Class10 Structural Bioinformatics Pt1 (11/01/24)

Janie Chang-Weinberg (A69037446)

What is in the PDB database anyway? Summary data grabbed from: https://www.rcsb.org/stats/summary

```
pdbstats <- read.csv("PDB_stats_summaryfile.csv", row.names = 1)
pdbstats
```

```
                          X.ray     EM    NMR Multiple.methods Neutron Other
Protein (only)          167,317 15,698 12,534               208      77    32
Protein/Oligosaccharide   9,645  2,639     34                 8       2     0
Protein/NA                8,735  4,718    286                 7       0     0
Nucleic acid (only)       2,869    138  1,507                14       3     1
Other                       170     10     33                 0       0     0
Oligosaccharide (only)       11      0      6                 1       0     4
                          Total
Protein (only)          195,866
Protein/Oligosaccharide  12,328
Protein/NA               13,746
Nucleic acid (only)       4,532
Other                       213
Oligosaccharide (only)       22
```

```
x <- pdbstats$Total
x
```

```
[1] "195,866" "12,328"  "13,746"  "4,532"   "213"     "22"
```

```
#as.numeric(x): Warning: NAs introduced by coercion[1]  NA  NA  NA  NA 213  22
```

```
gsub(',', '', x)
```

```
[1] "195866" "12328"  "13746"  "4532"   "213"    "22"
```

doesn't like the commas, ask google. can also use "claude" use gsub to replace commas with nothings

```
convertednocomma <- function(x){
  #remove comma
  x <- gsub(',', '', x)
  #convert to numeric
  x <- as.numeric(x)
  return(x)
}
```

```
convertednocomma(pdbstats$Total)
```

```
[1] 195866  12328  13746   4532    213     22
```

```
n.tot <- sum(convertednocomma(pdbstats$Total))
n.tot
```

```
[1] 226707
```

The `apply()` function is very useful as it can take any function and "Apply" it over either the ROWS or COLs of a data.frame.

```
colSums(apply(pdbstats, 2, convertednocomma))/n.tot
```

```
          X.ray              EM             NMR Multiple.methods
   0.8325592064    0.1023479646    0.0635181093    0.0010498132
        Neutron           Other           Total
   0.0003617003    0.0001632063    1.0000000000
```

```
#another way of importing data to avoid issues with commas in the numbas
#library(readr)
#read_csv("PDB_stats_summaryfile.csv")
```

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

```
n.xray <- sum(convertednocomma(pdbstats$X.ray))
n.em <- sum(convertednocomma(pdbstats$EM))

xraypercent <- ((n.xray/n.tot)*100)
empercent <- ((n.em/n.tot)*100)

xraypercent
```

```
[1] 83.25592
```

```
empercent
```

```
[1] 10.2348
```

83.26% (xray%), 10.23% (EM%)

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

```
n.totalprots <- 195866/n.tot

n.totalprots
```

```
[1] 0.863961
```

86.4%

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?

4,640

Using Mol*

Figure 1: My first image from Mol-star :D

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

> Cannot see Hydrogen at this resolution, the only atom being displayed here is oxygen.

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

> This water molecule's residue number is 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: with Aspartate residues and H2O mol

Use the Bio3D package for structural Bioinformatics

```
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
```

```
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>
```

```
pdbseq(pdb)[25]
```

```
 25
"D"
```

Q7. How many amino acids are there in this structure?

```
length(pdbseq(pdb))
```

```
[1] 198
```

There are 198 amino acids in this structure

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

HOH (water)

Q9. How many protein chains are in this structure?

2

```
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
```

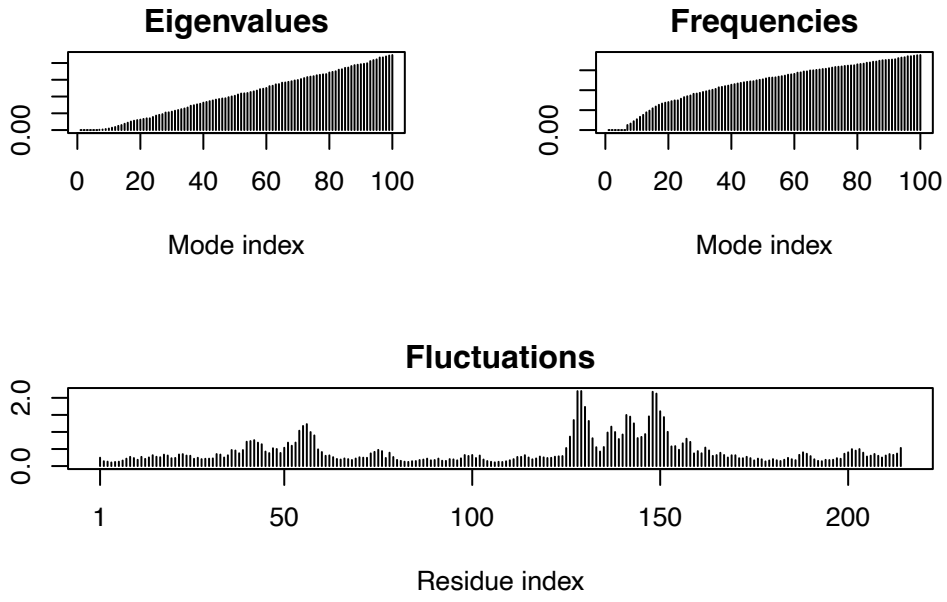
```
library(r3dmol)
source("https://tinyurl.com/viewpdb")

library(shiny)
view.pdb(pdb)
```

```
modes <- nma(adk)
```

```
 Building Hessian...        Done in 0.018 seconds.
 Diagonalizing Hessian...   Done in 0.342 seconds.
```

```
plot(modes)
```




```
mktrj(modes, pdb=adk, file='adk.pdb') #give it the pdb= thing to have proper seq
```

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

Q10. Which of the packages above is found only on BioConductor and not CRAN?

biocmanager

Q.11 Which of the above packages is not found on BioConductor or CRAN?

bio3d

Q.12 True or False? Functions from the devtools package can be used to install packages from GitHub and BitBucket?

True

```
library(bio3d)
aa <- get.seq("1ake_A")
```

```
Warning in get.seq("1ake_A"): Removing existing file: seqs.fasta

Fetching... Please wait. Done.
```

```
aa
```

```
              1        .         .         .         .         .         60
pdb|1AKE|A    MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVT
              1        .         .         .         .         .         60

             61        .         .         .         .         .         120
pdb|1AKE|A    DELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRI
             61        .         .         .         .         .         120

            121        .         .         .         .         .         180
pdb|1AKE|A    VGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIG
            121        .         .         .         .         .         180

            181        .         .         .   214
pdb|1AKE|A    YYSKEAEAGNTKYAKVDGTKPVAEVRADLEKILG
            181        .         .         .   214

Call:
  read.fasta(file = outfile)

Class:
  fasta

Alignment dimensions:
  1 sequence rows; 214 position columns (214 non-gap, 0 gap)

+ attr: id, ali, call
```

Q13. How many amino acids are in this sequence, i.e. how long is this sequence?

There are 214 AAs

##Blast or hmmer search

```
b <- blast.pdb(aa)
```

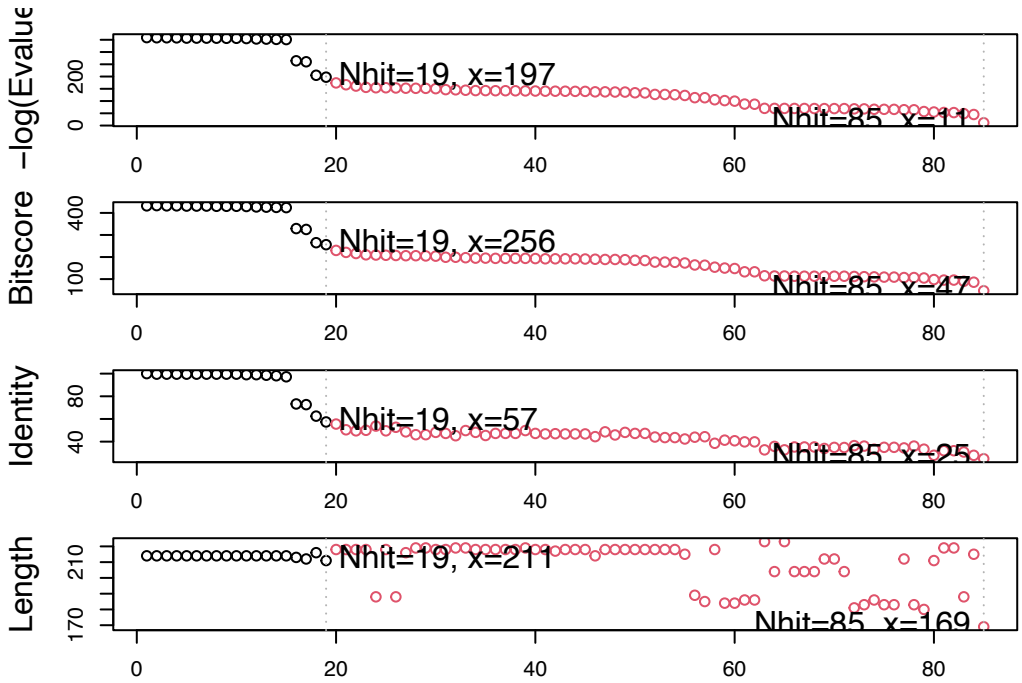
```
 Searching ... please wait (updates every 5 seconds) RID = JP0JPEGF016
 ..................
 Reporting 85 hits
```

##Plot a summary of search results:

```
hits <- plot(b)
```

```
  * Possible cutoff values:    197 11
            Yielding Nhits:    19 85

  * Chosen cutoff value of:    197
            Yielding Nhits:    19
```



##List some "Top Hits" and download their .pdb files:

```
head(hits$pdb.id)
```

```
[1] "1AKE_A" "8BQF_A" "4X8M_A" "6S36_A" "8Q2B_A" "8RJ9_A"
```

```
files <- get.pdb(hits$pdb.id, path="pdbs", split=TRUE, gzip=TRUE)
```

```
Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/1AKE.pdb.gz exists. Skipping download

Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/8BQF.pdb.gz exists. Skipping download

Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/4X8M.pdb.gz exists. Skipping download

Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/6S36.pdb.gz exists. Skipping download

Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/8Q2B.pdb.gz exists. Skipping download

Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/8RJ9.pdb.gz exists. Skipping download

Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/6RZE.pdb.gz exists. Skipping download

Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/4X8H.pdb.gz exists. Skipping download

Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/3HPR.pdb.gz exists. Skipping download

Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/1E4V.pdb.gz exists. Skipping download

Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/5EJE.pdb.gz exists. Skipping download
```

```
Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/1E4Y.pdb.gz exists. Skipping download

Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/3X2S.pdb.gz exists. Skipping download

Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/6HAP.pdb.gz exists. Skipping download

Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/6HAM.pdb.gz exists. Skipping download

Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/4K46.pdb.gz exists. Skipping download

Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/4NP6.pdb.gz exists. Skipping download

Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/3GMT.pdb.gz exists. Skipping download

Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/4PZL.pdb.gz exists. Skipping download


  |
  |                                                                      |   0%
  |
  |====                                                                  |   5%
  |
  |=======                                                               |  11%
  |
  |===========                                                           |  16%
  |
  |===============                                                       |  21%
  |
  |==================                                                    |  26%
  |
  |======================                                                |  32%
  |
```

```
  |=========================                                             |  37%
  |
  |============================                                          |  42%
  |
  |================================                                      |  47%
  |
  |====================================                                  |  53%
  |
  |========================================                              |  58%
  |
  |============================================                          |  63%
  |
  |================================================                      |  68%
  |
  |====================================================                  |  74%
  |
  |========================================================              |  79%
  |
  |============================================================          |  84%
  |
  |================================================================      |  89%
  |
  |====================================================================  |  95%
  |
  |======================================================================| 100%
```

##Align and Superimpose structures:

```
pdbs <- pdbaln(files, fit = TRUE, exefile="msa")
```

```
Reading PDB files:
pdbs/split_chain/1AKE_A.pdb
pdbs/split_chain/8BQF_A.pdb
pdbs/split_chain/4X8M_A.pdb
pdbs/split_chain/6S36_A.pdb
pdbs/split_chain/8Q2B_A.pdb
pdbs/split_chain/8RJ9_A.pdb
pdbs/split_chain/6RZE_A.pdb
pdbs/split_chain/4X8H_A.pdb
pdbs/split_chain/3HPR_A.pdb
pdbs/split_chain/1E4V_A.pdb
pdbs/split_chain/5EJE_A.pdb
```

```
pdbs/split_chain/1E4Y_A.pdb
pdbs/split_chain/3X2S_A.pdb
pdbs/split_chain/6HAP_A.pdb
pdbs/split_chain/6HAM_A.pdb
pdbs/split_chain/4K46_A.pdb
pdbs/split_chain/4NP6_A.pdb
pdbs/split_chain/3GMT_A.pdb
pdbs/split_chain/4PZL_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
.   PDB has ALT records, taking A only, rm.alt=TRUE
..   PDB has ALT records, taking A only, rm.alt=TRUE
.   PDB has ALT records, taking A only, rm.alt=TRUE
.   PDB has ALT records, taking A only, rm.alt=TRUE
.   PDB has ALT records, taking A only, rm.alt=TRUE
..   PDB has ALT records, taking A only, rm.alt=TRUE
..   PDB has ALT records, taking A only, rm.alt=TRUE
....   PDB has ALT records, taking A only, rm.alt=TRUE
.   PDB has ALT records, taking A only, rm.alt=TRUE
....

Extracting sequences

pdb/seq: 1   name: pdbs/split_chain/1AKE_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 2   name: pdbs/split_chain/8BQF_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 3   name: pdbs/split_chain/4X8M_A.pdb
pdb/seq: 4   name: pdbs/split_chain/6S36_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 5   name: pdbs/split_chain/8Q2B_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 6   name: pdbs/split_chain/8RJ9_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 7   name: pdbs/split_chain/6RZE_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 8   name: pdbs/split_chain/4X8H_A.pdb
pdb/seq: 9   name: pdbs/split_chain/3HPR_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 10   name: pdbs/split_chain/1E4V_A.pdb
pdb/seq: 11   name: pdbs/split_chain/5EJE_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 12   name: pdbs/split_chain/1E4Y_A.pdb
pdb/seq: 13   name: pdbs/split_chain/3X2S_A.pdb
```

```
pdb/seq: 14   name: pdbs/split_chain/6HAP_A.pdb
pdb/seq: 15   name: pdbs/split_chain/6HAM_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 16   name: pdbs/split_chain/4K46_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 17   name: pdbs/split_chain/4NP6_A.pdb
pdb/seq: 18   name: pdbs/split_chain/3GMT_A.pdb
pdb/seq: 19   name: pdbs/split_chain/4PZL_A.pdb
```

```
ids <- basename.pdb(pdbs$id)
#plot(pdbs, labels=ids)

#commenting out, as R will not render this figure. It is added here as a screenshot.
```

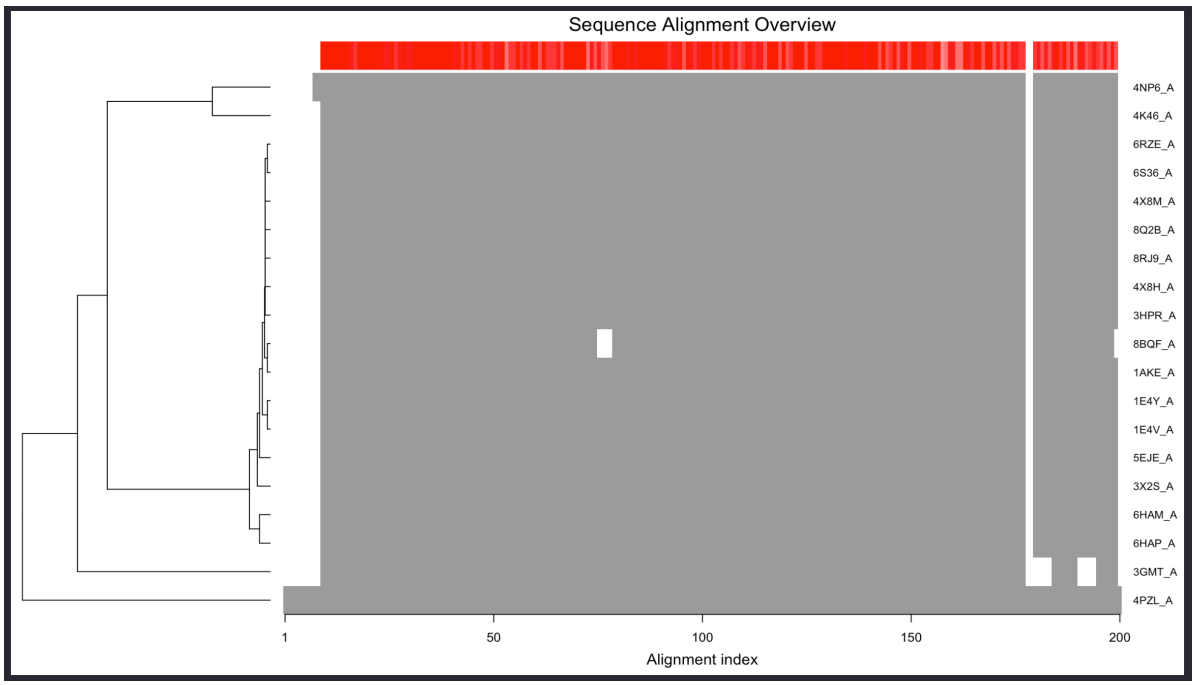


Figure 3: pdbs plot

##Annotate PDB Structures

```
anno <- pdb.annotate(ids)
unique(anno$source)
```

```
[1] "Escherichia coli"
```

```
[2] "Escherichia coli K-12"
[3] "Escherichia coli O139:H28 str. E24377A"
[4] "Escherichia coli str. K-12 substr. MDS42"
[5] "Photobacterium profundum"
[6] "Vibrio cholerae O1 biovar El Tor str. N16961"
[7] "Burkholderia pseudomallei 1710b"
[8] "Francisella tularensis subsp. tularensis SCHU S4"
```

```
anno
```

```
       structureId chainId macromoleculeType chainLength experimentalTechnique
1AKE_A        1AKE       A           Protein         214                X-ray
8BQF_A        8BQF       A           Protein         234                X-ray
4X8M_A        4X8M       A           Protein         214                X-ray
6S36_A        6S36       A           Protein         214                X-ray
8Q2B_A        8Q2B       A           Protein         214                X-ray
8RJ9_A        8RJ9       A           Protein         214                X-ray
6RZE_A        6RZE       A           Protein         214                X-ray
4X8H_A        4X8H       A           Protein         214                X-ray
3HPR_A        3HPR       A           Protein         214                X-ray
1E4V_A        1E4V       A           Protein         214                X-ray
5EJE_A        5EJE       A           Protein         214                X-ray
1E4Y_A        1E4Y       A           Protein         214                X-ray
3X2S_A        3X2S       A           Protein         214                X-ray
6HAP_A        6HAP       A           Protein         214                X-ray
6HAM_A        6HAM       A           Protein         214                X-ray
4K46_A        4K46       A           Protein         214                X-ray
4NP6_A        4NP6       A           Protein         217                X-ray
3GMT_A        3GMT       A           Protein         230                X-ray
4PZL_A        4PZL       A           Protein         242                X-ray
       resolution       scopDomain                                    pfam
1AKE_A      2.000 Adenylate kinase Adenylate kinase, active site lid (ADK_lid)
8BQF_A      2.050             <NA> Adenylate kinase, active site lid (ADK_lid)
4X8M_A      2.600             <NA>                    Adenylate kinase (ADK)
6S36_A      1.600             <NA>                    Adenylate kinase (ADK)
8Q2B_A      1.760             <NA> Adenylate kinase, active site lid (ADK_lid)
8RJ9_A      1.590             <NA>                    Adenylate kinase (ADK)
6RZE_A      1.690             <NA>                    Adenylate kinase (ADK)
4X8H_A      2.500             <NA> Adenylate kinase, active site lid (ADK_lid)
3HPR_A      2.000             <NA>                    Adenylate kinase (ADK)
1E4V_A      1.850 Adenylate kinase                    Adenylate kinase (ADK)
5EJE_A      1.900             <NA>                    Adenylate kinase (ADK)
```

1E4Y_A	1.850	Adenylate kinase	Adenylate kinase, active site lid (ADK_lid)	AP5	BIS(ADENOSINE)-5'-PENTAPHOSPHATE
3X2S_A	2.800	<NA>	Adenylate kinase (ADK)	AP5	BIS(ADENOSINE)-5'-PENTAPHOSPHATE
6HAP_A	2.700	<NA>	Adenylate kinase, active site lid (ADK_lid)	AP5	BIS(ADENOSINE)-5'-PENTAPHOSPHATE
6HAM_A	2.550	<NA>	Adenylate kinase (ADK)	AP5	BIS(ADENOSINE)-5'-PENTAPHOSPHATE
4K46_A	2.010	<NA>	Adenylate kinase, active site lid (ADK_lid)	ADP,AMP,PO4	ADENOSINE-5'-DIPHOSPHATE (2)
4NP6_A	2.004	<NA>	Adenylate kinase (ADK)	<NA>	<NA>
3GMT_A	2.100	<NA>	Adenylate kinase (ADK)	SO4 (2)	SODIUM ION (3),CHLORIDE ION (2)
4PZL_A	2.100	<NA>	Adenylate kinase (ADK)	CA,FMT,GOL	<NA>
1AKE_A	BIS(ADENOSINE)-5'-PENTAPHOSPHATE				
8BQF_A	BIS(ADENOSINE)-5'-PENTAPHOSPHATE				
4X8M_A	<NA>				
6S36_A	CHLORIDE ION (3),SODIUM ION,MAGNESIUM ION (2)				
8Q2B_A	BIS(ADENOSINE)-5'-PENTAPHOSPHATE,SULFATE ION,3[N-MORPHOLINO]PROPANE SULFONIC ACID				
8RJ9_A	ADENOSINE-5'-DIPHOSPHATE (2)				
6RZE_A	SODIUM ION (3),CHLORIDE ION (2)				
4X8H_A	<NA>				
3HPR_A	BIS(ADENOSINE)-5'-PENTAPHOSPHATE				
1E4V_A	BIS(ADENOSINE)-5'-PENTAPHOSPHATE				
5EJE_A	BIS(ADENOSINE)-5'-PENTAPHOSPHATE,COBALT (II) ION				
1E4Y_A	BIS(ADENOSINE)-5'-PENTAPHOSPHATE				
3X2S_A	N-(pyren-1-ylmethyl)acetamide (2),BIS(ADENOSINE)-5'-PENTAPHOSPHATE,MAGNESIUM ION				
6HAP_A	BIS(ADENOSINE)-5'-PENTAPHOSPHATE				

6HAM_A	BIS(ADENOSINE)-5'-PENTAPHOSPHATE
4K46_A	ADENOSINE-5'-DIPHOSPHATE,ADENOSINE MONOPHOSPHATE,PHOSPHATE ION
4NP6_A	<NA>
3GMT_A	SULFATE ION (2)
4PZL_A	CALCIUM ION,FORMIC ACID,GLYCEROL

	source
1AKE_A	Escherichia coli
8BQF_A	Escherichia coli
4X8M_A	Escherichia coli
6S36_A	Escherichia coli
8Q2B_A	Escherichia coli
8RJ9_A	Escherichia coli
6RZE_A	Escherichia coli
4X8H_A	Escherichia coli
3HPR_A	Escherichia coli K-12
1E4V_A	Escherichia coli
5EJE_A	Escherichia coli O139:H28 str. E24377A
1E4Y_A	Escherichia coli
3X2S_A	Escherichia coli str. K-12 substr. MDS42
6HAP_A	Escherichia coli O139:H28 str. E24377A
6HAM_A	Escherichia coli K-12
4K46_A	Photobacterium profundum
4NP6_A	Vibrio cholerae O1 biovar El Tor str. N16961
3GMT_A	Burkholderia pseudomallei 1710b
4PZL_A	Francisella tularensis subsp. tularensis SCHU S4

1AKE_A	STRUCTURE OF THE COMPLEX BETWEEN ADENYLATE KINASE FROM ESCHERICHIA COLI AND THE INHIBI
8BQF_A	
4X8M_A	
6S36_A	
8Q2B_A	E. coli Adenylate Kinase variant D158A (
8RJ9_A	E. coli adenylate kinase Asp84
6RZE_A	
4X8H_A	
3HPR_A	
1E4V_A	
5EJE_A	Cryst
1E4Y_A	
3X2S_A	
6HAP_A	
6HAM_A	
4K46_A	
4NP6_A	

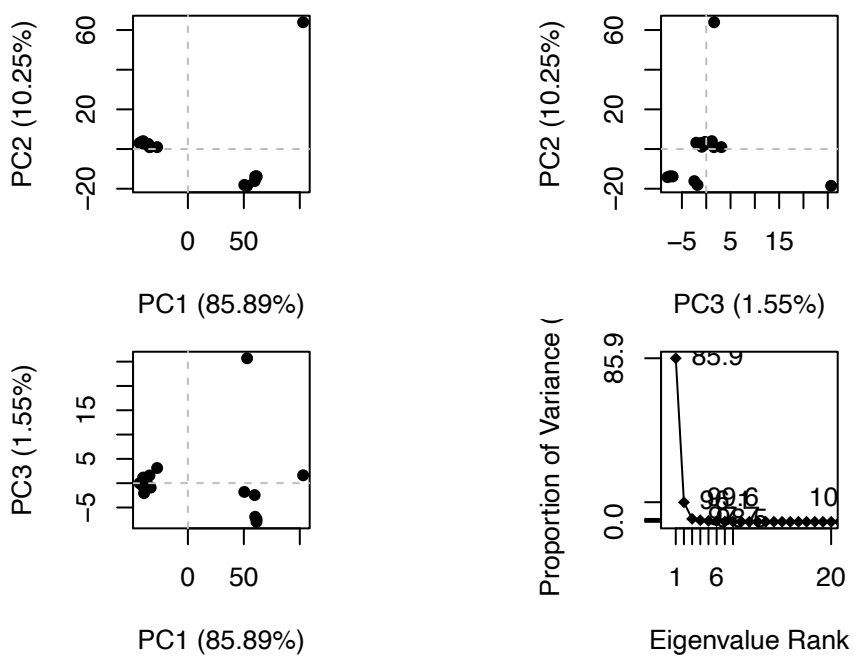
3GMT_A
4PZL_A The crys

	citation	rObserved	rFree
1AKE_A	Muller, C.W., et al. J Mol Biol (1992)	0.19600	NA
8BQF_A	Scheerer, D., et al. Proc Natl Acad Sci U S A (2023)	0.22073	0.25789
4X8M_A	Kovermann, M., et al. Nat Commun (2015)	0.24910	0.30890
6S36_A	Rogne, P., et al. Biochemistry (2019)	0.16320	0.23560
8Q2B_A	Nam, K., et al. J Chem Inf Model (2024)	0.18320	0.22440
8RJ9_A	Nam, K., et al. Sci Adv (2024)	0.15190	0.20290
6RZE_A	Rogne, P., et al. Biochemistry (2019)	0.18650	0.23500
4X8H_A	Kovermann, M., et al. Nat Commun (2015)	0.19610	0.28950
3HPR_A	Schrank, T.P., et al. Proc Natl Acad Sci U S A (2009)	0.21000	0.24320
1E4V_A	Muller, C.W., et al. Proteins (1993)	0.19600	NA
5EJE_A	Kovermann, M., et al. Proc Natl Acad Sci U S A (2017)	0.18890	0.23580
1E4Y_A	Muller, C.W., et al. Proteins (1993)	0.17800	NA
3X2S_A	Fujii, A., et al. Bioconjug Chem (2015)	0.20700	0.25600
6HAP_A	Kantaev, R., et al. J Phys Chem B (2018)	0.22630	0.27760
6HAM_A	Kantaev, R., et al. J Phys Chem B (2018)	0.20511	0.24325
4K46_A	Cho, Y.-J., et al. To be published	0.17000	0.22290
4NP6_A	Kim, Y., et al. To be published	0.18800	0.22200
3GMT_A	Buchko, G.W., et al. Biochem Biophys Res Commun (2010)	0.23800	0.29500
4PZL_A	Tan, K., et al. To be published	0.19360	0.23680

	rWork	spaceGroup
1AKE_A	0.19600	P 21 2 21
8BQF_A	0.21882	P 2 21 21
4X8M_A	0.24630	C 1 2 1
6S36_A	0.15940	C 1 2 1
8Q2B_A	0.18100	P 1 21 1
8RJ9_A	0.15010	P 21 21 2
6RZE_A	0.18190	C 1 2 1
4X8H_A	0.19140	C 1 2 1
3HPR_A	0.20620	P 21 21 2
1E4V_A	0.19600	P 21 2 21
5EJE_A	0.18630	P 21 2 21
1E4Y_A	0.17800	P 1 21 1
3X2S_A	0.20700	P 21 21 21
6HAP_A	0.22370	I 2 2 2
6HAM_A	0.20311	P 43
4K46_A	0.16730	P 21 21 21
4NP6_A	0.18600	P 43
3GMT_A	0.23500	P 1 21 1
4PZL_A	0.19130	P 32

##Perform PCA

```
pc.xray <- pca(pdbs)
plot(pc.xray)
```



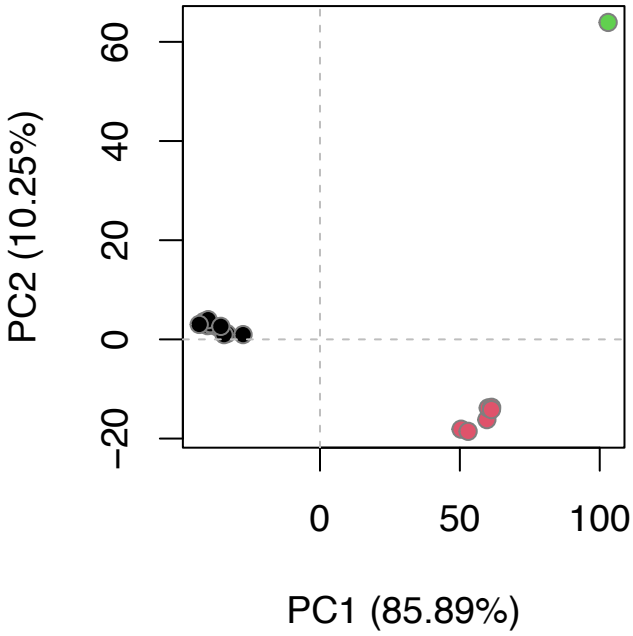
##Calculate RMSD and show structure-based clustering

```
rd <- rmsd(pdbs)
```

```
Warning in rmsd(pdbs): No indices provided, using the 199 non NA positions
```

```
hc.rd <- hclust(dist(rd))
grps.rd <- cutree(hc.rd, k=3)

plot(pc.xray, 1:2, col="grey50", bg=grps.rd, pch=21, cex=1)
```


PC2 (10.25%)
60
40
20
0
−20
0
50
100
PC1 (85.89%)