

Воронина Валерия Вадимовна

канд. техн. наук, доцент

ФГБОУ ВО «Ульяновский государственный
технический университет»

г. Ульяновск, Ульяновская область

Антонова Елена Ивановна

д-р биол. наук, профессор, директор

Научно-исследовательский центр фундаментальных

и прикладных проблем биоэкологии и биотехнологии

ФГБОУ ВО «Ульяновский государственный педагогический
университет им. И.Н. Ульянова»

г. Ульяновск, Ульяновская область

ИДЕНТИФИКАЦИЯ И КЛАССИФИКАЦИЯ СТРУКТУРНЫХ МОТИВОВ ДЛИННЫХ НЕКОДИРУЮЩИХ РНК

Аннотация: данное исследование посвящено аннотации и анализу взаимодействий третичной структуры РНК, в частности образованных удаленными друг от друга элементами вторичной структуры РНК. Показано, что наиболее распространенным типом окружения рассматриваемых контактов является взаимодействие стема и классической шпильки, а также выявлено, что наиболее распространенным третичным мотивом, содержащим дальнедействующие контакты, является мотив типа А-минор.

Ключевые слова: вторичная структура РНК, третичная структура РНК, третичные мотивы, длинные некодирующие РНК, lncRNA.

РНК обладает различными структурными, регуляторными и ферментативными функциями, которые обеспечивают выживание, развитие организмов как в условиях нормы, так и в случае развития заболеваний [8].

РНК – полимер, который синтезируется в процессе транскрипции с лидирующей нити ДНК направления 3'→5', в связи с этим все типы РНК имеют

направление 5'→3'. В процессе транскрипции обеспечивается перекодирование генетической информации с участков ДНК в полинуклетидную последовательность РНК. Функции молекулы РНК во многом обеспечиваются её способностью складываться в пространственные структуры. Для описания структуры РНК используется иерархический подход, согласно которому выделяют первичную, вторичную, третичную и четвертичную структуры. Первичная структура описывает нуклеотидную последовательность, вторичная структура описывает взаимодействия комплементарных нуклеотидов, третичная структура – укладку цепи в пространстве, а четвертичная – взаимодействие РНК с другими биомолекулами. Знание вторичной структуры РНК является одним из ключевых моментов понимания механизма сворачивания РНК и сборки в дальнейшем рибонуклеопротеидных комплексов (РНП-комплексов).

Структурно-функциональная организация РНК представлена трехмерной конформацией молекулы в пространстве в виде небольших структурных мотивов, образованных как каноническими (Уотс-Криковские, комплементарные, смысловые) так и неканоническими связями, задача которых состоит в стабилизации конформации молекулы [13]. В отличие от ДНК, двойные спирали РНК не обладают структурным полиморфизмом. Они могут принимать только спиральную конформацию А-формы с очень узкой большой бороздкой и неглубокой и широкой малой бороздкой, в которой малые или бороздчатые стороны А:U и G:C. Вместо этого РНК образует ряд структурных паттернов за счет спаривания оснований и укладки оснований, порождая различные мотивы и складки как строительные блоки трехмерной организации: двойные спирали, шпильки, внутренние петли, петли поцелуев, псевдоузлы, коаксиальные стопки и т. д.

В течение многих лет РНК рассматривалась в качестве посредника в транскрипции и трансляции. Однако в настоящее время хорошо известно, что транскриптом человека имеет более сложную организацию [8]. Белок-кодирующие последовательности (гены) составляют менее 2% от всего генома человека, а более 90% генома (некодирующая часть генома) человека активно транскриби-

руется в особый класс РНК – long non-coding RNA (lncRNA). Хотя раньше полагали, что эти 90% – транскрипционный «шум» или «эволюционный мусор», возникающий из-за вставки мобильных генетических элементов, последние данные свидетельствуют о том, что lncRNA могут играть важную биологическую роль в развитии и функционировании организмов [8; 18]. Изучение структурно-функциональной организации данного класса РНК во многом расширит понимание их роли в клеточной биологии и развитии заболеваний человека, а также на потенциальной возможности использования их в качестве биомаркеров. Следует отметить, что lncRNA – это транскрипты, не кодирующие белок и имеющие длину более 200 нуклеотидов [16]. Это позволяет их отличать от других видов нкРНК, таких как микроРНК (miRNA) – коротких интерферирующих РНК (siRNA), взаимодействующих с белком piwiРНК (piRNA), малых ядерных РНК (snoRNAs) и некоторых других коротких РНК.

В настоящее время идентифицированная вторичная структура РНК может быть получена в основном с помощью биологических экспериментов, таких как рентгеноструктурный анализ и ЯМР. Однако биологические экспериментальные методы дороги и трудны при измерении структур в больших масштабах [3; 10].

Учитывая вышесказанное, исследование пространственной структуры некодирующих РНК является важным и перспективным научным направлением [11]. Пространственные структуры таких молекул нередко определяют выполняемые ими функции. Однако технологии определения пространственных структур РНК сильно отстают от экспериментальных методов определения последовательности нуклеотидов по скорости генерации данных. Развитие высокопроизводительного секвенирования позволило разработать новые техники для экспериментального определения вторичных структур [15], открыв новый пласт задач по предсказанию пространственной структуры РНК, основываясь на данных о последовательности, а также информации о вторичной структуре. Для повышения качества решения таких задач можно опираться на консервативные мотивы в известных структурах РНК [20], которые часто служат при-

знаками и в задачах полногеномного поиска РНК-генов [17]. К тому же аннотация мотивов и анализ их структурного окружения улучшит понимание механизмов их образования и их ролей в биологических процессах. Например, отмечается важность мотива А-минор в действиях антибиотиков [9].

Таким образом, целью нашей работы является методами математического моделирования и статистики идентифицировать и классифицировать структурные мотивы вторичной структуры lncRNA. Аннотировать контакты, провести сравнительный анализ известных типов взаимодействий и мотивов, дополнить существующий массив знаний о структурах lncRNA из банка PDB [19] используя разметку дальнедействующих контактов (long-range interactions, между нуклеотидами) элементов вторичной структуры lncRNA на основе оригинальной классификации элементов вторичной структуры РНК [1].

Задачи исследования:

1) согласно разметке URSDb об элементах вторичной структуры провести сбор информации по дальнедействующим взаимодействиям между нуклеотидами в РНК: информация о самой паре нуклеотидов, о типах молекул РНК, об организмах [2];

2) построить графики распределения расстояний между атомами нуклеотидов против расстояния между нуклеотидами в последовательности для проверки валидности данных;

3) с помощью программы DSSR провести сбор информации о пересечении дальнедействующих контактов с третичными мотивами [21].

4) выявить наиболее часто встречающиеся пары дальнедействующих взаимодействий между элементами вторичной структуры и их пересечение с третичными мотивами.

В основе исследования лежит оригинальная математическая модель [1]. В настоящее время вторичная структура РНК описывается моделью Цукера-Мэтьюза-Тернера (модель ближайшего соседа, Nearest Neighbor Model, NNM [22]). Согласно этой модели, вторичная структура РНК разбивается на петли

(loops) и участки двойной спирали (стемя, stems). Основной задачей, для которой была создана эта модель, является задача прогнозирования вторичной структуры путем минимизации её свободной энергии. Но в контексте данной задачи указанная модель не применима к структурам РНК, содержащим псевдоузлы, т.е. участки с пересекающимися стемями. Их предсказание является NP-сложным [7]. Простейшие типы псевдоузлов (kissing hairpins, H-knots и др.) были ранее описаны в литературе, однако единой математической модели описания вторичной структуры при наличии псевдоузлов (подобного модели NNM для описания вторичных структур без псевдоузлов) не существовало. В работе [1] предложена новая модель вторичной структуры РНК, которая допускает псевдоузловые структуры. Также автор показывает, что в экспериментально определенных пространственных структурах функциональных РНК все псевдоузлы кроме двух простейших типов образованы с участием энергетически неустойчивых коротких стемов, состоящих из 2–3 спариваний оснований.

Согласно [1]: «Для описания третичной структуры РНК, обусловленной относительно слабыми взаимодействиями и образованной «поверх» вторичной структуры, принято использовать понятие третичного мотива: рекуррентного структурного элемента, выступающего в роли функциональной единицы и участвующего либо в стабилизации пространственной структуры, либо в узнавании других молекул [12]. Выделяют мотивы, содержащие неканонические спаривания оснований, изгибы сахара-фосфатного остова, стекинг участков двойной спирали и др. Также, различают локальные мотивы, соответствующие отдельным элементам вторичной структуры, и мотивы с дальним действием (long-range motifs), в образовании которых задействовано несколько различных петель и (или) стемов. Связь между мотивами с дальним действием и элементами вторичной структуры РНК остается слабо изученной. В работе впервые предложена классификация третичных мотивов РНК, объединяющая локальные мотивы и мотивы с дальним действием. На примере мотивов типа А-минор показано соответствие полученных структурных классов выполняемым функциям. Также

впервые показано, что в экспериментально определенных структурах РНК более половины мотивов типа А-минор существуют в кластерах».

Описанная в работе [1] модель и ее программная реализация были использованы в рамках данного исследования для идентификации и классификации структурных мотивов длинных некодирующих РНК.

В качестве исходных данных для разметки контактов были взяты файлы с кристаллографической информацией об РНК из избыточного подмножества структур, включающие структуры с разрешением не хуже 3 ангстрем ([14], версия 3.185_3.0A). Формат «cif» является стандартным текстовым форматом файла, содержащим кристаллографическую информацию, определенную Международным союзом кристаллографии [5]. Одной из секций является секция, содержащая данные о координатах атомов в пространстве. Именно на основе этой информации происходит анализ данных о структуре РНК. На рисунке 1 представлена вырезка данных этой секции.

Разметка контактов производилась средствами языка программирования Python. В работе использовались python-библиотеки pandas и matplotlib, инструмент DSSR и python-библиотека urslib2 базы данных URSDb [4].

Разметка контактов была реализована средствами urslib2, а именно методами RSS.SecStruct и RSS.Atompairs. С помощью RSS.SecStruct создается объект для каждой pdb-записи (каждого из рассматриваемых файлов). После этого для созданного объекта методом RSS.Atompairs размечаются взаимодействующие пары нуклеотидов на заданном расстоянии. Параметры вызова Atompairs – model, type1='R', type2='RDPLMW', dist=6. Здесь model – созданный на предыдущем шаге объект, type1 и type2 – типы взаимодействующих атомов (R = RNA, D = DNA, P = Protein, L = Ligand, M = Metal, W = Water), dist – максимальное расстояние в ангстремах (нами был выбран порог в 6 ангстрем).

```

loop_
  _atom_site.group_PDB
  _atom_site.id
  _atom_site.type_symbol
  _atom_site.label_atom_id
  _atom_site.label_alt_id
  _atom_site.label_comp_id
  _atom_site.label_asym_id
  _atom_site.label_entity_id
  _atom_site.label_seq_id
  _atom_site.pdbx_PDB_ins_code
  _atom_site.Cartn_x
  _atom_site.Cartn_y
  _atom_site.Cartn_z
  _atom_site.occupancy
  _atom_site.B_iso_or_equiv
  _atom_site.pdbx_formal_charge
  _atom_site.auth_seq_id
  _atom_site.auth_comp_id
  _atom_site.auth_asym_id
  _atom_site.auth_atom_id
  _atom_site.pdbx_PDB_model_num
ATOM 1 O OP3 . G A 1 1 ? 39.744 -8.666 15.919 1.00 37.17 ? 1 G A OP3 1
ATOM 2 P P . G A 1 1 ? 38.290 -8.613 15.633 1.00 36.97 ? 1 G A P 1
ATOM 3 O OP1 . G A 1 1 ? 37.849 -7.214 15.826 1.00 36.90 ? 1 G A OP1 1
ATOM 4 O OP2 . G A 1 1 ? 37.654 -9.692 16.422 1.00 37.13 ? 1 G A OP2 1
ATOM 5 O "O5'" . G A 1 1 ? 38.073 -9.003 14.097 1.00 35.67 ? 1 G A "O5'" 1
ATOM 6 C "C5'" . G A 1 1 ? 38.780 -8.329 13.061 1.00 33.71 ? 1 G A "C5'" 1
ATOM 7 C "C4'" . G A 1 1 ? 38.232 -8.704 11.695 1.00 32.09 ? 1 G A "C4'" 1
ATOM 8 O "O4'" . G A 1 1 ? 38.185 -10.147 11.546 1.00 31.48 ? 1 G A "O4'" 1
ATOM 9 C "C3'" . G A 1 1 ? 36.799 -8.268 11.422 1.00 31.20 ? 1 G A "C3'" 1
ATOM 10 O "O3'" . G A 1 1 ? 36.742 -6.900 11.037 1.00 30.36 ? 1 G A "O3'" 1
ATOM 11 C "C2'" . G A 1 1 ? 36.430 -9.207 10.280 1.00 30.85 ? 1 G A "C2'" 1
ATOM 12 O "O2'" . G A 1 1 ? 36.992 -8.816 9.040 1.00 30.61 ? 1 G A "O2'" 1

```

Рис. 1. Фрагмент исходного файла структуры 3d1l.

Секция с координатами атомов

В результате работы указанных методов был получен список данных следующего вида (листинг 1).

```

atompairs= [{ 'DSSR1': 'O.G.13.', 'DSSR2': 'O.GLN.61.', 'atom1': 'OP1', 'atom2':
'CA', 'dist': 4.302, 'type': 'RP'}, { 'DSSR1': 'O.U.19.', 'DSSR2': 'O.SER.5.', 'atom1':
«C2», 'atom2': 'CA', 'dist': 4.35, 'type': 'RP'}].

```

Листинг 1. Список взаимодействующих пар

Для удобства дальнейшей обработки полученный список был конвертирован в тип DataFrame, используя библиотеку pandas. Далее, для каждой пары был определен тип взаимодействия с точки зрения вторичной структуры РНК методом NuclRelation библиотеки urslib2. Возможные виды – SM – same, i.e. within

the same element, LC – local, i.e. adjacent stem and loop, NR – neighbor, i.e. neighbor stems (one loop apart) or neighbor loops (one stem apart), LR – long-range, i.e. from distant elements, NA – n/a. Полученные данные помещаются в исходный DataFrame с ключом r. Для удаления избыточной информации были сохранены только пары нуклеотидов с минимальным расстоянием между атомами. Затем таблица была дополнена следующими параметрами:

- расстояние между взаимодействующими нуклеотидами в цепи;
- типы элементов вторичной структуры, к которым принадлежат взаимодействующие нуклеотиды (возможные типы: S = Stem, H = Hairpin, B = Bulge, I = Internal loop, J = Multiple junction, C = Classical, I = Isolated, P = Pseudoknotted);
- информация о молекулах;
- информация об организмах.

Код описанной обработки представлен в листинге 2.

```
df_dist = pd.DataFrame(atompairs)
df_dist['r'] = df_dist.apply(lambda row: model.NuclRelation(row['DSSR1'],
row['DSSR2']), axis=1)
df3 = df2.groupby(['DSSR1', 'DSSR2', 'type', 'r', 'file'])['dist'].min().reset_index()
df3['SeqDist'] = df3.apply(lambda row: model.SeqDist(row['DSSR1'],
row['DSSR2']), axis=1)
df3['NuclSS1'] = df3.apply(lambda row: model.NuclSS(row['DSSR1']), axis=1)
df3['NuclSS2'] = df3.apply(lambda row: model.NuclSS(row['DSSR2']), axis=1)
df3['molecules1'] = df3.apply(lambda row: model.molecules[model.chains[row['DSSR1'].split('.')[0]]['MOL_ID']]['MOLECULE'],
axis=1)
df3['molecules2'] = df3.apply(lambda row: model.molecules[model.chains[row['DSSR2'].split('.')[0]]['MOL_ID']]['MOLECULE'],
axis=1)
```



```

df3['organism1'] = df3.apply(lambda row: model.molecules[model.chains[row['DSSR1'].split('.')[0]]['MOL_ID']]['ORGANISM_SCI
ENTIFIC'], axis=1)
df3['organism2'] = df3.apply(lambda row: model.molecules[model.chains[row['DSSR2'].split('.')[0]]['MOL_ID']]['ORGANISM_SCI
ENTIFIC'], axis=1)

```

Листинг 2. Код обработки данных.

В итоговый набор данных вошли только пары нуклеотидов между удаленными элементами вторичной структуры РНК (тип LR). Кроме того, такие пары были аннотированы третичными мотивами, размеченными программой DSSR, а именно: мотивами А-минор, рибозными молниями и стопками стекинга [6].

Полученные результаты.

В ходе анализа 1243 файлов PDB было обнаружено 93 180 пар нуклеотидов типа LR. Среди них 77 199 внутримолекулярных контактов и 15 981 – межмолекулярных. График распределения и гистограммы показывают, что наши данные валидны. По рисунку 2 мы видим, что на расстоянии менее 1,96 ангстрем располагается незначительное количество пар.

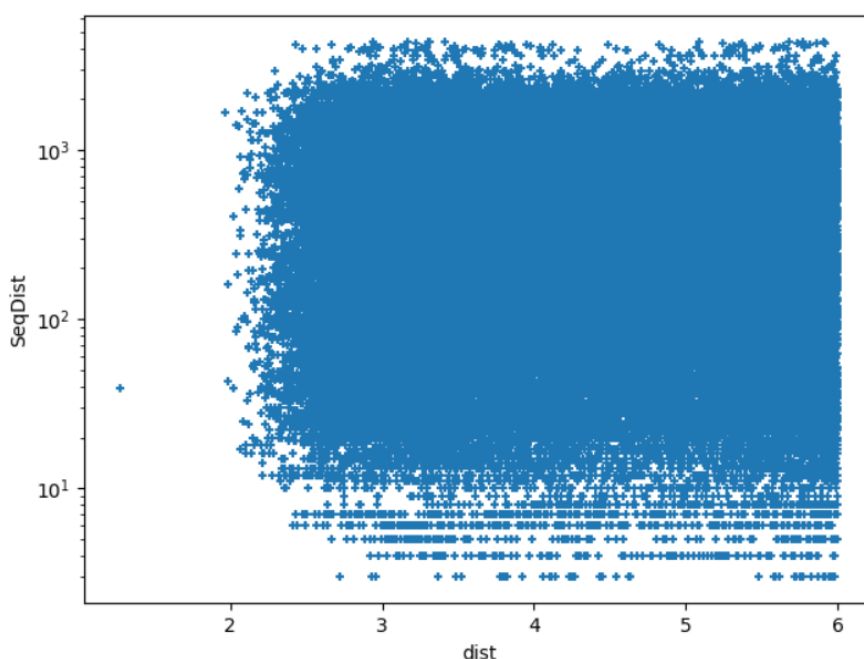


Рис. 2. График распределения контактов по двум типам расстояний.

По оси X минимальное расстояние между атомами взаимодействующих нуклеотидов. По оси Y расстояние между нуклеотидами в последовательности

В результате была удалена из анализируемых контактов пара: Vv.G.18., Vv.G.57. из файла bxa1.cif с расстоянием 1,276988 ангстрем.

На рисунках 3 и 4 представлены гистограммы расстояния между нуклеотидами в последовательности (SeqDist) и расстояния в ангстремах (dist) соответственно. Первая гистограмма показывает отсутствие в наших данных нефункциональных контактов, т.е. подтверждает, что мы рассматриваем взаимодействия между удаленными друг от друга нуклеотидами цепи.

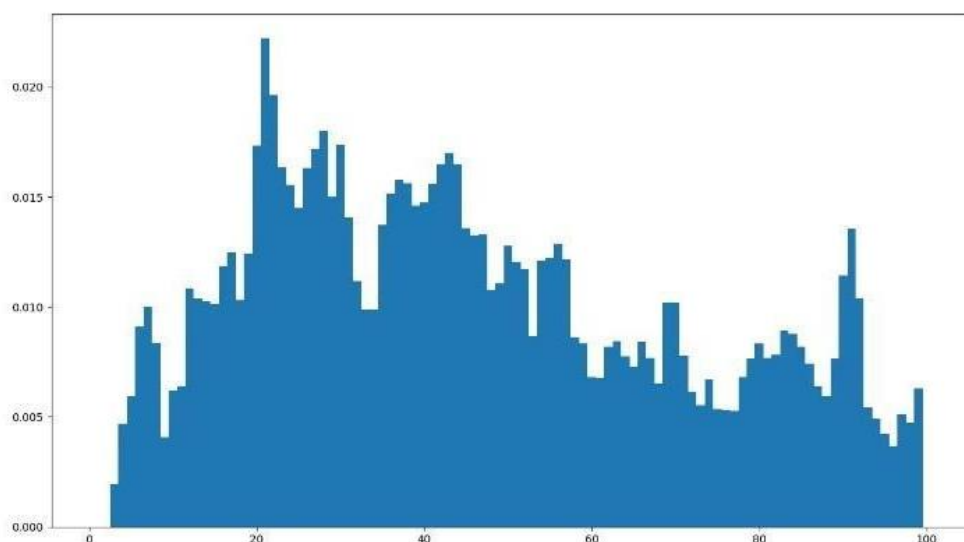


Рис. 3. Гистограмма расстояния между нуклеотидами в последовательности (срез до 400 нуклеотидов)

Рисунок 4 подтверждает гипотезу, что большинство контактов должно быть образовано водородными связями (водородная связь имеет характерную длину 3 ангстрем, а водородная связь через воду – 6 ангстрем).

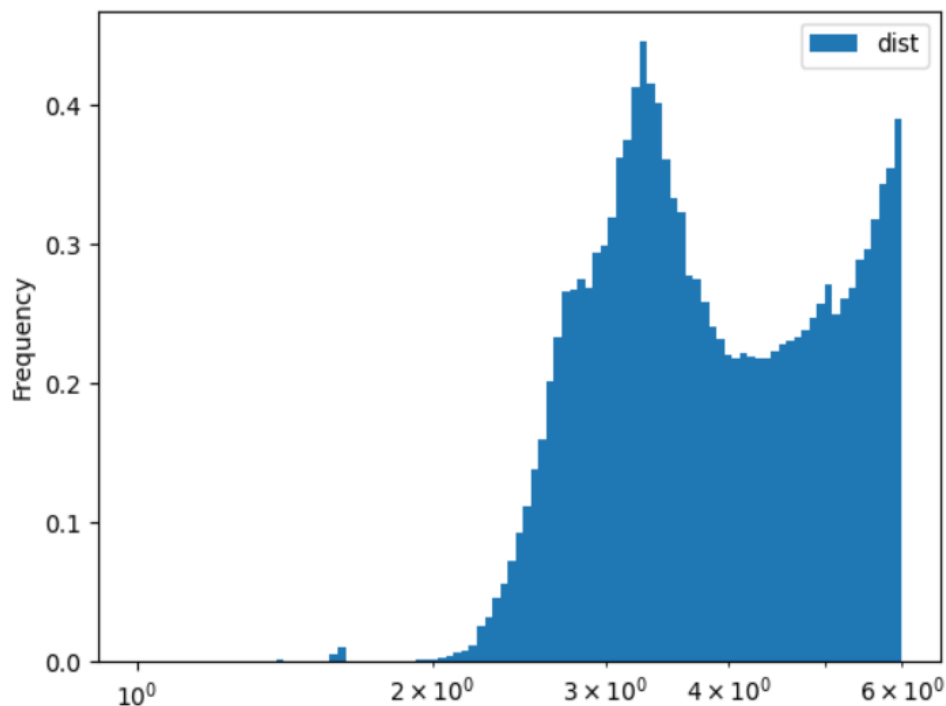


Рис. 4. Гистограмма минимального расстояния между атомами взаимодействующих нуклеотидов в ангстремах

В ходе исследования мотивов была выявлена неточность работы программы DSSR в части определения мотивов типа рибозная молния. Например, для файла 3jcs.cif отмечен мотив, в котором один и тот же нуклеотид указан дважды (листинг 3).

```
{'ID': 21, 'LENGTH': 4, 'SEQ': 'UGGG', 'NUCLS': [[['1.U.1367.', 'RES', 1366],
['1.G.1368.', 'RES', 1367], ['1.G.1368.', 'RES', 1367], ['1.G.1369.', 'RES', 1368]]]}
```

Листинг 3. Ошибка аннотации третичных мотивов в программе DSSR.

Визуальное подтверждение ошибочности аннотации мотива, выполненное программой ruMol представлено на рисунке 5. Подобное обнаружено также и с данными в файлах 4v8d.cif, 5tbw.cif и 6y2l.cif.

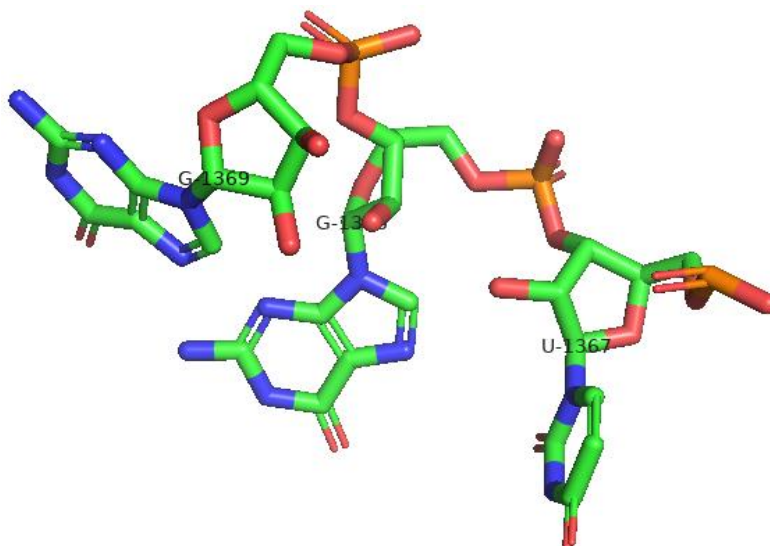


Рис. 5. Визуализация проблемного участка средствами ruMol

Если же говорить об элементах вторичной структуры, то статистика будет следующей (табл. 1). Чаще всего встречается взаимодействие стэм-классическая шпилька. Чуть реже – стэм-стэм.

Таблица 1

Статистика по элементам вторичной структуры

<i>Первый нуклеотид</i>	<i>Второй нуклеотид</i>	<i>% от общего количества</i>	<i>Класс</i>
S	HC	17,40	Наиболее часто встречающиеся
S	S	13,34	
S	IC	11,94	
JC	S	8,17	Умеренно распространенные
HC	HC	6,65	
HC	JC	4,61	Редко встречающиеся
IC	HC	4,32	

Все остальные типы взаимодействий элементов вторичной структуры могут быть отнесены к классу «Встречающиеся», так как их проценты минимальны.

Статистика по количеству найденных третичных мотивов, в которые входят оба нуклеотида LR-пары, представлена в таблице 2. Наиболее распространенным третичным мотивом, содержащим контакты типа LR, является мотив А-минор – всего около 7.2% всех пар входит в состав мотивов А-минор. Почти 4% пар нуклеотидов являются частью стопок стекинга оснований, а чуть боль-

ше 3% принадлежат рибозным молниям. Из результатов видно, что подавляющее количество рассматриваемых пар нуклеотидов не принадлежат ни одному из наиболее распространенных типов третичных мотивов РНК.

Таблица 2

Статистика по третичным мотивам

<i>Мотив</i>	<i>% пересеченных с данным мотивом пар</i>
А-миноры	7,2
стопки стэкинга	3,7
рибозные молнии	3,1

Если смотреть в разрезе семейств rfam, то по общему количеству дальнедействующих взаимодействий элементов вторичной структуры можно выделить следующие данные. На ribosomal-РНК приходится 91,5% всех взаимодействий. Остальные 8,5% – распределены по другим семействам. Если смотреть не по ribosomal-РНК, то наибольший процент нуклеотидов, принадлежащих взаимодействующим элементам вторичной структуры, приходится на RF00005 tRNA tRNA – 5,5%.

Среди них наибольшее количество следующих элементов (табл. 3, срез > 0,5%).

Таблица 3

Статистика для RF00005 tRNA tRNA

<i>Первый нуклеотид</i>	<i>Второй нуклеотид</i>	<i>Семейство rfam РНК первого нуклеотида</i>	<i>% от общего кол-ва</i>
НС	НС	RF00005 tRNA tRNA	1,76
S	S	RF00005 tRNA tRNA	0,92
НС	S	RF00005 tRNA tRNA	0,7

При этом на рибозные молнии приходится 2,7%. В разрезе элементов вторичной структуры данные следующие (см. табл. 4). На стопки стэкинга для данного семейства приходится 14,44%. В разрезе по элементам вторичной структуры данные следующие (см. табл. 5, срез > 1%).

На аминоры приходится 1,9%. В разрезе элементов вторичной структуры данные следующие (см. табл. 6, срез > 0,1%).

Таблица 4

Статистика по рибозным молниям для RF00005 tRNA tRNA

<i>Первый нуклеотид</i>	<i>Второй нуклеотид</i>	<i>Семейство rfam РНК первого нуклеотида</i>	<i>% от общего кол-ва</i>
SIP	JC	RF00005 tRNA tRNA	0,92285
S	JC	RF00005 tRNA tRNA	0,775194
SJP	JC	RF00005 tRNA tRNA	0,258398
HC	S	RF00005 tRNA tRNA	0,221484
SHP	IC	RF00005 tRNA tRNA	0,221484
S	S	RF00005 tRNA tRNA	0,147656
SHP	JC	RF00005 tRNA tRNA	0,110742

Таблица 5

Статистика по стопкам стэкинга для RF00005 tRNA tRNA

<i>Первый нуклеотид</i>	<i>Второй нуклеотид</i>	<i>Семейство rfam РНК первого нуклеотида</i>	<i>% от общего кол-ва</i>
HC	HC	RF00005 tRNA tRNA	10,7775
HC	S	RF00005 tRNA tRNA	1,738306

Таблица 6

Статистика по аминорам для RF00005 tRNA tRNA

<i>Первый нуклеотид</i>	<i>Второй нуклеотид</i>	<i>Семейство rfam РНК первого нуклеотида</i>	<i>% от общего кол-ва</i>
S	JC	RF00005 tRNA tRNA	0,56
SIP	JC	RF00005 tRNA tRNA	0,36
SHP	IC	RF00005 tRNA tRNA	0,33
HC	HC	RF00005 tRNA tRNA	0,23
SJP	JC	RF00005 tRNA tRNA	0,14

Таким образом, можно сделать следующие выводы:

1) на основе оригинальной классификации элементов вторичной структуры РНК была осуществлена аннотация дальнедействующих третичных контактов. Данные о таких контактах были систематизированы и пересечены с данными об известных типах третичных взаимодействий и мотивов в РНК. При-

мерно 17% всех контактов составили взаимодействия стэма и классической шпильки. Этому типу взаимодействий может быть присвоен класс «наиболее часто встречающиеся». Также были выявлены взаимодействия «умеренно распространенные», «редко встречающиеся» и «встречающиеся»;

2) для семейства RF00005 tRNA tRNA наиболее часто встречающимся типом взаимодействий элементов вторичной структуры является взаимодействие классических шпилек, которое входит в состав третичного мотива «стопки стэкинга»;

3) наиболее распространенным третичным мотивом, содержащим контакты типа LR, является мотив A-минор – всего около 7.2% всех пар входит в состав этого мотива. Почти 4% пар нуклеотидов являются частью стопок стекинга оснований, а чуть больше 3% принадлежат рибозным молниям;

4) около 68% LR-пар не пересеклись ни с одним из известных типов;

5) определен массив данных об элементах вторичной структуры, не входящих в известные типы третичных взаимодействия. В дальнейшем на этих данных можно осуществить поиск еще не открытых третичных взаимодействий;

б) выявлена неточность разметки мотивов средствами DSSR.

Полученная выборка взаимодействий является наиболее полной и подробной выборкой дальнедействующих третичных контактов РНК на данный момент. Полученные данные могут быть использованы в вычислительных задачах предсказания пространственной структуры РНК, в частности для уменьшения пространства поиска возможных подструктур.

Список литературы

1. Баулин Е.Ф. Классификация и идентификация структурных мотивов РНК: дис. ... канд. биол. наук. 03.01.09 – математическая биология, биоинформатика. Научные руководители: д-р физ.-мат. наук Михаил Абрамович Ройтберг, канд. физ.-мат. наук, д-р биол. наук Иван Владимирович Кулаковский. – Пущино, 2021.

2. Eugene Baulin, Victor Yacovlev, Denis Khachko, Sergei Spirin, Mikhail Roytberg, URS DataBase: universe of RNA structures and their motifs, Database, Volume 2016, 2016.

3. Novikova I.V., Hennelly S.P., Sanbonmatsu K.Y. Sizing up long non-coding RNAs: do lncRNAs have secondary and tertiary structure? // *Bioarchitecture* 2, 2012. 189–199. – doi: 10.4161/bioa.22592

4. Annotation of the local context of RNA secondary structure improves the classification and prediction of A-minors, Anna A. Shalybkova, Darya S. Mikhailova, Ivan V. Kulakovskiy, Iliia I. Fakhranurova, Eugene F. Baulin. *RNA* August 2021 27: 907–919. Published in Advance May 20, 2021.

5. Adams P.D., Afonine P.V., Baskaran K., Berman H.M., Berrisford J., Bricogne G., Brown D.G., Burley S.K., Chen M., Feng Z., Flensburg C., Gutmanas A., Hoch J.C., Ikegawa Y., Kengaku Y., Krissinel E., Kurisu G., Liang Y., Liebschner D., Mak L., Markley J.L., Moriarty N.W., Murshudov G.N., Noble M., Peisach E., Persikova I., Poon B.K., Sobolev O.V., Ulrich E.L., Velankar S., Vonrhein C., Westbrook J., Wojdyr M., Yokochi M. & Young J. Y. (2019). *Acta Cryst. D*75, 451–454.

6. Annotation of tertiary interactions in RNA structures reveals variations and correlations. Yurong Xin, Christian Laing, Neocles B. Leontis, Tamar Schlick. *RNA* 2008. 14: 2465–2477. Published in Advance October 28, 2008.

7. Classifying RNA pseudoknotted structures / A. Condon [et al.] // *Theoretical Computer Science*. – 2004. – June. – Vol. 320. №1. – P. 35–50. – DOI: 10.1016/j.tcs.2004.03. 042.

8. Engreitz J.M., Ollikainen N., Guttman M. Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression // *Nat. Rev. Mol. Cell Biol.*, 2016, 17, 756–770 <https://doi.org/10.1038/nrm.2016.126>.

9. Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics / A.P. Carter [et al.] // *Nature*. – 2000. – Sept. – Vol. 407. №6802. – P. 340–348. – DOI: 10.1038/35030019.

-
10. Fürtig B., Richter C., Wöhnert J., Schwalbe, H. (2003). NMR spectroscopy of RNA. *Chembiochem.* 4, 936–962. – doi: 10.1002/cbic.200300700.
11. Gorodkin J., Ruzzo W. L. RNA sequence, structure, and function: computational and Bioinformatic methods. – Springer, 2014. – DOI: 10.1007/978-1-62703-709-9.
12. Hendrix D.K., Brenner S.E., Holbrook S.R. RNA structural motifs: building blocks of a modular biomolecule // *Quarterly Reviews of Biophysics.* – 2005. – Aug. – Vol. 38. №3. – P. 221–243. – DOI: 10.1017/s0033583506004215.
13. Higgs P.G. (2000) RNA Secondary Structure: Physical and Computational Aspects. *Quart. Rev. Biophys.* 33, 199–253.
14. Leontis N.B., Zirbel C.L. (2012) Nonredundant 3D Structure Datasets for RNA Knowledge Extraction and Benchmarking. In: Leontis N., Westhof E. (eds) *RNA 3D Structure Analysis and Prediction. Nucleic Acids and Molecular Biology*, vol 27. Springer, Berlin, Heidelberg.
15. Low J.T., Weeks K.M. SHAPE-directed RNA secondary structure prediction // *Methods.* – 2010. – Oct. – Vol. 52. №2. – P. 150–158. – DOI: 10.1016/j.ymeth.2010.06.007.
16. Perkel J.M. Visiting «noncodarnia» // *BioTechniques* – 2013. – Vol. 54. N6. – P. 301, 303–304.
17. Regalia M. Prediction of signal recognition particle RNA genes // *Nucleic Acids Research.* – 2002. – Aug. – Vol. 30. №15. – P. 3368–3377. – DOI: 10.1093/nar/gkf468.
18. Rossi J.J. Ribozyme diagnostics comes of age // *Chemistry & Biology* (АНГЛ.) *рус.* – 2004. – Т. 11. №7. – С. 894–895. – doi:10.1016/j.chembiol.2004.07.002.
19. Stephen K. Burley, Helen M. Berman, Charmi Bhikadiya, Chunxiao Bi, Li Chen, Luigi Di Costanzo, Cole Christie, Ken Dalenberg, Jose M. Duarte, Shuchismita Dutta, Zukang Feng, Sutapa Ghosh, David S. Goodsell, Rachel K. Green, Vladimir Guranović, Dmytro Guzenko, Brian P. Hudson, Tara Kalro, Yuhe Liang, Robert

Lowe, Harry Namkoong, Ezra Peisach, Irina Periskova, Andreas Prlić, Chris Randle, Alexander Rose, Peter Rose, Raul Sala, Monica Sekharan, Chenghua Shao, Lihua Tan, Yi-Ping Tao, Yana Valasatava, Maria Voigt, John Westbrook, Jesse Woo, Huanwang Yang, Jasmine Young, Marina Zhuravleva, Christine Zardecki, RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy, *Nucleic Acids Research*, Volume 47, Issue D1, 08 January 2019, Pages D464-D474

20. Stochastic Sampling of Structural Contexts Improves the Scalability and Accuracy of RNA 3D Module Identification / R. Sarrazin-Gendron [et al.] // *Lecture Notes in Computer Science*. – Springer International Publishing, 2020. – P. 186–201. – DOI: 10.1007/978-3-030-45257-5_12.

21. Xiang-Jun Lu, Harmen J. Bussemaker, Wilma K. Olson, DSSR: an integrated software tool for dissecting the spatial structure of RNA, *Nucleic Acids Research*, Volume 43, Issue 21, 2 December 2015, Page e142.

22. Zuker M., Mathews D. H., Turner D. H. Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide // *RNA Biochemistry and Biotechnology*. – Springer Netherlands, 1999. – P. 11–43. – DOI: 10.1007/978-94-011-4485-8_2.

23. Воронина В.В. Анализ окружения дальнедействующих контактов в структурах некодирующих РНК / В.В. Воронина, Е.Ф. Баулин [Электронный ресурс]. – Режим доступа: <https://www.9111.ru/questions/77777777775012/>