

Exercise 3

Deadlines: Tuesday 2008.09.23 (copy) and Tuesday 2008.09.30 (corrected)

PROBLEM 1

Consider the sequences $\mathbf{v} = TACGGGTAT$ and $\mathbf{w} = GGACGTACG$. Assume that the match premium is $+1$ and that the mismatch and indel penalties are -1 .

a) Fill out the dynamic programming table for a *global* alignment between \mathbf{v} and \mathbf{w} . What is the score of the optimal global alignment and what alignment does this score correspond to? *Hint*: Draw arrows in the dynamic programming table that indicate the best choice for each entry (i.e. insertion, deletion or (mis)match). In case of ties, choose one by random. This makes it possible to backtrack one of the optimal alignments.

b) Fill out the dynamic programming table for a *local* alignment between \mathbf{v} and \mathbf{w} . What is the score of the optimal local alignment and what alignment does this score correspond to?

c) Suppose we use a gap penalty where it costs -20 to open a gap, and -1 to extend it. Scores of matches and mismatches are unchanged. What is the optimal global alignment in this case and what score does it achieve? *Hint*: You are not asked to fill in the dynamic programming table here.

PROBLEM 2

Write a naïve, brute force algorithm that solves the global alignment problem by searching through all paths from *source* to *sink*. Assume that you have a scoring matrix δ available to you. It is enough that your algorithm returns the score of the best path.

Can you say something about the time complexity of this algorithm? Discuss the difference in time complexity between this algorithm and the dynamic programming solution.

PROBLEM 3

a) Use the ExonChaining algorithm below to solve the exon chaining problem for the following putative exons (i.e. (start position, end position, weight)):

Exon 1: (1, 8, 7)

Exon 2: (2, 4, 3)

Exon 3: (3, 5, 5)

Exon 4: (6, 9, 3)

Exon 5: (7, 11, 4)

Exon 6: (10, 12, 5)

Draw a graph and fill in the dynamic programming array (i.e. \mathbf{s}). Show the path.

b) Consider the Exon chaining problem in the case when all intervals have the same weight. Design a greedy algorithm that finds an optimal solution for this limited case of the problem.

ExonChaining (G, n)

```
1  for  $i \leftarrow 0$  to  $2n$ 
2     $s_i \leftarrow 0$ 
3  for  $i \leftarrow 1$  to  $2n$ 
4    if vertex  $v_i$  in  $G$  corresponds to the right end of an interval  $I$ 
5       $j \leftarrow$  Index of vertex for left end of the interval  $I$ 
6       $w \leftarrow$  Weight of the interval  $I$ 
7       $s_i \leftarrow \max \{s_j + w, s_{i-1}\}$ 
8    else
9       $s_i \leftarrow s_{i-1}$ 
10 return  $s_{2n}$ 
```

PROBLEM 4

Let $\mathbf{v} = v_1v_2\dots v_n$ be a string (sequence) and let \mathbf{P} be a $4 \times m$ profile representing a multiple alignment.

a) Generalize the global sequence alignment algorithm (the one using a scoring matrix) for aligning a sequence against a profile by rewriting the corresponding recurrence.

b) Estimate the amount of time that your algorithm will take with respect to n and m .

PROBLEM 5

A string \mathbf{x} is called a *supersequence* of a string \mathbf{v} if \mathbf{v} is a subsequence of \mathbf{x} . For example, ABLUE is a supersequence for BLUE and ABLE.

Given strings \mathbf{v} and \mathbf{w} , describe an algorithm that will find the shortest sequence that is a supersequence for both \mathbf{v} and \mathbf{w} .