

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

root → ohne / | current directory | / ist nur wichtig wenn am Anfang

erstellt neue directory mit Name im root Verzeichnis
assignment13 in dem Ordner in dem man ist = current directory

c.) `rm -rf *.doc`

→ recursive force → löscht auch Schreibgeschützte Dateien
↳ keine Bestätigung, macht

könnte in dem Fall weggegangen werden
remove

→ löscht alle .doc Dokumente mit beliebigem Text ✓

[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

+ zeigt die Zeile an

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.

~~cp~~ `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!

- Listen, Slices

→ Funktionen:

Block: Funktionsbedingung } Doppelpunkt
if
- for - Schleifen

Question 2 Python (8 points)

Examine the following code snippet:

```
be range(10) → 0-9  
for i in range(1,11): Runde Klammern bei Range  
    for j in range(1,11): Doppelpunkt hinter for-Schleife  
        print i, j, i * j
```

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 (immer mit Zöckchen markieren)

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

```
for i in range(1,11) → zählt von 1-10  
    for j in range(i,11) → zählt von i bis 10  
        print i, '*', j, '=', i * j
```

was initial x was in j ist = und dann die Rechnung

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

→ Parameter
+ Rückgabewert

```
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
↑ stop
↓ 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	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					

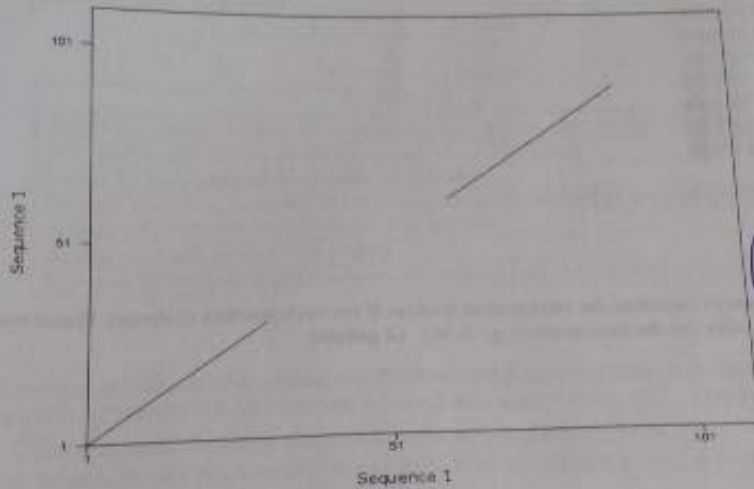
Nutzen darf man nur vergeben wenn man von Diagonal kommt

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

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)



(Normalerweise 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: Phylogenetics (10 points)

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

```

MVLGAACPTNVRAANS
MVLGPADRTNVRAANG
VQLGGEKPAVLAALWD
VHLTFEERSVITALLNG
APLSAENTKIRSAWA
MVLSETEWOLVLEVWA
GALTESONALVKESE
    
```

- kommt nur vor
 - konserviert: A
 - nicht konserviert:
 - aber 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 konserviert 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)

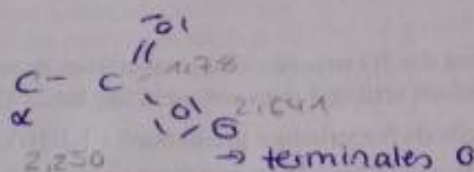
You are given a PDB entry excerpt:

ATOM	1	N	???	1	0.973	1.129	-2.010
ATOM	2	CA → Cd	???	1	2.250	0.643	-1.332
ATOM	3	C	???	1	1.780	0.000	0.000
ATOM	4	x Gxxxxx xxxxx xxxxxxxx	1	2.641	-0.376	0.824	
ATOM	5	OXT → Terminates	1	3.191	1.817	-1.106	
ATOM	6	CB → Cβ	1	4.875	1.267	-0.688	
ATOM	7	SG	1				

↓ Sing pos.

→ Cystein

- 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

