

Применение теории игр для задачи сворачивания рибонуклеиновых кислот

А.Н. Борисов, С.С. Яковлев

Аннотация. Разработано программное обеспечение RNAInSpace, которое позволяет моделировать процесс сворачивания рибонуклеиновой кислоты (РНК). Отличием разработанного метода от многих методов, применяемых для моделирования сворачивания, является то, что он основан на принципах теории игр. В качестве обоснования отказа от имеющихся методов, авторами ранее было показано, что использование метода Монте-Карло не может теоретически достичь достаточной точности при поиске третичной структуры РНК.

Ключевые слова: теория игр, сворачивание РНК, NP-полная задача, третичная структура РНК.

Введение

В работе [1] авторами были описаны два метода «Быстрое охлаждение» и «Х-тюнинг» для целенаправленного поиска третичной структуры фрагментов рибонуклеиновых кислот. Их отличием от других методов, основанных на стохастическом поиске, где как правило используются различные модификации метода Монте Карло [2-5], является отказ от использования случайности при выборе углов, на которые поворачиваются атомы в нуклеотидах при сворачивании цепи из рибонуклеиновых кислот (РНК) в компактную глобулу. Дело в том, что применение случайности делает поворот любого угла любого нуклеотида в цепи РНК равновероятным. Но предположение такой равновероятности в процессе сворачивания неправомерно. Наши исследования показывают, что успешность и скорость сворачивания РНК при моделировании (*in silico*) существенным образом зависит от стратегии поиска нативного состояния РНК (то есть состояния, находящегося в природном, не видоизмененном экспериментом состоянии). В этих стратегиях необходимо учитывать последовательность поворотов нуклеотидов на те, или иные углы и их комбинации между атомами.

Таким образом, искать третичную структуру РНК, используя случайность, фиксируя те повороты, которые уменьшают некоторую оценку энергии, это все равно, что учиться играть в шахматы, делая случайные ходы. А с учетом того, что невозможно перебрать все возможные состояния (NP-сложность), такой подход неперспективен. Поэтому в статье [1] нами данная проблематика была сформулирована как задача теории игр и принятия решений [6] и показано, что методы могут потенциально использоваться как для игры в шахматы, так и для игры с природой, находя третичную структуру РНК.

В этой статье мы опишем различные успешные стратегии игры, которые оптимально позволяют найти третичную структуру РНК. На пути к этому были проанализированы и другие стратегии, которые не были успешными. Детали их реализации будут опущены, и мы ограничимся в статье лишь кратким описанием неперспективных стратегий.

1. Сведения о предметной области

1.1. Общие сведения

Рибонуклеиновые кислоты состоят из отдельных нуклеотидов, каждый из которых состоит из рибозы ($C_5H_{10}O_5$) и одного из азоти-

стых оснований — аденина (а), цитозина (с), гуанина (g) или урацила (u). Атомы в нуклеотидах связаны ковалентными связями, и в зависимости от внешних условий могут, поворачиваясь в цепи РНК, принимать ту или иную конфигурацию. Данная конфигурация во многом определяется водородными связями, которые образуются между разными нуклеотидами. Знание конкретных водородных связей существенно для определения вторичной структуры.

На данный момент, зная первичную структуру, можно достаточно точно предсказать вторичную [7]. В задаче трехмерного сворачивания РНК нужно, зная вторичную структуру РНК, получить ее третичную структуру, то есть расположение всех атомов РНК в пространстве.

1.2. Углы поворотов в нуклеотиде

На Рис.1 показаны основные углы в нуклеотиде, которые подвержены вращению: 6 углов на главной цепи α , β , γ , δ , ϵ , ζ (обозначим их цифрами от 1 до 6) и 3 угла боковой цепи ν_1 , ν_2 , χ (обозначим эту совокупность одной цифрой 7). РНК изменяет свою форму путем поворотов этих углов. Не все комбинации углов возможны. Поэтому области углов коррелированы между собой. Но известны только общие закономерности, на основании которых нельзя сократить возможные комбинации. Даже поворот отдельно взятого угла на 0.1 градуса влияет на возможность провернуться цепи РНК по другим углам. Таким образом, существует астрономическое число вариаций, которые нужно проверить на одном шаге сворачивания.

2. Программное обеспечение RNAInSpace

Для расчетов и проведения экспериментов *in silico* С. Яковлевым было разработано специальное программное обеспечение (ПО) RNAInSpace. Оно позволяет визуализировать процесс сворачивания РНК в 3-мерном пространстве. А также аналитически следить за поиском третичной структуры РНК.

В следующем разделе будет алгоритмически описан метод «Медленное сжатие», лежащий в основе поиска третичной структуры РНК в ПО RNAInSpace. Здесь же описан применяемый подход на качественном уровне и некоторые

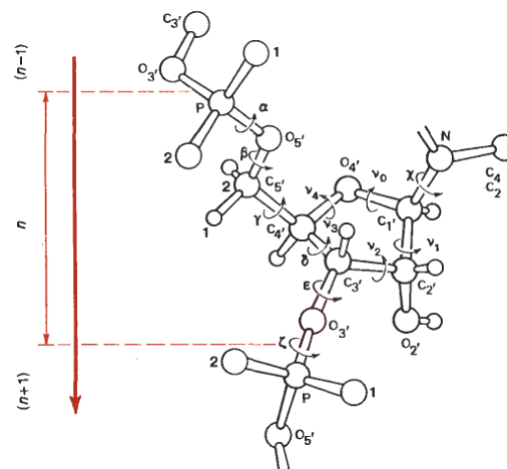


Рис. 1. Обозначения углов поворотов в нуклеотиде

Рисунок заимствован из Wolfram Saenger, Principles of Nucleic Acid Structure, 1984

его детали, чтобы пояснить, почему применяется именно этот метод.

2.1. Гладкость поверхности, образуемой множеством состояний

В работе [1] был представлен метод «Быстрое охлаждение» для поиска приблизительного (грубого) свернутого состояния цепи. По сути, этот метод является частным случаем метода «Медленное сжатие». Метод «Медленное сжатие» сформулирован таким образом, что в зависимости от задания ряда переменных можно получить различные стратегии осуществления поиска. И тогда метод «Быстрое охлаждение» представляет собой самую простую стратегию метода «Медленное сжатие».

С другой стороны, в работе [1] основную роль в поиске выполнял метод «Х-тюнинг». В отличие от метода «Быстрое охлаждение» метод «Х-тюнинг» осуществлял перебор положений не одного нуклеотида, а искал выгоднейшую комбинацию из двух нуклеотидов. При этом общим у этих методов было то, что они у каждого нуклеотида одновременно поворачивали все 9 углов.

Таким образом, метод «Быстрое охлаждение» был недостаточным, так как не находил требуемые комбинации углов, при которых цепь образовывала все требуемые водородные связи. В какой-то момент последовательные повороты приводили к тому, что значение функции полезности скатывалось в локальный минимум, из которого уже последовательными

поворотами нельзя было выбраться. А метод «Х-тюнинг» хоть и находил такие комбинации, но затрачивал много вычислительных ресурсов, так как нужно было принимать во внимание положение не одного нуклеотида, а сразу двух. Это приводило к перебору порядка миллиона возможных комбинаций значений углов только для осуществления одного поворота, а поиск мог длиться несколько суток. Общим у этих методов было то, что они довольно грубо изменяли комбинации углов между атомами нуклеотидов.

Так же как и в работе [2], комбинации углов взяты на основании рентгеноструктурного анализа большой рибосомной РНК вида *Haloarcula marismortui* [8]. Таким образом, получается база данных ≈ 2700 вариантов комбинаций 9 углов. Но в отличие от работы [2], в работе [1] вместо комбинаций сразу 27 углов (3 нуклеотидов) использовались комбинации 9 углов в зависимости от вида нуклеотида. Имеются для G - 829 вариаций углов, для C - 680, A - 697, U - 491.

Но и этого оказалось недостаточно – слишком мало комбинаций, чтобы можно было плавно осуществить поиск. А именно отсутствие гладкости поверхности, в которой мы осуществляем поиск, приводило к большому числу локальных минимумов, из-за которых метод «Быстрое охлаждение» не мог дальше работать. А метод «Х-тюнинг» для обеспечения гладкости искал комбинации 9 углов одного нуклеотида по отношению к 9-ти углам второго нуклеотида.

Но для обеспечения гладкости можно воспользоваться и другим способом. Один из способов это поворачивать только на один угол, а не сразу на девять. Таким образом, число комбинаций будет уже не 2700, а существенно больше за счет комбинирования. Но такая стратегия оказалась не очень успешной. А вот повороты на 2 угла сразу из девяти дали неплохие результаты. Большее число поворотов на три, четыре и т.д. углов сразу уже дают меньшую гладкость. Поэтому в основу разнообразия стратегий были положены различные комбинации именно двух углов.

2.2. Обобщенная функция полезности

В работе [1] функция полезности $f(x)$ играет такую же роль, как стремление энергетической

оценки к минимуму. Но основывается на близости нуклеотидов, которые должны образовать водородную связь. А именно $f(r, a)$, где r – расстояние между нуклеотидами, a – угол между донором и акцептором предполагаемой водородной связи. При некотором усреднении можно считать, что водородная связь образуется при расстоянии r между соответствующими атомами, меньшем $3.0 \text{ \AA}$, и угле a между донором и акцептором, меньше 20° . В методе «Быстрое охлаждение» функция полезности вычислялась для всех нуклеотидов последовательно от начала цепи до конца. Сама же функция полезности вычислялась как превышение уровня $r < 3.0$ и $a < 20^\circ$, при этом значения функции полезности для каждой пары нуклеотидов суммировались, получая, таким образом, совокупную дальность между парами нуклеотидов, которые должны образовать водородную связь.

В результате проведенных экспериментов стало очевидным, что образование водородной связи не у каждой пары нуклеотидов равноценно для достижения цели – образования всех водородных связей. Образование водородных связей у тех пар нуклеотидов, которые находятся ближе к середине цепи, играют более важную роль. Это связано с тем, что если образуется водородная цепь на концах цепи, то она потом должна опять быть разорвана при образовании водородных связей у пар нуклеотидов, находящихся ближе к центру цепи.

Это следствие применяемого геометрического способа вращения при моделировании, и эти выводы не стоит распространять на биологический ход сворачивания. Но, в то же время, нужно понимать, что моделировать процесс сворачивания так, как он идет в природе, мы не можем и не должны. В природе сворачивание – это параллельный процесс, который происходит в условиях хаотического колебания цепи РНК. Нам же важно, моделируя процесс сворачивания, разложить параллельное сворачивание на последовательные этапы, выстроив их в логическом порядке, идеализируя, таким образом, происходящий процесс.

Из-за этого наблюдения в методе «Медленное сворачивание» была так изменена функция полезности, чтобы она учитывала важность образования водородных связей у пар нуклеоти-

дов, расположенных ближе к центру цепи. А также, чтобы дальность водородных связей у пар, следующих друг за другом, была взаимосвязана. Опишем расчет функции полезности:

- Находим максимально отдаленное расстояние r среди 2 или 3 потенциальных пар атомов, образующих водородную связь в паре нуклеотидов:

$$\max R = \max (r1, r2, r3)$$

- Находим наибольший угол a :

$$\max A = \max (a1, a2, a3)$$

- Приводим расстояние и угол к одному «масштабу»: $\max R = (\max R - 3.0) * 10$. Вычитаем 3 ангстрема, как ограничение расстояния при образовании водородной связи

- Аналогично вычитаем 20 градусов:

$$\max A = (\max A - 20)$$

- Выбираем наибольшее значение близости расстояния или угла, приведенное к одному «масштабу»:

$$\text{if } (\max R > \max A) \{ \text{locScore} = \max R; \} \\ \text{else } \{ \text{locScore} = \max A; \}$$

- Когда $\text{locScore} \leq 0$ будут образованы водородные связи между всеми 2 или 3 парами атомов между одной парой нуклеотидов

- Задача решена, когда все заданные пары нуклеотидов образовали водородные связи. Тогда, учитывая выше описанное наблюдение, вид функции полезности $f(r,a)$ должен быть следующим:

$$\text{Score} = \frac{\sum_n (\text{Pairs}_n * \text{CompressionCoef}^{kt})}{\text{PrecisionCoef}} \quad (1)$$

где $\text{Pairs}_n (= \text{locScore})$ – оценка образования водородных связей между n -ой парой нуклеотидов; $kt = N - n$, то есть *обратное положение текущего нуклеотида* (N – общее число пар нуклеотидов, n – номер текущего нуклеотида); $\text{CompressionCoef} (=0.1, =5, =10)$ – коэффициент сжатия; $\text{PrecisionCoef} (=10)$ – коэффициент точности.

Заметим, что при коэффициенте сжатия = 1, функция полезности в методе «Медленное сжатие» становится такой же, как в методе «Быстрое охлаждение».

Кроме того, при моделировании сворачивания у РНК часто раньше соединяются концы фрагмента, а затем по направлению к середине фрагмента. Это создает принципиальную невозможность для прямых расчетов, так как

невозможно зафиксировать концы и рассчитать изменения внутри петли РНК так, чтобы не происходило движения на ее концах или не допустить разрыва цепи. Поэтому важно удерживать концы развернутыми до того, как образуются водородные связи, начиная с внутренней части фрагмента. Для этого, пока $\text{Pairs}_n > 10$, повороты, которые уменьшают оценку следующего нуклеотида быстрее текущего, не осуществляются. Другими словами, расстояния между парами нуклеотидов выдерживаются так, чтобы в каждой последующей паре расстояние было как минимум на 10 больше предыдущего, пока не снизится до меньше 10 (листинг 1).

Листинг 1. Ограничение сжатия

```
for (int k = 2; k < N+1; k++)
    if (Pairs[k - 1] > 10 and (Pairs[k - 1] + 10) > Pairs[k])
        StopRotate = 1; // Невыполнять такой поворот
```

3. Метод «Медленное сжатие»

В методе «Медленное сжатие» часто используется функция «Вращение одного нуклеотида», которую имеет смысл рассмотреть отдельно от самого метода. А затем перейти на основании применения этой функции к рассмотрению различных стратегий, применяемых для поиска третичной структуры РНК.

3.1. Вращение одного нуклеотида

На Рис. 2 представлен алгоритм, по которому выполняется вращение одного нуклеотида, в результате чего находится наилучший поворот, который уменьшает функцию пригодности.

3.2. Стратегии вращения цепи из нуклеотидов

В разделе 3.1. указывалось, что в основу стратегий были положены повороты различных комбинаций двух углов в нуклеотиде. Можно выделить три базовые стратегии:

1. Минимальная (St0) – комбинируются углы: 1 и 6 (далее, упрощая, без союза «и»), 7, 25, 34, 7
2. Начальная (St1) – комбинируются углы: 12, 7, 23, 34, 7, 45, 56, 7, 16
3. Полная (St2) – комбинируются углы: 16, 25, 34, 7, 13, 24, 36, 7, 12, 23, 45, 7, 14, 26, 35, 7, 15, 46, 56, 7

С другой стороны, указывалось, что в базе данных имеются для G - 829 вариаций углов,

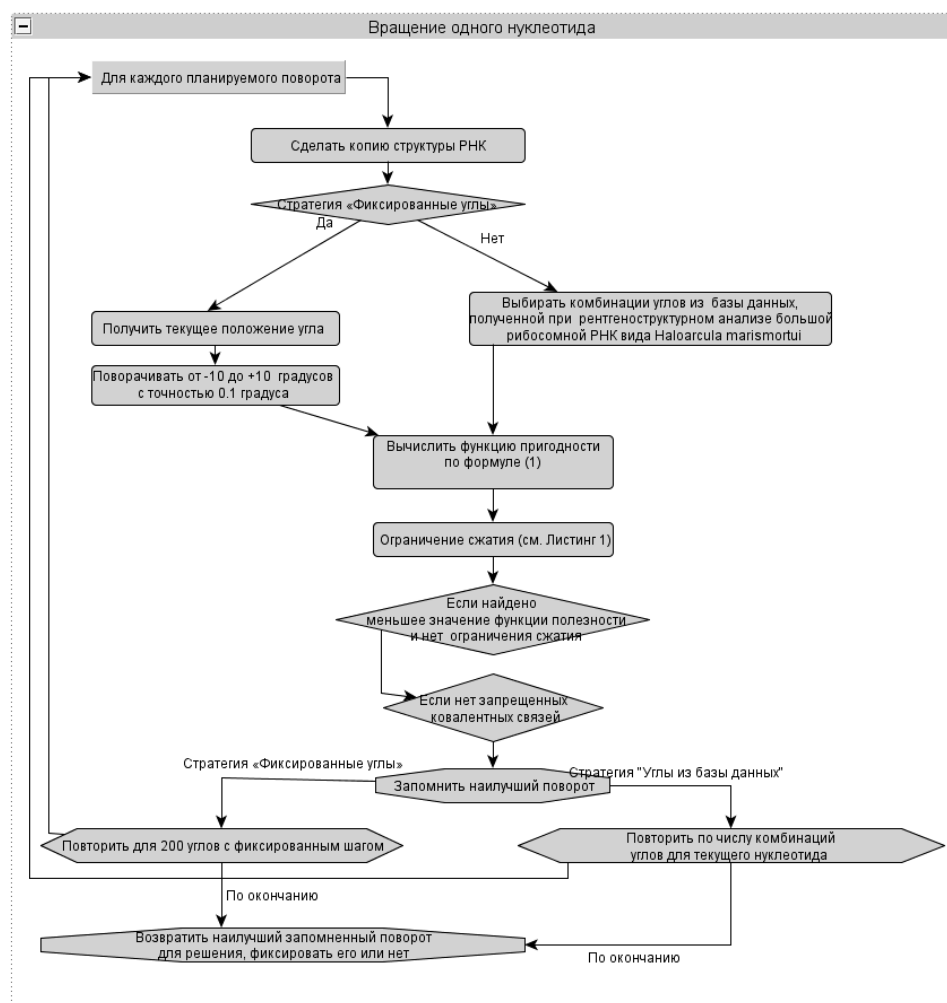


Рис. 2. Алгоритм нахождения наилучшей комбинации углов при вращении одного нуклеотида

для С - 680, А - 697, U - 491. Оказалось, что при разделении по виду нуклеотидов получаются недостаточно успешные стратегии. И лучше всего совмещать углы для пуринов (**a** и **g**, вместе 1526 вариантов) и пиримидинов (**c** и **u**, вместе 1171). Этот набор и был взят как основной. Но, кроме того, была модификация **Full**, где использовались все 2700 вариантов, вообще не различая вид нуклеотида. А также еще одна модификация **x2**, где, как в основной, различались пурины и пиримидины, но к базе из 2700 вариантов были добавлены еще варианты на основании другой модели 1JJ2 [9].

На Рис. 3 приведен алгоритм вращения цепи. Следует отдельно пояснить, что цепь разбивается на блоки, например, так:

6	14	7	8	9	10	11	12	13
5	15	6	14					
4	16	5	15					
3	17	4	16					
2	18	3	17					
1	19	2	18					

Здесь каждый номер означает номер нуклеотида, и каждая строка это отдельный блок. Первые две цифры в блоке это номера нуклеотидов, которые образуют водородные связи. Остальные цифры в блоке это номера нуклеотидов, положение которых наиболее влияет на возможность образования водородных связей у соответствующей пары. Последовательность вращения зависит от этой информации.

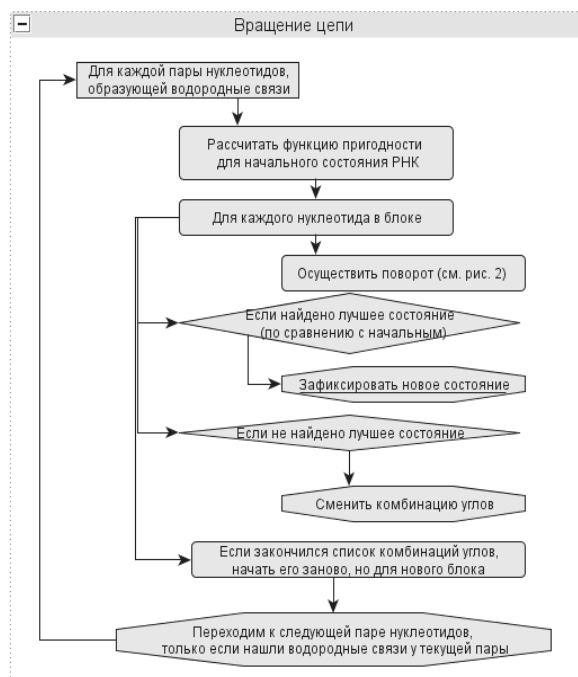


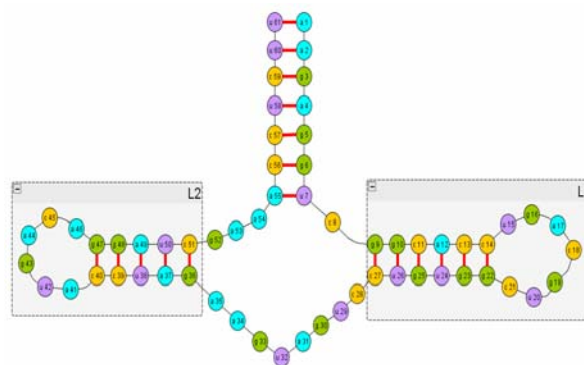
Рис. 3. Алгоритм вращения цепи РНК

4. Применение метода «Медленное сжатие»

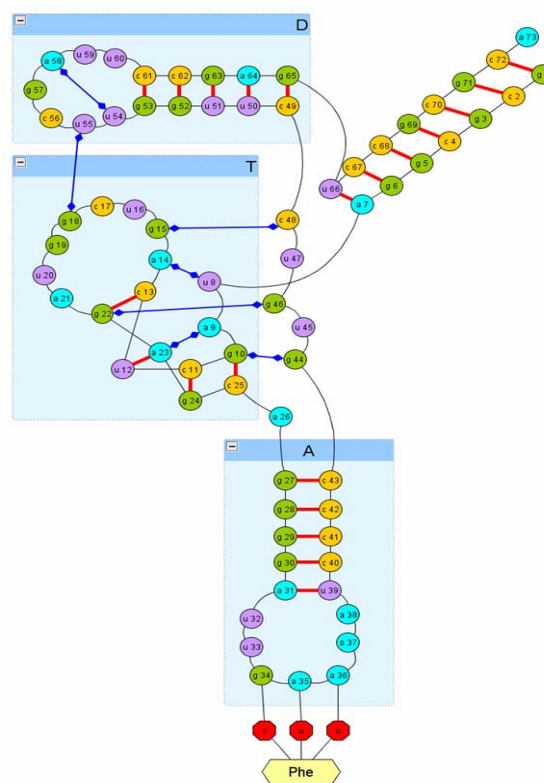
Продemonстрируем результаты применения метода «Медленное сжатие» на примере нахождения третичной структуры 5 различных фрагментов. В качестве первых двух были взяты фрагменты L1 и L2 вироидного рибозима RZ+ штамма *Chrysanthemum chlorotic mottle viroid* (секвенированный локус NC_003540 [10]) с классической структурой типа “головки молотка” (Hamerhead ribozyme) (Рис. 4, сверху). А в качестве еще трех D, A и T спирали транспортной РНК организма *Escherichia coli* (Рис. 4, внизу). В таблице приведены результаты применения различных стратегий при нахождении третичной структуры 5 фрагментов. Можно видеть, что ни одна стратегия не является абсолютно успешной. И хотя выводы нужно делать с осторожностью, можно видеть, что стратегия St0 не дает успешных результатов. Стратегия St2 чаще всего дает успешные результаты. Но в тех случаях, когда стратегия St2 не является успешной, может оказаться пригодной стратегия St1. Если же сравнить модификации этих стратегий, то для стратегии St1 модификации St1 Full и St1 x2 не дают существенно лучших

результатов. Но для стратегии St2 эти модификации иногда могут быть успешнее базового варианта St2.

Поэтому можно предложить следующую последовательность применения стратегий:



вторичная структура рибозима Hamerhead ribozyme



вторичная структура транспортной РНК Phe организма *Escherichia coli*

Рис. 4. Исходные данные метода «Медленное сжатие» на примере нахождения третичной структуры пяти различных фрагментов

Жирными линиями отмечены водородные связи, линиями с квадратами на концах – неканонические водородные связи между спиралями

Результаты применения метода «Медленное сжатие» с разными стратегиями

	St1 (№1)	St2 (№2)	St1x2	St2x2
L1	0-0-0-0-0-0	0-0-0-0-0-0	7.4-1.6-16.6-26.6-36.6	0-0.3-0-8.5-0-0
L2	0-1.7-0.8-0-0	0-0-11.8-21.8-31.8	0-0-0-0-0	0-0-0-0-0
D loop	0-0-0-0-0	0-0-0-0-0-0	0-0-0-0-0.2	3.0-0-0-0-0
A loop	10.8-20.8-30.8-40.8-50.8	0-0-0-0-0-0	0-40.1-50.1-60.1-70.1	0-0-0-0-0
T loop	0-0-1.8-0	30.3-40.3-50.3-60.3	0-0-0.2-9.5	12.6-22.6-32.6-42.6
	St0 Full	St1 Full	St2 Full	
L1	15.9-25.9-35.9-45.9-55.9-65.9	4.5-0-0-4.1-2.5-5.0	0-0-0-0-0-0	
L2	0-9.5-1.8-0-0	0-0-0-3.8-0	0-0-0-5.1-0	
D loop	10.0-8.2-0-2.3-7.2	0-0-0-0-0	0-0-0-0-0	
A loop	1.1-0-0-0-0	0-0-3.8-0-0	0-0-0-0-0	
T loop	41.5-51.5-61.5-71.5	0-0-2.6-7.7	8.4-18.2-28.2-38.2	

1. Применить стратегию St1 (она наиболее быстрая), если результат не удовлетворителен.
2. Применить стратегию St2, если результат и тут не удовлетворителен.
3. Применить модификации St2 Full или St x2.

Для четырех фрагментов были получены конечные результаты, где образованы все водородные связи, и только для Т петли были образованы водородные связи между 3 парами нуклеотидов из 4 необходимых.

5. Трудности моделирования многоспиральных цепей РНК

Следующим шагом в исследовании была попытка смоделировать многоспиральную цепь РНК, что еще в полной мере никому не удавалось. В отличие от сворачивания виридного рибозима организма *Chrysanthemum chlorotic mottle viroid* [13], которое в итоге нам удалось получить, полный ход сворачивания тРНК Phe организма *Escherichia coli* еще не установлен. Но далее показано, каким образом анализ критических точек сворачивания и их возможного расположения на пути сворачивания помогает выяснить истинный биологический путь сворачивания, ориентируясь на образование неканонических водородных связей.

5.1. Путь сворачивания и конечное третичное состояние виридного рибозима

Ряд экспериментов был выполнен для Hammerhead ribozyme. После того как фрагменты L1 и L2 были получены описанным выше методом, нужно было промоделировать сворачивание рибозима целиком. Для этого нужно было понять, что удерживает эти две спирали вместе. Оказывается, что у другого Hammerhead Ribozyme G12A, для которого был проведен рентгеноструктурный анализ [11], существуют неканонические водородные связи:

1. (с32-g37) - направляющая вращение одной спирали вокруг второй;
2. (a46-u19 и a46-u24) - фиксирующие две спирали компактно одну возле другой с помощью водородной связи между тремя нуклеотидами, так называемый тример, а не обычный димер.

Проведя сравнение первичных последовательностей и выравнивание вручную, ориентируясь на третичную структуру, были найдены водородные связи для исследуемого нами рибозима (см. ниже).

Совпадает:

1. центральный участок, соединяющий две петли CUGAUGA;
2. достаточно полно совпадает спираль L2 (но в ней имеются множество разрывов);

```

L3          * L1    L1c    L1    Center  L2    L2c    L2    * L3
12345678901 2 345678 9012345 678901 2345678 9012 3456 789 0 123 45678901234 56789
gggagcccuGU C acCgga UGugcUu uccGgu CUGAUGA_ G_UCC gUGA_ GGA_C aAA ACagggcuccc gaauu
aagagGU C ggCacc UGacgUc gguGuc CUGAUGAa GaUCC aUGAca GGAuC gAA ACcucuu
1234567 8 901234 5678901 234567 89012345 67890 123456 78901 234 5678901

```

3. есть совпадения в петле L1c (центр спирали L1);

4. есть совпадения в главной петле (отмечено*).

Не совпадает:

1. часть спирали L1, за исключением петли;

2. часть спирали L3, за исключением начала и главной петли.

Таким образом, для сохранения третичной структуры (а значит и функций рибозима) важно совпадение петель между спиралями, в то время как сами спирали могут быть разными. Но не произвольно разными, а попарно замененными (хоть и не полностью, но преимущественно, чтобы образовывать классические пары g-c, a-u).

Поэтому при моделировании требуется образовать дополнительные водородные связи c28-g33, a44 - u15, a44 - u20. Образование связи c28 - g33 описанным в данной статье способом не составило труда, это позволило существенно ограничить свободное вращение спиралей друг вокруг друга, запрещая множество зеркальных структур, которые теоретически можно было бы образовать без данной водородной связи.

Трудность составляет образование пары водородных связей a44 - u15, a44 - u20. Одна причина в том, что петли спиралей подходят так близко, что если бы они были хоть немного неточно промоделированы, то проверка на запрещенные ковалентные связи не позволяет требуемым нуклеотидам подойти достаточно близко друг к другу, чтобы образовать нужную водородную связь. Основная же проблема связана с конструированием водородной связи a44 - u20. Дело в том, что наличие связей a44 - u15, и главное, связей спиралей c14 - g22, c13 - g23, c40 - g47, c39 - g48, так сильно ограничивает возможность вращения нуклеотидов между нуклеотидами u20 и a44, что невозможно занять требуемые положения никакими возможными последовательными поворотами.

Но в данный момент уже получена модельная траектория сворачивания виридного рибозима организма *Chrysanthemum chlorotic mottle viroid* [13]. Основная трудность состоит в том, что в терминологии теории игр называется динамичностью функции пригодности. Дело в том, что за время сворачивания РНК проходит ряд стадий, в которых меняются взаимодейст-

вия, вызванные вновь образованными водородными связями. Поэтому из-за таких *критических точек*, в которых должна изменяться функция пригодности, невозможно полностью автоматизировать ход сворачивания. Кроме того, сам вид функции пригодности неизвестен и строится, исходя из ряда эвристик. Но необходимость образовать все ключевые водородные связи в РНК налагает ряд ограничений, соблюдение которых указывает наиболее вероятную траекторию сворачивания, опираясь на определенные *критические точки*. Таким образом, становится возможным судить о модели сворачивания, и становится понятно, что так называемые иерархические модели (hierarchical model, или их вариации, такие как framework model [12]) это лишь грубое приближение к действительности. Такие выводы делаются на основании принципиальной возможности создать условия для образования определенной комбинации водородных связей. Основную логику такого анализа мы проследим далее на примере сворачивания транспортной РНК, которая переносит фенилаланин (Phe) организма *Escherichia coli*.

5.2. Особенности сворачивания третичной структуры тРНК

Неканонические водородные связи между спиралями налагают достаточно жесткие ограничения на всю структуру тРНК (Рис. 4, справа). Можно сказать, что они даже предопределяют стабильную структуру тРНК, а также последовательные стадии сворачивания. А-спираль не имеет нестандартных водородных связей. Поэтому форма петли этой спирали, образованная нуклеотидами с номерами 32-38, может быть достаточно свободной. Форма петли в этом случае определяется лишь такой траекторией сворачивания, которая наиболее быстро из начального случайного положения позволит образовать спираль, созданную парами водородных связей 31-39, 30-40, 29-41, 28-42, 27-43.

Следующая стадия сворачивания после образования А-спирали с участием Т- и D- спиралей достаточно сложна. Иерархическая модель сворачивания говорит нам, что вначале образуются все вторичные структуры, то есть наряду с образованием А-спирали параллельно об-

разуются Т- и D-спирали. И лишь затем формируется полная третичная структура под воздействием ряда взаимодействий, из которых образование нестандартных водородных связей занимает важное место.

Но рассмотрим другие варианты возможных путей сворачивания, проходящих через определенные критические точки. В реальной структуре Т-спирали 18-й нуклеотид обращен из петли. Такое положение 18-го нуклеотида необходимо для создания водородных связей с нуклеотидом 55 из D-петли. Также важно то, что пара нуклеотидов 13-22, начинающая спираль, расположена под диагональным углом. Но если бы рядом не присутствовала D-петля, то такое положение 18-го нуклеотида равновероятно с любым другим его положением. Например, в моделировании при отсутствии взаимодействия с D-петлей была получена модель **К**, в которой 18-й нуклеотид обращен внутрь петли, а пара 13-22 стыкуется под прямым углом. Такое скручивание Т-спирали в условиях, когда она находится вне состава тРНК более вероятно.

Поэтому если бы образование Т- и D-спиралей происходило бы одновременно и отдельно друг от друга, то Т-спираль приняла бы положение, более близкое к такой модели **К**. Также нужно учесть, что последующие взаимодействия Т- и D-спиралей после их сворачивания не могут изменить положение 18-го нуклеотида, и соответственно, становится невозможным установить водородную связь между парой 18-55.

Следовательно, мы приходим к выводу, что сворачивание идет другим путем, а не таким, как постулируют иерархические модели сворачивания. А именно, вначале формируется D-спираль, после чего 18-й нуклеотид начинает взаимодействовать с 55-м, и только в условиях такого взаимодействия и начинает сворачиваться Т-спираль.

Заключение

Биоинформационные выводы

Будет ли иметь рибозим в реальности требуемую свободу вращения в аналогичных условиях, например, как было описано выше при моделировании создания водородной связи a44 - u20, мы не знаем. Но при математических вы-

числениях (*in silico*) он этой свободы принципиально не может иметь.

Дело в том, что при моделировании РНК представляется как дерево с множеством ветвей. И когда осуществляется один поворот, например, в 21 нуклеотиде, то неподвижными остаются только нуклеотиды от 1 до 20, а все нуклеотиды после 22-го изменяют свое положение в соответствии с поворотом. Если же при этом требуется сохранить уже образованные водородные связи c14 - g22, c13 - g23, c40 - g47, c39 - g48, то это становится невозможным. Это следствие того, что эти водородные связи находятся после 21-го нуклеотида, и соответственно, в моделировании они рвутся. Возможно применение метода Concerted Rotation Algorithm (CRA) [14]. При его применении можно оставить неподвижными нуклеотиды, но тогда существенно уменьшается число углов, на которые можно провернуть нуклеотиды, а это в свою очередь не позволяет найти такое положение в цепи, когда образуются водородные связи.

Таким образом, вычислительно эту проблему обойти нельзя. Впрочем, это может быть указанием, что это не только вычислительная проблема. Обдумывая, как же обойти эту проблему, мы пришли к выводу, что в процессе сворачивания необходимо выделить несколько различных фаз.

Фаза №1. До полного сворачивания спиралей необходима ориентация нуклеотидов в петлях спиралей на основании межспиральных водородных связей. Но при этом спирали должны лишь начать формироваться. В противном случае без зачатков спиралей, ориентация нуклеотидов в местах будущих петель спиралей приведет к тому, что нуклеотиды так приблизятся, что станет невозможным формирование спиралей. А без предварительной ориентации нуклеотидов в петлях, они останутся в случайном положении, а после формирования спиралей уже будет невозможно придать им требуемую ориентацию (спираль начнет разрушаться). Поэтому концепцию строго иерархических моделей (называемых *hierarchical model*, или их вариации, такие как *framework model* [12]) сворачивания невозможно использовать, как минимум, для моделирования (т.е. *in silico*).

Фаза №2. Полное образование спиралей. Если они закончат формироваться, то станет существенно затруднено (вплоть до невозможно-сти) изменение ориентации нуклеотидов, находящихся в петле спирали (так как это приводит к развалу спирали). А это, в свою очередь, не дает сформироваться межспиральным связям (нуклеотиды не могут подойти на такое близкое расстояние, не столкнувшись с другими атомами, так как это вызывает образование запрещенных ковалентных связей).

Фаза №3. Образование неканонических межспиральных водородных связей.

И, кроме того, указанные фазы могут циклически сменять друг друга до полного формирования третичной структуры РНК.

Кибернетические выводы

По-видимому, следует обратить внимание на то, что с течением времени при моделировании сворачивания макромолекул (белков, РНК) меняются различные парадигмы. Если ранее такое моделирование чаще исходило из биофизических предпосылок и использовались методы молекулярной динамики, то сейчас все чаще моделирование полностью отходит от физической составляющей процесса. Физические формулы заменяются приблизительными оценками энергии, которые основаны уже не на расчете реальной энергии, а на ее оценке, исходя из чисто геометрических соотношений между молекулами, их ориентаций и прочих коэффициентов, которые не имеют прямого отношения к физическому процессу.

В данной работе мы идем еще дальше – отказываемся от необходимости делать оценку энергии и затем ее минимизировать. Это связано с тем, что, основываясь на приблизительных и статистических оценках, невозможно получить такую точность, чтобы в моделировании атомы подошли друг к другу так, чтобы выполнялись условия образования всех необходимых водородных связей.

Далее, мы идеализируем процесс сворачивания, не рассматривая никакие другие взаимодействия, кроме водородных связей. Таким образом, в моделировании мы намеренно исходим из упрощения, идеализации, как бы отвечая на вопросы: «как пойдет ход сворачивания, если

РНК будет стремиться только к образованию водородных связей?» и «каков “чистый” вклад образования водородных связей в процесс сворачивания?».

Оказывается, что для моделирования односпиральных фрагментов РНК не нужно более ничего учитывать, и достаточно только образовывать водородные связи с той лишь погрешностью, что форма петли спирали будет произвольна. Но уточнить ее можно лишь в зависимости от того, какие другие фрагменты окружают эту спираль. В тоже время для многоспиральных фрагментов РНК это уже не совсем справедливо. Как минимум обязательно учитывать ряд неканонических водородных связей между различными спиралями.

С другой стороны, в случае многоспиральных фрагментов мы видим, что принципиально важен совершенно другой, не биологический, а математический аспект проблемы. Это ставит перед нами задачу математического доказательства пригодности предлагаемых методов для моделирования сворачивания. И это можно и нужно сделать сугубо кибернетическими и математическими подходами, без введения конкретной биологической проблематики.

В терминологии теории игр это означает, что нужно обеспечить возможность из любого начального положения игры достичь заданного конечного состояния. И если, например, в игре в шахматы мы знаем, как начинается игра и какое условие выигрыша, и знаем, что правилами шахмат обеспечено попадание из начала игры в конец, то здесь - при моделировании сворачивания макромолекул - нам нужно еще найти такие правила и доказать, что они обеспечивают процесс сворачивания.

В данной работе мы указали, что основной проблематикой нахождения таких правил является обеспечение гладкости поверхности, образуемой состояниями, в которых находится и изменяется цепь РНК.

Мы находимся еще в самом начале такого кибернетико-геометрического подхода, но уже сейчас можем достаточно быстро моделировать односпиральные фрагменты РНК без необходимости случайного перебора состояний. И этот подход позволяет понять как математические, так и биоинформационные основы и про-

блемы, которые мешают моделировать много-спиральные фрагменты РНК.

Литература

1. С.С. Яковлев, А. Н. Борисов. Целенаправленный поиск в задаче сворачивания третичной структуры РНК, Автоматика и вычислительная техника, 45-1, 1-10 Рига, 2011. <http://www.springerlink.com/content/j022770761177n51/> (англ.)
2. Rhiju Das, David Baker. Automated de novo prediction of native-like RNA tertiary structures // University of California, Berkeley, CA, 2007 <http://www.pnas.org/content/104/37/14664.full>
3. Jeffrey Skolnick, Andrzej Kolinski, Daisuke Kihara, Marcos Betancourt, Piotr Rotkiewicz, and Michal Boniecki. Ab Initio Protein Structure Prediction via a Combination of Threading, Lattice Folding, Clustering, and Structure Refinement, 2001 <http://kiharalab.org/paper/10.pdf>
4. Л. А. Зинченко, В. А. Шахнов. Особенности математического моделирования в задачах проектирования наносистем, Информационные технологии и вычислительные системы, Москва, 2009/04
5. Ж.Сетубал, Ж.Мейданис. Введение в вычислительную молекулярную биологию. НИЦ "Регулярная и хаотическая динамика", Институт компьютерных исследований, Серия: Биоинформатика и молекулярная биология, Москва-Ижевск, 420 с., 2007
6. Р.Л.Кини, Х.Райфа. Принятие решений при многих критериях: предпочтения и замещения. Москва, «Радио и связь», 560 с., 1981
7. I.L. Hofacker, W. Fontana, P.F. Stadler, S. Bonhoeffer, M. Tacker, P. Schuster. Fast Folding and Comparison of RNA Secondary Structures. Monatshefte f. Chemie 125: 167-188, 1994
8. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. Department of Molecular Biophysics & Biochemistry and Howard Hughes Medical Institute, New Haven, CT 06520-8114, USA. 2000. <http://www.pdb.org/pdb/explore/explore.do?pdbId=1FFK>
9. D.J. Klein, T.M. Schmeing, P.B. Moore, and T.A. Steitz, The kink-turn: a new RNA secondary structure motif, 2001. <http://www.pdb.org/pdb/explore/explore.do?pdbId=1JJ2>
10. US National Center of Biotechnological Information (NCBI) <http://www.ncbi.nlm.nih.gov/sites/entrez?Db=genome&Cmd=ShowDetailView&TermToSearch=16067>
11. Young-In Chi, Monika Martick, Monica Lares, Rosalind Kim, William G Scott, and Sung-Hou Kim, Capturing Hammerhead Ribozyme Structures in Action by Modulating General Base Catalysis, 2008. <http://www.pdb.org/pdb/explore/explore.do?structureId=2QUS>
12. Птицын О.Б. Стадийный механизм самоорганизации белковых молекул. Докл. АН СССР, 1973. Т.210
13. С.С. Яковлев, Модель сворачивания вириодного рибозима NC_003540 организма *Chrysanthemum chlorotic mottle viroid*, представлена в свободном доступе <http://www.youtube.com/watch?v=F5GHJ4eDnEU>
14. Jakob P. Ulmschneider and William L. Jorgensen, Monte Carlo backbone sampling for polypeptides with variable bond angles and dihedral angles using concerted rotations and a Gaussian bias, Chem. Phys. 118, 4261 (2003); doi: 10.1063/1.1542611, 2002

Борисов Аркадий Николаевич. Профессор кафедры имитации и моделирования Рижского технического университета. Окончил Рижский политехнический институт в 1964 году. Профессор, doctor habilitatus (Dr.habil.sc.comp.). Автор более 210 работ. Область научных интересов: нечеткие множества, нечеткая логика и методы вычислительного интеллекта
E-mail: Arkadijs.Borisovs@cs.rtu.lv

Яковлев Сергей Сергеевич. Проектировщик программного обеспечения, ведущий программист Латвийско-американского предприятия разработки коммерческих банковских систем SIA "DSW". Окончил Рижский технический университет в 2004 году. Автор 9 статей. Область научных интересов: искусственный интеллект, биоинформатика.
E-mail: tac@inbox.lv