

# Übungen zum Sequenzanalyse-Praktikum

Universität Bielefeld, SoSe 2014

Prof. Dr. Jens Stoye · M.Sc. Nina Luhmann · M.Sc. Linda Sundermann

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2014summer/SequaPrak>

praktikum-seqan@CeBiTec.Uni-Bielefeld.DE

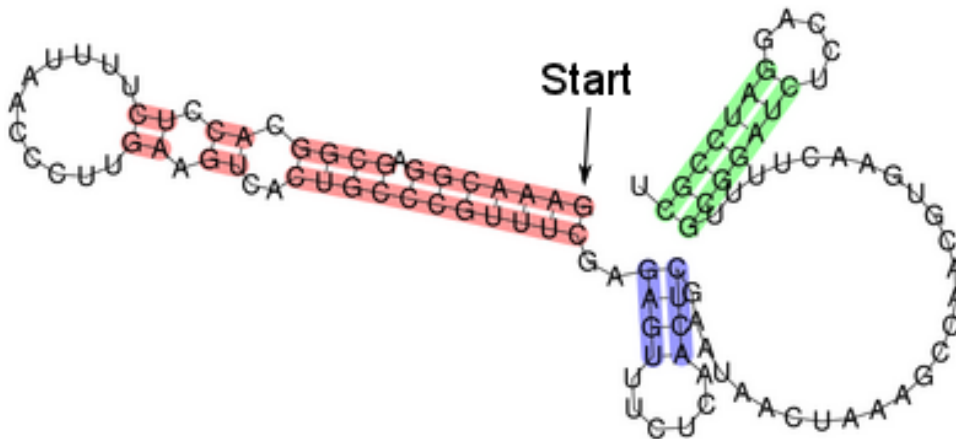
Übungsblatt 13 vom 14.07.2014

Abgabe am Donnerstag, den 17.07.2014

Der abstrakte Shape Ansatz abstrahiert von der konkreten Sekundärstruktur und klassifiziert Strukturen in Familien, die eine ähnliche Sekundärstruktur bzw. die gleiche Shapestruktur teilen (ein Beispiel ist die Kleeblattstruktur der tRNAs). Das Programm RNASHAPES definiert fünf verschiedene Abstraktionslevel für Shapes. Die Notation besteht aus eckigen Klammern '[' und ']' für gepaarte Regionen und Unterstriche '\_' für ungepaarte Regionen.

## Aufgabe 1 (RNA Shapes)

Schaue dir die Sekundärstruktur der *OxyS*-Familie aus der *Rfam*-Datenbank in der Abbildung an. Schreibe dir zunächst die Vienna-Dot-Bracket Darstellung für diese Struktur auf. Übersetze dann die Struktur in Shape Strings der Level 5, 3 und 1.



Die Abstraktionslevel sind die folgenden:

- **Level 5** abstrahiert von allen Helixunterbrechungen (Bulldges und Internal Loops) und ignoriert Einzelstrangregionen. Daher werden in Level 5 nur Hairpins und Multiloops dargestellt. Die tRNA Kleeblattstruktur, welche aus drei benachbarten Hairpins verschachtelt in einem Multiloop besteht, sähe so aus: `[ [ [ ] ] ]`.
- Zusätzlich zu Level 5, werden in **Level 3** alle Helixunterbrechungen durch ein extra Paar von verschachtelten, eckigen Klammern dargestellt. Ein Hairpin mit einem Buldges wäre also `[ [ ] ]`.
- **Level 1** stellt zusätzlich auch ungepaarte Regionen außer den innersten Loop eines Hairpins dar. Somit wäre der Hairpin mit einem Buldges mit ungepaarten Basen an der linken Seite nun `[ _ [ ] ]`.

Level	Shape für <i>OxyS</i>
5	
3	
1	

Ein weiteres Beispiel gibt es hier: <http://bibiserv.techfak.uni-bielefeld.de/rnashapes/manual.html>

## Aufgabe 2 (Nussinov)

Mit dem Nussinov Algorithmus können RNA-Sekundärstrukturen berechnet werden. Welches Kriterium soll dabei optimiert werden bzw. wie sieht die optimale Struktur aus?

Gegeben die RNA-Sequenz

```
>structure from turnip yellow mosaic virus
UUAGCUCGCCAGUUAGCGAGGUCUGUCCCCACACGACAGAUAAUCGGGUGCAACUCCCGCCCUUUUCCGAGGGUCAUCGGAACCA
```

ist die folgende Struktur nach Nussinov eine optimale Lösung:

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

Falte nun die gegebene Sequenz mit dem Tool RNAfold (<http://bibiserv.cebitec.uni-bielefeld.de/rnafold/>) und vergleiche die berechnete Struktur mit der gegebenen Nussinov-Struktur. Warum gibt es Unterschiede? Welche Struktur scheint die biologisch sinnvollere? Welche besondere Struktur kannst du im Ergebnis von RNAfold erkennen?