Class#9

Zainab Ashir

```
# Importing the dataset from PDB Statistics and adding them
CSV_file <- "https://bioboot.github.io/bimm143_W23/class-material/PDB.csv"
pdb= read.csv(CSV_file, row.names = 1)
```

Q1: What percentage of structures in the PDB are solved by X-Ray and Electron Microscopy?

It is 87.5% of X-ray for proteins only, and 5.4% of EM for proteins only.

Q2: What proportion of structures in the PDB are protein?

They are a totral of 174,642

Q3: Type HIV in the PDB website search box on the home page and determine how many HIV-1 protease structures are in the current PDB?

They are 204,352 structures of HIV-1

Q4: Water molecules normally have 3 atoms. Why do we see just one atom per water molecule in this structure?

Because this visualization 3D structures uses one atom to represent water just for having simplicity

There is a critical "conserved" water molecule in the binding site. Can you identify this water molecule? What residue number does this water molecule have?

yes I can, it has residue number of 308.

```
# Call Bio3D
library(bio3d)
```

```
pdb_1 <- read.pdb("1hsg")
```

```
  Note: Accessing on-line PDB file
```

```
pdb_1
```

```
 Call:  read.pdb(file = "1hsg")

   Total Models#: 1
     Total Atoms#: 1686,  XYZs#: 5058  Chains#: 2  (values: A B)

     Protein Atoms#: 1514  (residues/Calpha atoms#: 198)
     Nucleic acid Atoms#: 0  (residues/phosphate atoms#: 0)

     Non-protein/nucleic Atoms#: 172  (residues: 128)
     Non-protein/nucleic resid values: [ HOH (127), MK1 (1) ]

   Protein sequence:
      PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIGGFIKVRQYD
      QILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFPQITLWQRPLVTIKIGGQLKE
      ALLDTGADDTVLEEMSLPGRWKPKMIGGIGGFIKVRQYDQILIEICGHKAIGTVLVGPTP
      VNIIGRNLLTQIGCTLNF

+ attr: atom, xyz, seqres, helix, sheet,
        calpha, remark, call
```

Q7: How many amino acid residues are there in this pdb object?

##They are 198.

Q8: Name one of the two non-protein residues?

HOH and MK1

Q.9: How many protein chains are in this structure?

2 chains

```
attributes(pdb_1)
```

```
$names
[1] "atom"   "xyz"    "seqres" "helix"  "sheet"  "calpha" "remark" "call"

$class
[1] "pdb" "sse"
```

```
head(pdb_1$atom)
```

```
  type eleno elety  alt resid chain resno insert      x      y     z o     b
1 ATOM     1     N <NA>   PRO     A     1   <NA> 29.361 39.686 5.862 1 38.10
2 ATOM     2    CA <NA>   PRO     A     1   <NA> 30.307 38.663 5.319 1 40.62
3 ATOM     3     C <NA>   PRO     A     1   <NA> 29.760 38.071 4.022 1 42.64
4 ATOM     4     O <NA>   PRO     A     1   <NA> 28.600 38.302 3.676 1 43.40
5 ATOM     5    CB <NA>   PRO     A     1   <NA> 30.508 37.541 6.342 1 37.87
6 ATOM     6    CG <NA>   PRO     A     1   <NA> 29.296 37.591 7.162 1 38.40
  segid elesy charge
1  <NA>     N   <NA>
2  <NA>     C   <NA>
3  <NA>     C   <NA>
4  <NA>     O   <NA>
5  <NA>     C   <NA>
6  <NA>     C   <NA>
```

Q10. Which of the packages above is found only on BioConductor and not CRAN?

It would be "msa"

Q11. Which of the above packages is not found on BioConductor or CRAN?:

That would be "bio3d"

True or False? Functions from the devtools package can be used to install packages from GitHub and BitBucket?

That would be True

Q13. How many amino acids are in this sequence, i.e. how long is this sequence?

214 amino acids which makes its length 214

Q14: What do you note about this plot? Are the black and colored lines similar or different? Where do you think they differ most and why?

The black and colored lines are different beucase the black one represent different pdb than the red and freen ones, and the black one has the lowest spikes as if it is the baseline, and the colored ones show the actual observations. They differ since its probably the black ones have fluctuations ranging in low values, with its highes recording just slightly above 0.5, as fo rthe colored ones, thier fluctuations are showing high recordings that have reached 2.5-30.