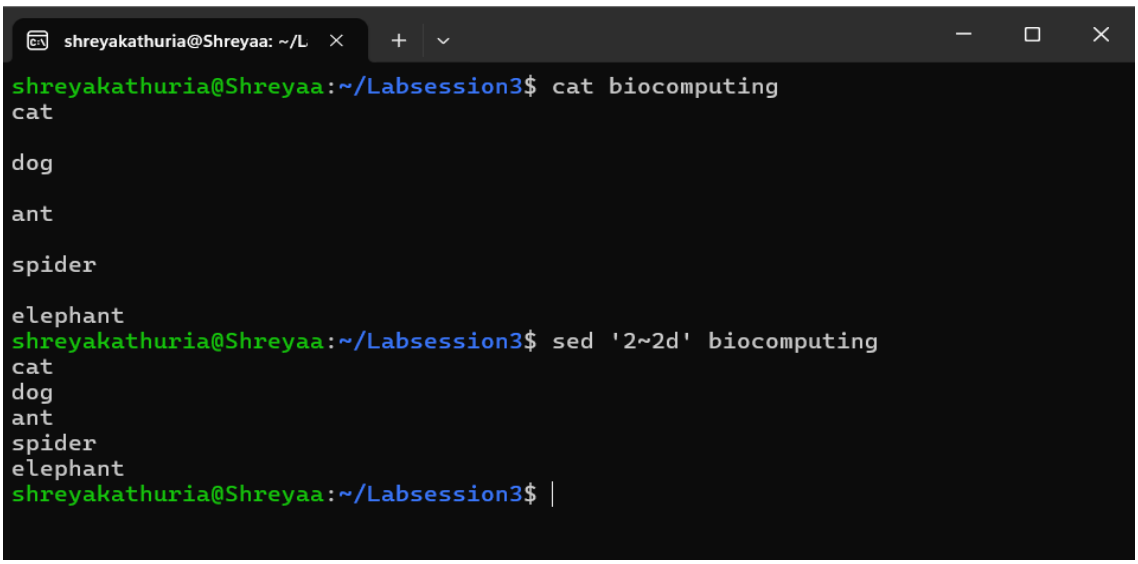
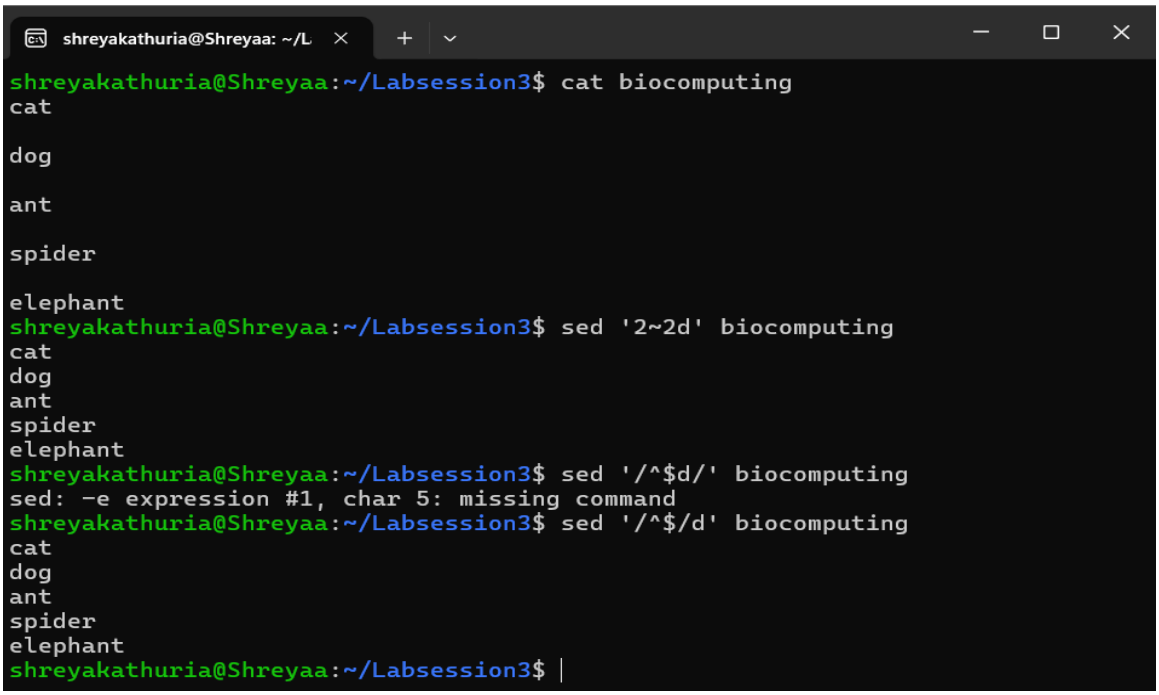
BIOCOMPUTING

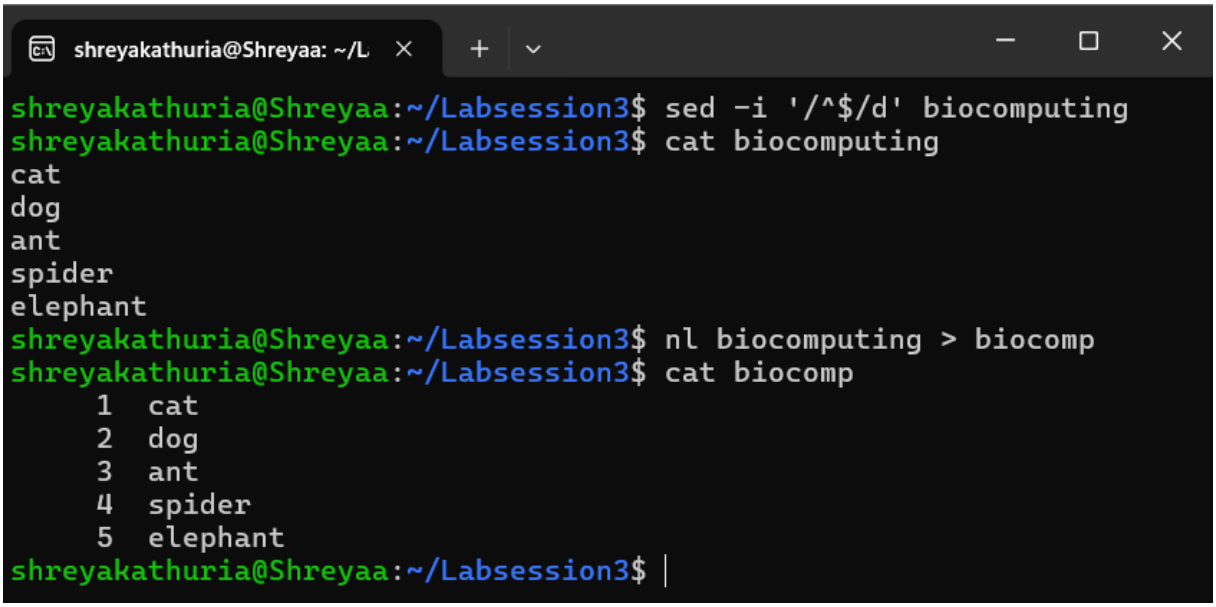
ASSIGNMENT 3

1. Create a file with some text written every alternate line using vi. Now delete all empty lines from file using sed (Hint use wildcards for beginning and end of lines)

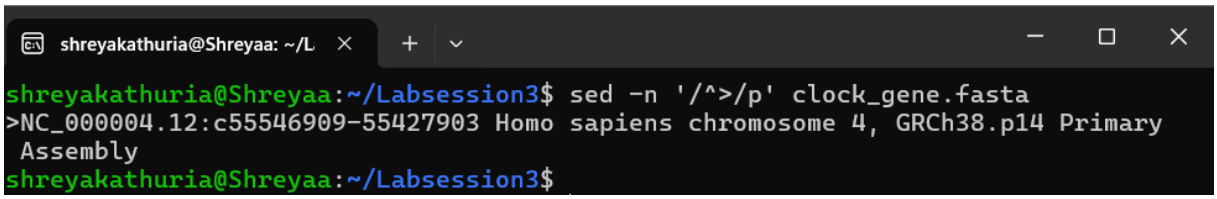




2. Using the same file created above, add line numbers in front of each line and save in another file.

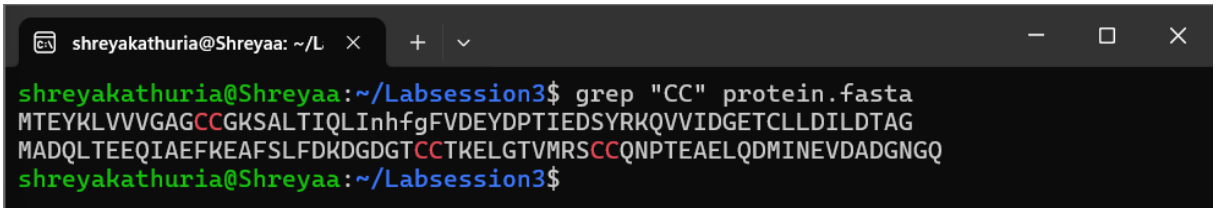


3. Print only the header lines from clock_gene.fasta using sed.



4. Print all headers from protein.fasta that contain the word CLOCK.

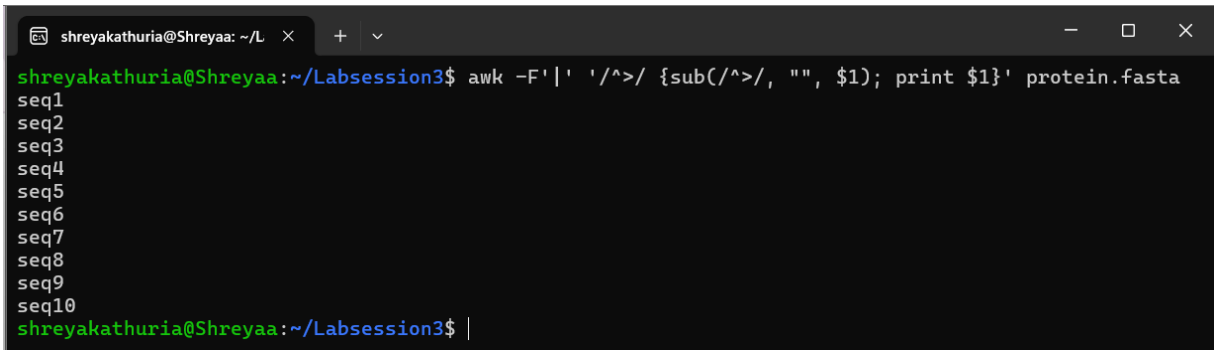
5. Extract sequences from protein.fasta that contain at least two consecutive C's



6. Count the total number of G's in clock_gene.fasta.

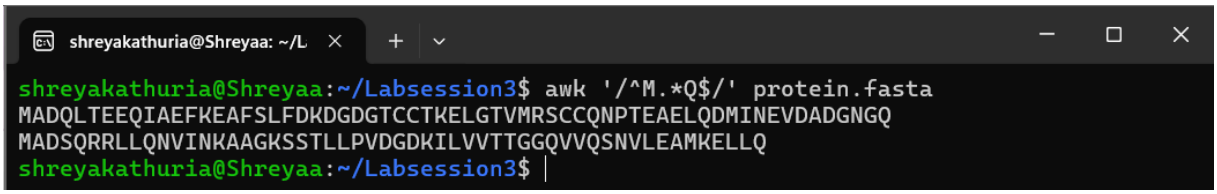
7. Print only lines 5 to 28 from clock_gene.fasta.

8. Print only the sequence ID (without >) from each header in protein.fasta.



Used gemini for this (I couldn't figure out how to remove headers (sub command)

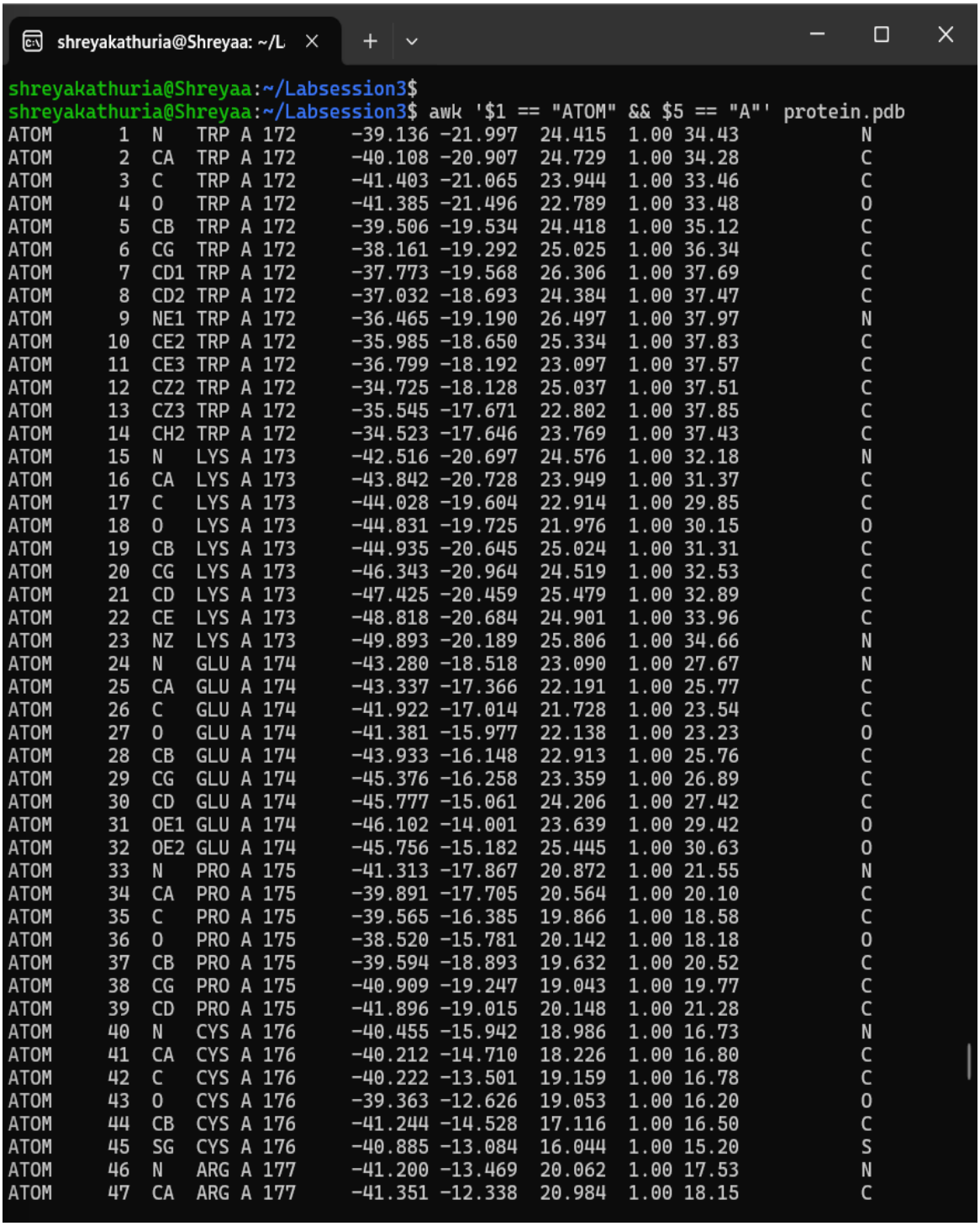
9a. From protein.fasta, extract sequence lines that start with M and end with Q.

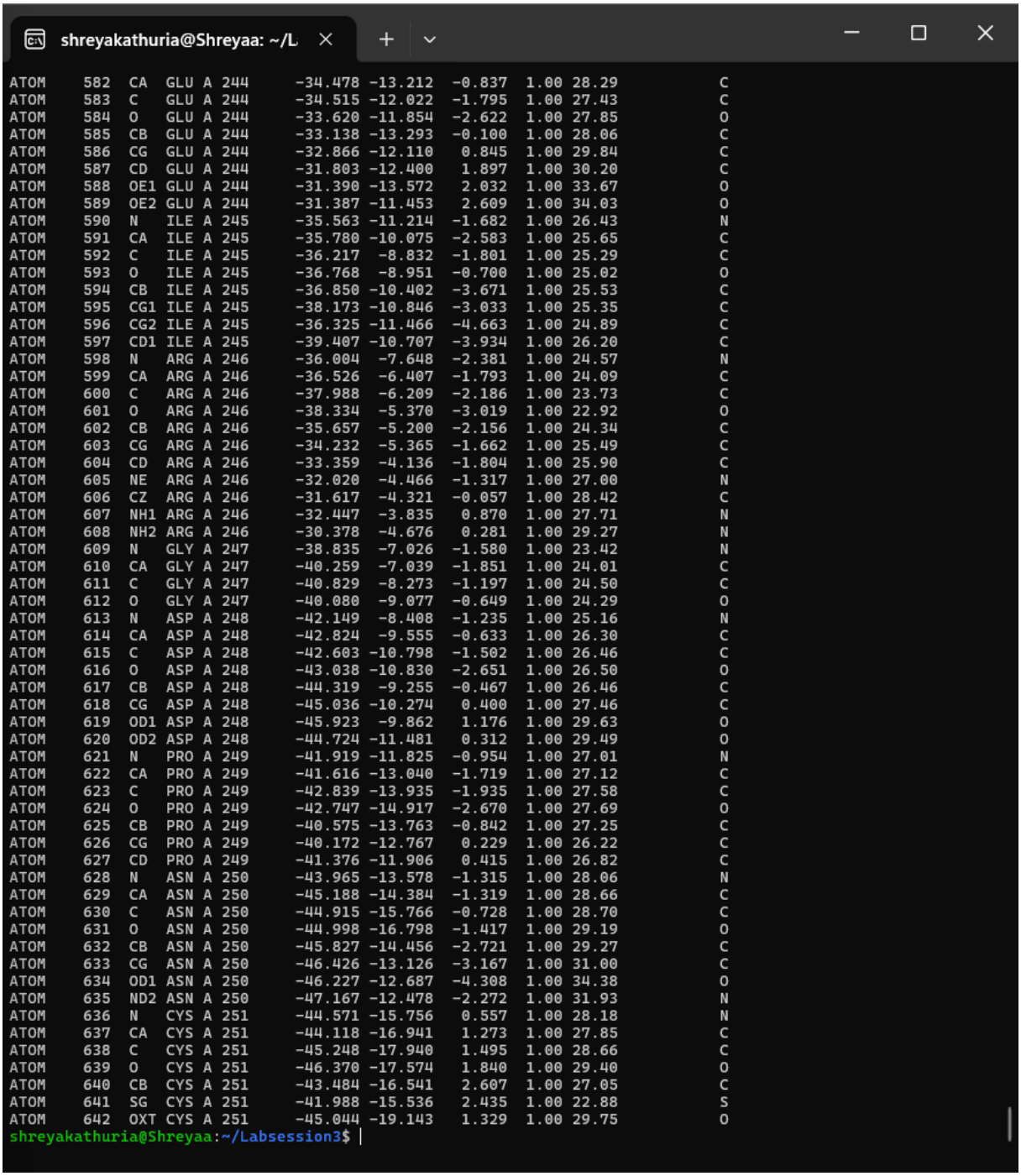


9. Find the length of each sequence in protein.fasta and print it alongside the sequenceID.

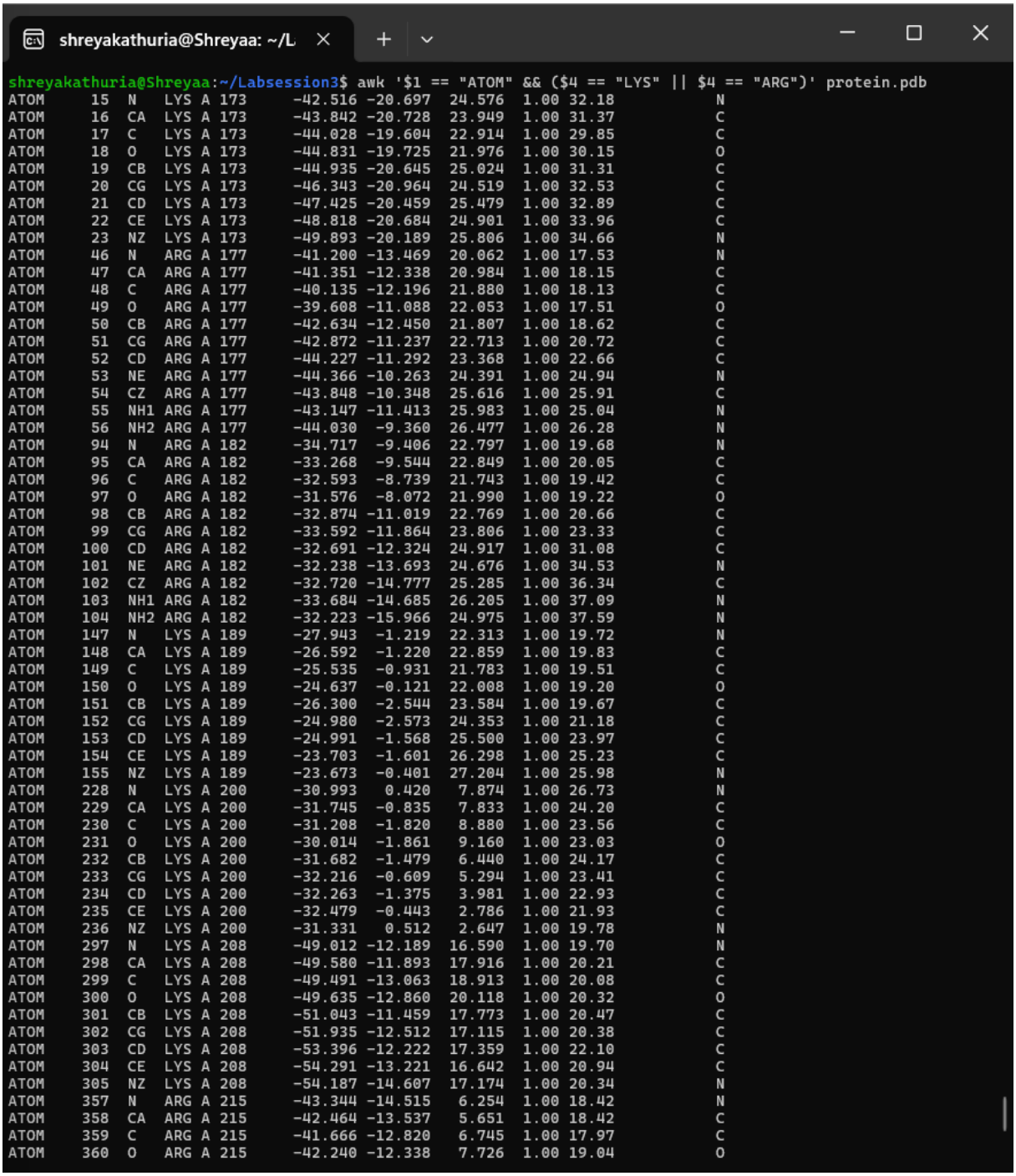
Used gemini for removing headers using sed.

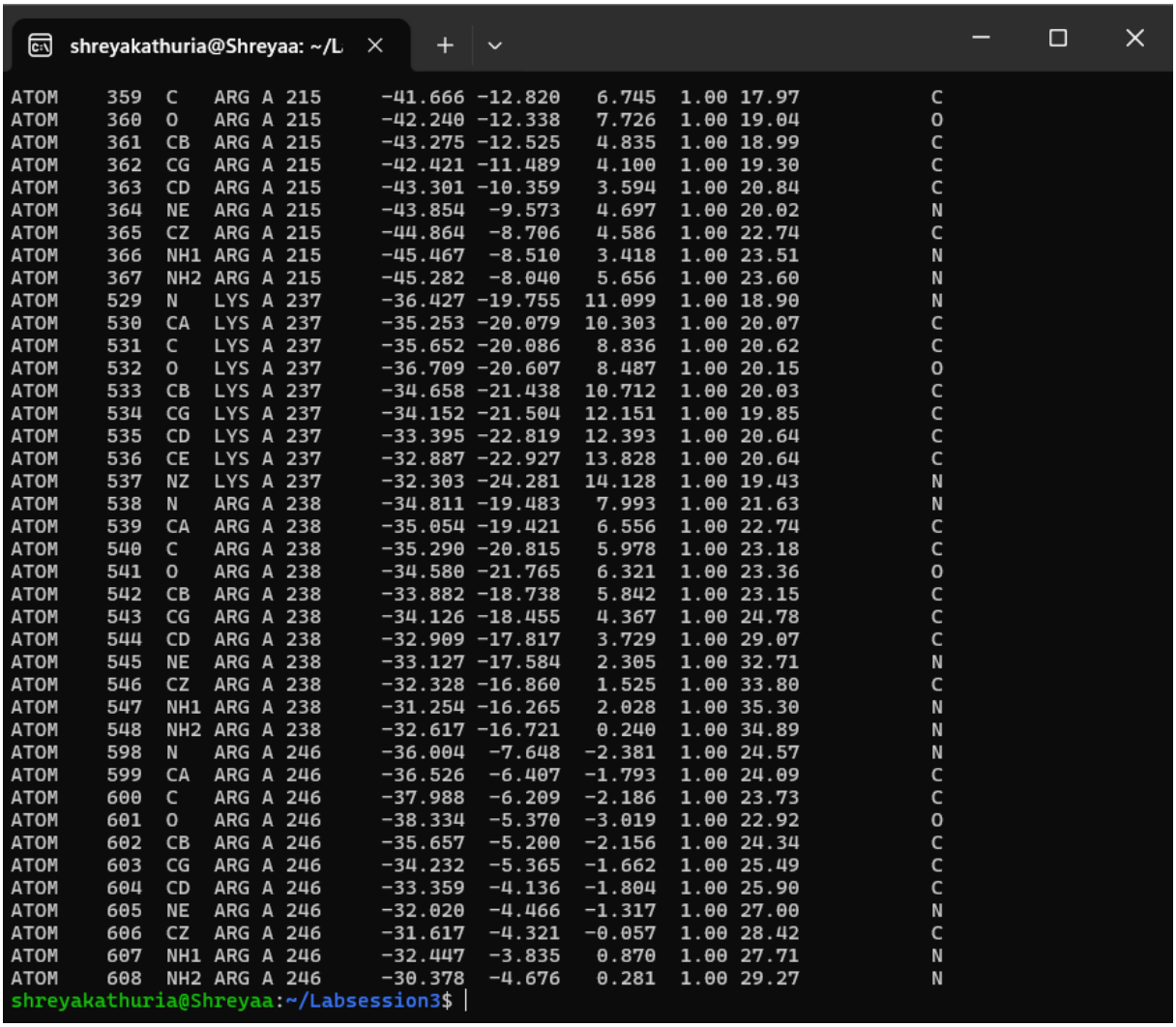
10. Print all ATOM lines from protein.pdb that belong to chain A only.




shreyakathuria@Shreyaa: ~/L
ATOM 582 CA GLU A 244 -34.478 -13.212 -0.837 1.00 28.29 C
ATOM 583 C GLU A 244 -34.515 -12.022 -1.795 1.00 27.43 C
ATOM 584 O GLU A 244 -33.620 -11.854 -2.622 1.00 27.85 O
ATOM 585 CB GLU A 244 -33.138 -13.293 -0.100 1.00 28.06 C
ATOM 586 CG GLU A 244 -32.866 -12.110 0.845 1.00 29.84 C
ATOM 587 CD GLU A 244 -31.803 -12.400 1.897 1.00 30.20 C
ATOM 588 OE1 GLU A 244 -31.390 -13.572 2.032 1.00 33.67 O
ATOM 589 OE2 GLU A 244 -31.387 -11.453 2.609 1.00 34.03 O
ATOM 590 N ILE A 245 -35.563 -11.214 -1.682 1.00 26.43 N
ATOM 591 CA ILE A 245 -35.780 -10.075 -2.583 1.00 25.65 C
ATOM 592 C ILE A 245 -36.217 -8.832 -1.801 1.00 25.29 C
ATOM 593 O ILE A 245 -36.768 -8.951 -0.700 1.00 25.02 O
ATOM 594 CB ILE A 245 -36.850 -10.402 -3.671 1.00 25.53 C
ATOM 595 CG1 ILE A 245 -38.173 -10.846 -3.033 1.00 25.35 C
ATOM 596 CG2 ILE A 245 -36.325 -11.466 -4.663 1.00 24.89 C
ATOM 597 CD1 ILE A 245 -39.407 -10.707 -3.934 1.00 26.20 C
ATOM 598 N ARG A 246 -36.004 -7.648 -2.381 1.00 24.57 N
ATOM 599 CA ARG A 246 -36.526 -6.407 -1.793 1.00 24.09 C
ATOM 600 C ARG A 246 -37.988 -6.209 -2.186 1.00 23.73 C
ATOM 601 O ARG A 246 -38.334 -5.370 -3.019 1.00 22.92 O
ATOM 602 CB ARG A 246 -35.657 -5.200 -2.156 1.00 24.34 C
ATOM 603 CG ARG A 246 -34.232 -5.365 -1.662 1.00 25.49 C
ATOM 604 CD ARG A 246 -33.359 -4.136 -1.804 1.00 25.90 C
ATOM 605 NE ARG A 246 -32.020 -4.466 -1.317 1.00 27.00 N
ATOM 606 CZ ARG A 246 -31.617 -4.321 -0.057 1.00 28.42 C
ATOM 607 NH1 ARG A 246 -32.447 -3.835 0.870 1.00 27.71 N
ATOM 608 NH2 ARG A 246 -30.378 -4.676 0.281 1.00 29.27 N
ATOM 609 N GLY A 247 -38.835 -7.026 -1.580 1.00 23.42 N
ATOM 610 CA GLY A 247 -40.259 -7.039 -1.851 1.00 24.01 C
ATOM 611 C GLY A 247 -40.829 -8.273 -1.197 1.00 24.50 C
ATOM 612 O GLY A 247 -40.080 -9.077 -0.649 1.00 24.29 O
ATOM 613 N ASP A 248 -42.149 -8.408 -1.235 1.00 25.16 N
ATOM 614 CA ASP A 248 -42.824 -9.555 -0.633 1.00 26.30 C
ATOM 615 C ASP A 248 -42.603 -10.798 -1.502 1.00 26.46 C
ATOM 616 O ASP A 248 -43.038 -10.830 -2.651 1.00 26.50 O
ATOM 617 CB ASP A 248 -44.319 -9.255 -0.467 1.00 26.46 C
ATOM 618 CG ASP A 248 -45.036 -10.274 0.400 1.00 27.46 C
ATOM 619 OD1 ASP A 248 -45.923 -9.862 1.176 1.00 29.63 O
ATOM 620 OD2 ASP A 248 -44.724 -11.481 0.312 1.00 29.49 O
ATOM 621 N PRO A 249 -41.919 -11.825 -0.954 1.00 27.01 N
ATOM 622 CA PRO A 249 -41.616 -13.040 -1.719 1.00 27.12 C
ATOM 623 C PRO A 249 -42.839 -13.935 -1.935 1.00 27.58 C
ATOM 624 O PRO A 249 -42.747 -14.917 -2.670 1.00 27.69 O
ATOM 625 CB PRO A 249 -40.575 -13.763 -0.842 1.00 27.25 C
ATOM 626 CG PRO A 249 -40.172 -12.767 0.229 1.00 26.22 C
ATOM 627 CD PRO A 249 -41.376 -11.906 0.415 1.00 26.82 C
ATOM 628 N ASN A 250 -43.965 -13.578 -1.315 1.00 28.06 N
ATOM 629 CA ASN A 250 -45.188 -14.384 -1.319 1.00 28.66 C
ATOM 630 C ASN A 250 -44.915 -15.766 -0.728 1.00 28.70 C
ATOM 631 O ASN A 250 -44.998 -16.798 -1.417 1.00 29.19 O
ATOM 632 CB ASN A 250 -45.827 -14.456 -2.721 1.00 29.27 C
ATOM 633 CG ASN A 250 -46.426 -13.126 -3.167 1.00 31.00 C
ATOM 634 OD1 ASN A 250 -46.227 -12.687 -4.308 1.00 34.38 O
ATOM 635 ND2 ASN A 250 -47.167 -12.478 -2.272 1.00 31.93 N
ATOM 636 N CYS A 251 -44.571 -15.756 0.557 1.00 28.18 N
ATOM 637 CA CYS A 251 -44.118 -16.941 1.273 1.00 27.85 C
ATOM 638 C CYS A 251 -45.248 -17.940 1.495 1.00 28.66 C
ATOM 639 O CYS A 251 -46.370 -17.574 1.840 1.00 29.40 O
ATOM 640 CB CYS A 251 -43.484 -16.541 2.607 1.00 27.05 C
ATOM 641 SG CYS A 251 -41.988 -15.536 2.435 1.00 22.88 S
ATOM 642 OXT CYS A 251 -45.044 -19.143 1.329 1.00 29.75 O
shreyakathuria@Shreyaa:~/Labsession3$

11. Extract all ATOM lines for residues LYS or ARG in protein.pdb.



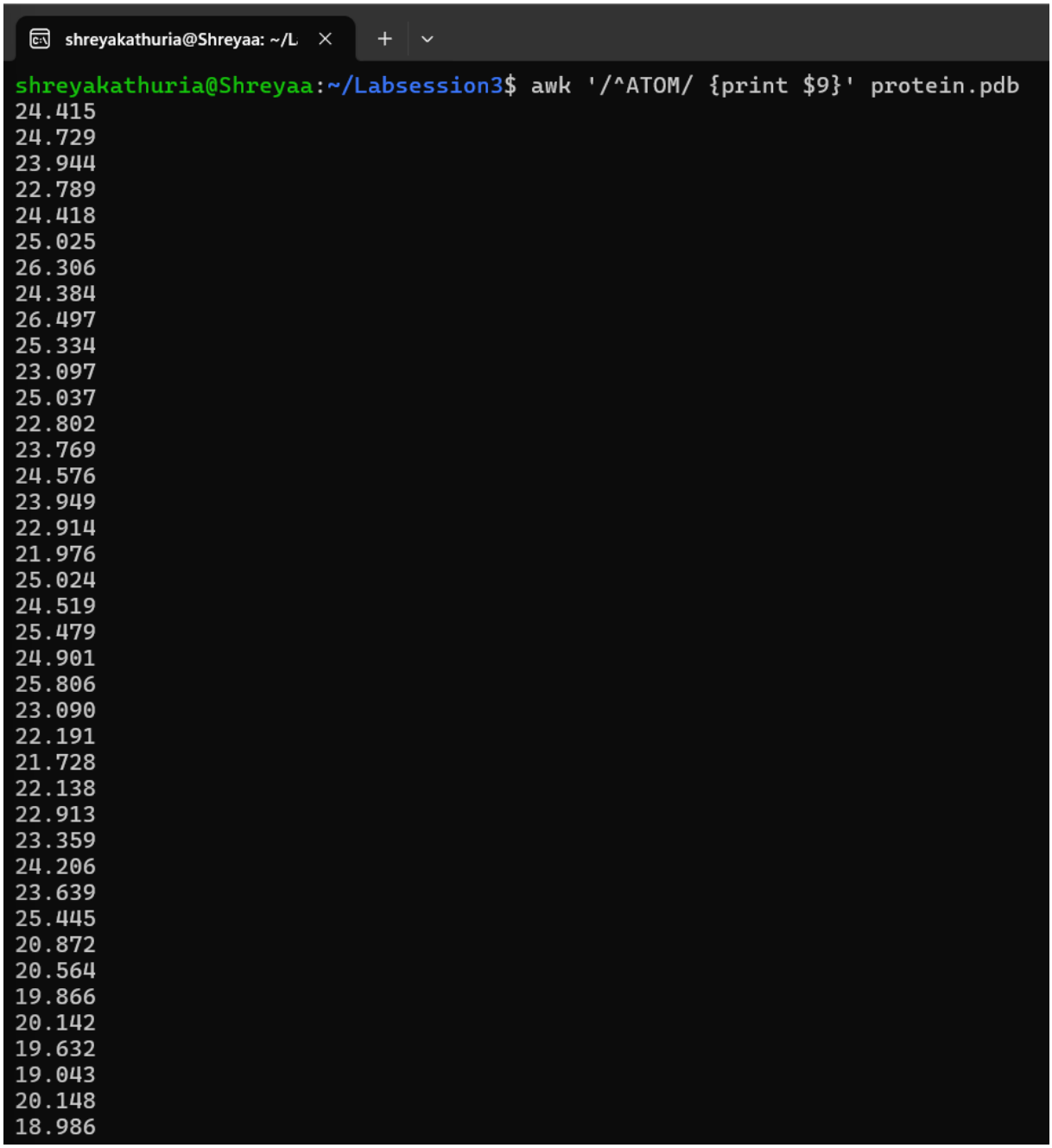

shreyakathuria@Shreyaa: ~/L
ATOM 359 C ARG A 215 -41.666 -12.820 6.745 1.00 17.97 C
ATOM 360 O ARG A 215 -42.240 -12.338 7.726 1.00 19.04 O
ATOM 361 CB ARG A 215 -43.275 -12.525 4.835 1.00 18.99 C
ATOM 362 CG ARG A 215 -42.421 -11.489 4.100 1.00 19.30 C
ATOM 363 CD ARG A 215 -43.301 -10.359 3.594 1.00 20.84 C
ATOM 364 NE ARG A 215 -43.854 -9.573 4.697 1.00 20.02 N
ATOM 365 CZ ARG A 215 -44.864 -8.706 4.586 1.00 22.74 C
ATOM 366 NH1 ARG A 215 -45.467 -8.510 3.418 1.00 23.51 N
ATOM 367 NH2 ARG A 215 -45.282 -8.040 5.656 1.00 23.60 N
ATOM 529 N LYS A 237 -36.427 -19.755 11.099 1.00 18.90 N
ATOM 530 CA LYS A 237 -35.253 -20.079 10.303 1.00 20.07 C
ATOM 531 C LYS A 237 -35.652 -20.086 8.836 1.00 20.62 C
ATOM 532 O LYS A 237 -36.709 -20.607 8.487 1.00 20.15 O
ATOM 533 CB LYS A 237 -34.658 -21.438 10.712 1.00 20.03 C
ATOM 534 CG LYS A 237 -34.152 -21.504 12.151 1.00 19.85 C
ATOM 535 CD LYS A 237 -33.395 -22.819 12.393 1.00 20.64 C
ATOM 536 CE LYS A 237 -32.887 -22.927 13.828 1.00 20.64 C
ATOM 537 NZ LYS A 237 -32.303 -24.281 14.128 1.00 19.43 N
ATOM 538 N ARG A 238 -34.811 -19.483 7.993 1.00 21.63 N
ATOM 539 CA ARG A 238 -35.054 -19.421 6.556 1.00 22.74 C
ATOM 540 C ARG A 238 -35.290 -20.815 5.978 1.00 23.18 C
ATOM 541 O ARG A 238 -34.580 -21.765 6.321 1.00 23.36 O
ATOM 542 CB ARG A 238 -33.882 -18.738 5.842 1.00 23.15 C
ATOM 543 CG ARG A 238 -34.126 -18.455 4.367 1.00 24.78 C
ATOM 544 CD ARG A 238 -32.909 -17.817 3.729 1.00 29.07 C
ATOM 545 NE ARG A 238 -33.127 -17.584 2.305 1.00 32.71 N
ATOM 546 CZ ARG A 238 -32.328 -16.860 1.525 1.00 33.80 C
ATOM 547 NH1 ARG A 238 -31.254 -16.265 2.028 1.00 35.30 N
ATOM 548 NH2 ARG A 238 -32.617 -16.721 0.240 1.00 34.89 N
ATOM 598 N ARG A 246 -36.004 -7.648 -2.381 1.00 24.57 N
ATOM 599 CA ARG A 246 -36.526 -6.407 -1.793 1.00 24.09 C
ATOM 600 C ARG A 246 -37.988 -6.209 -2.186 1.00 23.73 C
ATOM 601 O ARG A 246 -38.334 -5.370 -3.019 1.00 22.92 O
ATOM 602 CB ARG A 246 -35.657 -5.200 -2.156 1.00 24.34 C
ATOM 603 CG ARG A 246 -34.232 -5.365 -1.662 1.00 25.49 C
ATOM 604 CD ARG A 246 -33.359 -4.136 -1.804 1.00 25.90 C
ATOM 605 NE ARG A 246 -32.020 -4.466 -1.317 1.00 27.00 N
ATOM 606 CZ ARG A 246 -31.617 -4.321 -0.057 1.00 28.42 C
ATOM 607 NH1 ARG A 246 -32.447 -3.835 0.870 1.00 27.71 N
ATOM 608 NH2 ARG A 246 -30.378 -4.676 0.281 1.00 29.27 N
shreyakathuria@Shreyaa:~/Labsession3$

12. Replace every occurrence of LYS with ARG in protein.pdb.

shreyakathuria@Shreyaa: ~/L

```
ATOM    576  C   PRO A 243     -35.683 -15.295  -1.290  1.00 28.20           C
ATOM    577  O   PRO A 243     -36.259 -15.313  -0.194  1.00 27.71           O
ATOM    578  CB  PRO A 243     -37.167 -15.904  -3.244  1.00 28.46           C
ATOM    579  CG  PRO A 243     -38.367 -16.600  -2.761  1.00 28.86           C
ATOM    580  CD  PRO A 243     -37.933 -17.820  -2.003  1.00 28.47           C
ATOM    581  N   GLU A 244     -34.736 -14.419  -1.613  1.00 27.80           N
ATOM    582  CA  GLU A 244     -34.478 -13.212  -0.837  1.00 28.29           C
ATOM    583  C   GLU A 244     -34.515 -12.022  -1.795  1.00 27.43           C
ATOM    584  O   GLU A 244     -33.620 -11.854  -2.622  1.00 27.85           O
ATOM    585  CB  GLU A 244     -33.138 -13.293  -0.100  1.00 28.06           C
ATOM    586  CG  GLU A 244     -32.866 -12.110   0.845  1.00 29.84           C
ATOM    587  CD  GLU A 244     -31.803 -12.400   1.897  1.00 30.20           C
ATOM    588  OE1 GLU A 244     -31.390 -13.572   2.032  1.00 33.67           O
ATOM    589  OE2 GLU A 244     -31.387 -11.453   2.609  1.00 34.03           O
ATOM    590  N   ILE A 245     -35.563 -11.214  -1.682  1.00 26.43           N
ATOM    591  CA  ILE A 245     -35.780 -10.075  -2.583  1.00 25.65           C
ATOM    592  C   ILE A 245     -36.217  -8.832  -1.801  1.00 25.29           C
ATOM    593  O   ILE A 245     -36.768  -8.951  -0.700  1.00 25.02           O
ATOM    594  CB  ILE A 245     -36.850 -10.402  -3.671  1.00 25.53           C
ATOM    595  CG1 ILE A 245     -38.173 -10.846  -3.033  1.00 25.35           C
ATOM    596  CG2 ILE A 245     -36.325 -11.466  -4.663  1.00 24.89           C
ATOM    597  CD1 ILE A 245     -39.407 -10.707  -3.934  1.00 26.20           C
ATOM    598  N   ARG A 246     -36.004  -7.648  -2.381  1.00 24.57           N
ATOM    599  CA  ARG A 246     -36.526  -6.407  -1.793  1.00 24.09           C
ATOM    600  C   ARG A 246     -37.988  -6.209  -2.186  1.00 23.73           C
ATOM    601  O   ARG A 246     -38.334  -5.370  -3.019  1.00 22.92           O
ATOM    602  CB  ARG A 246     -35.657  -5.200  -2.156  1.00 24.34           C
ATOM    603  CG  ARG A 246     -34.232  -5.365  -1.662  1.00 25.49           C
ATOM    604  CD  ARG A 246     -33.359  -4.136  -1.804  1.00 25.90           C
ATOM    605  NE  ARG A 246     -32.020  -4.466  -1.317  1.00 27.00           N
ATOM    606  CZ  ARG A 246     -31.617  -4.321  -0.057  1.00 28.42           C
ATOM    607  NH1 ARG A 246     -32.447  -3.835   0.870  1.00 27.71           N
ATOM    608  NH2 ARG A 246     -30.378  -4.676   0.281  1.00 29.27           N
ATOM    609  N   GLY A 247     -38.835  -7.026  -1.580  1.00 23.42           N
ATOM    610  CA  GLY A 247     -40.259  -7.039  -1.851  1.00 24.01           C
ATOM    611  C   GLY A 247     -40.829  -8.273  -1.197  1.00 24.50           C
ATOM    612  O   GLY A 247     -40.080  -9.077  -0.649  1.00 24.29           O
ATOM    613  N   ASP A 248     -42.149  -8.408  -1.235  1.00 25.16           N
ATOM    614  CA  ASP A 248     -42.824  -9.555  -0.633  1.00 26.30           C
ATOM    615  C   ASP A 248     -42.603 -10.798  -1.502  1.00 26.46           C
ATOM    616  O   ASP A 248     -43.038 -10.830  -2.651  1.00 26.50           O
ATOM    617  CB  ASP A 248     -44.319  -9.255  -0.467  1.00 26.46           C
ATOM    618  CG  ASP A 248     -45.036 -10.274   0.400  1.00 27.46           C
ATOM    619  OD1 ASP A 248     -45.923  -9.862   1.176  1.00 29.63           O
ATOM    620  OD2 ASP A 248     -44.724 -11.481   0.312  1.00 29.49           O
ATOM    621  N   PRO A 249     -41.919 -11.825  -0.954  1.00 27.01           N
ATOM    622  CA  PRO A 249     -41.616 -13.040  -1.719  1.00 27.12           C
ATOM    623  C   PRO A 249     -42.839 -13.935  -1.935  1.00 27.58           C
ATOM    624  O   PRO A 249     -42.747 -14.917  -2.670  1.00 27.69           O
ATOM    625  CB  PRO A 249     -40.575 -13.763  -0.842  1.00 27.25           C
ATOM    626  CG  PRO A 249     -40.172 -12.767   0.229  1.00 26.22           C
ATOM    627  CD  PRO A 249     -41.376 -11.906   0.415  1.00 26.82           C
ATOM    628  N   ASN A 250     -43.965 -13.578  -1.315  1.00 28.06           N
ATOM    629  CA  ASN A 250     -45.188 -14.384  -1.319  1.00 28.66           C
ATOM    630  C   ASN A 250     -44.915 -15.766  -0.728  1.00 28.70           C
ATOM    631  O   ASN A 250     -44.998 -16.798  -1.417  1.00 29.19           O
ATOM    632  CB  ASN A 250     -45.827 -14.456  -2.721  1.00 29.27           C
ATOM    633  CG  ASN A 250     -46.426 -13.126  -3.167  1.00 31.00           C
ATOM    634  OD1 ASN A 250     -46.227 -12.687  -4.308  1.00 34.38           O
ATOM    635  ND2 ASN A 250     -47.167 -12.478  -2.272  1.00 31.93           N
ATOM    636  N   CYS A 251     -44.571 -15.756   0.557  1.00 28.18           N
ATOM    637  CA  CYS A 251     -44.118 -16.941   1.273  1.00 27.85           C
ATOM    638  C   CYS A 251     -45.248 -17.940   1.495  1.00 28.66           C
ATOM    639  O   CYS A 251     -46.370 -17.574   1.840  1.00 29.40           O
ATOM    640  CB  CYS A 251     -43.484 -16.541   2.607  1.00 27.05           C
ATOM    641  SG  CYS A 251     -41.988 -15.536   2.435  1.00 22.88           S
ATOM    642  OXT CYS A 251     -45.044 -19.143   1.329  1.00 29.75           O
```

13. Print only the z-coordinate (third number in coordinates) for each atom from protein.pdb.

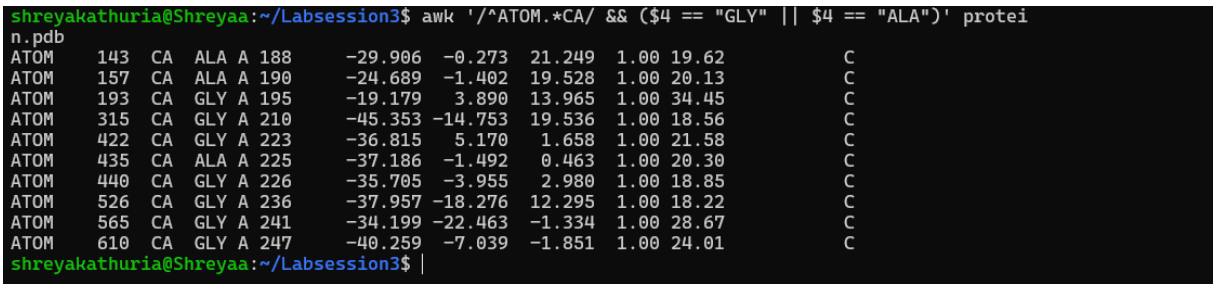


shreyakathuria@Shreyaa: ~/L

```
-3.033
-4.663
-3.934
-2.381
-1.793
-2.186
-3.019
-2.156
-1.662
-1.804
-1.317
-0.057
0.870
0.281
-1.580
-1.851
-1.197
-0.649
-1.235
-0.633
-1.502
-2.651
-0.467
0.400
1.176
0.312
-0.954
-1.719
-1.935
-2.670
-0.842
0.229
0.415
-1.315
-1.319
-0.728
-1.417
-2.721
-3.167
-4.308
-2.272
0.557
1.273
1.495
1.840
2.607
2.435
1.329
shreyakathuria@Shreyaa:~/Labsession3$
```

14. Count how many lines in protein.pdb contain a GLY residue.

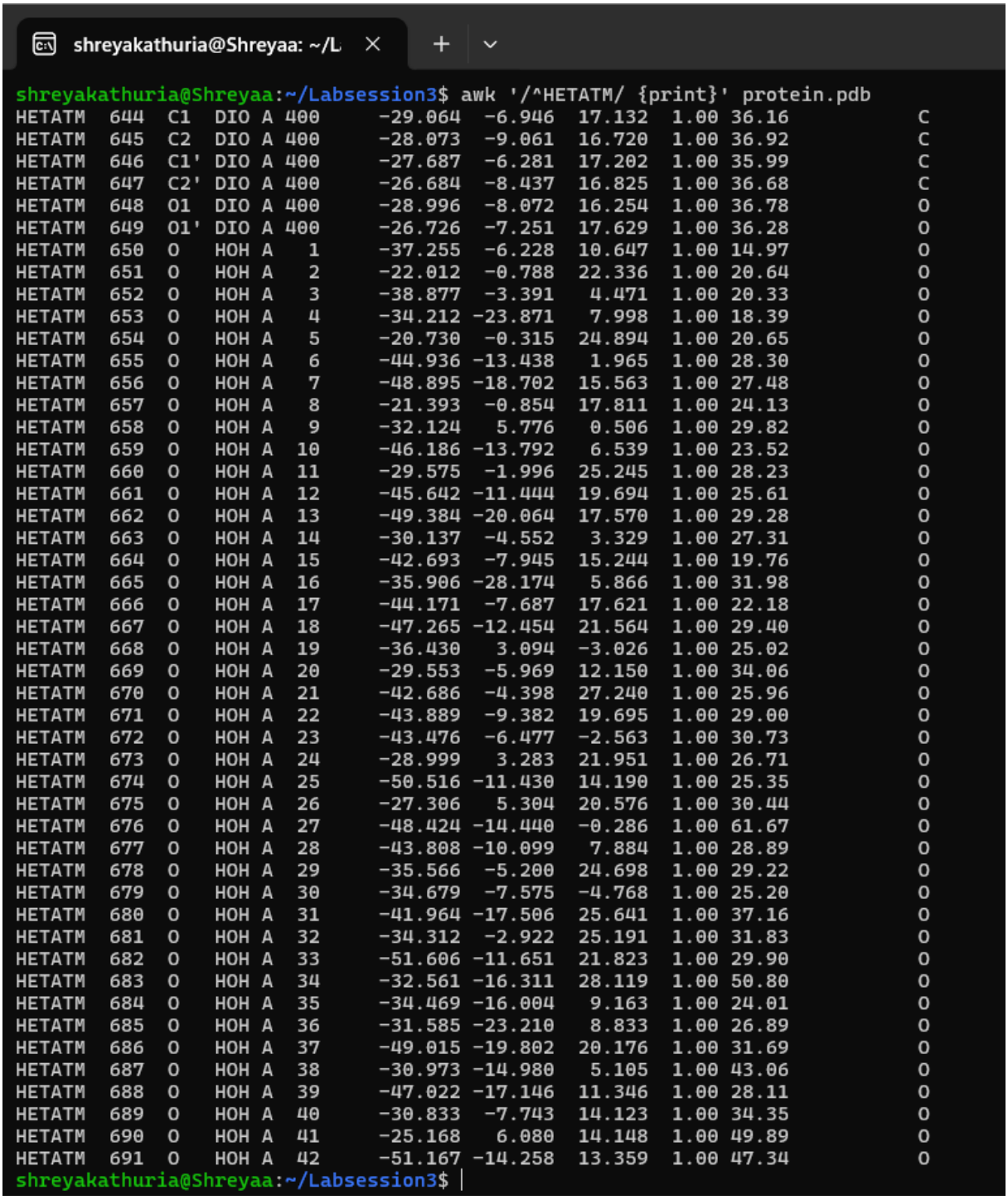
15. Print only the C-alpha (CA) atoms for residues ALA or GLY.



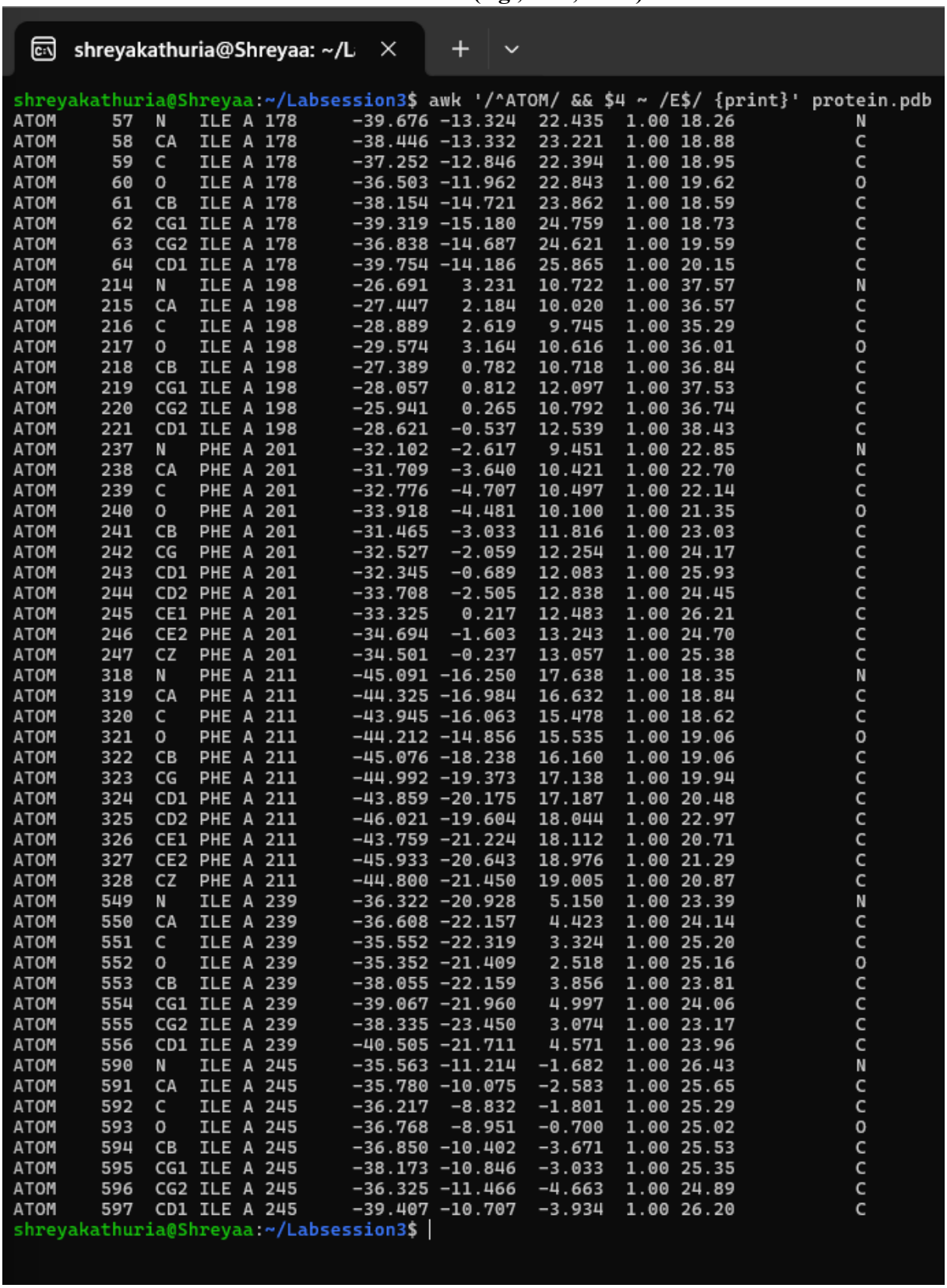
16. Count how many atoms are carbon (element C) in protein.pdb.

]

17. Print only the HETATM lines from protein.pdb.

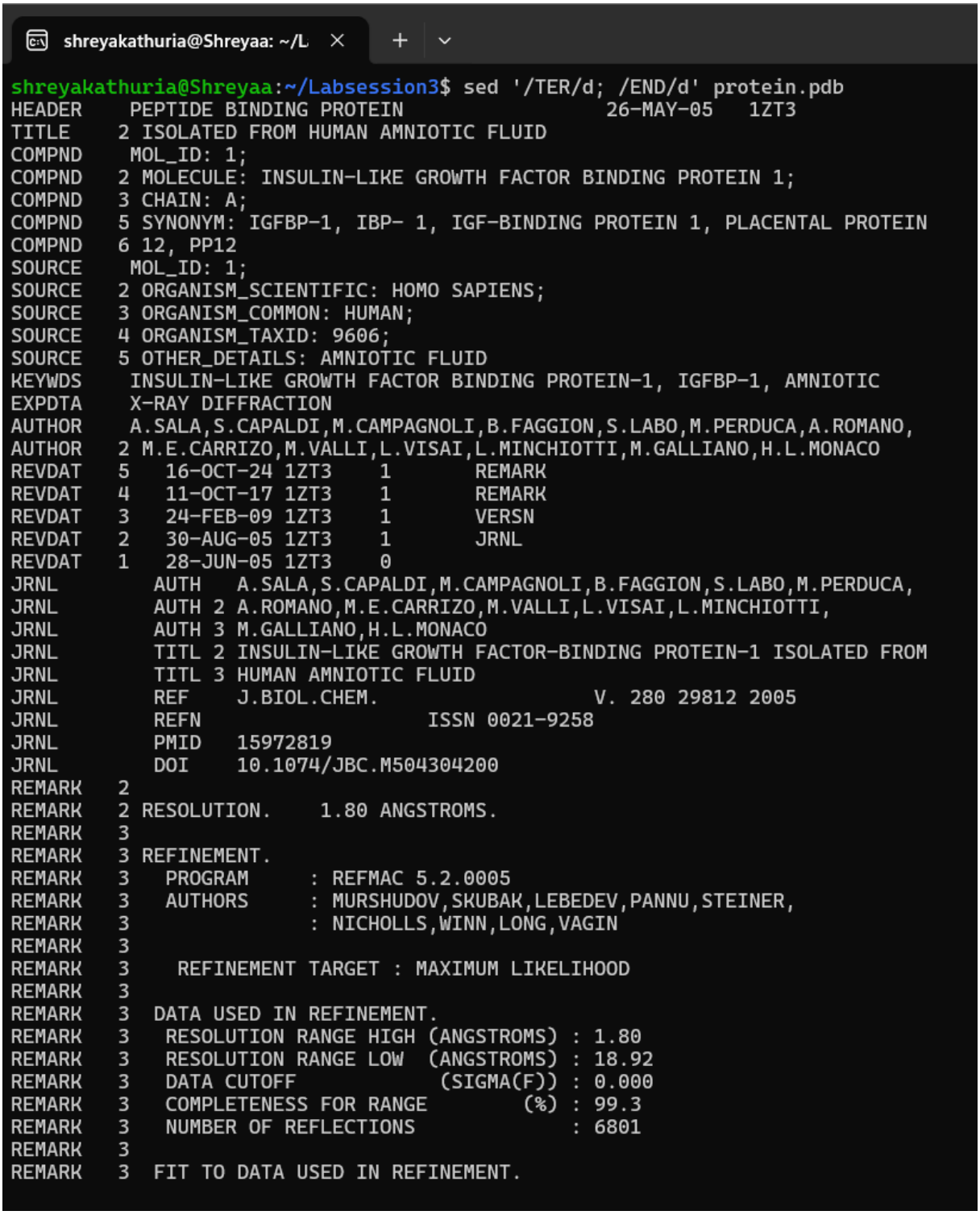


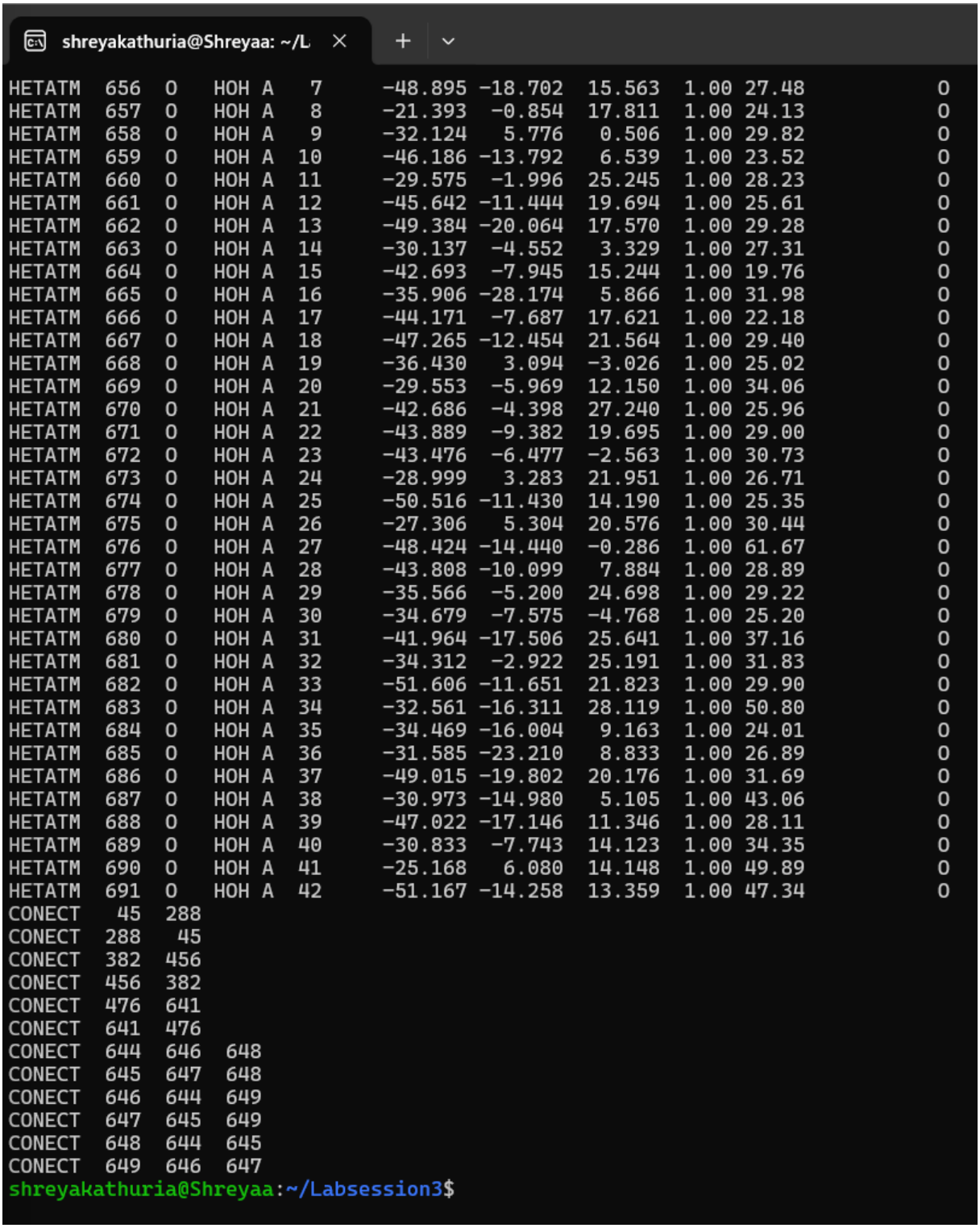
18. Extract all residue names that end with “E” (e.g., ILE, PHE).



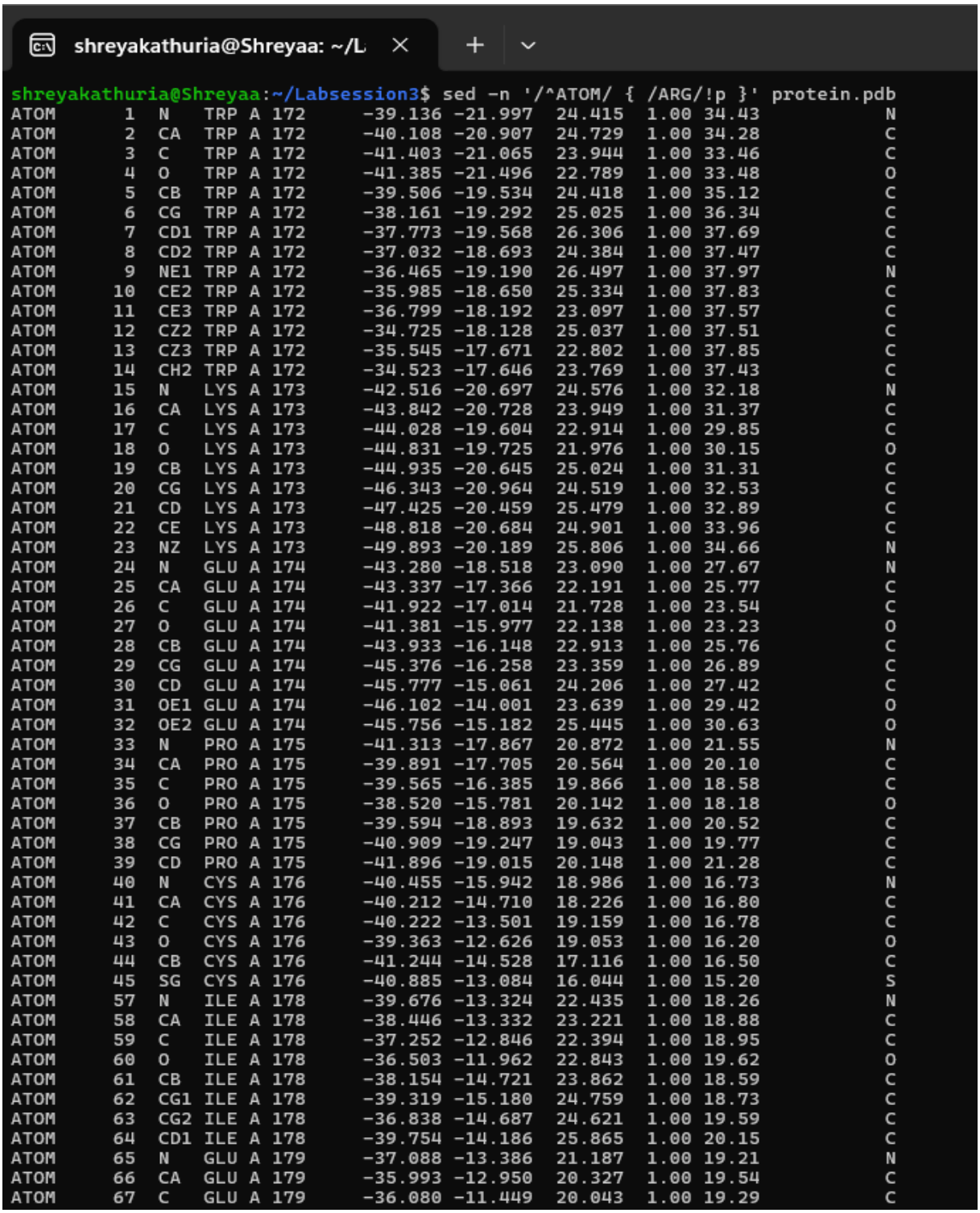
```
shreyakathuria@Shreyaa: ~/L
shreyakathuria@Shreyaa:~/Labsession3$ awk '/^ATOM/ && $4 ~ /E$/ {print}' protein.pdb
ATOM     57  N   ILE A 178     -39.676 -13.324  22.435  1.00 18.26           N
ATOM     58  CA  ILE A 178     -38.446 -13.332  23.221  1.00 18.88           C
ATOM     59  C   ILE A 178     -37.252 -12.846  22.394  1.00 18.95           C
ATOM     60  O   ILE A 178     -36.503 -11.962  22.843  1.00 19.62           O
ATOM     61  CB  ILE A 178     -38.154 -14.721  23.862  1.00 18.59           C
ATOM     62  CG1 ILE A 178     -39.319 -15.180  24.759  1.00 18.73           C
ATOM     63  CG2 ILE A 178     -36.838 -14.687  24.621  1.00 19.59           C
ATOM     64  CD1 ILE A 178     -39.754 -14.186  25.865  1.00 20.15           C
ATOM    214  N   ILE A 198     -26.691   3.231  10.722  1.00 37.57           N
ATOM    215  CA  ILE A 198     -27.447   2.184  10.020  1.00 36.57           C
ATOM    216  C   ILE A 198     -28.889   2.619   9.745  1.00 35.29           C
ATOM    217  O   ILE A 198     -29.574   3.164  10.616  1.00 36.01           O
ATOM    218  CB  ILE A 198     -27.389   0.782  10.718  1.00 36.84           C
ATOM    219  CG1 ILE A 198     -28.057   0.812  12.097  1.00 37.53           C
ATOM    220  CG2 ILE A 198     -25.941   0.265  10.792  1.00 36.74           C
ATOM    221  CD1 ILE A 198     -28.621  -0.537  12.539  1.00 38.43           C
ATOM    237  N   PHE A 201     -32.102  -2.617   9.451  1.00 22.85           N
ATOM    238  CA  PHE A 201     -31.709  -3.640  10.421  1.00 22.70           C
ATOM    239  C   PHE A 201     -32.776  -4.707  10.497  1.00 22.14           C
ATOM    240  O   PHE A 201     -33.918  -4.481  10.100  1.00 21.35           O
ATOM    241  CB  PHE A 201     -31.465  -3.033  11.816  1.00 23.03           C
ATOM    242  CG  PHE A 201     -32.527  -2.059  12.254  1.00 24.17           C
ATOM    243  CD1 PHE A 201     -32.345  -0.689  12.083  1.00 25.93           C
ATOM    244  CD2 PHE A 201     -33.708  -2.505  12.838  1.00 24.45           C
ATOM    245  CE1 PHE A 201     -33.325   0.217  12.483  1.00 26.21           C
ATOM    246  CE2 PHE A 201     -34.694  -1.603  13.243  1.00 24.70           C
ATOM    247  CZ  PHE A 201     -34.501  -0.237  13.057  1.00 25.38           C
ATOM    318  N   PHE A 211     -45.091 -16.250  17.638  1.00 18.35           N
ATOM    319  CA  PHE A 211     -44.325 -16.984  16.632  1.00 18.84           C
ATOM    320  C   PHE A 211     -43.945 -16.063  15.478  1.00 18.62           C
ATOM    321  O   PHE A 211     -44.212 -14.856  15.535  1.00 19.06           O
ATOM    322  CB  PHE A 211     -45.076 -18.238  16.160  1.00 19.06           C
ATOM    323  CG  PHE A 211     -44.992 -19.373  17.138  1.00 19.94           C
ATOM    324  CD1 PHE A 211     -43.859 -20.175  17.187  1.00 20.48           C
ATOM    325  CD2 PHE A 211     -46.021 -19.604  18.044  1.00 22.97           C
ATOM    326  CE1 PHE A 211     -43.759 -21.224  18.112  1.00 20.71           C
ATOM    327  CE2 PHE A 211     -45.933 -20.643  18.976  1.00 21.29           C
ATOM    328  CZ  PHE A 211     -44.800 -21.450  19.005  1.00 20.87           C
ATOM    549  N   ILE A 239     -36.322 -20.928   5.150  1.00 23.39           N
ATOM    550  CA  ILE A 239     -36.608 -22.157   4.423  1.00 24.14           C
ATOM    551  C   ILE A 239     -35.552 -22.319   3.324  1.00 25.20           C
ATOM    552  O   ILE A 239     -35.352 -21.409   2.518  1.00 25.16           O
ATOM    553  CB  ILE A 239     -38.055 -22.159   3.856  1.00 23.81           C
ATOM    554  CG1 ILE A 239     -39.067 -21.960   4.997  1.00 24.06           C
ATOM    555  CG2 ILE A 239     -38.335 -23.450   3.074  1.00 23.17           C
ATOM    556  CD1 ILE A 239     -40.505 -21.711   4.571  1.00 23.96           C
ATOM    590  N   ILE A 245     -35.563 -11.214  -1.682  1.00 26.43           N
ATOM    591  CA  ILE A 245     -35.780 -10.075  -2.583  1.00 25.65           C
ATOM    592  C   ILE A 245     -36.217  -8.832  -1.801  1.00 25.29           C
ATOM    593  O   ILE A 245     -36.768  -8.951  -0.700  1.00 25.02           O
ATOM    594  CB  ILE A 245     -36.850 -10.402  -3.671  1.00 25.53           C
ATOM    595  CG1 ILE A 245     -38.173 -10.846  -3.033  1.00 25.35           C
ATOM    596  CG2 ILE A 245     -36.325 -11.466  -4.663  1.00 24.89           C
ATOM    597  CD1 ILE A 245     -39.407 -10.707  -3.934  1.00 26.20           C
shreyakathuria@Shreyaa:~/Labsession3$
```

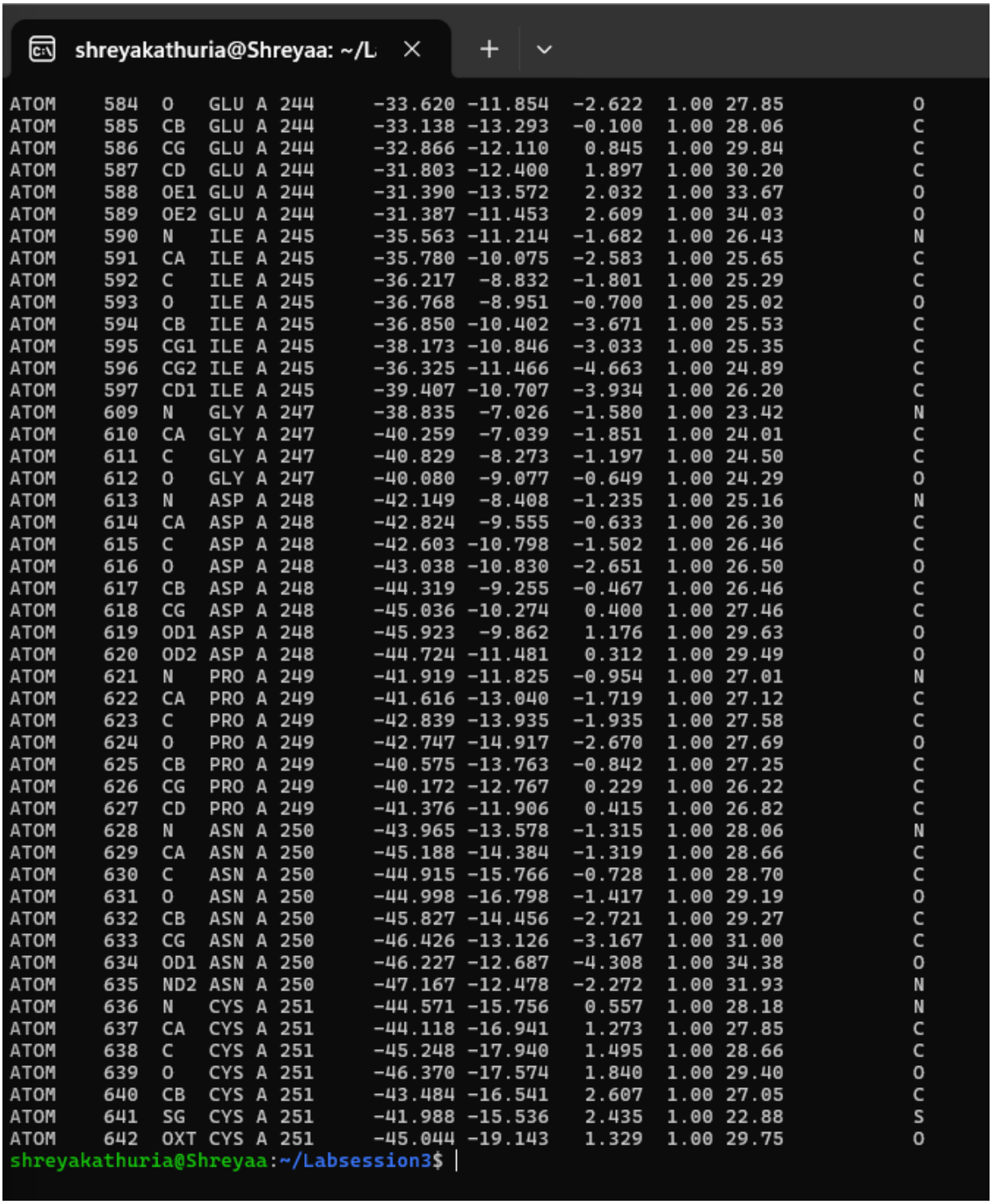
19. Delete all the lines that contain TER or END from protein.pdb.



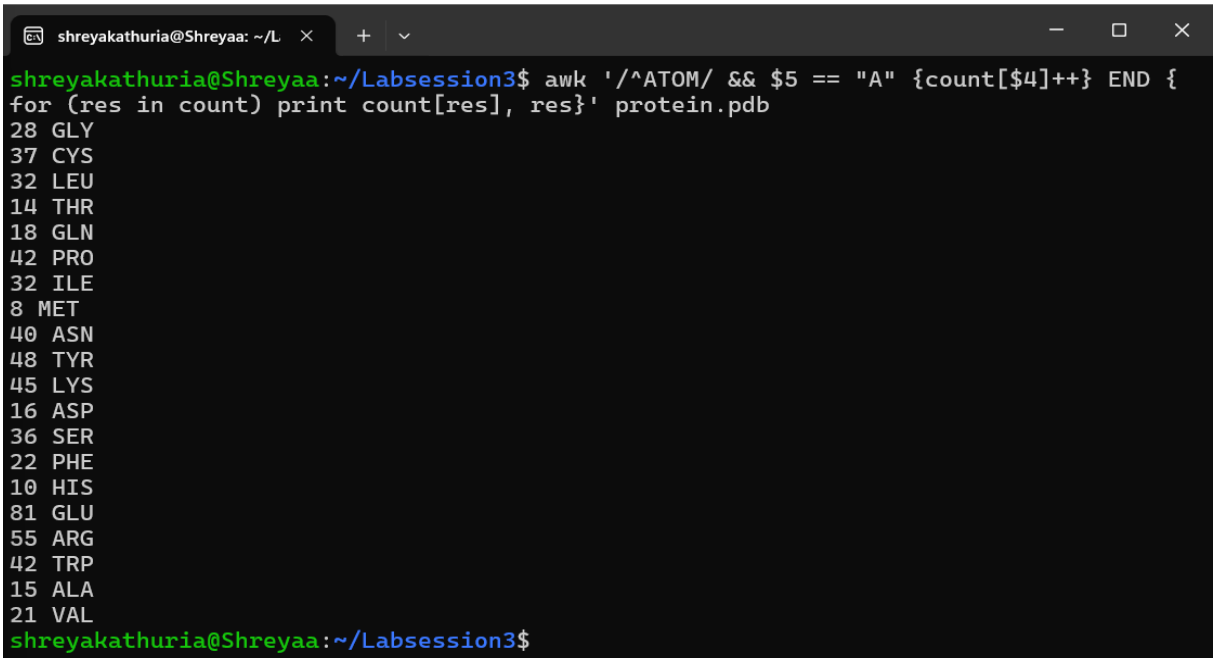

shreyakathuria@Shreyaa: ~/L
HETATM 656 O HOH A 7 -48.895 -18.702 15.563 1.00 27.48 O
HETATM 657 O HOH A 8 -21.393 -0.854 17.811 1.00 24.13 O
HETATM 658 O HOH A 9 -32.124 5.776 0.506 1.00 29.82 O
HETATM 659 O HOH A 10 -46.186 -13.792 6.539 1.00 23.52 O
HETATM 660 O HOH A 11 -29.575 -1.996 25.245 1.00 28.23 O
HETATM 661 O HOH A 12 -45.642 -11.444 19.694 1.00 25.61 O
HETATM 662 O HOH A 13 -49.384 -20.064 17.570 1.00 29.28 O
HETATM 663 O HOH A 14 -30.137 -4.552 3.329 1.00 27.31 O
HETATM 664 O HOH A 15 -42.693 -7.945 15.244 1.00 19.76 O
HETATM 665 O HOH A 16 -35.906 -28.174 5.866 1.00 31.98 O
HETATM 666 O HOH A 17 -44.171 -7.687 17.621 1.00 22.18 O
HETATM 667 O HOH A 18 -47.265 -12.454 21.564 1.00 29.40 O
HETATM 668 O HOH A 19 -36.430 3.094 -3.026 1.00 25.02 O
HETATM 669 O HOH A 20 -29.553 -5.969 12.150 1.00 34.06 O
HETATM 670 O HOH A 21 -42.686 -4.398 27.240 1.00 25.96 O
HETATM 671 O HOH A 22 -43.889 -9.382 19.695 1.00 29.00 O
HETATM 672 O HOH A 23 -43.476 -6.477 -2.563 1.00 30.73 O
HETATM 673 O HOH A 24 -28.999 3.283 21.951 1.00 26.71 O
HETATM 674 O HOH A 25 -50.516 -11.430 14.190 1.00 25.35 O
HETATM 675 O HOH A 26 -27.306 5.304 20.576 1.00 30.44 O
HETATM 676 O HOH A 27 -48.424 -14.440 -0.286 1.00 61.67 O
HETATM 677 O HOH A 28 -43.808 -10.099 7.884 1.00 28.89 O
HETATM 678 O HOH A 29 -35.566 -5.200 24.698 1.00 29.22 O
HETATM 679 O HOH A 30 -34.679 -7.575 -4.768 1.00 25.20 O
HETATM 680 O HOH A 31 -41.964 -17.506 25.641 1.00 37.16 O
HETATM 681 O HOH A 32 -34.312 -2.922 25.191 1.00 31.83 O
HETATM 682 O HOH A 33 -51.606 -11.651 21.823 1.00 29.90 O
HETATM 683 O HOH A 34 -32.561 -16.311 28.119 1.00 50.80 O
HETATM 684 O HOH A 35 -34.469 -16.004 9.163 1.00 24.01 O
HETATM 685 O HOH A 36 -31.585 -23.210 8.833 1.00 26.89 O
HETATM 686 O HOH A 37 -49.015 -19.802 20.176 1.00 31.69 O
HETATM 687 O HOH A 38 -30.973 -14.980 5.105 1.00 43.06 O
HETATM 688 O HOH A 39 -47.022 -17.146 11.346 1.00 28.11 O
HETATM 689 O HOH A 40 -30.833 -7.743 14.123 1.00 34.35 O
HETATM 690 O HOH A 41 -25.168 6.080 14.148 1.00 49.89 O
HETATM 691 O HOH A 42 -51.167 -14.258 13.359 1.00 47.34 O
CONECT 45 288
CONECT 288 45
CONECT 382 456
CONECT 456 382
CONECT 476 641
CONECT 641 476
CONECT 644 646 648
CONECT 645 647 648
CONECT 646 644 649
CONECT 647 645 649
CONECT 648 644 645
CONECT 649 646 647
shreyakathuria@Shreyaa:~/Labsession3$

20. From protein.pdb, print only the ATOM lines that do not belong to residue ARG.



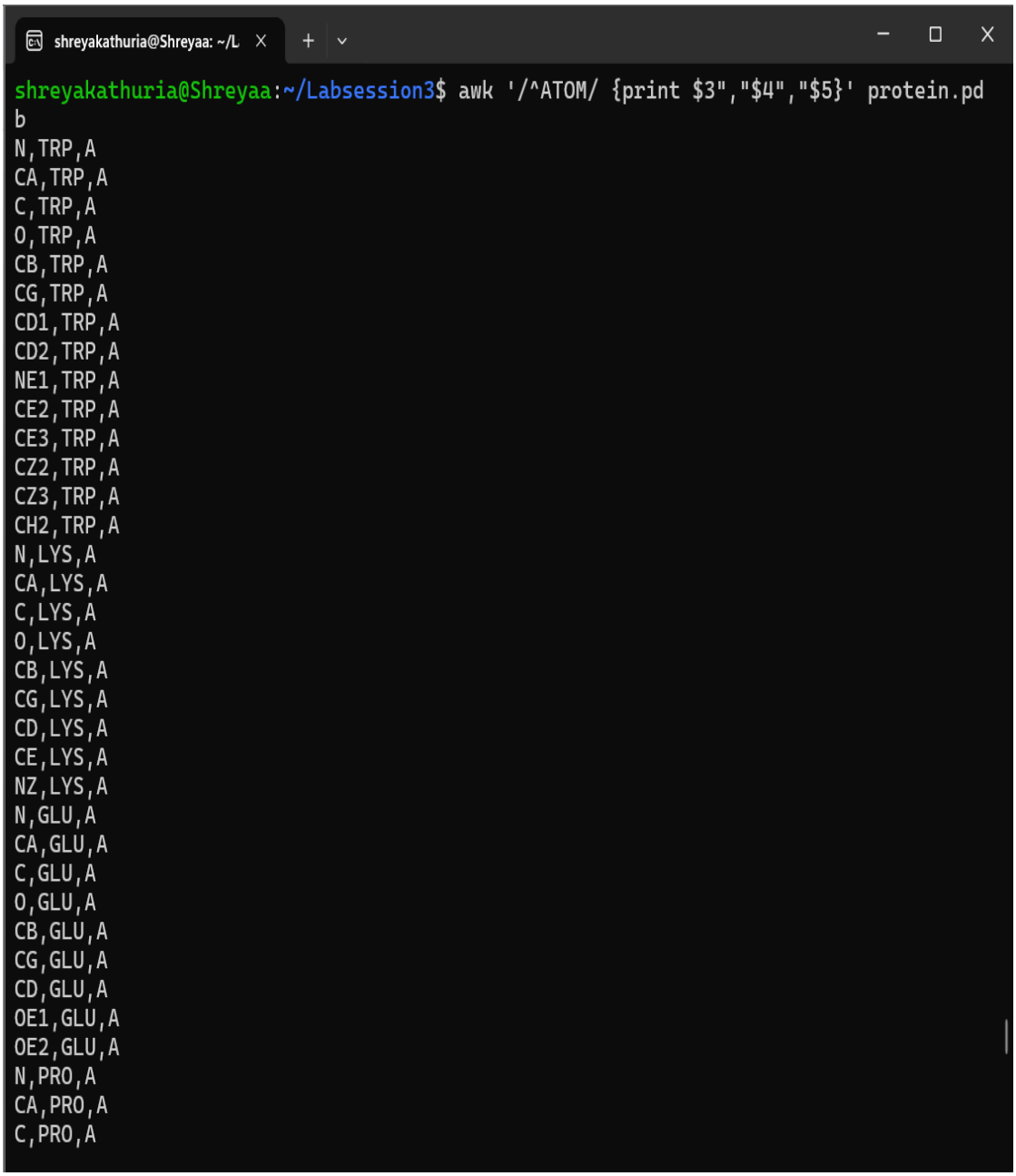

shreyakathuria@Shreyaa: ~/L
ATOM 584 O GLU A 244 -33.620 -11.854 -2.622 1.00 27.85 O
ATOM 585 CB GLU A 244 -33.138 -13.293 -0.100 1.00 28.06 C
ATOM 586 CG GLU A 244 -32.866 -12.110 0.845 1.00 29.84 C
ATOM 587 CD GLU A 244 -31.803 -12.400 1.897 1.00 30.20 C
ATOM 588 OE1 GLU A 244 -31.390 -13.572 2.032 1.00 33.67 O
ATOM 589 OE2 GLU A 244 -31.387 -11.453 2.609 1.00 34.03 O
ATOM 590 N ILE A 245 -35.563 -11.214 -1.682 1.00 26.43 N
ATOM 591 CA ILE A 245 -35.780 -10.075 -2.583 1.00 25.65 C
ATOM 592 C ILE A 245 -36.217 -8.832 -1.801 1.00 25.29 C
ATOM 593 O ILE A 245 -36.768 -8.951 -0.700 1.00 25.02 O
ATOM 594 CB ILE A 245 -36.850 -10.402 -3.671 1.00 25.53 C
ATOM 595 CG1 ILE A 245 -38.173 -10.846 -3.033 1.00 25.35 C
ATOM 596 CG2 ILE A 245 -36.325 -11.466 -4.663 1.00 24.89 C
ATOM 597 CD1 ILE A 245 -39.407 -10.707 -3.934 1.00 26.20 C
ATOM 609 N GLY A 247 -38.835 -7.026 -1.580 1.00 23.42 N
ATOM 610 CA GLY A 247 -40.259 -7.039 -1.851 1.00 24.01 C
ATOM 611 C GLY A 247 -40.829 -8.273 -1.197 1.00 24.50 C
ATOM 612 O GLY A 247 -40.080 -9.077 -0.649 1.00 24.29 O
ATOM 613 N ASP A 248 -42.149 -8.408 -1.235 1.00 25.16 N
ATOM 614 CA ASP A 248 -42.824 -9.555 -0.633 1.00 26.30 C
ATOM 615 C ASP A 248 -42.603 -10.798 -1.502 1.00 26.46 C
ATOM 616 O ASP A 248 -43.038 -10.830 -2.651 1.00 26.50 O
ATOM 617 CB ASP A 248 -44.319 -9.255 -0.467 1.00 26.46 C
ATOM 618 CG ASP A 248 -45.036 -10.274 0.400 1.00 27.46 C
ATOM 619 OD1 ASP A 248 -45.923 -9.862 1.176 1.00 29.63 O
ATOM 620 OD2 ASP A 248 -44.724 -11.481 0.312 1.00 29.49 O
ATOM 621 N PRO A 249 -41.919 -11.825 -0.954 1.00 27.01 N
ATOM 622 CA PRO A 249 -41.616 -13.040 -1.719 1.00 27.12 C
ATOM 623 C PRO A 249 -42.839 -13.935 -1.935 1.00 27.58 C
ATOM 624 O PRO A 249 -42.747 -14.917 -2.670 1.00 27.69 O
ATOM 625 CB PRO A 249 -40.575 -13.763 -0.842 1.00 27.25 C
ATOM 626 CG PRO A 249 -40.172 -12.767 0.229 1.00 26.22 C
ATOM 627 CD PRO A 249 -41.376 -11.906 0.415 1.00 26.82 C
ATOM 628 N ASN A 250 -43.965 -13.578 -1.315 1.00 28.06 N
ATOM 629 CA ASN A 250 -45.188 -14.384 -1.319 1.00 28.66 C
ATOM 630 C ASN A 250 -44.915 -15.766 -0.728 1.00 28.70 C
ATOM 631 O ASN A 250 -44.998 -16.798 -1.417 1.00 29.19 O
ATOM 632 CB ASN A 250 -45.827 -14.456 -2.721 1.00 29.27 C
ATOM 633 CG ASN A 250 -46.426 -13.126 -3.167 1.00 31.00 C
ATOM 634 OD1 ASN A 250 -46.227 -12.687 -4.308 1.00 34.38 O
ATOM 635 ND2 ASN A 250 -47.167 -12.478 -2.272 1.00 31.93 N
ATOM 636 N CYS A 251 -44.571 -15.756 0.557 1.00 28.18 N
ATOM 637 CA CYS A 251 -44.118 -16.941 1.273 1.00 27.85 C
ATOM 638 C CYS A 251 -45.248 -17.940 1.495 1.00 28.66 C
ATOM 639 O CYS A 251 -46.370 -17.574 1.840 1.00 29.40 O
ATOM 640 CB CYS A 251 -43.484 -16.541 2.607 1.00 27.05 C
ATOM 641 SG CYS A 251 -41.988 -15.536 2.435 1.00 22.88 S
ATOM 642 OXT CYS A 251 -45.044 -19.143 1.329 1.00 29.75 O
shreyakathuria@Shreyaa:~/Labsession3$

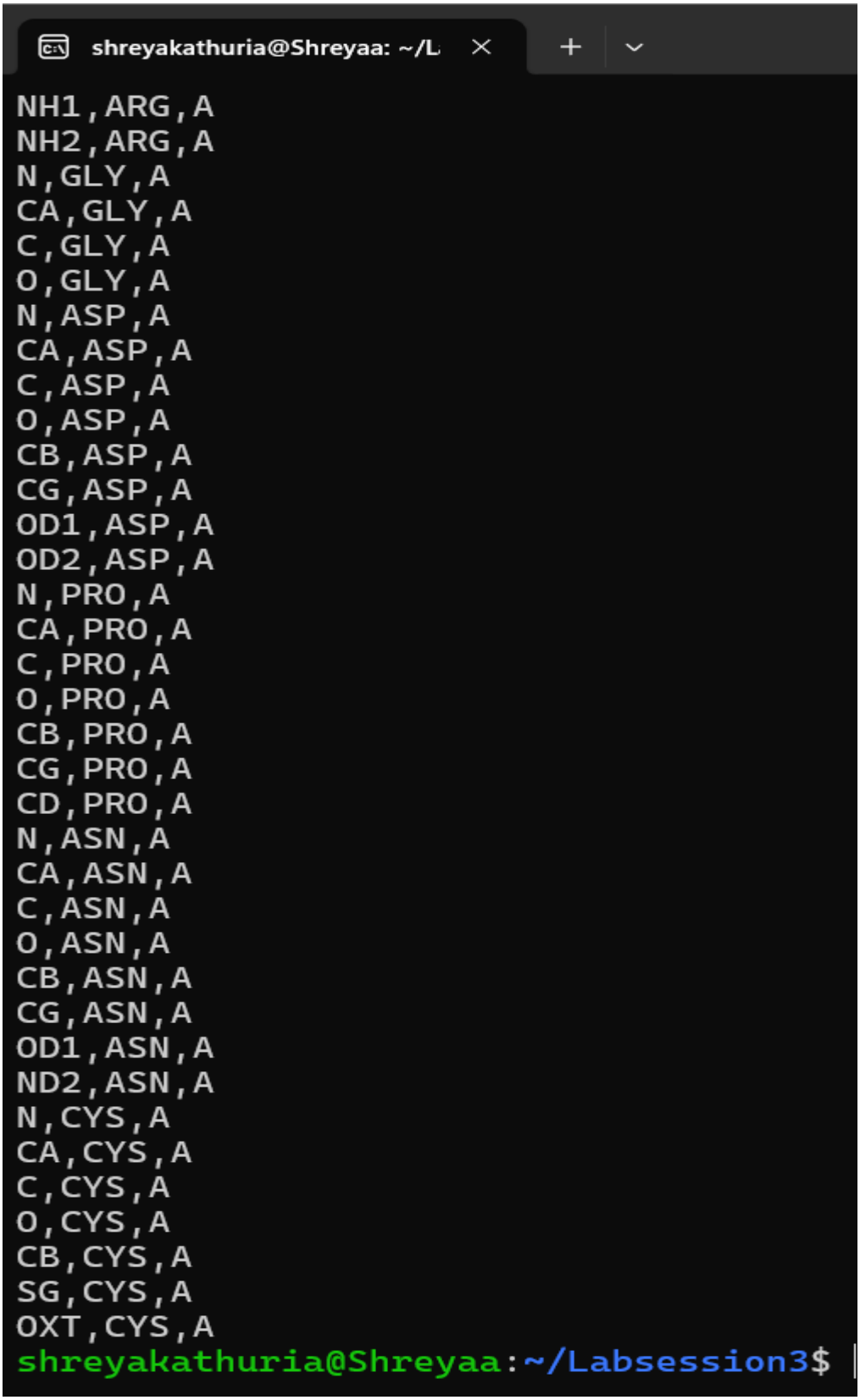
21. Extract all residues and their frequencies from chain A.



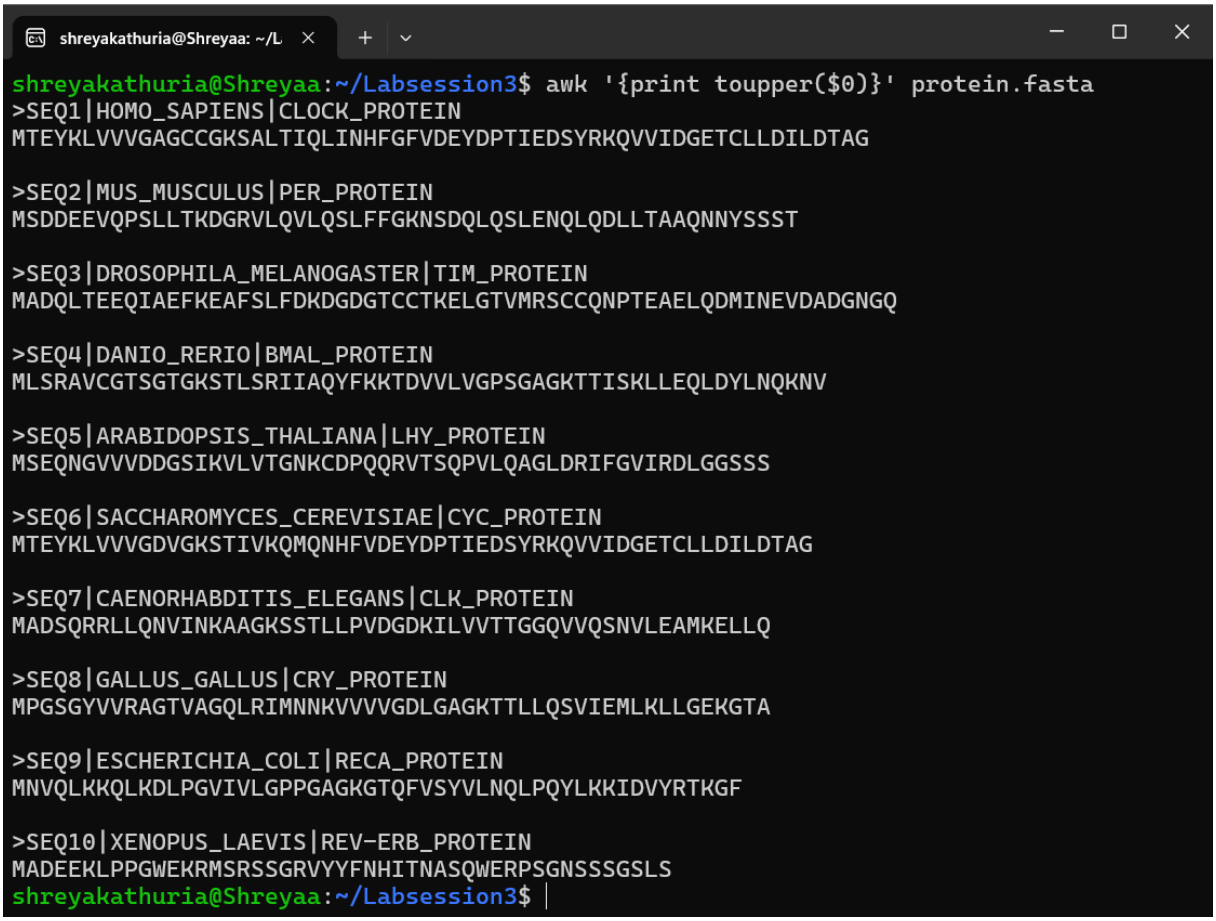
Used gemini for loop.

22. From protein.pdb, print only atom name, residue name, and chain ID.





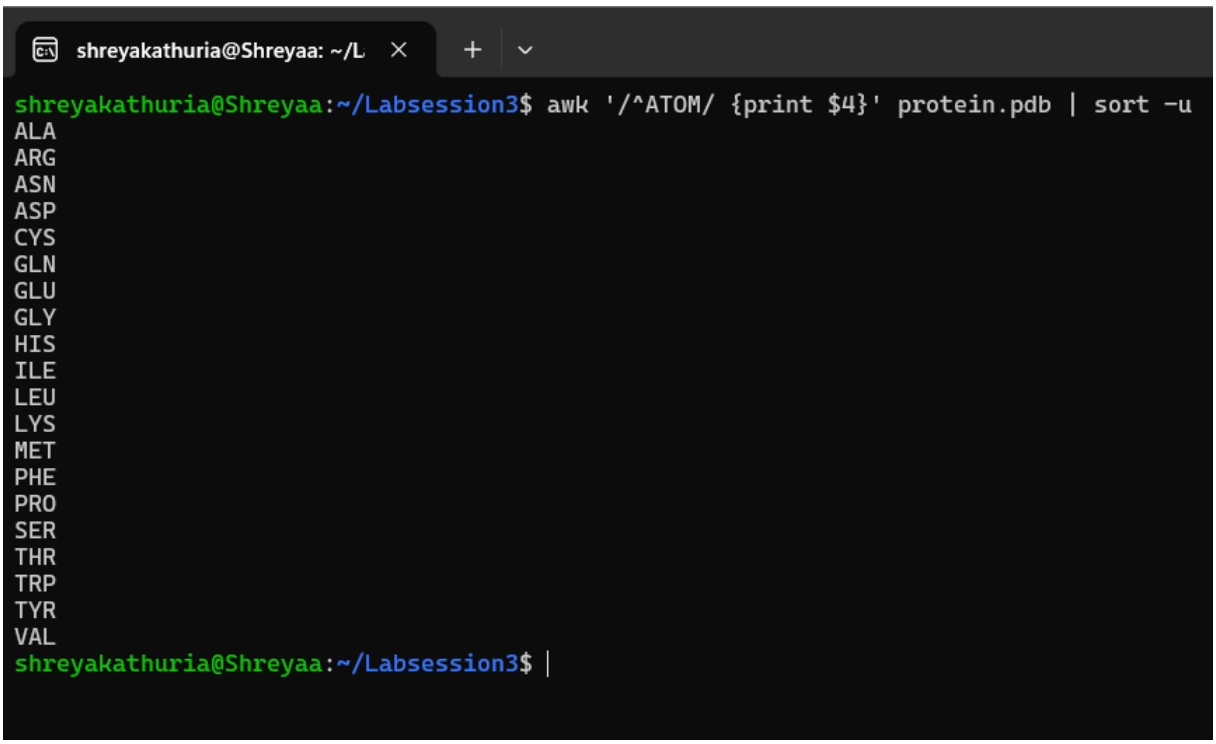
22. Replace all lowercase letters in sequences of protein.fasta with uppercase



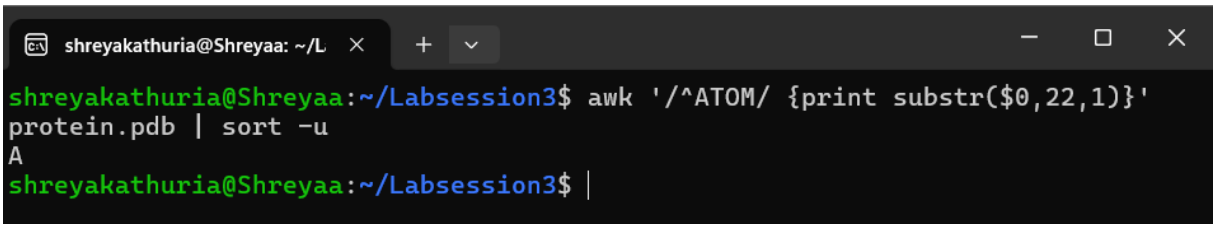
23. Find the sequence(s) in protein.fasta with the maximum length.

I used gemini for executing an if command and I am still not sure about how it works. (in removing headers)

24. Extract unique residue names from protein.pdb and sort them alphabetically.



25. Find how many distinct chains are present in protein.pdb.



Used gemini for Substr command and for knowing that in a standard pdb file 22nd column gives you the chain

26. From clock_gene.fasta, count nucleotide frequencies (A, T, G, C) separately.