Class 10 Structural Biochem and Alpha Fold

Matthew White

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

```
                         X.ray     EM    NMR Multiple.methods Neutron Other
Protein (only)         167,317 15,698 12,534              208      77    32
Protein/Oligosaccharide  9,645  2,639     34                8       2     0
Protein/NA               8,735  4,718    286                7       0     0
Nucleic acid (only)      2,869    138  1,507               14       3     1
Other                      170     10     33                0       0     0
Oligosaccharide (only)      11      0      6                1       0     4
                         Total
Protein (only)         195,866
Protein/Oligosaccharide  12,328
Protein/NA               13,746
Nucleic acid (only)       4,532
Other                       213
Oligosaccharide (only)       22
```

```
#pdbstats_numeric <- as.numeric(pdb_stats$c("X.ray", "EM", "NMR", "Total"))
```

```
x <- pdbstats$Total
#Write function that can remove commas then make character into a numeric
convert_comma_numbers <- function(x) {
  #remove commas
x <- gsub(",", "", x)

x <- as.numeric(x)

return(x)
}
```

```
convert_comma_numbers(x)
```

```
[1] 195866  12328  13746   4532    213     22
```

```
n.tot <- sum(convert_comma_numbers(pdbstats$Total))
n.tot
```

```
[1] 226707
```

Alternative mode of loading data that does not require the conversion function

```
#this readr package with read _ csv instead of .csv will read in values as numeric even with
library(readr)
pdbstats_numeric <- read_csv("Data Export Summary.csv")
```

```
Rows: 6 Columns: 8
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (1): Molecular Type
dbl (3): Multiple methods, Neutron, Other
num (4): X-ray, EM, NMR, Total

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
pdbstats_numeric
```

```
# A tibble: 6 x 8
  `Molecular Type`   `X-ray`    EM   NMR `Multiple methods` Neutron Other  Total
  <chr>                <dbl> <dbl> <dbl>              <dbl>   <dbl> <dbl>  <dbl>
1 Protein (only)      167317 15698 12534                208      77    32 195866
2 Protein/Oligosacc~    9645  2639    34                  8       2     0  12328
3 Protein/NA            8735  4718   286                  7       0     0  13746
4 Nucleic acid (onl~    2869   138  1507                 14       3     1   4532
5 Other                  170    10    33                  0       0     0    213
6 Oligosaccharide (~      11     0     6                  1       0     4     22
```

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

```
sum(pdbstats_numeric$`X-ray`)/sum(pdbstats_numeric$Total)
```

```
[1] 0.8325592
```

```
sum(pdbstats_numeric$EM)/sum(pdbstats_numeric$Total)
```

```
[1] 0.102348
```

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

```
pdbstats_numeric$Total
```

```
[1] 195866  12328  13746   4532    213     22
```

```
195866/sum(pdbstats_numeric$Total)
```

```
[1] 0.863961
```

```
#library(dplyr)
#pdbstats_numeric |>
 # filter(`Molecular Type`=="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?

4,563 diff structures!

##Using Mol*

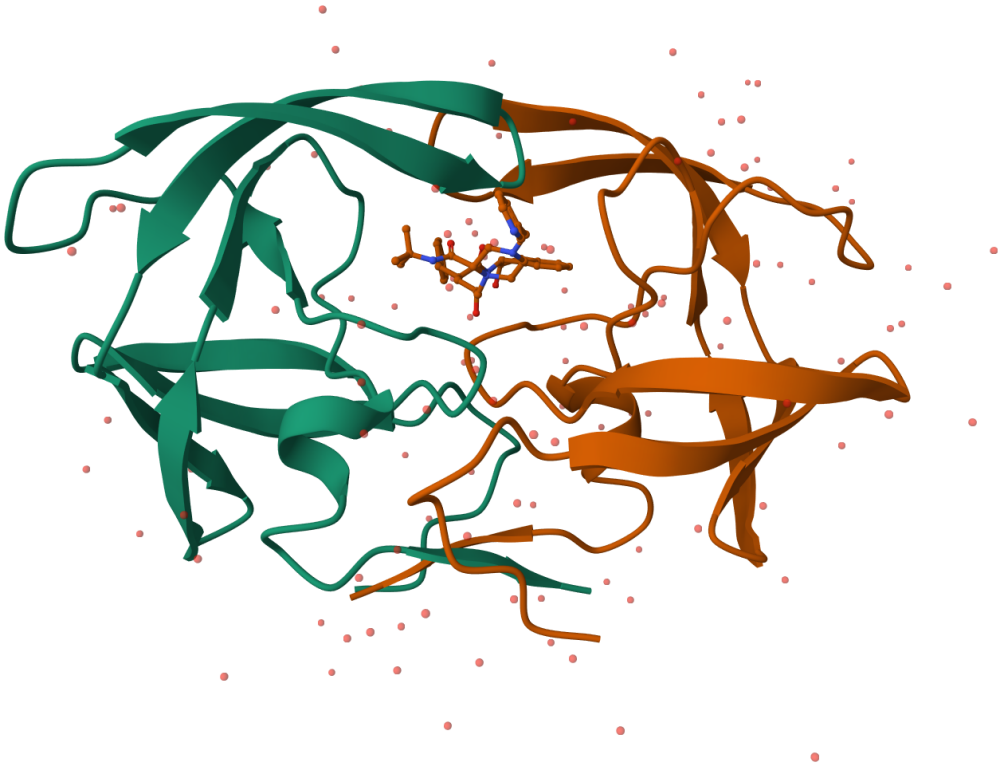


Figure 1: My first image from Mol-star

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

The structure does not show hydrogen anywhere since it is smaller than the 2 Angstrom resolution of the group's protein structure.

> 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

this is water 308 in the structure.

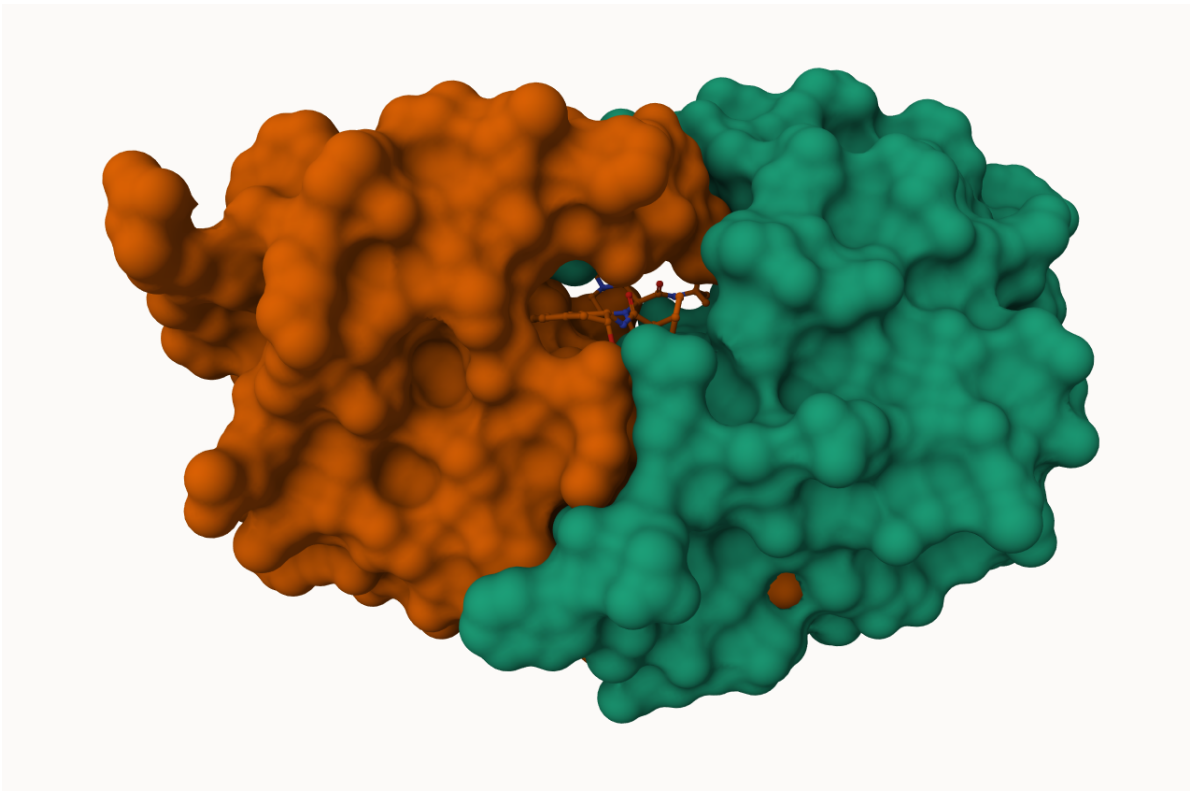


Figure 2: HIV inhibitor drug MK1 fit inside hsg protein

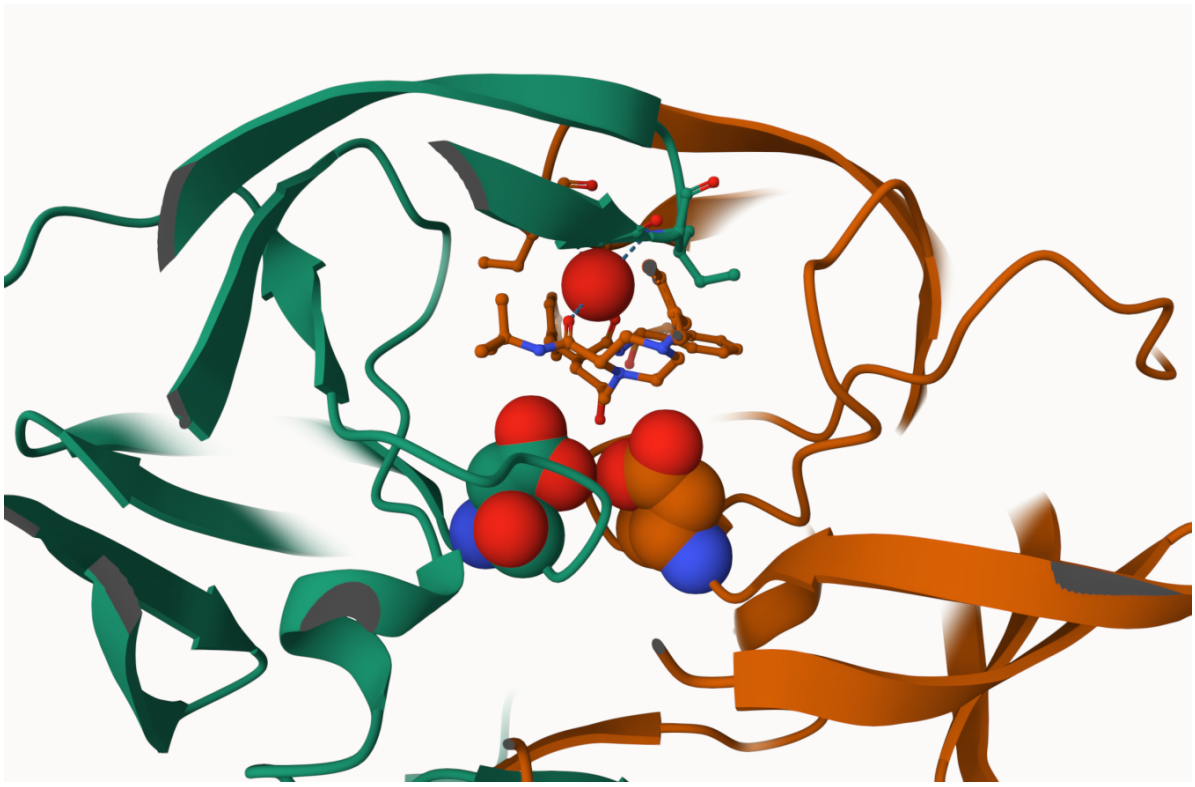


Figure 3: catalytic zone and stabilizing water 308

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
```

```
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>
```

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

```
 25
"D"
```

Q7: How many amino acid residues are there in this pdb object?

198

Q8: Name one of the two non-protein residues?

HOH and MK1 >Q9: How many protein chains are in this structure?

2

```
length(pdbseq(pdb))
```

```
[1] 198
```

##Functional dynamics prediction

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

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

```
summary(adk)
```

```
 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) ]

+ attr: atom, xyz, seqres, helix, sheet,
        calpha, remark, call
```

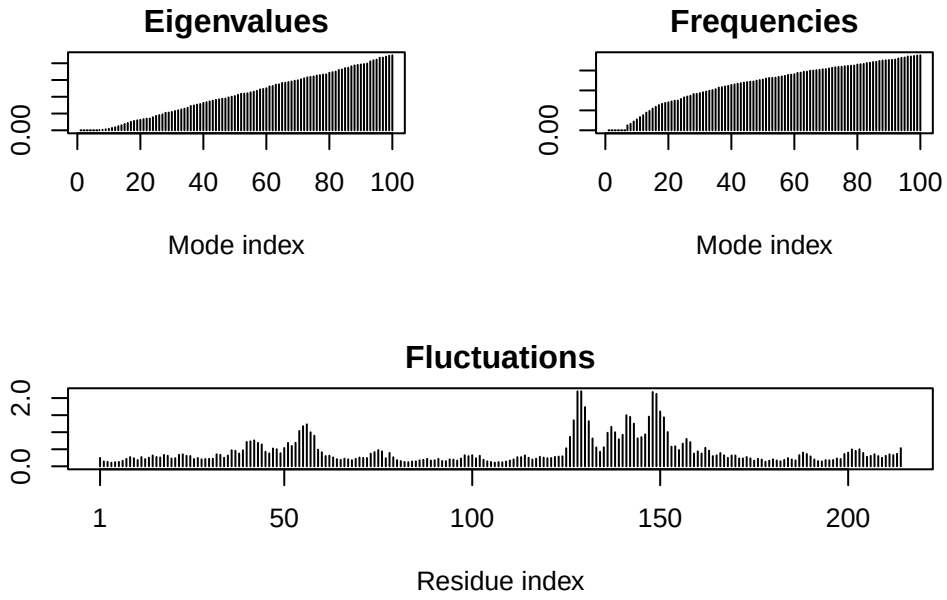
```
source("https://tinyurl.com/viewpdb")
library(r3dmol)

#view.pdb(pdb, backgroundColor = "yellow")
```

```
#Tells fluctuation
modes <- nma(adk)
```

```
 Building Hessian...        Done in 0.025 seconds.
 Diagonalizing Hessian...   Done in 0.282 seconds.
```

```
plot(modes)
```




```
#The 3 lines of code that give us trajectory file that can view in Mol-star to show flexibili

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

```
  Note: Accessing on-line PDB file

Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
/var/folders/yx/txx1yxp11mv_frz5vxdqg1dh0000gn/T//Rtmp9PHoyw/6s36.pdb exists.
Skipping download
```

```
   PDB has ALT records, taking A only, rm.alt=TRUE
```

```
modes <- nma(adk)
```

```
 Building Hessian...        Done in 0.012 seconds.
 Diagonalizing Hessian...   Done in 0.274 seconds.
```

```
mktrj(modes, file = "adk.pdb")
```