

Lösungsvorschläge für die Übungsaufgaben – Blatt 7

Zürich, 11. Januar 2006

Lösung zu Aufgabe 19

- (a) Eine mögliche Wahl der Städte-Kodierungen, die die beiden Bedingungen erfüllt, ist in der folgenden Tabelle gezeigt:

Stadt	Kodierung
START	AAA TTT
x	CCC GGG
y	TAT ATA
z	GCG CGC
ZIEL	ATC ACC

Man beachte, dass eine kürzere Kodierung nicht möglich ist: Die Länge der Kodierung muss gerade sein, damit auch die Strassen kodiert werden können. Eine Länge von vier reicht nicht aus, da Bedingung 1 fordert, dass sich die Kodierungen an mindestens vier Positionen unterscheiden müssen. Damit könnten wir jedes der vier Symbole **A, C, G, T** höchstens einmal pro Position verwenden, dies reicht offenbar nicht aus, um fünf Städte-Kodierungen zu finden.

- (b) Aus den unter (a) angegebenen Städte-Kodierungen ergeben sich eindeutig die folgenden Strassen-Kodierungen:

Strasse	Kodierung
START $\rightarrow y$	AAA ATA
$x \rightarrow$ START	CCC TTT
$x \rightarrow$ ZIEL	CCC TAG
$y \rightarrow$ START	ATA TTT
$y \rightarrow z$	ATA CGC
$y \rightarrow$ ZIEL	ATA TAG
$z \rightarrow x$	GCG GGG
ZIEL $\rightarrow y$	TGG ATA

Die Kodierung **AAA ATA** der Strasse von START nach y ergibt sich zum Beispiel durch Hintereinanderhängen des Komplements **AAA** der zweiten Hälfte der Kodierung von START und des Komplements **ATA** der ersten Hälfte der Kodierung von y .

Lösung zu Aufgabe 20

Mit den Kodierungen aus der Lösung zu Aufgabe 19 ergibt sich die folgende konkrete Vorgehensweise gemäss des Adleman-Algorithmus:

1. Gebe DNA-Kodierungen von allen Städten und Strassen (als Einzelstränge) in ein Reagenzglas T , also die DNA-Einzelstränge AAA TTT, CCC GGG, TAT ATA, GCG CGC, ATC ACC, AAA ATA, CCC TTT, CCC TAG, ATA TTT, ATA CGC, ATA TAG, GCG GGG und TGG ATA.
2. Wiederhole $(2n \cdot \log_2 n)$ -mal, also 30-mal, die Operation Amplify(T), um mindestens $n^{2n} = 5^{10}$ Kopien von jedem dieser DNA-Stränge zu erhalten.
3. Erzeuge mit Concatenate(T) eine grosse Menge von doppelsträngigen DNA-Sequenzen, die unterschiedlich lange Wege in dem Strassennetz repräsentieren.
4. Wende die Operation Length-Separate($T, 30$) an. (30 ist die 5-fache Länge der Kodierung einer einzelnen Stadt.) Dann bleiben in T nur die Kodierungen solcher Wege erhalten, die genau 5 Städte lang sind.
5. Wende Separate-Prefix($T, AAA TTT$) an, um nur diejenigen DNA-Sequenzen in T zu behalten, die Kodierungen von Wegen entsprechen, die in START beginnen.
6. Wende Separate-Suffix($T, ATC ACC$) an, um nur diejenigen DNA-Sequenzen in T zu behalten, die Kodierungen von Wegen entsprechen, die in ZIEL enden.
7. Wende Separate($T, CCC GGG$), Separate($T, TAT ATA$) und Separate($T, GCG CGC$) an. Damit bleiben nur diejenigen Wege in T übrig, die jede Stadt mindestens einmal enthalten.
8. Untersuche den Inhalt von T mit Empty?(T) und gib die Antwort JA aus. (Es sollte noch eine DNA-Sequenz in T enthalten sein.)

Lösung zu Bonus-Aufgabe 7

Die Bedingungen aus Aufgaben 19 (a) reichen nicht aus, um zu garantieren, dass jeder gebildete DNA-Doppelstrang einem Weg durch das Strassennetz entspricht.

Dies liegt daran, dass wir zusätzlich sicherstellen müssen, dass sich die Strassenkodierungen nur an den gewünschten Positionen an die Städtekodierungen anlagern können. Sonst könnte zum Beispiel die folgende Situation auftreten: Wir betrachten einen Ausschnitt aus einem Strassennetz, bestehend aus drei Städten mit den Kodierungen $x = \text{ACGT TGCA}$, $y = \text{TGAC GTTG}$ und $z = \text{CATG AGTT}$, so dass zwischen diesen drei Städten nur eine Strasse von y nach z besteht. Dann enthält unsere Kodierung auch die Strassenkodierung $e = \text{CAAC GTAC}$. Hiermit kann sich nun der folgende Doppelstrang bilden:

```
A C G T T G C A T G A C G T T G
      C A A C G T A C
```

Dieser Doppelstrang besteht aus den Kodierungen der Städte x und y sowie der Strassenkodierung e . Damit repräsentiert dieser Doppelstrang eine Strasse von x nach y , die in dem Strassennetz nicht existiert.

Zusätzlich sind bei dem Entwurf solcher DNA-Kodierungen noch viele biologische Besonderheiten zu berücksichtigen, zum Beispiel sind lange Folgen desselben Symbols wie **AAAAAAAAAAAAAAAAAAAA** nicht möglich, weil das DNA-Molekül dann instabil wird. Aus diesen Gründen ist der Entwurf solcher Kodierungen in der Praxis eine sehr anspruchsvolle Aufgabe.