Class10

Kyle Wittkop (A18592410)

Table of contents

PBD statistics

The Protein Data bank (PDB) is the main repository of biomolecular structutres. Let's see what it contains.

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

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

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

```
Rows: 6 Columns: 9
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (1): Molecular Type
dbl (4): Integrative, 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.
```

```
XrayEm<-sum(c(Dataexp$`X-ray`,Dataexp$EM))

(XrayEm/sum(Dataexp$Total))*100
```

```
[1] 93.7892
```

93.7892%

The commas in these numbers leads to the numbers here being read as characters.

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

```
Dataexp
```

```
# 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)       178795 21825 12773         343                  226      84
2 Protein/Oligosacch~   10363  3564    34           8                   11       1
3 Protein/NA             9106  6335   287          24                    7       0
4 Nucleic acid (only)    3132   221  1566           3                   15       3
5 Other                   175    25    33           4                    0       0
6 Oligosaccharide (o~      11     0     6           0                    1       0
# i 2 more variables: Other <dbl>, Total <dbl>
```

```
sum(Dataexp[1,9]) / sum(Dataexp$Total) * 100
```

```
[1] 85.96889
```

85.96889%

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

Visualizing the HIV-1 protease structure

We can use Molstar viewer online : https://molstar.org/viewer/

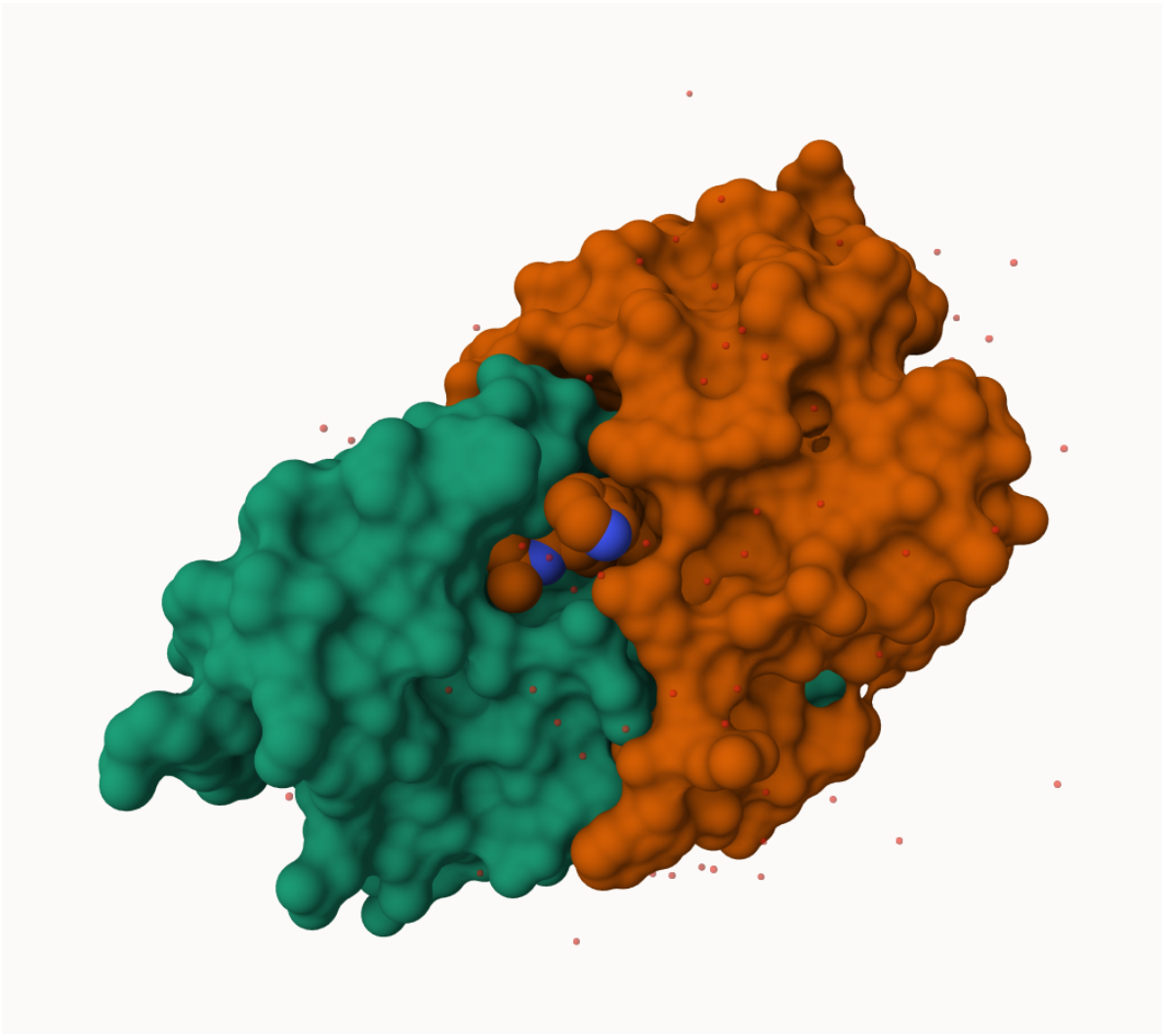


Figure 1: HIV-Pr with surface display showing ligand

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

the Hydrogen atoms are not shown, only the oxygen atom 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.

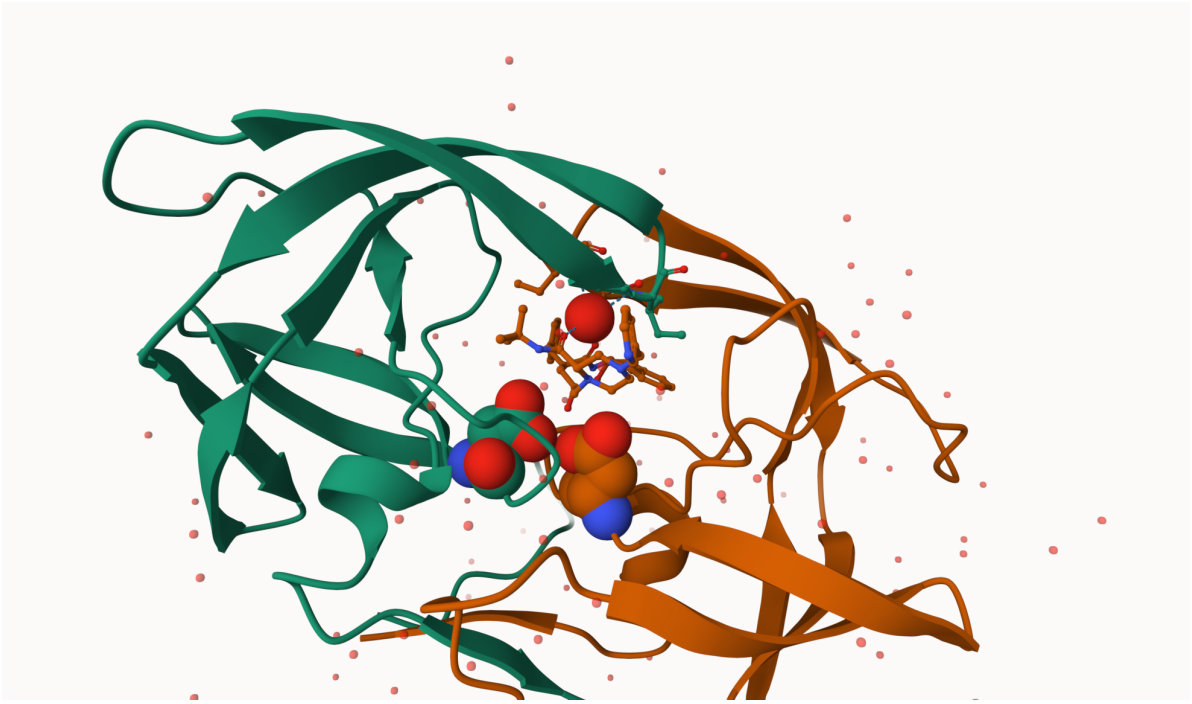


Figure 2: Image showing the catalytic ASP25 amino acid in both chains of the HIV-PR dimer along with the inhibitor and important active site water

Introduction to Bio3D in R

```
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
```

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

128

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

HOH (127)

Q9: How many protein chains are in this structure?

2 chains

```
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>
```

```
#library(bio3dview)
#library(NGLVieweR)

#view.pdb(pdb) |>
  #setSpin()
```

```
#sele <- atom.select(pdb, resno=25)

#view.pdb(pdb, cols=c("navy","teal"),
        # highlight = sele,
        # highlight.style = "spacefill") |>
  #setRock()
```

Predicting functional motions of a single structure

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

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

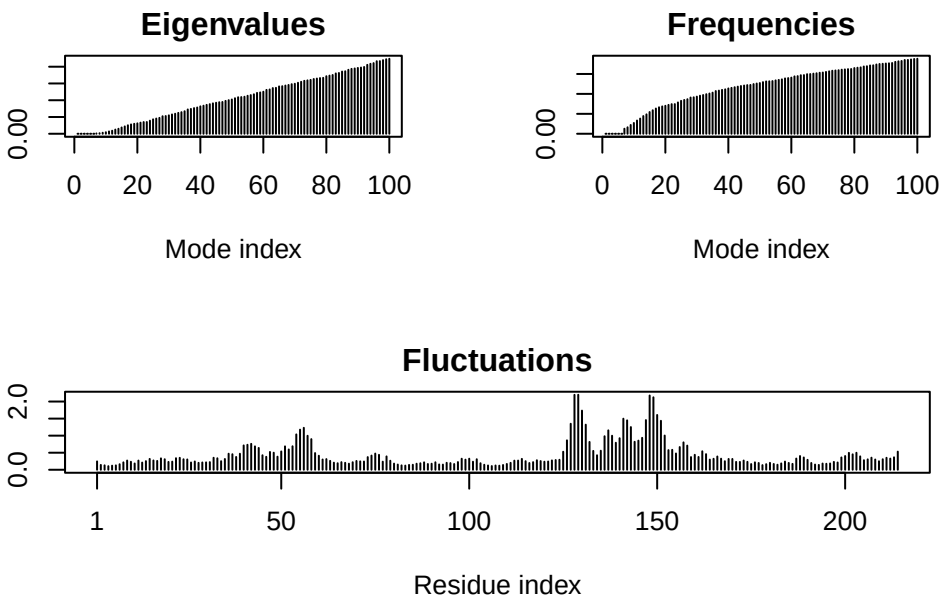
```
adk
```

```
 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) ]

   Protein sequence:
      MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVT
      DELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDKI
      VGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIG
      YYSKEAEAGNTKYAKVDGTKPVAEVRADLEKILG

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

```
m <- nma(adk)
```

```
 Building Hessian...        Done in 0.021 seconds.
 Diagonalizing Hessian...   Done in 0.086 seconds.
```

```
plot(m)
```



write out our results with a trajectory/movie of predicted motions

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

Comparative Analysis with PCA

First step - find an ADK sequence

```
#library(bio3d)
#id <- "1ake_A" ## Change this to run a different analysis
#aa <- get.seq(id)
```