

# Übungsblatt 2

### Aufgabe 2.1 Bewerten von Alignments (20 %)

In der Vorlesung haben wir verschiedene Möglichkeiten besprochen um das Alignment von Sequenzen zu bewerten. Nun wollen wir das Alignment der folgenden Aminosäuresequenzen bewerten:

1. Prozentuale Übereinstimmung

|   |   |   |   |   |   |
|---|---|---|---|---|---|
| A | A | K | M | W | V |
| A | S | K | M | V | V |

(1)

$$T = \frac{s}{n}, \quad a \hat{=} \text{korrekt aligned}, \quad n \hat{=} \text{Gesamtzahl}$$

Hier:  $I = \frac{4}{7} \approx 57\%$

Nach der Faustregel, dass  $I > 30\%$  für homologe Sequenzen sind diese beiden Sequenzen homolog, allerdings ist eine Aminosäuresequenz von 7 Nukleotiden zu kurz, um eine sinnvolle Aussage treffen zu können.

2. Score:  $4 + 1 + 5 + 5 - 1 + 4 = 18$

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |   |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|---|
| Ala | 4   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |   |
| Arg | -1  | 5   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |   |
| Asn | -2  | 0   | 6   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |   |
| Asp | -2  | -2  | 1   | 6   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |   |
| Cys | 0   | -3  | -3  | -3  | 9   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |   |
| Gln | -1  | 1   | 0   | 0   | -3  | 5   |     |     |     |     |     |     |     |     |     |     |     |     |     |   |
| Glu | -1  | 0   | 0   | 2   | -4  | 2   | 5   |     |     |     |     |     |     |     |     |     |     |     |     |   |
| Gly | 0   | -2  | 0   | -1  | -3  | -2  | -2  | 6   |     |     |     |     |     |     |     |     |     |     |     |   |
| His | -2  | 0   | 1   | -1  | -3  | 0   | 0   | -2  | 8   |     |     |     |     |     |     |     |     |     |     |   |
| Ile | -1  | -3  | -3  | -3  | -1  | -3  | -3  | -4  | -3  | 4   |     |     |     |     |     |     |     |     |     |   |
| Leu | -1  | -2  | -3  | -4  | -1  | -2  | -3  | -4  | -3  | 2   | 4   |     |     |     |     |     |     |     |     |   |
| Lys | -1  | 2   | 0   | -1  | -3  | 1   | 1   | -2  | -1  | -3  | -2  | 5   |     |     |     |     |     |     |     |   |
| Met | -1  | -1  | -2  | -3  | -1  | 0   | -2  | -3  | -2  | 1   | 2   | -1  | 5   |     |     |     |     |     |     |   |
| Phe | -2  | -3  | -3  | -3  | -2  | -3  | -3  | -3  | -1  | 0   | -3  | 0   | -3  | 0   | 6   |     |     |     |     |   |
| Pro | -1  | -2  | -2  | -1  | -3  | -1  | -1  | -2  | -2  | -3  | -3  | -1  | -2  | -4  | 7   |     |     |     |     |   |
| Ser | 1   | 1   | 1   | 0   | -1  | 0   | 0   | 0   | -1  | -2  | -2  | 0   | -1  | -2  | -1  | 4   |     |     |     |   |
| Thr | 0   | -1  | 0   | -1  | -1  | -1  | -1  | -2  | -2  | -1  | -1  | -1  | -2  | -1  | 1   | 5   |     |     |     |   |
| Tyr | -3  | -3  | -4  | -4  | -2  | -2  | -3  | -2  | -3  | -2  | -3  | -1  | -1  | -4  | -3  | -2  | 11  |     |     |   |
| Val | -2  | -2  | -2  | -3  | -2  | -1  | -2  | -3  | 2   | -1  | -1  | -2  | -1  | 3   | -3  | -2  | 2   | 7   |     |   |
|     | 0   | -3  | -3  | -3  | -1  | -2  | -2  | -3  | -3  | 3   | 1   | -2  | 1   | -1  | -2  | -2  | 0   | -3  | -1  | 4 |
|     | Ala | Arg | Asn | Asp | Cys | Gln | Glu | Gly | His | Ile | Leu | Lys | Met | Phe | Pro | Ser | Thr | Tyr | Val |   |

Für eine so kurze Sequenz ist dies ein relativ hoher Score, weswegen von einer Homologie auszugehen ist.

### 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?

siehe uebung 2.R

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

Finden Sie das optimale Alignment der folgenden beiden DNA Sequenzen.

$$\begin{array}{ccccc} G & C & A & T & G \\ G & A & T & T & A \end{array} \quad (2)$$

|   | A        | G        | C        | T        | -        |
|---|----------|----------|----------|----------|----------|
| A | +1       | -1       | -1       | -1       | $\sigma$ |
| G | -1       | +1       | -1       | -1       | $\sigma$ |
| C | -1       | -1       | +1       | -1       | $\sigma$ |
| T | -1       | -1       | -1       | +1       | $\sigma$ |
| - | $\sigma$ | $\sigma$ | $\sigma$ | $\sigma$ | 0        |

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

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

$$6 = 0$$

### Alignments:

I - G C A T - G -  
- G - A T T - A

|    |   |   |   |   |   |   |   |   |
|----|---|---|---|---|---|---|---|---|
| II | - | G | C | A | T | G | - | - |
|    | - | G | - | A | T | - | T | A |

III - G C A - T - G  
- G - A T T A -

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

$$G = -1$$

## Alignment

|   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|
| I | - | G | C | A | T | - | G |
|   | - | G | - | A | T | T | A |

## Aufgabe 2.5 Analyse einer unbekannten Sequenz (20%)

### 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?
- Zu welchem Organismus gehört diese Sequenz wahrscheinlich?
- Zu welchem Protein gehört diese Sequenz wahrscheinlich?
- Wie groß ist die Percentage Identity mit diesem Protein?
- Wie lautet der E-Score und wie interpretieren Sie ihn?

Es wird **blastp** genutzt, da die vorliegende Sequenz aus den Buchstaben des einzelnen Aminosäuren besteht und somit die Primärsequenz eines Proteins beschreibt. Demnach sollte diese Sequenz mit anderen Proteinen der Datenbank verglichen werden.

Die Sequenz gehört sehr wahrscheinlich zur

**Hexokinase 1** der **Bäckerhefe**  
(*Saccharomyces cerevisiae*)

Beim besten Alignment beträgt die Percentage Identity 99,79%

Der E-Score liegt bei 0,0, womit es nahezu ausgeschlossen ist, dass die Übereinstimmung zufällig auftritt.

|                                     |                                                                   |                                   |     |     |      |     |        |     |                             |
|-------------------------------------|-------------------------------------------------------------------|-----------------------------------|-----|-----|------|-----|--------|-----|-----------------------------|
| <input checked="" type="checkbox"/> | <a href="#">hexokinase 1 [Saccharomyces cerevisiae S288C]</a>     | <a href="#">Saccharomyces ...</a> | 953 | 953 | 100% | 0.0 | 99.79% | 485 | <a href="#">NP_116711.3</a> |
| <input checked="" type="checkbox"/> | <a href="#">hexokinase isoenzyme 1 [Saccharomyces cerevisiae]</a> | <a href="#">Saccharomyces ...</a> | 953 | 953 | 100% | 0.0 | 99.59% | 485 | <a href="#">AEP68292.1</a>  |
| <input checked="" type="checkbox"/> | <a href="#">Hxk1p [Saccharomyces cerevisiae YJM693]</a>           | <a href="#">Saccharomyces ...</a> | 951 | 951 | 100% | 0.0 | 99.38% | 485 | <a href="#">AJU37745.1</a>  |
| <input checked="" type="checkbox"/> | <a href="#">Hxk1p [Saccharomyces cerevisiae YJM1129]</a>          | <a href="#">Saccharomyces ...</a> | 951 | 951 | 100% | 0.0 | 99.38% | 485 | <a href="#">AJU38858.1</a>  |
| <input checked="" type="checkbox"/> | <a href="#">Hxk1p [Saccharomyces cerevisiae YJM1386]</a>          | <a href="#">Saccharomyces ...</a> | 951 | 951 | 100% | 0.0 | 99.38% | 485 | <a href="#">AJV23225.1</a>  |
| <input checked="" type="checkbox"/> | <a href="#">hexokinase isoenzyme 1 [Saccharomyces cerevisiae]</a> | <a href="#">Saccharomyces ...</a> | 950 | 950 | 100% | 0.0 | 99.38% | 485 | <a href="#">AEP68272.1</a>  |
| <input checked="" type="checkbox"/> | <a href="#">hexokinase isoenzyme 1 [Saccharomyces cerevisiae]</a> | <a href="#">Saccharomyces ...</a> | 950 | 950 | 100% | 0.0 | 99.38% | 485 | <a href="#">AEP68293.1</a>  |
| <input checked="" type="checkbox"/> | <a href="#">Hxk1p [Saccharomyces cerevisiae FostersB]</a>         | <a href="#">Saccharomyces ...</a> | 949 | 949 | 100% | 0.0 | 99.18% | 485 | <a href="#">EGA58845.1</a>  |
| <input checked="" type="checkbox"/> | <a href="#">hexokinase isoenzyme 1 [Saccharomyces cerevisiae]</a> | <a href="#">Saccharomyces ...</a> | 949 | 949 | 100% | 0.0 | 99.18% | 485 | <a href="#">AEP68270.1</a>  |
| <input checked="" type="checkbox"/> | <a href="#">Hxk1p [Saccharomyces cerevisiae YJM195]</a>           | <a href="#">Saccharomyces ...</a> | 949 | 949 | 100% | 0.0 | 98.97% | 485 | <a href="#">AJU35571.1</a>  |

### Teilaufgabe 2.5.2 Mutation identifizieren

Sie befürchten, dass das von Ihnen sequenzierte Protein eine schädliche Punktmutation aufweist. Finden Sie mit Hilfe bioinformatischer Programme heraus an welcher Stelle in der Sequenz diese auftritt und welche Aminosäure verändert wurde.

Unter der detaillierten Ansicht des besten Alignments in BLAST ist es möglich, das Alignment grafisch zu untersuchen, wobei Mismatches rot hervorgehoben werden. So zeigt sich, dass folgende Punktmutation auftritt:

An Stelle **401** ist im Wildtyp ein **Alanin (A)**, in der untersuchten Sequenz aber eine **Asparaginsäure (D)**

