# Übungsblatt 2

Sequenz Alignment

Deadline: 11.5.2022 um 20:00 MEZ

Bioinformatik für Biochemiestudierende

Dr. Florian Klimm

Sommersemester 2022

### Aufgabe 2.1 Bewerten von Alignments



# Teilaufgabe 2.1.1 Prozentuale Übereinstimmung

### wenn, I > 30% wahrscheinlich zwei homologe Proteine

I < 20% unwahrscheinlich, dass zwei homologe Proteine

20% < I < 30% Graubereich

Teilaufgabe 2.1.2 Bewertung mit Substitutionsmatrix

	Α	Α	K	М	W	V
Α	<mark>4</mark>	4	-1	-1	-3	0
S	1	1	0	-1	-3	-2
K	-1	-1	<mark>5</mark>	-1	-3	-2
М	-1	-1	-1	<mark>5</mark>	-1	1
V	0	0	-2	1	<mark>-3</mark>	4
V	0	0	-2	1	-3	<mark>4</mark>

### Aufgabe 2.4 Needleman-Wunsch Algorithmus (30%)

Finden Sie das optimale Alignment der folgenden beiden DNA Sequenzen

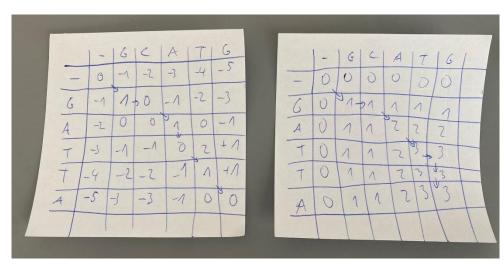


Wenden Sie (manuell, nicht am Computer) den Needleman-Wunsch Algorithmus an. Als Substitutionsmatrix verwenden Sie.. wobei  $\sigma$  die gap penalty ist.

Vergleichen Sie die Alignments mit  $\sigma = -1$  und  $\sigma = 0$ .

Links  $\sigma = -1$ 

Rechts  $\sigma = 0$ 



Vergleich  $\sigma = -1$  und  $\sigma = 0$ 

 $\sigma$  = -1:

- überwiegend negative Werte

Zahlenintervall von (-5 bis 2)

 $\sigma = 0$ 

- überwiegend positive Werte

# Aufgabe 2.5 Analyse einer unbekannten Sequenz (20%)

Sie haben die folgende Proteinsequenz sequenziert:
MVHLGPKKPQARKGSMADVPKELMDEIHQLEDMFTVDSETLRKVVKHFID
ELNKGLTKKGGNIPMIPGWVMEFPTGKESGNYLAIDLGGTNLRVVLVKLS
GNHTFDTTQSKYKLPHDMRTTKHQEELWSFIADSLKDFMVEQELLNTKDT
LPLGFTFSYPASQNKINEGILQRWTKGFDIPNVEGHDVVPLLQNEISKRE
LPIEIVALINDTVGTLIASYYTDPETKMGVIFGTGVNGAFYDVVSDIEKL
EGKLADDIPSNSPMAINCEYGSFDNEHLVLPRTKYDVAVDEQSPRPGQQA
FEKMTSGYYLGELLRLVLLELNEKGLMLKDQDLSKLKQPYIMDTSYPARI
EDDPFENLEDTDDIFQKDFGVKTTLPERKLIRRLCELIGTRAARLAVCGI
DAICQKRGYKTGHIAADGSVYNKYPGFKEAAAKGLRDIYGWTGDASKDPI
TIVPAEDGSGAGAAVIAALSEKRIAEGKSLGIIGA

Nun wollen Sie diese bioinformatisch untersuchen

Genutzt wurde folgende Seite: https://www.ebi.ac.uk/Tools/sss/ncbiblast/

#### Teilaufgabe 2.5.1 BLAST

• Welche BLAST Variante nutzen Sie? Warum?

blastp Variante

• Zu welchem Organismus gehört diese Sequenz wahrscheinlich?

Saccharomyces cerevisiae (Backhefe)

Zu welchem Protein gehört diese Sequenz wahrscheinlich?

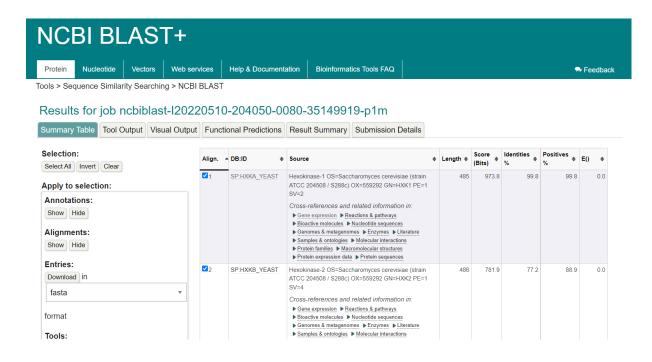
Hexokinase-1

• Wie groß ist die Percentage Identity mit diesem Protein?

99,8%

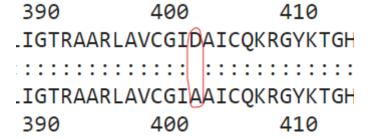
• Wie lautet der E-Score und wie interpretieren Sie ihn?

Wir haben einen Score von 0.0 da, E-Value < 0.01 wahrscheinlich homologe Sequenzen haben



#### Teilaufgabe 2.5.2 Mutation identifizieren

Hexokinase-1



Aminosäure an der Position 401 wurde D mit A ausgetauscht (Hier findet die Punktmutation statt)

\_

G <mark>A</mark> -> D G <mark>C</mark> -> A

Eine Punktmutation C zu A an der Stelle 1202 (401 \* 3)

Mit der Code-Sonne kommt folgendes raus: