Class 10: Structural Bioinformatics Pt 1

Bernice Lozada (A16297973)

What's in the PDB Database?

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

This database contains:

```
pdb_csv <- read.csv("Class10.csv", row.names = 1)
pdb_csv
```

```
                         X.ray     EM    NMR Multiple.methods Neutron Other
Protein (only)         161,663 12,592 12,337              200      74    32
Protein/Oligosaccharide  9,348  2,167     34                8       2     0
Protein/NA               8,404  3,924    286                7       0     0
Nucleic acid (only)      2,758    125  1,477               14       3     1
Other                      164      9     33                0       0     0
Oligosaccharide (only)      11      0      6                1       0     4
                          Total
Protein (only)          186,898
Protein/Oligosaccharide  11,559
Protein/NA               12,621
Nucleic acid (only)       4,378
Other                       206
Oligosaccharide (only)       22
```

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

First need to get rid of the commas - can you find a function to get rid of the commas?

```
x <- pdb_csv$X.ray
as.numeric(sub(",","",x))
```

```
[1] 161663   9348   8404   2758    164     11
```

Let's write a function for this and then use `apply()` to work on the entire table of data.

```
sumcomma <- function(x) {
  sum(as.numeric(sub(",","",x)))
}
apply(pdb_csv, 2, sumcomma)
```

```
        X.ray            EM           NMR Multiple.methods
       182348         18817         14173           230
      Neutron         Other         Total
           79            37        215684
```

```
(apply(pdb_csv, 2, sumcomma))/sumcomma(pdb_csv$Total) *100
```

```
        X.ray            EM           NMR Multiple.methods
  84.54405519    8.72433746    6.57118748    0.10663749
      Neutron         Other         Total
   0.03662766    0.01715473  100.00000000
```

84.5% of structures are solved by X-ray and 8.7% are solved by electron microscopy.

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

```
pdb_total <- sumcomma(pdb_csv$Total)
(as.numeric(sub(",","",pdb_csv[1, "Total"])))/pdb_total
```

```
[1] 0.8665362
```

0.86 of the structures in the PDB are protein.

> 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 about 0.0075118 structures in the PDB.

Visualizing the HIV-1 protease structure

I want to insert my image from Mol* here.

Here, I insert my image with the Asp in both Chain A and B as spacefill into the program.

Working with the bio3d package

```
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
```

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

```
 25
"D"
```

Predicting functional motions of a single structure

We can do a bioinformatics prediction of functional motions (i.e. flexibility/dynamics):

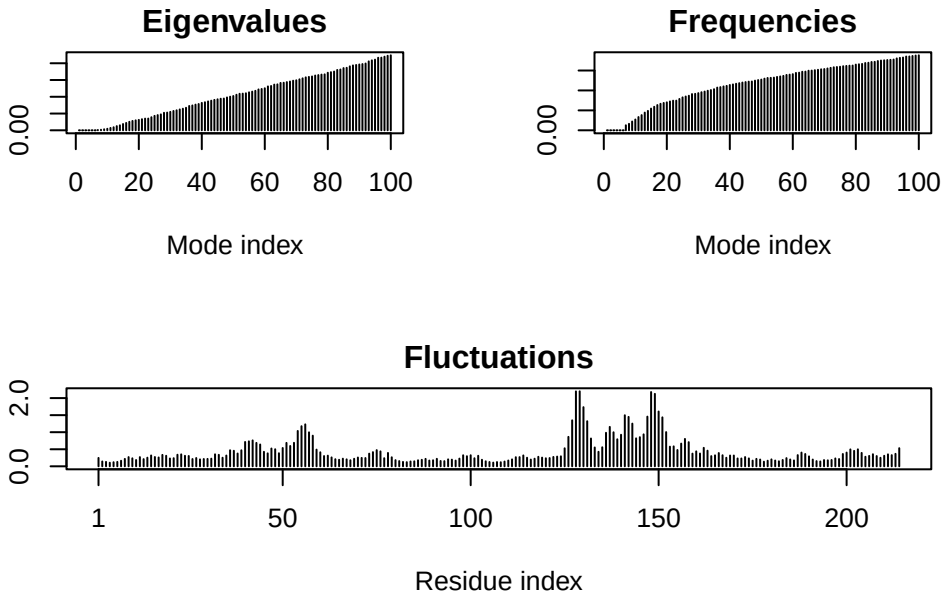
```
pdb_1 <- read.pdb("6s36")
```

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

```
# Normal mode analysis = NMA
m <- nma(pdb_1)
```

```
 Building Hessian...        Done in 0.063 seconds.
 Diagonalizing Hessian...   Done in 0.581 seconds.
```

```
plot(m)
```




```
# Writes a pdb file to working directory
mktrj(m, file="adk_m7.pdb")
```