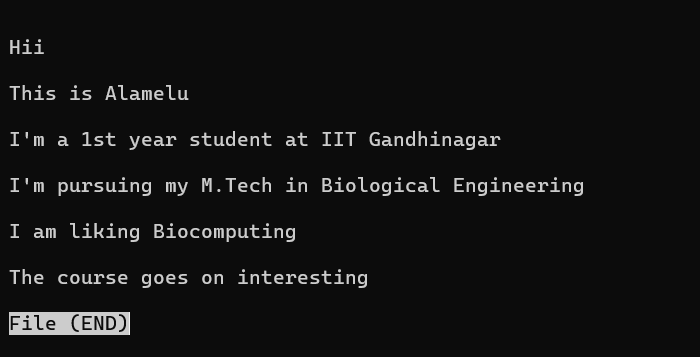
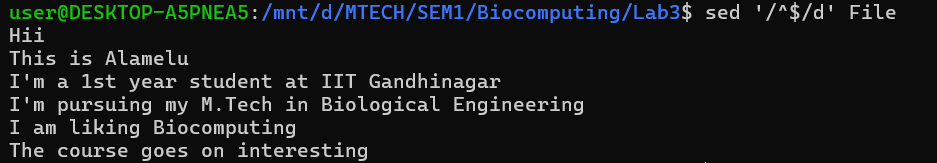
I have used CHATGPT for some difficult questions between 20 and 26 and I have also got hints for some questions from 1-20 from these AI tools only to understand logic which are different from the Lab session 3 codes about sed and awk.

1)

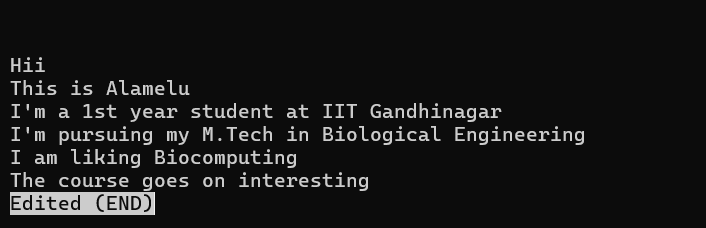




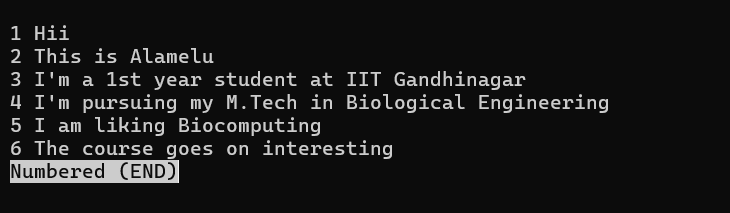


2)









3)



4)



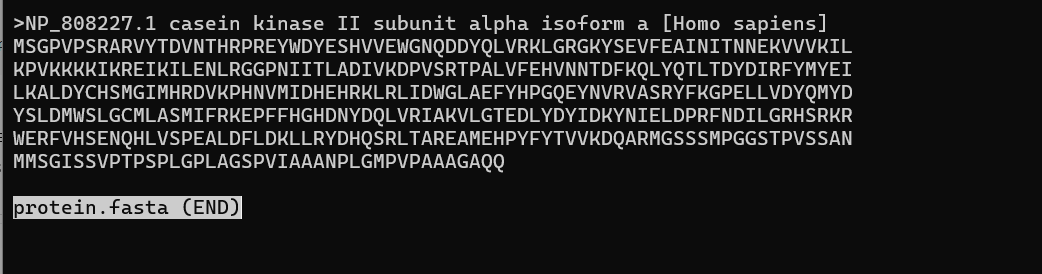
No headers found in protein.fasta that contains the word clock

5)



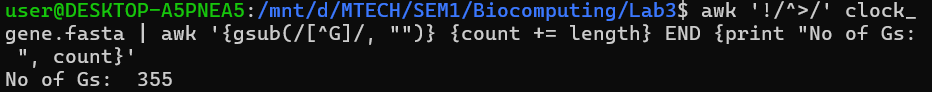


(less protein.fasta was executed to confirm if no lines with CC were present)

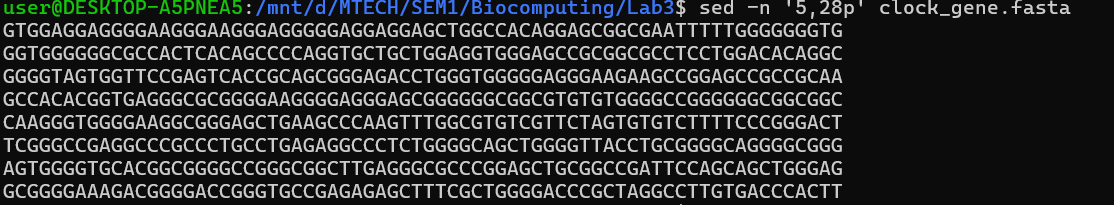


There are no such lines in protein.fasta with two consecutive CC s.

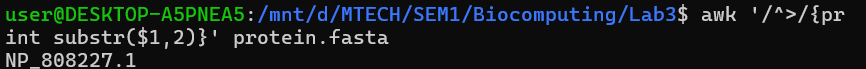
6)



7)



8)



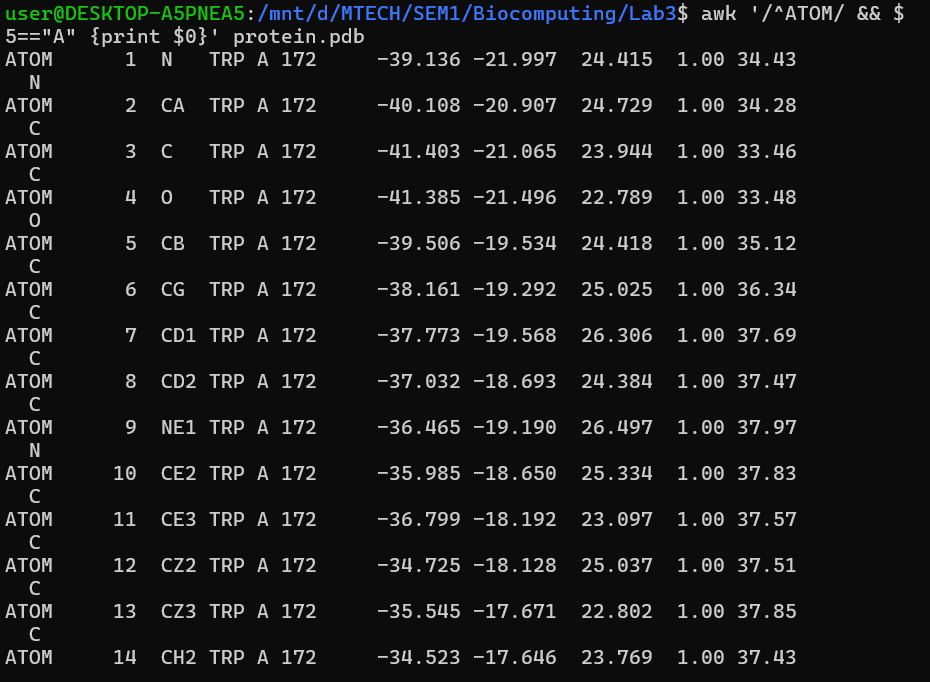
9)1)



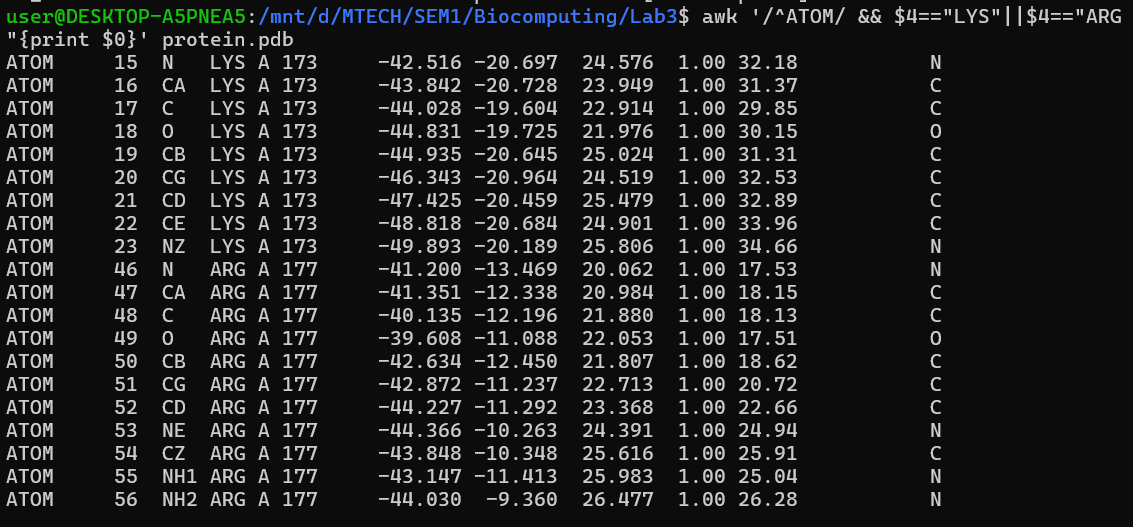
9)2)



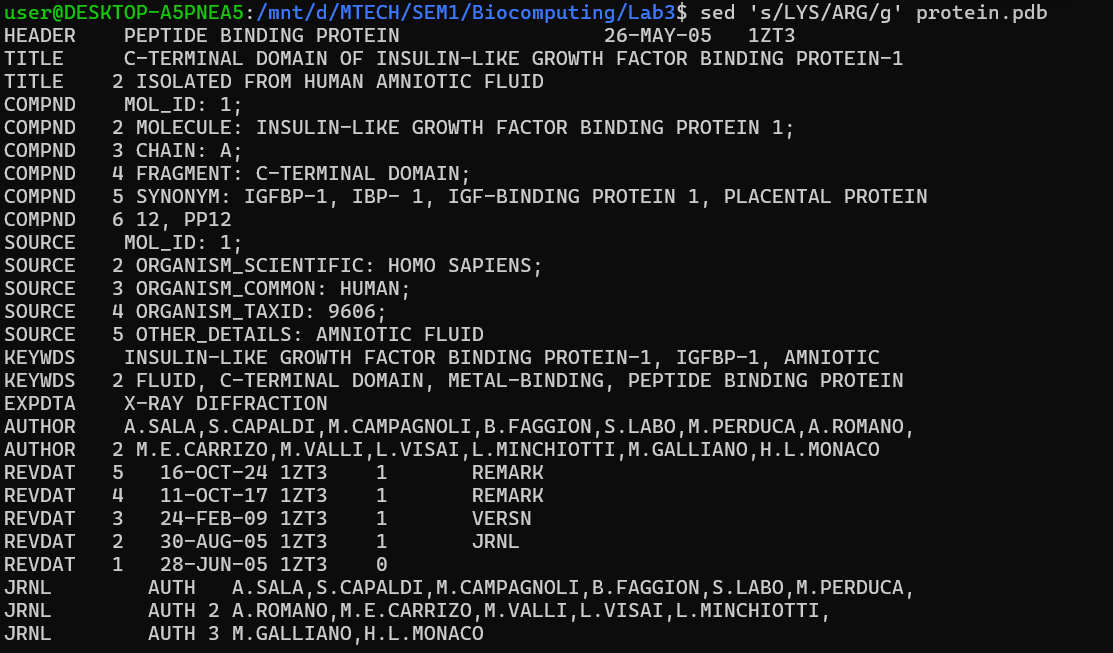
10)



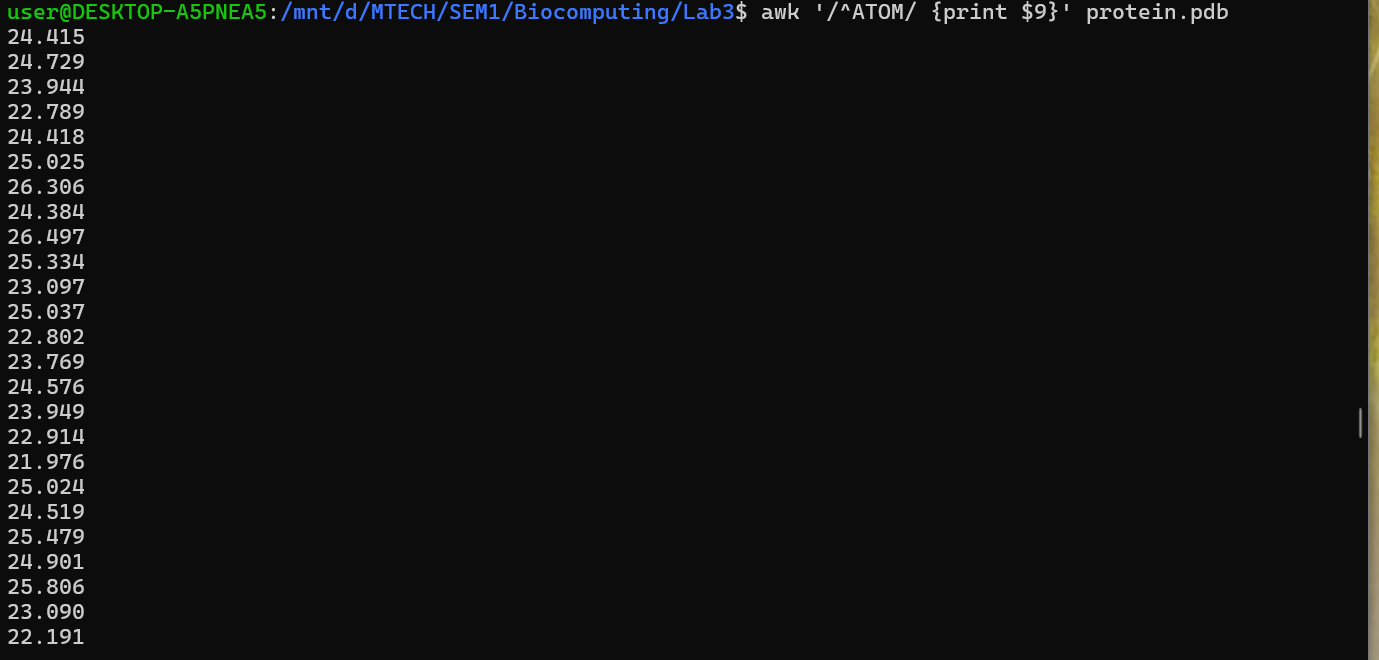
11)



12)



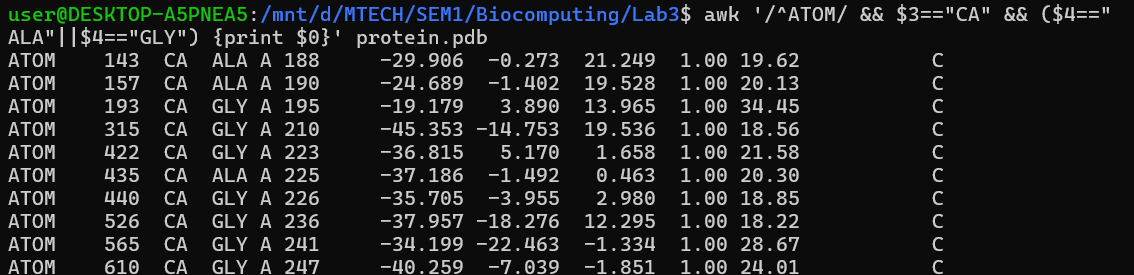
13)



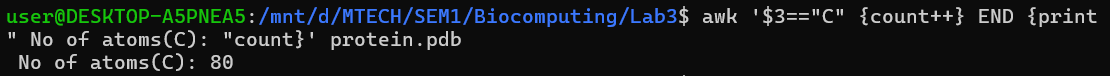
14)



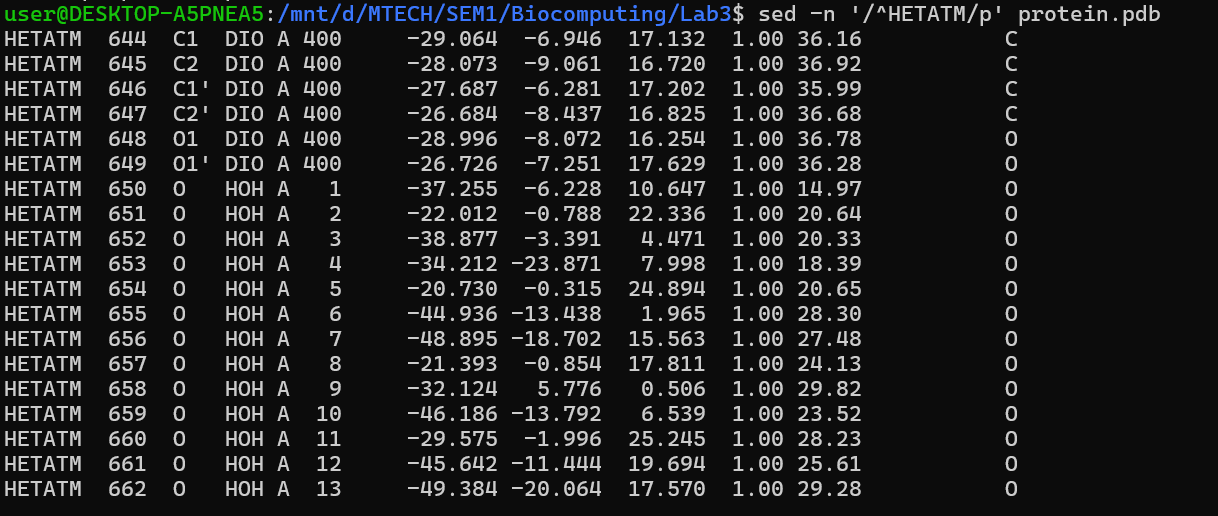
15)



16)

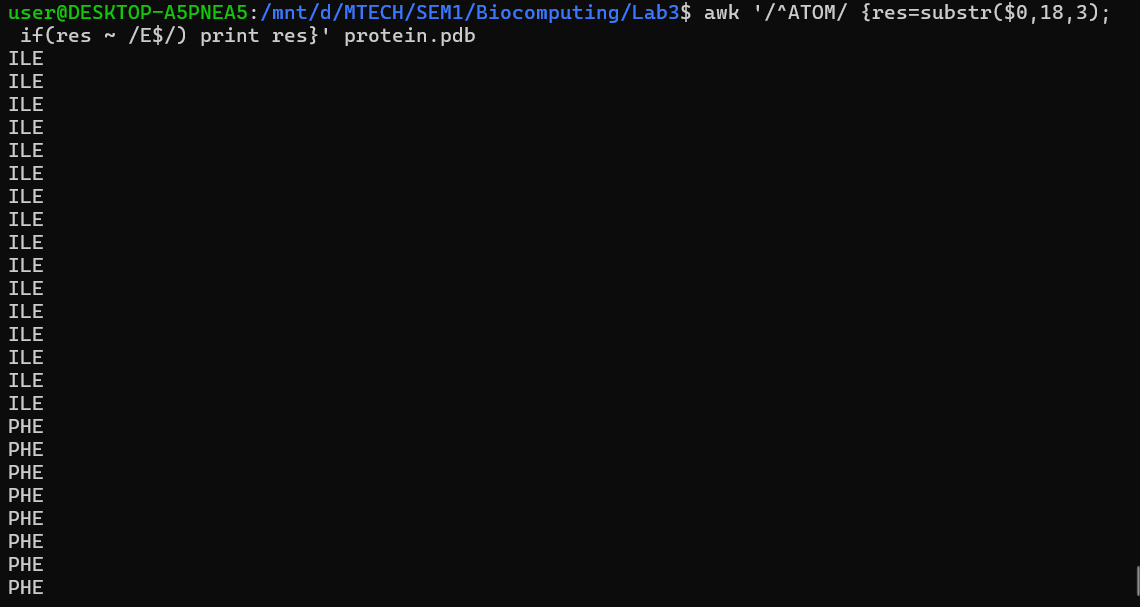


17)

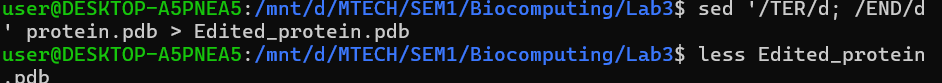


18)

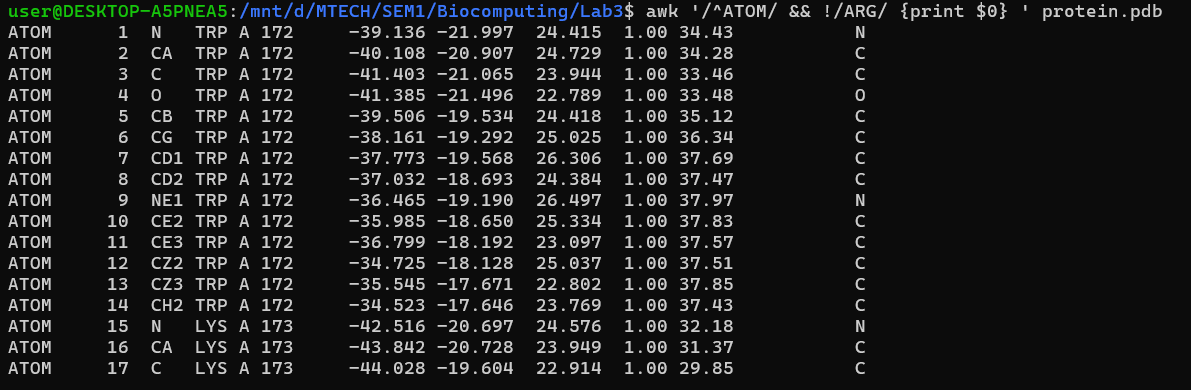
awk '$1=="ATOM" {res=substr($0,18,3); if(res ~ /E$/) print res}' protein.pdb | sort | uniq (CODE from ChatGPT)



19)



20)

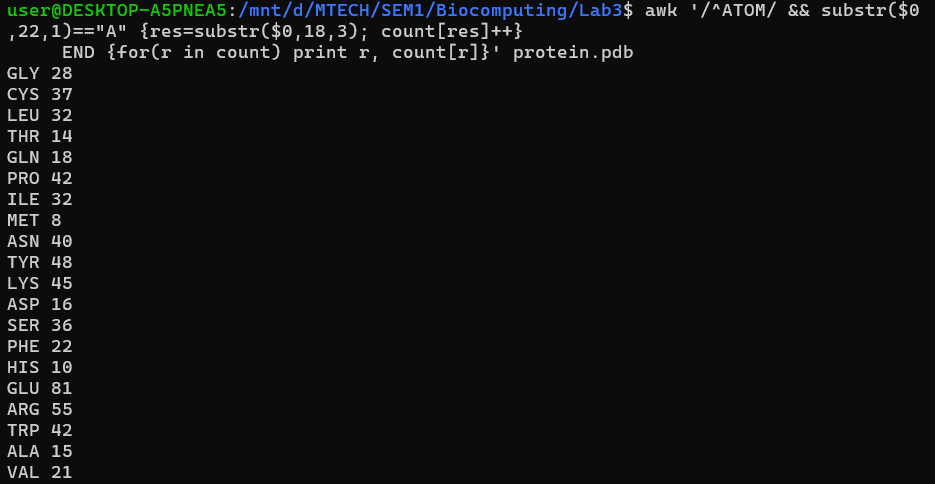


21)

CODE FROM CHATGPT:

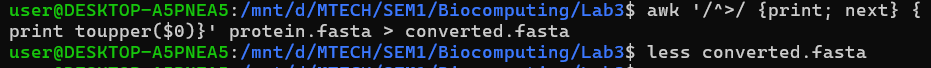
**awk '$1=="ATOM" && substr($0,22,1)=="A" {res=substr($0,18,3); count[res]++}**

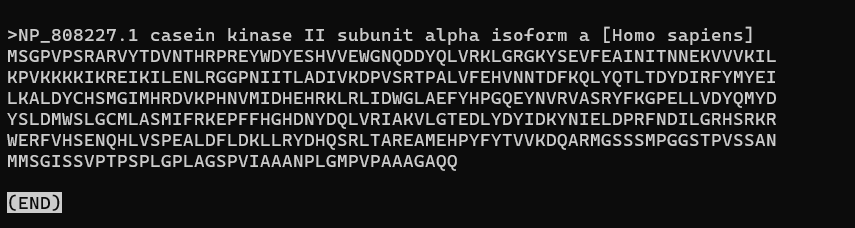
**END {for(r in count) print r, count[r]}'**



22)Toupper function to convert lower to uppercase was identified from ChatGPT

Usage of /^>/ {print; next} → if line is a header (starts with >), print as-is.(was got from ChatGPT)





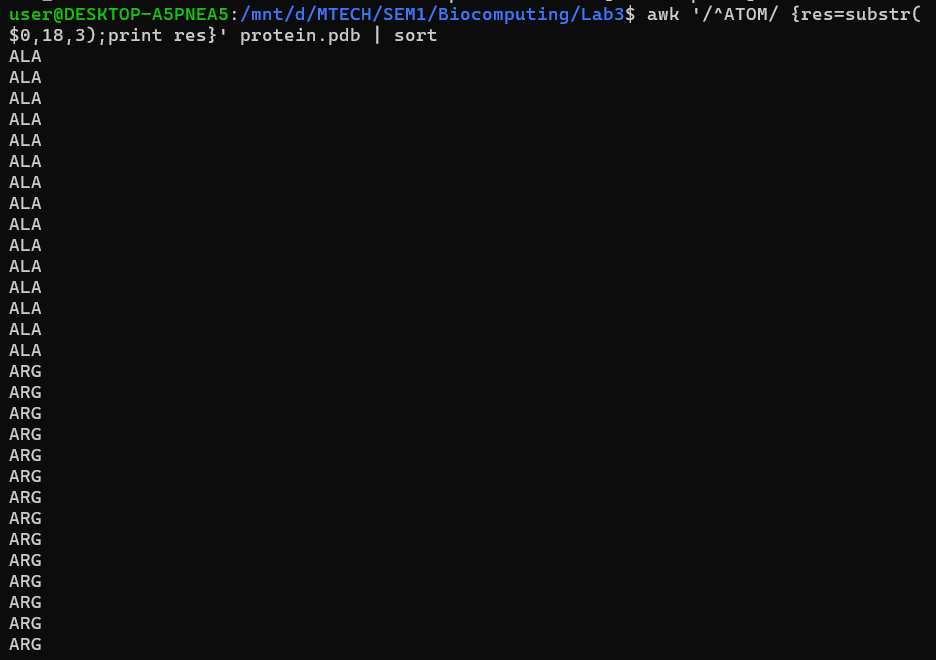
23)



There is only one sequence in this file, hence it is the one with maximum length(I have solved this logically without code)

24)

awk '$1=="ATOM" {res=substr($0,18,3); print res}' protein.pdb | sort | uniq (code from GPT and I have referred the sort usage from it)



25)



INFERENCE: Only A chain is present in the protein.pdb file

26)

The code was complicated, so I got the complete code from CHATGPT, but I’ve understood the logic of the code

