Class 09

PDB Exploration

We will start by load a csv file from PDB

To read this file we are going to use read.csv

```
library(readr)
pdb_stats <- read_csv("Data Export Summary.csv")
```

```
Rows: 6 Columns: 8
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (1): Molecular Type
dbl (3): Multiple methods, Neutron, Other
num (4): X-ray, EM, NMR, Total

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
pdb_stats
```

```
# A tibble: 6 x 8
  `Molecular Type`   `X-ray`    EM   NMR `Multiple methods` Neutron Other  Total
  <chr>                <dbl> <dbl> <dbl>              <dbl>   <dbl> <dbl>  <dbl>
1 Protein (only)      154766 10155 12187                191      72    32 177403
2 Protein/Oligosacc~    9083  1802    32                  7       1     0  10925
3 Protein/NA            8110  3176   283                  6       0     0  11575
4 Nucleic acid (onl~    2664    94  1450                 12       2     1   4223
5 Other                  163     9    32                  0       0     0    204
6 Oligosaccharide (~      11     0     6                  1       0     4     22
```

We are going to explore the data

```
pdb_stats$X.ray
```

```
Warning: Unknown or uninitialised column: `X.ray`.
```

```
NULL
```

We are going to use gsub to remove commas

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

```
Warning: Unknown or uninitialised column: `X.ray`.
```

```
numeric(0)
```

```
as.numeric(gsub(",", "", pdb_stats$EM))
```

```
[1] 10155  1802  3176    94     9     0
```

I use sum command to get sum

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

```
Warning: Unknown or uninitialised column: `X.ray`.
```

```
n_em = sum(as.numeric(gsub(",", "", pdb_stats$EM)))
```

```
n_total = sum(as.numeric(gsub(",", "", pdb_stats$Total)))
```

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

```
((n_xray + n_em)/n_total) * 100
```

```
[1] 7.455763
```

The percentage of structures solved by X-Ray and EM is 92.99%

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

```
n_proteino = as.numeric(gsub(",", "", pdb_stats["Protein (only)","Total"]))
```

```
Warning: NAs introduced by coercion
```

```
n_proteino/n_total
```

```
[1] NA
```

0.86 proportion of the structures are proteins

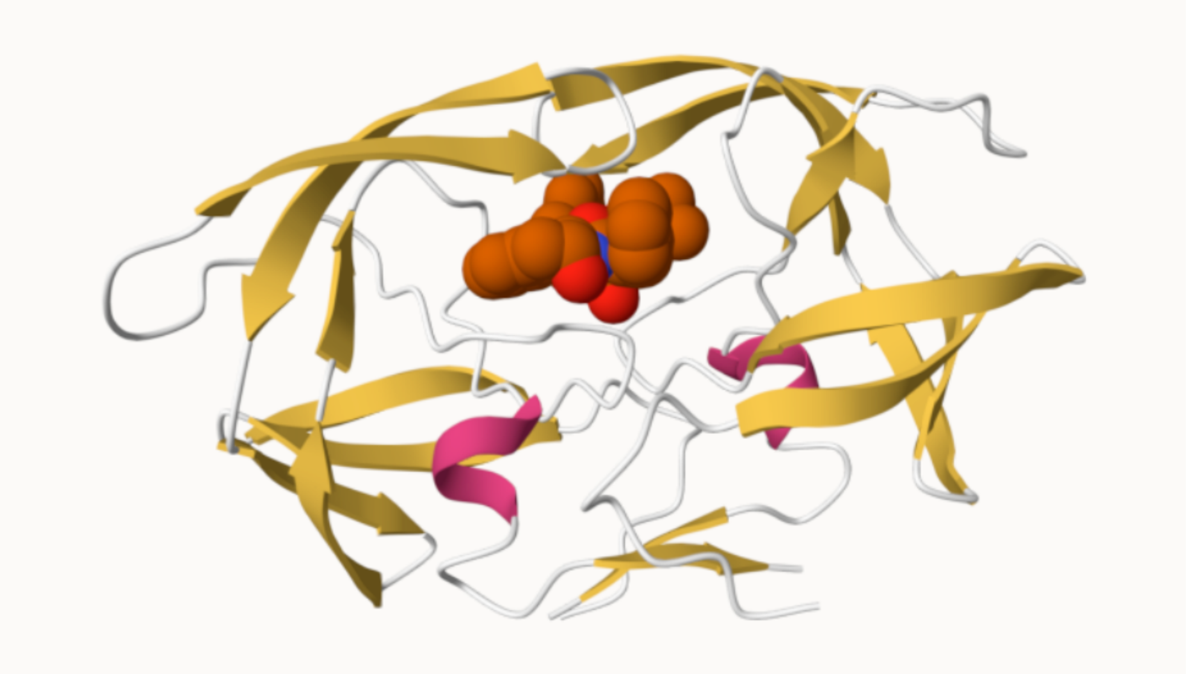
```
sum(as.numeric(gsub(",", "", pdb_stats$Total)))
```

```
[1] 204352
```

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?

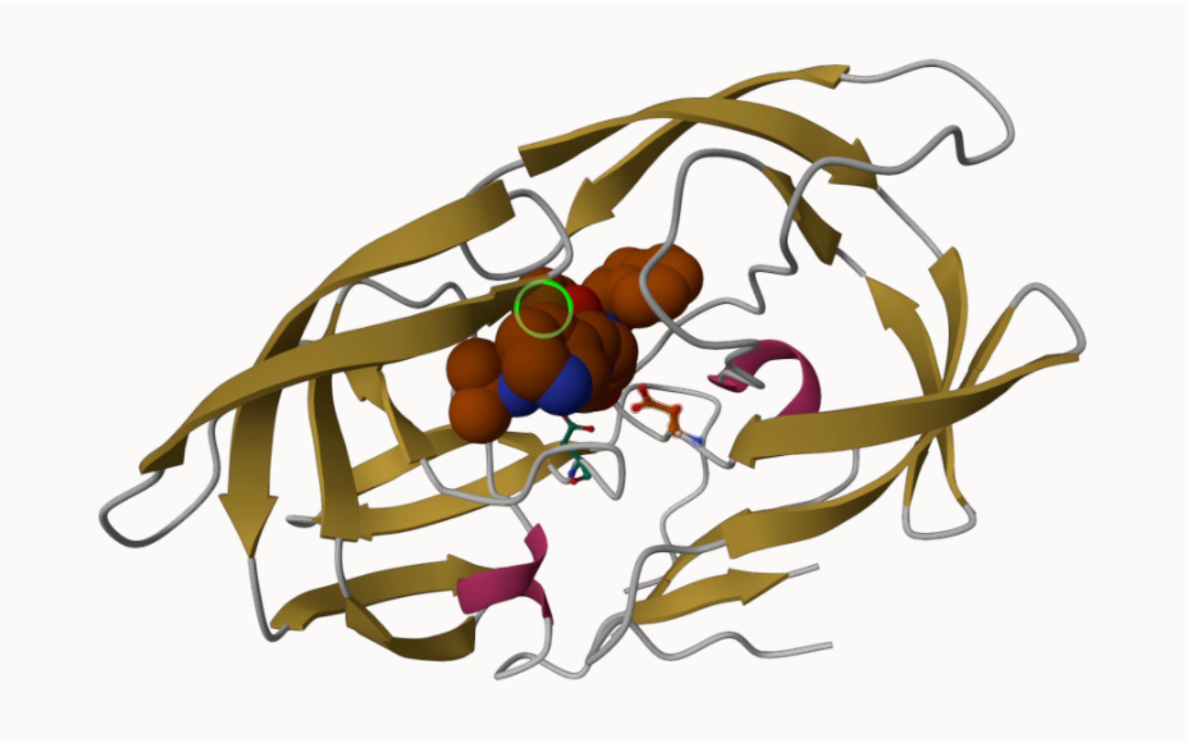
Too difficult

2. Visualizing the HIV-1 protease structure



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 308



Intro to BIO3d in R

```
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
```

```r
attributes(pdb)
```

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

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

```r
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

```r
adk = read.pdb("6s36")
```

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

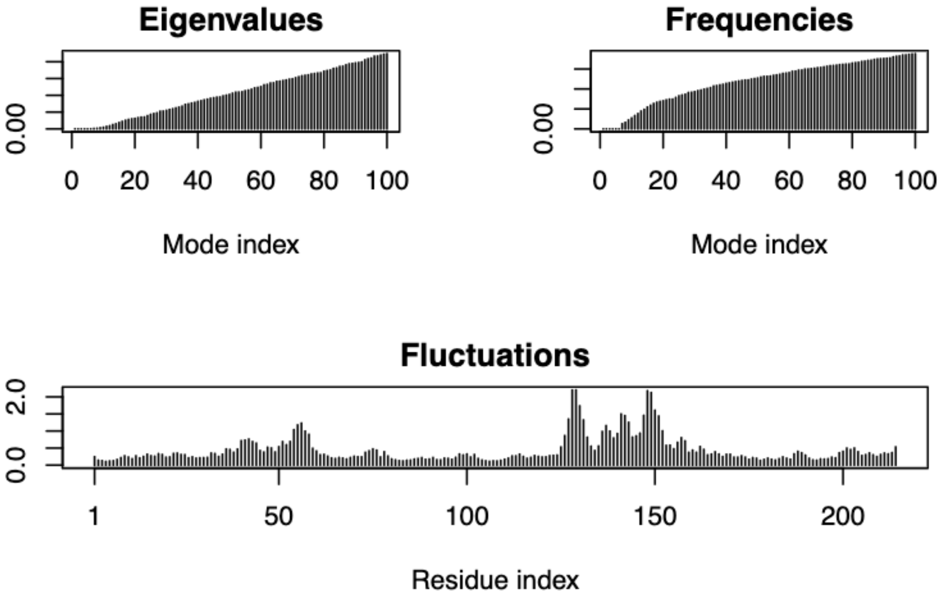
```r
adk
```

```
 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
```

```r
m = nma(adk)
```

```
 Building Hessian...        Done in 0.046 seconds.
 Diagonalizing Hessian...   Done in 0.499 seconds.
```

```r
plot(m)
```



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