# Class 9: Structural Bioinformatics pt. 1

Chloe Wong (PID: A16893383)

The main database for structural data is called the PDB (Protein Data Bank). Let's see what it contains:

Data from alternate link: https://tinyurl.com/pdbstats24

Read this into R

```
pdbdb <- read.csv("pdb_stats.csv")
pdbdb</pre>
```

	Molecular.Type	X.ray	EM	NMR	Multiple.methods	Neutron	Other
1	Protein (only)	167,192	15,572	12,529	208	77	32
2	Protein/Oligosaccharide	9,639	2,635	34	8	2	0
3	Protein/NA	8,730	4,697	286	7	0	0
4	Nucleic acid (only)	2,869	137	1,507	14	3	1
5	Other	170	10	33	0	0	0
6	Oligosaccharide (only)	11	0	6	1	0	4
	Total						
1	195,610						
2	12,318						
3	13,720						
4	4,531						
5	213						
6	22						

and answer the following questions:

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

(answered below)

## pdbdb\$Total

```
[1] "195,610" "12,318" "13,720" "4,531" "213" "22"
```

I need to remove the comma and convert to numeric to do math:

```
as.numeric(sub(",","", pdbdb$Total))
```

```
[1] 195610 12318 13720 4531 213 22
```

I could turn this into a function to fic the whole table or any future table I read like this:

```
x <- pdbdb$Total
as.numeric (sub(",", "", x))</pre>
```

[1] 195610 12318 13720 4531 213 22

```
comma2numeric <- function(x) {
  as.numeric(sub(",", "", x))
}</pre>
```

#### Test it

```
comma2numeric(pdbdb$X.ray)
```

[1] 167192 9639 8730 2869 170 11

```
apply(pdbdb, 2, comma2numeric)
```

Warning in FUN(newX[, i], ...): NAs introduced by coercion

	Molecular.Type	X.ray	EM	NMR	${\tt Multiple.methods}$	Neutron	0ther	Total
[1,]	NA	167192	15572	12529	208	77	32	195610
[2,]	N <b>A</b>	9639	2635	34	8	2	0	12318
[3,]	N <b>A</b>	8730	4697	286	7	0	0	13720
[4,]	NA.	2869	137	1507	14	3	1	4531
[5,]	N <b>A</b>	170	10	33	0	0	0	213
[6,]	NA	11	0	6	1,	0	4	22

# Or try a different read/import function

```
#/ message: false
library(readr)
pdbdb <- read_csv("pdb_Stats.csv")</pre>
Rows: 6 Columns: 8
-- Column specification --
Delimiter: ","
chr (1): Molecular Type
dbl (3): 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.
sum(pdbdb$Total)
[1] 226414
     Q1: What percentage of structures in the PDB are solved by X-Ray and Electron
     Microscopy.
sum(pdbdb$'X-ray')/sum(pdbdb$Total) * 100
[1] 83.30359
sum(pdbdb$EM)/sum(pdbdb$Total) * 100
[1] 10.18091
     Q2: What proportion of structures in the PDB are protein?
sum(pdbdb$Total[1])/sum(pdbdb$Total) * 100
[1] 86.39483
```

# Mol\*

Mol\* (pronounced "molstar") is a new web-based molecular viewer that we will need to learn the basics of here.

https://molstar.org/viewer/

We will use PDB code: 1HSG



Some more custom images:



Figure 1: The all important catalytic ASP25 amino acids

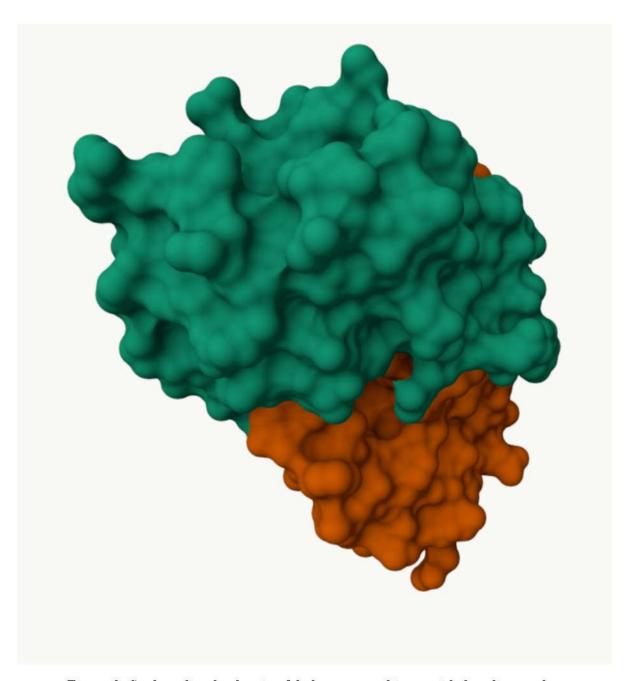
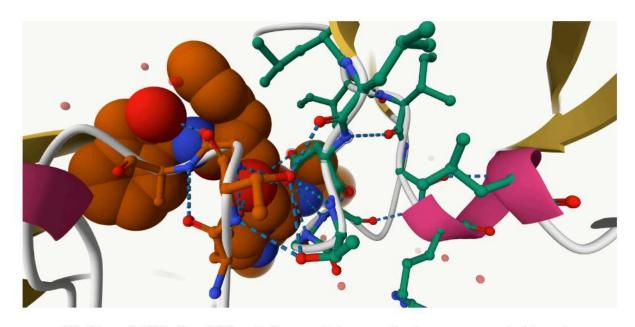


Figure 2: Surface dispaly showing Merk compound in peptide bonding pocket



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?

2294

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

There is just one atom per water molecule, oxygen. in this structure because protein have hydrogen in their structure which is not shown in the image.

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

Yes you can identify this water molecule. This water molecule has residue number 324.

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. Discussion Topic: Can you think of a way in which indinavir, or even larger ligands and substrates, could enter the binding site?

Images of the figures are shown above. Indinavir or even larger ligands and substrates could enter the binding site through conformational flexibility. Flexible loops in its structure (flap regions) are able to change confirmations to allow ligand to enter the binding site.

##The Bio3D package

The bio3d package allows us to do all sorts of structurl bioinformatics work in R. Let's start with how it can read these PDB files:

```
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
      {\tt ALLDTGADDTVLEEMSLPGRWKPKMIGGIGGFIKVRQYDQILIEICGHKAIGTVLVGPTP}
      VNIIGRNLLTQIGCTLNF
+ attr: atom, xyz, segres, helix, sheet,
        calpha, remark, call
attributes(pdb)
$names
[1] "atom"
                       "seqres" "helix" "sheet" "calpha" "remark" "call"
             ^{\rm H}{\rm Z}{\rm V}{\rm Z}^{\rm H}
$class
[1] "pdb" "sse"
```

## head(pdb\$atom)

```
type eleno elety alt resid chain resno insert
                                                                X
                                                                        У
                                                                               Z O
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
                    C <NA>
                                                    <NA> 29.760 38.071 4.022 1 42.64
             3
                              PRO
                                        A
                                               1
4 ATOM
             4
                    0 < NA >
                              PRO
                                               1
                                                    <NA> 28.600 38.302 3.676 1 43.40
5 ATOM
                  CB <NA>
                              PRO
                                                    <NA> 30.508 37.541 6.342 1 37.87
                                        A
                                               1
                  CG <NA>
                                                    <NA> 29.296 37.591 7.162 1 38.40
6 ATOM
             6
                              PRO
                                        A
                                               1
  segid elesy charge
   \langle NA \rangle
             N
                  <NA>
1
2
   <NA>
              C
                  <NA>
3
   \langle NA \rangle
              C
                  < NA >
   <NA>
              0
                  <NA>
   \langle NA \rangle
              \mathbb{C}
                  <NA>
   <NA>
              C
                   <NA>
```

### pdbseq(pdb) [25]

25 "D"

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

# sum (pdb\$calpha)

[1] 198

# length(pdbseq(pdb))

[1] 198

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

HOH and MK1

Q9: How many protein chains are in this structure?

2

#### unique(pdb\$atom\$chain)

```
[1] "A" "B"
```

##Predicting functional motions of a single structure

Let's do a bioinformatics prediction of functional motions- i.e. the movements that one of these molecules needs to make to do its stuff.

```
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
```

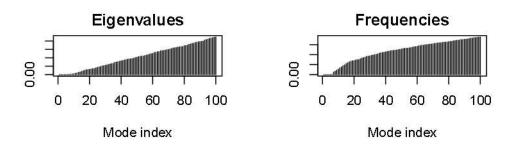
DELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDKI VGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIG YYSKEAEAGNTKYAKVDGTKPVAEVRADLEKILG

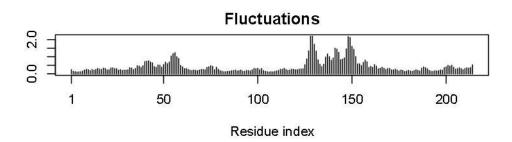
```
+ attr: atom, xyz, seqres, helix, sheet, calpha, remark, call
```

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

Building Hessian... Done in 0.03 seconds. Diagonalizing Hessian... Done in 0.25 seconds.

### plot(m)





Write out multi-model PDB file that we can use to make an animation of the predicted motions.

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

I can open this in Mol\* to play the trajectory...

##Comparative analysis of protein structures

### library(bio3d)

Here we will find and analyze all ADK structures in the PDB database.

We will start with a single database accession id: "lake\_A"

```
id <- "lake_A"
aa <- get.seq(id)</pre>
```

Warning in get.seq(id): Removing existing file: seqs.fasta

Fetching... Please wait. Done.

I ran these cmds in the R brain/console

install.packages("BiocManager") BiocManager::install("msa")

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

  The msa package is from BioConductor.
- Q13. How many amino acids are in this sequence, i.e. how long is this sequence? 214 positions.

#### length(aa)

[1] 3

#### attributes (aa)

\$names

[1] "id" "ali" "call"

\$class

[1] "fasta"

aa\$id

[1] "pdb | 1AKE | A"

#### aa\$ali

```
[,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13]
pdb|1AKE|A "M"
                          ^{\rm H}{
m I}{}^{\rm H}
                                 nIn
                                       ^{\rm II}L
                                             ^{\rm H}L^{\rm H}
                                                    ^{\rm H}G^{\rm H}
                                                          пАп
                                                                 пРп
                                                                               пАп
              [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21]
                                                                           [,22] [,23] [,24]
                                                                                   пKп
pdb | 1AKE | A "G"
                      пΤп
                             "Q"
                                             пQп
                                                    ηFη
                                                            πIπ
                                                                                          пүп
                                     пΑп
                                                                   пМп
                                                                           пЕп
              [,25] [,26] [,27] [,28] [,29] [,30] [,31] [,32]
                                                                           [,33]
                                                                                   [,34]
                                                                                          [,35]
pdb|1AKE|A "G"
                     пΙп
                             прп
                                     пQп
                                            πIπ
                                                    "S"
                                                            пТп
                                                                   ^{\rm H}{
m G}^{\rm H}
                                                                           пЪп
                                                                                   пМп
                                                                                          սԼո
              [,36] [,37] [,38] [,39] [,40] [,41] [,42] [,43] [,44] [,45] [,46]
```

```
\Pi \Lambda \Pi
                                                            ngn
                                                                     пGп
pdb | 1AKE | A "R"
                         "A"
                                 HAH
                                                   пKп
                                                                             ngn
                                                                                      ^{\rm H}E^{\rm H}
                                                                                               nŢ,n
                                                                                                       пGп
                [,47] [,48] [,49] [,50] [,51] [,52] [,53] [,54]
                                                                                      [,55] [,56] [,57]
pdb|1AKE|A "K"
                         "Q"
                                 ндп
                                          \pi K \pi
                                                   ^{\rm H}D^{\rm H}
                                                            пТп
                                                                     пМп
                                                                             ^{\rm H}{\rm D}^{\rm H}
                                                                                      пдп
                                                                                               пGп
                                                                                                       \pi K \pi
                [,58] [,59] [,60] [,61] [,62] [,63] [,64] [,65]
                                                                                      [,66] [,67] [,68]
pdb|1AKE|A "L"
                         uVu
                                 пТп
                                          ^{\rm H}{\rm D}^{\rm H}
                                                   пΕп
                                                            ^{\rm H}L^{\rm H}
                                                                    пУп
                                                                             пΙп
                                                                                      пап
                                                                                              ^{\rm n} {\rm L}^{\rm n}
                                                                                                       uvu
                [,69] [,70] [,71] [,72] [,73] [,74] [,75]
                                                                             [,76]
                                                                                      [,77] [,78] [,79]
pdb|1AKE|A "K"
                                 ^{\rm H}R^{\rm H}
                                          ^{\rm H} I ^{\rm H}
                                                                    ^{\rm H}\rm E^{\rm H}
                                                                             ^{\rm H}{\rm D}^{\rm H}
                                                                                      ^{\rm H}C^{\rm H}
                         ^{\rm H}E^{\rm H}
                                                   пΑп
                                                            пQп
                                                                                              ^{\rm II}R^{\rm II}
                                                                                                       пNп
                [,80] [,81] [,82] [,83] [,84] [,85] [,86]
                                                                             [,87]
                                                                                      [,88]
                                                                                              [,89] [,90]
pdb|1AKE|A "G"
                         пFп
                                 nt.n
                                          nt.n
                                                   ^{\rm H}D^{\rm H}
                                                            ^{\rm H}{\rm G}^{\rm H}
                                                                    nEn
                                                                             прп
                                                                                      пRп
                                                                                               пТп
                                                                                                       птп
                [,91] [,92] [,93] [,94] [,95] [,96] [,97] [,98]
                                                                                     [,99] [,100] [,101]
pdb|1AKE|A "P"
                                          \pi D \pi
                                                   ^{\Pi}A^{\Pi}
                                                            пМи
                                                                                              ^{\rm H}{\rm G}^{\rm H}
                         "Q"
                                 "A"
                                                                    ^{\rm II}{\rm K}^{\rm II}
                                                                             nEn
                                                                                      пдπ
                                                                                                         пIп
                [,102] [,103] [,104] [,105] [,106] [,107]
                                                                             [,108] [,109] [,110]
pdb|1AKE|A "N"
                          пVп
                                    ^{\rm H}{\rm D}^{\rm H}
                                              пүп
                                                         πVn
                                                                   ^{\rm n} {\rm L}^{\rm n}
                                                                             пЕп
                                                                                       "F"
                                                                                                 ^{\rm H}{\rm D}^{\rm H}
                [,111] [,112] [,113] [,114] [,115] [,116] [,117] [,118] [,119]
pdb|1AKE|A "V"
                          прп
                                    \pi D \pi
                                               пЕп
                                                         ^{\rm n}L^{\rm n}
                                                                   иТи
                                                                             пVп
                                                                                       пДп
                                                                                                 "R"
                [,120] [,121] [,122] [,123] [,124] [,125] [,126] [,127] [,128]
pdb | 1AKE | A "I"
                          пУп
                                    \pi G \pi
                                               "R"
                                                         "R"
                                                                   \Pi V \Pi
                                                                             пНп
                                                                                       пдп
                                                                                                 прп
                [,129] [,130] [,131] [,132] [,133] [,134] [,135] [,136] [,137]
pdb|1AKE|A "S"
                          ^{\mathrm{H}}\mathrm{G}^{\mathrm{H}}
                                    "R"
                                               пуп
                                                         пүп
                                                                   пНп
                                                                             пУп
                                                                                       \pi K \pi
                                                                                                 ^{\rm n}F^{\rm n}
                [,138] [,139] [,140] [,141] [,142] [,143] [,144] [,145] [,146]
pdb|1AKE|A "N"
                          прп
                                    прп
                                               пKп
                                                         пVп
                                                                   ^{\rm H}E^{\rm H}
                                                                             ^{\rm H}G^{\rm H}
                                                                                       \pi K \pi
                                                                                                 ^{\rm II}D^{\rm II}
                [,147] [,148] [,149] [,150]
                                                         [,151] [,152]
                                                                             [,153] [,154]
                                                                                                 [,155]
pdb|1AKE|A "D"
                                                                                       пТп
                          пVп
                                    пТп
                                               ^{\rm H}G^{\rm H}
                                                         ^{\rm n}E^{\rm n}
                                                                   "E"
                                                                             nŢ.n
                                                                                                 пТп
                [,156] [,157] [,158] [,159] [,160] [,161] [,162] [,163] [,164]
pdb|1AKE|A "R"
                          пKп
                                    пДп
                                               пDп
                                                         "Q"
                                                                   nEn
                                                                             nEn
                                                                                       птп
                                                                                                 пVп
                [,165] [,166] [,167] [,168] [,169] [,170] [,171] [,172] [,173]
pdb|1AKE|A "R"
                          _{\rm II} K_{\rm II}
                                    "R"
                                               ^{\rm H}L^{\rm H}
                                                         пVп
                                                                   nEn
                                                                             пүп
                                                                                       пНп
                                                                                                 "Q"
                [,174] [,175] [,176] [,177]
                                                        [,178] [,179] [,180] [,181] [,182]
pdb|1AKE|A "M"
                                                                                       пүп
                          пТп
                                    "A"
                                               при
                                                         nt.n
                                                                   птп
                                                                             пGп
                                                                                                 пүп
                [,183] [,184] [,185] [,186] [,187] [,188]
                                                                             [,189] [,190] [,191]
pdb|1AKE|A "S"
                          \pi K \pi
                                    nEn
                                               HAH
                                                         ^{\rm H}E^{\rm H}
                                                                   пдп
                                                                             иGи
                                                                                       пИп
                                                                                                 птп
                [,192] [,193] [,194] [,195] [,196] [,197] [,198] [,199] [,200]
pdb|1AKE|A "K"
                                                                                       иТи
                          пүп
                                    нди
                                               пКп
                                                         пVп
                                                                   прп
                                                                             ^{\rm H}G^{\rm H}
                                                                                                 \pi K \pi
                [,201] [,202] [,203] [,204] [,205] [,206] [,207] [,208] [,209]
pdb|1AKE|A "P"
                          uVu
                                               nEn
                                                         пVп
                                                                   "R"
                                                                             "A"
                                                                                       ^{\rm H}{\rm D}^{\rm H}
                                                                                                  սLս
                                    "A"
                [,210]
                          [,211]
                                    [,212]
                                              [,213]
                                                         [,214]
pdb | 1AKE | A "E"
                          _{\rm II}K_{\rm II}
                                    \pi I \pi
                                               nLn
                                                         пGп
```

#### aa\$call

read.fasta(file = outfile)

```
ncol(aa$ali)
[1] 214
#b <- blast.pdb(aa)
#hits <- plot(b)
#hits
#hits$pdb.id
#head(hits$pdb.id)
Pre-calculated results:
hits <- NULL
hits$pdb.id <- c('1AKE A', '6S36 A', '6RZE A', '3HPR A', '1E4V A', '5EJE A', '1E4Y A', '3X2S A', '6H
#Download related PDB files
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 exists. Skipping download
Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/6836.pdb exists. Skipping download
Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/6RZE.pdb exists. Skipping download
Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/3HPR.pdb exists. Skipping download
Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/1E4V.pdb exists. Skipping download
```

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

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

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

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

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

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

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

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



	6	54%
    	jë g	62%
  ===================================	Ĭ	69%
  ===================================		77%
 		85%
  ===================================	Î	92%
 		100%

Next we will use the pdbaln() function to align and also optionally fit (i.e. superpose) the identified PDB structures.

```
#Align releated PDBs
pdbs <- pdbaln(files, fit = TRUE, exefile="msa")</pre>
```

```
Reading PDB files:
pdbs/split_chain/1AKE_A.pdb
pdbs/split_chain/6S36_A.pdb
pdbs/split_chain/6RZE_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/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
. . .
```

16

#### 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/6836 A.pdb PDB has ALT records, taking A only, rm.alt=TRUE pdb/seq: 3 name: pdbs/split\_chain/6RZE\_A.pdb PDB has ALT records, taking A only, rm.alt=TRUE name: pdbs/split\_chain/3HPR\_A.pdb pdb/seq: 4 PDB has ALT records, taking A only, rm.alt=TRUE pdb/seq: 5 name: pdbs/split chain/1E4V A.pdb pdb/seq: 6 name: pdbs/split chain/5EJE A.pdb PDB has ALT records, taking A only, rm.alt=TRUE pdb/seq: 7 name: pdbs/split\_chain/1E4Y\_A.pdb pdb/seq: 8 name: pdbs/split\_chain/3X2S\_A.pdb pdb/seq: 9 name: pdbs/split\_chain/6HAP\_A.pdb pdb/seq: 10 name: pdbs/split chain/6HAM A.pdb PDB has ALT records, taking A only, rm.alt=TRUE pdb/seq: 11 name: pdbs/split chain/4K46 A.pdb PDB has ALT records, taking A only, rm.alt=TRUE pdb/seq: 12 name: pdbs/split\_chain/3GMT\_A.pdb pdb/seq: 13 name: pdbs/split\_chain/4PZL\_A.pdb

#### pdbs

[Truncated\_Name:1]1AKE\_A.pdb
[Truncated\_Name:2]6S36\_A.pdb
[Truncated\_Name:3]6RZE\_A.pdb
[Truncated\_Name:4]3HPR\_A.pdb
[Truncated\_Name:5]1E4V\_A.pdb
[Truncated\_Name:6]5EJE\_A.pdb
[Truncated\_Name:7]1E4Y\_A.pdb
[Truncated\_Name:8]3X2S\_A.pdb
[Truncated\_Name:9]6HAP\_A.pdb
[Truncated\_Name:10]6HAM\_A.pdb
[Truncated\_Name:11]4K46\_A.pdb
[Truncated\_Name:11]3GMT\_A.pdb
[Truncated\_Name:12]3GMT\_A.pdb

40 1 -MRIILLGAPGAGKGTQAQFIMEKYGIPQIS -MRIILLGAPGAGKGTQAQFIMEKYGIPQIS -MRIILLGAPGAGKGTQAQFIMEKYGIPQIS -MRIILLGAPGAGKGTQAQFIMEKYGIPQIS ---MRIILLGAPVAGKGTQAQFIMEKYGIPQIS -MRIILLGAPGAGKGTQAQFIMEKYGIPQIS ---MRIILLGALVAGKGTQAQFIMEKYGIPQIS ----MRIILLGAPGAGKGTQAQFIMEKYGIPQIS -MRIILLGAPGAGKGTQAQFIMEKYGIPQIS -MRIILLGAPGAGKGTQAQFIMEKYGIPQIS -MRIILLGAPGAGKGTQAQFIMAKFGIPQIS ----MRLILLGAPGAGKGTQANFIKEKFGIPQIS TENLYFQSNAMRIILLGAPGAGKGTQAKIIEQKYNIAHIS \*\*^\*\*\*\* \*\*\*\*\*\* \* 1 40

	41					80
[Truncated_Name:1]1AKE_A.pdb	TGDMLRA	AVKSGSE	LGKQAKDI	IMDAGKLVT	DELVIALV	<i>I</i> KE
[Truncated_Name:2]6836_A.pdb	TGDMLRA	AVKSGSE	LGKQAKDI	IMDAGKLVT	DELVIALV	/KE
[Truncated_Name:3]6RZE_A.pdb	TGDMLRA	AVKSGSE	LGKQAKDI	IMDAGKLVT	DELVIALV	/KE
[Truncated_Name: 4] 3HPR_A.pdb	TGDMLRA	AVKSGSE	LGKQAKDI	IMDAGKLVT	DELVIALV	/KE
[Truncated_Name:5]1E4V_A.pdb	TGDMLRA	AVKSGSE	LGKQAKDI	IMDAGKLVT	DELVIALV	/KE
[Truncated_Name: 6] 5EJE_A.pdb	TGDMLRA	AVKSGSE	LGKQAKDI	IMDACKLVT	DELVIALV	<i>I</i> KE
[Truncated_Name:7]1E4Y_A.pdb				IMDAGKLVT		
[Truncated_Name:8]3X2S_A.pdb				IMDCGKLVT		
[Truncated_Name:9]6HAP_A.pdb	TGDMLRA	AVKSGSE	LGKQAKDI	IMDAGKLVT	DELVIALV	/RE
[Truncated_Name: 10] 6HAM_A.pdb	TGDMLRA	AIKSGSE	LGKQAKDI	IMDAGKLVT	DEIIIALV	/KE
[Truncated_Name:11]4K46_A.pdb				VIDAGQLVS		
[Truncated_Name: 12] 3GMT_A.pdb				YMDEGKLVP		
[Truncated_Name:13]4PZL_A.pdb				VLDAGELVS		
	****^*	^* *^		^* **		k^^
	41	5% 5%	***			80
	81	20 30	÷			120
[Truncated_Name:1]1AKE_A.pdb	RIAQEDC	RNGFLLD	GFPRTIPO	QADAMKEAG	INVDYVLE	≟FD
[Truncated Name:2]6S36 A.pdb	RIAQEDC	RNGFLLD	GFPRTIPO	QADAMKEAG	INVDYVLE	≟FD
[Truncated_Name:3]6RZE_A.pdb	RIAQEDC	RNGFLLD	GFPRTIPO	QADAMKEAG	INVDYVLE	£FD
[Truncated_Name: 4] 3HPR_A.pdb	74 M			QADAMKEAG		
[Truncated_Name:5]1E4V_A.pdb				QADAMKEAG		
[Truncated_Name:6]5EJE_A.pdb	73			QADAMKEAG		
[Truncated_Name:7]1E4Y_A.pdb				QADAMKEAG		
[Truncated_Name:8]3X2S_A.pdb				QADAMKEAG		
[Truncated_Name:9]6HAP_A.pdb	Anno 111 - 112 - 1-20 -			QADAMKEAG		
[Truncated_Name: 10] 6HAM_A.pdb				QADAMKEAG		
[Truncated_Name:11]4K46_A.pdb				QADGLKEVG		
[Truncated_Name: 12] 3GMT_A.pdb				QADAMKEAG		
[Truncated_Name:13]4PZL_A.pdb				, QAQELDKLG		
	*^ *		* ***	150 H50 <sub>10</sub> .	^ ^**^^	
	81	200 200				120
	121		Ÿ			160
[Truncated_Name:1]1AKE_A.pdb	VPDELIV	DRIVGRR	VHAPSGRV	VYHVKFNPP	KVEGKDDV	/TG
[Truncated_Name:2]6S36_A.pdb	VPDELIV	DKIVGRR	VHAPSGRV	VYHVKFNPP	KVEGKDDV	/TG
[Truncated_Name:3]6RZE_A.pdb	VPDELIV	DAIVGRR	VHAPSGRV	VYHVKFNPP	KVEGKDDV	/TG
[Truncated_Name: 4] 3HPR_A.pdb	VPDELIV	DRIVGRR	VHAPSGRV	VYHVKFNPP	KVEGKDD(	GTG
[Truncated_Name:5]1E4V_A.pdb	VPDELIV	DRIVGRR	VHAPSGRV	VYHVKFNPP	KVEGKDDV	/TG
[Truncated_Name:6]5EJE_A.pdb				VYHVKFNPP		
[Truncated_Name:7]1E4Y_A.pdb				VYHVKFNPP		

[Truncated_Name:8]3X2S_A.pdb [Truncated_Name:9]6HAP_A.pdb [Truncated_Name:10]6HAM_A.pdb [Truncated_Name:11]4K46_A.pdb [Truncated_Name:12]3GMT_A.pdb [Truncated_Name:13]4PZL_A.pdb	VADSVIVERMAGRRAHLASGRTYHNVYNPPKVEGKDDVTG VPFSEIIERMSGRRTHPASGRTYHVKFNPPKVEGKDDVTG
[Truncated_Name:1]1AKE_A.pdb [Truncated_Name:2]6S36_A.pdb [Truncated_Name:3]6RZE_A.pdb [Truncated_Name:4]3HPR_A.pdb [Truncated_Name:5]1E4V_A.pdb [Truncated_Name:6]5EJE_A.pdb [Truncated_Name:7]1E4Y_A.pdb [Truncated_Name:8]3X2S_A.pdb [Truncated_Name:9]6HAP_A.pdb [Truncated_Name:10]6HAM_A.pdb [Truncated_Name:11]4K46_A.pdb [Truncated_Name:12]3GMT_A.pdb [Truncated_Name:13]4PZL_A.pdb	EDLVIREDDKEETVLARLGVYHNQTAPLIAYYGKEAEAGN EPLVQRDDDKEETVKKRLDVYEAQTKPLITYYGDWARRGA EPLITRTDDNEDTVKQRLSVYHAQTAKLIDFYRNFSSTNT * * * * * * * * * * * * * * *
[Truncated_Name:1]1AKE_A.pdb [Truncated_Name:2]6S36_A.pdb [Truncated_Name:3]6RZE_A.pdb [Truncated_Name:4]3HPR_A.pdb [Truncated_Name:5]1E4V_A.pdb [Truncated_Name:6]5EJE_A.pdb [Truncated_Name:7]1E4Y_A.pdb [Truncated_Name:8]3X2S_A.pdb [Truncated_Name:9]6HAP_A.pdb [Truncated_Name:10]6HAM_A.pdb [Truncated_Name:11]4K46_A.pdb [Truncated_Name:12]3GMT_A.pdb [Truncated_Name:13]4PZL_A.pdb	TQYLKFDGTKAVAEVSAELEKALA- ENGLKAPAYRKISG-

```
Call:
   pdbaln(files = files, fit = TRUE, exefile = "msa")

Class:
   pdbs, fasta

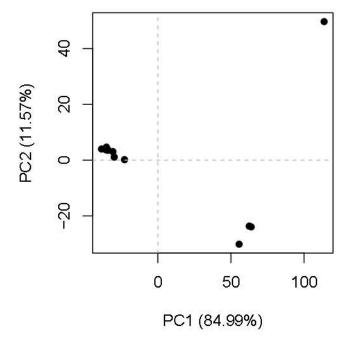
Alignment dimensions:
   13 sequence rows; 227 position columns (204 non-gap, 23 gap)

+ attr: xyz, resno, b, chain, id, ali, resid, sse, call

ids <- basename.pdb(pdbs$id)</pre>
```

# **Principal Component Analysis**

```
#Perform PCA
pc.xray <- pca(pdbs)
plot(pc.xray, pc.axes = c(1,2))</pre>
```

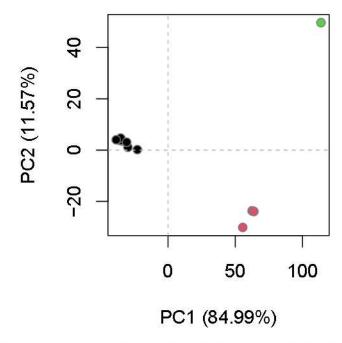


```
# Calculate RMSD
rd <- rmsd(pdbs)
```

Warning in rmsd(pdbs): No indices provided, using the 204 non NA positions

```
# Structure-based clustering
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)</pre>
```



To visualize the major structural variations in the ensemble the function mktrj() can be used to generate a trajectory PDB file by interpolating along a give PC (eigenvector):

```
pc1 <- mktrj(pc.xray, pc=1, file="pc_1.pdb")

#Plotting results with ggplot2
library(ggplot2)
library(ggrepel)

df <- data.frame(PC1=pc.xray$z[,1],</pre>
```

