Class 10

Renny Ng (A98061553)

```
read.csv("Data Export Summary.csv")
```

```
           Molecular.Type   X.ray     EM    NMR Multiple.methods Neutron Other
1          Protein (only) 161,663 12,592 12,337              200      74    32
2 Protein/Oligosaccharide   9,348  2,167     34                8       2     0
3              Protein/NA   8,404  3,924    286                7       0     0
4     Nucleic acid (only)   2,758    125  1,477               14       3     1
5                   Other     164      9     33                0       0     0
6  Oligosaccharide (only)      11      0      6                1       0     4
    Total
1 186,898
2  11,559
3  12,621
4   4,378
5     206
6      22
```

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

There are 215,684 structures in the PDB 182,348 structures are solved by X-Ray, and 18,817 structures are solved by EM Therefore, 93.26% of the structures are solved.

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

Of the 215,684 structures in the PDB, there are 186,898 proteins (only) structures. This accounts for 86.65% of all structures in the PDB.

> 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?

There are currently 4,284 structures.

Figure 1: Visualizing the HIV-1 protease structure

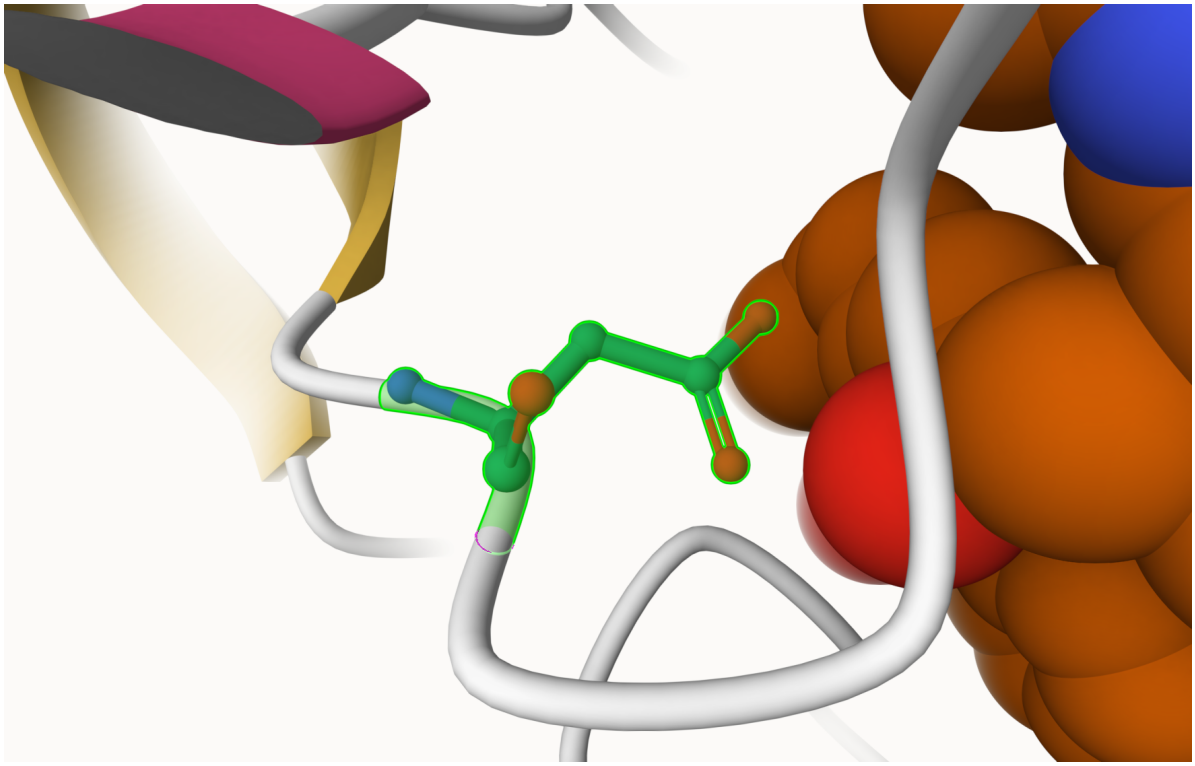


Figure 2: A closer look at the HIV-1 protease structure

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

The two hydrogen atoms are simply protons, so they do not show up.

> Q5: 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

This is HOH 332

> Q6: Generate and save a figure clearly showing the two distinct chains of HIV-protease along with the ligand. You might also consider showing the catalytic residues ASP 25 in each chain and the critical water (we recommend "Ball & Stick" for these side-chains). Add this figure to your Quarto document.

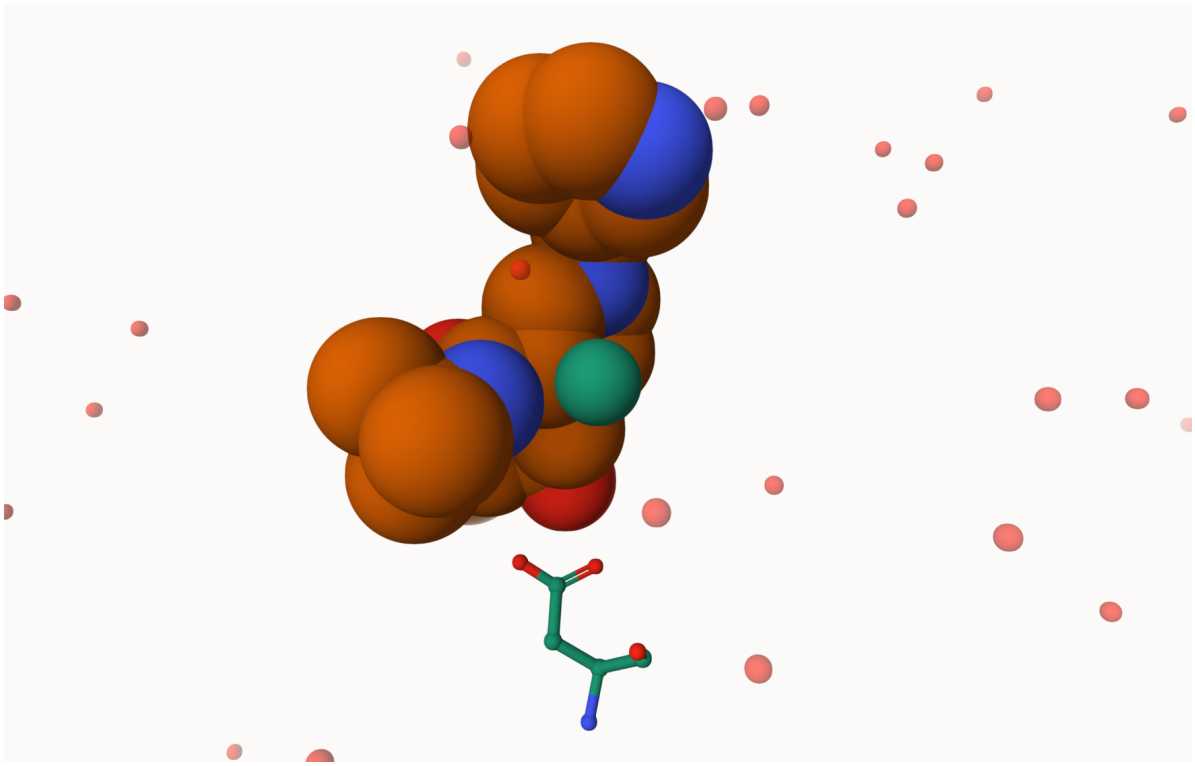


Figure 3: Distinct chains of the HIV-protease along with ligase; critcial water is shown in green

```
library(bio3d)
```

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

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

```
pdb
```

```
 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?

There are 198 amino acid residues.

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

Water is one of the non-protein residues.

Q9: How many protein chains are in this structure?

There are two chains in this structure.

```
adk <- read.pdb("6s36")
```

```
  Note: Accessing on-line PDB file
   PDB has ALT records, taking A only, rm.alt=TRUE
```

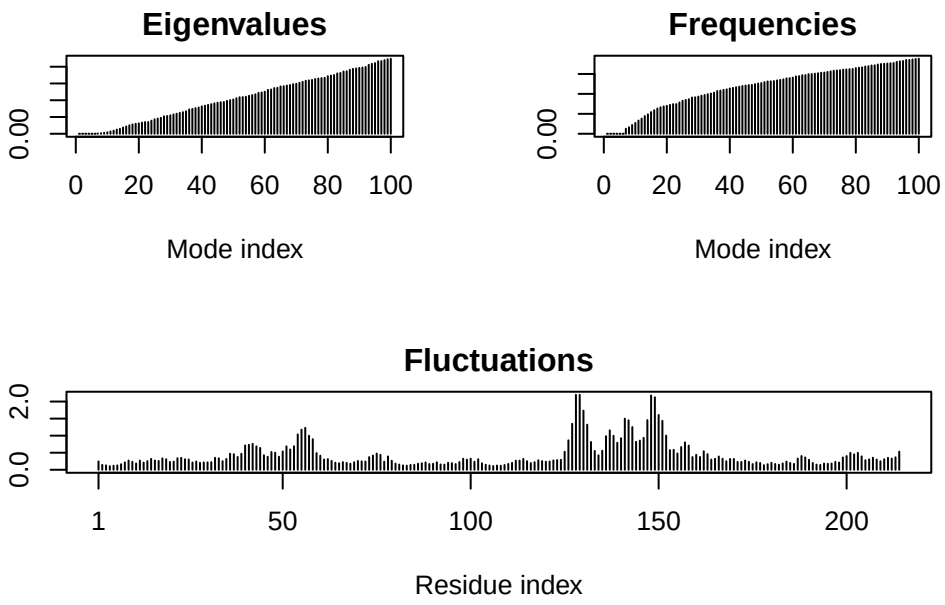
```
adk
```

```
 Call:  read.pdb(file = "6s36")

   Total Models#: 1
     Total Atoms#: 1898,  XYZs#: 5694  Chains#: 1  (values: A)

     Protein Atoms#: 1654  (residues/Calpha atoms#: 214)
     Nucleic acid Atoms#: 0  (residues/phosphate atoms#: 0)

     Non-protein/nucleic Atoms#: 244  (residues: 244)
     Non-protein/nucleic resid values: [ CL (3), HOH (238), MG (2), NA (1) ]

   Protein sequence:
      MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVT
      DELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDKI
      VGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIG
      YYSKEAEAGNTKYAKVDGTKPVAEVRADLEKILG

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

```
m <- nma(adk)
```

```
 Building Hessian...        Done in 0.013 seconds.
 Diagonalizing Hessian...   Done in 0.27 seconds.
```

```
plot(m)
```




```
mktrj(m, file="adk_m7.pdb")
```

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

Only "BiocManager" is found on BioConductor.

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

devtools is not found on BioConductor or CRAN.

Q12. True or False? Functions from the devtools package can be used to install packages from GitHub and BitBucket?

True

```
aa <- get.seq("1ake_A")
```

```
Warning in get.seq("1ake_A"): Removing existing file: seqs.fasta
```

```
Fetching... Please wait. Done.
```

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

There are 214 amino acid residues in this sequence.

> 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 dissimilar, in that they deviate in the magnitude of their fluctuations. Generally, the colored lines are higher in magnitude compared to the black lines. These results appear to indicate multiple conformational states for Adk.