class 10 Comparative analysis of structure

Ruofan Kang (A17236920)

We need some packages for todays class. These include 'bio3d' and 'msa'. The 'mea'package is from BioConductor. These packages focus on genomics type work and are managed by the 'BiocManager'package.

Install 'install.packages("BiocManager")' and then 'BiocManager::install("msa")' all entered in the R "brain" console.

```
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
```

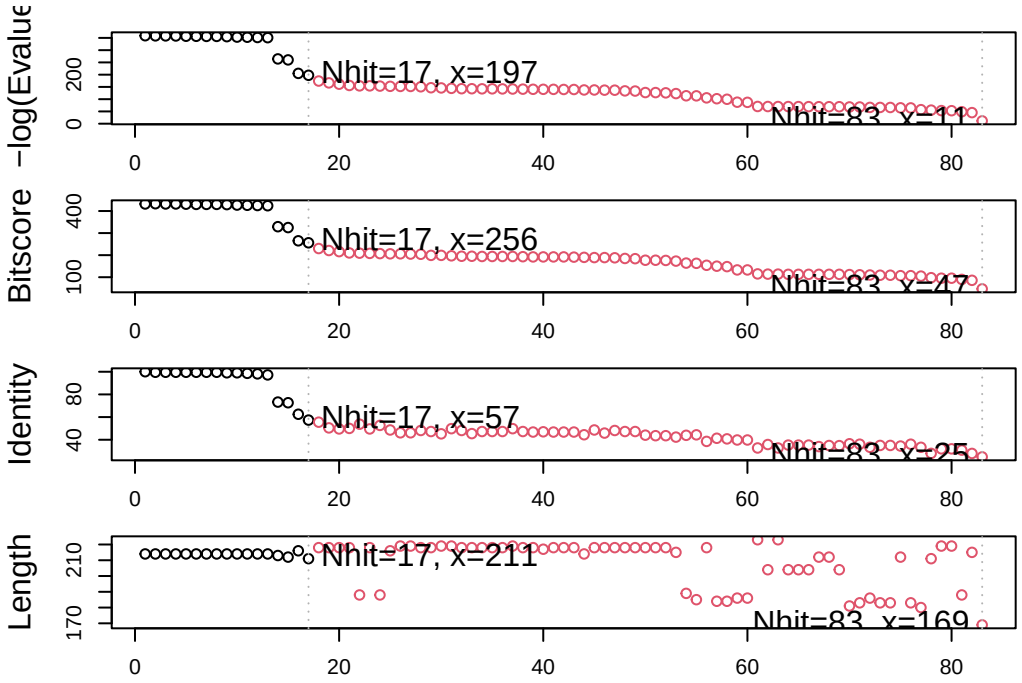
Now I can search the PDB databse for related sequences:

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

```
Searching ... please wait (updates every 5 seconds) RID = N4P6N8AS016
.......................................
Reporting 83 hits
```

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

```
 * Possible cutoff values:    197 11
          Yielding Nhits:    17 83

 * Chosen cutoff value of:    197
          Yielding Nhits:    17
```




```
attributes(b)
```

```
$names
[1] "hit.tbl" "raw"     "url"

$class
[1] "blast"
```

```
head(b$hit.tbl)
```

```
      queryid subjectids identity alignmentlength mismatches gapopens q.start
1 Query_63445     1AKE_A  100.000             214          0        0       1
2 Query_63445     8BQF_A   99.533             214          1        0       1
3 Query_63445     4X8M_A   99.533             214          1        0       1
4 Query_63445     6S36_A   99.533             214          1        0       1
5 Query_63445     6RZE_A   99.533             214          1        0       1
6 Query_63445     4X8H_A   99.533             214          1        0       1
  q.end s.start s.end    evalue bitscore positives mlog.evalue pdb.id    acc
1   214       1   214 1.46e-156      432    100.00    358.8248 1AKE_A 1AKE_A
2   214      21   234 2.38e-156      433    100.00    358.3362 8BQF_A 8BQF_A
```

```
3   214       1   214 2.60e-156      432    100.00    358.2478 4X8M_A 4X8M_A
4   214       1   214 3.82e-156      432    100.00    357.8630 6S36_A 6S36_A
5   214       1   214 1.10e-155      431     99.53    356.8054 6RZE_A 6RZE_A
6   214       1   214 1.44e-155      430     99.53    356.5360 4X8H_A 4X8H_A
```

There are the related structures in the PDB database that we found via a BLAST search…

```
hits$pdb.id
```

```
 [1] "1AKE_A" "8BQF_A" "4X8M_A" "6S36_A" "6RZE_A" "4X8H_A" "3HPR_A" "1E4V_A"
 [9] "5EJE_A" "1E4Y_A" "3X2S_A" "6HAP_A" "6HAM_A" "4K46_A" "4NP6_A" "3GMT_A"
[17] "4PZL_A"
```

Side-note: Lets annotate these structures (in other words find out what thet are, what soeies they are from, stuff about the experiment they were solved in etc.)

For this we can use the 'pdb.annotate()'

```
anno <- pdb.annotate(hits$pdb.id)
```

```
#attributes(anno)
head(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
6RZE_A        6RZE       A           Protein         214                 X-ray
4X8H_A        4X8H       A           Protein         214                 X-ray
       resolution       scopDomain                                      pfam
1AKE_A       2.00 Adenylate kinase Adenylate kinase, active site lid (ADK_lid)
8BQF_A       2.05             <NA> Adenylate kinase, active site lid (ADK_lid)
4X8M_A       2.60             <NA> Adenylate kinase, active site lid (ADK_lid)
6S36_A       1.60             <NA> Adenylate kinase, active site lid (ADK_lid)
6RZE_A       1.69             <NA> Adenylate kinase, active site lid (ADK_lid)
4X8H_A       2.50             <NA> Adenylate kinase, active site lid (ADK_lid)
              ligandId                                    ligandName
1AKE_A             AP5             BIS(ADENOSINE)-5'-PENTAPHOSPHATE
8BQF_A             AP5             BIS(ADENOSINE)-5'-PENTAPHOSPHATE
4X8M_A            <NA>                                         <NA>
```

```
6S36_A CL (3),NA,MG (2) CHLORIDE ION (3),SODIUM ION,MAGNESIUM ION (2)
6RZE_A    NA (3),CL (2)              SODIUM ION (3),CHLORIDE ION (2)
4X8H_A             <NA>                                      <NA>
                source
1AKE_A Escherichia coli
8BQF_A Escherichia coli
4X8M_A Escherichia coli
6S36_A Escherichia coli
6RZE_A Escherichia coli
4X8H_A Escherichia coli

1AKE_A STRUCTURE OF THE COMPLEX BETWEEN ADENYLATE KINASE FROM ESCHERICHIA COLI AND THE INHIBI
8BQF_A
4X8M_A
6S36_A
6RZE_A
4X8H_A
                                                    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
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
         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
6RZE_A 0.18190    C 1 2 1
4X8H_A 0.19140    C 1 2 1
```

Now we can download all these structures for further analysis with the'get.pdb()' function.

```
# Download releated 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.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/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%
  |
  |====                                                                  |   6%
  |
  |========                                                              |  12%
  |
  |============                                                          |  18%
  |
  |================                                                      |  24%
  |
  |====================                                                  |  29%
  |
  |========================                                              |  35%
  |
  |============================                                          |  41%
  |
  |================================                                      |  47%
  |
  |====================================                                  |  53%
  |
  |========================================                              |  59%
  |
  |============================================                          |  65%
  |
  |================================================                      |  71%
  |
  |====================================================                  |  76%
  |
  |============================================================          |  82%
  |
  |================================================================      |  88%
  |
  |====================================================================  |  94%
  |
```

```
  |======================================================================| 100%
```

Now we have all these related structures we can Align and Supperpose...

```
# Align releated PDBs
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/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
....

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/6RZE_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 6   name: pdbs/split_chain/4X8H_A.pdb
pdb/seq: 7   name: pdbs/split_chain/3HPR_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 8   name: pdbs/split_chain/1E4V_A.pdb
pdb/seq: 9   name: pdbs/split_chain/5EJE_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 10   name: pdbs/split_chain/1E4Y_A.pdb
pdb/seq: 11   name: pdbs/split_chain/3X2S_A.pdb
pdb/seq: 12   name: pdbs/split_chain/6HAP_A.pdb
pdb/seq: 13   name: pdbs/split_chain/6HAM_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 14   name: pdbs/split_chain/4K46_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 15   name: pdbs/split_chain/4NP6_A.pdb
pdb/seq: 16   name: pdbs/split_chain/3GMT_A.pdb
pdb/seq: 17   name: pdbs/split_chain/4PZL_A.pdb
```

```
pdbs
```

```
                                1        .        .        .         40
[Truncated_Name:1]1AKE_A.pdb    ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:2]8BQF_A.pdb    ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:3]4X8M_A.pdb    ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:4]6S36_A.pdb    ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:5]6RZE_A.pdb    ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:6]4X8H_A.pdb    ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:7]3HPR_A.pdb    ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:8]1E4V_A.pdb    ----------MRIILLGAPVAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:9]5EJE_A.pdb    ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:10]1E4Y_A.pdb   ----------MRIILLGALVAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:11]3X2S_A.pdb   ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:12]6HAP_A.pdb   ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:13]6HAM_A.pdb   ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:14]4K46_A.pdb   ----------MRIILLGAPGAGKGTQAQFIMAKFGIPQIS
[Truncated_Name:15]4NP6_A.pdb   --------NAMRIILLGAPGAGKGTQAQFIMEKFGIPQIS
[Truncated_Name:16]3GMT_A.pdb   ----------MRLILLGAPGAGKGTQANFIKEKFGIPQIS
[Truncated_Name:17]4PZL_A.pdb   TENLYFQSNAMRIILLGAPGAGKGTQAKIIEQKYNIAHIS
```

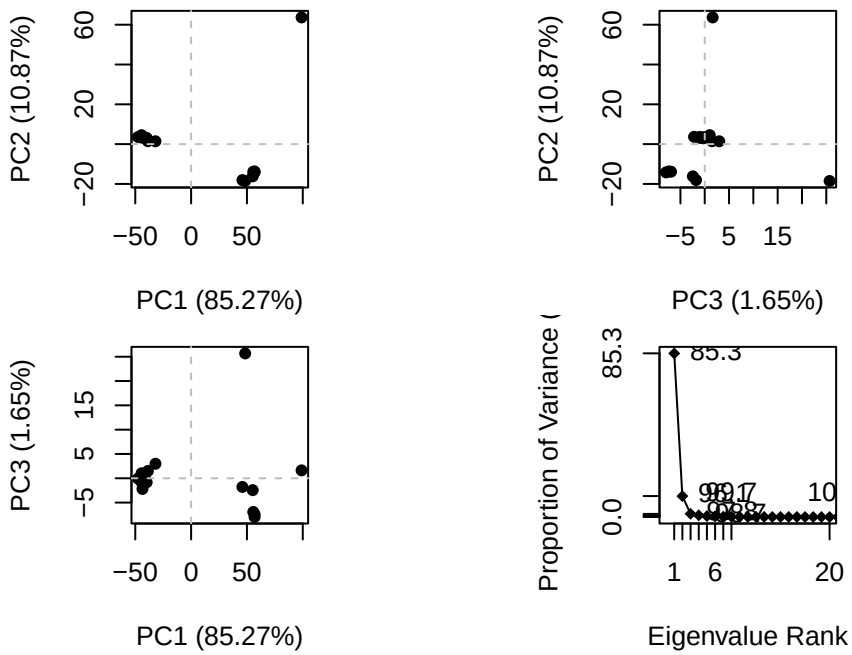
```
                                  **^*****  *******  *  *^ *  **
                                1         .         .         .        40

                               41         .         .         .        80
[Truncated_Name:1]1AKE_A.pdb    TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:2]8BQF_A.pdb    TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:3]4X8M_A.pdb    TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:4]6S36_A.pdb    TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:5]6RZE_A.pdb    TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:6]4X8H_A.pdb    TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:7]3HPR_A.pdb    TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:8]1E4V_A.pdb    TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:9]5EJE_A.pdb    TGDMLRAAVKSGSELGKQAKDIMDACKLVTDELVIALVKE
[Truncated_Name:10]1E4Y_A.pdb   TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:11]3X2S_A.pdb   TGDMLRAAVKSGSELGKQAKDIMDCGKLVTDELVIALVKE
[Truncated_Name:12]6HAP_A.pdb   TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVRE
[Truncated_Name:13]6HAM_A.pdb   TGDMLRAAIKSGSELGKQAKDIMDAGKLVTDEIIIALVKE
[Truncated_Name:14]4K46_A.pdb   TGDMLRAAIKAGTELGKQAKSVIDAGQLVSDDIILGLVKE
[Truncated_Name:15]4NP6_A.pdb   TGDMLRAAIKAGTELGKQAKAVIDAGQLVSDDIILGLIKE
[Truncated_Name:16]3GMT_A.pdb   TGDMLRAAVKAGTPLGVEAKTYMDEGKLVPDSLIIGLVKE
[Truncated_Name:17]4PZL_A.pdb   TGDMIRETIKSGSALGQELKKVLDAGELVSDEFIIKIVKD
                                ****^*  ^* *^ **   *  ^*   ** *  ^^ ^^^^
                               41         .         .         .        80

                               81         .         .         .        120
[Truncated_Name:1]1AKE_A.pdb    RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:2]8BQF_A.pdb    RIAQE----GFLLDGFPRTIPQADAMKEAGINVDYVIEFD
[Truncated_Name:3]4X8M_A.pdb    RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:4]6S36_A.pdb    RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:5]6RZE_A.pdb    RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:6]4X8H_A.pdb    RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:7]3HPR_A.pdb    RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:8]1E4V_A.pdb    RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:9]5EJE_A.pdb    RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:10]1E4Y_A.pdb   RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:11]3X2S_A.pdb   RIAQEDSRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:12]6HAP_A.pdb   RICQEDSRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:13]6HAM_A.pdb   RICQEDSRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:14]4K46_A.pdb   RIAQDDCAKGFLLDGFPRTIPQADGLKEVGVVVDYVIEFD
[Truncated_Name:15]4NP6_A.pdb   RIAQADCEKGFLLDGFPRTIPQADGLKEMGINVDYVIEFD
[Truncated_Name:16]3GMT_A.pdb   RLKEADCANGYLFDGFPRTIAQADAMKEAGVAIDYVLEID
[Truncated_Name:17]4PZL_A.pdb   RISKNDCNNGFLLDGVPRTIPQAQELDKLGVNIDYIVEVD
                                *^       *^* ** **** **  ^   *^ ^**^^* *
```

```
                                81        .         .         .         120

                               121        .         .         .         160
[Truncated_Name:1]1AKE_A.pdb     VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:2]8BQF_A.pdb     VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:3]4X8M_A.pdb     VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:4]6S36_A.pdb     VPDELIVDKIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:5]6RZE_A.pdb     VPDELIVDAIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:6]4X8H_A.pdb     VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:7]3HPR_A.pdb     VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDGTG
[Truncated_Name:8]1E4V_A.pdb     VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:9]5EJE_A.pdb     VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:10]1E4Y_A.pdb    VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:11]3X2S_A.pdb    VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:12]6HAP_A.pdb    VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:13]6HAM_A.pdb    VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:14]4K46_A.pdb    VADSVIVERMAGRRAHLASGRTYHNVYNPPKVEGKDDVTG
[Truncated_Name:15]4NP6_A.pdb    VADDVIVERMAGRRAHLPSGRTYHVVYNPPKVEGKDDVTG
[Truncated_Name:16]3GMT_A.pdb    VPFSEIIERMSGRRTHPASGRTYHVKFNPPKVEGKDDVTG
[Truncated_Name:17]4PZL_A.pdb    VADNLLIERITGRRIHPASGRTYHTKFNPPKVADKDDVTG
                                 *    ^^^ ^ *** *  *** **  ^*****  *** **
                               121        .         .         .         160

                               161        .         .         .         200
[Truncated_Name:1]1AKE_A.pdb     EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:2]8BQF_A.pdb     EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:3]4X8M_A.pdb     EELTTRKDDQEETVRKRLVEWHQMTAPLIGYYSKEAEAGN
[Truncated_Name:4]6S36_A.pdb     EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:5]6RZE_A.pdb     EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:6]4X8H_A.pdb     EELTTRKDDQEETVRKRLVEYHQMTAALIGYYSKEAEAGN
[Truncated_Name:7]3HPR_A.pdb     EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:8]1E4V_A.pdb     EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:9]5EJE_A.pdb     EELTTRKDDQEECVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:10]1E4Y_A.pdb    EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:11]3X2S_A.pdb    EELTTRKDDQEETVRKRLCEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:12]6HAP_A.pdb    EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:13]6HAM_A.pdb    EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:14]4K46_A.pdb    EDLVIREDDKEETVLARLGVYHNQTAPLIAYYGKEAEAGN
[Truncated_Name:15]4NP6_A.pdb    EDLVIREDDKEETVRARLNVYHTQTAPLIEYYGKEAAAGK
[Truncated_Name:16]3GMT_A.pdb    EPLVQRDDDKEETVKKRLDVYEAQTKPLITYYGDWARRGA
[Truncated_Name:17]4PZL_A.pdb    EPLITRTDDNEDTVKQRLSVYHAQTAKLIDFYRNFSSTNT
                                 * *  * ** *^ *  **  ^   *  ** ^*
                               161        .         .         .         200
```

```
                              201        .         .      227
[Truncated_Name:1]1AKE_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:2]8BQF_A.pdb    T--KYAKVDGTKPVAEVRADLEKIL--
[Truncated_Name:3]4X8M_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:4]6S36_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:5]6RZE_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:6]4X8H_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:7]3HPR_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:8]1E4V_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:9]5EJE_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:10]1E4Y_A.pdb   T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:11]3X2S_A.pdb   T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:12]6HAP_A.pdb   T--KYAKVDGTKPVCEVRADLEKILG-
[Truncated_Name:13]6HAM_A.pdb   T--KYAKVDGTKPVCEVRADLEKILG-
[Truncated_Name:14]4K46_A.pdb   T--QYLKFDGTKAVAEVSAELEKALA-
[Truncated_Name:15]4NP6_A.pdb   T--QYLKFDGTKQVSEVSADIAKALA-
[Truncated_Name:16]3GMT_A.pdb   E-------NGLKAPA-----YRKISG-
[Truncated_Name:17]4PZL_A.pdb   KIPKYIKINGDQAVEKVSQDIFDQLNK
                                         *
                              201        .         .      227

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

Class:
  pdbs, fasta

Alignment dimensions:
  17 sequence rows; 227 position columns (199 non-gap, 28 gap)

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

Principal Cpmponent Alalysis

```
# Perform PCA
pc.xray <- pca(pdbs)
plot(pc.xray)
```




```
# Visualize first principal component
pc1 <- mktrj(pc.xray, pc=1, file="pc_1.pdb")
```

Untitled

Ruofan Kang (A17236920)

2023-11-13

Interpreting Results

HIV_pr

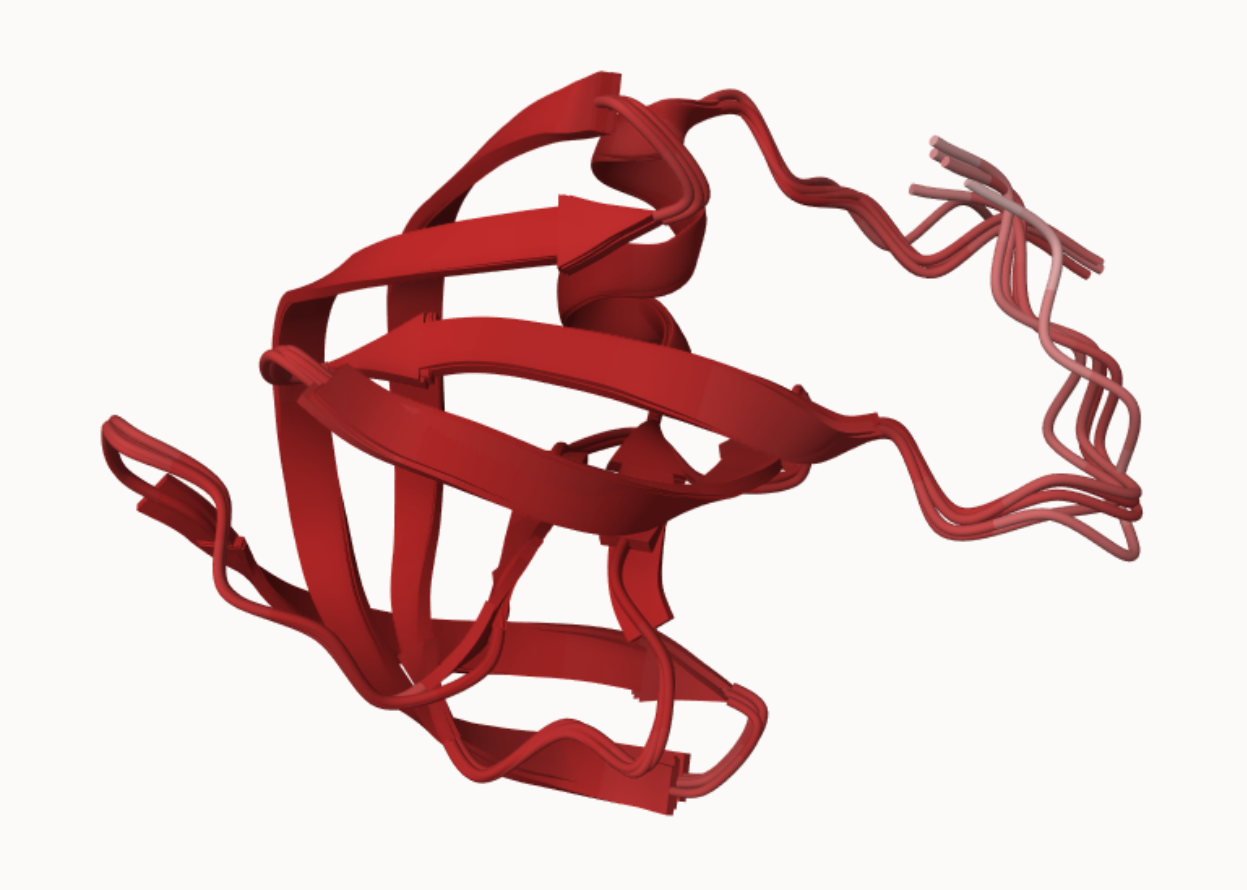


Figure 1: HIVPRDIMER

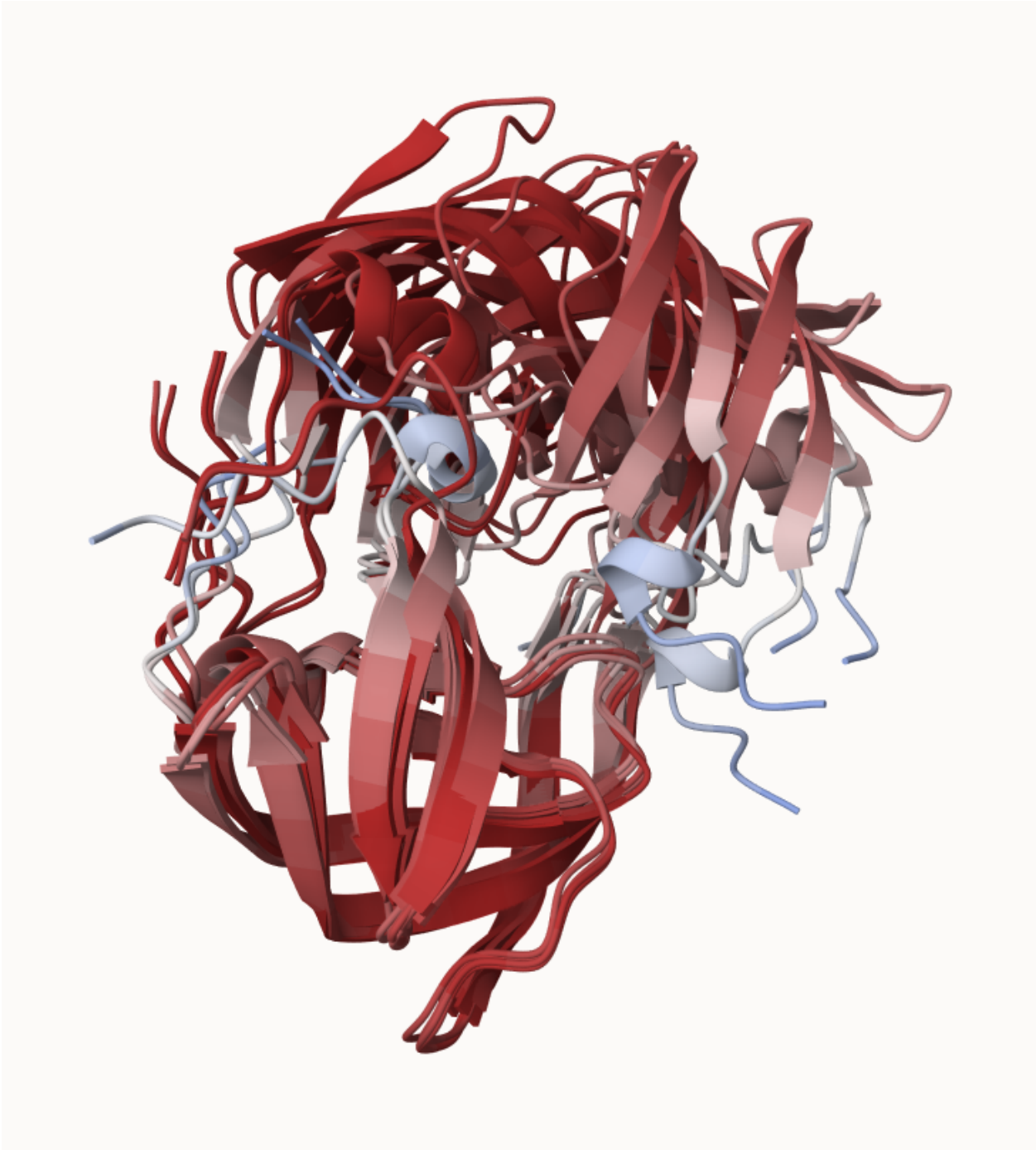


Figure 2: HIVPRDIMER

HIV_pr_dimer

8. Custom analysis of resulting models

```
# Change this for YOUR results dir name
results_dir <- "hivprdimer_23119/"
```

```
# File names for all PDB models
pdb_files <- list.files(path=results_dir,
                        pattern="*.pdb",
                        full.names = TRUE)

# Print our PDB file names
basename(pdb_files)
```

```
## [1] "HIVprDimer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_1_seed_000.pdb"
## [2] "HIVprDimer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000.pdb"
## [3] "HIVprDimer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000.pdb"
## [4] "HIVprDimer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_2_seed_000.pdb"
## [5] "HIVprDimer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_3_seed_000.pdb"
```

```
library(bio3d)

# Read all data from Models
#  and superpose/fit coords
pdbs <- pdbaln(pdb_files, fit=TRUE, exefile="msa")
```

```
## Reading PDB files:
## hivprdimer_23119/HIVprDimer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_1_seed_000.pdb
## hivprdimer_23119/HIVprDimer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000.pdb
## hivprdimer_23119/HIVprDimer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000.pdb
## hivprdimer_23119/HIVprDimer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_2_seed_000.pdb
## hivprdimer_23119/HIVprDimer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_3_seed_000.pdb
## .....
##
## Extracting sequences
##
## pdb/seq: 1   name: hivprdimer_23119/HIVprDimer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model
## pdb/seq: 2   name: hivprdimer_23119/HIVprDimer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model
## pdb/seq: 3   name: hivprdimer_23119/HIVprDimer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model
## pdb/seq: 4   name: hivprdimer_23119/HIVprDimer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model
## pdb/seq: 5   name: hivprdimer_23119/HIVprDimer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model
```

```
pdbs
```

```
##                                 1        .        .        .        .         50
## [Truncated_Name:1]HIVprDimer   PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
## [Truncated_Name:2]HIVprDimer   PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
## [Truncated_Name:3]HIVprDimer   PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
## [Truncated_Name:4]HIVprDimer   PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
```

```
## [Truncated_Name:5]HIVprDimer   PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
##                                **************************************************
##                                1        .         .         .         .         50
##
##                               51        .         .         .         .         100
## [Truncated_Name:1]HIVprDimer   GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
## [Truncated_Name:2]HIVprDimer   GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
## [Truncated_Name:3]HIVprDimer   GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
## [Truncated_Name:4]HIVprDimer   GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
## [Truncated_Name:5]HIVprDimer   GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
##                                **************************************************
##                               51        .         .         .         .         100
##
##                              101        .         .         .         .         150
## [Truncated_Name:1]HIVprDimer   QITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIG
## [Truncated_Name:2]HIVprDimer   QITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIG
## [Truncated_Name:3]HIVprDimer   QITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIG
## [Truncated_Name:4]HIVprDimer   QITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIG
## [Truncated_Name:5]HIVprDimer   QITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIG
##                                **************************************************
##                              101        .         .         .         .         150
##
##                              151        .         .         .         .       198
## [Truncated_Name:1]HIVprDimer   GFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
## [Truncated_Name:2]HIVprDimer   GFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
## [Truncated_Name:3]HIVprDimer   GFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
## [Truncated_Name:4]HIVprDimer   GFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
## [Truncated_Name:5]HIVprDimer   GFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
##                                *********************************************
##                              151        .         .         .         .       198
##
## Call:
##   pdbaln(files = pdb_files, fit = TRUE, exefile = "msa")
##
## Class:
##   pdbs, fasta
##
## Alignment dimensions:
##   5 sequence rows; 198 position columns (198 non-gap, 0 gap)
##
## + attr: xyz, resno, b, chain, id, ali, resid, sse, call
```

```
rd <- rmsd(pdbs, fit=T)
```

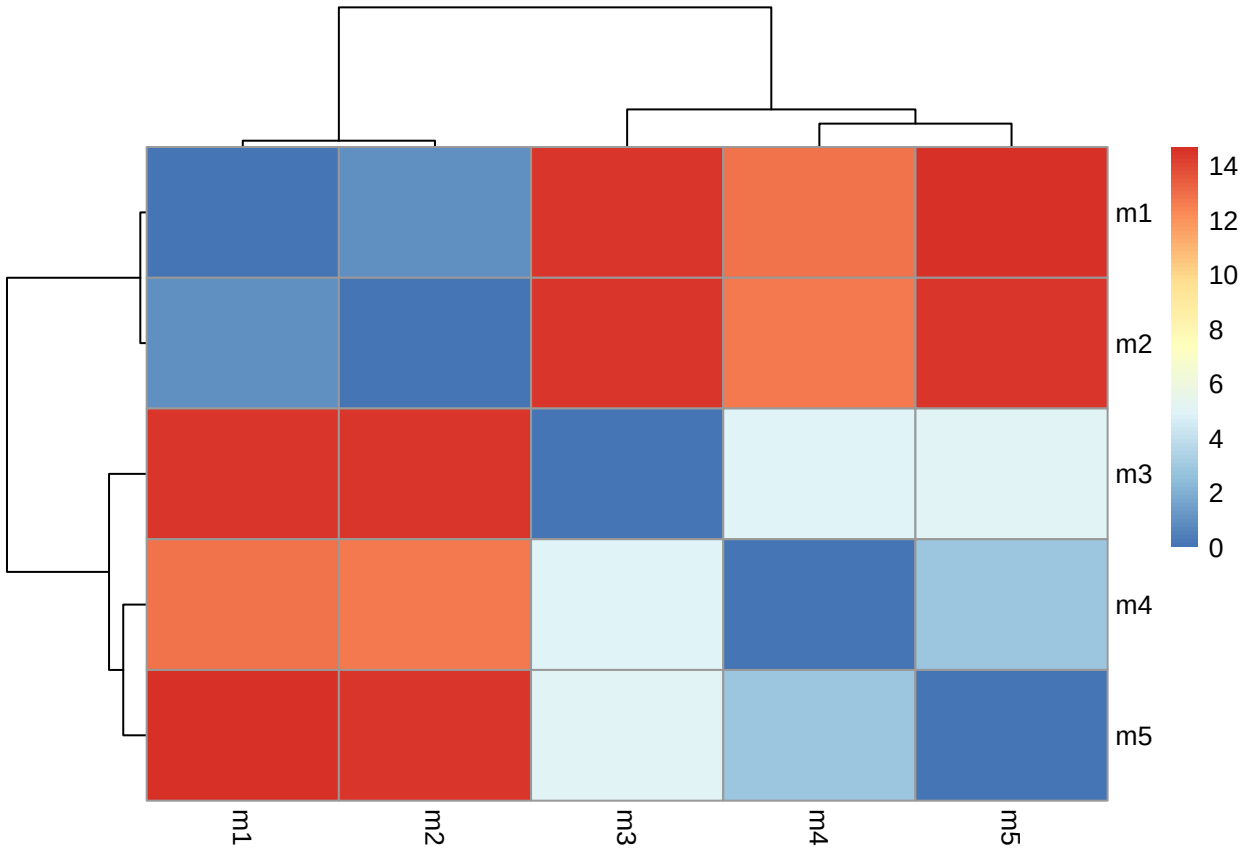
```
## Warning in rmsd(pdbs, fit = T): No indices provided, using the 198 non NA positions
```

```
range(rd)
```

```
## [1]  0.000 14.689
```

```
library(pheatmap)
```

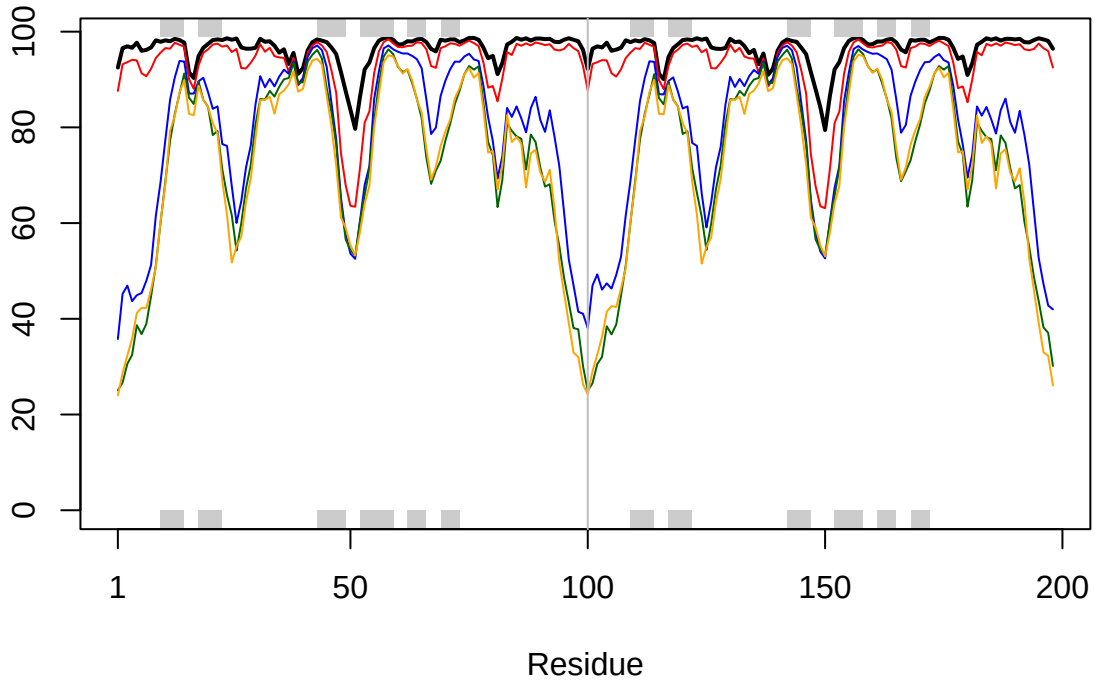
```
colnames(rd) <- paste0("m",1:5)
rownames(rd) <- paste0("m",1:5)
pheatmap(rd)
```




```
# Read a reference PDB structure
pdb <- read.pdb("1hsg")
```

```
##   Note: Accessing on-line PDB file
```

```
plotb3(pdbs$b[1,], typ="l", lwd=2, sse=pdb)
points(pdbs$b[2,], typ="l", col="red")
points(pdbs$b[3,], typ="l", col="blue")
points(pdbs$b[4,], typ="l", col="darkgreen")
points(pdbs$b[5,], typ="l", col="orange")
abline(v=100, col="gray")
```




```
core <- core.find(pdbs)
```

```
##  core size 197 of 198  vol = 6066.233
##  core size 196 of 198  vol = 5264.821
##  core size 195 of 198  vol = 4843.551
##  core size 194 of 198  vol = 4525.203
##  core size 193 of 198  vol = 4270.845
##  core size 192 of 198  vol = 4025.976
##  core size 191 of 198  vol = 3813.952
##  core size 190 of 198  vol = 3643.35
##  core size 189 of 198  vol = 3488.401
##  core size 188 of 198  vol = 3321.841
##  core size 187 of 198  vol = 3190.327
##  core size 186 of 198  vol = 3089.083
##  core size 185 of 198  vol = 2996.516
##  core size 184 of 198  vol = 2935.398
##  core size 183 of 198  vol = 2875.538
##  core size 182 of 198  vol = 2821.483
##  core size 181 of 198  vol = 2767.813
##  core size 180 of 198  vol = 2725.993
##  core size 179 of 198  vol = 2692.74
##  core size 178 of 198  vol = 2682.168
##  core size 177 of 198  vol = 2689.154
##  core size 176 of 198  vol = 2703.881
##  core size 175 of 198  vol = 2716.139
```

```
##  core size 174 of 198  vol = 2744.776
##  core size 173 of 198  vol = 2761.593
##  core size 172 of 198  vol = 2748.261
##  core size 171 of 198  vol = 2736.041
##  core size 170 of 198  vol = 2719.659
##  core size 169 of 198  vol = 2694.179
##  core size 168 of 198  vol = 2658.705
##  core size 167 of 198  vol = 2616.718
##  core size 166 of 198  vol = 2547.711
##  core size 165 of 198  vol = 2481.8
##  core size 164 of 198  vol = 2427.951
##  core size 163 of 198  vol = 2355.01
##  core size 162 of 198  vol = 2278.031
##  core size 161 of 198  vol = 2205.985
##  core size 160 of 198  vol = 2143.334
##  core size 159 of 198  vol = 2077.427
##  core size 158 of 198  vol = 1996.191
##  core size 157 of 198  vol = 1919.247
##  core size 156 of 198  vol = 1860.365
##  core size 155 of 198  vol = 1801.404
##  core size 154 of 198  vol = 1746.289
##  core size 153 of 198  vol = 1689.186
##  core size 152 of 198  vol = 1623.896
##  core size 151 of 198  vol = 1568.913
##  core size 150 of 198  vol = 1503.345
##  core size 149 of 198  vol = 1453.429
##  core size 148 of 198  vol = 1403.326
##  core size 147 of 198  vol = 1359.356
##  core size 146 of 198  vol = 1308.036
##  core size 145 of 198  vol = 1263.862
##  core size 144 of 198  vol = 1221.11
##  core size 143 of 198  vol = 1171.505
##  core size 142 of 198  vol = 1124.942
##  core size 141 of 198  vol = 1087.374
##  core size 140 of 198  vol = 1037.898
##  core size 139 of 198  vol = 991.863
##  core size 138 of 198  vol = 962.958
##  core size 137 of 198  vol = 930.192
##  core size 136 of 198  vol = 889.362
##  core size 135 of 198  vol = 848.906
##  core size 134 of 198  vol = 808.421
##  core size 133 of 198  vol = 768.835
##  core size 132 of 198  vol = 727.114
##  core size 131 of 198  vol = 688.64
##  core size 130 of 198  vol = 651.078
##  core size 129 of 198  vol = 608.118
##  core size 128 of 198  vol = 577.423
##  core size 127 of 198  vol = 548.005
##  core size 126 of 198  vol = 523.303
##  core size 125 of 198  vol = 497.184
##  core size 124 of 198  vol = 476.785
##  core size 123 of 198  vol = 443.582
##  core size 122 of 198  vol = 412.39
##  core size 121 of 198  vol = 386.325
```

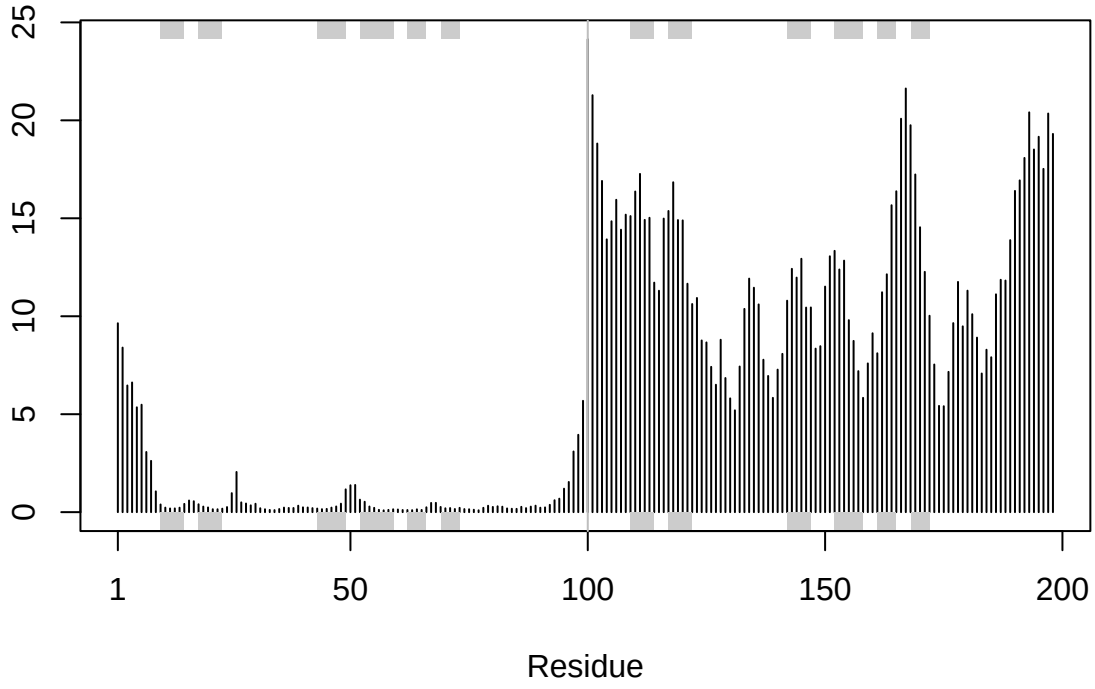
```
##  core size 120 of 198  vol = 353.601
##  core size 119 of 198  vol = 334.355
##  core size 118 of 198  vol = 312.861
##  core size 117 of 198  vol = 291.346
##  core size 116 of 198  vol = 272.077
##  core size 115 of 198  vol = 248.387
##  core size 114 of 198  vol = 232.533
##  core size 113 of 198  vol = 218.326
##  core size 112 of 198  vol = 202.098
##  core size 111 of 198  vol = 187.139
##  core size 110 of 198  vol = 172.408
##  core size 109 of 198  vol = 163.544
##  core size 108 of 198  vol = 153.544
##  core size 107 of 198  vol = 143.197
##  core size 106 of 198  vol = 134.351
##  core size 105 of 198  vol = 126.065
##  core size 104 of 198  vol = 118.219
##  core size 103 of 198  vol = 109.086
##  core size 102 of 198  vol = 99.658
##  core size 101 of 198  vol = 89.865
##  core size 100 of 198  vol = 79.565
##  core size 99 of 198  vol = 68.357
##  core size 98 of 198  vol = 61.965
##  core size 97 of 198  vol = 56.477
##  core size 96 of 198  vol = 46.39
##  core size 95 of 198  vol = 39.706
##  core size 94 of 198  vol = 29.392
##  core size 93 of 198  vol = 22.444
##  core size 92 of 198  vol = 14.094
##  core size 91 of 198  vol = 9.164
##  core size 90 of 198  vol = 4.916
##  core size 89 of 198  vol = 3.449
##  core size 88 of 198  vol = 2.593
##  core size 87 of 198  vol = 1.8
##  core size 86 of 198  vol = 1.498
##  core size 85 of 198  vol = 1.292
##  core size 84 of 198  vol = 1.13
##  core size 83 of 198  vol = 0.97
##  core size 82 of 198  vol = 0.825
##  core size 81 of 198  vol = 0.691
##  core size 80 of 198  vol = 0.56
##  core size 79 of 198  vol = 0.507
##  core size 78 of 198  vol = 0.457
##  FINISHED: Min vol ( 0.5 ) reached
```

```
core.inds <- print(core, vol=0.5)
```

```
## # 79 positions (cumulative volume <= 0.5 Angstrom^3)
##   start end length
## 1    10  24     15
## 2    27  48     22
## 3    53  94     42
```

```
xyz <- pdbfit(pdbs, core.inds, outpath="corefit_structures")
```

```
rf <- rmsf(xyz)

plotb3(rf, sse=pdb)
abline(v=100, col="gray", ylab="RMSF")
```




```
library(jsonlite)

# Listing of all PAE JSON files
pae_files <- list.files(path=results_dir,
                        pattern=".*model.*\\.json",
                        full.names = TRUE)
```

```
pae1 <- read_json(pae_files[1],simplifyVector = TRUE)
pae5 <- read_json(pae_files[5],simplifyVector = TRUE)

attributes(pae1)
```

```
## $names
## [1] "plddt"    "max_pae" "pae"     "ptm"     "iptm"
```

```
# Per-residue pLDDT scores
#  same as B-factor of PDB..
head(pae1$plddt)
```

```
## [1] 92.50 96.56 96.94 96.62 97.69 96.00
```

How about the other models, what are thir max PAE scores?

```
pae1$max_pae
```

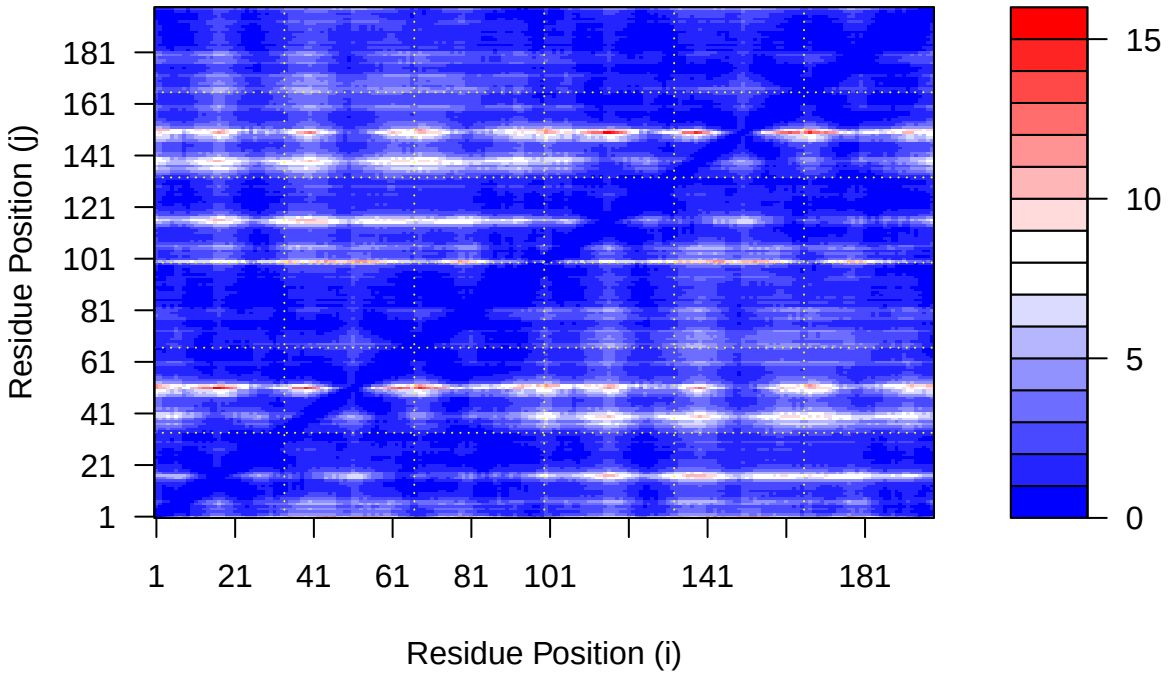
```
## [1] 15.54688
```

```
pae5$max_pae
```

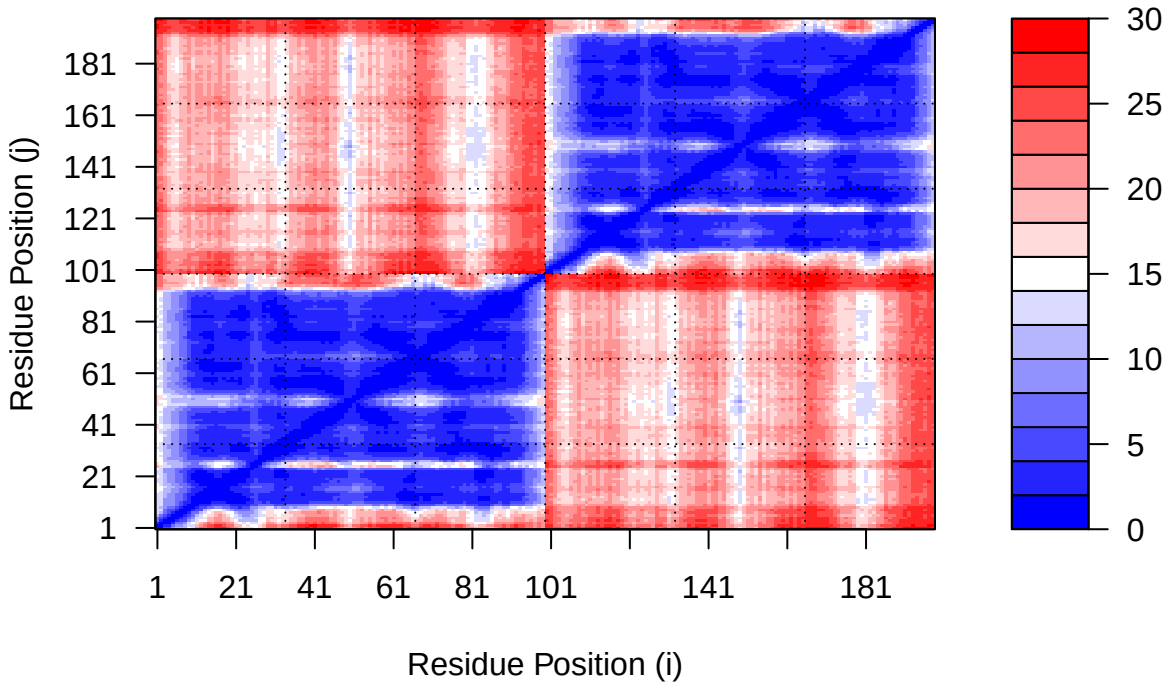
```
## [1] 29.29688
```

```
library(bio3d)
```

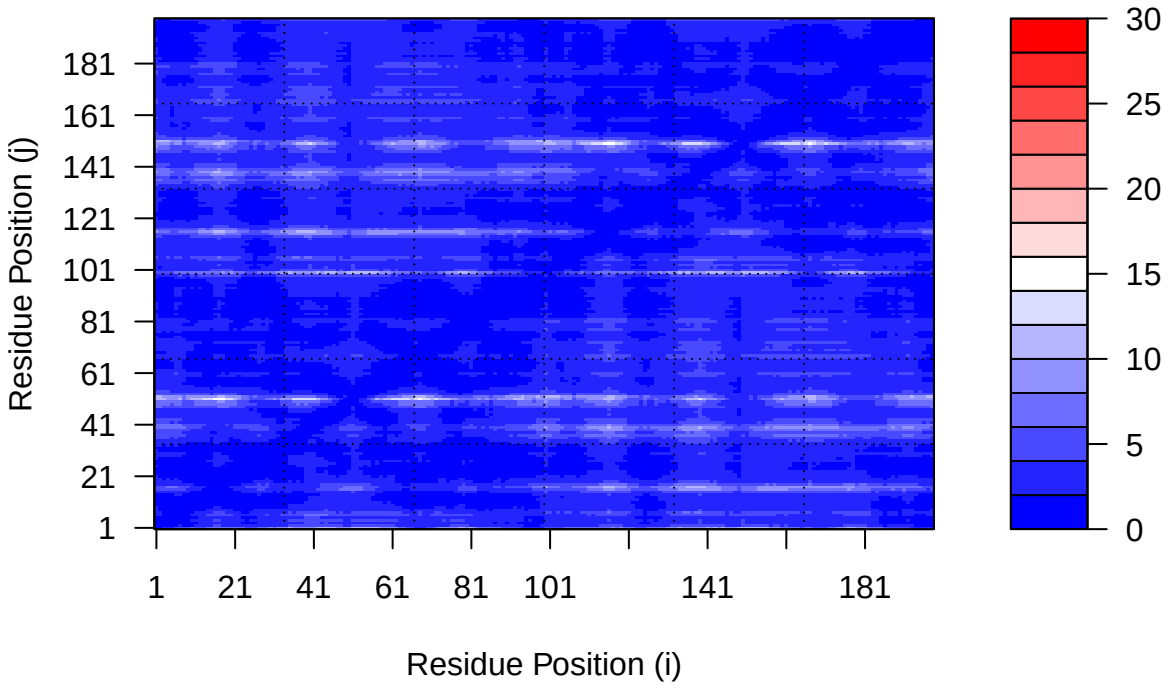
```
plot.dmat(pae1$pae, xlab="Residue Position (i)", ylab="Residue Position (j)")
```



```
plot.dmat(pae5$pae,
          xlab="Residue Position (i)",
          ylab="Residue Position (j)",
          grid.col = "black",
          zlim=c(0,30))
```




```
plot.dmat(pae1$pae,
          xlab="Residue Position (i)",
          ylab="Residue Position (j)",
          grid.col = "black",
          zlim=c(0,30))
```



Residue conservation from alignment file

```
aln_file <- list.files(path=results_dir,
                       pattern=".a3m$",
                        full.names = TRUE)
aln_file
```

```
## [1] "hivprdimer_23119/HIVprDimer_23119.a3m"
```

```
aln <- read.fasta(aln_file[1], to.upper = TRUE)
```

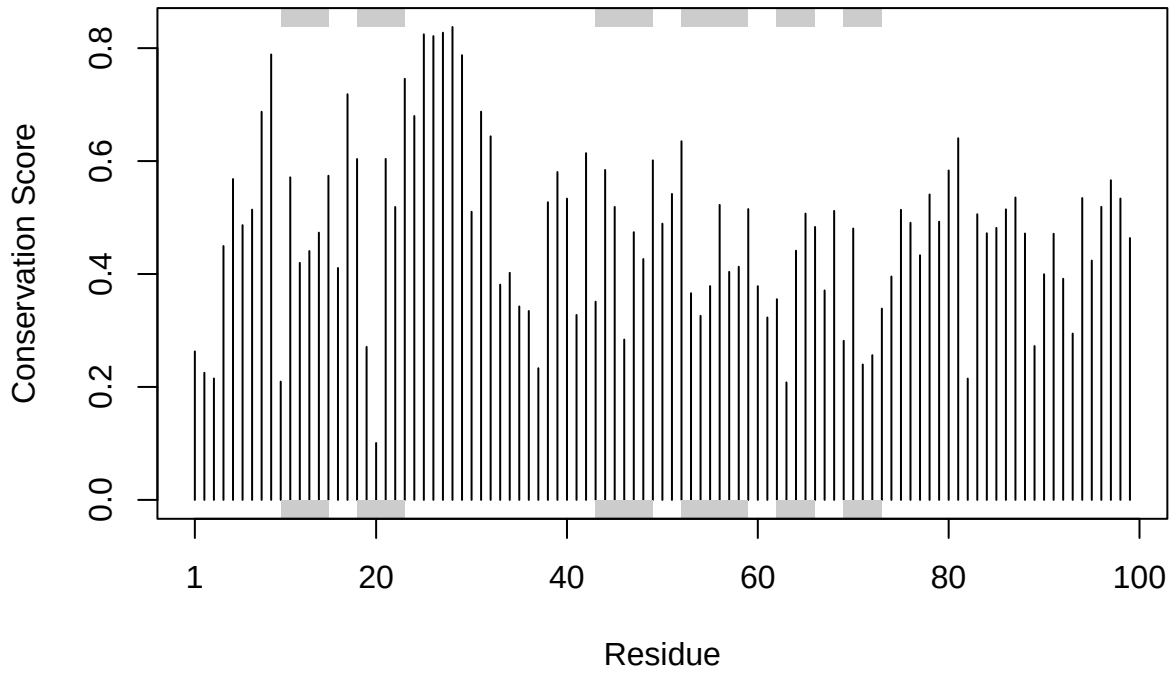
```
## [1] " ** Duplicated sequence id's: 101 **"
## [2] " ** Duplicated sequence id's: 101 **"
```

```
dim(aln$ali)
```

```
## [1] 5378  132
```

```
sim <- conserv(aln)
```

```
plotb3(sim[1:99], sse=trim.pdb(pdb, chain="A"),
       ylab="Conservation Score")
```




```
con <- consensus(aln, cutoff = 0.9)
con$seq
```

```
##   [1] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
##  [19] "-" "-" "-" "-" "-" "-" "D" "T" "G" "A" "-" "-" "-" "-" "-" "-" "-" "-"
##  [37] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
##  [55] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
##  [73] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
##  [91] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
## [109] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
## [127] "-" "-" "-" "-" "-" "-"
```

```
m1.pdb <- read.pdb(pdb_files[1])
occ <- vec2resno(c(sim[1:99], sim[1:99]), m1.pdb$atom$resno)
write.pdb(m1.pdb, o=occ, file="m1_conserv.pdb")
```

Now display this PDB in MOL* and color by Occupancy column

The conserved active site residues are highlighted.

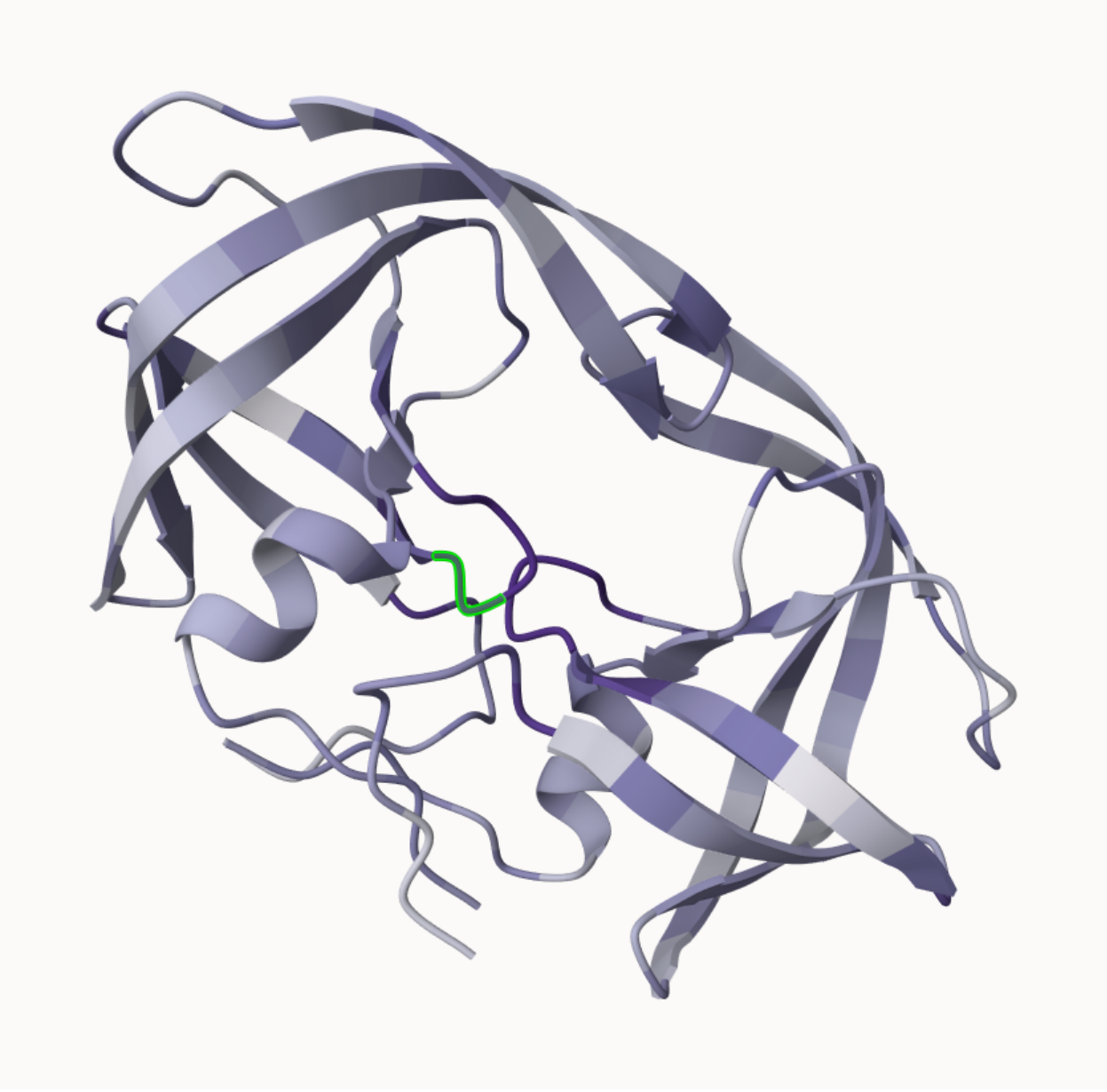


Figure 3: CONSERV