

## Algorithmische Bioinformatik Übungsblatt 8

Ausgabe: 07. Januar 2020 · Besprechung: 14. Januar 2020

**Aufgabe 8.1** Wir haben den Verdacht, dass das Muster GSCGCG im Genom von *Corynebacterium diphtheriae* (siehe Blatt 7) überrepräsentiert ist, und wollen dies untersuchen.

1. Welchen Strings entspricht dieses Muster (berücksichtige auch das reverse Komplement!)?
2. Beschreibe (z.B. mit einer Skizze) einen NFA, der Texte akzeptiert, die mit einer Instanz dieses Musters (also einem String laut erster Teilaufgabe) enden.
3. Beschreibe einen DFA dazu.
4. Beschreibe den DAA dazu.
5. Beschreibe den PAA, den man aus dem DAA kombiniert mit dem M0-Modell erhält.

**Aufgabe 8.2** Die Konstruktion  $[DAA + \text{Textmodell} \mapsto PAA]$  kann man selbst implementieren, aber die Verwendung von MoSDi (in diesem Fall `mosdi-stat`) spart Zeit. Bestimme die Eigenschaften des Motivs aus Aufgabe 8.1 (Anzahl der Vorkommen, reverses Komplement beachten!; p-Wert im M0-Modell und im M2-Modell).