмикутов. Ген *lysQ* кодирует пермеазу. Возможно, что у единственного вида *Lactococcus lactis* эта пермеаза сменила специфичность, поскольку мы наблюдаем зависимость от концентрации гистидина регуляцию, которая отсутствует у других фирмикутов. Для гена *hisZ* такая регуляция наблюдается у бацилл, листерий и единственной клостридии, но у листерий триплексы не образуются. Для гена *trpB* регуляция предсказана только у двух представителей фирмикутов: *Bacillus halodurans* и *Bacillus thuringiensis*. В регуляции не участвуют триплексы и структура мало консервативна. Это редкий случай регуляции этого гена как первого в опероне, еще он регулируется таким образом у единственной актинобактерии *Corynebacterium diphtheriae*. У гамма-протеобактерий он регулируется в составе длинного оперона *trpEGDCFBA*.

Для гена *ilvD* эта регуляция наблюдается у *Staphylococcus* и *Listeria*. Однако для гена *ilvD* у *Geobacillus thermodenitrificans* наблюдается классическая регуляция, когда антитерминатор непосредственно пересекается с терминатором и триплекс не образуется. Здесь предварительный результат был представлен в [4].

Для генов *hisG*, *ilvD*, *trpE* у *Bacteroides* по всему их спектру регуляция уже в ее традиционном варианте, когда антитерминатор непосредственно пересекается с терминатором, нашлась в более, чем 48 случаях.

Для генов *hisS*, *trpE*, *ilvD* у Thermotogae и Chloroflexi в ряде случаев также нашлась регуляция, у которой антитерминатор непосредственно пересекается с терминатором.

СПИСОК ЛИТЕРАТУРЫ

1. Селиверстов А.В., Любецкий В.А. Алгоритмы для поиска много боксовых сигналов и их применение // Труды 51-й научной конференции МФТИ «Современные проблемы фундаментальных и прикладных наук». – 2008. (в данном сборнике)
2. Seliverstov A.V., Lyubetsky V.A. Translation regulation of intron containing genes in chloroplasts // Journal of Bioinformatics and Computational Biology. – 2006. – V. 4, N. 4. – P. 783-793.
3. Vitreschak A.G., Lyubetskaya E.V., Shirshin M.A., Gelfand M.S., Lyubetsky V.A. Attenuation regulation of amino acid biosynthetic operons in proteobacteria: comparative genomics analysis // FEMS Microbiology Letters. – 2004. – V. 234. – P. 357-370.
4. Lyubetsky V.A., Seliverstov A.V. Amino acid biosynthesis attenuation in Bacteria // Proceedings of the fourth international conference on bioinformatics of genome regulation and structure. – 2004. – V. 1. – P. 307-310.