

# **Разработка алгоритма синтаксического анализа через умножение матриц**

Анна Явейн

Куратор: Григорьев Семен

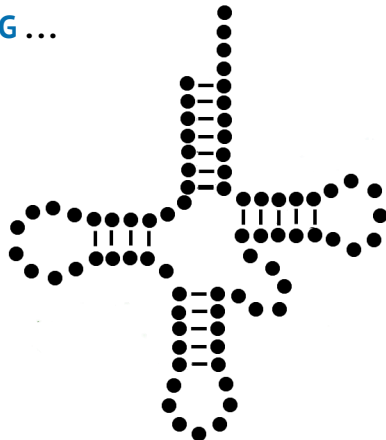
2016

## РНК:

- Первичная структура

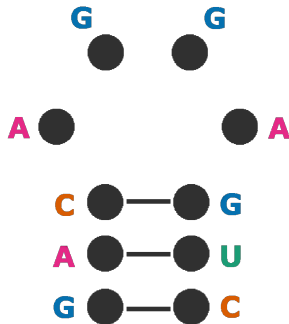
... ACCUUCUAAAG ...

- Вторичная структура



# Представление вторичной структуры

G A C A G G A G U C  
[ ( [ \* \* \* \* ] ) ]

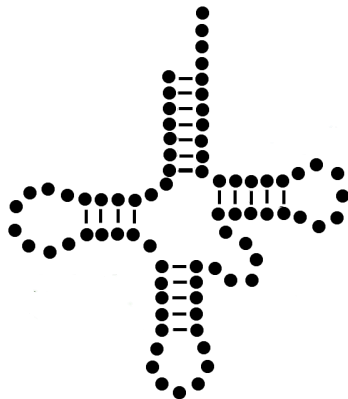


$(V, \Sigma, R, S)$

**Контекстно-свободная грамматика:**

$N \mapsto A$

$N \mapsto A \mid B$



$(V, \Sigma, R, S)$

## Булева грамматика:

$N \mapsto A$

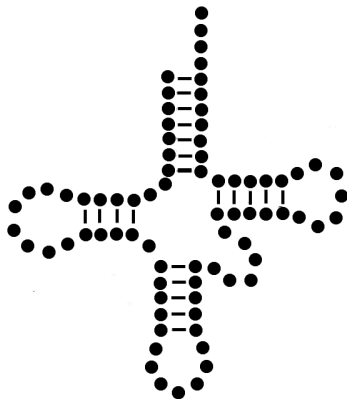
$N \mapsto A \mid B$

$N \mapsto A$

$N \mapsto A \mid B$

$N \mapsto A \& B$

$N \mapsto \neg A$

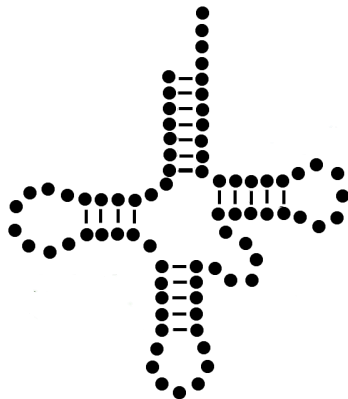


$(V, \Sigma, R, S)$

## Стохастическая грамматика:

$N \mapsto A, \quad p_{N \mapsto A}$

$N \mapsto A|B, \quad p_{N \mapsto A|B}$



# Задача

Дано:

- **Первичная** структура РНК.
- Вид **вторичной** структуры задан **грамматикой**.

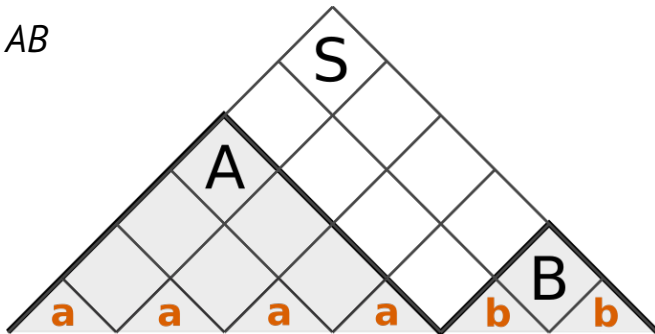
Хотим найти:

- **Подстроку** РНК, которая сворачивается во **вторичную структуру** данного вида.

Усложнение:

- Первичная структура в виде **графа**.

Алгоритм Кока – Янгера – Касами (CYK).

$$S \mapsto AB$$




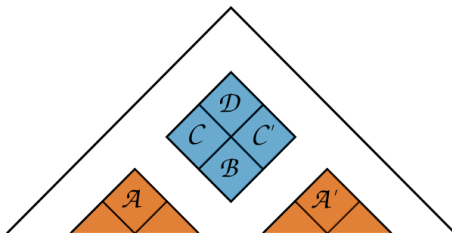
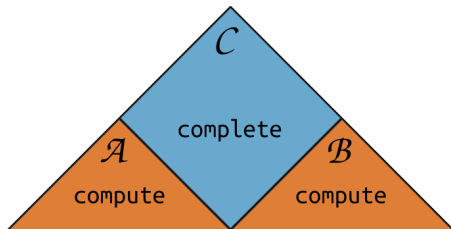
- Не предназначен для **поиска**.
- $O(n^3 |G|)$ .
- **Булева / стохастическая** грамматика.
- На входе не строка, а **граф**.

**compute:**

```
compute(A)
compute(B)
complete(C)
```

**complete:**

```
complete(B)
C += A * B
complete(C)
C' += B * A'
complete(C')
D += A * C'
D += C * A'
complete(D)
```



- Булева / стохастическая грамматика.
  - $O\left(MM(n) \log^{O(1)}(n) |G|\right)$ .
  - Рекурсивный.
  - Много **перемножений** небольших **матриц**.
  - Плохо **параллелится**.
- 
- Реализовано на **F#**

# Решение (переработанный алгоритм)

**complete:**

$$C \quad += \quad A * B$$

$$C' \quad += \quad B * A'$$

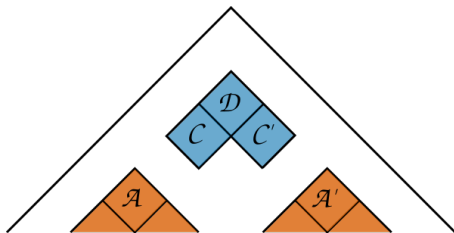
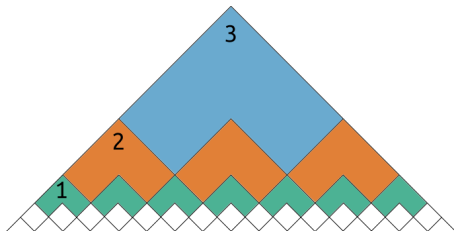
complete( $C$ )

complete( $C'$ )

$$D \quad += \quad A * C'$$

$$D \quad += \quad C * A'$$

complete( $D$ )



# Решение (переработанный алгоритм)

- **Булева / стохастическая** грамматика.
  - $O\left(MM(n) \log^{O(1)} n |G|\right)$ .
  - Чуть менее рекурсивный.
  - **Параллельное умножение** матриц.
- 
- Реализовано на **F#**

- Эффективно перемножать много маленьких матриц.
- Не линейный вход.

`github.com/YaccConstructor/  
YaccConstructor/tree/bio_cyk`