

1 Предсказание вторичных структур для отдельных последовательностей

Анализ последовательности CP000627.1/583753-583892

Последовательность РНК:

```
GCUUGGCCUUAACUCCGAGCUUACCGCGCUAAGUUUA  
AACCUUUAUAUAUGCGUUGUAAGCCAGUGACCGCUUG  
UCACAAGGGCAGAAUUGGAAAUGAUUUUGCCUCCCGU  
AUUUGGAAAGGUGUUCUGUGGCGCAACAA
```

В процессе вычислений были построены две вторичные структуры с использованием методов минимальной свободной энергии (MFE) и центроидного метода.

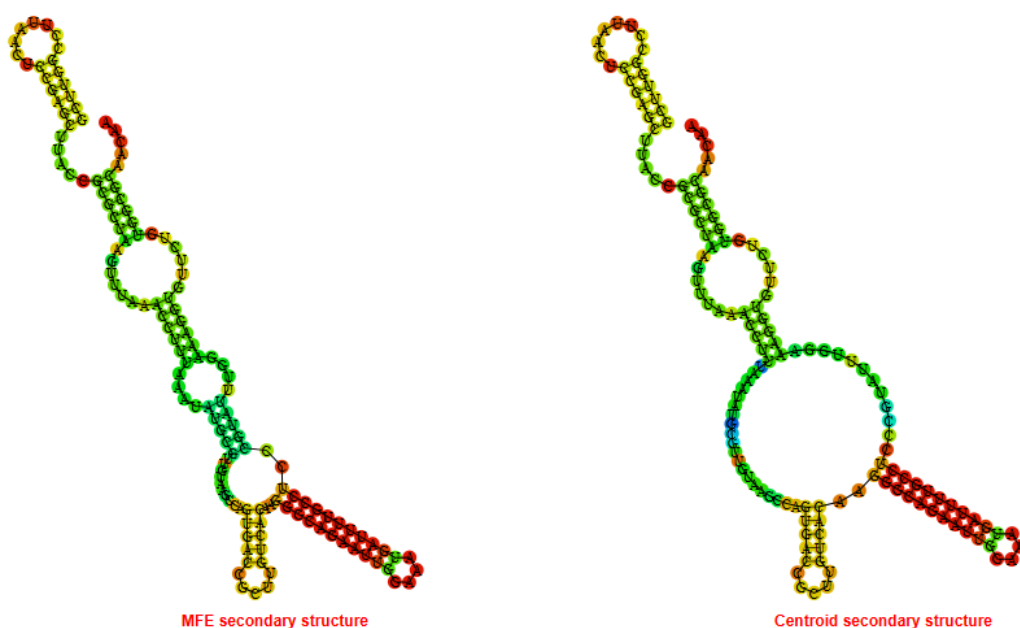


Рис. 1: Вторичные структуры CP000627.1/583753-583892

При анализе обеих структур, сгенерированных [RNAfold](#), результат для CP000627.1/583753-583892 можно посмотреть [здесь](#), [Vienna record MFE](#), [Vienna record Centroid](#), можно сделать следующие наблюдения:

При прогнозировании вторичной структуры РНК широко используются как метод минимальной свободной энергии (MFE), так и метод центроида, но они различаются по целям оптимизации и полученным структурам. Ниже приведен анализ этих методов на основе предоставленных данных:

- Структура MFE (минимальная свободная энергия) соответствует конфигурации, которая имеет минимально возможную свободную энергию. Для этой последовательности структура MFE имеет свободную энергию: -43.90 ккал/моль
(((((((.....)))))).....(((((((..... .(((((((.....((((.....((((.....))))))..((((((((((((.....))))))))))..)))))....)))))).....)))))).....
- Структура центроида представляет собой структуру, которая минимизирует расстояние до всех структур в термодинамическом ансамбле. Обычно ее считают средней структурой. Здесь структура центроида имеет более высокую свободную энергию по

сравнению с MFE -37.50 ккал/моль
 ((((((.....)))))).....((((..... .((((.....((((.....)
))))..(((((((((((.....))))))))))..... ..))))).....)))))).....

Обозначения с точками и скобками показывают различия в моделях спаривания оснований. В структуре MFE существует более обширная конфигурация спаривания оснований, что приводит к более отрицательной (более низкой) свободной энергии. Однако структура центроида представляет собой конфигурацию, которая более типична для всего ансамбля, а не является наиболее энергетически стабильной.

Это различие возникает из-за того, что структура MFE направлена исключительно на минимизацию свободной энергии, в то время как структура центроида направлена на представление "средней" структуры в ансамбле, которая может иметь менее обширное спаривание оснований.

Термодинамический ансамбль:

Свободная энергия термодинамического ансамбля, которая является средневзвешенной свободной энергией по всем доступным структурам, рассчитывается как: -45.73 ккал/моль

Это значение немного ниже энергии MFE, что предполагает, что ансамбль включает некоторые конфигурации с еще более низкой свободной энергией, чем сама структура MFE. Однако эти структуры встречаются с низкой вероятностью или являются частью менее стабильных областей в ансамбле. Кроме того, разнообразие ансамбля 35.27% указывает на значительную изменчивость возможных структур, отражая структурную гетерогенность в молекуле РНК.

Структура инвертированной последовательности:

Вторичная структура РНК зависит от последовательности оснований, поскольку каждое спаривание оснований и взаимодействие стекирования специфично для последовательности. Инвертированная последовательность полностью изменит потенциал спаривания оснований и, вероятно, приведет к другой вторичной структуре. Поэтому инвертированная последовательность вряд ли сформирует ту же самую структуру или будет иметь ту же самую свободную энергию, поскольку взаимодействия спаривания оснований и мотивы вторичной структуры будут перестроены.

Анализ последовательности CP000247.1/832790-832934

Последовательность РНК:

```
UUAACCACUAAACACUCUAGCCUCUGCACCUGGGUCA  
ACUGAUACGGUGCUUUGGCCGUGACAAUGCUCGUAAA  
GAUUGCCACCAGGGCGAAGGAAGAAAUGACUUCGCCU  
CCCGUAUCUGGAAAGGUGUACAUGGCUUCACAAC
```

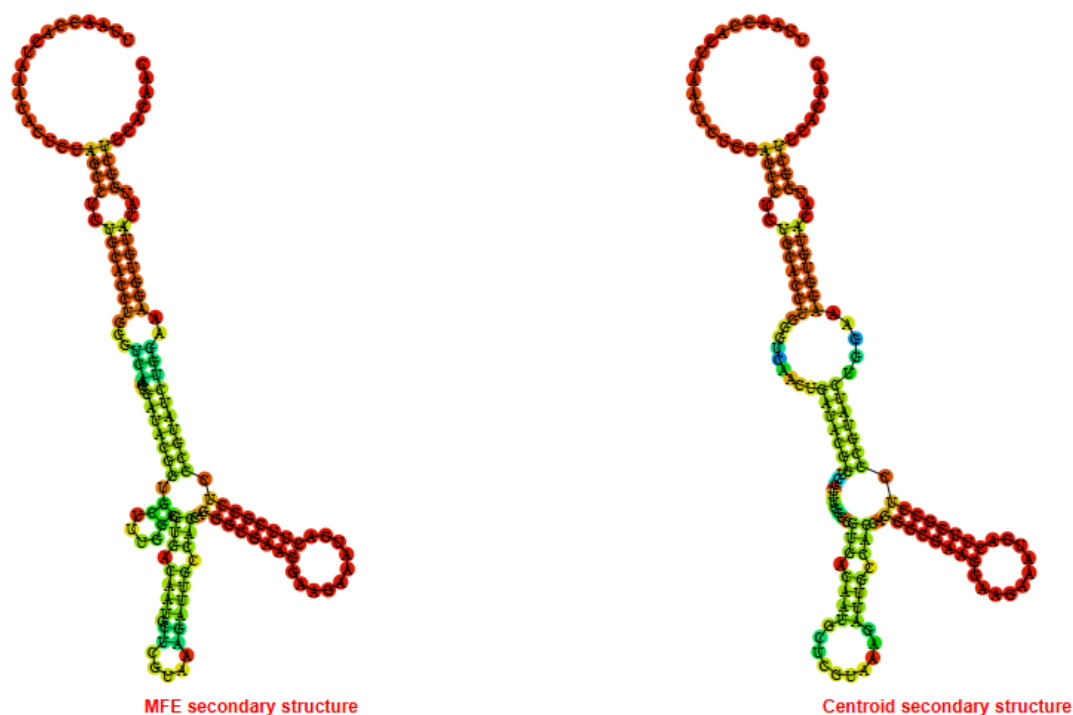


Рис. 2: Вторичные структуры CP000247.1/832790-832934

Результат и Vienna record MFE, Vienna record Centroid

Минимальная свободная энергия (MFE) и центроидные структуры предоставляют различные способы представления оптимальной вторичной структуры последовательности РНК:

- Структура MFE определяется путем нахождения конфигурации, которая дает минимально возможную свободную энергию для данной последовательности, в данном случае -41.50 ккал/моль. Эта структура отдает приоритет минимизации энергии и, как правило, уникальна, с парами оснований, расположенными для достижения минимальной энергии.

```
.....((((..(((((((...((( ...(((((((..((....)).(((..(((..((....)  
))))).))..(((((((.....)))))) .)))))))).))))))....
```

- Центроидная структура представляет собой структуру с наибольшей вероятностью в термодинамическом ансамбле. Хотя она также стремится к низкой энергии, она учитывает вероятность по распределению потенциальных структур, в результате чего получается структура со свободной энергией -38.50 ккал/моль, что немного выше, чем MFE.

```
.....((((..(((((((..... ...(((((((.....((..(((.....  
..))))).))..(((((((.....)))))) .))))))....))))))....
```

Различия между структурами MFE и центроидными структурами возникают из их критериев оптимизации. Структура MFE фокусируется на минимизации энергии независимо от стабильности всего ансамбля, в то время как структура центроида оптимизирует для наиболее вероятной конфигурации в ансамбле, включая стабильность по возможным флуктуациям. Следовательно, структура центроида может демонстрировать вариации в конфигурациях петель или спаривании оснований, которые обеспечивают баланс между энергией и структурной стабильностью, отсюда ее немного более высокая энергия.

Термодинамический ансамбль:

Энергия термодинамического ансамбля, рассчитанная как $-43,32$ ккал/моль, является средней метрикой, отражающей все вероятные структуры, которые может принять РНК. Эта энергия ансамбля обычно лежит между значениями центроида и MFE и дает представление об общей структурной стабильности в различных конфигурациях, а не о какой-либо одной структуре.

Структура инвертированной последовательности:

Инвертирование последовательности РНК, скорее всего, не приведет к идентичной структуре с теми же энергетическими свойствами. Формирование вторичной структуры в значительной степени зависит от специфичных для последовательности пар оснований и стековых взаимодействий, которые не полностью отражаются при инверсии. Следовательно, инвертированная последовательность может принять отчетливую вторичную структуру с различными характеристиками стабильности и складчатости.

2 Предсказание структуры на основе выравнивания (align&fold)

Метод align&fold позволяет оценить, какие элементы структуры сохраняются, а какие изменяются. Выравнивание позволяет выявить общие черты между последовательностями и предсказать вторичную структуру для их консенсуса.

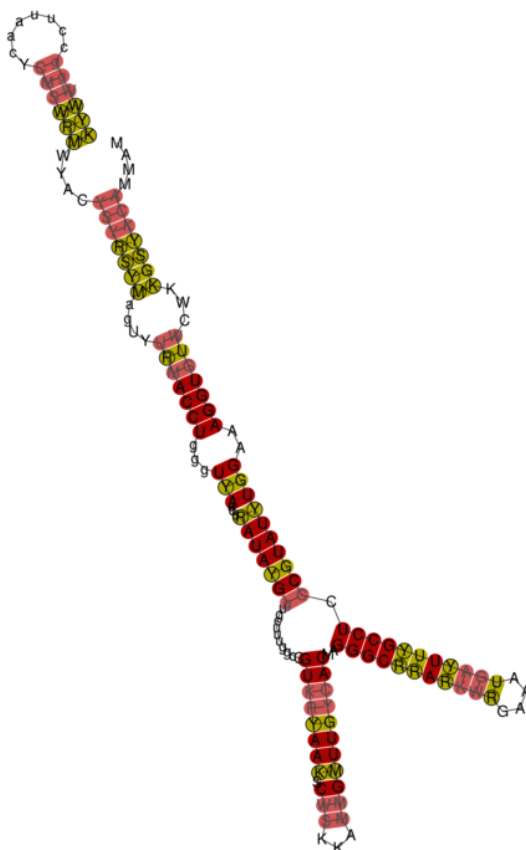


Рис. 3: Реализация метода align&fold для двух РНК с вторичными структурами

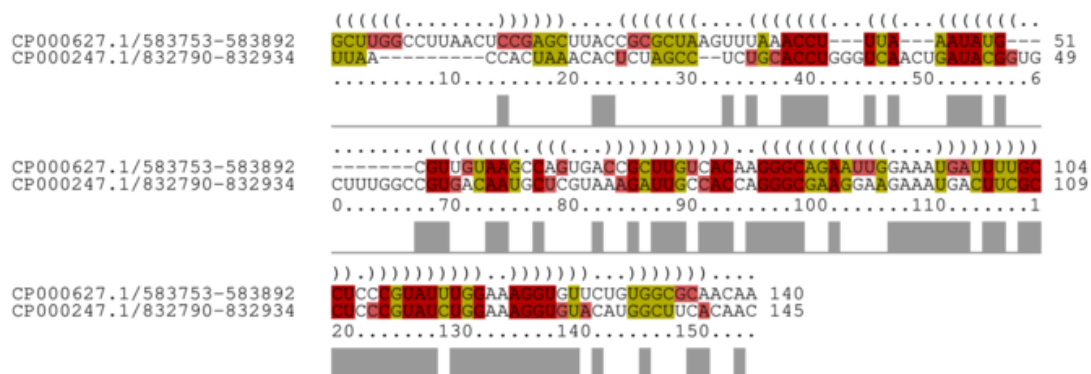


Рис. 4: Выровненные последовательности вторичных структур для двух РНК

Результаты (также можно посмотреть тут) выравнивания показали:

- Идентичные нуклеотиды, например, локусы GG, обозначены красным.
- Желтым отмечены места, где нуклеотиды различаются.

3 Сравнение структур, построенных двумя методами

[Ipyrb](#), [Plot](#)

Консервативные элементы обозначены зелёным цветом, что указывает на области, сохраняющие свою структуру в обоих РНК.

Также отмечены unknotted base pairs, свидетельствуя о том, что совместная загрузка файлов прошлого этапа, отразила общие черты структуры.

