Class 09

Darby Patterson

We first need to read the file using `read.csv`.

```
pdb_data = read.csv("Data Export Summary.csv", row.names=1)
pdb_data
```

	X.ray	EM	NMR	Multiple.methods	Neutron	Other
Protein (only)	154,766	10,155	12,187	191	72	32
Protein/Oligosaccharide	9,083	1,802	32	7	1	0
Protein/NA	8,110	3,176	283	6	0	0
Nucleic acid (only)	2,664	94	1,450	12	2	1
Other	163	9	32	0	0	0
Oligosaccharide (only)	11	0	6	1	0	4

	Total
Protein (only)	177,403
Protein/Oligosaccharide	10,925
Protein/NA	11,575
Nucleic acid (only)	4,223
Other	204
Oligosaccharide (only)	22

We need to extract the commas from within the data set as well.

```
#sum(pdb_data$X.ray)
#as.numeric(pdb_data$X.ray)
nocommaxray = as.numeric(gsub(',','', pdb_data$X.ray))
nocommaEM = as.numeric(gsub(',','', pdb_data$EM))
```

```
n_xray= sum(nocommaxray)
n_EM = sum(nocommaEM)
n_Total = sum(as.numeric(gsub(',','', pdb_data$Total)))
```

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

```
percent_xray_EM= (n_xray + n_EM)/ n_Total
percent_xray_EM
```

```
[1] 0.9299297
```

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

```
proteins_total = as.numeric(gsub(',','', pdb_data[1,7]))

proteins_total / n_Total
```

```
[1] 0.8681246
```

Visualizing HIV-1 Protease Structure

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 were 5 structures with HIV-1 protease structures in the PDB.

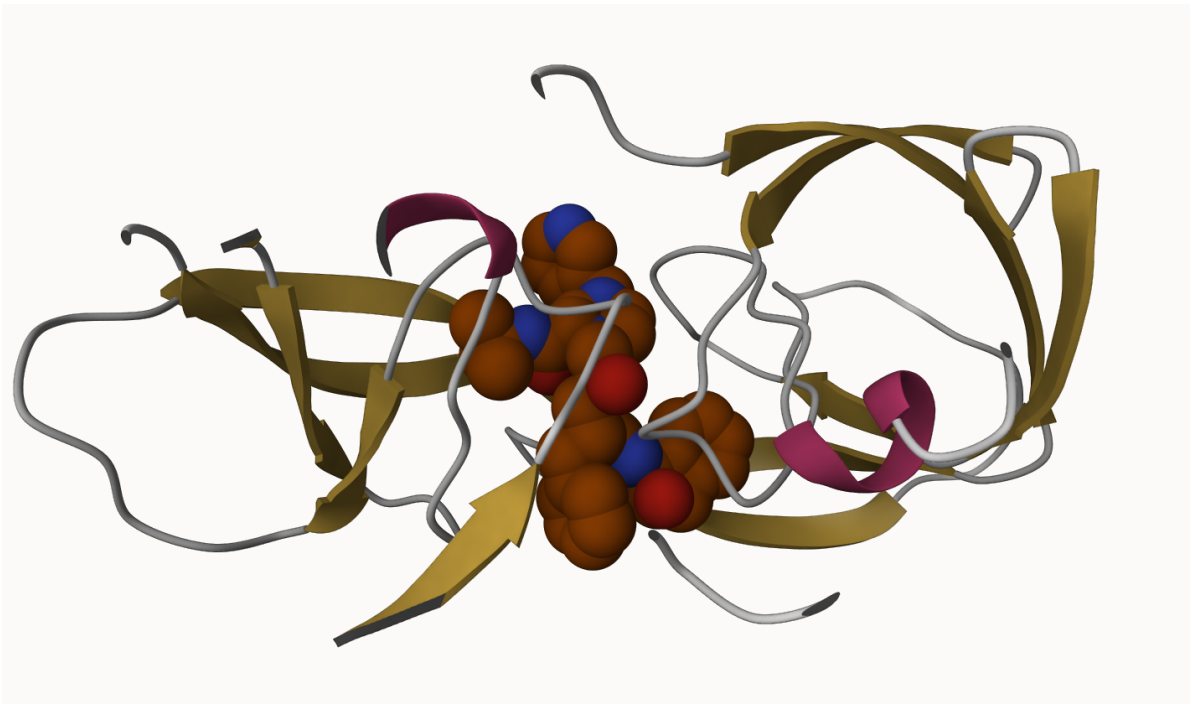
Q4: Water molecules normally have 3 atoms. Why do we see just one atom per water molecule in this structure?

For clarity, the visualization of water is simplified to one small molecule so that it doesn't get too confusing when viewing the molecule.

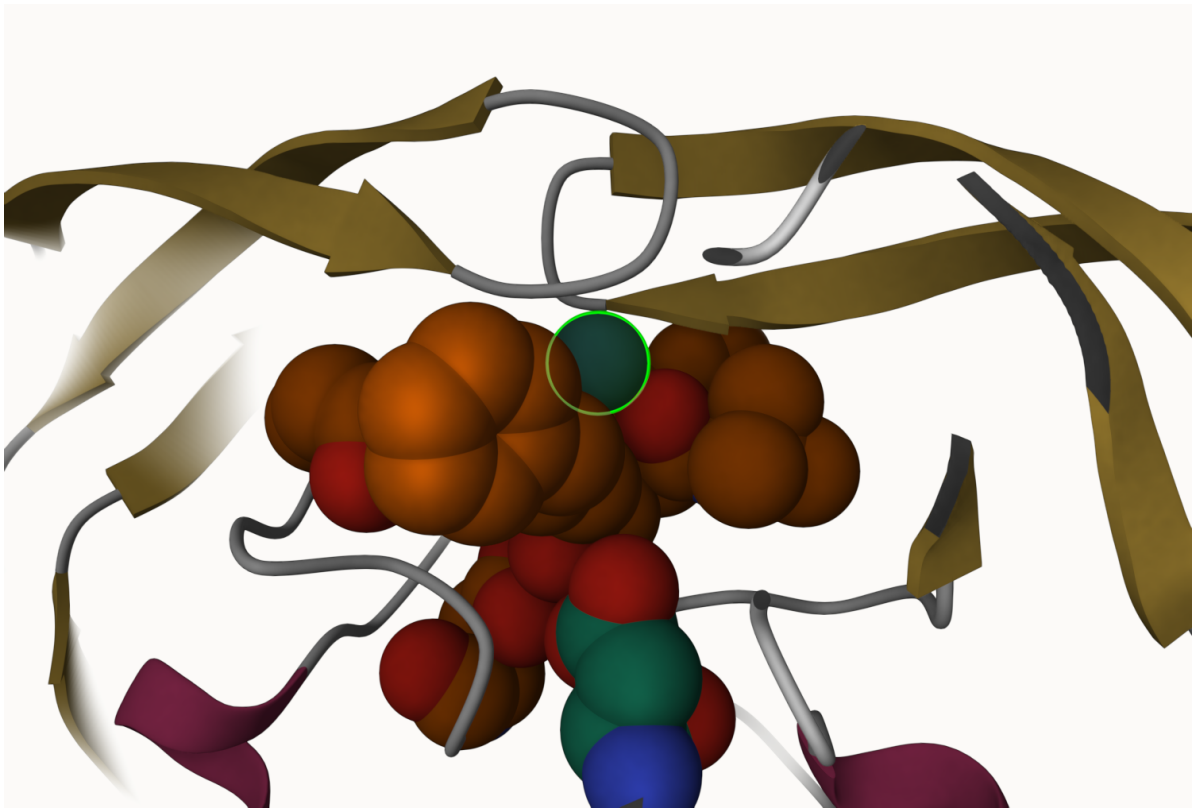
Q5: There is a critical "conserved" water molecule in the binding site. Can you identify this water molecule? What residue number does this water molecule have?

Residue number 308, it is conserved because it closely interacts with the ligands.

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.



Critical water molecule highlighted in green in the figure below.



3. Introduction to Bio3D in R

In your existing Rmarkdown document load the Bio3D package by typing in a new code chunk:

Hide

```
library(bio3d)
```

```
#install.packages('bio3d')
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?

198

```
attributes(pdb)
```

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

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

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

HOH (water), and MK1

Q9: How many protein chains are in this structure?

```
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

```
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) ]
```

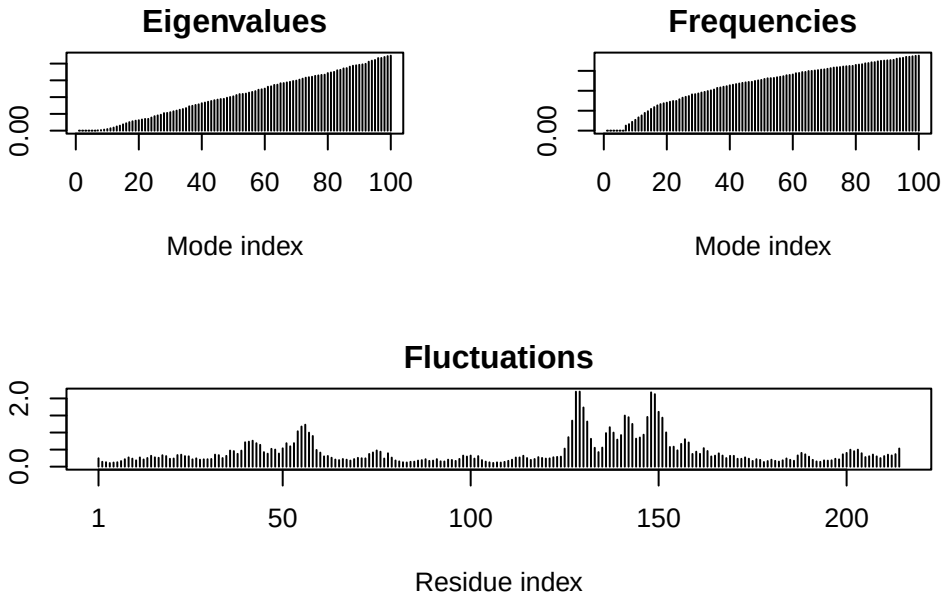
```
   Protein sequence:
      MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVT
      DELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDKI
      VGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIG
      YYSKEAEAGNTKYAKVDGTKPVAEVRADLEKILG

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

```
# Perform flexiblity prediction
m <- nma(adk)
```

```
 Building Hessian...        Done in 0.052 seconds.
 Diagonalizing Hessian...   Done in 1.358 seconds.
```

```
plot(m)
```




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