Class 09: Structural Bioinformatics

AUTHOR
Hyeseung (Frankie) Son PID: A16025601

Introduction to the RCSB Protein Data Bank (PDB)

```
db <- read.csv("Data Export Summary.csv")
db
```

```
           Molecular.Type  X.ray    EM   NMR Multiple.methods Neutron Other
1          Protein (only) 154766 10155 12187              191      72    32
2 Protein/Oligosaccharide   9083  1802    32                7       1     0
3              Protein/NA   8110  3176   283                6       0     0
4     Nucleic acid (only)   2664    94  1450               12       2     1
5                   Other    163     9    32                0       0     0
6  Oligosaccharide (only)     11     0     6                1       0     4
   Total
1 177403
2  10925
3  11575
4   4223
5    204
6     22
```

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

```
xray.tot <- sum(as.numeric(gsub(",", "", db$X.ray)))
xray.tot
```

```
[1] 174797
```

```
em.tot <- sum(as.numeric(gsub(",", "", db$EM)))
em.tot
```

```
[1] 15236
```

We found the sum of each column. Let's create a working snippet and a new function

```
#I will work with `x` as the input

sum_comma <- function(x) {
  # Substitue the comma and convert to numeric
  sum(as.numeric(gsub(",", "", x)))
```

```
}
```

For X ray: We'll round the answer to two decimal places

```
round (sum_comma(db$X.ray) / sum_comma(db$Total), 2)
```

```
[1] 0.86
```

86% of structures are solved by X-ray

For EM: we'll round the answer to 2 decimal places as well.

```
round( sum_comma(db$EM)/ sum_comma(db$Total), 2)
```

```
[1] 0.07
```

7% of structures are solved by EM.

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

```
round (sum_comma(db$Total[1]) / sum_comma(db$Total), 2)
```

```
[1] 0.87
```

87% of structures in the PDB file are protein.

Visualizing the HIV-1 protease structure

> Q3. insert this image

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

The resolution is too low on the struture to see H atoms. We need a resolution of below 1 Angstroms to visualize the H atoms.

> 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

HOH308

Working with Structures in R

```
library(bio3d)
```

```
Warning: package 'bio3d' was built under R version 4.2.3
```

```
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 residues

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

HOH127

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

2 protein chains, (A,B)

```
attributes(pdb)
```

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

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

```
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>
```

REad an ADK structure

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

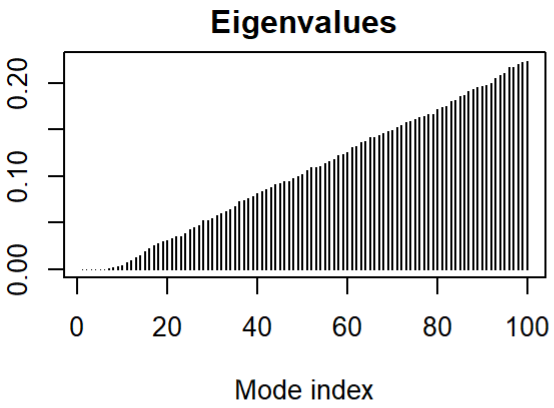
```
  Note: Accessing on-line PDB file
   PDB has ALT records, taking A only, rm.alt=TRUE
```

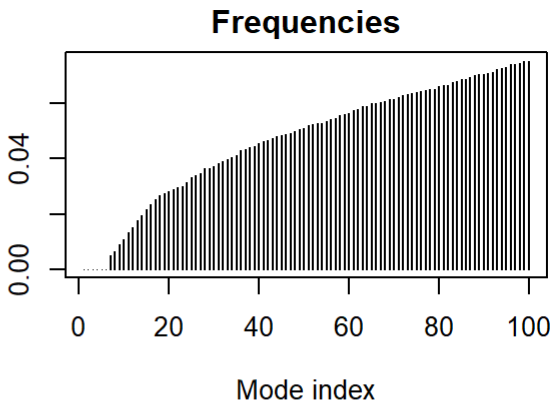
Perform a prediction of flexibility with a technique called NMA

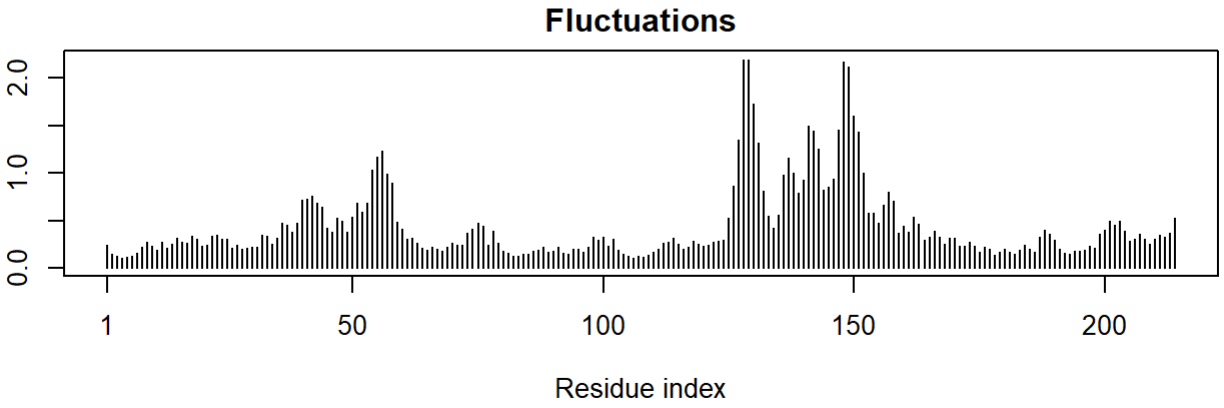
```
m <- nma(adk)
```

```
 Building Hessian...        Done in 0.06 seconds.
 Diagonalizing Hessian...   Done in 0.52 seconds.
```

```
plot(m)
```







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

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