

Übungen zur Vorlesung Bioinformatik I

Wintersemester 2004/2005

Abgabe bis 31.1. 2005, 12.00 Uhr, INF 580

1. (a) Gegeben sei folgender Abschnitt aus einem multiplen Sequenzalignment:

```
LKRPTFKQIVQLI
LKRPTFKQVVQLI
LKRPTFKQIVQLI
VKRPSFSKIVEKI
```

Berechne den Score für dieses Alignment nach der *Sum-of-Pairs* (SP) Methode. Benutze dafür die BLOSUM62-Matrix (s. Vorlesungsskript). Der Rechenweg ist anzugeben! (2P)

(b) Welche Nachteile hat diese Methode? Berechne zum Vergleich den optimalen Score, bei dem die Aminosäuren in allen Spalten identisch sind. Schlage eine Verbesserung für den SP-Score vor. (2P)

2. (**Divide-and-Conquer**) Beim Divide-and-Conquer-Verfahren wird der Suchraum durch Spalten der Sequenz stark eingeschränkt. Bekanntlich ist die Komplexität für ein optimales Multiples Sequenzalignment $\mathcal{O}(L^N \times (2^N - 1))$. Dabei ist L die Länge der Sequenzen, und N ihre Anzahl.

(a) Nimm an, dass 5 Sequenzen der Länge 240 aliniert werden sollen. Das Alignment von 2 Sequenzen dieser Länge dauert 0,5 s. Um wieviel schneller ist der DCA-Algorithmus, wenn die Sequenz in 2, 5 oder 8 gleich große Stücke geteilt wird? Vernachlässige die Zeit, die zum Finden der Spaltstellen benötigt wird. (2P)

(b) Warum darf die Zeit zum Auffinden der Spaltstellen eigentlich nicht vernachlässigt werden? (1P)