**Aufgabe 2.1 Bewerten von Alignments (20 %)**  
In der Vorlesung haben wir verschiedene Möglichkeiten besprochen um das Align-  
ment von Sequenzen zu bewerten. Nun wollen wir das Alignment der folgenden  
Aminosäuresequenzen bewerten:

A A K M W V  
A S K M V V (1)

**Teilaufgabe 2.1.1 Prozentuale Übereinstimmung**  
Berechnen Sie die prozentuale Übereinstimmung der beiden Sequenzen. Sind die-  
se Sequenzen wahrscheinlich homolog? Begründen Sie Ihre Einschätzung.

M = Anzahl übereinstimmender Symbole

L = Länge der Sequenz

Mit einem I-Wert von weit über 30 % handelt es sich der Faustregel zufolge wahrscheinlich um homologe Sequenzen. Da die Sequenzen jedoch sehr kurz sind, lässt sich nicht vollkommen ausschließen, dass sich die Sequenzen trotz des hohen I-Wertes unabhängig entwickelt haben.

**Teilaufgabe 2.1.2 Bewertung mit Substitutionsmatrix**  
Recherchieren Sie die Werte der BLOSUM62 Substitutionsmatrix. Nutzen Sie diese  
um das Alignment zu bewerten.

**Aufgabe 2.2 Computergestütztes Bewerten von Alignments (30 %)**  
Schreiben Sie ein R das ein Alignment aus einer Textdatei lädt und die prozentuale  
Übereinstimmung berechnet. Wenden Sie dies auf die Textdatei alignedSequencesCalmodulin.txt  
an. Wie hoch ist die prozentuale Übereinstimmung beider Sequenzen?

Leider habe ich nach einigem Probieren keine Lösung bekommen, und frage mich, ob es daran liegt, dass ich die Sequenzen nicht als Vektoren vorliegen habe. Denn der Befehl %in% müsste mir ja eigentlich FALSE oder TRUE für alle Elemente geben, und ich bekomme als Antwort nur FALSE. Mit dem Operator == bin ich leider auch nicht weitergekommen. Es wäre super, wenn wir in der Übung nochmal For loops besprechen könnten, weil das wahrscheinlich eine Möglichkeit wäre, die characters paarweise zu vergleichen.

**Aufgabe 2.4 Needleman–Wunsch Algorithmus (30 %)**  
Finden Sie das optimale Alignment der folgenden beiden DNA Sequenzen

G C A T G  
G A T T A (2)

Wenden Sie (manuell, nicht am Computer) den Needleman-Wunsch Algorithmus  
an. Als Substitutionsmatrix verwenden Sie

**A G C T -  
A** +1 -1 -1 -1 σ  
**G** -1 +1 -1 -1 σ  
**C** -1 -1 +1 -1 σ  
**T** -1 -1 -1 +1 σ  
**-** σ σ σ σ 0

wobei σ die gap penalty ist. Vergleichen Sie die Alignments mit σ = −1 und σ = 0.

σ = −1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | - | G | A | T | T | A |
| - | 0 | -1 | -2 | -3 | -4 | -5 |
| G | -1 | +1 | 0 | -1 | -2 | -3 |
| C | -2 | 0 | 0 | -1 | -2 | -3 |
| A | -3 | -1 | +1 | 0 | -1 | 0 |
| T | -4 | -2 | 0 | +2 | +3 | +2 |
| G | -5 | -1 | -1 | 1 | 0 | +2 |

GCAT-G

G-ATTA

σ = 0

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | - | G | A | T | T | A |
| - | 0 | 0 | 0 | 0 | 0 | 0 |
| G | 0 | +1 | +1 | +1 | +1 | +1 |
| C | 0 | +1 | +1 | +1 | +1 | +1 |
| A | 0 | +1 | +2 | +2 | +2 | +2 |
| T | 0 | +1 | +2 | +3 | +3 | +3 |
| G | 0 | +1 | +2 | +3 | +3 | +3 |

Möglichkeit 1:

GCA-T-G

G-ATTA-

Möglichkeit 2:

GCA-TG-

G-ATT-A

Möglichkeit 3:

GCATG--

G-AT-TA

Möglichkeit 4:

GCAT--G

G-ATTA-

Möglichkeit 5:

GCAT-G-

G-ATT-A

**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.

**Teilaufgabe 2.5.1 BLAST**  
Nutzen Sie BLAST um die Sequenz mit Proteinsequenzen in der Datenbank zu  
vergleichen. Zeigen Sie einen Screenshot der Ergebnisliste.  
Beantworten Sie insbesondere die folgenden Fragen:

**• Welche BLAST Variante nutzen Sie? Warum?**

BLASTp, da eine Proteinsequenz vorliegt und diese mit Proteinsequenzen verglichen werden soll

**• Zu welchem Organismus gehört diese Sequenz wahrscheinlich?**

Saccharomyces cerevisiae

**• Zu welchem Protein gehört diese Sequenz wahrscheinlich?**

Hexokinase 1

**• Wie groß ist die Percentage Identity mit diesem Protein?**

99.79 %

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

Der E-Score für das Alignment ist Null, daher handelt es sich bei dem sequenzierten Protein höchstwahrscheinlich um Hexokinase 1.

**Teilaufgabe 2.5.2 Mutation identifizieren**  
Sie befürchten, dass das von Ihnen sequenzierte Protein eine schädliche Punkt-  
mutation aufweist. Finden Sie mit Hilfe bioinformatischer Programme heraus an  
welcher Stelle in der Sequenz diese auftritt und welche Aminosäure verändert  
wurde.

Alignment der Sequenzen mit Needle:

Table

Description automatically generated with medium confidence

Die Punktmutation führt zu einem Austausch der Aminosäure Alanin (A) gegen Asparaginsäure (D) an Aminosäure-Position 401 in dem von mir sequenzierten Protein. Ein Blick auf die Codonsonde verrät, dass die vorliegende Punktmutation ein Cytidin gegen ein Adenin austauscht (1202. Nukleotid des ORFs).