BIOINFORMATICS
2019

The Composition of Dense Neural Networks and Formal Grammars for Secondary Structure Analysis

Polina Lunina, **Semyon Grigorev**

JetBrains Research, Programming Languages and Tools Lab
Saint Petersburg University

Febrary !!!, 2019

Problem Statement

- Sequences classification
  - Lexer and parser
  - Translator
  - Types mapping
  - Headers files processing
  - ...
- Unify kernels on client side
  - Currently native Brahma.FSharp's kernel and kernel loaded by type provider are different types
- Improve mechanism of kernels composition

Secondary structure!!!

- Key is secondary structure: Infernal, other works
- Problem: hight variability.
- Solutions: PCFG, CM

Our receip: Parsing + DNN

- Idea: not secondary structure modelling, but features extraction!

Our receip: Parsing + DNN

- Idea description. Figure

Grammar

```
s1: stem<s0>
any_str : any_smb*[2..10]
s0: any_str | any_str stem<s0> s0
any_smb: A | T | C | G
stem1<s>:                    \\ stem of height exactly 1
      A s T | G s C | T s A | C s G
stem2<s>:                    \\ stem of height exactly 2
      stem1< stem1<s> >
stem<s>:                     \\ stem of height 3 or more
      A stem<s> T
    | T stem<s> A
    | C stem<s> G
    | G stem<s> C
    | stem1< stem2<s> >
```

Grammar

```
s1: stem<s0>
any_str : any_smb*[2..10]
s0: any_str | any_str stem<s0> s0
any_smb: A | T | C | G
stem1<s>:                    \\ stem of height exactly 1
      A s T | G s C | T s A | C s G
stem2<s>:                    \\ stem of height exactly 2
      stem1< stem1<s> >
stem<s>:                     \\ stem of height 3 or more
      A stem<s> T
    | T stem<s> A
    | C stem<s> G
    | G stem<s> C
    | stem1< stem2<s> >
```

Grammar

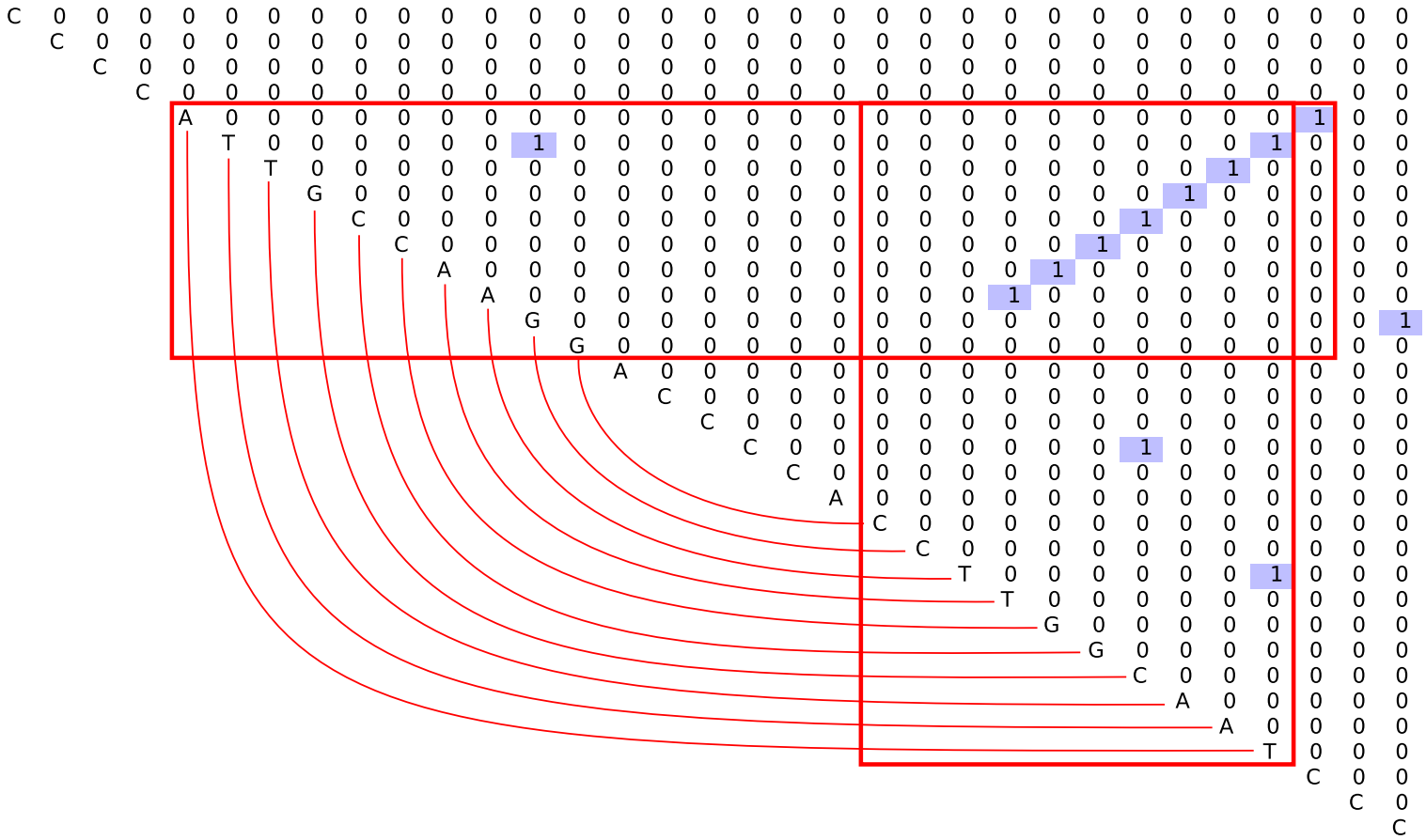
```
s1: stem<s0>
any_str : any_smb*[2..10]
s0: any_str | any_str stem<s0> s0
any_smb: A | T | C | G
stem1<s>:                    \\ stem of height exactly 1
      A s T | G s C | T s A | C s G
stem2<s>:                    \\ stem of height exactly 2
      stem1< stem1<s> >
stem<s>:                     \\ stem of height 3 or more
      A stem<s> T
    | T stem<s> A
    | C stem<s> G
    | G stem<s> C
    | stem1< stem2<s> >
```

Grammar

```
s1: stem<s0>
any_str : any_smb*[2..10]
s0: any_str | any_str stem<s0> s0
any_smb: A | T | C | G
stem1<s>:                    \\ stem of height exactly 1
      A s T | G s C | T s A | C s G
stem2<s>:                    \\ stem of height exactly 2
      stem1< stem1<s> >
stem<s>:                     \\ stem of height 3 or more
      A stem<s> T
    | T stem<s> A
    | C stem<s> G
    | G stem<s> C
    | stem1< stem2<s> >
```

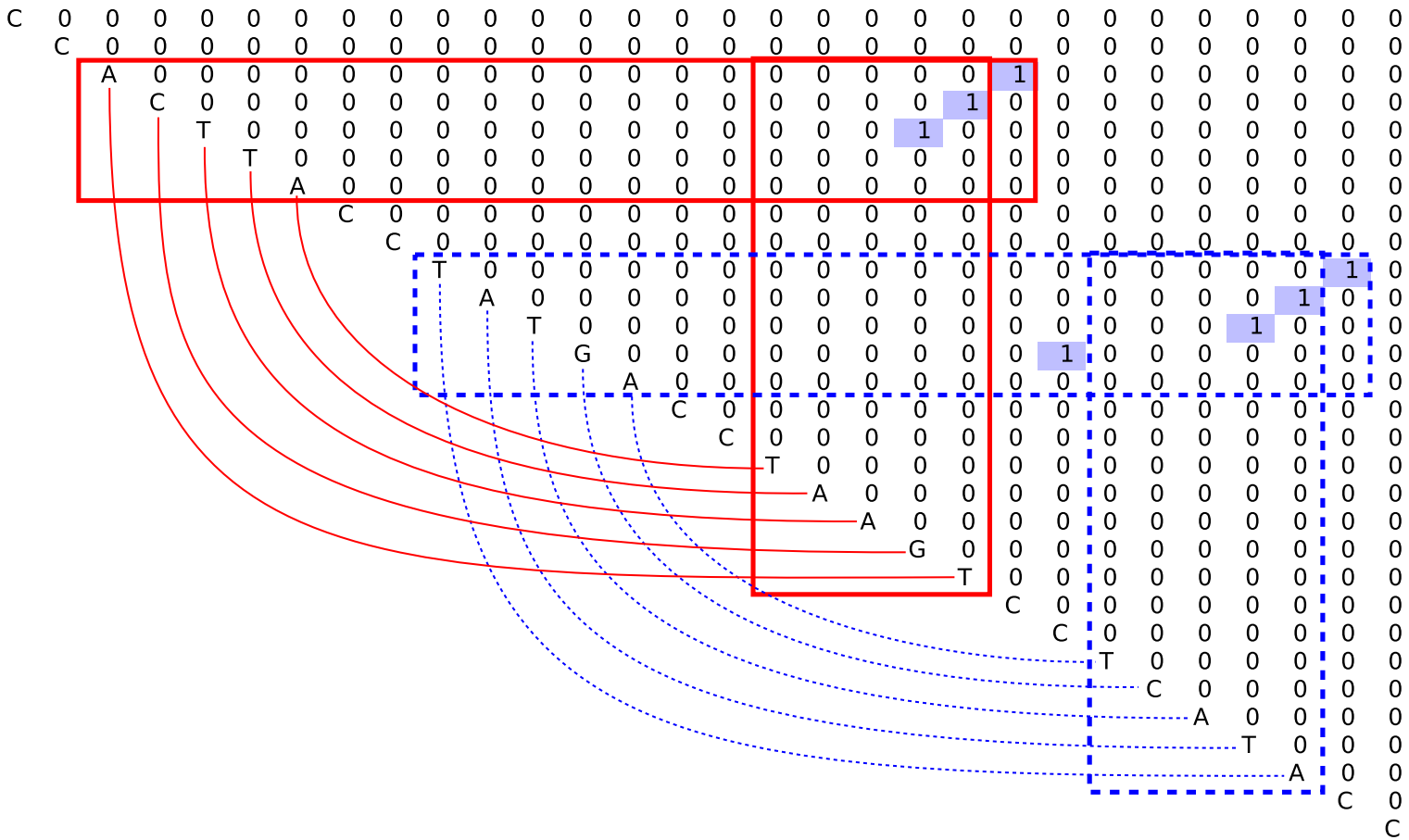
Example 1: Stem

$$\omega_1 = \texttt{CCCCATTGCCAAGGACCCCACCTTGGCAATCCC}$$



Example 2: Pseudoknot

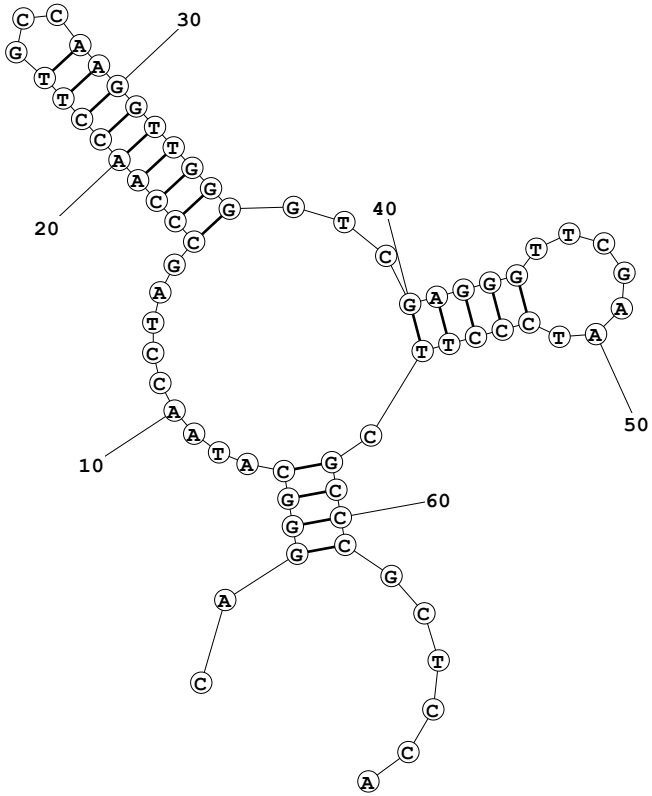
$$\omega_2 = \texttt{CCACTTACCTATGACCTAAGTCCTCATACC}$$

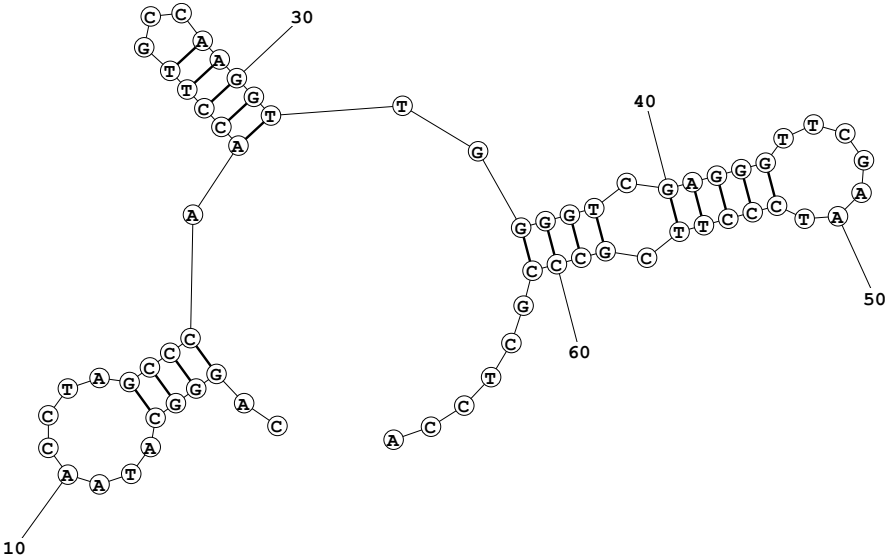


Example 3: real tRNA

$$\omega_3 = \texttt{CAGGGCATAACCTAGCCCAACCTTGCCAAGG}$$
$$\texttt{TTGGGGTCGAGGGTTCGAATCCCTTCGCCCGCTCCA}^1$$

Predicted secondary structures[2]

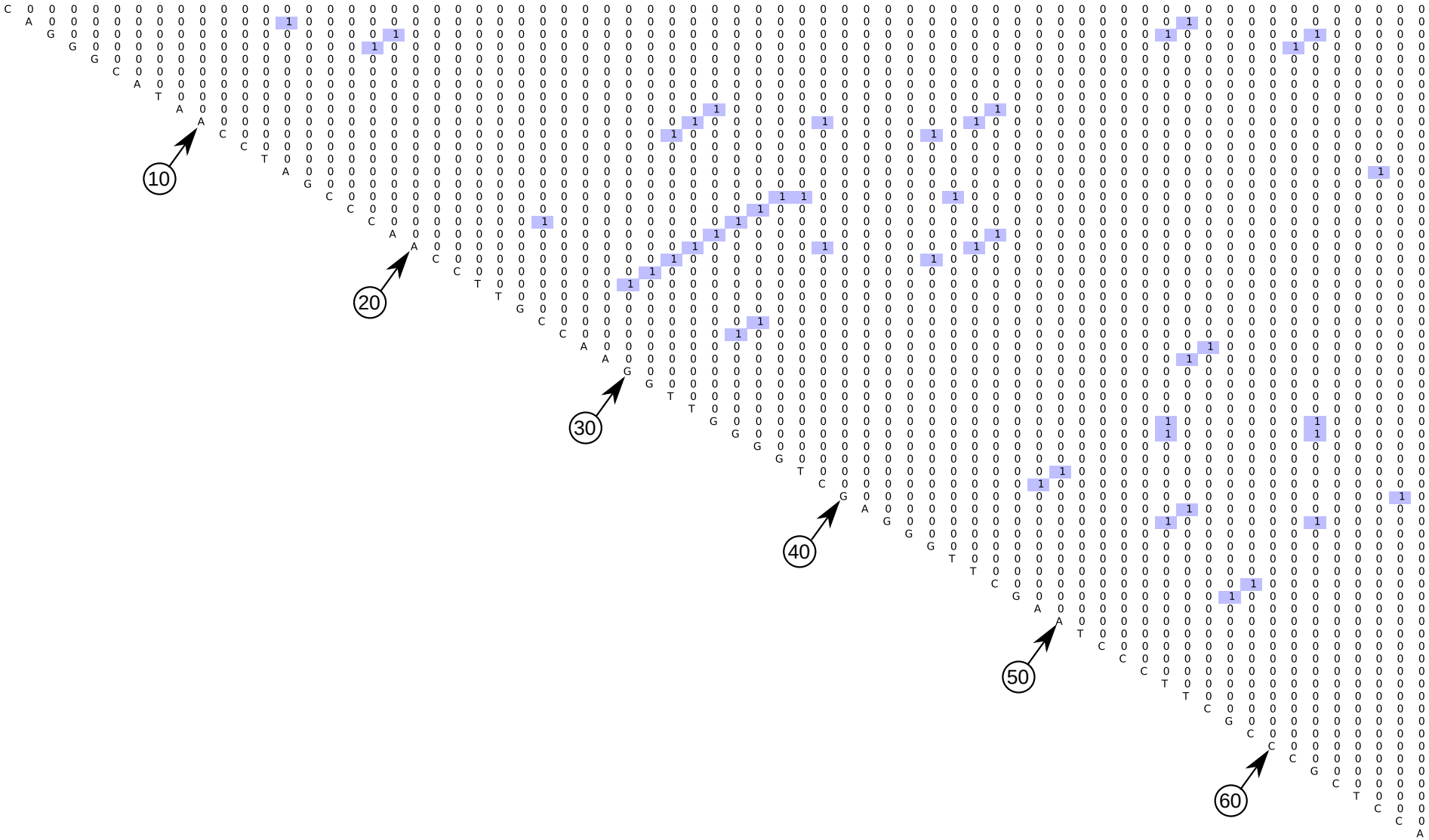




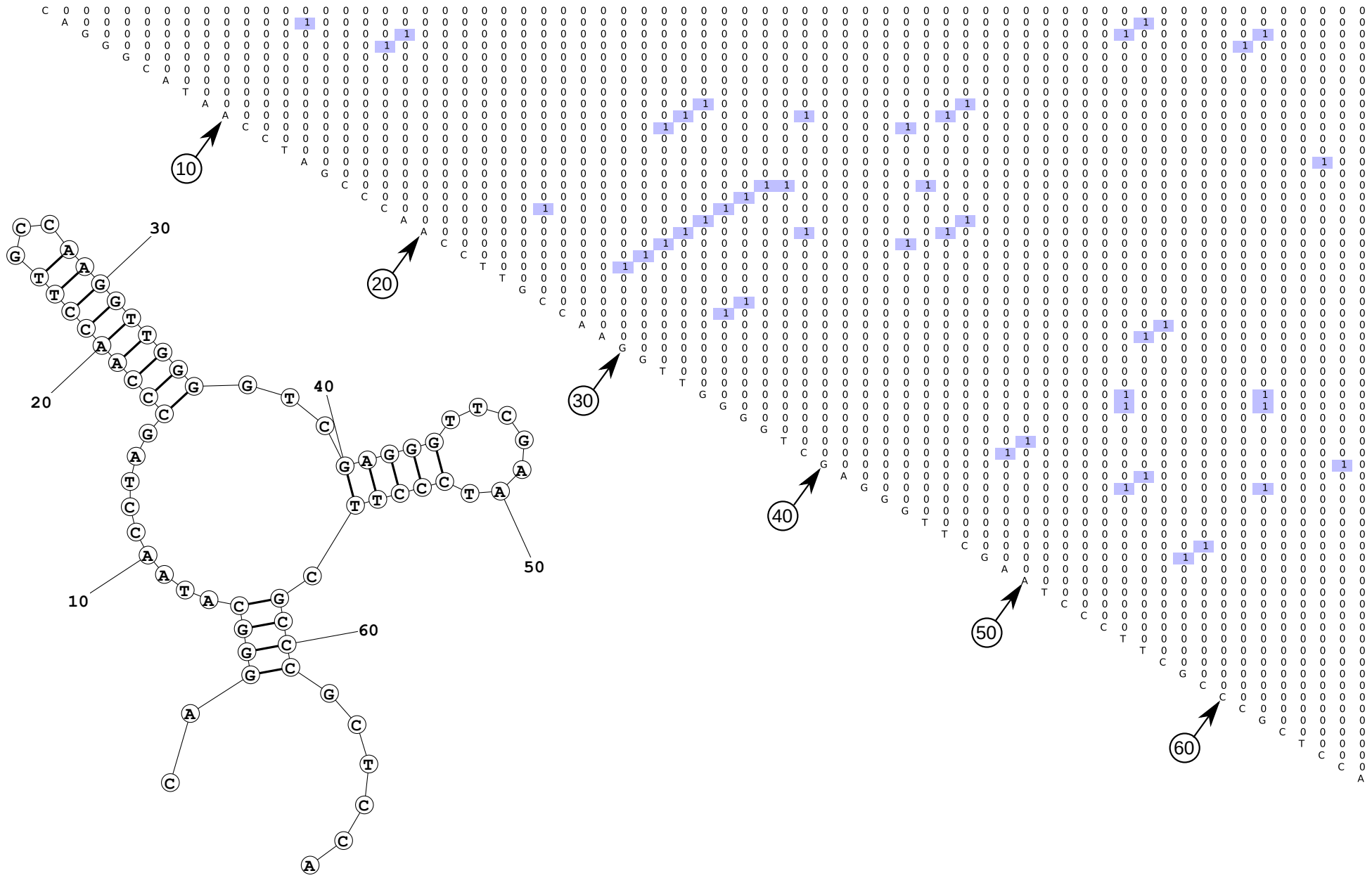
[1] Novosphingobium aromaticivorans from GtRNAdb.

[2] Results are given by using the Fold Web Server with default settings.

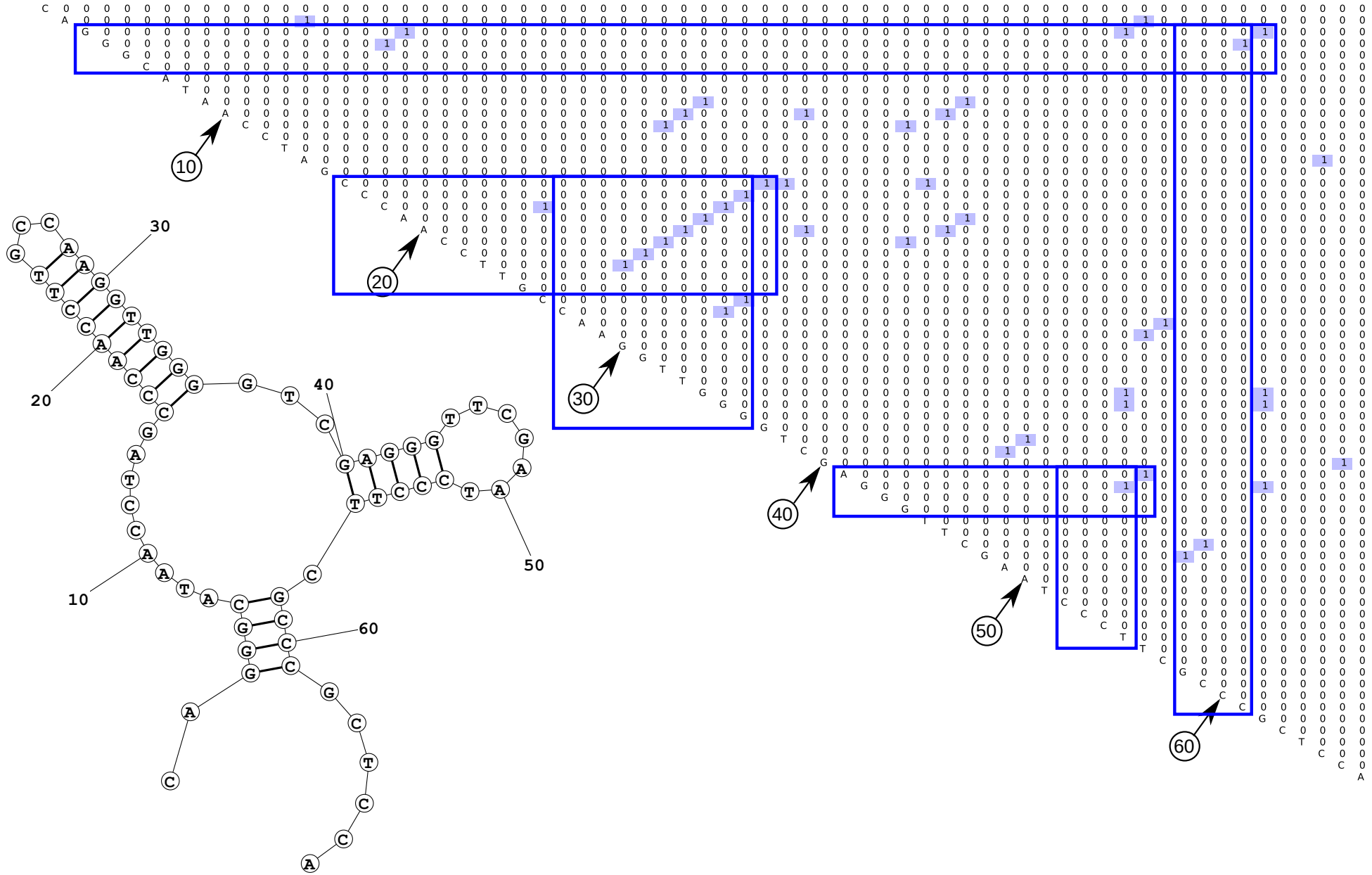
Example 3: real tRNA



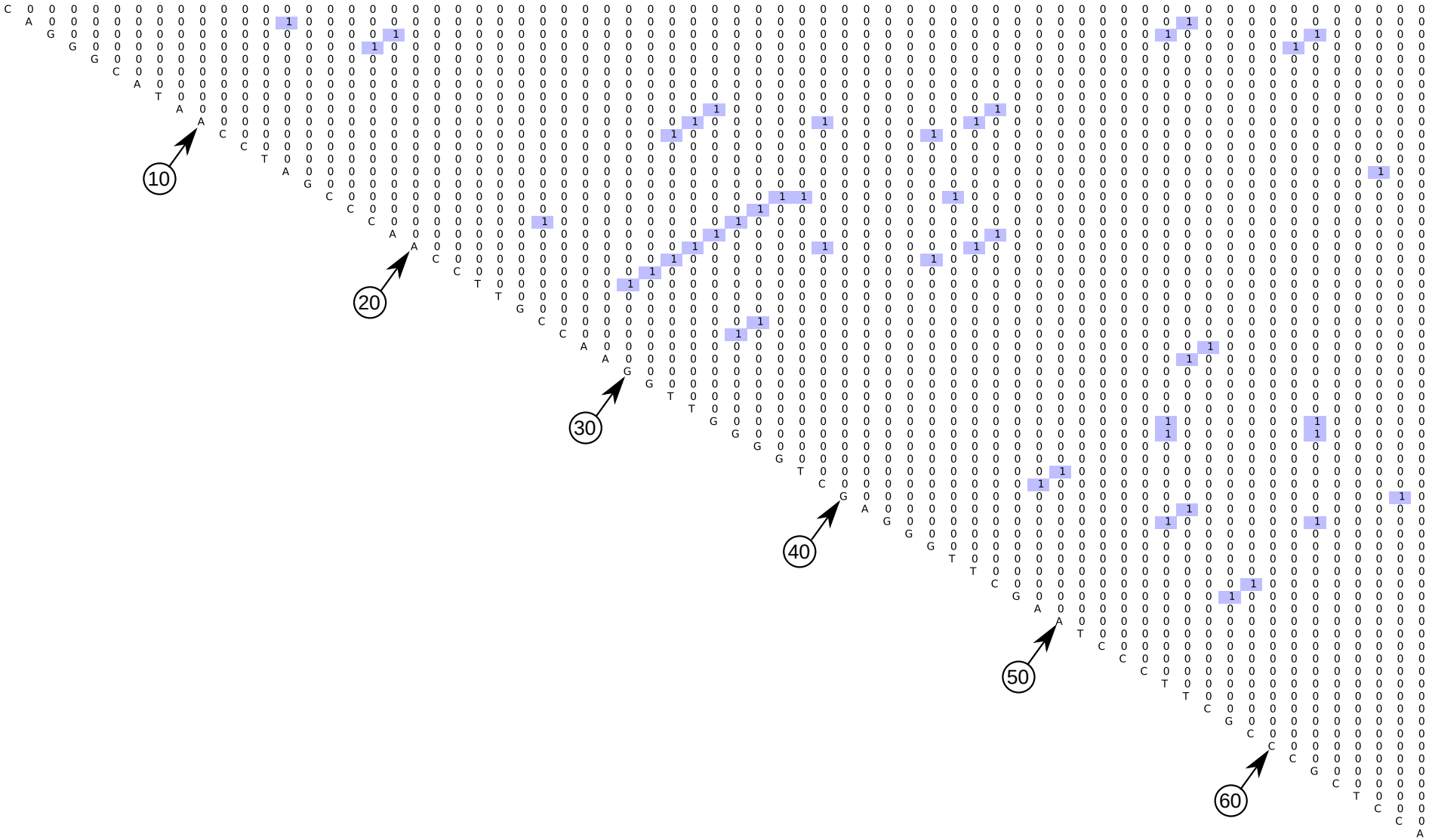
Example 3: real tRNA



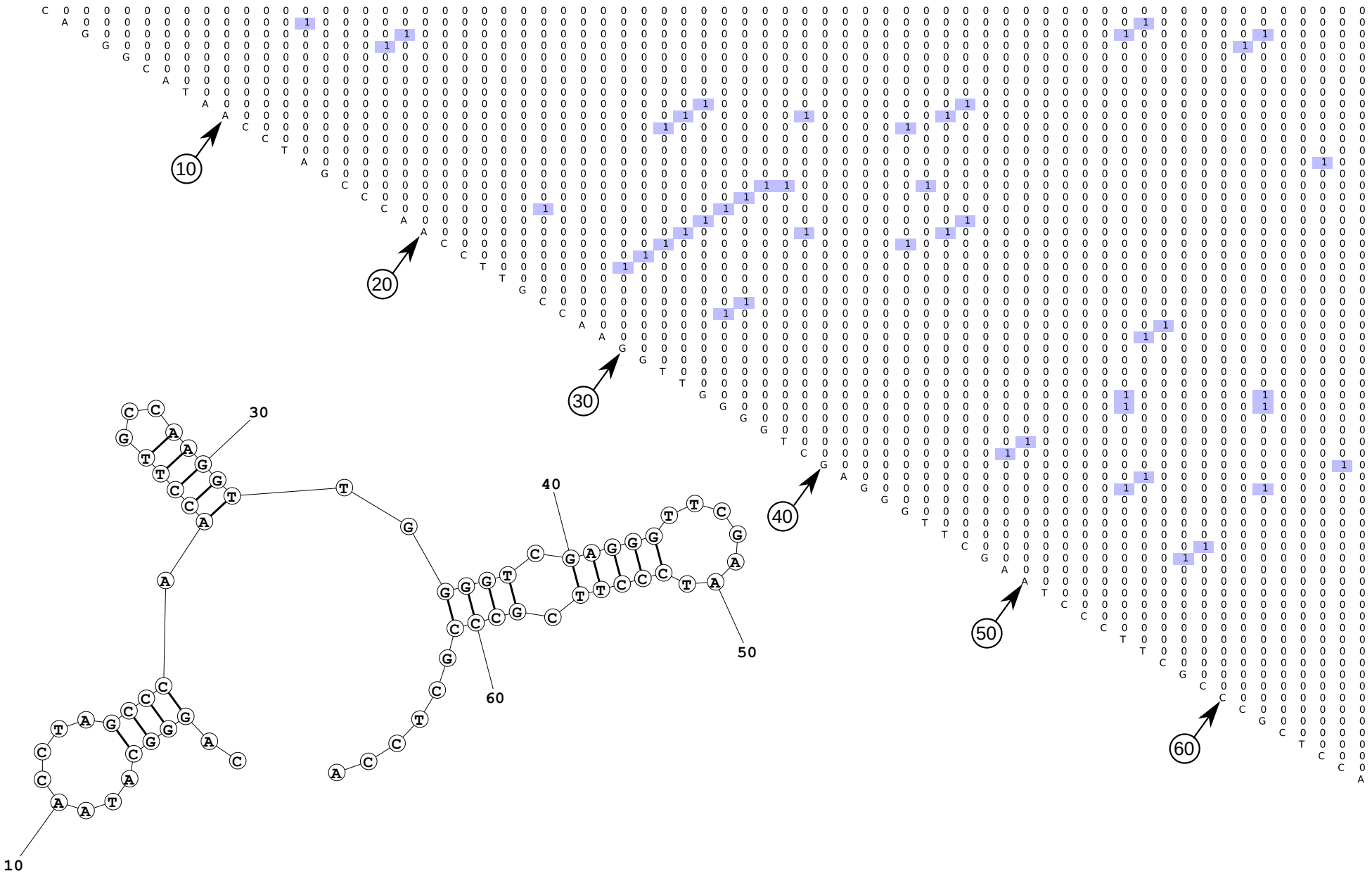
Example 3: real tRNA



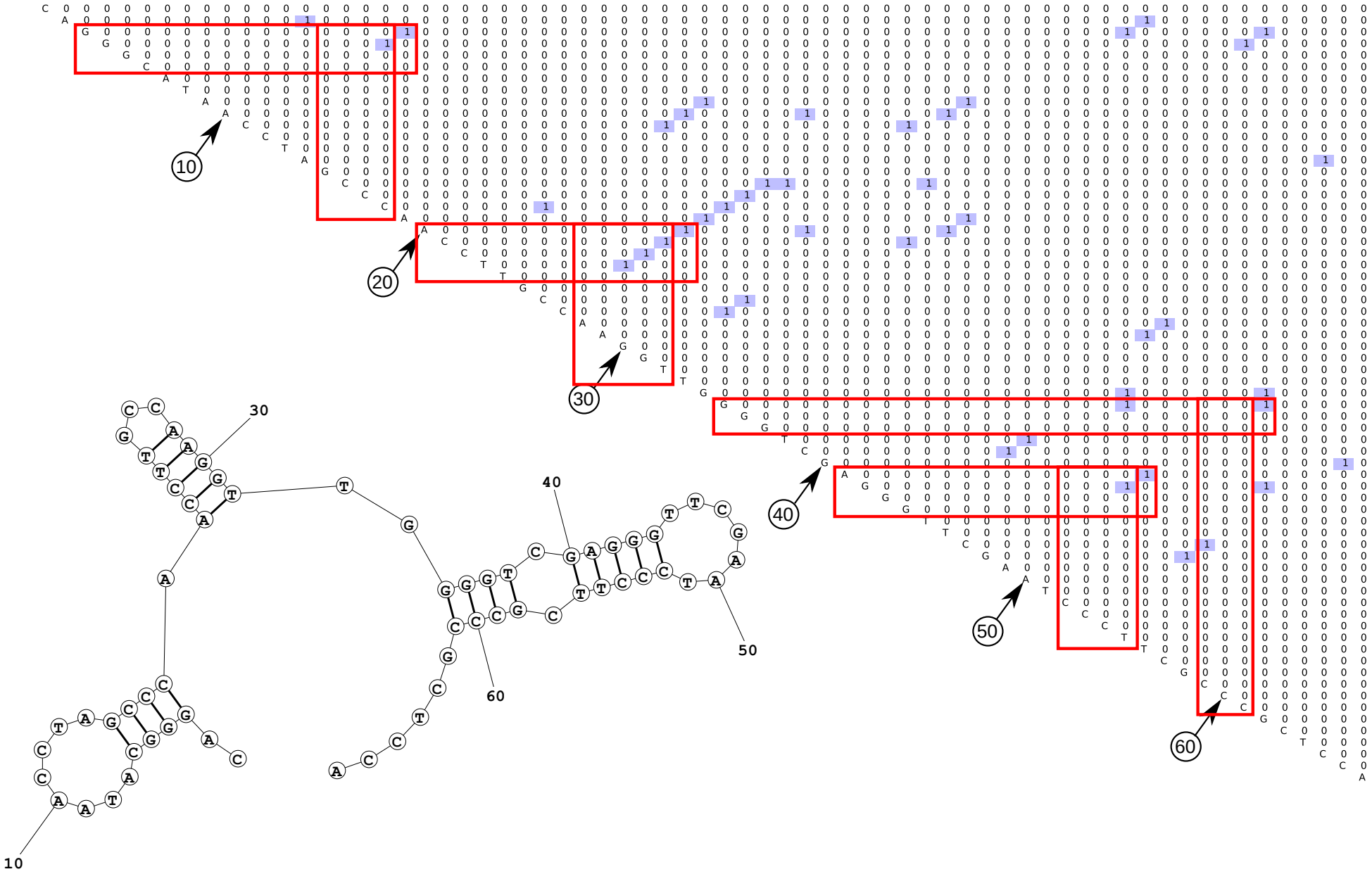
Example 3: real tRNA



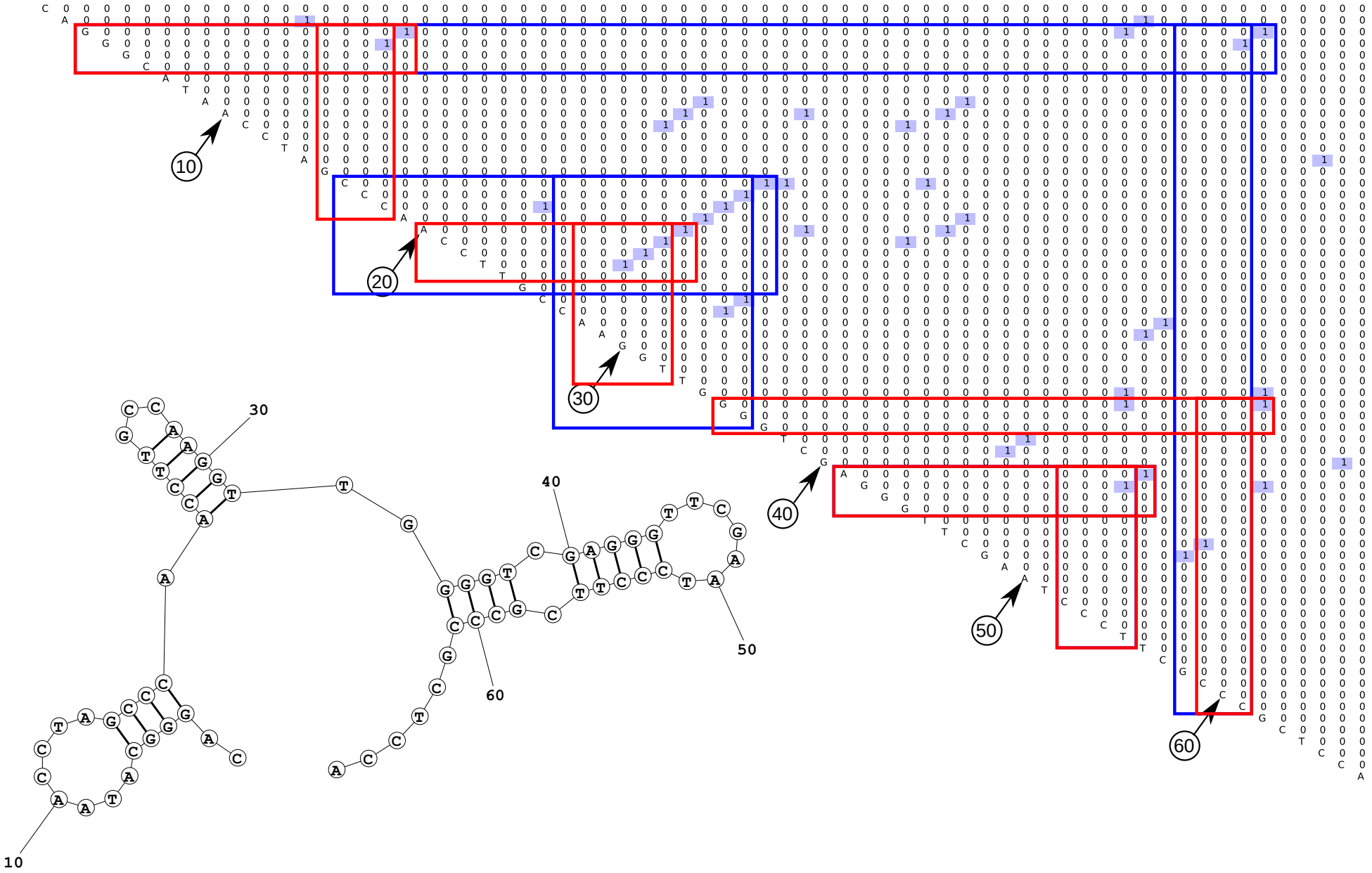
Example 3: real tRNA



Example 3: real tRNA



Example 3: real tRNA



Evaluation setup

- Compile-time metaprogramming for types creation
  - Type provider is a **function which constructs type**

Evaluation results

- Compile-time metaprogramming for types creation
  - Type provider is a **function which constructs type**

Future work

- Compile-time metaprogramming for types creation
  - Type provider is a **function which constructs type**

Summary

- F# OpenCL C type provider
  - Type-safe integration of existing OpenCL C code in F# applications
  - Proof of concept

- Source code on GitHub:
  `https://github.com/YaccConstructor/Brahma.FSharp`
- Package on NuGet:
  `https://www.nuget.org/packages/Brahma.FSharp/`

Contact Information

- Semyon Grigorev:
  - s.v.grigoriev@spbu.ru
  - Semen.Grigorev@jetbrains.com
- Polind Lunina: lunina_polina@mail.ru

Thanks!