Bioinformatics I

WS 15/16

Tutor: Alexander Seitz

Jonas Ditz

& Benjamin Schroeder

Blatt 7

(Abgabe am 26. October 2015)

Exercise 1 - *Hirschberg algorithm for alignment in linear space*

With two given Sequences (Xand Y), the scoring of s(a,a) = 4, s(a,b)=-3 and a gap penalty g=-3, the Hirschberg algorithm was started.

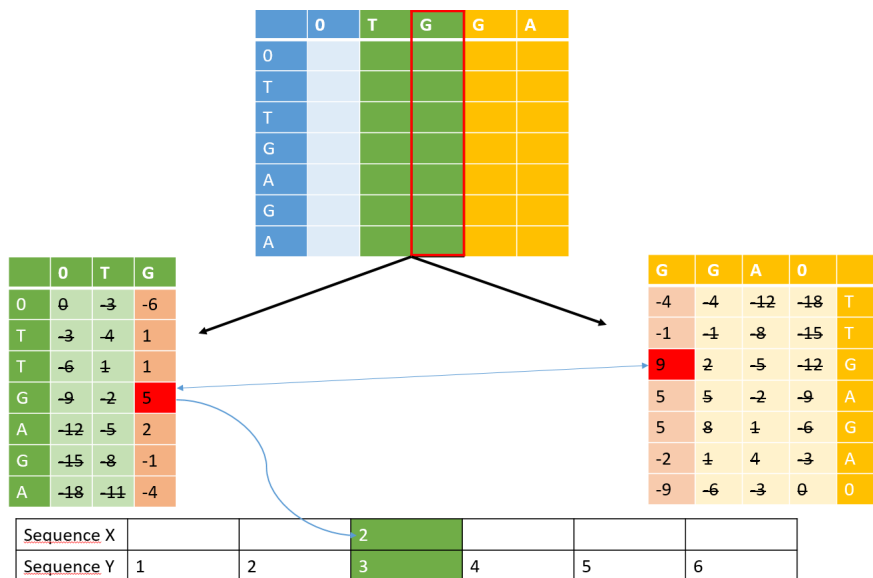First, the Sequence X was split into two parts and

Figure 1: In the first step the sequence X was split into two parts and for both a Needleman Wunsch global alignment was calculated. It's important to state, that the Needleman Wunsch of the suffix of the sequence splitting was calculated reverse. The Maximum which was found after the substraction of both maxima was used to identify two letters, whrch are to be aligned.

Exercise 2 - *Do scoring matrices have expected score smaller than zero?*

The substitution matrices were evaluated according to the expectancy value formula given in the lecture, the substitution matrices were evaluated with the assumption of uniform distributed amino acids in two sequences.

$$\sum_{a,b \in \Sigma} p_a p_b s(a,b) <=> (\sum_{a,b \in \Sigma} s(a,b)) p_a p_b$$
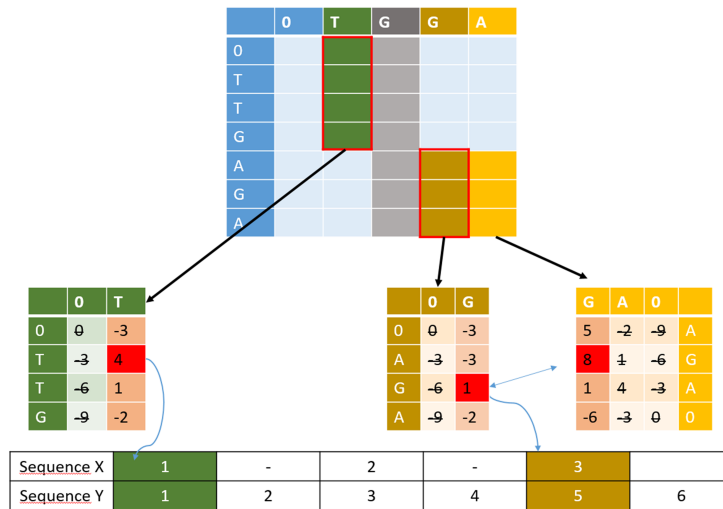
Figure 2: BLAAAAAAAAAAA

For the determination of one expectance value, p_a and p_b were each substituted with $1/20$, for the probabilistic occurrence in sequences with uniform distributed amino acids. The sum was calculated via R and the total summation of the substitution matrix .The result for each matrix can be seen in table .

Substitution Matrix	Expectance Value
BLSOUM50	-1.155
BLOSUM52	-1.065
BLOSUM80	-2.3275
PAM250	-1.14
PAMN	0.155

Table 1: Result of the calculation with each given substitution matrix according to equation

As a conclusion we want to state, that all substitution matrices are possible feasible matrices. Except of the PAMN matrix, which has a positive expectancy value, which could lead to wrong conclusion. One Aspect we also want to state, is that we did not calculate the values for the matrices with the unknown sign x or X. and the B or Z sign, which stands for two amino acids. Although it would be possbile to use the unknown sign, if they are also uniform distributed in a sequence. But as a result of the wanted comparison, between matrices which didn't contain these signs, we did not use these columns and rows in our calculations.
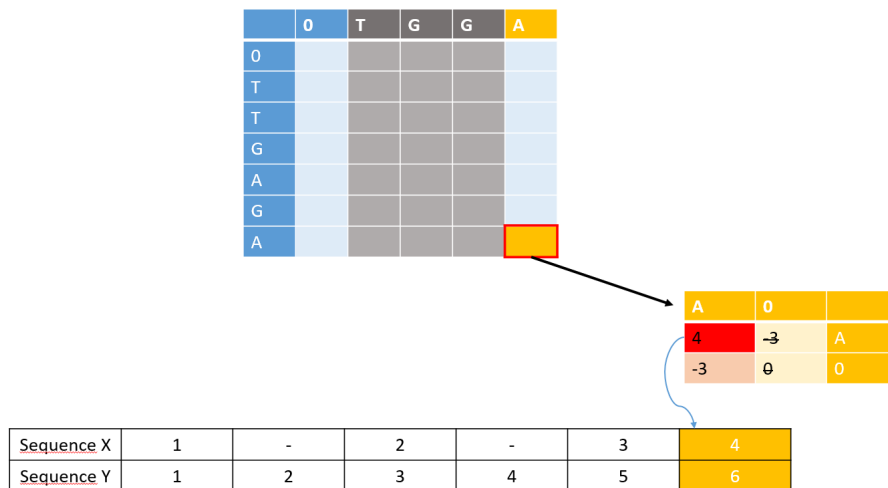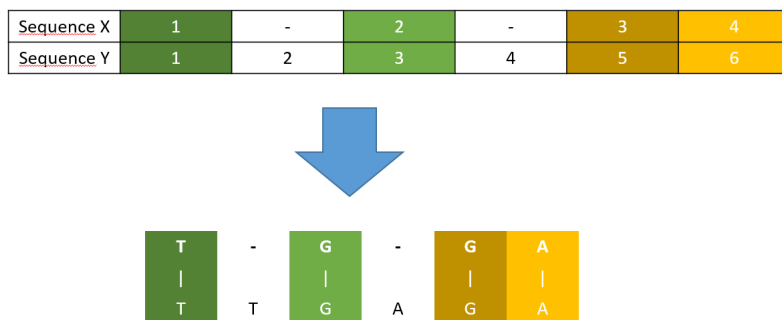
Figure 3: BLAAAAAAAAAAA

Figure 4: BLAAAAAAAAAAA

Exercise 3 - *Needleman-Wunsch algorithm for affine gap scores*