Lab10: Structural Bioinformatics

Duy An Le (PID: A16400411)

The main repository of biomolecular structure is the PDB <www.rcsb.org>

Download the csv file of PDB contents by experimental method and molecular type.

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

sum.comma <- function(x){
  sum(as.numeric(gsub(",", "", x)))
}
```

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

```
sum_col <- apply(pdb_data, 2, sum.comma)
percent.xray.em <- (sum_col["X.ray"] + sum_col["EM"])/sum_col["Total"]
```

The percentage of X-ray and EM structures is 93.3435247.

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

```
percent.protein <- as.numeric(gsub(",","",pdb_data[1,"Total"]))/sum.comma(pdb_data$Total)
```

The proportion of structures that are protein 0.8658848

> 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 4445 structures.

Visualizing the HIV-1 protease structure

Mol* viewer is now everywhere https://molstar.org/viewer/

Insert an image from Mol* here:

1HSG (HIV)

Working with the bio3d package

```
library(bio3d)

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

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

```
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>
```

```
pdbseq(pdb)[25]
```

```
 25
"D"
```

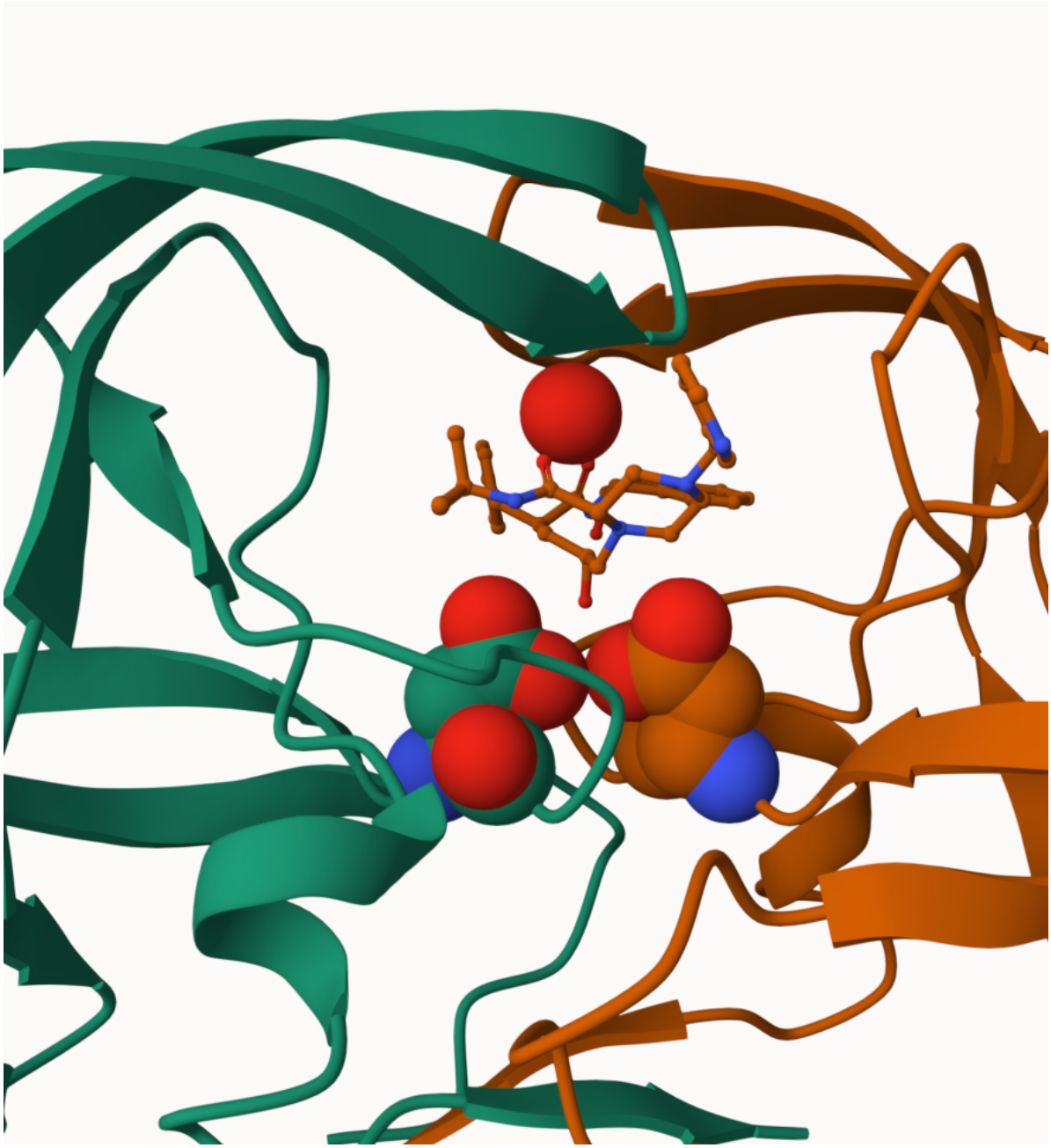


Figure 1: My first molecular image with aspartic acid and water highlighted.

Predicting function motions of a single structure

We can do bioinformatics prediction of functional motions (ie. flexibility, dynamics)

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

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

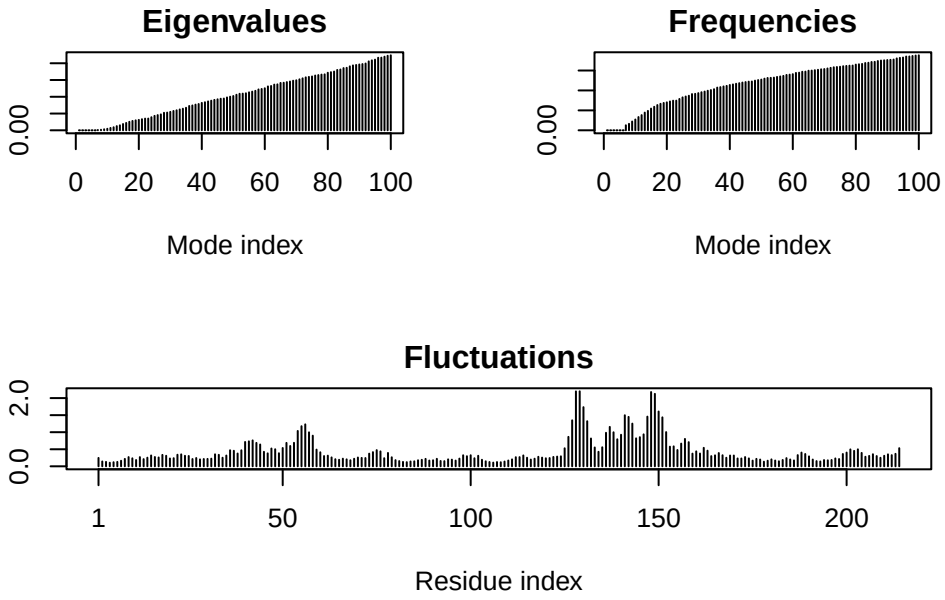
```
pdb
```

```
 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(pdb)
```

```
 Building Hessian...        Done in 0.038 seconds.
 Diagonalizing Hessian...   Done in 0.483 seconds.
```

```
plot(m)
```




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