

УДК: 577.214.5

Классификация элементов вторичной структуры РНК

©2012 Баулин Е.Ф.^{*1,2}, Астахова Т.В.¹, Ройтберг М.А.^{**1}

¹Институт математических проблем биологии, Российская академия наук, Пущино,
Московская область, Институтская, 4, 142290, Россия

²НИУ Высшая школа экономики, Москва, Мясницкая 20, 101990, Россия

Аннотация. Предложена классификация элементов вторичной структуры РНК, которая является обобщением классификации петель, предложенной Мэтьюзом и Тернером. Эта классификация позволяет единообразно описывать структуры как содержащие, так и не содержащие псевдоузлы.

Ключевые слова: РНК, вторичная структура, спираль, псевдоузел, петля.

ВВЕДЕНИЕ

Изучение вторичной структуры РНК, в частности, предсказание вторичной структуры по последовательности путем минимизации энергии, - одна из классических задач биоинформатики [1–2]. В рамках общепринятой в настоящее время модели Цукера-Мэтьюза-Тернера (Nearest Neighbour Model, NNM, см. [3–4]) структура РНК разбивается на петли (loops). Каждая петля относится к определенному классу, для каждого класса дается формула вычисления энергии петли, зависящая от длин соответствующих участков РНК, нуклеотидов и т. п.

К сожалению, эта модель неприменима при изучении структур, которые содержат псевдоузлы. Проблема предсказания структур РНК, которые содержат псевдоузлы, изучена значительно хуже, чем проблема предсказания «классических» структур, не содержащих псевдоузлов (pseudoknot-free structures) [5]. При этом в настоящее время не существует общепринятой классификации псевдоузлов. Описаны простейшие виды псевдоузлов (kissing hairpins, H-structures и др, см. [6]), псевдоузлы, играющие важную роль в клеточных процессах, см. например, [7]); в [8] дан анализ классов псевдоузлов, введенных в различных работах по предсказанию вторичной структуры РНК по ее последовательности. В базе данных [9, 10] собраны примеры псевдоузлов в экспериментально определенных структурах РНК. Однако единого способа описания элементов вторичной структуры при наличии псевдоузлов, подобного тому, который был предложен в [3–4] для описания структур, которые не содержат псевдоузлов, в настоящее время нет.

Цель настоящей работы – восполнить этот пробел. Мы предлагаем способ описания элементов вторичной структуры, пригодный как для описания классических петель, так и для описания псевдоузлов. Этот способ основан на разложении плоских графов в элементарные циклы, ср. с работой [11].

Статья имеет следующую структуру. В разделе «Основные определения» вводится вся необходимая терминология. В разделе «Спирали и петли» вводится основное для предлагаемого подхода понятие петли, обобщающее понятие петли по Мэтьюзу-Тернеру. Далее, в разделе «Структура петель» доказываются утверждения,

* baulin@rambler.ru

** mroytberg@impb.psn.ru, mroytberg@lpm.org.ru

позволяющие установить, каков общий вид петель и ввести их классификацию, которая согласована с классификацией Мэтьюза-Тернера. В «Заключении» обсуждаются перспективы практического применения предложенных определений.

ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ

Молекулу РНК мы будем представлять, как последовательность нуклеотидов, иначе говоря, как символьную последовательность в алфавите $\{A, C, G, U\}$. Каждый нуклеотид в молекуле имеет свой номер от 1 до L , где L – длина последовательности.

Связь – это пара нуклеотидов (i, j) , где $i < j$, которая образует водородную связь. При этом допускаются не только связи между комплементарными нуклеотидами и G-U связи, но и неканонические связи, см. [12].

Спираль – это последовательность пар нуклеотидов вида $(i, j), (i + 1, j - 1), \dots, (i + k, j - k)$ такая, что

- 1) $i < j, i + k < j - k, k \geq 1$;
- 2) все пары вида $(i + x, j - x)$, где $x = 0, \dots, k$, образуют связи, причём связи (i, j) и $(i + k, j - k)$ – Уотсон-Криковские связи (УК-связи), т. е. связи между комплементарными нуклеотидами.

Участок цепи $[i, i + k]$ будем называть *левым крылом* спирали, соответственно участок $[j - k, j]$ будем называть *правым крылом* спирали.

Пару (i, j) будем называть *внешней парой* спирали или *торцом* спирали, пару $(i + k, j - k)$ будем называть *внутренней парой* спирали.

Замечание. В популярной программе 3DNA [12] используется другое определение спирали. Описанный ниже подход применим к любому определению спирали. Говоря неформально, мы считаем, что на последовательности РНК каким-то образом уже размечены левые и правые крылья спиралей и установлено соответствие между крыльями одной спирали.

Нить – это такой участок цепи $[i, j]$, где $i < j$, что

- 1) не существует такой связи (k, t) , что $i \leq k \leq j$ или $i \leq t \leq j$.
- 2) существуют пары, в которые входят нуклеотиды i и j .

Вторичная структура РНК – это такое множество связей, что

- 1) каждый нуклеотид входит не более, чем в одну связь;
- 2) каждая пара входит в некоторую спираль.

Отметим, что в экспериментально определенных пространственных структурах РНК есть значительное число водородных связей, не входящих в спирали [13], роль таких связей в настоящее время изучена слабо. Мы исходим из предположения, что полезно отдельно рассматривать «базовую» вторичную структуру, образованную спиралью, и (над этой структурой) – одиночные водородные связи.

СПИРАЛИ И ПЕТЛИ

Здесь и далее будем считать фиксированной цепь РНК с заданной на ней вторичной структурой. Эту цепь можно рассматривать как чередующуюся последовательность нитей (некоторые нити могут иметь нулевую длину) и крыльев. Для удобства изложения мы будем считать, что перед первым и после последнего нуклеотида цепи добавлены крылья «внешней спирали».

С каждой спиралью связан *внутренний* по отношению к ней участок цепи – участок между концом левого крыла и началом правого крыла, иначе говоря, – между нуклеотидами, образующими внутреннюю пару спирали. Для фиктивной внешней спирали внутренним участком является вся исходная последовательность РНК.

Пусть H – спираль и (i, j) – ее внутренняя пара.

Определение 1. Позиция цепи t – *внутренняя* для спирали H (синоним: *лежит внутри H*), если $i < t < j$. Фрагмент цепи – *внутренний* для спирали H (синоним: *лежит*

внутри H), если все его позиции – внутренние для спирали H . Спираль H_1 лежит внутри спирали H (является внутренней для H), если все позиции ее крыльев – внутренние для H .

Определение 2. Позиция цепи t принадлежит спирали H , если она внутренняя для H и не существует спирали H_1 , лежащей внутри H , такой, что $x < t < y$, где (x, y) – внешняя пара (торец) H .

Определение 3. Петля спирали H – это множество всех позиций, которые принадлежат спирали H .

Очевидно, каждая позиция, не входящая в связь, принадлежит хотя бы одной петле – обычной или внешней. При этом если какая-то позиция нити (крыла) принадлежит некоторой петле, то и вся нить (все крыло) принадлежит этой петле.

Если в структуре нет псевдоузлов, то каждая петля в смысле определения 3 является петлей по Мэтьюзу-Тернеру и наоборот. При этом каждая нить принадлежит ровно одной петле (возможно, внешней), а ни одно крыло не принадлежит какой-либо петле. Для структур с псевдоузлами оба эти свойства нарушаются.

СТРУКТУРА ПЕТЕЛЬ

Определение 4. Пусть H – спираль и (u, v) – ее внутренняя пара. Участок, $[i, j]$ называется замкнутым относительно H , если

- 1) $[i, j]$ лежит внутри H ;
- 2) не существует таких связей (k, t) , что $(i \leq k \leq j < t < v)$ или $(u < k < i \leq t \leq j)$;
- 3) существуют связи (i, k) и (t, j) , где $k \leq j$; $u \leq t$.
- 4) не существует отличного от $[i, j]$ участка $[i', j']$ такого, что $i \leq i' < j' \leq j$ и участок $[i', j']$ удовлетворяет условиям 1) - 3).

Пара нуклеотидов (i, j) называется торцом замкнутого относительно H участка.

Утверждение 1. Пусть $Z = [f, g]$ – участок, замкнутый относительно спирали H ; (u, v) – внутренняя пара спирали H . Тогда:

1. Участок Z целиком лежит внутри спирали H .
2. Крыло либо целиком лежит в Z , либо целиком лежит вне Z .
3. Замкнутый относительно H участок начинается левым крылом некоторой спирали H_1 , лежащей внутри H , и заканчивается правым крылом некоторой спирали H_2 , лежащей внутри H .
4. Если $H_1=H_2$ – это одна и та же спираль, то торец (s, t) участка Z – это торец спирали H . В противном случае s – это начало левого крыла спирали H_1 , t – это конец правого крыла спирали H_2 .

Доказательство – следует из определения 4 и того, что крылья не пересекаются.

Определение 5. Пусть Z – это участок, замкнутый относительно спирали H . Участок Z называется простым, если его торец – это торец некоторой спирали и сложным в противном случае. Сложные участки для краткости будем называть блоками.

Утверждение 2. Пусть H – спираль; (u, v) – ее внутренняя пара. Тогда

1. Никакие два участка, замкнутых относительно H , не пересекаются.
2. Пусть позиция t лежит внутри спирали H . Позиция t НЕ принадлежит спирали H тогда и только тогда, когда t лежит внутри некоторого участка Z , замкнутого относительно H (т. е. лежит в Z , но не входит в его торец).

Доказательство – следует из определений 1, 2 и 4.

Определение 6. Пусть H – спираль и (u, v) – ее внутренняя пара. Пусть $(s_1, t_1), \dots, (s_n, t_n)$ – торцы всех участков, замкнутых относительно H ; $s_1 < t_1 < \dots < s_n < t_n$. Для удобства пусть $t_0 = u$; $s_{n+1} = v$. Пусть k – целое; $1 \leq k \leq n + 1$. Тогда k -я грань петли H – это фрагмент $[t_{k-1}+1, s_k-1]$.

Замечание. Если $s_k = t_{k-1}+1$, то k -я грань петли H – пустой отрезок.

Утверждение 3. Пусть H – спираль и (u, v) – ее внутренняя пара. Пусть $(s_1, t_1), \dots, (s_n, t_n)$ – торцы всех участков, замкнутых относительно H ; $s_1 < t_1 < \dots < s_n < t_n$. Для удобства пусть $t_0 = u$; $s_{n+1} = v$. Тогда петля спирали H – это объединение торцов всех участков, замкнутых относительно H , и расположенных между ними граней.

Доказательство – следует из утверждения 2.

Утверждение 4. Пусть H – спираль и (u, v) – ее внутренняя пара и позиция x принадлежит грани (t, s) петли спирали H . Тогда

1. Позиция x либо не участвует в связи, либо принадлежит крылу спирали H' , другое крыло которой лежит вне спирали H .

2. Если x принадлежит нити (крылу спирали), то все позиции этой нити (этого крыла) принадлежат той же грани петли спирали H .

Доказательство – следует – следует из определения граней, п. 1.2 и того, что крылья не пересекаются.

Утверждения 3 и 4 описывают возможные структуры петель. Отметим, что в случае структур, которые не содержат псевдоузлов, все замкнутые участки – простые и каждая грань состоит из единственного одностороннего участка.

Поэтому можно дать такое определение.

Определение 7. Петля называется *классической*, если она не содержит крыльев и торцов блоков. Петля называется *изолированной*, если она не содержит крыльев, и *узловой*, если она содержит крылья.

Спираль называется *узловой*, если ее петля – узловая.

Классификацию петель по Мэтьюзу-Тернеру на введенное нами обобщение, основываясь на количестве торцов, входящих в петлю. Отметим, что в нашем случае торцы могут быть как торцами спиралей (иными словами – простых замкнутых участков), так и концами блоков (сложных замкнутых участков).

Определение 8. Петля называется *тупиковой* (hairpin), если она не содержит торцов и, соответственно имеет одну грань. Петля называется *внутренней* (internal), если она содержит ровно один торец, и, соответственно, имеет две грани. Петля называется *ветвящейся* (multiple), если она содержит более одного торца, и, соответственно, более двух граней.

ЗАКЛЮЧЕНИЕ

Нами предложено новое определение петли, которое, с одной стороны, является обобщением понятия петли по Мэтьюзу-Тернеру, а с другой позволяет разбить на петли произвольную вторичную структуру, а не только структуры, не содержащие псевдоузлов. Возможность такого описания важна при создании баз данных экспериментально определенных структур РНК

Определение основывается на наблюдении, что определение Мэтьюза-Тернеру устанавливает взаимно-однозначное соответствие между спиралью и петлей, причем каждая петля по Мэтьюзу-Тернеру состоит из односторонних участков, которые разделены торцами спиралей. Основные отличия наших петель от петель Мэтьюза-Тернера состоят в следующем.

1) В петлях для вторичных структур общего вида торцы могут быть торцами как спиралей, так и сложных замкнутых участков (в петлях Мэтьюза-Тернера – только торцы спиралей).

2) Грань петли – это чередующаяся последовательность петель и крыльев спиралей (в петлях Мэтьюза-Тернера грани не содержат крыльев спиралей).

3) Односторонний участок РНК может принадлежать нескольким петлям (в петлях Мэтьюза-Тернера – только одной петле).

На основе новых определений предложена классификация петель, являющаяся обобщением классификации Мэтьюза-Тернера. Представляется интересным выяснить,

какие из возможных видов петель реализуются в экспериментально определенных структурах РНК.

Работа выполнена в рамках государственного контракта 07.514.11.4004 шифр 2011-1.4-514-008-009 при финансовой поддержке Минобрнауки России.

СПИСОК ЛИТЕРАТУРЫ

1. *Mathematical methods for DNA sequences*. Ed. Waterman M.S. Boca Raton: CRC Press, 1989. 293 p.
2. Nussinov R., Jacobson A.B. Fast algorithm for predicting the secondary structure of single-stranded RNA. *Proc. Natl Acad. Sci. USA*. 1980. V. 77. P. 6309–6313.
3. Xia T., SantaLucia J.Jr., Burkard M.E., Kierzek R., Schroeder S.J., Jiao X., Cox C., Turner D.H. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson–Crick base pairs. *Biochemistry*. 1998. V. 37. P. 4735.
4. Zuker M., Mathews D.H., Turner D.H. Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In: *RNA Biochemistry and Biotechnology*. Eds. J. Barciszewski, B.F.C. Clark. Boston: Kluwer Academic Publishers, 1999. P. 11–43. (NATO ASI Series).
5. Mathews D.H., Turner D.H. Prediction of RNA secondary structure by free energy minimization. *Current Opinion in Structural Biology*. 2006. V. 16. P. 270–278.
6. Pleij, C.W.A. RNA Pseudoknots. In: *The RNA world*. Eds. Gesteland R.F., Atkins J.F. Cold Spring Harbor Laboratory Press, 1993. P. 609–613.
7. Gulyaev A.P., Olsthoorn R.C. A family of non-classical pseudoknots in influenza A and B viruses. *RNA Biol*. 2010. V. 7. P. 125–129.
8. Condon A., Davy B., Rastegari B., Zhao S., Tarrant F. Classifying RNA pseudoknotted structures. *Theoretical Computer Science*. 2004. V. 320. P. 35–50.
9. Batenburg F.H.D. van, Gulyaev A.P., Pleij C.W.A., Ng J., Oliehoek J. Pseudobase: a database with RNA pseudoknots. *Nucl. Acids Res*. 2000. V. 28. P. 201–204.
10. Taufer M., Licon A., Araiza R., Mireles D., Batenburg F.H.D. van, Gulyaev A.P., Leung M.-Y. PseudoBase++: an extension of PseudoBase for easy searching, formatting and visualization of Pseudoknots. *Nucl. Acids Res*. 2009. V. 37. P. D127–D136.
11. Haslinger C., Stadler P.F. RNA Structures with Pseudo-knots: Graph-theoretical, Combinatorial, and Statistical Properties. *Bulletin of Mathematical Biology*. 1999. V. 61. P. 437–467.
12. Zheng G., Lu X.-J., Olson W.K. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res*. 2003. V. 31. P. 5108–5121.
13. Baulin E., Ivankov D., Roytberg M. Statistics of RNA structures. In: *Proceedings of Moscow Conference on Computational Molecular Biology (July 21-24, 2011. Moscow)*. 2011. P. 325–326.

Материал поступил в редакцию 22.10.2012, опубликован 26.10.2012.