CLASS 11 - BIMM 143

Abzael Verduzco

We need some packages for todays class. These include `bio3d` and `msa`.

The `msa` package is from BioConductor. These packages focus on genomics type work and are managed by the `BiocManager` package.

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

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

MSA

> Q11. Which of the above packages is not found on BioConductor or CRAN?

> Q12. 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?

It is 214 amino acids long

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

#hits <- plot(b)

#attributes(b)
#head(b$hit.tbl)

hits <- NULL
hits$pdb.id <- c('1AKE_A','6S36_A','6RZE_A','3HPR_A','1E4V_A','5EJE_A','1E4Y_A','3X2S_A','
```

Side-note: Lets annotate these structures (in other word find out what they are, what 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
6S36_A	6S36	A	Protein	214	X-ray
6RZE_A	6RZE	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

	resolution	scopDomain	pfam	ligandId
1AKE_A	2.00	Adenylate kinase	Adenylate kinase (ADK)	AP5
6S36_A	1.60	<NA>	Adenylate kinase (ADK)	CL (3),NA,MG (2)
6RZE_A	1.69	<NA>	Adenylate kinase (ADK)	NA (3),CL (2)
3HPR_A	2.00	<NA>	Adenylate kinase (ADK)	AP5
1E4V_A	1.85	Adenylate kinase	Adenylate kinase (ADK)	AP5
5EJE_A	1.90	<NA>	Adenylate kinase (ADK)	AP5,CO

	ligandName
1AKE_A	BIS(ADENOSINE)-5'-PENTAPHOSPHATE
6S36_A	CHLORIDE ION (3),SODIUM ION,MAGNESIUM ION (2)
6RZE_A	SODIUM ION (3),CHLORIDE ION (2)
3HPR_A	BIS(ADENOSINE)-5'-PENTAPHOSPHATE
1E4V_A	BIS(ADENOSINE)-5'-PENTAPHOSPHATE
5EJE_A	BIS(ADENOSINE)-5'-PENTAPHOSPHATE,COBALT (II) ION

	source
1AKE_A	Escherichia coli
6S36_A	Escherichia coli
6RZE_A	Escherichia coli
3HPR_A	Escherichia coli K-12
1E4V_A	Escherichia coli
5EJE_A	Escherichia coli O139:H28 str. E24377A

```
1AKE_A STRUCTURE OF THE COMPLEX BETWEEN ADENYLATE KINASE FROM ESCHERICHIA COLI AND THE INHIBI
6S36_A
6RZE_A
```

```
3HPR_A
1E4V_A
5EJE_A                                                                  Cryst
                                                 citation rObserved  rFree
1AKE_A          Muller, C.W., et al. J Mol Biol (1992)    0.1960      NA
6S36_A          Rogne, P., et al. Biochemistry (2019)     0.1632 0.2356
6RZE_A          Rogne, P., et al. Biochemistry (2019)     0.1865 0.2350
3HPR_A Schrank, T.P., et al. Proc Natl Acad Sci U S A (2009)    0.2100 0.2432
1E4V_A          Muller, C.W., et al. Proteins (1993)      0.1960      NA
5EJE_A Kovermann, M., et al. Proc Natl Acad Sci U S A (2017)    0.1889 0.2358
        rWork spaceGroup
1AKE_A 0.1960  P 21 2 21
6S36_A 0.1594    C 1 2 1
6RZE_A 0.1819    C 1 2 1
3HPR_A 0.2062  P 21 21 2
1E4V_A 0.1960  P 21 2 21
5EJE_A 0.1863  P 21 2 21
```

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

```
#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.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/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/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%
  |
  |=====                                                                 |   8%
  |
  |==========                                                            |  15%
  |
  |===============                                                       |  23%
  |
  |=====================                                                 |  31%
  |
  |==========================                                            |  38%
  |
  |===============================                                       |  46%
  |
  |====================================                                  |  54%
  |
  |=========================================                             |  62%
```

```
  |
  |==================================================                 |  69%
  |
  |=====================================================              |  77%
  |
  |==========================================================         |  85%
  |
  |===============================================================    |  92%
  |
  |===================================================================| 100%
```

Now we have all these realated structures we can Align and Superpose...

```
  # Align related PDBs
  pdbs <- pdbaln(files, fit = TRUE, exefile="msa")
```

```
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
..   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/6S36_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
pdb/seq: 4   name: pdbs/split_chain/3HPR_A.pdb
    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

```
                                 1         .         .         .        40
[Truncated_Name:1]1AKE_A.pdb    ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:2]6S36_A.pdb    ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:3]6RZE_A.pdb    ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:4]3HPR_A.pdb    ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:5]1E4V_A.pdb    ----------MRIILLGAPVAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:6]5EJE_A.pdb    ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:7]1E4Y_A.pdb    ----------MRIILLGALVAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:8]3X2S_A.pdb    ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:9]6HAP_A.pdb    ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:10]6HAM_A.pdb   ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:11]4K46_A.pdb   ----------MRIILLGAPGAGKGTQAQFIMAKFGIPQIS
[Truncated_Name:12]3GMT_A.pdb   ----------MRLILLGAPGAGKGTQANFIKEKFGIPQIS
[Truncated_Name:13]4PZL_A.pdb   TENLYFQSNAMRIILLGAPGAGKGTQAKIIEQKYNIAHIS
                                          **^*****  *******  *  *^ *  **
                                 1         .         .         .        40

                                41        .         .         .        80
[Truncated_Name:1]1AKE_A.pdb    TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
```

7

```
[Truncated_Name:2]6S36_A.pdb       TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:3]6RZE_A.pdb       TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:4]3HPR_A.pdb       TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:5]1E4V_A.pdb       TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:6]5EJE_A.pdb       TGDMLRAAVKSGSELGKQAKDIMDACKLVTDELVIALVKE
[Truncated_Name:7]1E4Y_A.pdb       TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:8]3X2S_A.pdb       TGDMLRAAVKSGSELGKQAKDIMDCGKLVTDELVIALVKE
[Truncated_Name:9]6HAP_A.pdb       TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVRE
[Truncated_Name:10]6HAM_A.pdb      TGDMLRAAIKSGSELGKQAKDIMDAGKLVTDEIIIALVKE
[Truncated_Name:11]4K46_A.pdb      TGDMLRAAIKAGTELGKQAKSVIDAGQLVSDDIILGLVKE
[Truncated_Name:12]3GMT_A.pdb      TGDMLRAAVKAGTPLGVEAKTYMDEGKLVPDSLIIGLVKE
[Truncated_Name:13]4PZL_A.pdb      TGDMIRETIKSGSALGQELKKVLDAGELVSDEFIIKIVKD
                                   ****^*  ^* *^ **    *  ^*    ** *  ^^ ^*^^
                                   41          .         .         .        80


                                   81          .         .         .        120
[Truncated_Name:1]1AKE_A.pdb       RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:2]6S36_A.pdb       RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:3]6RZE_A.pdb       RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:4]3HPR_A.pdb       RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:5]1E4V_A.pdb       RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:6]5EJE_A.pdb       RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:7]1E4Y_A.pdb       RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:8]3X2S_A.pdb       RIAQEDSRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:9]6HAP_A.pdb       RICQEDSRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:10]6HAM_A.pdb      RICQEDSRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:11]4K46_A.pdb      RIAQDDCAKGFLLDGFPRTIPQADGLKEVGVVVDYVIEFD
[Truncated_Name:12]3GMT_A.pdb      RLKEADCANGYLFDGFPRTIAQADAMKEAGVAIDYVLEID
[Truncated_Name:13]4PZL_A.pdb      RISKNDCNNGFLLDGVPRTIPQAQELDKLGVNIDYIVEVD
                                   *^     *   *^* ** **** **  ^    *^ ^**^^* *
                                   81          .         .         .        120


                                   121         .         .         .        160
[Truncated_Name:1]1AKE_A.pdb       VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:2]6S36_A.pdb       VPDELIVDKIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:3]6RZE_A.pdb       VPDELIVDAIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:4]3HPR_A.pdb       VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDGTG
[Truncated_Name:5]1E4V_A.pdb       VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:6]5EJE_A.pdb       VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:7]1E4Y_A.pdb       VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:8]3X2S_A.pdb       VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:9]6HAP_A.pdb       VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:10]6HAM_A.pdb      VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
```

```
[Truncated_Name:11]4K46_A.pdb    VADSVIVERMAGRRAHLASGRTYHNVYNPPKVEGKDDVTG
[Truncated_Name:12]3GMT_A.pdb    VPFSEIIERMSGRRTHPASGRTYHVKFNPPKVEGKDDVTG
[Truncated_Name:13]4PZL_A.pdb    VADNLLIERITGRRIHPASGRTYHTKFNPPKVADKDDVTG
                                 *     ^^^ ^ *** *  *** **  ^*****  *** **
                              121         .         .         .        160


                              161         .         .         .        200
[Truncated_Name:1]1AKE_A.pdb     EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:2]6S36_A.pdb     EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:3]6RZE_A.pdb     EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:4]3HPR_A.pdb     EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:5]1E4V_A.pdb     EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:6]5EJE_A.pdb     EELTTRKDDQEECVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:7]1E4Y_A.pdb     EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:8]3X2S_A.pdb     EELTTRKDDQEETVRKRLCEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:9]6HAP_A.pdb     EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:10]6HAM_A.pdb    EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:11]4K46_A.pdb    EDLVIREDDKEETVLARLGVYHNQTAPLIAYYGKEAEAGN
[Truncated_Name:12]3GMT_A.pdb    EPLVQRDDDKEETVKKRLDVYEAQTKPLITYYGDWARRGA
[Truncated_Name:13]4PZL_A.pdb    EPLITRTDDNEDTVKQRLSVYHAQTAKLIDFYRNFSSTNT
                                 * *   * **  *^ *  **   *   *  ** ^*
                              161         .         .         .        200


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


Call:
  pdbaln(files = files, fit = TRUE, exefile = "msa")
```
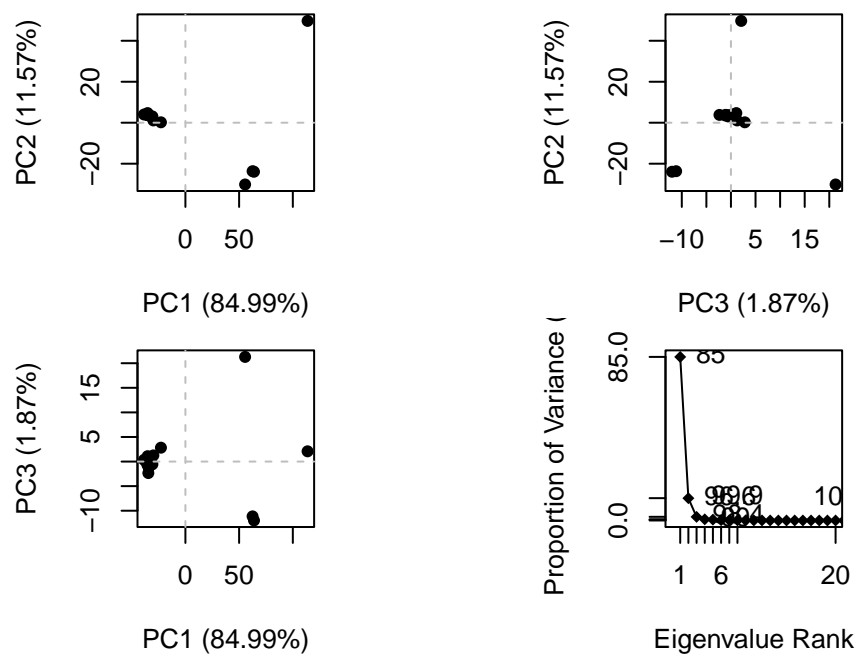
9

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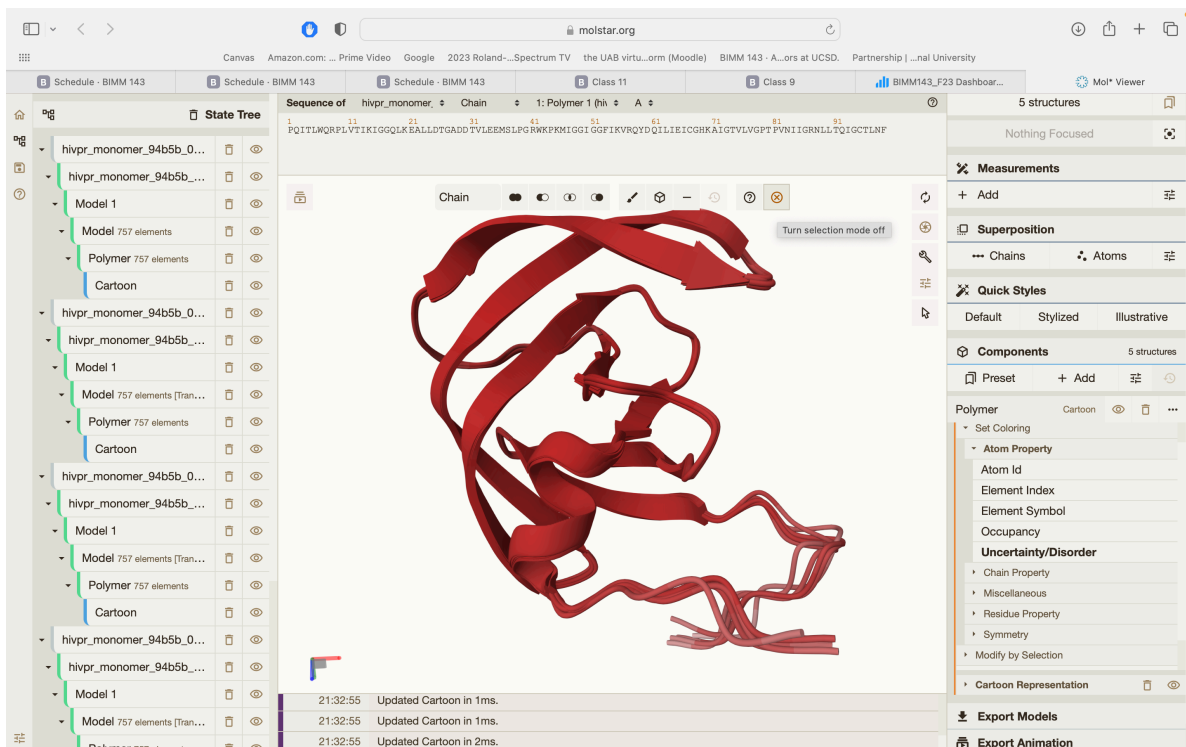
```
# Perform PCA
pc.xray <- pca(pdbs)
plot(pc.xray)
```



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

Class 11 Primer

10

```r
# Get the current working directory
current_dir <- getwd()

# Print the current working directory
print(current_dir)
```

```
[1] "/Users/abzael/Desktop/BIMM 143/Class 11 - Bimm 143/hivpr_dimer_23119"
```

```r
library(bio3d)

results_dir <- current_dir
```

```r
# 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] "hivpr_dimer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_1_seed_000.pdb"
[2] "hivpr_dimer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000.pdb"
[3] "hivpr_dimer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000.pdb"
[4] "hivpr_dimer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_2_seed_000.pdb"
[5] "hivpr_dimer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_3_seed_000.pdb"
[6] "pc_1.pdb"
```

```r
library(bio3d)

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

```
Reading PDB files:
/Users/abzael/Desktop/BIMM 143/Class 11 - Bimm 143/hivpr_dimer_23119/hivpr_dimer_23119_unrela
/Users/abzael/Desktop/BIMM 143/Class 11 - Bimm 143/hivpr_dimer_23119/hivpr_dimer_23119_unrela
/Users/abzael/Desktop/BIMM 143/Class 11 - Bimm 143/hivpr_dimer_23119/hivpr_dimer_23119_unrela
/Users/abzael/Desktop/BIMM 143/Class 11 - Bimm 143/hivpr_dimer_23119/hivpr_dimer_23119_unrela
/Users/abzael/Desktop/BIMM 143/Class 11 - Bimm 143/hivpr_dimer_23119/hivpr_dimer_23119_unrela
/Users/abzael/Desktop/BIMM 143/Class 11 - Bimm 143/hivpr_dimer_23119/pc_1.pdb
......

Extracting sequences

pdb/seq: 1   name: /Users/abzael/Desktop/BIMM 143/Class 11 - Bimm 143/hivpr_dimer_23119/hivpr
pdb/seq: 2   name: /Users/abzael/Desktop/BIMM 143/Class 11 - Bimm 143/hivpr_dimer_23119/hivpr
pdb/seq: 3   name: /Users/abzael/Desktop/BIMM 143/Class 11 - Bimm 143/hivpr_dimer_23119/hivpr
pdb/seq: 4   name: /Users/abzael/Desktop/BIMM 143/Class 11 - Bimm 143/hivpr_dimer_23119/hivpr
pdb/seq: 5   name: /Users/abzael/Desktop/BIMM 143/Class 11 - Bimm 143/hivpr_dimer_23119/hivpr
pdb/seq: 6   name: /Users/abzael/Desktop/BIMM 143/Class 11 - Bimm 143/hivpr_dimer_23119/pc_1
```

```r
pdbs
```

```
                        1          .         .         .         .          50
[Truncated_Name:1]hivpr_dime   PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
[Truncated_Name:2]hivpr_dime   PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
[Truncated_Name:3]hivpr_dime   PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
[Truncated_Name:4]hivpr_dime   PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
[Truncated_Name:5]hivpr_dime   PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
[Truncated_Name:6]pc_1.pdb     AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
                               ^^    *    ^*           ^       ^^
                        1          .         .         .         .          50
```

```
                                        51          .          .          .          .          100
[Truncated_Name:1]hivpr_dime    GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
[Truncated_Name:2]hivpr_dime    GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
[Truncated_Name:3]hivpr_dime    GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
[Truncated_Name:4]hivpr_dime    GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
[Truncated_Name:5]hivpr_dime    GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
[Truncated_Name:6]pc_1.pdb      AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
                                ^^                 ^  * ^      ^          ^        ^
                                        51          .          .          .          .          100


                                        101         .          .          .          .          150
[Truncated_Name:1]hivpr_dime    QITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIG
[Truncated_Name:2]hivpr_dime    QITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIG
[Truncated_Name:3]hivpr_dime    QITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIG
[Truncated_Name:4]hivpr_dime    QITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIG
[Truncated_Name:5]hivpr_dime    QITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIG
[Truncated_Name:6]pc_1.pdb      AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
                                          ^^      *      ^*              ^         ^^ ^
                                        101         .          .          .          .          150


                                        151         .          .          .          .          200
[Truncated_Name:1]hivpr_dime    GFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF--
[Truncated_Name:2]hivpr_dime    GFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF--
[Truncated_Name:3]hivpr_dime    GFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF--
[Truncated_Name:4]hivpr_dime    GFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF--
[Truncated_Name:5]hivpr_dime    GFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF--
[Truncated_Name:6]pc_1.pdb      AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
                                ^                  ^  * ^      ^          ^        ^
                                        151         .          .          .          .          200


                                        201  204
[Truncated_Name:1]hivpr_dime    ----
[Truncated_Name:2]hivpr_dime    ----
[Truncated_Name:3]hivpr_dime    ----
[Truncated_Name:4]hivpr_dime    ----
[Truncated_Name:5]hivpr_dime    ----
[Truncated_Name:6]pc_1.pdb      AAAA

                                        201  204


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

13

```
Class:
  pdbs, fasta

Alignment dimensions:
  6 sequence rows; 204 position columns (198 non-gap, 6 gap)

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

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

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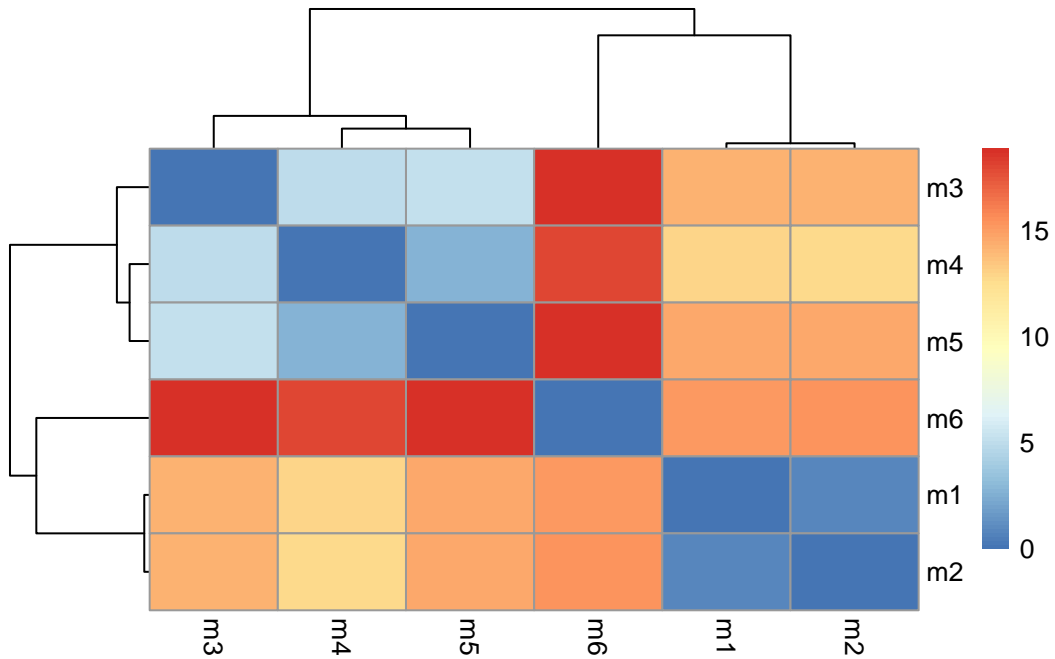
```r
range(rd)
```

```
[1]  0.000 18.842
```

```r
library(pheatmap)

colnames(rd) <- paste0("m",1:6)
rownames(rd) <- paste0("m",1:6)
pheatmap(rd)
```
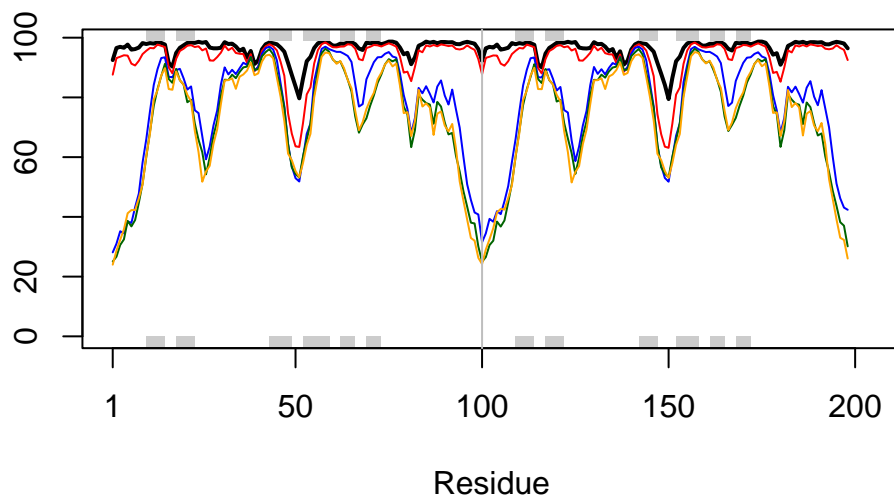
```
# Read a reference PDB structure
pdb <- read.pdb("1hsg")
```

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 = 24427.32
core size 196 of 198   vol = 22911.68
core size 195 of 198   vol = 22073.89
core size 194 of 198   vol = 21328.92
core size 193 of 198   vol = 20700.65
core size 192 of 198   vol = 20139.24
core size 191 of 198   vol = 19522.56
core size 190 of 198   vol = 18995.53
core size 189 of 198   vol = 18506.97
core size 188 of 198   vol = 18075.76
core size 187 of 198   vol = 17688.64
core size 186 of 198   vol = 17441.08
core size 185 of 198   vol = 17119.3
core size 184 of 198   vol = 16601.69
core size 183 of 198   vol = 16186.29
core size 182 of 198   vol = 15692.09
core size 181 of 198   vol = 15269
core size 180 of 198   vol = 14895.94
core size 179 of 198   vol = 14575.27
core size 178 of 198   vol = 14350.03
```

```
core size 177 of 198   vol = 14196.26
core size 176 of 198   vol = 14015.18
core size 175 of 198   vol = 13840.46
core size 174 of 198   vol = 13723.29
core size 173 of 198   vol = 13619.93
core size 172 of 198   vol = 13508.79
core size 171 of 198   vol = 13388.4
core size 170 of 198   vol = 13208.23
core size 169 of 198   vol = 12983.51
core size 168 of 198   vol = 12795.17
core size 167 of 198   vol = 12541.98
core size 166 of 198   vol = 12309.71
core size 165 of 198   vol = 12074.45
core size 164 of 198   vol = 11782.37
core size 163 of 198   vol = 11466.05
core size 162 of 198   vol = 11221.85
core size 161 of 198   vol = 10993.56
core size 160 of 198   vol = 10819.07
core size 159 of 198   vol = 10536.22
core size 158 of 198   vol = 10220.11
core size 157 of 198   vol = 10003.66
core size 156 of 198   vol = 9711.317
core size 155 of 198   vol = 9403.547
core size 154 of 198   vol = 9096.128
core size 153 of 198   vol = 8815.405
core size 152 of 198   vol = 8594.691
core size 151 of 198   vol = 8321.921
core size 150 of 198   vol = 8080.948
core size 149 of 198   vol = 7912.441
core size 148 of 198   vol = 7615.273
core size 147 of 198   vol = 7412.695
core size 146 of 198   vol = 7255.47
core size 145 of 198   vol = 6986.149
core size 144 of 198   vol = 6731.547
core size 143 of 198   vol = 6512.064
core size 142 of 198   vol = 6257.187
core size 141 of 198   vol = 6057.258
core size 140 of 198   vol = 5860.267
core size 139 of 198   vol = 5740.221
core size 138 of 198   vol = 5570.237
core size 137 of 198   vol = 5340.966
core size 136 of 198   vol = 5137.386
core size 135 of 198   vol = 4964.342
```

```
core size 134 of 198   vol = 4801.664
core size 133 of 198   vol = 4587.367
core size 132 of 198   vol = 4414.136
core size 131 of 198   vol = 4264.741
core size 130 of 198   vol = 4044.26
core size 129 of 198   vol = 3906.161
core size 128 of 198   vol = 3770.191
core size 127 of 198   vol = 3665.324
core size 126 of 198   vol = 3554.272
core size 125 of 198   vol = 3390.298
core size 124 of 198   vol = 3204.642
core size 123 of 198   vol = 3095.267
core size 122 of 198   vol = 3013.63
core size 121 of 198   vol = 2925.413
core size 120 of 198   vol = 2700.296
core size 119 of 198   vol = 2624.459
core size 118 of 198   vol = 2546.108
core size 117 of 198   vol = 2429.493
core size 116 of 198   vol = 2352.066
core size 115 of 198   vol = 2275.06
core size 114 of 198   vol = 2204.646
core size 113 of 198   vol = 2112.21
core size 112 of 198   vol = 2025.003
core size 111 of 198   vol = 1971.788
core size 110 of 198   vol = 1891.315
core size 109 of 198   vol = 1836.58
core size 108 of 198   vol = 1789.663
core size 107 of 198   vol = 1776.844
core size 106 of 198   vol = 1734.373
core size 105 of 198   vol = 1640.862
core size 104 of 198   vol = 1599.153
core size 103 of 198   vol = 1482.273
core size 102 of 198   vol = 1367.169
core size 101 of 198   vol = 1291.91
core size 100 of 198   vol = 1094.704
core size 99 of 198  vol = 1024.606
core size 98 of 198  vol = 953.94
core size 97 of 198  vol = 846.573
core size 96 of 198  vol = 796.648
core size 95 of 198  vol = 734.069
core size 94 of 198  vol = 695.998
core size 93 of 198  vol = 632.564
core size 92 of 198  vol = 525.33
```

```
core size 91 of 198  vol = 460.598
core size 90 of 198  vol = 354.311
core size 89 of 198  vol = 230.417
core size 88 of 198  vol = 101.122
core size 87 of 198  vol = 69.969
core size 86 of 198  vol = 59.291
core size 85 of 198  vol = 43.767
core size 84 of 198  vol = 35.855
core size 83 of 198  vol = 33.212
core size 82 of 198  vol = 30.233
core size 81 of 198  vol = 27.778
core size 80 of 198  vol = 25.42
core size 79 of 198  vol = 23.667
core size 78 of 198  vol = 21.865
core size 77 of 198  vol = 19.865
core size 76 of 198  vol = 18.464
core size 75 of 198  vol = 17.167
core size 74 of 198  vol = 15.886
core size 73 of 198  vol = 15.092
core size 72 of 198  vol = 14.312
core size 71 of 198  vol = 13.348
core size 70 of 198  vol = 12.457
core size 69 of 198  vol = 11.556
core size 68 of 198  vol = 10.73
core size 67 of 198  vol = 10.053
core size 66 of 198  vol = 9.304
core size 65 of 198  vol = 8.735
core size 64 of 198  vol = 8.272
core size 63 of 198  vol = 7.726
core size 62 of 198  vol = 7.186
core size 61 of 198  vol = 6.74
core size 60 of 198  vol = 6.328
core size 59 of 198  vol = 5.888
core size 58 of 198  vol = 5.505
core size 57 of 198  vol = 5.175
core size 56 of 198  vol = 4.912
core size 55 of 198  vol = 4.602
core size 54 of 198  vol = 4.415
core size 53 of 198  vol = 4.114
core size 52 of 198  vol = 3.815
core size 51 of 198  vol = 3.647
core size 50 of 198  vol = 3.46
core size 49 of 198  vol = 3.287
```

```
core size 48 of 198  vol = 3.13
core size 47 of 198  vol = 2.855
core size 46 of 198  vol = 2.701
core size 45 of 198  vol = 2.559
core size 44 of 198  vol = 2.445
core size 43 of 198  vol = 2.282
core size 42 of 198  vol = 2.076
core size 41 of 198  vol = 1.895
core size 40 of 198  vol = 1.785
core size 39 of 198  vol = 1.693
core size 38 of 198  vol = 1.577
core size 37 of 198  vol = 1.449
core size 36 of 198  vol = 1.329
core size 35 of 198  vol = 1.264
core size 34 of 198  vol = 1.198
core size 33 of 198  vol = 1.118
core size 32 of 198  vol = 1.062
core size 31 of 198  vol = 1.042
core size 30 of 198  vol = 1.026
core size 29 of 198  vol = 0.927
core size 28 of 198  vol = 0.866
core size 27 of 198  vol = 0.811
core size 26 of 198  vol = 0.767
core size 25 of 198  vol = 0.702
core size 24 of 198  vol = 0.628
core size 23 of 198  vol = 0.559
core size 22 of 198  vol = 0.5
FINISHED: Min vol ( 0.5 ) reached
```

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

```
# 23 positions (cumulative volume <= 0.5 Angstrom^3)
   start end length
1     32  35      4
2     37  38      2
3     41  42      2
4     56  64      9
5     77  77      1
6     83  85      3
7     88  88      1
8     91  91      1
```

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

Dimer