

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)

```
shreyakathuria@Shreyaa: ~/L x + v
shreyakathuria@Shreyaa:~/Labsession3$ cat biocomputing
cat
dog
ant
spider

elephant
shreyakathuria@Shreyaa:~/Labsession3$ sed '2~2d' biocomputing
cat
dog
ant
spider
elephant
shreyakathuria@Shreyaa:~/Labsession3$ |
```

```
shreyakathuria@Shreyaa: ~/L x + v
shreyakathuria@Shreyaa:~/Labsession3$ cat biocomputing
cat
dog
ant
spider

elephant
shreyakathuria@Shreyaa:~/Labsession3$ sed '2~2d' biocomputing
cat
dog
ant
spider
elephant
shreyakathuria@Shreyaa:~/Labsession3$ sed '/^$/d' biocomputing
sed: -e expression #1, char 5: missing command
shreyakathuria@Shreyaa:~/Labsession3$ sed '/^$/d' biocomputing
cat
dog
ant
spider
elephant
shreyakathuria@Shreyaa:~/Labsession3$ |
```

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

```
shreyakathuria@Shreyaa: ~/L x + v
shreyakathuria@Shreyaa:~/Labsession3$ sed -i '/^$/d' biocomputing
shreyakathuria@Shreyaa:~/Labsession3$ cat biocomputing
cat
dog
ant
spider
elephant
shreyakathuria@Shreyaa:~/Labsession3$ nl biocomputing > biocomp
shreyakathuria@Shreyaa:~/Labsession3$ cat biocomp
1 cat
2 dog
3 ant
4 spider
5 elephant
shreyakathuria@Shreyaa:~/Labsession3$ |
```

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

```
shreyakathuria@Shreyaa: ~/L x + v
shreyakathuria@Shreyaa:~/Labsession3$ sed -n '/^>/p' clock_gene.fasta
>NC_000004.12:c55546909-55427903 Homo sapiens chromosome 4, GRCh38.p14 Primary
Assembly
shreyakathuria@Shreyaa:~/Labsession3$
```

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

```
shreyakathuria@Shreyaa: ~/L x + v
shreyakathuria@Shreyaa:~/Labsession3$ grep "^>.*CLOCK" protein.fasta
>seq1|Homo_sapiens|CLOCK_protein
shreyakathuria@Shreyaa:~/Labsession3$ |
```

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

```
shreyakathuria@Shreyaa: ~/L x + v
shreyakathuria@Shreyaa:~/Labsession3$ grep "CC" protein.fasta
MTEYKLVVVGAGCCGKSALTIQLInhfgFVDEYDPTIEDSYRKQVVIDGETCLLDILDTAG
MADQLTEEQIAEFKEAFSLFDKDGDTCTKELGTVMRSCCQNPTEAELQDMINEVDADGNGQ
shreyakathuria@Shreyaa:~/Labsession3$
```

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

```
shreyakathuria@Shreyaa: ~/L  × + v - □ ×
shreyakathuria@Shreyaa:~/Labsession3$ grep -o "G" clock_gene.fasta | wc -l
23472
shreyakathuria@Shreyaa:~/Labsession3$ |
```

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

```
shreyakathuria@Shreyaa: ~/L  × + v -
shreyakathuria@Shreyaa:~/Labsession3$ sed -n '5,28p' clock_gene.fasta
GTGGAGGAGGGGAAGGGAAGGGAGGGGAGGAGGAGCTGGCCACAGGAGCGGCGAATTTTTGGGGGGGTG
GGTGGGGGGCGCCACTCACAGCCCCAGGTGCTGCTGGAGGTGGGAGCCGCGGCGCCTCCTGGACACAGGC
GGGGTAGTGGTTCGAGTCAACGCAGCGGGAGACCTGGGTGGGGGAGGGAAGAAGCCGGAGCCGCCGAA
GCCACACGGTGAGGGCGCGGGGAAGGGGAGGGAGCGGGGGGCGGCGTGTGTGGGGCCGGGGGGCGGCGGC
CAAGGGTGGGGAAGGCGGGAGCTGAAGCCCAAGTTTGGCGTGTCTGTCTAGTGTGTCTTTCCCGGGACT
TCGGGCGGAGGCCCGCCCTGCCTGAGAGGCCCTCTGGGGCAGCTGGGGTTACCTGCGGGGCAGGGGCGGG
AGTGGGGTGCACGGCGGGGCCGGCGGCTTGAGGGCGCCCGAGCTGCGGCCGATTCCAGCAGCTGGGAG
GCGGGGAAAGACGGGGACCGGGTGCCGAGAGAGCTTTCGCTGGGGACCCGCTAGGCCTTGTGACCCACTT
TATTCCTGTCACTACTCGGGCACGTTTGGAGCAGCGCCCAATGGGGCGCCGGGGCGGCCAGCTCCTCCGG
GGAACCCCCGCCCTCCCGGCGCCCGGCCGCGTGGCCGAGTCCGCAGTCCGAACGGCCGCCGTTGCCGGC
CGCGGGCTGGTTCGGTTAGTGGTGGTGGTTCGGGGTTCCGTTCTAGGCAGCGCGCGGCTATTAGCGTC
TGACTCCAGCGACCGCGCGGGTTTCGAGGGTTGGCGGCGAGGCGCTCGGTTTCTCTTCCGTCCACC
CGCGCTTCCCGTTCCCGTCCCGTCCCGTCCCGTGGCGTGTGTAGATTCTTTCCGCCAGTGAAGCTGGGTTTC
TGGAGTTGGCTCTGGCGCTCTGGCCCTGGAGTGTAATTTCTACACGCAGCGCCGAGAGTTTATATTC
TTTGAAAGTGTGTGTAGCTTTGTAGAGTCTCTTGTGATGGTAGGTGAGCCTAATTCTGCAAGATAAA
AGCCTAGTCTCTGACCTGGCAGATGAAAGATCAAATCAGATTGTGGTTTCCTGCTATTAGAATGCCGTGC
TATTAGACTTTAAGGCTTTTAGCCTTCTTTAAAAAATAAAAAAATTTTACAGTGGAAGAAAAGCACAA
GAAGTAAACTTTTACAGTCGTTGATTTGACTATAACGCTGATCCCCCAAATCAAAGGTAATTTCACTTT
GAAGATTGCGTCTGATTTGTAGCTTTAAGCGATTAGAGAAAATGTGCAATATTTCCCTCTACCTGTTT
GAAATAAAACATTCTTAAAGGATGTAATTTAGATAATGAATTGCTTTCTGAAACTTATCCCTTGGA
CACCTCAAATCTGATTGGTTTTCAAAGCTTGGGGGAAGGAAATTAATTCCTGTGATAAGTGGTTGGCTGA
ACAGAATGTCTTGAAAGGTTAGCCTGTAGCATTAGGAGAAAATACCTAATGTAAACGACGAGTTAATGGGT
GCAGCACACCAGCGTGGCACATTTATACATATGTGACAAACCTGCACGTTGTGCACATGTACCCTAGAAC
TTAAAGTATAATAATAAAAAAAGTAAAAAAGTTAGCCTGAAGAAAGCAGACTGAAAATGTTCT
shreyakathuria@Shreyaa:~/Labsession3$ |
```

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

```
shreyakathuria@Shreyaa: ~/L  × + v - □ ×
shreyakathuria@Shreyaa:~/Labsession3$ awk -F'|' '/^>/ {sub(/^>/, "", $1); print $1}' protein.fasta
seq1
seq2
seq3
seq4
seq5
seq6
seq7
seq8
seq9
seq10
shreyakathuria@Shreyaa:~/Labsession3$ |
```

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.

```
shreyakathuria@Shreyaa: ~/L  × + v
shreyakathuria@Shreyaa:~/Labsession3$ awk '/^M.*Q$/' protein.fasta
MADQLTEEQIAEFKEAFSLFDKDGDTCTKELGTVMRSCQNPTAEELQDMINEVDADGNGQ
MADSQRRLQLQNVINKAAGKSSTLLPVDGDKILVTTGGQVVQSNVLEAMKELLQ
shreyakathuria@Shreyaa:~/Labsession3$ |
```

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

```
shreyakathuria@Shreyaa:~/Labsession3$ awk '/^>/ {if (seqlen > 0) print ID, seqlen; ID=$0; seqlen=0; next} {seqlen+=length($0)} END {print ID, seqlen}' prote
in.fasta | sed 's/^>/'
seq1|Homo_sapiens|CLOCK_protein 61
seq2|Mus_musculus|PER_protein 56
seq3|Drosophila_melanogaster|TIM_protein 63
seq4|Danio_rerio|BMAL_protein 58
seq5|Arabidopsis_thaliana|LHY_protein 54
seq6|Saccharomyces_cerevisiae|CYC_protein 57
seq7|Caenorhabditis_elegans|CLW_protein 54
seq8|Gallus_gallus|CRY_protein 54
seq9|Escherichia_coli|RecA_protein 52
seq10|Xenopus_laevis|REV-ERB_protein 47
shreyakathuria@Shreyaa:~/Labsession3$
```

Used gemini for removing headers using sed.

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

```
shreyakathuria@Shreyaa: ~/L × + ∨ − □ ×
shreyakathuria@Shreyaa:~/Labsession3$
shreyakathuria@Shreyaa:~/Labsession3$ awk '$1 == "ATOM" && $5 == "A"' protein.pdb
ATOM      1  N   TRP A 172    -39.136 -21.997  24.415  1.00 34.43    N
ATOM      2  CA  TRP A 172    -40.108 -20.907  24.729  1.00 34.28    C
ATOM      3  C   TRP A 172    -41.403 -21.065  23.944  1.00 33.46    C
ATOM      4  O   TRP A 172    -41.385 -21.496  22.789  1.00 33.48    O
ATOM      5  CB  TRP A 172    -39.506 -19.534  24.418  1.00 35.12    C
ATOM      6  CG  TRP A 172    -38.161 -19.292  25.025  1.00 36.34    C
ATOM      7  CD1 TRP A 172    -37.773 -19.568  26.306  1.00 37.69    C
ATOM      8  CD2 TRP A 172    -37.032 -18.693  24.384  1.00 37.47    C
ATOM      9  NE1 TRP A 172    -36.465 -19.190  26.497  1.00 37.97    N
ATOM     10  CE2 TRP A 172    -35.985 -18.650  25.334  1.00 37.83    C
ATOM     11  CE3 TRP A 172    -36.799 -18.192  23.097  1.00 37.57    C
ATOM     12  CZ2 TRP A 172    -34.725 -18.128  25.037  1.00 37.51    C
ATOM     13  CZ3 TRP A 172    -35.545 -17.671  22.802  1.00 37.85    C
ATOM     14  CH2 TRP A 172    -34.523 -17.646  23.769  1.00 37.43    C
ATOM     15  N   LYS A 173    -42.516 -20.697  24.576  1.00 32.18    N
ATOM     16  CA  LYS A 173    -43.842 -20.728  23.949  1.00 31.37    C
ATOM     17  C   LYS A 173    -44.028 -19.604  22.914  1.00 29.85    C
ATOM     18  O   LYS A 173    -44.831 -19.725  21.976  1.00 30.15    O
ATOM     19  CB  LYS A 173    -44.935 -20.645  25.024  1.00 31.31    C
ATOM     20  CG  LYS A 173    -46.343 -20.964  24.519  1.00 32.53    C
ATOM     21  CD  LYS A 173    -47.425 -20.459  25.479  1.00 32.89    C
ATOM     22  CE  LYS A 173    -48.818 -20.684  24.901  1.00 33.96    C
ATOM     23  NZ  LYS A 173    -49.893 -20.189  25.806  1.00 34.66    N
ATOM     24  N   GLU A 174    -43.280 -18.518  23.090  1.00 27.67    N
ATOM     25  CA  GLU A 174    -43.337 -17.366  22.191  1.00 25.77    C
ATOM     26  C   GLU A 174    -41.922 -17.014  21.728  1.00 23.54    C
ATOM     27  O   GLU A 174    -41.381 -15.977  22.138  1.00 23.23    O
ATOM     28  CB  GLU A 174    -43.933 -16.148  22.913  1.00 25.76    C
ATOM     29  CG  GLU A 174    -45.376 -16.258  23.359  1.00 26.89    C
ATOM     30  CD  GLU A 174    -45.777 -15.061  24.206  1.00 27.42    C
ATOM     31  OE1 GLU A 174    -46.102 -14.001  23.639  1.00 29.42    O
ATOM     32  OE2 GLU A 174    -45.756 -15.182  25.445  1.00 30.63    O
ATOM     33  N   PRO A 175    -41.313 -17.867  20.872  1.00 21.55    N
ATOM     34  CA  PRO A 175    -39.891 -17.705  20.564  1.00 20.10    C
ATOM     35  C   PRO A 175    -39.565 -16.385  19.866  1.00 18.58    C
ATOM     36  O   PRO A 175    -38.520 -15.781  20.142  1.00 18.18    O
ATOM     37  CB  PRO A 175    -39.594 -18.893  19.632  1.00 20.52    C
ATOM     38  CG  PRO A 175    -40.909 -19.247  19.043  1.00 19.77    C
ATOM     39  CD  PRO A 175    -41.896 -19.015  20.148  1.00 21.28    C
ATOM     40  N   CYS A 176    -40.455 -15.942  18.986  1.00 16.73    N
ATOM     41  CA  CYS A 176    -40.212 -14.710  18.226  1.00 16.80    C
ATOM     42  C   CYS A 176    -40.222 -13.501  19.159  1.00 16.78    C
ATOM     43  O   CYS A 176    -39.363 -12.626  19.053  1.00 16.20    O
ATOM     44  CB  CYS A 176    -41.244 -14.528  17.116  1.00 16.50    C
ATOM     45  SG  CYS A 176    -40.885 -13.084  16.044  1.00 15.20    S
ATOM     46  N   ARG A 177    -41.200 -13.469  20.062  1.00 17.53    N
ATOM     47  CA  ARG A 177    -41.351 -12.338  20.984  1.00 18.15    C
```

```
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 × + ▾
shreyakathuria@Shreyaa:~/Labsession3$ awk '$1 == "ATOM" && ($4 == "LYS" || $4 == "ARG")' protein.pdb
ATOM      15  N   LYS A 173    -42.516 -20.697  24.576  1.00 32.18    N
ATOM      16  CA  LYS A 173    -43.842 -20.728  23.949  1.00 31.37    C
ATOM      17  C   LYS A 173    -44.028 -19.604  22.914  1.00 29.85    C
ATOM      18  O   LYS A 173    -44.831 -19.725  21.976  1.00 30.15    O
ATOM      19  CB  LYS A 173    -44.935 -20.645  25.024  1.00 31.31    C
ATOM      20  CG  LYS A 173    -46.343 -20.964  24.519  1.00 32.53    C
ATOM      21  CD  LYS A 173    -47.425 -20.459  25.479  1.00 32.89    C
ATOM      22  CE  LYS A 173    -48.818 -20.684  24.901  1.00 33.96    C
ATOM      23  NZ  LYS A 173    -49.893 -20.189  25.806  1.00 34.66    N
ATOM      46  N   ARG A 177    -41.200 -13.469  20.062  1.00 17.53    N
ATOM      47  CA  ARG A 177    -41.351 -12.338  20.984  1.00 18.15    C
ATOM      48  C   ARG A 177    -40.135 -12.196  21.880  1.00 18.13    C
ATOM      49  O   ARG A 177    -39.608 -11.088  22.053  1.00 17.51    O
ATOM      50  CB  ARG A 177    -42.634 -12.450  21.807  1.00 18.62    C
ATOM      51  CG  ARG A 177    -42.872 -11.237  22.713  1.00 20.72    C
ATOM      52  CD  ARG A 177    -44.227 -11.292  23.368  1.00 22.66    C
ATOM      53  NE  ARG A 177    -44.366 -10.263  24.391  1.00 24.94    N
ATOM      54  CZ  ARG A 177    -43.848 -10.348  25.616  1.00 25.91    C
ATOM      55  NH1  ARG A 177    -43.147 -11.413  25.983  1.00 25.04    N
ATOM      56  NH2  ARG A 177    -44.030 -9.360  26.477  1.00 26.28    N
ATOM      94  N   ARG A 182    -34.717 -9.406  22.797  1.00 19.68    N
ATOM      95  CA  ARG A 182    -33.268 -9.544  22.849  1.00 20.05    C
ATOM      96  C   ARG A 182    -32.593 -8.739  21.743  1.00 19.42    C
ATOM      97  O   ARG A 182    -31.576 -8.072  21.990  1.00 19.22    O
ATOM      98  CB  ARG A 182    -32.874 -11.019  22.769  1.00 20.66    C
ATOM      99  CG  ARG A 182    -33.592 -11.864  23.806  1.00 23.33    C
ATOM     100  CD  ARG A 182    -32.691 -12.324  24.917  1.00 31.08    C
ATOM     101  NE  ARG A 182    -32.238 -13.693  24.676  1.00 34.53    N
ATOM     102  CZ  ARG A 182    -32.720 -14.777  25.285  1.00 36.34    C
ATOM     103  NH1  ARG A 182    -33.684 -14.685  26.205  1.00 37.09    N
ATOM     104  NH2  ARG A 182    -32.223 -15.966  24.975  1.00 37.59    N
ATOM     147  N   LYS A 189    -27.943 -1.219  22.313  1.00 19.72    N
ATOM     148  CA  LYS A 189    -26.592 -1.220  22.859  1.00 19.83    C
ATOM     149  C   LYS A 189    -25.535 -0.931  21.783  1.00 19.51    C
ATOM     150  O   LYS A 189    -24.637 -0.121  22.008  1.00 19.20    O
ATOM     151  CB  LYS A 189    -26.300 -2.544  23.584  1.00 19.67    C
ATOM     152  CG  LYS A 189    -24.980 -2.573  24.353  1.00 21.18    C
ATOM     153  CD  LYS A 189    -24.991 -1.568  25.500  1.00 23.97    C
ATOM     154  CE  LYS A 189    -23.703 -1.601  26.298  1.00 25.23    C
ATOM     155  NZ  LYS A 189    -23.673 -0.401  27.204  1.00 25.98    N
ATOM     228  N   LYS A 200    -30.993  0.420  7.874  1.00 26.73    N
ATOM     229  CA  LYS A 200    -31.745 -0.835  7.833  1.00 24.20    C
ATOM     230  C   LYS A 200    -31.208 -1.820  8.880  1.00 23.56    C
ATOM     231  O   LYS A 200    -30.014 -1.861  9.160  1.00 23.03    O
ATOM     232  CB  LYS A 200    -31.682 -1.479  6.440  1.00 24.17    C
ATOM     233  CG  LYS A 200    -32.216 -0.609  5.294  1.00 23.41    C
ATOM     234  CD  LYS A 200    -32.263 -1.375  3.981  1.00 22.93    C
ATOM     235  CE  LYS A 200    -32.479 -0.443  2.786  1.00 21.93    C
ATOM     236  NZ  LYS A 200    -31.331  0.512  2.647  1.00 19.78    N
ATOM     297  N   LYS A 208    -49.012 -12.189  16.590  1.00 19.70    N
ATOM     298  CA  LYS A 208    -49.580 -11.893  17.916  1.00 20.21    C
ATOM     299  C   LYS A 208    -49.491 -13.063  18.913  1.00 20.08    C
ATOM     300  O   LYS A 208    -49.635 -12.860  20.118  1.00 20.32    O
ATOM     301  CB  LYS A 208    -51.043 -11.459  17.773  1.00 20.47    C
ATOM     302  CG  LYS A 208    -51.935 -12.512  17.115  1.00 20.38    C
ATOM     303  CD  LYS A 208    -53.396 -12.222  17.359  1.00 22.10    C
ATOM     304  CE  LYS A 208    -54.291 -13.221  16.642  1.00 20.94    C
ATOM     305  NZ  LYS A 208    -54.187 -14.607  17.174  1.00 20.34    N
ATOM     357  N   ARG A 215    -43.344 -14.515  6.254  1.00 18.42    N
ATOM     358  CA  ARG A 215    -42.464 -13.537  5.651  1.00 18.42    C
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
```



```
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  ×  +  ▾
shreyakathuria@Shreyaa:~/Lab$ sed 's/LYS/ARG/g' protein.pdb
HEADER      PEPTIDE BINDING PROTEIN                      26-MAY-05   1ZT3
TITLE       C-TERMINAL DOMAIN OF INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN-1
TITLE       2 ISOLATED FROM HUMAN AMNIOTIC FLUID
COMPND      MOL_ID: 1;
COMPND      2 MOLECULE: INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 1;
COMPND      3 CHAIN: A;
COMPND      4 FRAGMENT: C-TERMINAL DOMAIN;
COMPND      5 SYNONYM: IGFBP-1, IBP- 1, IGF-BINDING PROTEIN 1, PLACENTAL PROTEIN
COMPND      6 12, PP12
SOURCE      MOL_ID: 1;
SOURCE      2 ORGANISM_SCIENTIFIC: HOMO SAPIENS;
SOURCE      3 ORGANISM_COMMON: HUMAN;
SOURCE      4 ORGANISM_TAXID: 9606;
SOURCE      5 OTHER_DETAILS: AMNIOTIC FLUID
KEYWDS      INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN-1, IGFBP-1, AMNIOTIC
KEYWDS      2 FLUID, C-TERMINAL DOMAIN, METAL-BINDING, PEPTIDE BINDING PROTEIN
EXPDTA      X-RAY DIFFRACTION
AUTHOR      A.SALA,S.CAPALDI,M.CAMPAGNOLI,B.FAGGION,S.LABO,M.PERDUCA,A.ROMANO,
AUTHOR      2 M.E.CARRIZO,M.VALLI,L.VISAI,L.MINCHIOTTI,M.GALLIANO,H.L.MONACO
REVDAT      5   16-OCT-24 1ZT3   1      REMARK
REVDAT      4   11-OCT-17 1ZT3   1      REMARK
REVDAT      3   24-FEB-09 1ZT3   1      VERSN
REVDAT      2   30-AUG-05 1ZT3   1      JRNL
REVDAT      1   28-JUN-05 1ZT3   0
JRNL         AUTH   A.SALA,S.CAPALDI,M.CAMPAGNOLI,B.FAGGION,S.LABO,M.PERDUCA,
JRNL         AUTH 2 A.ROMANO,M.E.CARRIZO,M.VALLI,L.VISAI,L.MINCHIOTTI,
JRNL         AUTH 3 M.GALLIANO,H.L.MONACO
JRNL         TITL   STRUCTURE AND PROPERTIES OF THE C-TERMINAL DOMAIN OF
JRNL         TITL 2 INSULIN-LIKE GROWTH FACTOR-BINDING PROTEIN-1 ISOLATED FROM
JRNL         TITL 3 HUMAN AMNIOTIC FLUID
JRNL         REF    J.BIOL.CHEM.                      V. 280 29812 2005
JRNL         REFN   ISSN 0021-9258
JRNL         PMID   15972819
JRNL         DOI    10.1074/JBC.M504304200
REMARK      2
REMARK      2 RESOLUTION.      1.80 ANGSTROMS.
REMARK      3
REMARK      3 REFINEMENT.
REMARK      3   PROGRAM      : REFMAC 5.2.0005
REMARK      3   AUTHORS      : MURSHUDOV,SKUBAK,LEBEDEV,PANNU,STEINER,
REMARK      3                   : NICHOLLS,WINN,LONG,VAGIN
REMARK      3
```



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 × + v
shreyakathuria@Shreyaa:~/Labsession3$ awk '/^ATOM/ {print $9}' protein.pdb
24.415
24.729
23.944
22.789
24.418
25.025
26.306
24.384
26.497
25.334
23.097
25.037
22.802
23.769
24.576
23.949
22.914
21.976
25.024
24.519
25.479
24.901
25.806
23.090
22.191
21.728
22.138
22.913
23.359
24.206
23.639
25.445
20.872
20.564
19.866
20.142
19.632
19.043
20.148
18.986
```



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.

```
shreyakathuria@Shreyaa: ~/L × + v
shreyakathuria@Shreyaa:~/Labsession3$ grep 'GLY' protein.pdb | wc -l
33
shreyakathuria@Shreyaa:~/Labsession3$ |
```

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

```
shreyakathuria@Shreyaa:~/Labsession3$ awk '/^ATOM.*CA/ && ($4 == "GLY" || $4 == "ALA")' protein.pdb
ATOM 143 CA ALA A 188 -29.906 -0.273 21.249 1.00 19.62 C
ATOM 157 CA ALA A 190 -24.689 -1.402 19.528 1.00 20.13 C
ATOM 193 CA GLY A 195 -19.179 3.890 13.965 1.00 34.45 C
ATOM 315 CA GLY A 210 -45.353 -14.753 19.536 1.00 18.56 C
ATOM 422 CA GLY A 223 -36.815 5.170 1.658 1.00 21.58 C
ATOM 435 CA ALA A 225 -37.186 -1.492 0.463 1.00 20.30 C
ATOM 440 CA GLY A 226 -35.705 -3.955 2.980 1.00 18.85 C
ATOM 526 CA GLY A 236 -37.957 -18.276 12.295 1.00 18.22 C
ATOM 565 CA GLY A 241 -34.199 -22.463 -1.334 1.00 28.67 C
ATOM 610 CA GLY A 247 -40.259 -7.039 -1.851 1.00 24.01 C
shreyakathuria@Shreyaa:~/Labsession3$ |
```

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

]

```
shreyakathuria@Shreyaa: ~/L × + v
shreyakathuria@Shreyaa:~/Labsession3$ grep ' C ' protein.pdb | wc -l
405
shreyakathuria@Shreyaa:~/Labsession3$ |
```

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

```
shreyakathuria@Shreyaa: ~/L × + v
shreyakathuria@Shreyaa:~/Labsession3$ awk '/^HETATM/ {print}' protein.pdb
HETATM 644 C1 DIO A 400 -29.064 -6.946 17.132 1.00 36.16 C
HETATM 645 C2 DIO A 400 -28.073 -9.061 16.720 1.00 36.92 C
HETATM 646 C1' DIO A 400 -27.687 -6.281 17.202 1.00 35.99 C
HETATM 647 C2' DIO A 400 -26.684 -8.437 16.825 1.00 36.68 C
HETATM 648 O1 DIO A 400 -28.996 -8.072 16.254 1.00 36.78 O
HETATM 649 O1' DIO A 400 -26.726 -7.251 17.629 1.00 36.28 O
HETATM 650 O HOH A 1 -37.255 -6.228 10.647 1.00 14.97 O
HETATM 651 O HOH A 2 -22.012 -0.788 22.336 1.00 20.64 O
HETATM 652 O HOH A 3 -38.877 -3.391 4.471 1.00 20.33 O
HETATM 653 O HOH A 4 -34.212 -23.871 7.998 1.00 18.39 O
HETATM 654 O HOH A 5 -20.730 -0.315 24.894 1.00 20.65 O
HETATM 655 O HOH A 6 -44.936 -13.438 1.965 1.00 28.30 O
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
shreyakathuria@Shreyaa:~/Labsession3$
```

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

```
shreyakathuria@Shreyaa: ~/L × + v
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 × + v
shreyakathuria@Shreyaa:~/Labssession3$ sed '/TER/d; /END/d' protein.pdb
HEADER      PEPTIDE BINDING PROTEIN                      26-MAY-05   1ZT3
TITLE       2 ISOLATED FROM HUMAN AMNIOTIC FLUID
COMPND      MOL_ID: 1;
COMPND      2 MOLECULE: INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 1;
COMPND      3 CHAIN: A;
COMPND      5 SYNONYM: IGFBP-1, IBP- 1, IGF-BINDING PROTEIN 1, PLACENTAL PROTEIN
COMPND      6 12, PP12
SOURCE      MOL_ID: 1;
SOURCE      2 ORGANISM_SCIENTIFIC: HOMO SAPIENS;
SOURCE      3 ORGANISM_COMMON: HUMAN;
SOURCE      4 ORGANISM_TAXID: 9606;
SOURCE      5 OTHER_DETAILS: AMNIOTIC FLUID
KEYWDS      INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN-1, IGFBP-1, AMNIOTIC
EXPDTA      X-RAY DIFFRACTION
AUTHOR      A.SALA,S.CAPALDI,M.CAMPAGNOLI,B.FAGGION,S.LABO,M.PERDUCA,A.ROMANO,
AUTHOR      2 M.E.CARRIZO,M.VALLI,L.VISAI,L.MINCHIOTTI,M.GALLIANO,H.L.MONACO
REVDAT      5 16-OCT-24 1ZT3      1      REMARK
REVDAT      4 11-OCT-17 1ZT3      1      REMARK
REVDAT      3 24-FEB-09 1ZT3      1      VERSN
REVDAT      2 30-AUG-05 1ZT3      1      JRNL
REVDAT      1 28-JUN-05 1ZT3      0
JRNL         AUTH  A.SALA,S.CAPALDI,M.CAMPAGNOLI,B.FAGGION,S.LABO,M.PERDUCA,
JRNL         AUTH 2 A.ROMANO,M.E.CARRIZO,M.VALLI,L.VISAI,L.MINCHIOTTI,
JRNL         AUTH 3 M.GALLIANO,H.L.MONACO
JRNL         TITL 2 INSULIN-LIKE GROWTH FACTOR-BINDING PROTEIN-1 ISOLATED FROM
JRNL         TITL 3 HUMAN AMNIOTIC FLUID
JRNL         REF   J.BIOL.CHEM.                      V. 280 29812 2005
JRNL         REFN                      ISSN 0021-9258
JRNL         PMID  15972819
JRNL         DOI   10.1074/JBC.M504304200
REMARK      2
REMARK      2 RESOLUTION.      1.80 ANGSTROMS.
REMARK      3
REMARK      3 REFINEMENT.
REMARK      3 PROGRAM      : REFMAC 5.2.0005
REMARK      3 AUTHORS      : MURSHUDOV,SKUBAK,LEBEDEV,PANNU,STEINER,
REMARK      3              : NICHOLLS,WINN,LONG,VAGIN
REMARK      3
REMARK      3 REFINEMENT TARGET : MAXIMUM LIKELIHOOD
REMARK      3
REMARK      3 DATA USED IN REFINEMENT.
REMARK      3 RESOLUTION RANGE HIGH (ANGSTROMS) : 1.80
REMARK      3 RESOLUTION RANGE LOW (ANGSTROMS)    : 18.92
REMARK      3 DATA CUTOFF (SIGMA(F)) : 0.000
REMARK      3 COMPLETENESS FOR RANGE (%) : 99.3
REMARK      3 NUMBER OF REFLECTIONS : 6801
REMARK      3
REMARK      3 FIT TO DATA USED IN REFINEMENT.
```

```
shreyakathuria@Shreyaa: ~/L × + v
HETATM 656 0 HOH A 7 -48.895 -18.702 15.563 1.00 27.48 0
HETATM 657 0 HOH A 8 -21.393 -0.854 17.811 1.00 24.13 0
HETATM 658 0 HOH A 9 -32.124 5.776 0.506 1.00 29.82 0
HETATM 659 0 HOH A 10 -46.186 -13.792 6.539 1.00 23.52 0
HETATM 660 0 HOH A 11 -29.575 -1.996 25.245 1.00 28.23 0
HETATM 661 0 HOH A 12 -45.642 -11.444 19.694 1.00 25.61 0
HETATM 662 0 HOH A 13 -49.384 -20.064 17.570 1.00 29.28 0
HETATM 663 0 HOH A 14 -30.137 -4.552 3.329 1.00 27.31 0
HETATM 664 0 HOH A 15 -42.693 -7.945 15.244 1.00 19.76 0
HETATM 665 0 HOH A 16 -35.906 -28.174 5.866 1.00 31.98 0
HETATM 666 0 HOH A 17 -44.171 -7.687 17.621 1.00 22.18 0
HETATM 667 0 HOH A 18 -47.265 -12.454 21.564 1.00 29.40 0
HETATM 668 0 HOH A 19 -36.430 3.094 -3.026 1.00 25.02 0
HETATM 669 0 HOH A 20 -29.553 -5.969 12.150 1.00 34.06 0
HETATM 670 0 HOH A 21 -42.686 -4.398 27.240 1.00 25.96 0
HETATM 671 0 HOH A 22 -43.889 -9.382 19.695 1.00 29.00 0
HETATM 672 0 HOH A 23 -43.476 -6.477 -2.563 1.00 30.73 0
HETATM 673 0 HOH A 24 -28.999 3.283 21.951 1.00 26.71 0
HETATM 674 0 HOH A 25 -50.516 -11.430 14.190 1.00 25.35 0
HETATM 675 0 HOH A 26 -27.306 5.304 20.576 1.00 30.44 0
HETATM 676 0 HOH A 27 -48.424 -14.440 -0.286 1.00 61.67 0
HETATM 677 0 HOH A 28 -43.808 -10.099 7.884 1.00 28.89 0
HETATM 678 0 HOH A 29 -35.566 -5.200 24.698 1.00 29.22 0
HETATM 679 0 HOH A 30 -34.679 -7.575 -4.768 1.00 25.20 0
HETATM 680 0 HOH A 31 -41.964 -17.506 25.641 1.00 37.16 0
HETATM 681 0 HOH A 32 -34.312 -2.922 25.191 1.00 31.83 0
HETATM 682 0 HOH A 33 -51.606 -11.651 21.823 1.00 29.90 0
HETATM 683 0 HOH A 34 -32.561 -16.311 28.119 1.00 50.80 0
HETATM 684 0 HOH A 35 -34.469 -16.004 9.163 1.00 24.01 0
HETATM 685 0 HOH A 36 -31.585 -23.210 8.833 1.00 26.89 0
HETATM 686 0 HOH A 37 -49.015 -19.802 20.176 1.00 31.69 0
HETATM 687 0 HOH A 38 -30.973 -14.980 5.105 1.00 43.06 0
HETATM 688 0 HOH A 39 -47.022 -17.146 11.346 1.00 28.11 0
HETATM 689 0 HOH A 40 -30.833 -7.743 14.123 1.00 34.35 0
HETATM 690 0 HOH A 41 -25.168 6.080 14.148 1.00 49.89 0
HETATM 691 0 HOH A 42 -51.167 -14.258 13.359 1.00 47.34 0
CONNECT 45 288
CONNECT 288 45
CONNECT 382 456
CONNECT 456 382
CONNECT 476 641
CONNECT 641 476
CONNECT 644 646 648
CONNECT 645 647 648
CONNECT 646 644 649
CONNECT 647 645 649
CONNECT 648 644 645
CONNECT 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 × + v
shreyakathuria@Shreyaa:~/Labsession3$ sed -n '/^ATOM/ { /ARG/!p }' protein.pdb
ATOM      1  N   TRP A 172      -39.136 -21.997  24.415  1.00 34.43      N
ATOM      2  CA  TRP A 172      -40.108 -20.907  24.729  1.00 34.28      C
ATOM      3  C   TRP A 172      -41.403 -21.065  23.944  1.00 33.46      C
ATOM      4  O   TRP A 172      -41.385 -21.496  22.789  1.00 33.48      O
ATOM      5  CB  TRP A 172      -39.506 -19.534  24.418  1.00 35.12      C
ATOM      6  CG  TRP A 172      -38.161 -19.292  25.025  1.00 36.34      C
ATOM      7  CD1 TRP A 172      -37.773 -19.568  26.306  1.00 37.69      C
ATOM      8  CD2 TRP A 172      -37.032 -18.693  24.384  1.00 37.47      C
ATOM      9  NE1 TRP A 172      -36.465 -19.190  26.497  1.00 37.97      N
ATOM     10  CE2 TRP A 172      -35.985 -18.650  25.334  1.00 37.83      C
ATOM     11  CE3 TRP A 172      -36.799 -18.192  23.097  1.00 37.57      C
ATOM     12  CZ2 TRP A 172      -34.725 -18.128  25.037  1.00 37.51      C
ATOM     13  CZ3 TRP A 172      -35.545 -17.671  22.802  1.00 37.85      C
ATOM     14  CH2 TRP A 172      -34.523 -17.646  23.769  1.00 37.43      C
ATOM     15  N   LYS A 173      -42.516 -20.697  24.576  1.00 32.18      N
ATOM     16  CA  LYS A 173      -43.842 -20.728  23.949  1.00 31.37      C
ATOM     17  C   LYS A 173      -44.028 -19.604  22.914  1.00 29.85      C
ATOM     18  O   LYS A 173      -44.831 -19.725  21.976  1.00 30.15      O
ATOM     19  CB  LYS A 173      -44.935 -20.645  25.024  1.00 31.31      C
ATOM     20  CG  LYS A 173      -46.343 -20.964  24.519  1.00 32.53      C
ATOM     21  CD  LYS A 173      -47.425 -20.459  25.479  1.00 32.89      C
ATOM     22  CE  LYS A 173      -48.818 -20.684  24.901  1.00 33.96      C
ATOM     23  NZ  LYS A 173      -49.893 -20.189  25.806  1.00 34.66      N
ATOM     24  N   GLU A 174      -43.280 -18.518  23.090  1.00 27.67      N
ATOM     25  CA  GLU A 174      -43.337 -17.366  22.191  1.00 25.77      C
ATOM     26  C   GLU A 174      -41.922 -17.014  21.728  1.00 23.54      C
ATOM     27  O   GLU A 174      -41.381 -15.977  22.138  1.00 23.23      O
ATOM     28  CB  GLU A 174      -43.933 -16.148  22.913  1.00 25.76      C
ATOM     29  CG  GLU A 174      -45.376 -16.258  23.359  1.00 26.89      C
ATOM     30  CD  GLU A 174      -45.777 -15.061  24.206  1.00 27.42      C
ATOM     31  OE1 GLU A 174      -46.102 -14.001  23.639  1.00 29.42      O
ATOM     32  OE2 GLU A 174      -45.756 -15.182  25.445  1.00 30.63      O
ATOM     33  N   PRO A 175      -41.313 -17.867  20.872  1.00 21.55      N
ATOM     34  CA  PRO A 175      -39.891 -17.705  20.564  1.00 20.10      C
ATOM     35  C   PRO A 175      -39.565 -16.385  19.866  1.00 18.58      C
ATOM     36  O   PRO A 175      -38.520 -15.781  20.142  1.00 18.18      O
ATOM     37  CB  PRO A 175      -39.594 -18.893  19.632  1.00 20.52      C
ATOM     38  CG  PRO A 175      -40.909 -19.247  19.043  1.00 19.77      C
ATOM     39  CD  PRO A 175      -41.896 -19.015  20.148  1.00 21.28      C
ATOM     40  N   CYS A 176      -40.455 -15.942  18.986  1.00 16.73      N
ATOM     41  CA  CYS A 176      -40.212 -14.710  18.226  1.00 16.80      C
ATOM     42  C   CYS A 176      -40.222 -13.501  19.159  1.00 16.78      C
ATOM     43  O   CYS A 176      -39.363 -12.626  19.053  1.00 16.20      O
ATOM     44  CB  CYS A 176      -41.244 -14.528  17.116  1.00 16.50      C
ATOM     45  SG  CYS A 176      -40.885 -13.084  16.044  1.00 15.20      S
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     65  N   GLU A 179      -37.088 -13.386  21.187  1.00 19.21      N
ATOM     66  CA  GLU A 179      -35.993 -12.950  20.327  1.00 19.54      C
ATOM     67  C   GLU A 179      -36.080 -11.449  20.043  1.00 19.29      C
```



shreyakathuria@Shreyaa: ~/L X



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.

```
shreyakathuria@Shreyaa: ~/L × + v
shreyakathuria@Shreyaa:~/Labsession3$ awk '/^ATOM/ && $5 == "A" {count[$4]++} END {
for (res in count) print count[res], res}' protein.pdb
28 GLY
37 CYS
32 LEU
14 THR
18 GLN
42 PRO
32 ILE
8 MET
40 ASN
48 TYR
45 LYS
16 ASP
36 SER
22 PHE
10 HIS
81 GLU
55 ARG
42 TRP
15 ALA
21 VAL
shreyakathuria@Shreyaa:~/Labsession3$
```

Used gemini for loop.

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

```
shreyakathuria@Shreyaa: ~/L  ×  +  v
shreyakathuria@Shreyaa:~/Labsession3$ awk '/^ATOM/ {print $3,"$4","$5}' protein.pdb
b
N,TRP,A
CA,TRP,A
C,TRP,A
O,TRP,A
CB,TRP,A
CG,TRP,A
CD1,TRP,A
CD2,TRP,A
NE1,TRP,A
CE2,TRP,A
CE3,TRP,A
CZ2,TRP,A
CZ3,TRP,A
CH2,TRP,A
N,LYS,A
CA,LYS,A
C,LYS,A
O,LYS,A
CB,LYS,A
CG,LYS,A
CD,LYS,A
CE,LYS,A
NZ,LYS,A
N,GLU,A
CA,GLU,A
C,GLU,A
O,GLU,A
CB,GLU,A
CG,GLU,A
CD,GLU,A
OE1,GLU,A
OE2,GLU,A
N,PRO,A
CA,PRO,A
C,PRO,A
```

```
shreyakathuria@Shreyaa: ~/L × + v
NH1, ARG, A
NH2, ARG, A
N, GLY, A
CA, GLY, A
C, GLY, A
O, GLY, A
N, ASP, A
CA, ASP, A
C, ASP, A
O, ASP, A
CB, ASP, A
CG, ASP, A
OD1, ASP, A
OD2, ASP, A
N, PRO, A
CA, PRO, A
C, PRO, A
O, PRO, A
CB, PRO, A
CG, PRO, A
CD, PRO, A
N, ASN, A
CA, ASN, A
C, ASN, A
O, ASN, A
CB, ASN, A
CG, ASN, A
OD1, ASN, A
ND2, ASN, A
N, CYS, A
CA, CYS, A
C, CYS, A
O, CYS, A
CB, CYS, A
SG, CYS, A
OXT, CYS, A
shreyakathuria@Shreyaa: ~/Labsession3$
```

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


```
shreyakathuria@Shreyaa: ~/L x + v
shreyakathuria@Shreyaa:~/Labsession3$ awk '{print toupper($0)}' protein.fasta
>SEQ1|HOMO_SAPIENS|CLOCK_PROTEIN
MTEYKLVVVGAGCCGKSALT IQLINHF GFVDEYDPTIEDSYRKQVVIDGETCLLDILDTAG

>SEQ2|MUS_MUSCULUS|PER_PROTEIN
MSDDEEVQPSLLTKDGRVLQVLQSLFFGKNSDQLQSLNQLQDLLTAAQNMYSSST

>SEQ3|DROSOPHILA_MELANOGASTER|TIM_PROTEIN
MADQLTEEQIAEFKEAFSLFDKDGDTCTKELGTVMRSCCQNPTEAELQDMINEVDADGNGQ

>SEQ4|DANIO_RERIO|BMAL_PROTEIN
MLSRVAVCGTSGTGKSTLSRIIAQYFKKTDVVLVGP SGAGKTTISKLLEQLDYLNQKNV

>SEQ5|ARABIDOPSIS_THALIANA|LHY_PROTEIN
MSEQNGVVVDGSGIKVLVTGNKCDPQQRVTSQPVLQAGLDRIFGVIRD LGSSS

>SEQ6|SACCHAROMYCES_CEREVISIAE|CYC_PROTEIN
MTEYKLVVVG DVGKSTIVKQMQNH FVDEYDPTIEDSYRKQVVIDGETCLLDILDTAG

>SEQ7|CAENORHABDITIS_ELEGANS|CLK_PROTEIN
MADSQRRLQN VINKAAGKSTLLPVDGDKILVTTGGQVVQSNVLEAMKELLQ

>SEQ8|GALLUS_GALLUS|CRY_PROTEIN
MPGSGYVVRAGTVAGQLRIMNNKVVVVDLGAGKTTLLQSVIEMLKLLGEKGTA

>SEQ9|ESCHERICHIA_COLI|RECA_PROTEIN
MNVQLKKQLKDLPGVIVLGP PGAGKGTQFVS YVLNQLPQYLKKIDVYRTKGF

>SEQ10|XENOPUS_LAEVIS|REV-ERB_PROTEIN
MADEEKLP PGWEKRMSRSGRVYFNHITNASQWERPSGNSSSGSLS
shreyakathuria@Shreyaa:~/Labsession3$ |
```

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

```
shreyakathuria@Shreyaa:~$ cd Labsession3/
shreyakathuria@Shreyaa:~/Labsession3$ awk '/^>/ {if (seqlen > maxlen) {maxlen=seqlen; maxseq=header} seqlen=0; header=$0
; next} {seqlen+=length($0)} END {if (seqlen > maxlen) {maxseq=header} print maxseq}' protein.fasta
>seq3|Drosophila_melanogaster|TIM_protein
shreyakathuria@Shreyaa:~/Labsession3$
```

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.

```
shreyakathuria@Shreyaa: ~/L  ×  +  v
shreyakathuria@Shreyaa:~/Labsession3$ awk '/^ATOM/ {print $4}' protein.pdb | sort -u
ALA
ARG
ASN
ASP
CYS
GLN
GLU
GLY
HIS
ILE
LEU
LYS
MET
PHE
PRO
SER
THR
TRP
TYR
VAL
shreyakathuria@Shreyaa:~/Labsession3$ |
```

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

```
shreyakathuria@Shreyaa: ~/L  ×  +  v  -  □  ×
shreyakathuria@Shreyaa:~/Labsession3$ awk '/^ATOM/ {print substr($0,22,1)}'
protein.pdb | sort -u
A
shreyakathuria@Shreyaa:~/Labsession3$ |
```

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.

```
shreyakathuria@Shreyaa: ~/L  ×  +  v  -  □  ×
shreyakathuria@Shreyaa:~/Labsession3$ sed '/^>/d' clock_gene.fasta | grep -o "A" | wc -l
35332
shreyakathuria@Shreyaa:~/Labsession3$ sed '/^>/d' clock_gene.fasta | grep -o "T" | wc -l
39197
shreyakathuria@Shreyaa:~/Labsession3$ sed '/^>/d' clock_gene.fasta | grep -o "G" | wc -l
23471
shreyakathuria@Shreyaa:~/Labsession3$ sed '/^>/d' clock_gene.fasta | grep -o "C" | wc -l
21007
shreyakathuria@Shreyaa:~/Labsession3$ |
```