Class 9

AUTHOR
Charlie Rezanka (A15837296)

PBD Statistics

```
data <- read.csv("Data Export Summary.csv", row.names=1)
head(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
```

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

```
#create a function with X as an imput
sum_comma <- function(data) {
  #Substitute comma and convert to numeric
  sum(as.numeric(gsub(",","", data)))
}
```

For X Ray

```
round(sum_comma(data$X.ray)/sum_comma(data$Total), 3)*100
```

```
[1] 85.5
```

For EM

```
round(sum_comma(data$EM)/sum_comma(data$Total), 3)*100
```

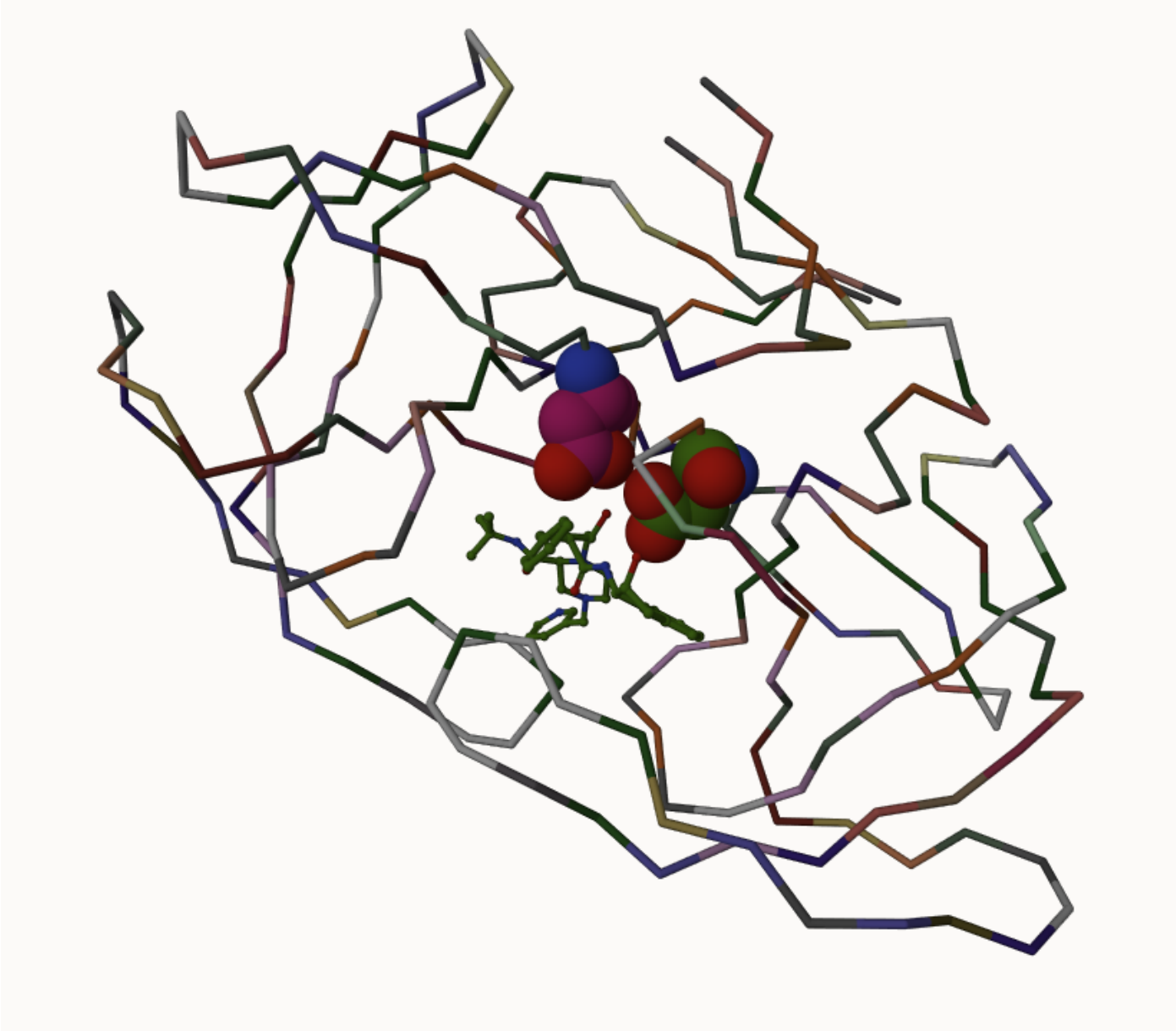
```
[1] 7.5
```

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

```
round(sum_comma(data$Total[1])/sum_comma(data$Total), 3)*100
```

```
[1] 86.8
```

> Q3: Insert your HIV image.



HIV-PR structure from MERK with critical residues highlighted. Protein structure is color coded to residue.

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

Hydrogen molecules are too small and abundant to be efficiently displayed

> 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 binds to to ASP25 on each chain

#Working with Structures 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
```

```
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>
```

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

There are 190 residues.

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

Water (shown as HOH)

> Q9: How many protein chains are in this structure?

There are 2 chains.

Predicting Movements of a Structure

First, read the ADK structure

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

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

```
adk
```

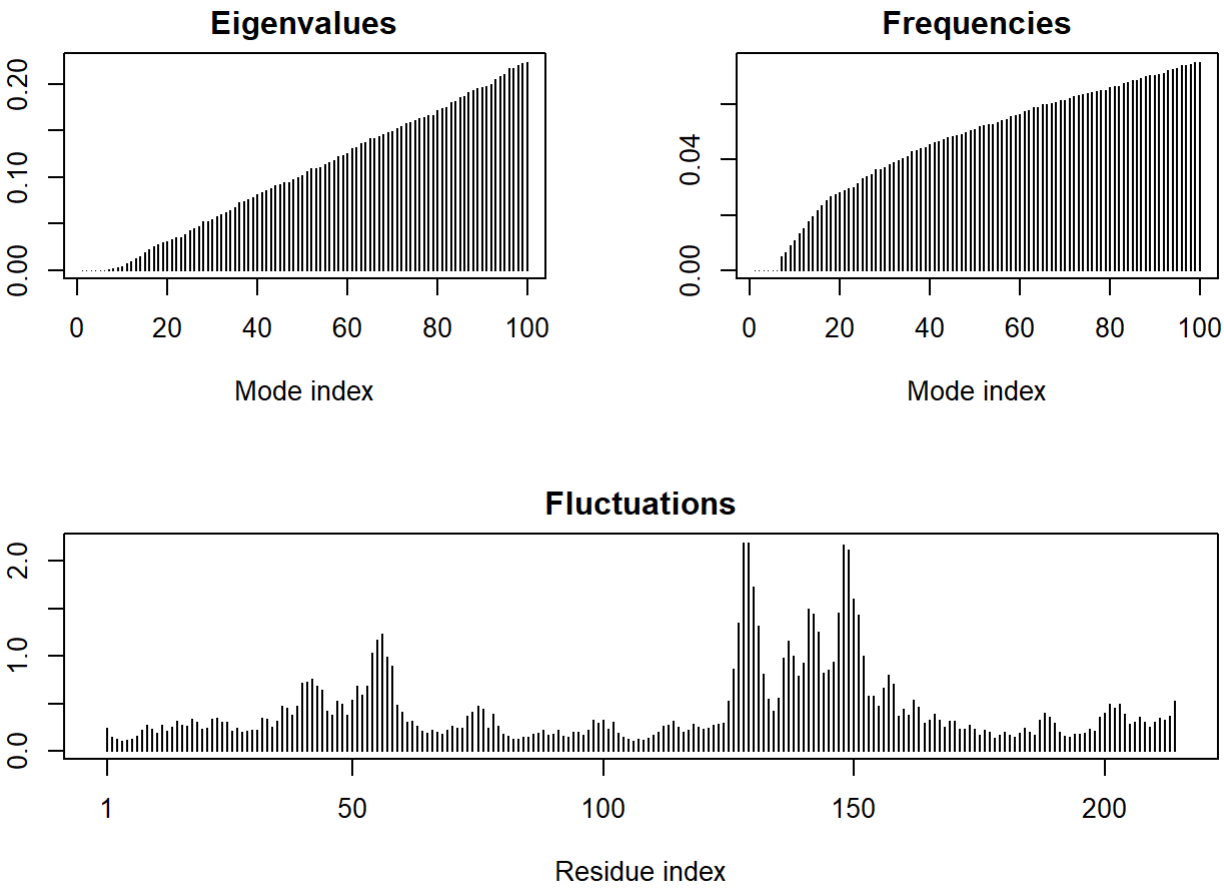
```
 Call:  read.pdb(file = "6s36")

    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 a prediction of flexibility with a technique called NMA (normal mode analysis), which is meant to predict flexibility and motion.

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

```
Building Hessian...        Done in 0.1 seconds.
Diagonalizing Hessian...   Done in 0.81 seconds.
```

```
plot(m)
```



Write out a “movie” (trajectory) of the motion for viewing in MOlstar

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