Class 9: Structual Bioinformatics

Youn Soo Na (PID: A17014731)

The main database for structural data is called the PDB (Protein Data Bank). Let's see what it contains!

Data from: https://www.rcsb.org/stats Or from alternate link: https//tinyurl.com/pdbstats24

```r
pdb24 <- read.csv("pdb_stats.csv", row.names=1)
head(pdb24)
```

```
                          X.ray     EM    NMR Multiple.methods Neutron Other
Protein (only)          167,192 15,572 12,529              208      77    32
Protein/Oligosaccharide   9,639  2,635     34                8       2     0
Protein/NA                8,730  4,697    286                7       0     0
Nucleic acid (only)       2,869    137  1,507               14       3     1
Other                       170     10     33                0       0     0
Oligosaccharide (only)       11      0      6                1       0     4
                          Total
Protein (only)          195,610
Protein/Oligosaccharide  12,318
Protein/NA               13,720
Nucleic acid (only)       4,531
Other                       213
Oligosaccharide (only)       22
```

```r
# Some of the "numeric" values are actually characters
# We need to change them to numeric values.
# as.numeric( sub(",", "", pdb24$Total))
# I could run this into a function to fix the whole table or any future table
# I read like this:
# x <- pdb24$Total
# as.numeric( sub(",", "", x) )
```

```
 comma2numeric <- function(x) {
   as.numeric( sub(",", "", x) )
 }

# Test it.
# comma2numeric(pdb24$X.ray)
# head(pdb24)

pdb24test <- apply(pdb24, 2, comma2numeric)
head(pdb24test)
```

```
      X.ray    EM   NMR Multiple.methods Neutron Other  Total
[1,] 167192 15572 12529              208      77    32 195610
[2,]   9639  2635    34                8       2     0  12318
[3,]   8730  4697   286                7       0     0  13720
[4,]   2869   137  1507               14       3     1   4531
[5,]    170    10    33                0       0     0    213
[6,]     11     0     6                1       0     4     22
```

```
## try a different read/import function:
library(readr)
pdbdb <- read_csv("pdb_stats.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.
```

```
sum(pdbdb$Total)
```

```
[1] 226414
```

And answer the following questions:

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

```
(sum(pdbdb$`X-ray`)/sum(pdbdb$Total) * 100)+(sum(pdbdb$EM)/sum(pdbdb$Total) *100)
```

```
[1] 93.4845
```

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

```
# library(dplyr)
# pdbdb %>%
#   filter(rowSums(sapply(., function(x) grepl("protein", x, ignore.case = TRUE))) > 0)
pdbdb$Total[1]/sum(pdbdb$Total) * 100
```

```
[1] 86.39483
```

> 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?

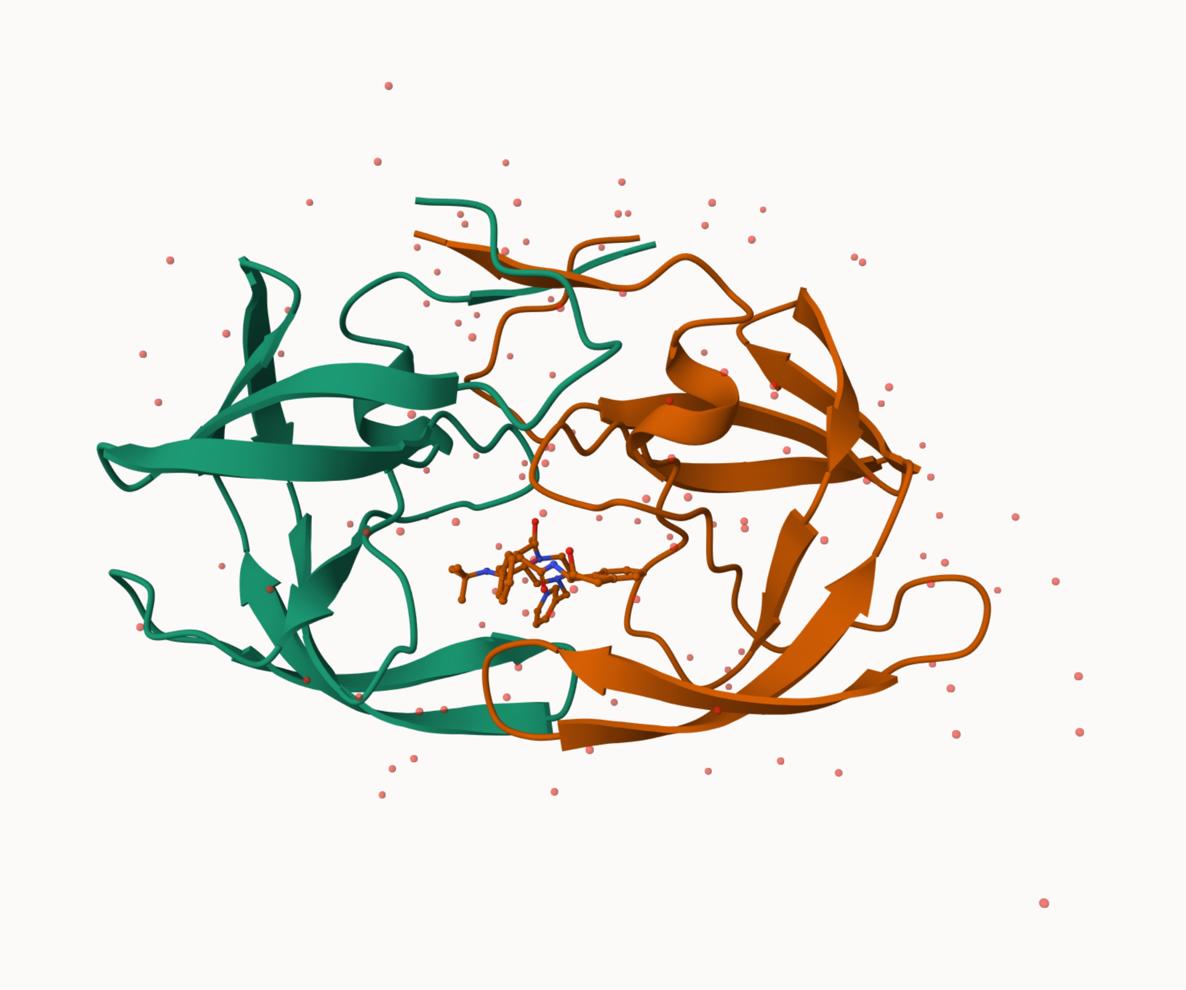
4553

Mol*

Mol* (pronounced "molstar") is a new web-based molecular viewer that we will need to learn the basics of here.

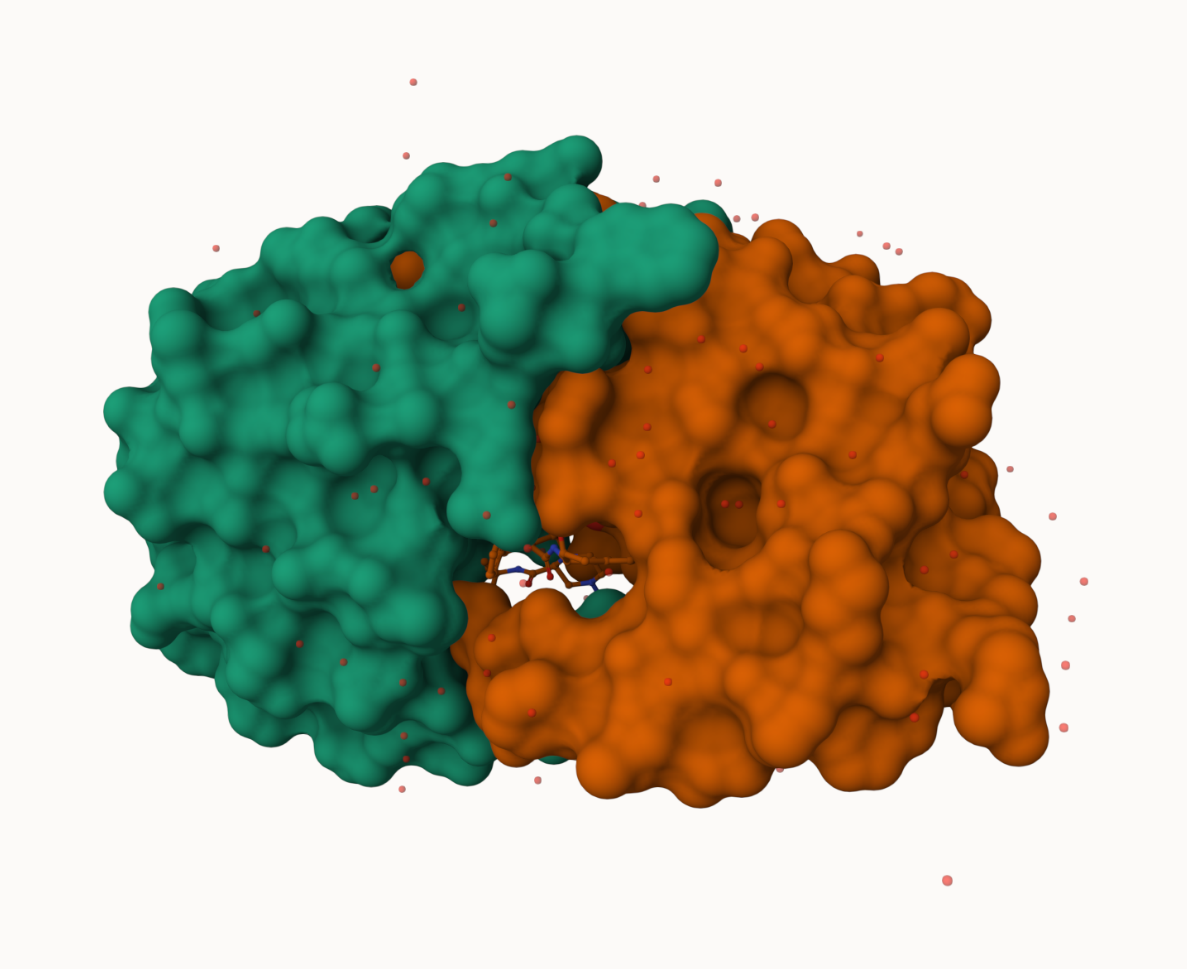
https://molstar.org/viewer/

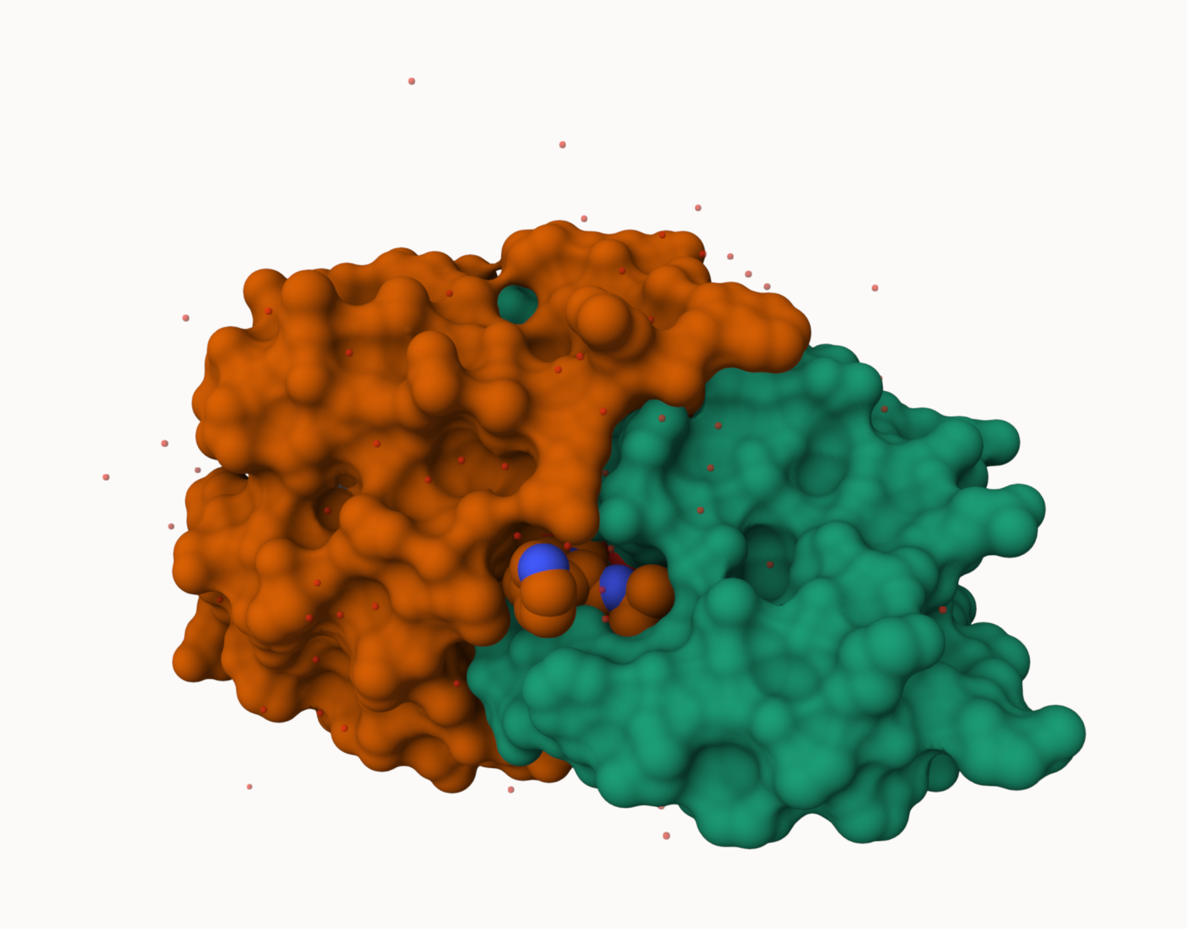
We will use PDB code: 1HSG



Note: This is an aspartic protease that uses 2 aspartic acid

Here are some more custome images:





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

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.

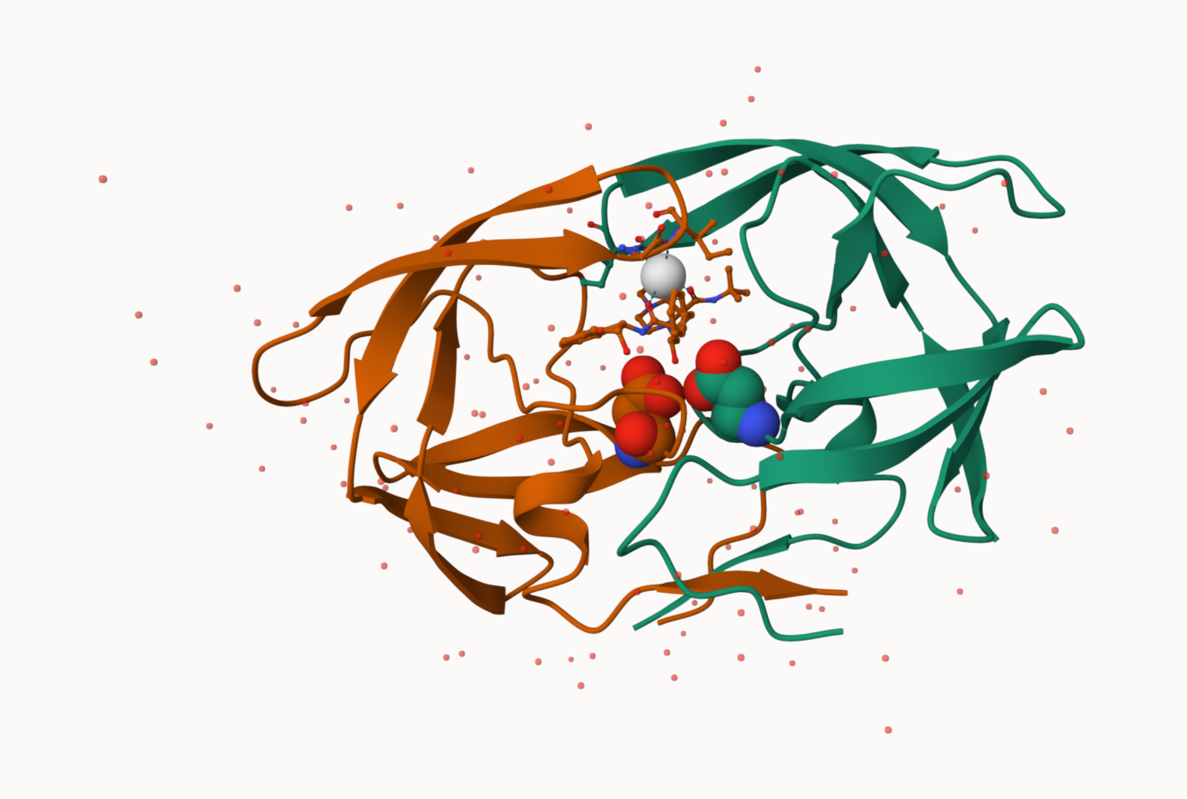


Figure 1: Water molecule is represented as a white ball