

Class_09

Dan Vu (PID: A17380158)

```
library(tidyverse)
library(readr)
pdbdata <- read_csv("~/Downloads/Data Export Summary.csv")
head(pdbdata)
```

```
# A tibble: 6 x 9
  `Molecular Type`      `X-ray`      EM      NMR Integrative `Multiple methods` Neutron
  <chr>                <dbl> <dbl> <dbl>          <dbl>          <dbl> <dbl>
1 Protein (only)      176204 20299 12708          342          218    83
2 Protein/Oligosacch~  10279  3385    34           8           11     1
3 Protein/NA           9007  5897   287          24           7     0
4 Nucleic acid (only)  3066   200  1553           2          15     3
5 Other                173    13    33           3           0     0
6 Oligosaccharide (o~    11     0     6           0           1     0
# i 2 more variables: Other <dbl>, Total <dbl>
```

Download a CSV file from the PDB site (accessible from “Analyze” > “PDB Statistics” > “by Experimental Method and Molecular Type”. Move this CSV file into your RStudio project and use it to answer the following questions:

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

```
n.total <- sum(pdbdata$Total)
n.xray <- sum(pdbdata$`X-ray`)
n.EM <- sum(pdbdata$EM)
```

X-Ray Percentage

```
round(n.xray/n.total * 100,2)
```

```
[1] 81.48
```

81.48 percent

EM Percentage

```
round(n.EM/n.total * 100,2)
```

```
[1] 12.22
```

12.22 percent

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

```
n.protein <- pdldata$Total[1]  
n.protein
```

```
[1] 209886
```

```
round(n.protein/n.total * 100,2)
```

```
[1] 86.05
```

86.05 percent

Exploring PDB Structures

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?

Package for structural bioinformatics

```
library(bio3d)  
  
hiv <- read.pdb("1hsg")
```

Note: Accessing on-line PDB file

hiv

```
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
```

Let's first use the Mol* viewer to explore this structure.

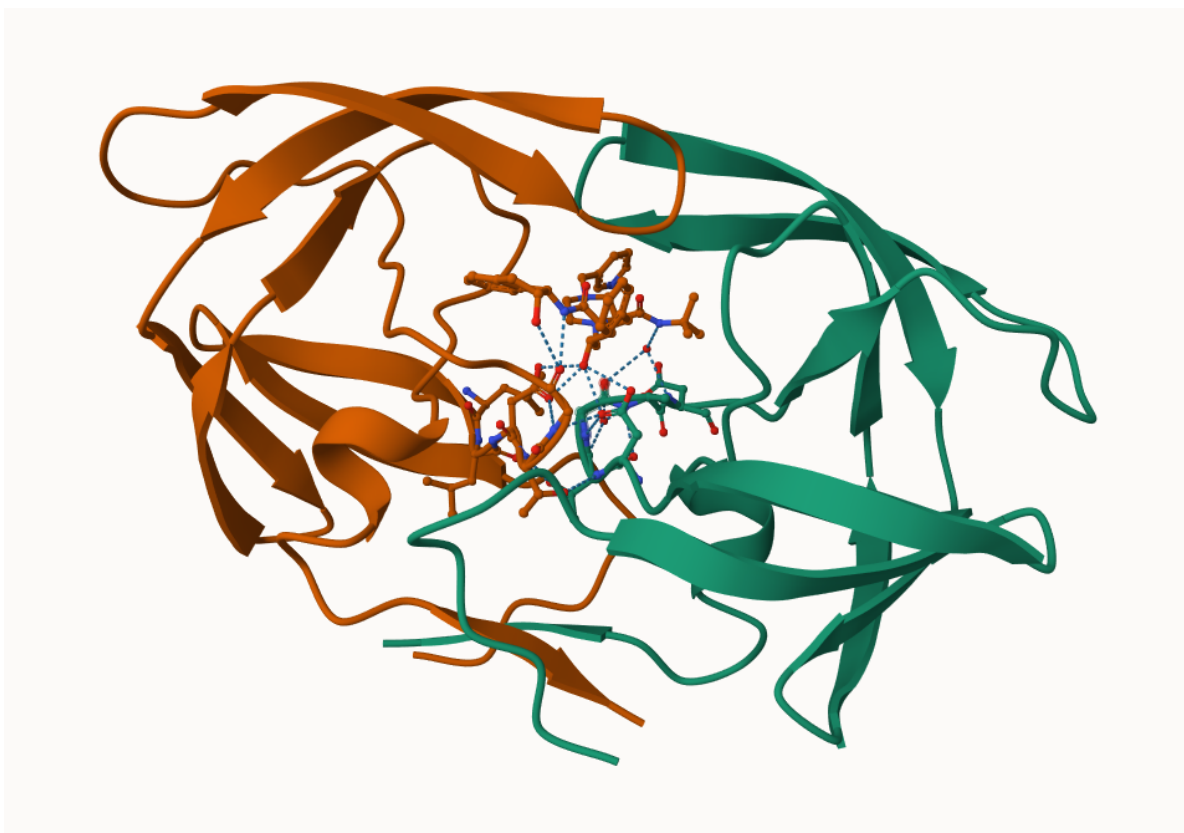


Figure 1: My first view of HIV-Pr