Structural Bioinformatics part 1

Jason (A15796973)

```
stats = read.csv("Data Export Summary.csv", row.names=1)
stats
```

```
                         X.ray     EM    NMR Multiple.methods Neutron Other
Protein (only)         158,844 11,759 12,296               197      73    32
Protein/Oligosaccharide  9,260  2,054     34                 8       1     0
Protein/NA               8,307  3,667    284                 7       0     0
Nucleic acid (only)      2,730    113  1,467                13       3     1
Other                      164      9     32                 0       0     0
Oligosaccharide (only)      11      0      6                 1       0     4
                         Total
Protein (only)         183,201
Protein/Oligosaccharide  11,357
Protein/NA               12,265
Nucleic acid (only)       4,327
Other                       205
Oligosaccharide (only)       22
```

```
rn.comma = function(x){

  as.numeric(gsub(",", "", x) )
}
```

```
pdbstats = apply(stats, 2, rn.comma)
pdbstats
```

```
     X.ray    EM   NMR Multiple.methods Neutron Other  Total
[1,] 158844 11759 12296              197      73    32 183201
[2,]   9260  2054    34                8       1     0  11357
```

```
[3,]   8307  3667   284                7       0      0  12265
[4,]   2730   113  1467               13       3      1   4327
[5,]    164     9    32                0       0      0    205
[6,]     11     0     6                1       0      4     22
```

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

```
xray = sum(pdbstats[,1])
em = sum(pdbstats[,2])
total = sum(pdbstats[,7])

(xray + em) / total
```

```
[1] 0.9315962
```

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

```
proteins = sum(pdbstats[1, 7])
proteins / total
```

```
[1] 0.8667026
```

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?

SKIPPED!

```
library(readr)
```

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

Hydrogen is too small for the current resolution, so only the oxygen is shown.

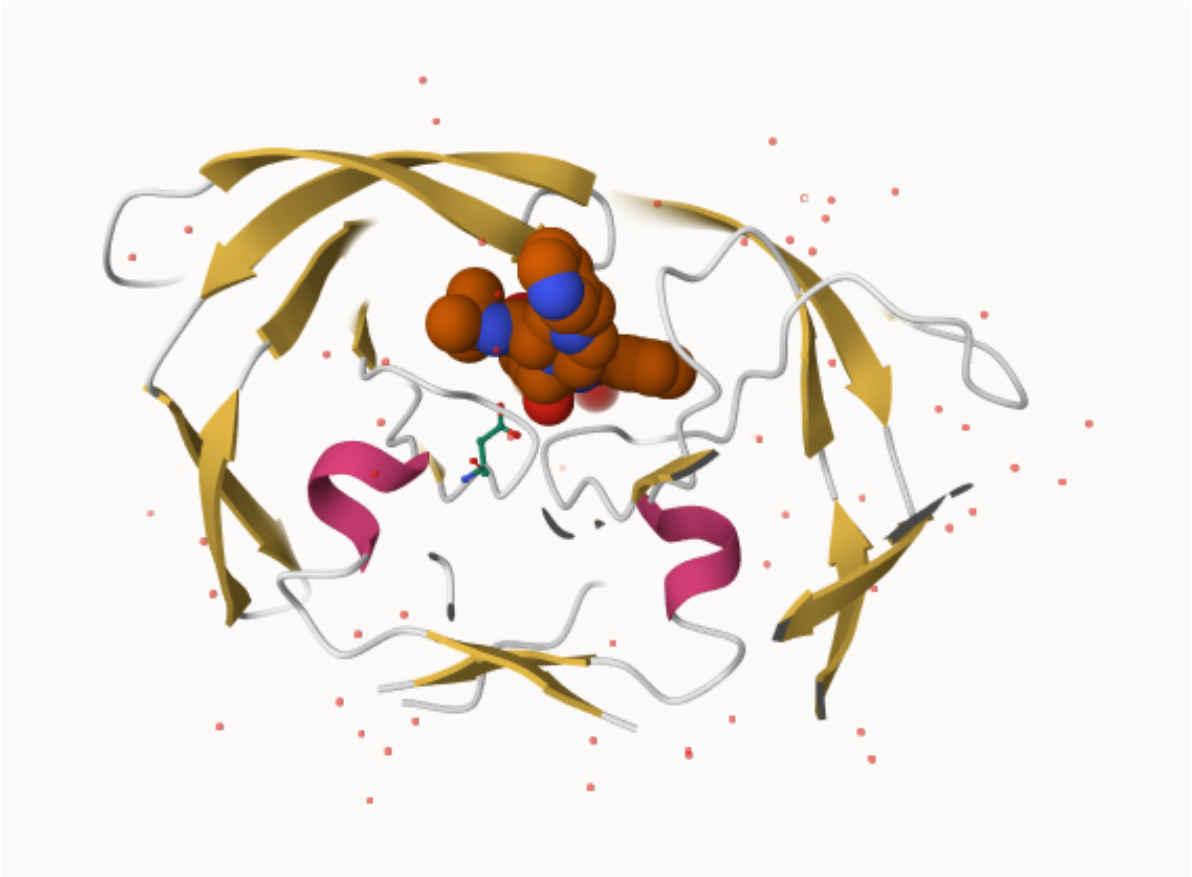
Q5: There is a critical "conserved" water molecule in the binding site. Can you identify this water molecule? What residue number does this water molecule have

HOH 308

Q6: Generate and save a figure clearly showing the two distinct chains of HIV-protease along with the ligand. You might also consider showing the catalytic residues ASP 25 in each chain

and the critical water (we recommend "Ball & Stick" for these side-chains). Add this figure to your Quarto document.

Here is the figure



The bio3d package for structural bioinformatics

```
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>
```

Predicting functional motions of a single structure

Let's finish today with a bioinformatics calculation to predict the functional motinos of a PDB structure

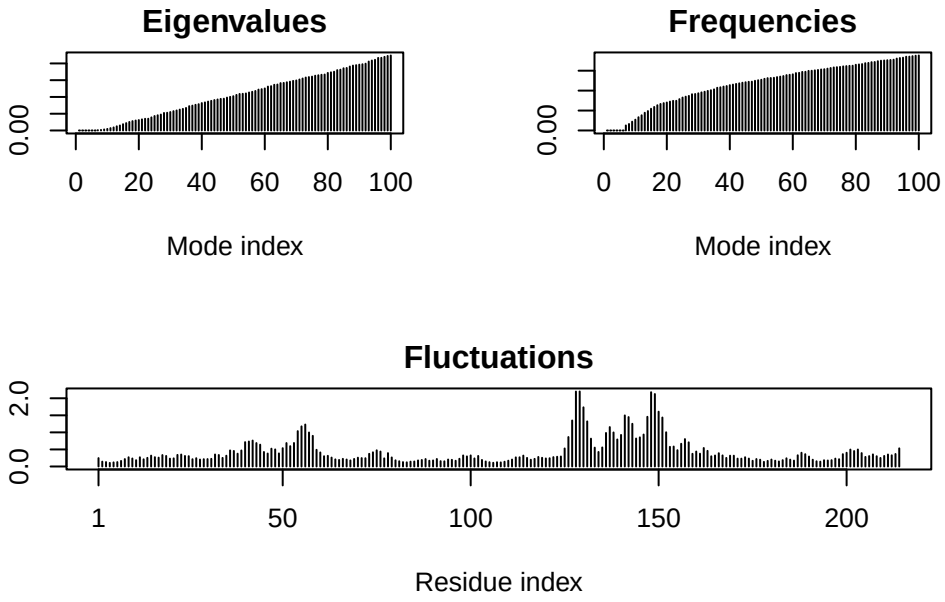
```
adk = read.pdb("6s36")
```

```
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 0.49 seconds.
```

```
plot(m)
```




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