

Foundations of Sequence Analysis
Winter semester 2003/2004

Exercises

Exercise 4, Discussion: 12/08/2003.

1. (Maximal Matches Model)

Given the sequences $u = AGTGCACACATC$ and $t = ATCACACTTAGC$.

- (a) Calculate the left-to-right partition of u w.r.t. t .
- (b) Calculate the right-to-left partition of u w.r.t. t .
- (c) Calculate the left-to-right partition of t w.r.t. u .
- (d) Calculate the right-to-left partition of t w.r.t. u .

2. Prove that the left-to-right partition is a minimal partition. (Hint: Proof by contradiction.)

3. (q-Gram Model)

- (a) Develop a Haskell function to compute the *q-gram profile*.

Given a window size q and a sequence u , a *q-gram profile* is the function

$g_q u :: \text{String} \rightarrow \text{Int}$,

where $g_q u w$ is the number of different positions in u where the sequence $w \in \mathbf{A}^q$ begins. (The function does not necessarily need to be optimal in terms of time and space complexity.)

- (b) Given two strings u and v , implement a Java or C program $qgdist(q, u, v)$ for the computation of the *q-gram distance* according to the algorithm given in Section 3.6 of the lecture notes.