

Class 11: Comparative analysis of structures

Destiny (A16340362)

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

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

`BiocManager::install("msa")` all entered in the R “brain” console

```
library(bio3d)

aa <- get.seq("lake_A")
```

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

Fetching... Please wait. Done.

```
aa
```

```

      1      .      .      .      .      .      .      60
pdb|1AKE|A  MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLV
      1      .      .      .      .      .      .      60

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

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

      181      .      .      .      214
pdb|1AKE|A  YYSKEAEAGNTKYAKVDGTPVAEVRADLEKILG
```

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 database for related sequences

```
#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',
```

```
hits$ pdb.id
```

```
[1] "1AKE_A" "6S36_A" "6RZE_A" "3HPR_A" "1E4V_A" "5EJE_A" "1E4Y_A" "3X2S_A"
```

```
[9] "6HAP_A" "6HAM_A" "4K46_A" "3GMT_A" "4PZL_A"
```

Side note: Let's annotate these structures (in other words, find out what they are, what species they're from. stuff about the experiment they were solved in, etc.)

For this 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 0139:H28 str. E24377A

1AKE_A	STRUCTURE OF THE COMPLEX BETWEEN ADENYLATE KINASE FROM ESCHERICHIA COLI AND THE INHIBIT
6S36_A	
6RZE_A	
3HPR_A	
1E4V_A	
5EJE_A	

	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		

Cryst

```
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.

```
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
```

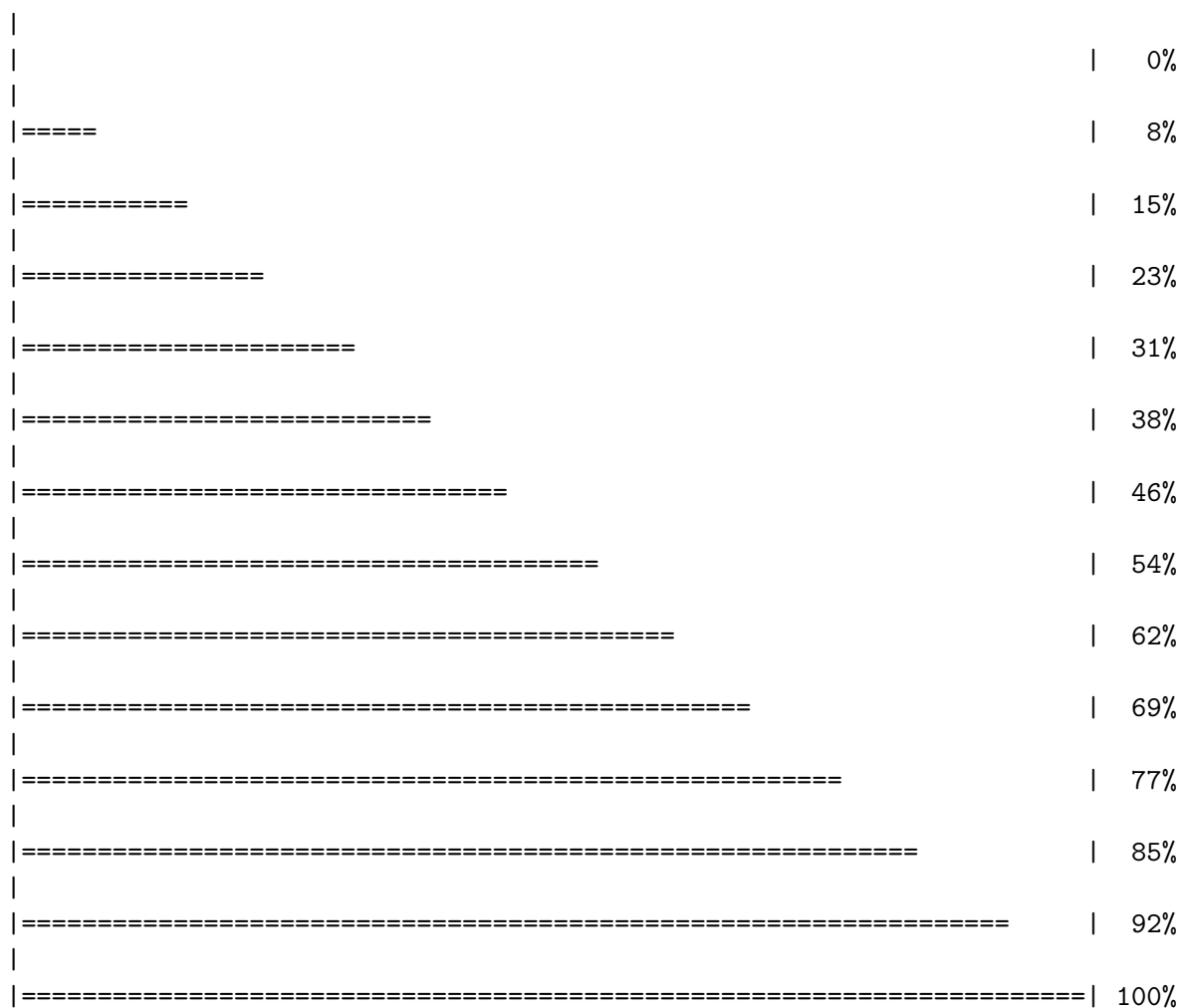
```
Warning in get.pdb(hits$ pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/6HAP.pdb.gz exists. Skipping download
```

```
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



Now we have all these related structures we can Align and superimpose

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

Reading PDB files:

```
pdbbs/split_chain/1AKE_A.pdb
pdbbs/split_chain/6S36_A.pdb
pdbbs/split_chain/6RZE_A.pdb
pdbbs/split_chain/3HPR_A.pdb
pdbbs/split_chain/1E4V_A.pdb
pdbbs/split_chain/5EJE_A.pdb
pdbbs/split_chain/1E4Y_A.pdb
pdbbs/split_chain/3X2S_A.pdb
pdbbs/split_chain/6HAP_A.pdb
pdbbs/split_chain/6HAM_A.pdb
pdbbs/split_chain/4K46_A.pdb
pdbbs/split_chain/3GMT_A.pdb
pdbbs/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: pdbbs/split_chain/1AKE_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 2   name: pdbbs/split_chain/6S36_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 3   name: pdbbs/split_chain/6RZE_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 4   name: pdbbs/split_chain/3HPR_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 5   name: pdbbs/split_chain/1E4V_A.pdb
pdb/seq: 6   name: pdbbs/split_chain/5EJE_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 7   name: pdbbs/split_chain/1E4Y_A.pdb
pdb/seq: 8   name: pdbbs/split_chain/3X2S_A.pdb
pdb/seq: 9   name: pdbbs/split_chain/6HAP_A.pdb
pdb/seq: 10  name: pdbbs/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		TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVKE			
[Truncated_Name:2] 6S36_A.pdb		TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVKE			
[Truncated_Name:3] 6RZE_A.pdb		TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVKE			
[Truncated_Name:4] 3HPR_A.pdb		TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVKE			
[Truncated_Name:5] 1E4V_A.pdb		TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVKE			
[Truncated_Name:6] 5EJE_A.pdb		TGDMLRAAVKSGSELGKQAKDIMDACKLVDELVIALVKE			
[Truncated_Name:7] 1E4Y_A.pdb		TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVKE			
[Truncated_Name:8] 3X2S_A.pdb		TGDMLRAAVKSGSELGKQAKDIMDCGKLVDELVIALVKE			
[Truncated_Name:9] 6HAP_A.pdb		TGDMLRAAVKSGSELGKQAKDIMDAGKLVDELVIALVRE			
[Truncated_Name:10] 6HAM_A.pdb		TGDMLRAAIIKSGSELGKQAKDIMDAGKLVDEIIIALVKE			
[Truncated_Name:11] 4K46_A.pdb		TGDMLRAAIIKAGTELGKQAKSVIDAGQLVSDDIILGLVKE			
[Truncated_Name:12] 3GMT_A.pdb		TGDMLRAAVKAGTPLGVEAKTYMDEGKLVPSLIIGLVKE			
[Truncated_Name:13] 4PZL_A.pdb		TGDMIRETIKSGSALGQELKKVLDAGELVSDEFIIVKVD			
		****~* ~* *~** * ~* ** * ~*~*~			
	41	.	.	.	80

	81	.	.	.	120
[Truncated_Name:1] 1AKE_A.pdb	RIAQEDCRNGFLLDGFPR	TIPQADAMKEAGINVDYVLEFD			
[Truncated_Name:2] 6S36_A.pdb	RIAQEDCRNGFLLDGFPR	TIPQADAMKEAGINVDYVLEFD			
[Truncated_Name:3] 6RZE_A.pdb	RIAQEDCRNGFLLDGFPR	TIPQADAMKEAGINVDYVLEFD			
[Truncated_Name:4] 3HPR_A.pdb	RIAQEDCRNGFLLDGFPR	TIPQADAMKEAGINVDYVLEFD			
[Truncated_Name:5] 1E4V_A.pdb	RIAQEDCRNGFLLDGFPR	TIPQADAMKEAGINVDYVLEFD			
[Truncated_Name:6] 5EJE_A.pdb	RIAQEDCRNGFLLDGFPR	TIPQADAMKEAGINVDYVLEFD			
[Truncated_Name:7] 1E4Y_A.pdb	RIAQEDCRNGFLLDGFPR	TIPQADAMKEAGINVDYVLEFD			
[Truncated_Name:8] 3X2S_A.pdb	RIAQEDSRNGFLLDGFPR	TIPQADAMKEAGINVDYVLEFD			
[Truncated_Name:9] 6HAP_A.pdb	RICQEDSRNGFLLDGFPR	TIPQADAMKEAGINVDYVLEFD			
[Truncated_Name:10] 6HAM_A.pdb	RICQEDSRNGFLLDGFPR	TIPQADAMKEAGINVDYVLEFD			
[Truncated_Name:11] 4K46_A.pdb	RIAQDDCAKGFLLDGFPR	TIPQADGLKEVGVVVDYVIEFD			
[Truncated_Name:12] 3GMT_A.pdb	RLKEADCANGYLFDFPR	TIPQADAMKEAGVAIDYVLEID			
[Truncated_Name:13] 4PZL_A.pdb	RISKNCNNGFLLDGVPR	TIPQAQELDKLGVNIDYIVEVD			
	*^	* *~* ** ***** ** ^ *^~**~* *			
	81	.	.	.	120
	121	.	.	.	160
[Truncated_Name:1] 1AKE_A.pdb	VPDELIVDRIVGRRVHAP	SGRVYHVKNPPKVEGKDDVTG			
[Truncated_Name:2] 6S36_A.pdb	VPDELIVDKIVGRRVHAP	SGRVYHVKNPPKVEGKDDVTG			
[Truncated_Name:3] 6RZE_A.pdb	VPDELIVDAIVGRRVHAP	SGRVYHVKNPPKVEGKDDVTG			
[Truncated_Name:4] 3HPR_A.pdb	VPDELIVDRIVGRRVHAP	SGRVYHVKNPPKVEGKDDGTG			
[Truncated_Name:5] 1E4V_A.pdb	VPDELIVDRIVGRRVHAP	SGRVYHVKNPPKVEGKDDVTG			
[Truncated_Name:6] 5EJE_A.pdb	VPDELIVDRIVGRRVHAP	SGRVYHVKNPPKVEGKDDVTG			
[Truncated_Name:7] 1E4Y_A.pdb	VPDELIVDRIVGRRVHAP	SGRVYHVKNPPKVEGKDDVTG			
[Truncated_Name:8] 3X2S_A.pdb	VPDELIVDRIVGRRVHAP	SGRVYHVKNPPKVEGKDDVTG			
[Truncated_Name:9] 6HAP_A.pdb	VPDELIVDRIVGRRVHAP	SGRVYHVKNPPKVEGKDDVTG			
[Truncated_Name:10] 6HAM_A.pdb	VPDELIVDRIVGRRVHAP	SGRVYHVKNPPKVEGKDDVTG			
[Truncated_Name:11] 4K46_A.pdb	VADSVIVERMAGRRRAHL	ASGRTYHNVYNPPKVEGKDDVTG			
[Truncated_Name:12] 3GMT_A.pdb	VPFSEIIERMSGRRTHP	ASGRTYHVKNPPKVEGKDDVTG			
[Truncated_Name:13] 4PZL_A.pdb	VADNLLIERITGRRIH	PASGRTYHTKFNPPKVADKDDVTG			
	*	^~^ ^ *** * *** ** ^***** *** **			
	121	.	.	.	160
	161	.	.	.	200
[Truncated_Name:1] 1AKE_A.pdb	EELTTRKDDQEETVRKRL	VEYHQMTAPLIGYYSKEAEAGN			
[Truncated_Name:2] 6S36_A.pdb	EELTTRKDDQEETVRKRL	VEYHQMTAPLIGYYSKEAEAGN			
[Truncated_Name:3] 6RZE_A.pdb	EELTTRKDDQEETVRKRL	VEYHQMTAPLIGYYSKEAEAGN			
[Truncated_Name:4] 3HPR_A.pdb	EELTTRKDDQEETVRKRL	VEYHQMTAPLIGYYSKEAEAGN			
[Truncated_Name:5] 1E4V_A.pdb	EELTTRKDDQEETVRKRL	VEYHQMTAPLIGYYSKEAEAGN			
[Truncated_Name:6] 5EJE_A.pdb	EELTTRKDDQEECVRKRL	VEYHQMTAPLIGYYSKEAEAGN			
[Truncated_Name:7] 1E4Y_A.pdb	EELTTRKDDQEETVRKRL	VEYHQMTAPLIGYYSKEAEAGN			


```

[Truncated_Name:8]3X2S_A.pdb      EELTTRKDDQEETVRKRLCEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:9]6HAP_A.pdb      EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:10]6HAM_A.pdb     EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:11]4K46_A.pdb     EDLVIREDDKEETVLARLGVYHNQ TAPLIAYYGKEAEAGN
[Truncated_Name:12]3GMT_A.pdb     EPLVQRDDDKEETVKKRLDVYEAQ TKPLITYYGDWARRGA
[Truncated_Name:13]4PZL_A.pdb     EPLITRTDDNEDTVKQRLSVYHAQ TAKLIDFYRNFSSNTNT
                                     * * * * * ^ * * * * ^ *
161                               . . .                               200

201                               . . .                               227
[Truncated_Name:1]1AKE_A.pdb      T--KYAKVDG TKPVAEVRADLEKILG-
[Truncated_Name:2]6S36_A.pdb      T--KYAKVDG TKPVAEVRADLEKILG-
[Truncated_Name:3]6RZE_A.pdb      T--KYAKVDG TKPVAEVRADLEKILG-
[Truncated_Name:4]3HPR_A.pdb      T--KYAKVDG TKPVAEVRADLEKILG-
[Truncated_Name:5]1E4V_A.pdb      T--KYAKVDG TKPVAEVRADLEKILG-
[Truncated_Name:6]5EJE_A.pdb      T--KYAKVDG TKPVAEVRADLEKILG-
[Truncated_Name:7]1E4Y_A.pdb      T--KYAKVDG TKPVAEVRADLEKILG-
[Truncated_Name:8]3X2S_A.pdb      T--KYAKVDG TKPVAEVRADLEKILG-
[Truncated_Name:9]6HAP_A.pdb      T--KYAKVDG TKPVCEVRADLEKILG-
[Truncated_Name:10]6HAM_A.pdb     T--KYAKVDG TKPVCEVRADLEKILG-
[Truncated_Name:11]4K46_A.pdb     T--QYLKFDG TKAVEVSAELEKALA-
[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")
```

Class:

```
pdbs, fasta
```

Alignment dimensions:

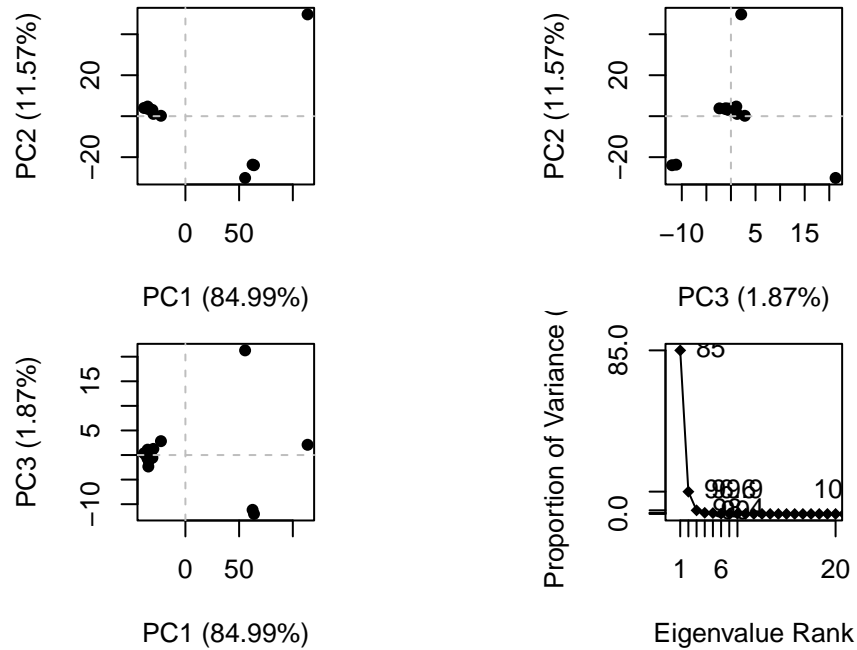
```
13 sequence rows; 227 position columns (204 non-gap, 23 gap)
```

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



##Principal Component Analysis

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



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

Custom analysis of resulting models

```
results_dir <- "hivpr_dimer_23119"

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

library(bio3d)

# Read all data from Models
```

```
# and superpose/fit coords
```

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

Reading PDB files:

```
hivpr_dimer_23119/hivpr_dimer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_1_seed_001
hivpr_dimer_23119/hivpr_dimer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_001
hivpr_dimer_23119/hivpr_dimer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_001
hivpr_dimer_23119/hivpr_dimer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_2_seed_001
hivpr_dimer_23119/hivpr_dimer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_3_seed_001
.....
```

Extracting sequences

```
pdb/seq: 1 name: hivpr_dimer_23119/hivpr_dimer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_1_seed_001
pdb/seq: 2 name: hivpr_dimer_23119/hivpr_dimer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_001
pdb/seq: 3 name: hivpr_dimer_23119/hivpr_dimer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_001
pdb/seq: 4 name: hivpr_dimer_23119/hivpr_dimer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_2_seed_001
pdb/seq: 5 name: hivpr_dimer_23119/hivpr_dimer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_3_seed_001
```

pdbbs

```
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
*****
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
*****
51 . . . . 100
```

```

101 . . . . 150
[Truncated_Name:1]hivpr_dime QITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPMIGGIG
[Truncated_Name:2]hivpr_dime QITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPMIGGIG
[Truncated_Name:3]hivpr_dime QITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPMIGGIG
[Truncated_Name:4]hivpr_dime QITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPMIGGIG
[Truncated_Name:5]hivpr_dime QITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPMIGGIG
*****
101 . . . . 150

151 . . . . 198
[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
*****
151 . . . . 198

```

Call:

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

Class:

```
pdb, 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(pdb)
```

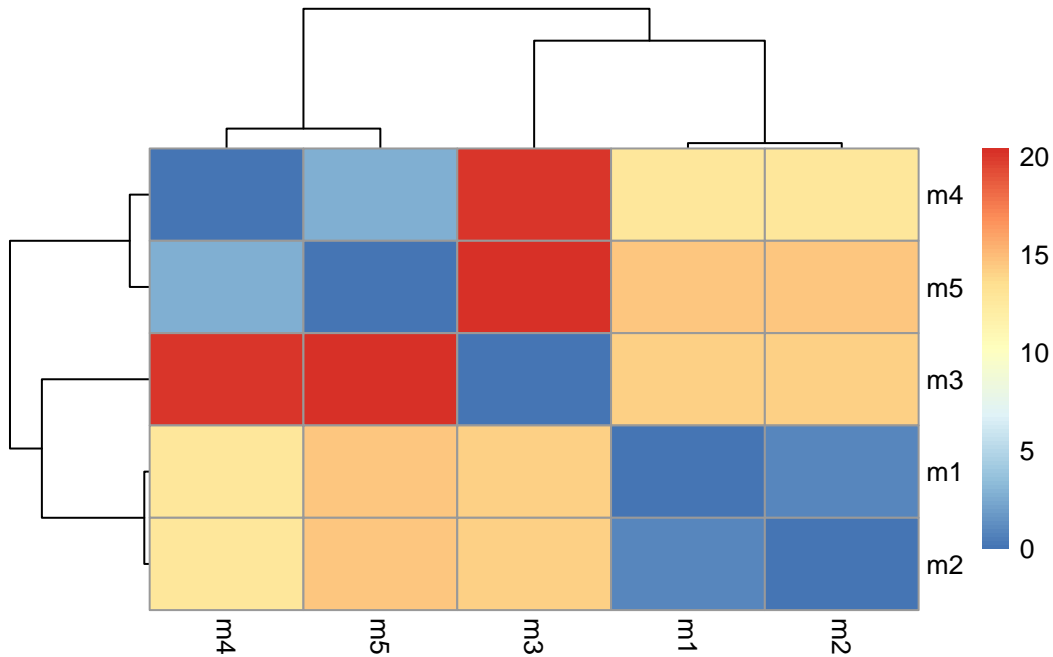
Warning in rmsd(pdb): No indices provided, using the 198 non NA positions

```
range(rd)
```

```
[1] 0.000 20.431
```

```
library(pheatmap)
```

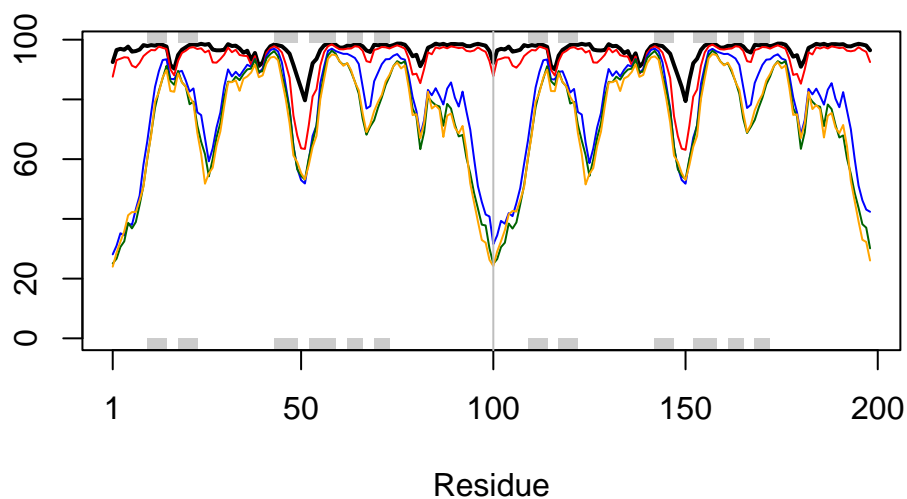
```
colnames(rd) <- paste0("m",1:5)
rownames(rd) <- paste0("m",1:5)
pheatmap(rd)
```



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

Note: Accessing on-line PDB file

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



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

```
core size 197 of 198 vol = 6154.856
core size 196 of 198 vol = 5399.691
core size 195 of 198 vol = 5074.812
core size 194 of 198 vol = 4802.535
core size 193 of 198 vol = 4520.271
core size 192 of 198 vol = 4305.376
core size 191 of 198 vol = 4089.806
core size 190 of 198 vol = 3886.157
core size 189 of 198 vol = 3758.332
core size 188 of 198 vol = 3620.19
core size 187 of 198 vol = 3496.708
core size 186 of 198 vol = 3389.995
core size 185 of 198 vol = 3320.123
core size 184 of 198 vol = 3258.693
core size 183 of 198 vol = 3208.601
core size 182 of 198 vol = 3156.745
core size 181 of 198 vol = 3141.677
core size 180 of 198 vol = 3136.582
core size 179 of 198 vol = 3155.527
core size 178 of 198 vol = 3185.368
```

core size 177 of 198 vol = 3204.492
core size 176 of 198 vol = 3211.981
core size 175 of 198 vol = 3234.994
core size 174 of 198 vol = 3244.062
core size 173 of 198 vol = 3237.844
core size 172 of 198 vol = 3218.77
core size 171 of 198 vol = 3180.743
core size 170 of 198 vol = 3130.369
core size 169 of 198 vol = 3067.881
core size 168 of 198 vol = 2989.546
core size 167 of 198 vol = 2928.271
core size 166 of 198 vol = 2851.193
core size 165 of 198 vol = 2780.877
core size 164 of 198 vol = 2708.433
core size 163 of 198 vol = 2636.515
core size 162 of 198 vol = 2563.25
core size 161 of 198 vol = 2478.024
core size 160 of 198 vol = 2404.793
core size 159 of 198 vol = 2330.997
core size 158 of 198 vol = 2250.477
core size 157 of 198 vol = 2159.432
core size 156 of 198 vol = 2070.759
core size 155 of 198 vol = 1983.579
core size 154 of 198 vol = 1917.913
core size 153 of 198 vol = 1842.556
core size 152 of 198 vol = 1775.398
core size 151 of 198 vol = 1695.133
core size 150 of 198 vol = 1632.172
core size 149 of 198 vol = 1570.391
core size 148 of 198 vol = 1497.238
core size 147 of 198 vol = 1434.802
core size 146 of 198 vol = 1367.705
core size 145 of 198 vol = 1302.596
core size 144 of 198 vol = 1251.985
core size 143 of 198 vol = 1207.975
core size 142 of 198 vol = 1167.112
core size 141 of 198 vol = 1118.27
core size 140 of 198 vol = 1081.663
core size 139 of 198 vol = 1029.749
core size 138 of 198 vol = 981.765
core size 137 of 198 vol = 944.445
core size 136 of 198 vol = 899.223
core size 135 of 198 vol = 859.402

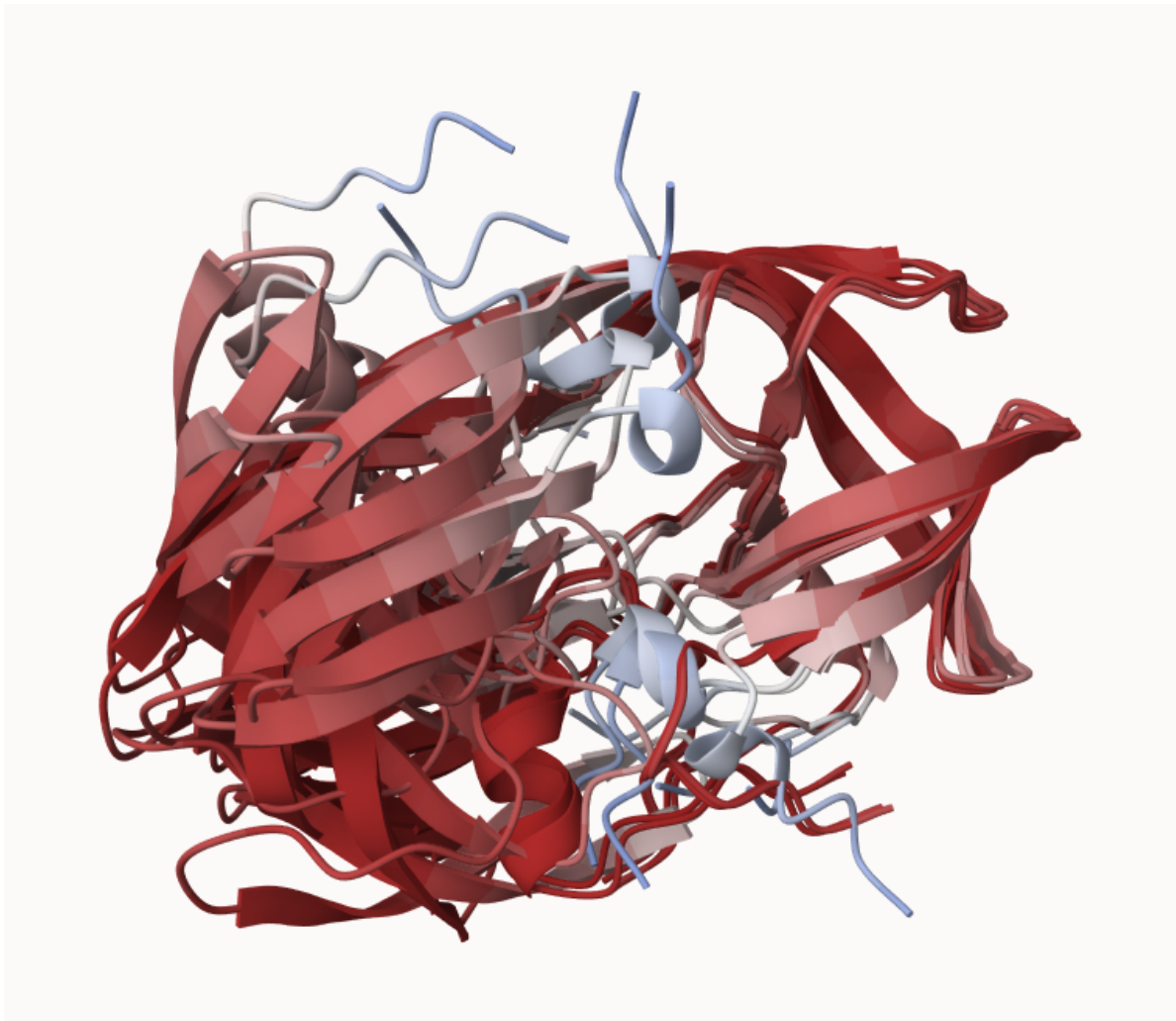
core size 134 of 198 vol = 814.693
core size 133 of 198 vol = 771.861
core size 132 of 198 vol = 733.807
core size 131 of 198 vol = 702.053
core size 130 of 198 vol = 658.757
core size 129 of 198 vol = 622.574
core size 128 of 198 vol = 578.29
core size 127 of 198 vol = 543.07
core size 126 of 198 vol = 510.934
core size 125 of 198 vol = 481.595
core size 124 of 198 vol = 464.672
core size 123 of 198 vol = 451.721
core size 122 of 198 vol = 430.417
core size 121 of 198 vol = 409.141
core size 120 of 198 vol = 378.942
core size 119 of 198 vol = 348.325
core size 118 of 198 vol = 324.739
core size 117 of 198 vol = 312.394
core size 116 of 198 vol = 300.89
core size 115 of 198 vol = 279.976
core size 114 of 198 vol = 263.434
core size 113 of 198 vol = 250.263
core size 112 of 198 vol = 229.592
core size 111 of 198 vol = 209.929
core size 110 of 198 vol = 196.379
core size 109 of 198 vol = 180.628
core size 108 of 198 vol = 167.088
core size 107 of 198 vol = 155.875
core size 106 of 198 vol = 142.595
core size 105 of 198 vol = 128.923
core size 104 of 198 vol = 114.054
core size 103 of 198 vol = 100.936
core size 102 of 198 vol = 90.43
core size 101 of 198 vol = 81.972
core size 100 of 198 vol = 74.017
core size 99 of 198 vol = 66.855
core size 98 of 198 vol = 59.525
core size 97 of 198 vol = 52.263
core size 96 of 198 vol = 43.699
core size 95 of 198 vol = 35.813
core size 94 of 198 vol = 28.888
core size 93 of 198 vol = 20.692
core size 92 of 198 vol = 14.976

```
core size 91 of 198 vol = 9.146
core size 90 of 198 vol = 5.232
core size 89 of 198 vol = 3.53
core size 88 of 198 vol = 2.657
core size 87 of 198 vol = 1.998
core size 86 of 198 vol = 1.333
core size 85 of 198 vol = 1.141
core size 84 of 198 vol = 1.012
core size 83 of 198 vol = 0.891
core size 82 of 198 vol = 0.749
core size 81 of 198 vol = 0.618
core size 80 of 198 vol = 0.538
core size 79 of 198 vol = 0.479
FINISHED: Min vol ( 0.5 ) reached
```

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

```
# 80 positions (cumulative volume <= 0.5 Angstrom^3)
  start end length
1    10  25     16
2    27  48     22
3    53  94     42
```

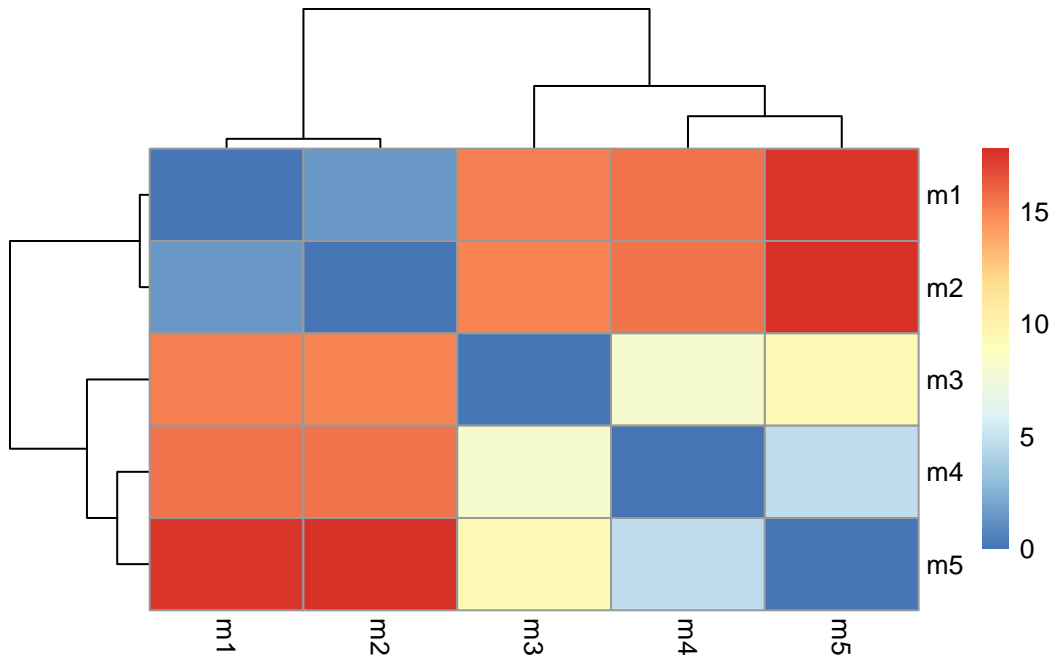
```
xyz <- pdbfit(pdb, core.inds, outpath="corefit_structures")
```



```
rd <- rmsd(xyz)
```

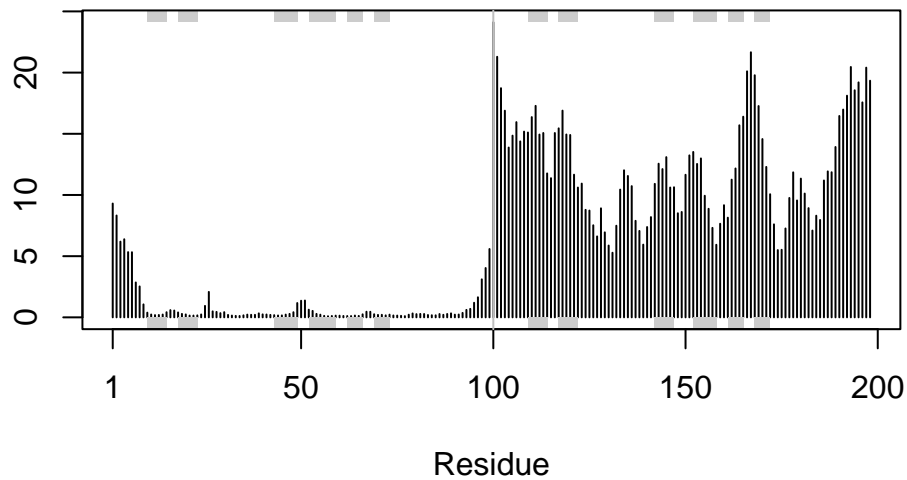
Warning in rmsd(xyz): No indices provided, using the 198 non NA positions

```
# Change the names for easy reference  
colnames(rd) <- paste0("m",1:5)  
rownames(rd) <- paste0("m",1:5)  
pheatmap(rd)
```



```
rf <- rmsf(xyz)

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



Predicted Alignment Error for domains

```
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
```

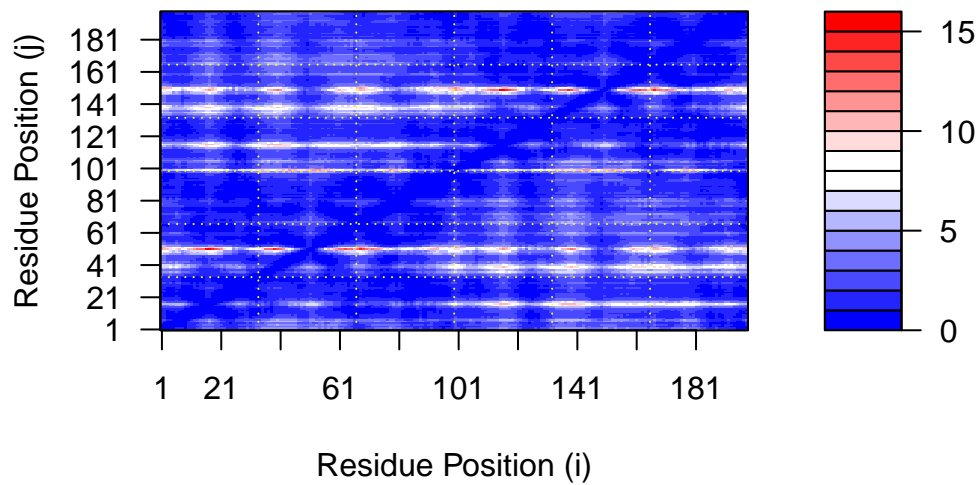
```
pae1$max_pae
```

```
[1] 15.54688
```

```
pae5$max_pae
```

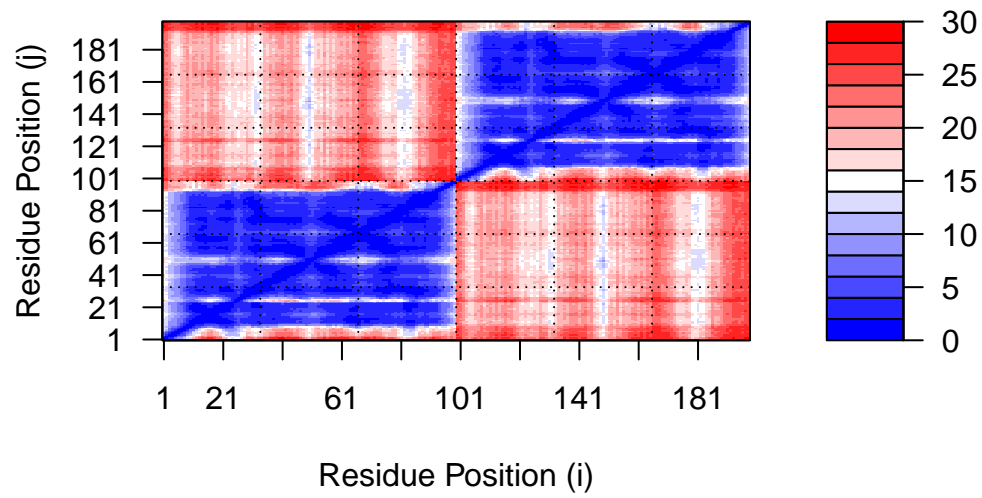
```
[1] 29.29688
```

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

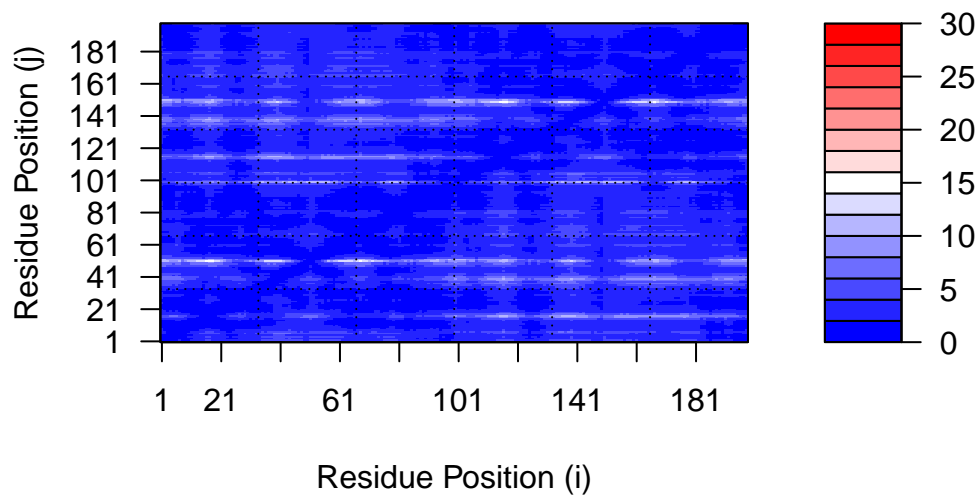


```
plot.dmat(pae5$pae,  
          xlab="Residue Position (i)",  
          ylab="Residue Position (j)",  
          grid.col = "black",
```

```
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] "hivpr_dimer_23119/hivpr_dimer_23119.a3m"
```

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

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

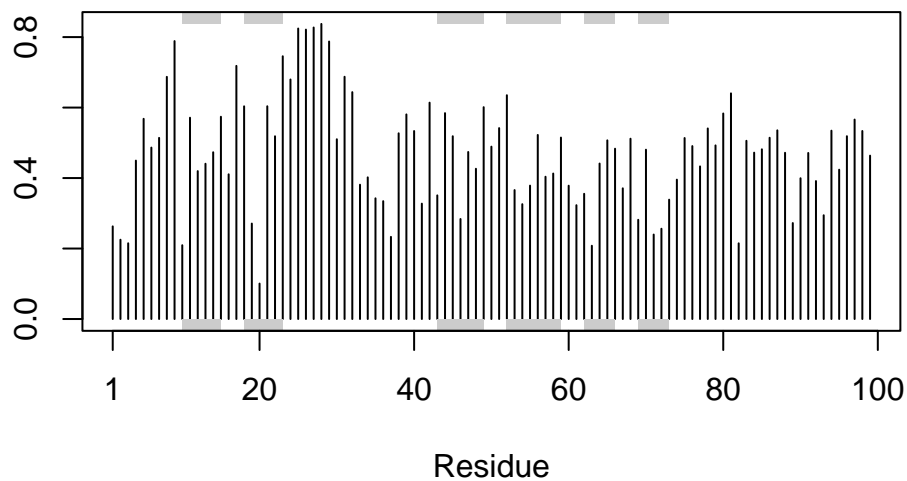
```
[2] " ** Duplicated sequence id's: 101 **"
```

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

```
[1] 5378 132
```



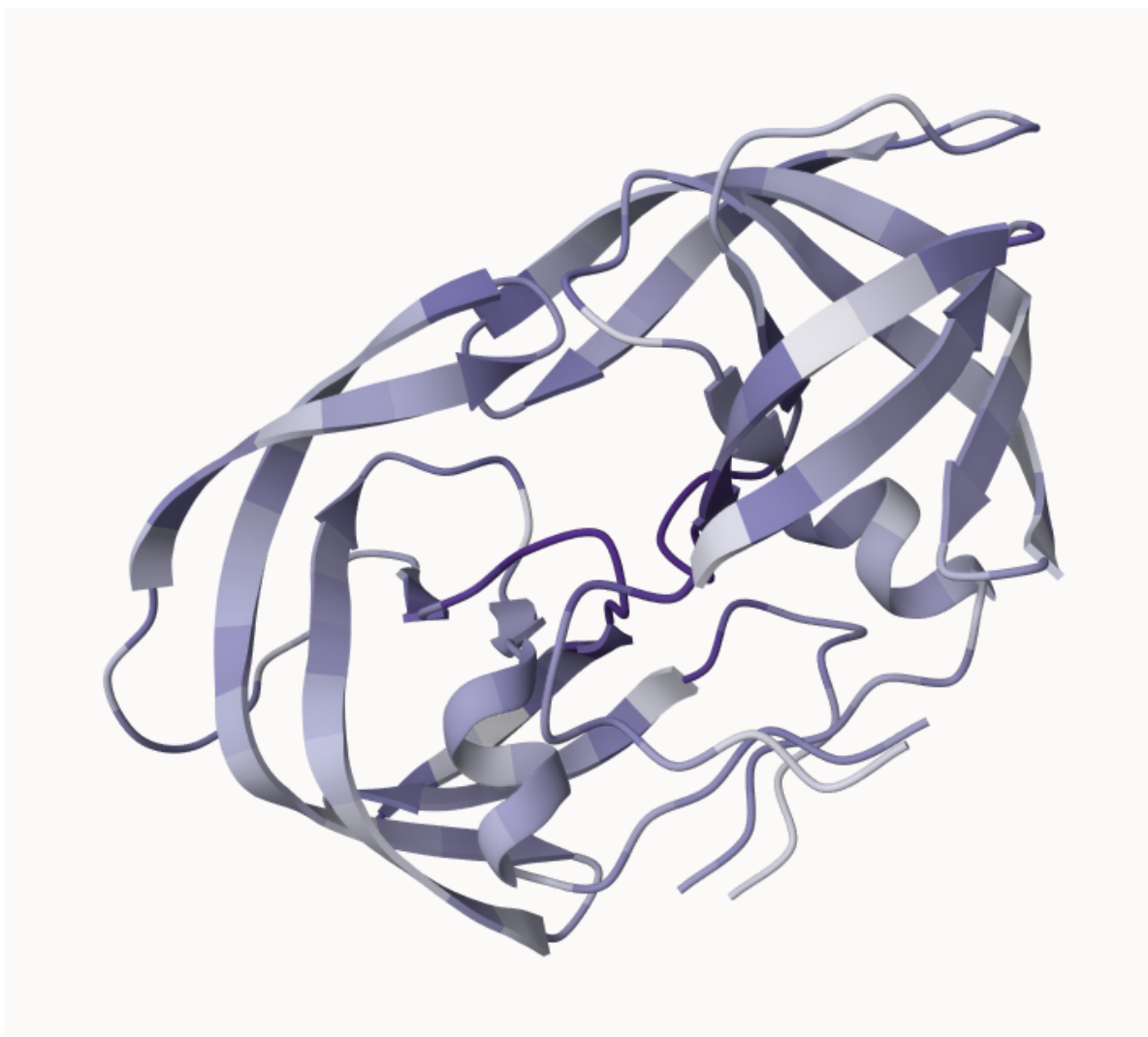
```
sim <- conserv(aln)
plotb3(sim[1:99], sse=trim.pdb(pdb, chain="A"))
```



```
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")
```



```
sessionInfo()
```

```
R version 4.3.1 (2023-06-16)  
Platform: aarch64-apple-darwin20 (64-bit)  
Running under: macOS Sonoma 14.1
```

```
Matrix products: default
```

```
BLAS:   /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/lib/libRblas.0.dylib  
LAPACK: /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/lib/libRlapack.dylib;
```

```
locale:
```

```
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
```

```
time zone: America/Los_Angeles
tzcode source: internal
```

```
attached base packages:
```

```
[1] stats      graphics  grDevices  utils      datasets  methods    base
```

```
other attached packages:
```

```
[1] jsonlite_1.8.7  pheatmap_1.0.12 bio3d_2.4-4
```

```
loaded via a namespace (and not attached):
```

[1] crayon_1.5.2	httr_1.4.7	cli_3.6.1
[4] knitr_1.45	rlang_1.1.1	xfun_0.41
[7] glue_1.6.2	S4Vectors_0.40.1	colorspace_2.1-0
[10] RCurl_1.98-1.13	Biostrings_2.70.1	htmltools_0.5.7
[13] stats4_4.3.1	scales_1.2.1	rmarkdown_2.25
[16] grid_4.3.1	munsell_0.5.0	evaluate_0.23
[19] bitops_1.0-7	fastmap_1.1.1	lifecycle_1.0.3
[22] yaml_2.3.7	IRanges_2.36.0	GenomeInfoDb_1.38.0
[25] compiler_4.3.1	RColorBrewer_1.1-3	Rcpp_1.0.11
[28] XVector_0.42.0	digest_0.6.33	R6_2.5.1
[31] GenomeInfoDbData_1.2.11	curl_5.1.0	parallel_4.3.1
[34] gtable_0.3.4	tools_4.3.1	zlibbioc_1.48.0
[37] msa_1.34.0	BiocGenerics_0.48.1	