# Intro to Structural Bioinformatics

Andy Hsu

## RCSB's Protein Data Base

Today we will explore the PDB, one of the oldest protein databases out there. Taking a quick look at the PDB statistics, we can see that 93.2% of all structures in the PDB are solved by Xray Cyrstallography or Electron Microscopy. We can also see that proteins constitute 86.7% of the PDB.

```
pdbstat <- read.csv("Data Export Summary.csv",row.names=1)
pdbstat[] <- lapply(pdbstat, function(x) as.numeric(gsub(",", "", x)))

sum(pdbstat[,1:2])/sum(pdbstat[,7])
```

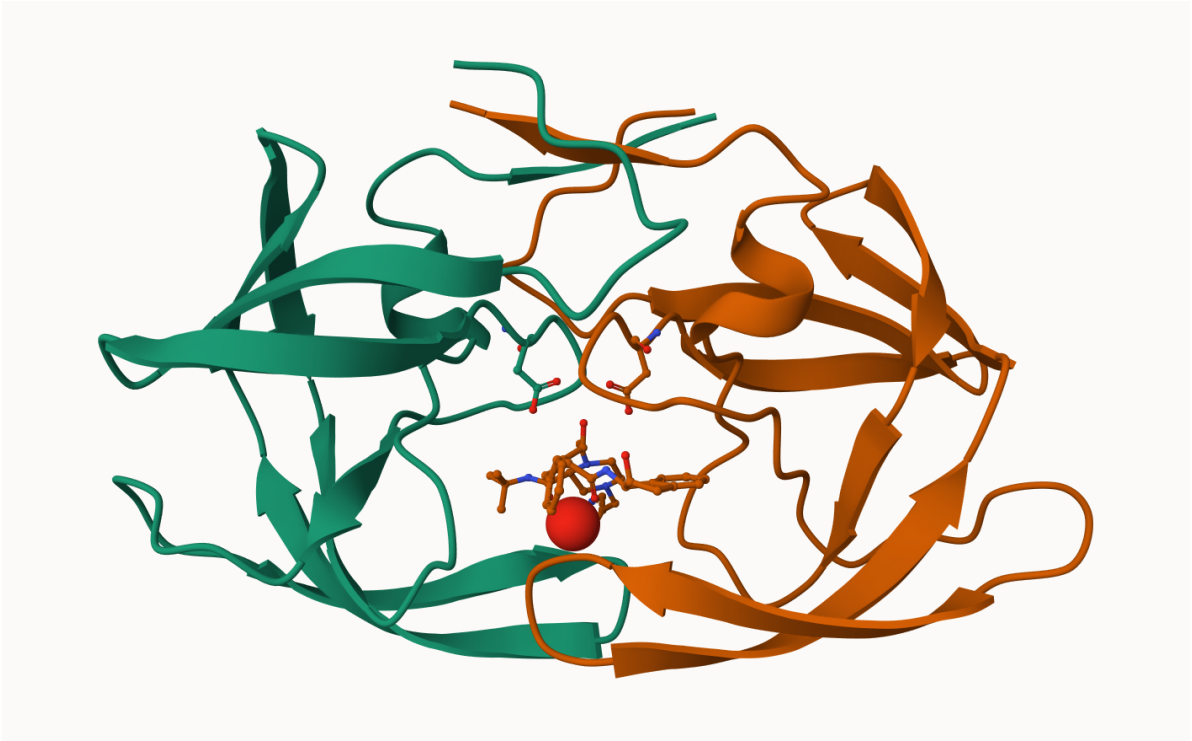
```
[1] 0.9315962
```

```
pdbstat[1,7]/sum(pdbstat[,7])
```

```
[1] 0.8667026
```

### Visualizing an HIV Protease Structure

Using Mol* (https://molstar.org), we can view the structure of a particular HIV protease, 1HSG. After performing some edits to the representation like displaying HOH 308 and both D25, we obtain a capture of the structure.



An important thing to note with the display of water molecules is that hydrogen atoms are smaller than the resolution of most imaging techniques, and so will not show in structure files.

### Using Bio3D to Read PDB Files

The package Bio3D can be used to read PDB files and tell us information about the structure within 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
```

Reading from the results, we find that this structure has 198 residues, 2 chains, and that HOH is one of the non-protein residues in the structure.

Finally, we can use Bio3D to predict the functional motions of a certain PDB structures. Let's use the 6S36 protein in this case.

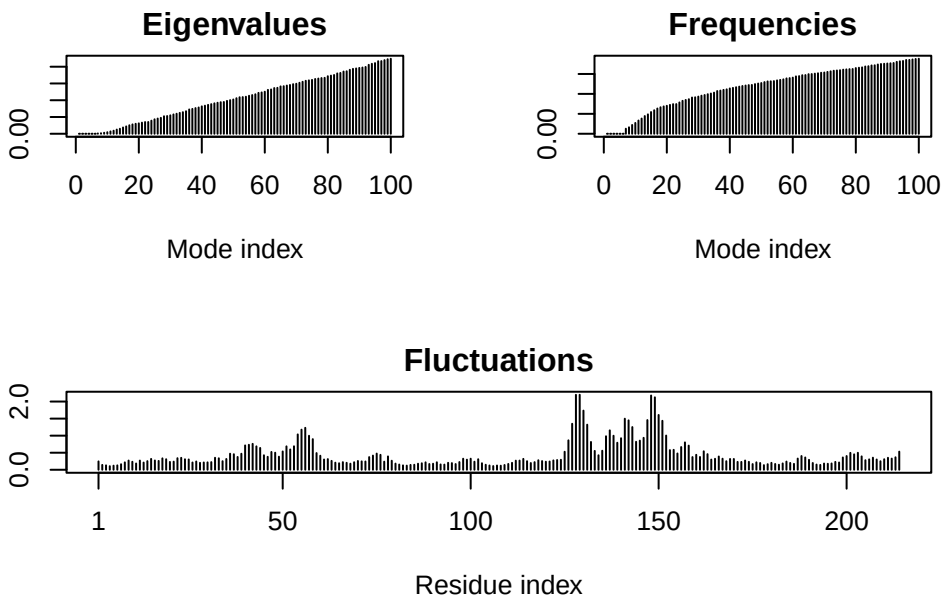
```
adk <- read.pdb("6s36")
```

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

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

```
Building Hessian...        Done in 0.08 seconds.
Diagonalizing Hessian...   Done in 0.93 seconds.
```

```
plot(m)
```







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