

Question 1 UNIX (8 points)

What effect do these UNIX commands have? Describe in one sentence: (4 points)

a.) `cd ../..` change directory

.. eins höher in die parent directory
 • = noch eins höher ✓

b.) `mkdir` `assignment13`

erstellt neue directory mit Name im root Verzeichnis
 assignment13 in dem Ordner in dem man ist = current directory
 / ist nur wichtig wenn am Anfang

c.) `rm -rf *.doc`

recursively force → löscht auch Schreibgeschützte Dateien
 ↳ keine Bestätigung, macht
 löscht alle .doc Dokumente mit beliebigem Text ✓
 + zeigt die Zeile an

[rm -r kann verschluckt werden in denen noch Daten sind]

d.) `grep protein wheat.fasta | wc -w`

man durchsucht wheat.fasta nach Protein
 zählt die Nummer an wörtern dem wort Protein
 worte in den Zeilen in denen "protein" vorkommt

Write the UNIX commands to use if you want to do the following: (4 points)

e.) Display the manual for grep.

`man grep` ✓

f.) Copy the file `bachelor.thesis.alphaversion.txt` to `pure.nonsense.txt`.

`cp x y` ✓

g.) Find all lines in `bachelor.thesis.betaversion.txt` that contain 'obscure' and put them into `less_nonsense.txt`.

`grep obscure x > less_nonsense.txt`
 redirection operator

h.) Unpack and decompress `archive.tar.gz`.

`tar xzf archive.tar.gz` ✓

entweder 2 mal z
 oder garnicht!

- Lists, slices

→ Funktionen:

Block: Funktionsbedingung } Doppelpunkt
- if
- for - Schleifen

Question 2 Python (8 points)

Examine the following code snippet:

```
be = range(10)
for i in range(1, 11):
    for j in range(1, 11):
        print i, j, i * j
```

→ immer ord Einrückten
Runde Klammern bei Range
Doppelpunkt hinter forschleife
beachten bei " " (immer von Zeichen hinter range)

a.) Spot all errors and explain why these are errors! (2 points)

[→ Optionen und nicht Bedingungen

: → Argument

wenn eine for-Schleife geöffnet wird kommt ein Doppelpunkt

allgemein auf Einrückung achten; strings; Anführungszeichen

b.) What would the corrected code print? (2 points)

```
for i in range(1, 11):
    for j in range(i, 11):
        print i, 'x', j, '=', i * j
```

→ zählt von 1-10

→ zählt von i bis 10

was ist x was j ist = und dann die Rechnung

c.) Define a Python function dividebyten that divides its argument by ten. (2 points)

→ Parameter + Rückgabewert

def

```
def dividebyten(x):
    return x/10
```

d.) You are given a Python string as variable named seq. Give one (!) line of code that prints the content of seq backwards. (2 points)

```
print seq[::-1]
```

Start ↑
↓ Ende

fängt bei 1 an

↓

1 · 1 , 1 · 2 , 1 · 3 ... 1 · 10

→ dann 2:

2 · 2 , 2 · 3 ... 2 · 10

3 · 3 ... 3 · 10

4 · 4 ... 4 · 10

5 · 5 ...

3

10 · 10

Question 3 Alignments (6 points)

Fill the distance matrix for an alignment of the Sequences using the Needleman-Wunsch algorithm. Use the following distance function and a gap cost of 1.

$$\text{distance}(a, b) = \begin{cases} 0, & \text{if } a = b \\ 1, & \text{else} \end{cases}$$

and highlight the alignment path in the traceback matrix and give the alignment:

DP-Matrix

A

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

Traceback-Matrix

	-	A	G	C	T
-					
A					
T					
G					
C					
A					

Nullen darf man nur vergeben wenn man von Diagonal kommt

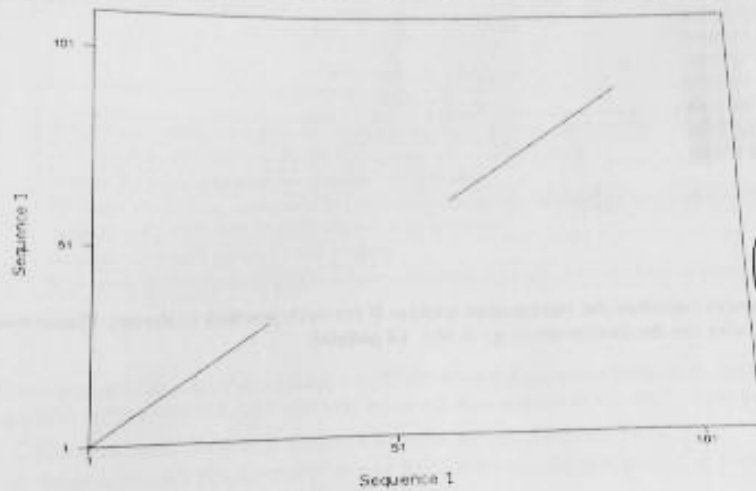
A A T G C A
 B A - G C T
 2 = 0 1 0 0 1

A
B

1 von rei
 - 1 von oben
 0 diagonal

Question 4 Alignment2 (6 points)

a.) What can you deduce about the sequence of the protein for which a dotplot is shown here? (3 points)



(Vormalerweise wäre Hauptdiagonale durchgängig)

Heterodimer ??

Sequenz, die Inversionen enthält

Fenster: Länge 5 → man sch. auf SAS an

Flasum ?? = score

threshold = Fenster in Dotplot anschauen

b.) Explain how the parameters window size and threshold influence the dotplot. (3 points)

Question 5 Phylogenies (10 points)

a.) Give the consensus sequence for this multiple sequence alignment. (2 points)

```

MVLGAADPTNVKAAWS
MVLGPADKTNVKAAG
VQLSGEENKAVLALWD
VHLTFEERKSVITALNG
APLSAENETKIRSAWA
MVLSEETWOLNLFVWA
GALTESONALVKESE
    
```

- kommt nur vor: A
 - nicht konserviert: a
 - meist mehrfach: a
 - sonst: .

mvLs.aeht.vb aaW.

b.) For the given sequences calculate the information content R for each position (column). Please round each result to two digits after the decimal point (e.g.: 0.56). (4 points)

es war kommt nicht dran

12,48

c.) Draw the phylogenetic tree from the following CLUSTALW output. Describe your method. Note that the length of the edges of the tree is not relevant for the drawing. Label the taxa in the tree! (4 points)

```

{
gummybear:0.29688,
pooh:0.39062)
:0.03906,
{
brownbear:0.06875,
grizzly:0.05625)
:0.16667,
teddybear:0.30208)
:0.07031,
{
moonbear:0.16406,
coati:0.21094)
:0.14844);
    
```

was in einer Klammer ist



Question 6 Protein structure elucidation & prediction (6 points)

a.) Which of the following statements are true for which method? Check all statements that are correct for the respective method. (3 points)

	XRD	NMR
Requires computer models to be fitted to experimental electron density data	☒	☐
Also yields information about the dynamics of the system	☐	☒
Suited for huge proteins or protein complexes	☒	☐
Proteins should be isotopically labeled as the predominant naturally occurring isotope ^{12}C does not exhibit absorption spectra.	☐	☒
Requires crystallisation of the protein	☒	☒
The protein database (PDB) contains structures analysed with this method	☒	☒

b.) You are given the task of predicting the 3D structure of a protein from its sequence. Searching the sequence against UniProt with BLAST (default settings) does not yield any hits. (3 points)

- Can you apply threading methods for structure prediction? *Nein*
- What additional requirements would have to be met for threading, i.e. what additional data would you need? *Template, = Schabene*
- If threading is not applicable, which other methods could be used?
Ab Initio, Rosetta

Question 7 PDB Format (6 points)

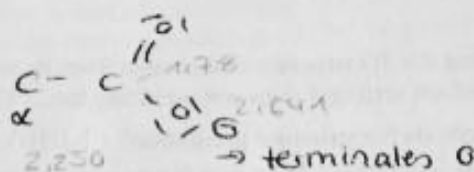
You are given a PDB entry excerpt:

ATOM	1	N	???	1	0.973	1.129	-2.010
ATOM	2	CA → Cα	???	1	2.250	0.643	-1.332
ATOM	3	C	???	1	1.780	0.000	0.000
ATOM	4	X	xxxxxx	xxxxxx	xxxxxx	xxxxxx	xxxxxx
ATOM	5	OXT → O	???	1	2.641	-0.376	0.824
ATOM	6	CB → Cβ	???	1	3.191	1.817	-1.106
ATOM	7	SG	???	1	4.875	1.267	-0.688

↓ Sing pos.

→ Cysteine

- a.) Give the name and the structure of the given amino acid, hidden by question marks (3-letter code). (2 points)



- b.) Which atom is hidden by 'xxx'? Print the complete ATOM record under the assumption that a carbonyl bond has a length of 124 pm. (Please disregard the proper number of blanks.) (4 points)

→ O

$$\begin{array}{r}
 1.78 \\
 - 1.24 \\
 \hline
 0.54
 \end{array}$$

Question 8 Ab-initio prediction of protein structures (6 points)

a.) In which case a the switch to a new conformation is accepted with a 100% probability in the Metropolis-Monte-Carlo algorithm? (3 points)

- wenn der neue score schlechter ist von Temp. abhängig.
- wenn neuer score besser ist akzeptieren wir den immer

b.) Describe the role of the parameter T (absolute temperature) in ROSETTA in one sentence. (3 points)

- Von der Temperatur hängt ab, mit welcher Wahrscheinlichkeit ein schlechterer Score akzeptiert wird (simulated annealing)

Question 9 Miscellaneous (4 points)
Check all that apply.

	true	false
The protein database contains only gene sequences.	☒	☒
Repeated execution of ROSETTA with the same input will yield different results.	☒	☒
Runtime complexity of the Needleman-Wunsch algorithm runtime is $O(n^2m^2)$	☒	☒
Proline is often found in the middle of α -helices.	☒	☒
The UniProt database contains curated and uncurated protein sequences	☒	☒
Rotamers are conformers and thus represent local energy minima.	☒	☒

wg zufall

$O(n^2)$

Proline ist ein helixbrecher