class09(1-9)

Izabelle Querubin

1. Introduction to the RCSB Protein Data Bank (PDB)

```
pdb_data <- read.csv("MolecType.csv", header = TRUE, stringsAsFactors = FALSE, row.names =
pdb_data$X.ray <- gsub(",", "", pdb_data$X.ray)
pdb_data$EM <- gsub(",", "", pdb_data$EM)
pdb_data$NMR <- gsub(",", "", pdb_data$NMR)
pdb_data$Total <- gsub(",", "", pdb_data$Total)
head(pdb_data)
```

```
                         X.ray    EM   NMR Multiple.methods Neutron Other
Protein (only)          154904 10218 12189              191      72    32
Protein/Oligosaccharide   9089  1805    32                7       1     0
Protein/NA                8129  3184   283                6       0     0
Nucleic acid (only)       2675    94  1450               12       2     1
Other                      163     9    32                0       0     0
Oligosaccharide (only)      11     0     6                1       0     4
                         Total
Protein (only)          177606
Protein/Oligosaccharide  10934
Protein/NA               11602
Nucleic acid (only)       4234
Other                      204
Oligosaccharide (only)      22
```

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

```
pdb_data$X.ray <- as.integer(pdb_data$X.ray)
pdb_data$EM <- as.integer(pdb_data$EM)
pdb_data$NMR <- as.integer(pdb_data$NMR)
```

```
pdb_data$Total <- as.integer(pdb_data$Total)
head(pdb_data)
```

```
                       X.ray    EM   NMR Multiple.methods Neutron Other
Protein (only)        154904 10218 12189               191      72    32
Protein/Oligosaccharide 9089  1805    32                 7       1     0
Protein/NA              8129  3184   283                 6       0     0
Nucleic acid (only)     2675    94  1450                12       2     1
Other                    163     9    32                 0       0     0
Oligosaccharide (only)    11     0     6                 1       0     4
                       Total
Protein (only)        177606
Protein/Oligosaccharide 10934
Protein/NA              11602
Nucleic acid (only)      4234
Other                     204
Oligosaccharide (only)     22
```

```
percent_xray <- sum(pdb_data$X.ray, pdb_data$EM, na.rm = TRUE) / sum(pdb_data$Total) * 100
percent_xray
```

```
[1] 93.00056
```

93% of structures in the PDB are solved by X-Ray and Electron Microscopy.

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

```
sum(pdb_data$Total[1], pdb_data$Total[2], pdb_data$Total[3]) / sum(pdb_data$Total)
```

```
[1] 0.9782016
```

The proportion of structures in the PDB are protein is 0.98.

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 2,067 HIV-1 protease structures in the current PDB.

2. Visualizing the HIV-1 protease structure

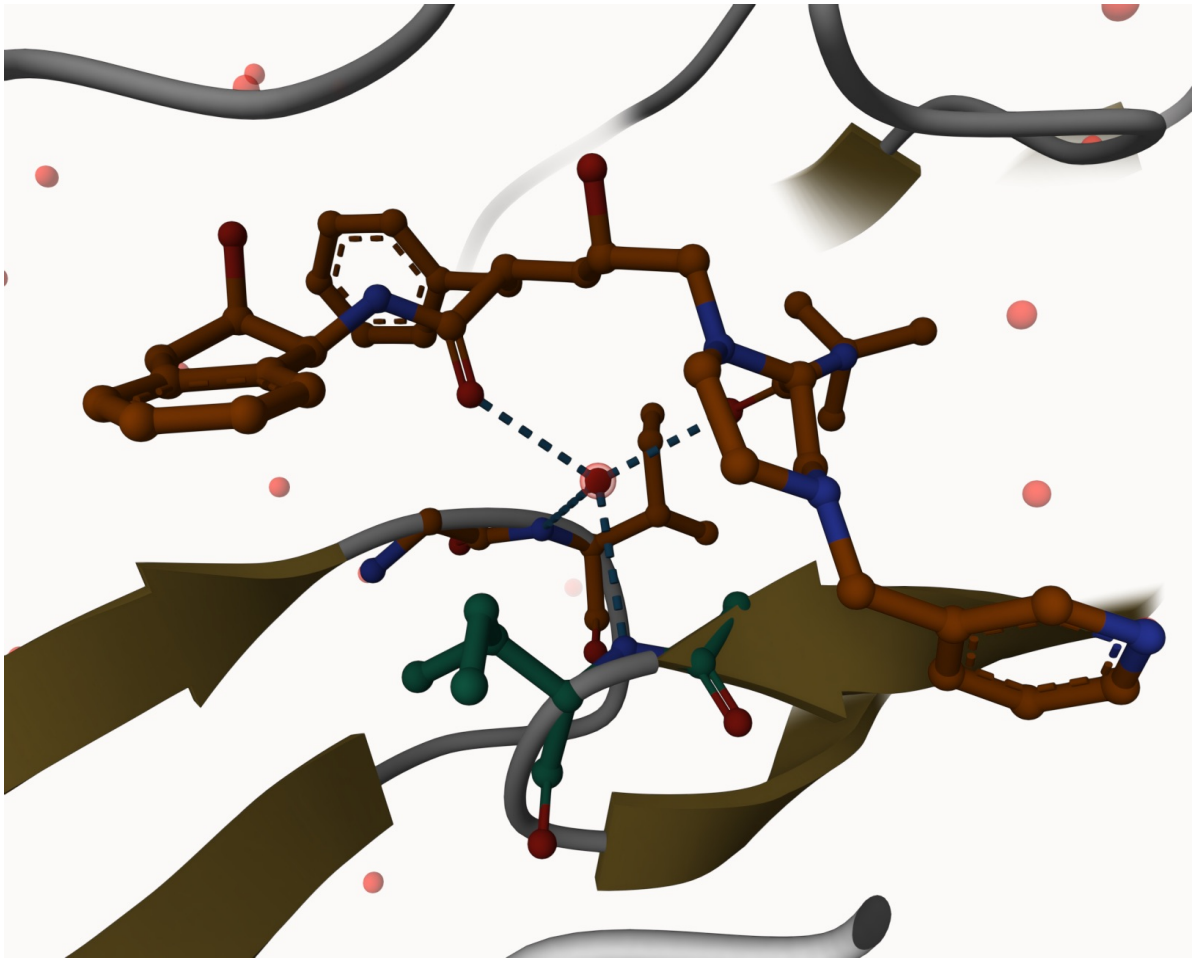
Q4: Water molecules normally have 3 atoms. Why do we see just one atom per water molecule in this structure?

In this structure, we see only one atom per water molecule due to the PDB format. As a way to conserve space and reduce the size of PDB files, water molecules are only represented by one atom.

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.



3. 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 acids residues are there in this pdb object?

198

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

HOH

Q9. How many protein chains are in this structure?

2

```
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>
```

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
```

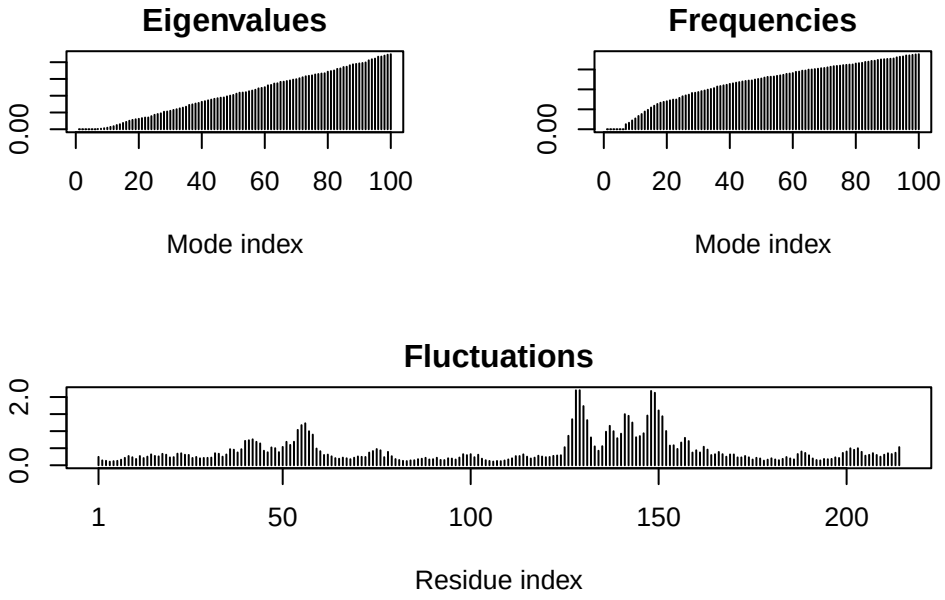
```
      VGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIG
      YYSKEAEAGNTKYAKVDGTKPVAEVRADLEKILG

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

```
# Perform flexibility prediction
m <- nma(adk)
```

```
 Building Hessian...        Done in 0.033 seconds.
 Diagonalizing Hessian...   Done in 0.335 seconds.
```

```
plot(m)
```




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