# Abschlussklausur Modul „Sequence Analysis“

Anwendungsbezogene Bioinformatik und Biostatistik 2022-1  
Dozent: Christoph Knorr  
05.04.2022

**Dotplot**

You have these two nucleotide sequences

S1: = „AGTTGT“ und S2:= „AGTTTGG“,

Make a Dotplot for these two sequences and use a Windowsize of 2. What information can you read out of this specific plot?  
Which information can be generally read out of a Dotplot? /8

**Score and Alignment**

You have this Alignment:

K A T Z E N

K O T Z E \_

Calculate the Evaluation according to the following scoring scheme:

Match := 0, Mismatch := 1, Gap := 2,

Is this scheme one over the distance or the score? Please tell in your answer when a scheme about distance and when a scheme about score is optimal. /4

**Optimal Alignments**

You have the following sequences:

S1 := „KLAUS“ and S2 := „FAGUN“

Calculate the optimal local alignment and the score with an optimal algorithm. Use the following costs:

Match := 5, Mismatch :=-2 und Gap := -3. /4

**Multiple Sequence Alignments**

What is calculated with the ClustalW algorithm. Please state the three phases of this algorithm. /4

Define the concept of heuristics in bioinformatics. /2

**Homology**

Explain the terms Homology, Paralogy and Orthology in relation to biological sequences.

/4.5

**PAM**

The entries in a PAM matrix can be negative, positive or zero. What is the meaning of this?

/3

You look at two sequences with an identity of 60%. You want to make an Alignment with the help of a PAM Matrix – Which matrix would you use? Please explain your answer. /2

**PARSEN**

State the basic algorithm for parsing files. /2

You have the file down below. Is this file formatted according to Fasta standards? Please explain your answer. /2

>ProteinName1  
SYFPHEITI   
>ProteinName2  
YFPHEIEETI

ProteinName3  
SYFPHPIDTKEVTI

/35,5 ==> % ==> Note: