

1 Bioinformatik Übung 2

Sarah Lange

1.1

A A K M W V

A S K M V V

I = prozentuale Übereinstimmung | M = Anzahl der Übereinstimmungen | L = Länge der Sequenz

$$M = 4 \quad L = 6$$

$$I = \frac{100 \cdot M}{L}$$
$$I = \frac{100 \cdot 4}{6} = 66,7\%$$

Bei der prozentualen Berechnung gilt als Faustregel, wenn $I > 30\%$, dann sind die Proteine wahrscheinlich homolog. Da bei der Berechnung ein Prozentwert von $66,7\%$ heraus kam, kann also davon ausgegangen werden, dass die zwei Sequenzen homolog sind.

1.2

[illegible]

: Abbildung 1

BLOSUM62 Matrix

Quelle: https://students.aertslab.org/ebook/chapter6_alignments.html (5.5.22, 12:15 Uhr)

A A K M W V

A S K M V V $\Rightarrow +4 -1 +5 +5 -3 +4 = 14$

1.3

siehe Code

1.4 Needleman-Wunsch-Algorithmus

$\sigma = -1$	-	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	+1	0
G	-5	-3	-1	+1	+1	0

Abbildung 2: Needleman-Wunsch-Algorithmus mit $\sigma = -1$

Optimales Alignment :

G C A T G -

G G A T T A

$\sigma = 0$	-	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

Abbildung 3: Needleman-Wunsch-Algorithmus mit $\sigma = 0$

Optimales Alignment :

G G A - T - G

G - A A T A -

1.5

1.5.1

The screenshot shows the NCBI BLAST search results for a protein sequence. The job title is "Protein Sequence" with RID 79V4FXJW013. The program used is BLASTP. The database is nr. The query ID is lcl|Query_19967. The description is "unnamed protein product". The molecule type is amino acid. The query length is 485. The filter results section shows the organism filter set to "only top 20 will appear" and the "Add organism" button. The "Sequences producing significant alignments" table is displayed with columns: Description, Scientific Name, Max Score, Total Score, Query Cover, E value, Per. Ident, Acc. Len, and Accession. The table lists several sequences from Saccharomyces cerevisiae, including hexokinase 1 and hexokinase isoenzyme 1, with high scores and low E values.

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
hexokinase 1 [Saccharomyces cerevisiae S288C]	Saccharomyces...	953	953	100%	0.0	99.79%	485	NP_116711.3
hexokinase isoenzyme 1 [Saccharomyces cerevisiae]	Saccharomyces...	953	953	100%	0.0	99.59%	485	AEF68292.1
Hxk1p [Saccharomyces cerevisiae YJM693]	Saccharomyces...	951	951	100%	0.0	99.38%	485	AJU37745.1
Hxk1p [Saccharomyces cerevisiae YJM1129]	Saccharomyces...	951	951	100%	0.0	99.38%	485	AJU38858.1
Hxk1p [Saccharomyces cerevisiae YJM1386]	Saccharomyces...	951	951	100%	0.0	99.38%	485	AJV23225.1
hexokinase isoenzyme 1 [Saccharomyces cerevisiae]	Saccharomyces...	950	950	100%	0.0	99.38%	485	AEF68272.1
hexokinase isoenzyme 1 [Saccharomyces cerevisiae]	Saccharomyces...	950	950	100%	0.0	99.38%	485	AEF68293.1
Hxk1p [Saccharomyces cerevisiae FostersB]	Saccharomyces...	949	949	100%	0.0	99.18%	485	EGA58845.1

Abbildung 4: Ergebnisliste der BLAST-Suche

1. Welche BLAST Variante nutzen Sie? Warum?

- Verwendung von blastp
- vergleicht Proteinsequenz mit der Proteinen in Datenbank
- kann daher anzeigen, zu welchem Protein diese Sequenz gehört, wonach in der Aufgabenstellung gefragt wird
- kann zusätzlich Organismen mit ähnlichen Proteinen anzeigen, auch danach ist gefragt

2. Zu welchem Organismus gehört diese Sequenz wahrscheinlich?

- Saccharomyces cerevisiae (Backhefe)

3. Zu welchem Protein gehört diese Sequenz wahrscheinlich?

- Hexokinase I

4. Wie groß ist die Percentage Identity mit diesem Protein?

- 99.79%

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

- E-Score ist hier 0.0
- dies deutet darauf hin, dass die Proteine wahrscheinlich homolog sind

1.5.2

Alignment view Pairwise with dots for identities [Restore defaults](#)

99 sequences selected

[Download](#) [GenPept](#) [Graphics](#)

hexokinase 1 [Saccharomyces cerevisiae S288C]
 Sequence ID: [NP_116711.3](#) Length: 485 Number of Matches: 1
[See 7 more title\(s\)](#) [See all Identical Proteins\(IPG\)](#)

Range 1: 1 to 485 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
953 bits(2464)	0.0	Compositional matrix adjust.	484/485(99%)	484/485(99%)	0/485(0%)
Query 1	MVHLGPKKPQARKGSMADVPKELMDEIHQLEDMTVDSETLRKVVKHFIDELNKGKTKKG	60			
Sbjct 1	60			
Query 61	GNIPMIPGWVMEFPTGKESGNLYLAIDLGGTNLRVVLKLSGNHTFDTTQSKYKLPMDMRT	120			
Sbjct 61	120			
Query 121	TKHQEELWSFIADSLKDFMVEQELLNTKDTLPLGFTFSYPASQNKINEGILQRWTKGFDI	180			
Sbjct 121	180			
Query 181	PNVEGHDVPLQLQNEISKRELPIEIVALINDTVGTLIASYYTDPETKMGVIFGTGVNGAF	240			
Sbjct 181	240			
Query 241	YDVVSDIEKLEGKLADDIPSNSPMAINCEYGSFDNEHLVLPRTKYDVAVDEQSPRPGQQA	300			
Sbjct 241	300			
Query 301	FEKMTSGYYlgellrlvlllelnekglmlKDQLSKLKQPYIMDTSYPARIEDDPFENLED	360			
Sbjct 301	360			
Query 361	TDDIFQKDFGVKTTLPERKLIRRLCELIGTRAARLAVCGIDAICQKRGYKTGHIAADGSV	420			
Sbjct 361 A	420			
Query 421	YNKYPGFKEAAAKGLRDIYGTGDASKDPITIVPAEDGSGAGAAVIAALSEKRIAEGKSL	480			
Sbjct 421	480			
Query 481	GIIGA 485				
Sbjct 481 485				

Abbildung 5: Alignment der vorgegebenen Sequenz mit der Vergleichssequenz, in rot ist die Mutation markiert. Wurde mithilfe von BLAST erstellt

Die Aminosäure A ist zur Aminosäure D mutiert. Wo genau, ist anhand der Grafik abzulesen.