class 09

Yinuo Song

PDB statistics

To read the csv file downloaded from PDB, we are going to use the command "`read.csv`"

```
pdb_stats <- read.csv('Data Export Summary.csv',row.names=1)
```

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

I need to sum all the elements of the x-ray column

```
as.numeric(gsub(",","",pdb_stats$X.ray))
```

```
[1] 154766   9083   8110   2664    163     11
```

```
n_xray <- sum(as.numeric(gsub(",","",pdb_stats$X.ray)))
n_em <- sum(as.numeric(gsub(",","",pdb_stats$EM)))
n_total <- sum(as.numeric(gsub(",","",pdb_stats$Total)))
```

```
(n_xray+n_em)/n_total
```

```
[1] 0.9299297
```

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

Q2 What proportion of structures in the PDB are protein?

```
total_protein <- as.numeric(gsub(",","",pdb_stats[1,7]))
```

Protein

```
total_protein/n_total
```

```
[1] 0.8681246
```

86.81 percentage of structures in the PDB are protein.

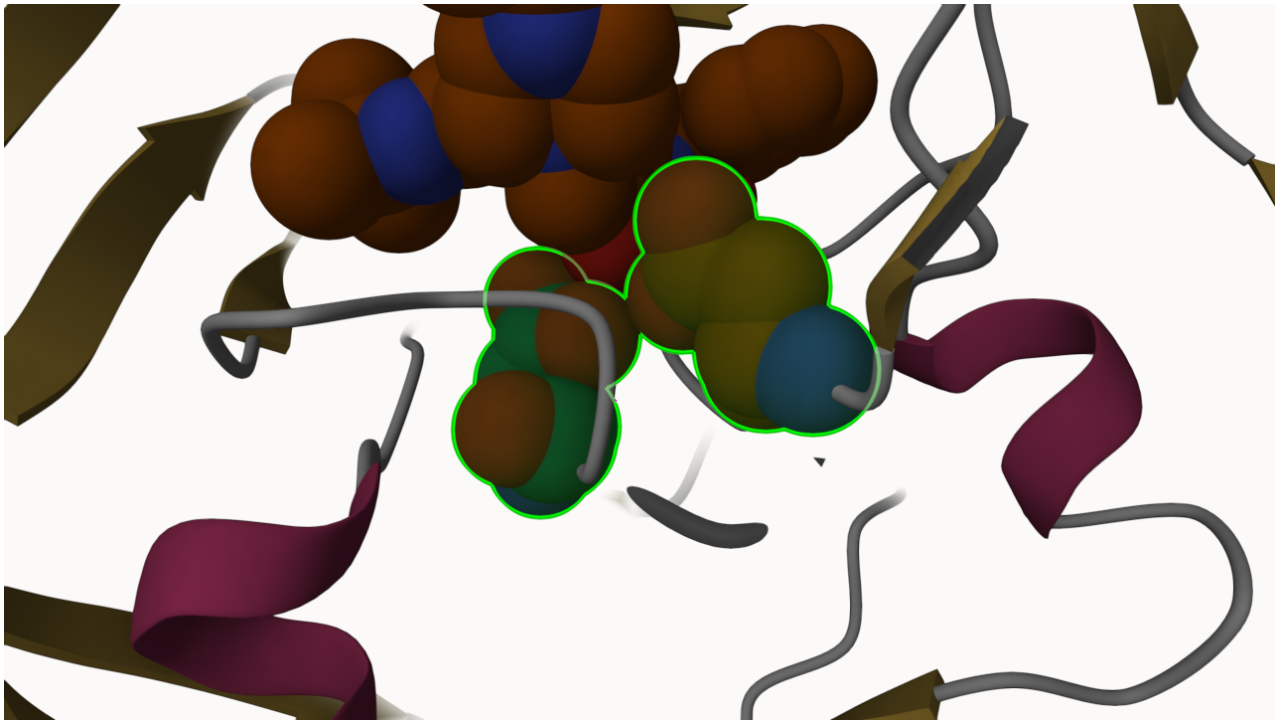
Q3 Type HIV in the PDB website search box on the home page and determine how many HIV-1 protease structures are in the current PDB?

There are 5 structures in the current PDB.

2. Visualizing the HIV-1 protease structure

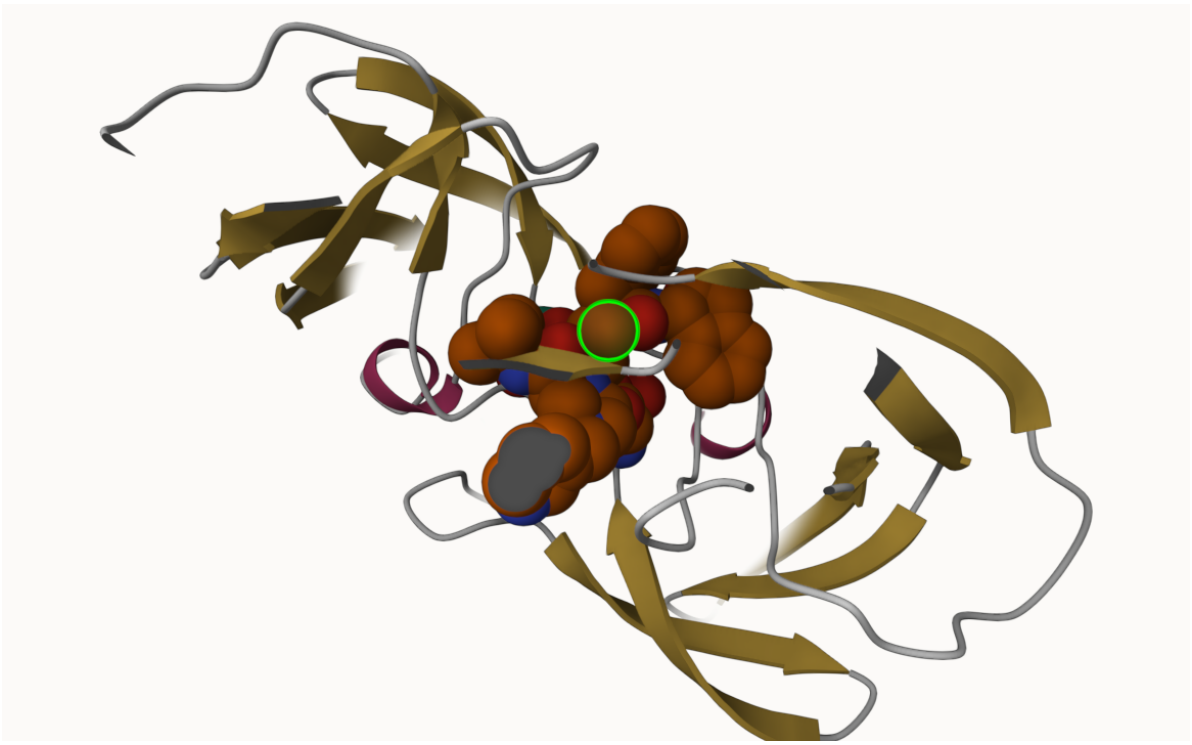
figure 1

figure 2(Asp 25)



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



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

Because the structure of a water molecule compared to an amino acid structure is too small.

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

The water molecule highlighted in figure 3 is HOH 308.

3. Introduction to Bio3D in R

```
library(bio3d)
```

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

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

```
attributes(pdb)
```

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

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

```
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?

198

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

nucleic Atoms

Q9: How many protein chains are in this structure?

2

```
head(pdb$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>
```

Predicting functional motions of a single structure by NMA

```
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
```

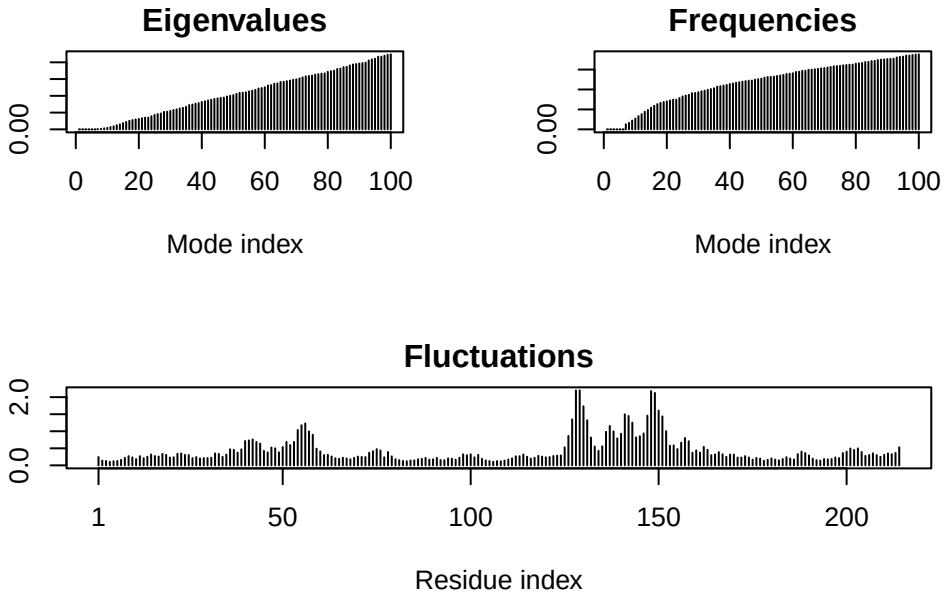
```
      VGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIG
      YYSKEAEAGNTKYAKVDGTKPVAEVRADLEKILG

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

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

```
 Building Hessian...        Done in 0.029 seconds.
 Diagonalizing Hessian...   Done in 0.347 seconds.
```

```
plot(m)
```




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

View the file in Mol Viewer