Class09Lab

Peyton Chiu (PID:A18145937)

The main database for structural biology is called the PDB. Let's havve a look at what it contains.

Download CS file from the PDB site accessible from :"Analyze" > "PDB Statistics" > "by Experimental Method and Molecular Type" "Analyze" > "PDB Statistics" > "by Experimental Method and Molecular Type"

```
pdb.df <- read.csv("Data Export Summary.csv")
```

```
library(readr)
pdb.df <-read_csv("Data Export Summary.csv")
pdb.df
```

```
# 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>
```

> Q1: What percentage of structures in the PDB are solved by X-Ray and Electron Microscopy.Give your answers in two sig figs

```
total_structures <- sum(pdb.df$Total)
n.Xray <- sum(pdb.df$`X-ray`)
ans1 <-n.Xray/total_structures * 100
```

```
n.EM <- sum(pdb.df$EM)
ans2<-n.EM/total_structures * 100
```

There are 81.48 percent X-ray structures in the PDB and there are 12.22 percent EM sturtures.

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

```
total_structures <- sum(pdb.df$Total)
total_protein <- pdb.df[1,9]
protein_prop <- total_protein/total_structures
protein_prop
```

```
      Total
1 0.8605059
```

86.05% of all the structures are 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?

There are 1,150 HIV-1 protease structures

Exploring PDB structures

Package for structural infomatics

```
library(bio3d)
```

```
Warning: package 'bio3d' was built under R version 4.4.3
```

```
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 Mol* viewer tp explore this structure

Figure 1: My first view of HIV-Pr

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 haveQ4: Water molecules normally have 3 atoms. Why do we see just one atom per water molecule in this structure?

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.