

# БИ ДЗ-4. Вторичная структура РНК

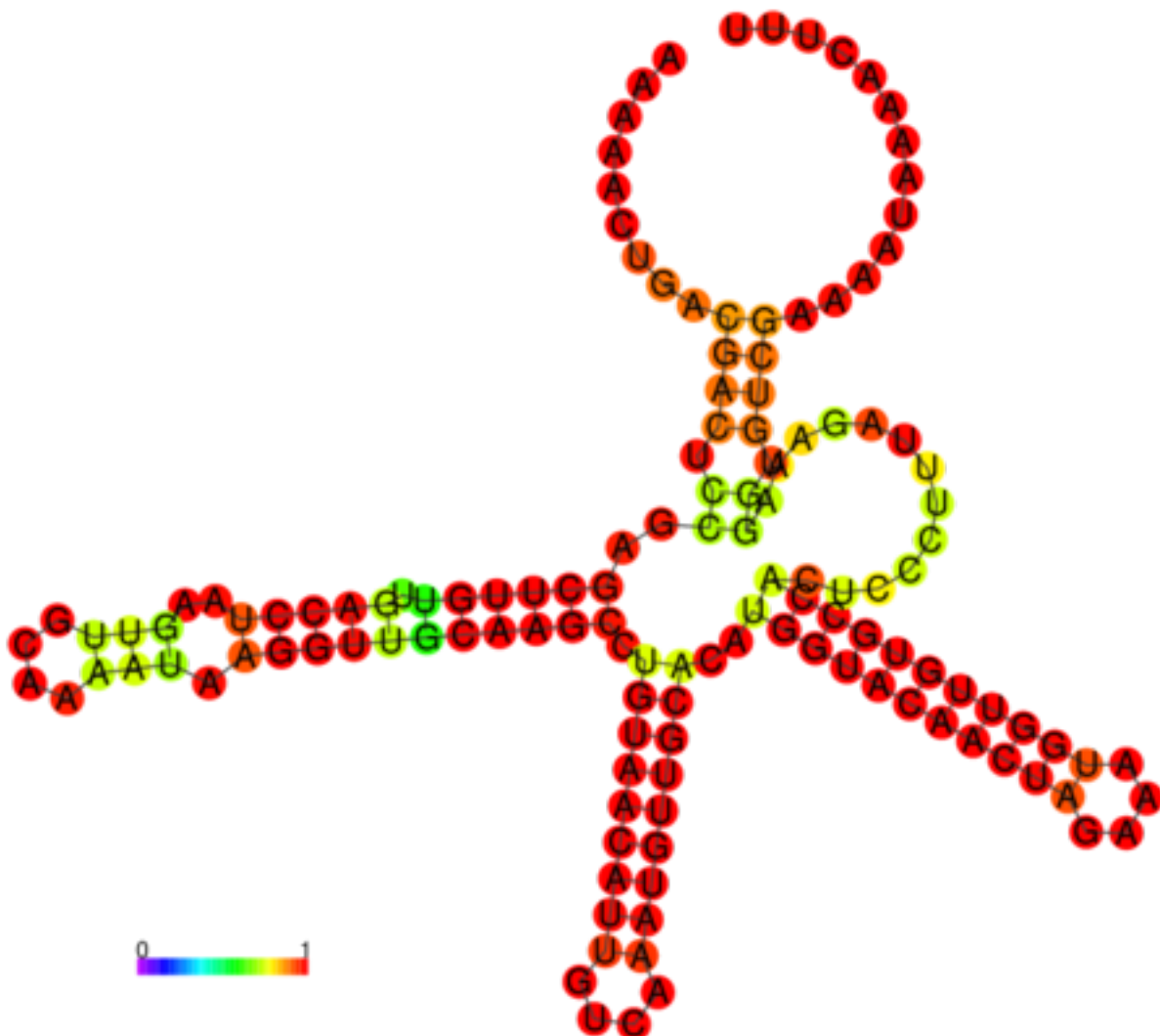
## Задание 1.

1. Взяла последнюю последовательность из файла V0.txt и сделала предсказание вторичной структуры РНК (<http://nibiru.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>).

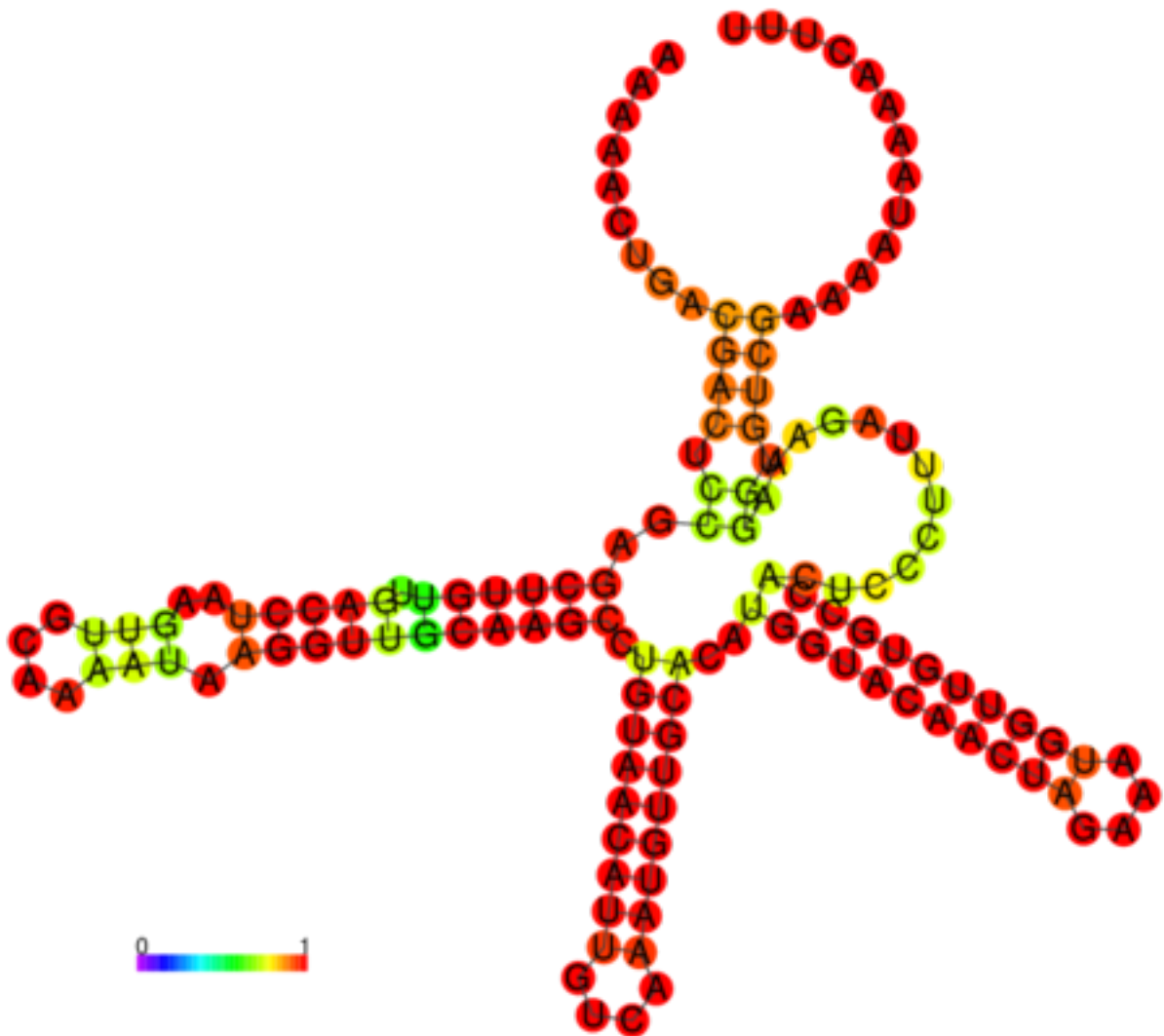
>CP000436.1/881170-881033

AAAAACUGACGACUCCGAGCUUGUUGACCUAAGUUGCAAAAUAAGGUUGCAAGCCUGUAACAUUGUCAAAUGUUGCA  
CAUGGUACAACUAGAAUGGUUGUGCCACUCCUUUAGAAAGGUGUCGAAAAUAAAACUUU

2. Скачала картинки
  - MFE structure drawing encoding base-pair probabilities



- Centroid structure drawing encoding base-pair probabilities



3. Да, одинаковые картинки получились

4. Повторяю процедуру для второй последовательности

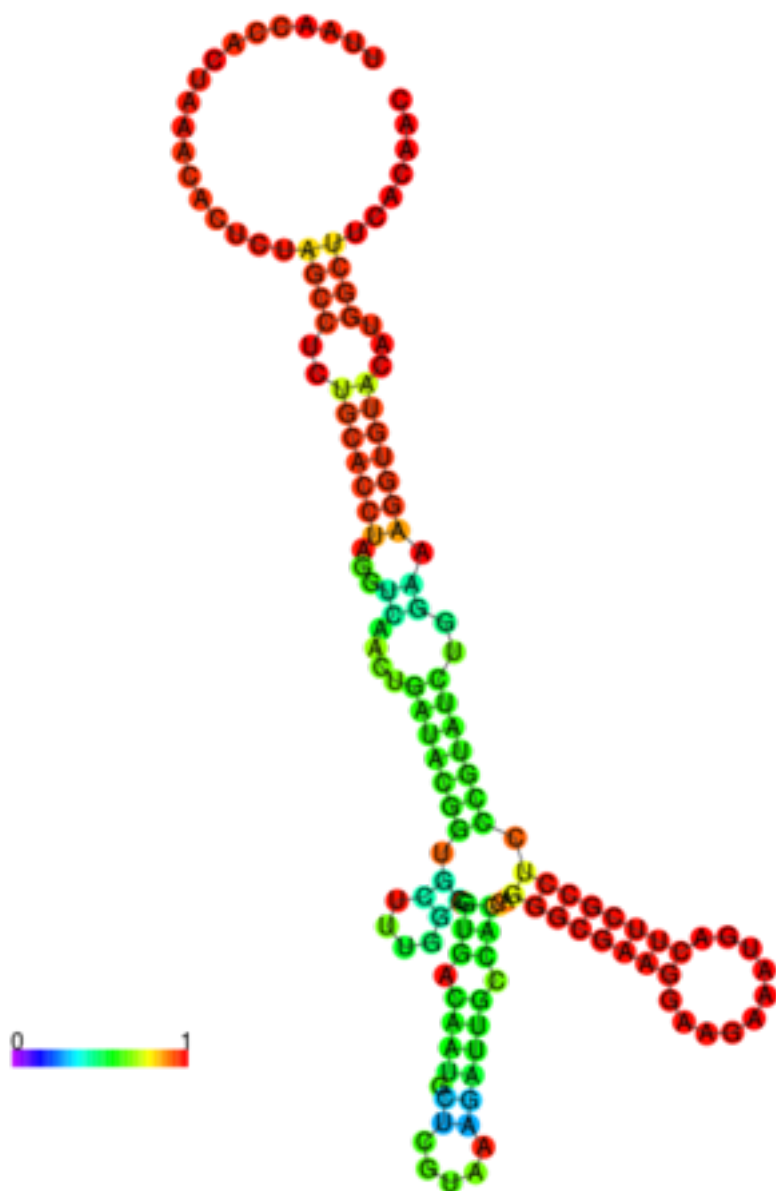
```
>CP000034.1/768837-768693
```

```
UUAACCACUAAACACUCUAGCCUCUGCACCUAGGUCAACUGAUACGGUGCUUUGGCCGUG
```

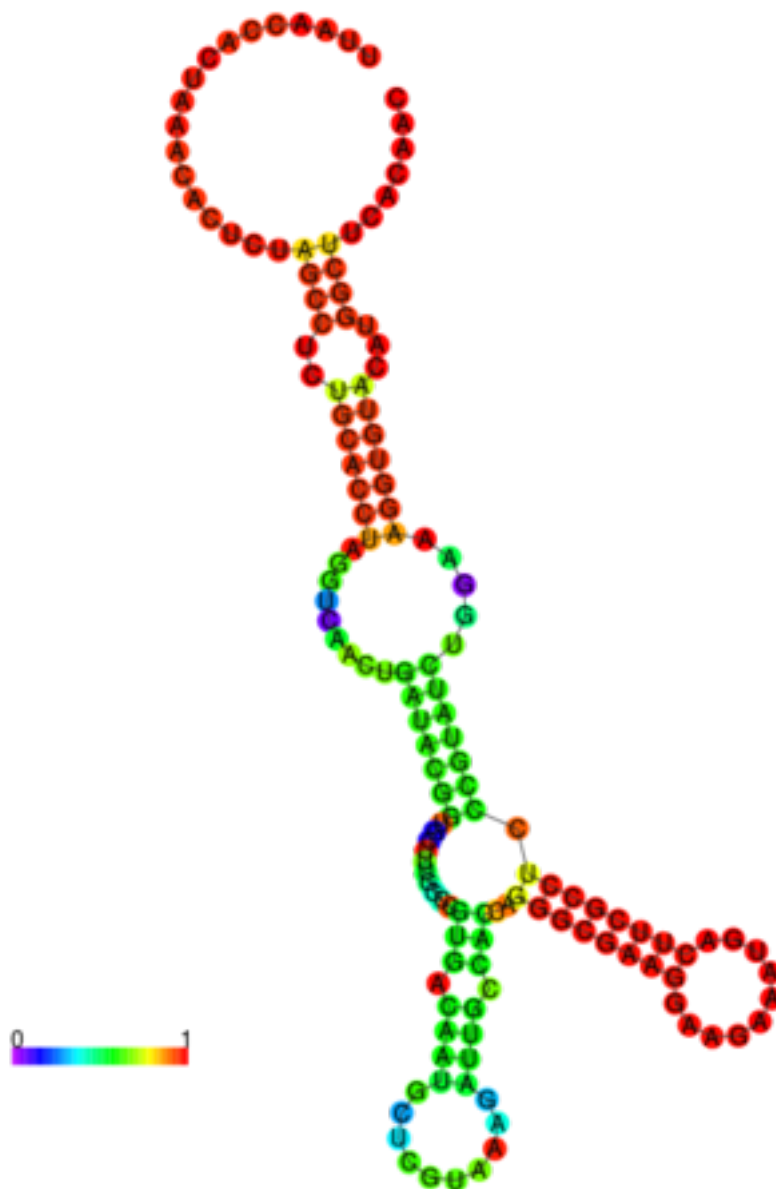
```
ACAAUGCUCGUAAAGAUUGCCACCAGGGCGAAGGAAGAAAUGACUUCGCCUCCCGUAUCU
```

```
GGAAAGGUGUACAUGGCUUCACAAC
```

- MFE structure drawing encoding base-pair probabilities



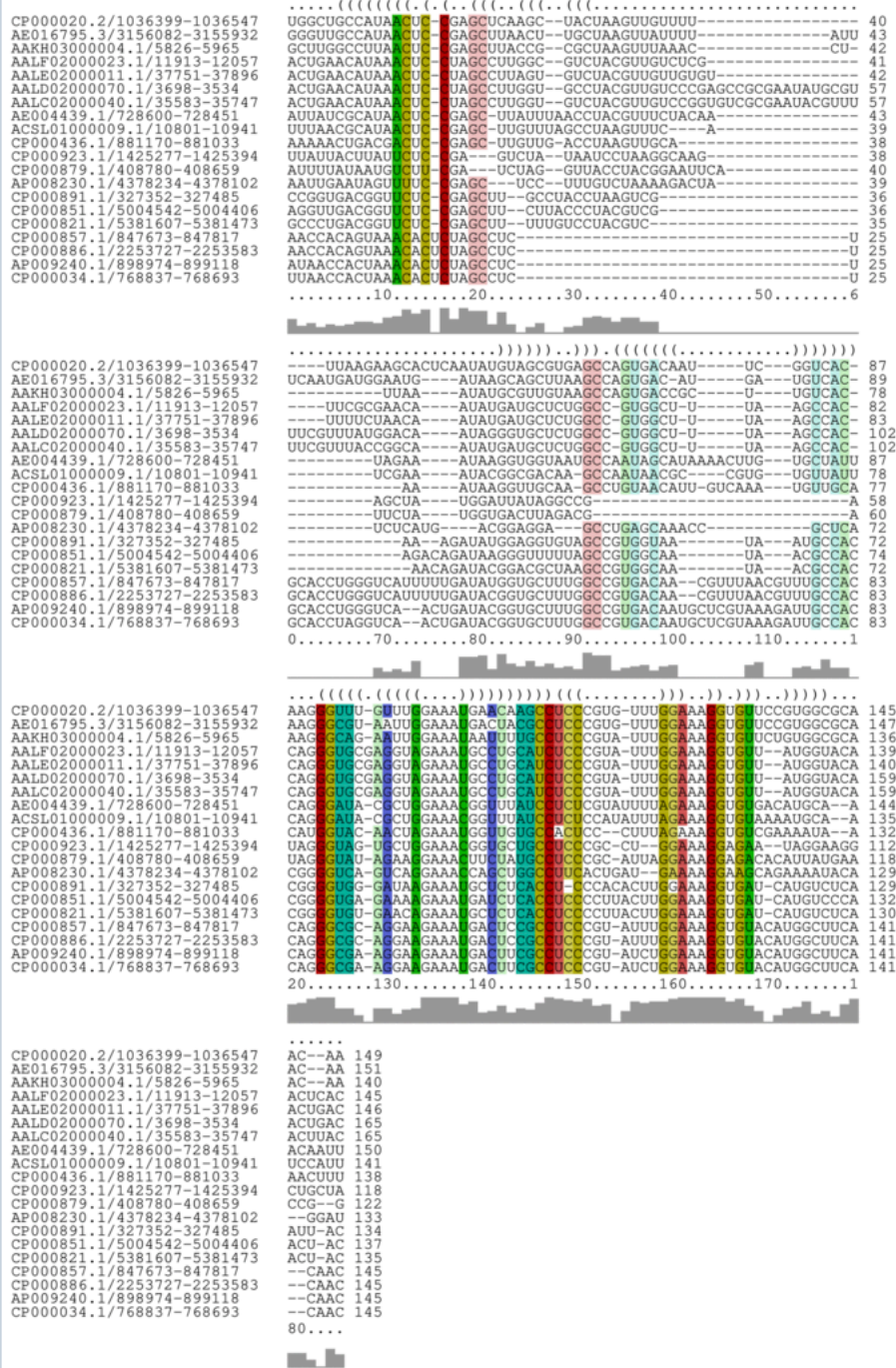
- Centroid structure drawing encoding base-pair probabilities



Вывод: предсказания вторичных структур РНК обоими методами выдает схожие результаты. Красные участки - зоны наиболее точного предсказания, то есть наиболее консервативные. Следовательно, первая РНК более консервативна, чем вторая.

## Задание 2.

1. Выполнила выравнивания 20 последних последовательностей ( <https://rna.informatik.uni-freiburg.de/LocARNA/Input.jsp>).
2. Скачала три файла: Stockholm, FASTA and structure.
3. Вот что видно на сайте:



Вывод: Выравнивание хорошее, что говорит о достаточно высоком уровне консервативности анализируемых молекул.

### Задание 3. Сравнение вторичных структур РНК, полученных двумя разными методами.

1. Сравнение по длине ([https://colab.research.google.com/drive/1dQZQgV2Y63VPjoqrqPC1B8Y5xG\\_HfzPv?usp=sharing](https://colab.research.google.com/drive/1dQZQgV2Y63VPjoqrqPC1B8Y5xG_HfzPv?usp=sharing)).
2. Запустила код:

```
sequence =
"AUAACCACUAAACACUCUAGCCUCUGCACCUGGGUCAACUGAUACGGUGCUUUGGCCGUGACAAUGCUCGUAAAGA
UUGCCACCAGGGCGAAGGAAGAAUGACUUCGCCUCCCGUAUCUGGAAAGGUGUACAUGGCUUCACAACAUAACCAC
UAAACACUCUAGCCUCUGCACCUGGGUCA--
ACUGAUACGGUGCUUUGGCCGUGACAAUGCUCGUAAAGAUUGCCACCAGGGCGAAGGAAGAAUGACUUCGCCUCCC
```

```

GUAUCUGGAAAGGUGUACAUGGCUUCACAACAUAACCACUAAAACACUCUAGCCUC-----
-----UGCACCUGGGUCA--
ACUGAUACGGUGCUUUGGCCGUGACAAUGCUCGUAAGAUUGCCACCAGGGCGA-
AGGAAGAAAUGACUUCGCCUCCCGU-AUCUGGAAAGGUGUACAUGGCUUCA--
CAACAUAACCACUAAAACACUCUAGCCUC-----
UGCACCUGGGUCA--ACUGAUACGGUGCUUUGGCCGUGACAAUGCUCGUAAGAUUGCCACCAGGGCGA-
AGGAAGAAAUGACUUCGCCUCCCGU-AUCUGGAAAGGUGUACAUGGCUUCA--CAAC"

structure = ".....(((((((.(.(.(((.(.(.
(((.....)))))))).)).
(((((((.....))))))..(((((((.....)))))))))
(((.....)))..)))..)))....."

for i in range(len(sequence)):
    if sequence[i:i+1] == "-":
        structure = structure[:i] + "." + structure[i:]
print(structure)

```

### 3. Output:

```

.....(((((((.(.(.(((.(.(.
(((.....)))))))).)).
(((((((.....))))))..(((((((.....)))))))))
(((.....)))..)))..))).....
.....

```

### 4. Делаю график (<https://e-rna.org/r-chie/>).

- Вставила последовательность из файла result.stk (Стокгольмского):

```

CP000436.1/881170-881033    CAUGGUAC-AACUAGAAAUGGUUGUGCCACUCC--
CUUUAGAAAGGUGUCGAAAAUA--AAACUUU

```

- также в следующую ячейку вставила output из предыдущего пункта:

```

.....(((((((.(.(.(((.(.(.
(((.....)))))))).)).
(((((((.....))))))..(((((((.....)))))))))
(((.....)))..)))..))).....
.....

```

- Нажимаю кнопку Plot.

 гг - не делает ничего, и даже на почту отчет не присылает

## Задание 4. Оценка структуры

1. Запустила job в R-scape (<http://eddylib.org/R-scape/>).
2. Получила результат.

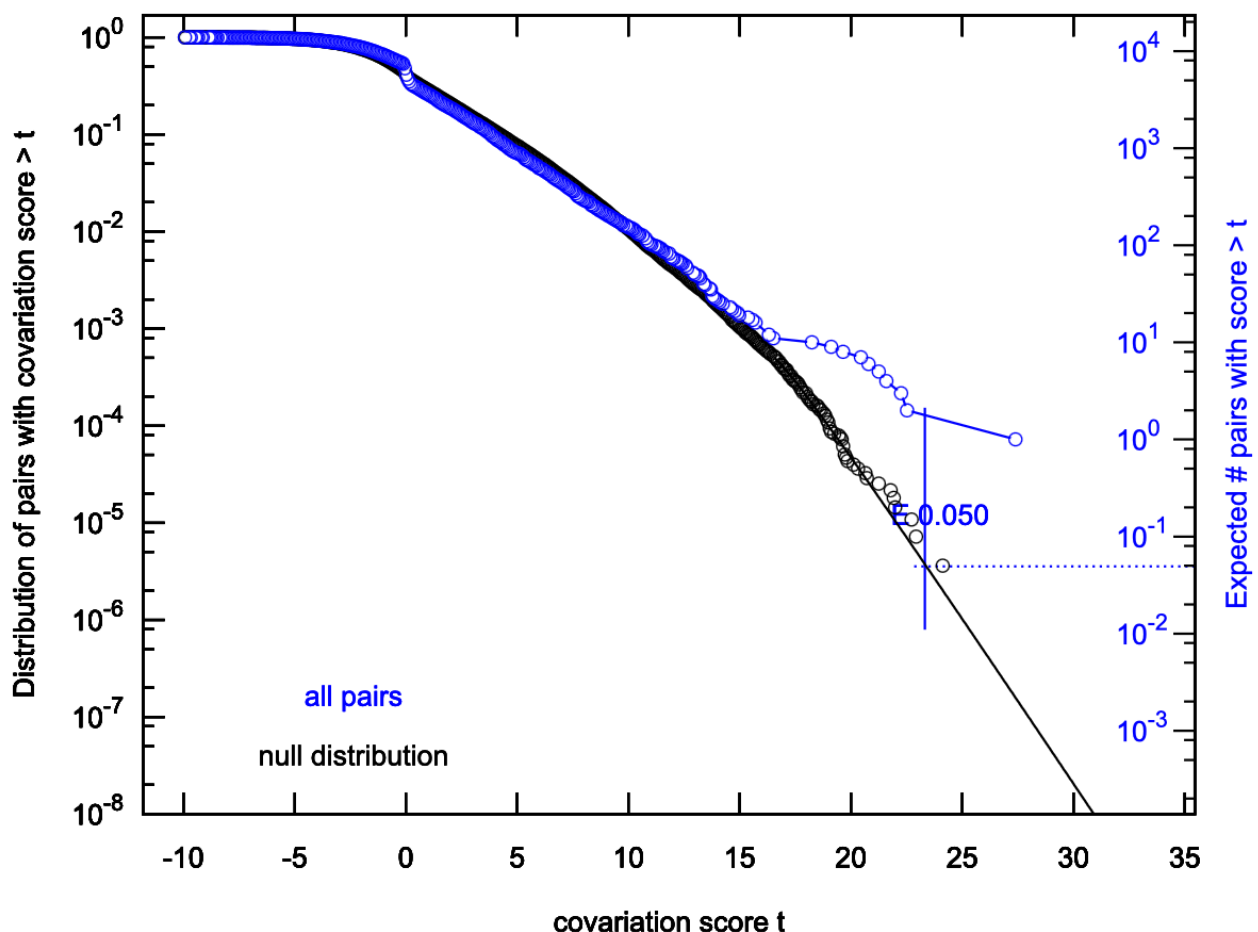
## List of covarying basepairs

in given structure	Left base	Right base	Covariation Score	E-value	Substitutions	Power
~	127	145	27.41575	0.00207881	12	0.40

Significantly covarying pairs present in the structure are marked green. Other covarying pairs are marked orange if both residues are unpaired in the structure or there is no structure present. Orange covarying pairs could be an indication of an under annotated structure or pseudoknots. Black covarying pairs could indicate covariation supports for an alternative structure, tertiary interactions, or false positives.

- \* Base pair in the structure
- ~ Both residues unpaired in the structure, or no structure is present
- At least one residue is involved in other pairing in the structure

output (seqs 20 alen 167 avgid 53 bpairs 0)



Вывод: наблюдается меньше пар с высокими ковариационными баллами, чем ожидается в правой части графиков. Здесь (в правой части) это может указывать на **низкую консервативность** РНК. В левой части графика, где черная и синие линии пересекаются, можно сказать, что уровень консервативности РНК достаточно высок.

## Задание 5. Поиск известной структуры по последовательности.

Загрузила последовательность (вторую из первого задания) на Rfam.



## Search Rfam

Text search

Sequence search

Batch search

Keyword

Taxonomy

Entry type

### Sequence search

Find Rfam families within your sequence of interest. Paste your nucleotide sequence below to search for Rfam families as well as similar sequences from [RNAcentral](#). [More...](#)

Powered by [RNAcentral](#) | Local alignment using [nhmmer](#)

UUAACCACUAAACACUCUAGCCUCUGCACCUGAGGUCAACUGAUACGGUGCUUUGGCC  
GUGACAAUGCUCGUAAAGAUUGCCACCAGGGCGAAGGAAGAAUGACUUCGCCUCCC  
GUAUCUGGAAAGGUGUACAUGGCUUCACAAC



Examples: [lysine riboswitch](#) [16S](#) [SNORD3A](#)

Обнаружены известные структуры:

Shigella dysenteriae Sd197 Moco (molybdenum cofactor) riboswitch  
(URS0000AB92A0\_300267)

И другие.

## Задание 6. Command line practice

1. Take Stockholm file (the alignment produced by locaRNA).

```
/Users/rivikta/Downloads/result.stk
```

2. In command line prompt, type a command that will show you all lines starting with the same ID as you used in Task 1 (any of the two).

```
grep "^CP000436" /Users/rivikta/Downloads/result.stk
```

3. Make a screenshot and include in your report.

```
(base) rivikta@Riviktas-MacBook-Pro ~ % grep "^CP000436" /Users/rivikta/Downloads/result.stk
CP000436.1/881170-881033   AAAACUGACGACUC-CGAGC-UUGUUG-ACCUAAGUUGCA-----
-----AA----AUAAGGUUGCAA-GCCUGUAACAUU-GUCAA--UGUUGCA
CP000436.1/881170-881033   CAUGGUAC-AACUAGAAUGGUUGUGCCACUCC--CUUUAGAAAGGUGUCGAA
AAUA--AAACUUU
```

4. Then, type a command that count number of lines NOT containing this ID.

```
grep -v "^CP000436" /Users/rivikta/Downloads/result.stk | wc -l
```

5. Make a screenshot and include it in your report.

```
(base) rivikta@Riviktas-MacBook-Pro ~ % grep -v "^CP000436" /Users/rivikta/Downloads/result.stk | wc -l
44
```

6. Bonustask (0.5 compensatory points): write a command that can make a FASTA record from the output from (2).

```
grep "^CP000436" /Users/rivikta/Downloads/result.stk | awk '{print
">$1\n"$0}' > /Users/rivikta/Downloads/output.fasta
```



```
output.fasta — Edited ✓
>CP000436.1/881170-881033nCP000436.1/881170-881033
1
AAAAACUGACGACUC-CGAGC-UUGUUG-
ACCUAAGUUGCA-----AA----
5. Se
nce
AUAAGGUUGCAA-GCCUGUAACAUU-GUCAA--UGUUGCA
>CP000436.1/881170-881033nCP000436.1/881170-881033
CAUGGUAC-AACUAGAAUGGUUGGCCACUCC--
CUUUAGAAAGGUGUCGAAAUA--AAACUUU
K
```