

Sara Kemmler 5760949
Robin Bonkaß 5769588

1	2	Σ

Übungsblatt Nr. 01

(Abgabetermin 05.05.22)

Aufgabe 1

a)

Die scoring-Matrizen für BLASTn, welche keinen match score von 1 verwenden sind:

	A	G	C	T
A	2	-3	-3	-3
G	-3	2	-3	-3
C	-3	-3	2	-3
T	-3	-3	-3	2

Verhältnis zwischen match und mismatch score: 2:3

	A	G	C	T
A	4	-5	-5	-5
G	-5	4	-5	-5
C	-5	-5	4	-5
T	-5	-5	-5	4

Verhältnis zwischen match und mismatch score: 4:5

b)

Bedingung für eine scoring-Matrix:

$$\sum_{a,b \in \Sigma} p_a p_b S(a,b) < 0$$

Da gilt, dass die Wahrscheinlichkeiten p_a und p_b für alle $a, b \in \{A, G, C, T\}$ gleich sein sollen, kann folgendes daraus geschlossen werden:

$$p_a p_b = \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16} = 0,0625$$

Dies ist der Fall, da alle Wahrscheinlichkeiten ($p_A = p_G = p_C = p_T = 0,25$) sein müssen. Für die obige Bedingung ergibt sich dann folgendes:

$$\begin{aligned} \sum_{a,b \in \Sigma} \frac{1}{16} S(a,b) &< 0 \\ \iff \frac{1}{16} \sum_{a,b \in \Sigma} S(a,b) &< 0 \end{aligned}$$

Da unsere Summe aber insgesamt echt kleiner als 0 sein muss, kann auch der Bruch weggekürzt werden, da dieser unsere Summe hinsichtlich des Vorzeichens nicht verändert. Damit ergibt sich schließlich:

$$\sum_{a,b \in \Sigma} S(a,b) < 0$$

Unser Verhältnis zwischen match- und mismatch-Anzahl liegt bei 4:12, da wir bei einer 4x4-Matrix 4 Matches und 12 Mismatches erreichen. Die folgende Gleichung zeigt die Relation zwischen k (Mismatch) und m (Match), wobei $m \in \mathbb{N}$ gilt:

$$\begin{aligned}4m + 12k &< 0 \\12k &< -4m \\k &< \frac{-4m}{12} \\k &< \frac{-1}{3}m\end{aligned}$$

Als konkretes Beispiel für die Werte $m = 3$ und $k = -2$ ergibt sich folgende Matrix:

	A	G	C	T
A	3	-2	-2	-2
G	-2	3	-2	-2
C	-2	-2	3	-2
T	-2	-2	-2	3

4

Aufgabe 2

umgesetzt in file Sara_Kemmler_Robin_Bonkass_A1.py

8

Aufgabe 3

umgesetzt in file Sara_Kemmler_Robin_Bonkass_A1.py

8