

Softwarewerkzeuge der Bioinformatik

Prof. Dr. Volkhard Helms
PD Dr. Michael Hutter, Markus Hollander,
Marie Detzler
Winter Semester 2020/2021

Saarland University
Center for Bioinformatics

Exercise Sheet 4

Sequence Analysis: Motifs

Learning objective: *The goal is to learn the definition of sequence motifs, how they are visualised, and how to employ various motif databases and tools. Additionally, you are going to construct a PSSM and find the corresponding consensus sequence.*

Exercise 4.1: PRINTS and PROSITE

- a) Look up the protein cytochrom C (**P00044**) in PRINTS (<http://bioinf.man.ac.uk/dbbrowser/PRINTS/index.php>). UniProtKB might be useful for this.
 - i. How many motifs are there?
 - ii. How many proteins were used to generate the motifs?
 - iii. Does **P00044** contain all motifs?
 - iv. Locate the motifs in the amino acid sequence.
- b) Compare the potential motifs of the Clustal Omega alignment of exercise 3.3 with the entries for FOSB_MOUSE in the databases Datenbanken PROSITE (<http://www.expasy.org/prosite>) and PRINTS.
- c) Which PRINTS motifs correspond approximately to the PROSITE signature?

Hint:

Accession numbers:

Proteins have different accessions in UniProtKB, PROSITE and PRINTS.

Components of PROSITE signatures:

- [] Amino acids that can be found at this position
- next position
- x any amino acid
- () number of repetitions
- { } amino acids that are not allowed at this position

Exercise 4.2: MEME and MAST

<http://alternate.meme-suite.org>

- a) Use the proteins that were used to generate the PRINTS motifs for FOSB_MOUSE to generate MEME motifs.
- b) Compare the MEME motifs with the PRINTS motifs.
- c) Use MAST to test if FOSB_MOUSE contains the MEME motifs.

Exercise 4.3: Position-specific scoring matrix (PSSM)

Compute the PSSM and consensus sequence of the following alignment. (X represents all amino acids not listed here.)

S E Q V E N C E	alignment matrix →	1	2	3	4	5	6	7	8		
S E Q Y E M C E		C	E	I	M	N	Q	S	V	Y	X
S I Q V I N S E		E	I	M	N	Q	S	V	Y	X	
S E Q V E - C E		I	M	N	Q	S	V	Y	X		
S E Q V E M C I		M	N	Q	S	V	Y	X			
S E Q V E N S I		N	Q	S	V	Y	X				
* * .		Q	S	V	Y	X					
consensus:		S	V	Y	X						

	1	2	3	4	5	6	7	8
C								
E								
I								
M								
N								
Q								
S								
V								
Y								
X								

PSSM

$\forall i : p_i = 0.05$
a priori probability

Have fun!