

class 11: Structural Bioinformatics pt2

Mudit

Today, before delving into structure prediction with alphafold we will finish off our previous lab10 “comparitive structure analysis” section

```
library(bio3d)

id <- "1ake_A"

aa <- get.seq(id)
```

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

Fetching... Please wait. Done.

aa

```

      1      .      .      .      .      .      60
pdb|1AKE|A  MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMMLRAAVKSGSELGKQAKDIMDAGKLV
      1      .      .      .      .      .      60

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

      121      .      .      .      .      .      180
pdb|1AKE|A  VGRRVHAPSGRVYHVKFNPVKVEGKDDVTGEELTRKDDQEETVRKRLVEYHQMTAPLIG
      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
```

```
# PDB search the sequence
b <- blast.pdb(aa)
```

```
Searching ... please wait (updates every 5 seconds) RID = K3S6EA87013
.....
Reporting 85 hits
```

```
# check what is in b
attributes(b)
```

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

```
$class
[1] "blast"
```

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

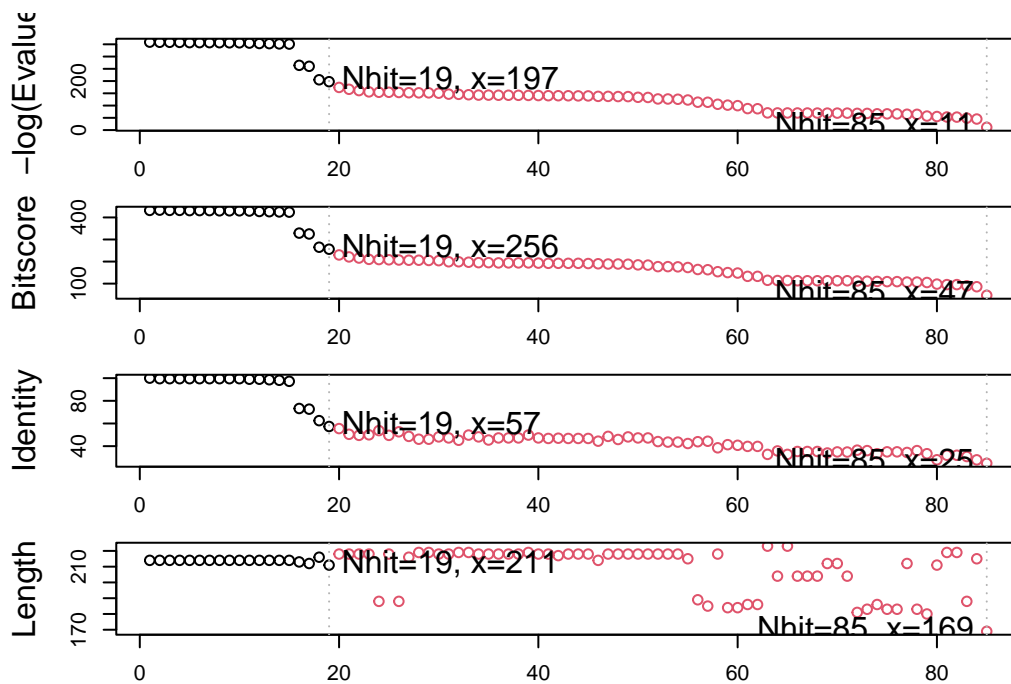
	queryid	subjectids	identity	alignmentlength	mismatches	gapopens	q.start		
1	Query_2098351	1AKE_A	100.000	214	0	0	1		
2	Query_2098351	8BQF_A	99.533	214	1	0	1		
3	Query_2098351	4X8M_A	99.533	214	1	0	1		
4	Query_2098351	6S36_A	99.533	214	1	0	1		
5	Query_2098351	8Q2B_A	99.533	214	1	0	1		
6	Query_2098351	8RJ9_A	99.533	214	1	0	1		
	q.end	s.start	s.end	eval	bitscore	positives	mlog.eval	pdb.id	acc
1	214	1	214	1.58e-156	432	100.00	358.7458	1AKE_A	1AKE_A
2	214	21	234	2.58e-156	433	100.00	358.2555	8BQF_A	8BQF_A

3	214	1	214	2.82e-156	432	100.00	358.1665	4X8M_A	4X8M_A
4	214	1	214	4.14e-156	432	100.00	357.7826	6S36_A	6S36_A
5	214	1	214	1.10e-155	431	99.53	356.8054	8Q2B_A	8Q2B_A
6	214	1	214	1.10e-155	431	99.53	356.8054	8RJ9_A	8RJ9_A

```
hist <- plot(b)
```

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

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

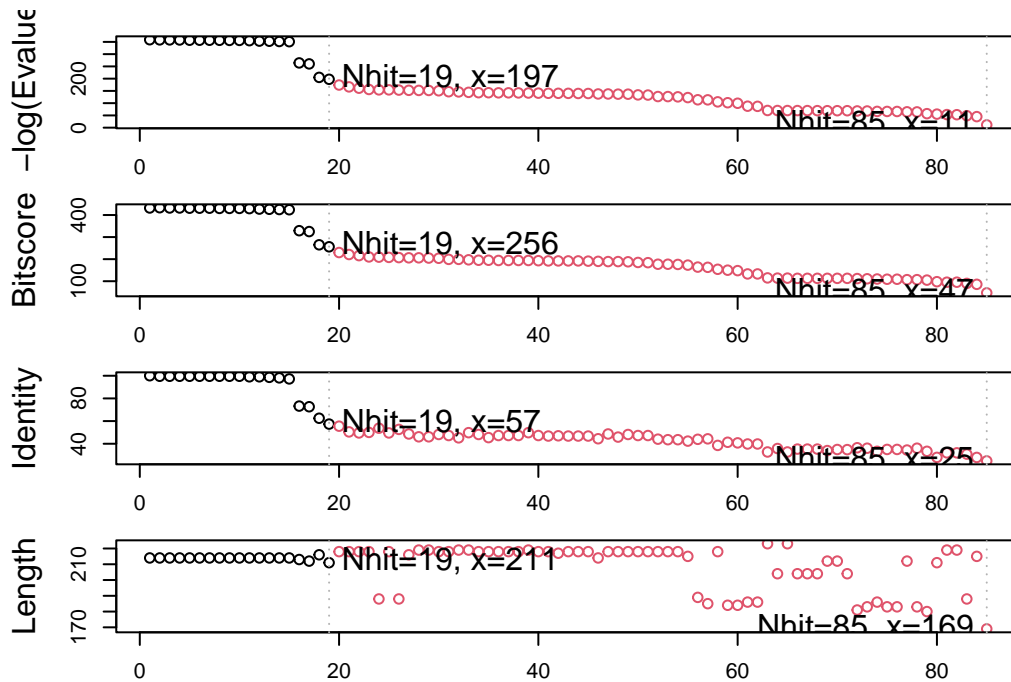


-log <- larger the better

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

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

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



```
attributes(hits)
```

```
$names
[1] "hits"    "pdb.id"  "acc"     "inds"

$class
[1] "blast"
```

Top hits that we like from our blast results:

```
hits$pdb.id
```

```
[1] "1AKE_A" "8BQF_A" "4X8M_A" "6S36_A" "8Q2B_A" "8RJ9_A" "6RZE_A" "4X8H_A"
[9] "3HPR_A" "1E4V_A" "5EJE_A" "1E4Y_A" "3X2S_A" "6HAP_A" "6HAM_A" "4K46_A"
[17] "4NP6_A" "3GMT_A" "4PZL_A"
```

```
# Download releated PDB files
files <- get.pdb(hits$pdb.id, path="pds", split=TRUE, gzip=TRUE)
```

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

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

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

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

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

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

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

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

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

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

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

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

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

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

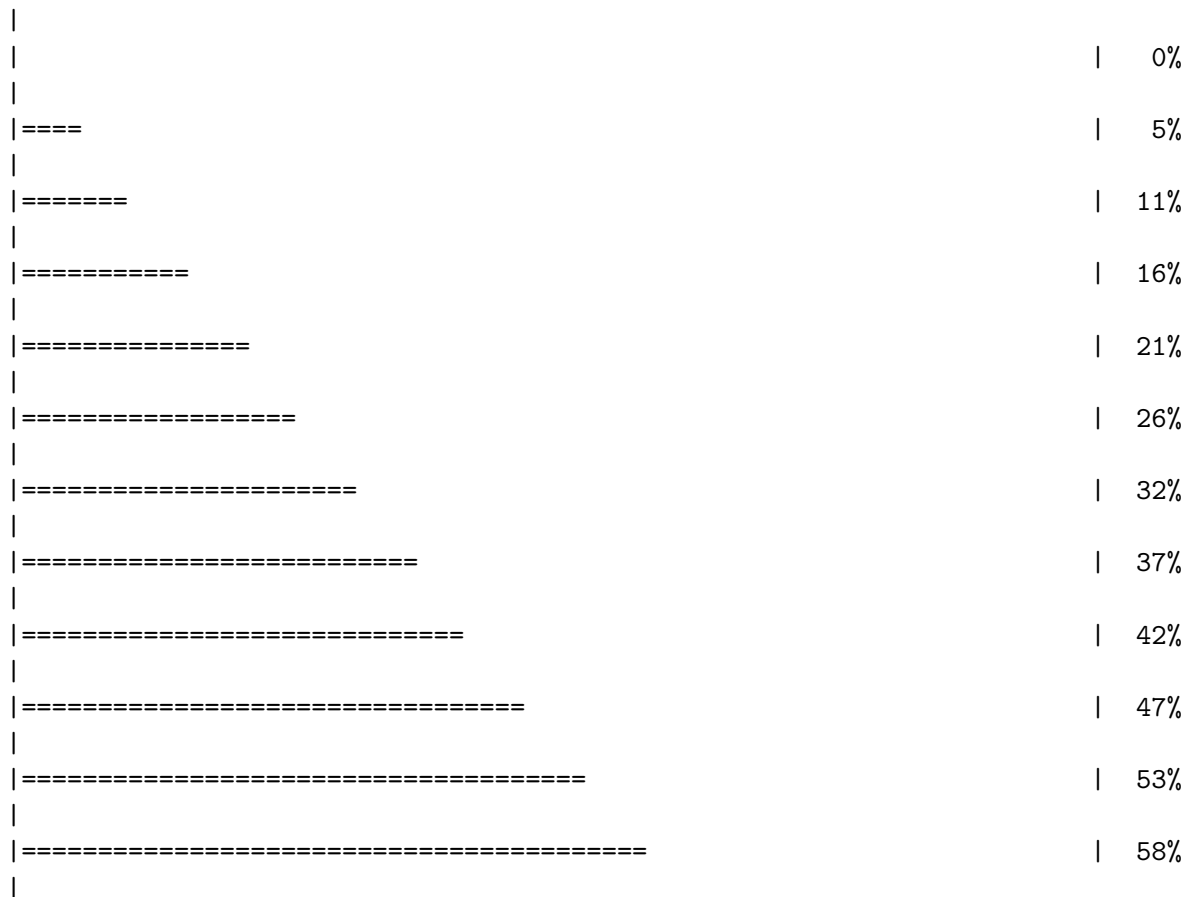
Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/6HAM.pdb.gz exists. Skipping download

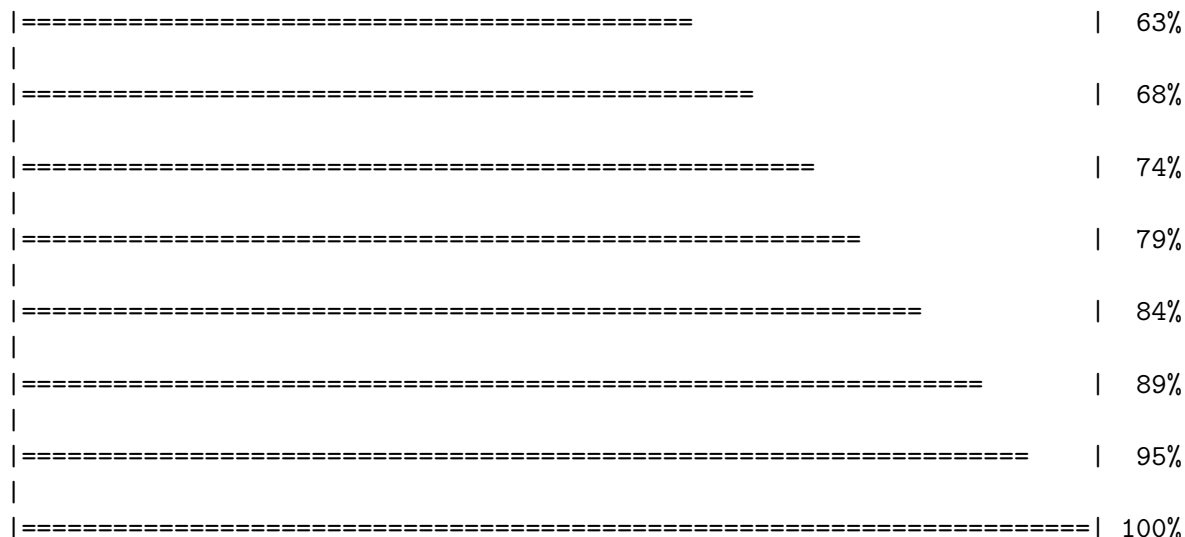
Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/4K46.pdb.gz exists. Skipping download

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

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

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





I have now found and downloaded all ADK structures in the PDB database but viewing them is difficult as they need to be aligned and superposed.

(Need MUSCLE but need package) I am going to install BiocManager package from CRAN

Then I can use `BiocManager::install()` to install any bioconductor package.

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

Reading PDB files:

```
pdbbs/split_chain/1AKE_A.pdb
pdbbs/split_chain/8BQF_A.pdb
pdbbs/split_chain/4X8M_A.pdb
pdbbs/split_chain/6S36_A.pdb
pdbbs/split_chain/8Q2B_A.pdb
pdbbs/split_chain/8RJ9_A.pdb
pdbbs/split_chain/6RZE_A.pdb
pdbbs/split_chain/4X8H_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/4NP6_A.pdb
```

```

pdbc/split_chain/3GMT_A.pdb
pdbc/split_chain/4PZL_A.pdb
    PDB has ALT records, taking A only, rm.alt=TRUE
.   PDB has ALT records, taking A only, rm.alt=TRUE
..  PDB has ALT records, taking A only, rm.alt=TRUE
.   PDB has ALT records, taking A only, rm.alt=TRUE
.   PDB has ALT records, taking A only, rm.alt=TRUE
.   PDB has ALT records, taking A only, rm.alt=TRUE
..  PDB has ALT records, taking A only, rm.alt=TRUE
..  PDB has ALT records, taking A only, rm.alt=TRUE
.... PDB has ALT records, taking A only, rm.alt=TRUE
.   PDB has ALT records, taking A only, rm.alt=TRUE
....

```

Extracting sequences

```

pdb/seq: 1   name: pdbc/split_chain/1AKE_A.pdb
    PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 2   name: pdbc/split_chain/8BQF_A.pdb
    PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 3   name: pdbc/split_chain/4X8M_A.pdb
pdb/seq: 4   name: pdbc/split_chain/6S36_A.pdb
    PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 5   name: pdbc/split_chain/8Q2B_A.pdb
    PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 6   name: pdbc/split_chain/8RJ9_A.pdb
    PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 7   name: pdbc/split_chain/6RZE_A.pdb
    PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 8   name: pdbc/split_chain/4X8H_A.pdb
pdb/seq: 9   name: pdbc/split_chain/3HPR_A.pdb
    PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 10  name: pdbc/split_chain/1E4V_A.pdb
pdb/seq: 11  name: pdbc/split_chain/5EJE_A.pdb
    PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 12  name: pdbc/split_chain/1E4Y_A.pdb
pdb/seq: 13  name: pdbc/split_chain/3X2S_A.pdb
pdb/seq: 14  name: pdbc/split_chain/6HAP_A.pdb
pdb/seq: 15  name: pdbc/split_chain/6HAM_A.pdb
    PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 16  name: pdbc/split_chain/4K46_A.pdb
    PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 17  name: pdbc/split_chain/4NP6_A.pdb

```


pdb/seq: 18 name: pdbs/split_chain/3GMT_A.pdb
 pdb/seq: 19 name: pdbs/split_chain/4PZL_A.pdb

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] 8Q2B_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS			
[Truncated_Name:6] 8RJ9_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS			
[Truncated_Name:7] 6RZE_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS			
[Truncated_Name:8] 4X8H_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS			
[Truncated_Name:9] 3HPR_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS			
[Truncated_Name:10] 1E4V_A.pdb	-----	MRIILLGAPVAGKGTQAQFIMEKYGIPQIS			
[Truncated_Name:11] 5EJE_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS			
[Truncated_Name:12] 1E4Y_A.pdb	-----	MRIILLGALVAGKGTQAQFIMEKYGIPQIS			
[Truncated_Name:13] 3X2S_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS			
[Truncated_Name:14] 6HAP_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS			
[Truncated_Name:15] 6HAM_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS			
[Truncated_Name:16] 4K46_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMAKFGIPQIS			
[Truncated_Name:17] 4NP6_A.pdb	-----	NAMRIILLGAPGAGKGTQAQFIMEKFGIPQIS			
[Truncated_Name:18] 3GMT_A.pdb	-----	MRLILLGAPGAGKGTQANFIKEKFGIPQIS			
[Truncated_Name:19] 4PZL_A.pdb		TENLYFQSNAMRIILLGAPGAGKGTQAKIIEQYNI AHIS			
		~*** ***** * *~* **			
	1	.	.	.	40
	41	.	.	.	80
[Truncated_Name:1] 1AKE_A.pdb		TGDM LRAAVKSGSELGKQAKDIMDAGKLVTD ELVIALVKE			
[Truncated_Name:2] 8BQF_A.pdb		TGDM LRAAVKSGSELGKQAKDIMDAGKLVTD ELVIALVKE			
[Truncated_Name:3] 4X8M_A.pdb		TGDM LRAAVKSGSELGKQAKDIMDAGKLVTD ELVIALVKE			
[Truncated_Name:4] 6S36_A.pdb		TGDM LRAAVKSGSELGKQAKDIMDAGKLVTD ELVIALVKE			
[Truncated_Name:5] 8Q2B_A.pdb		TGDM LRAAVKSGSELGKQAKDIMDAGKLVTD ELVIALVKE			
[Truncated_Name:6] 8RJ9_A.pdb		TGDM LRAAVKSGSELGKQAKDIMDAGKLVTD ELVIALVKE			
[Truncated_Name:7] 6RZE_A.pdb		TGDM LRAAVKSGSELGKQAKDIMDAGKLVTD ELVIALVKE			
[Truncated_Name:8] 4X8H_A.pdb		TGDM LRAAVKSGSELGKQAKDIMDAGKLVTD ELVIALVKE			
[Truncated_Name:9] 3HPR_A.pdb		TGDM LRAAVKSGSELGKQAKDIMDAGKLVTD ELVIALVKE			
[Truncated_Name:10] 1E4V_A.pdb		TGDM LRAAVKSGSELGKQAKDIMDAGKLVTD ELVIALVKE			
[Truncated_Name:11] 5EJE_A.pdb		TGDM LRAAVKSGSELGKQAKDIMDAGKLVTD ELVIALVKE			
[Truncated_Name:12] 1E4Y_A.pdb		TGDM LRAAVKSGSELGKQAKDIMDAGKLVTD ELVIALVKE			
[Truncated_Name:13] 3X2S_A.pdb		TGDM LRAAVKSGSELGKQAKDIMDCGKLVTD ELVIALVKE			

[Truncated_Name:14] 6HAP_A.pdb	TGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVRE
[Truncated_Name:15] 6HAM_A.pdb	TGDMRLRAAIKSGSELGKQAKDIMDAGKLVTDDEIIIALVKE
[Truncated_Name:16] 4K46_A.pdb	TGDMRLRAAIKAGTELGKQAKSVIDAGQLVSDDIILGLVKE
[Truncated_Name:17] 4NP6_A.pdb	TGDMRLRAAIKAGTELGKQAKAVIDAGQLVSDDIILGLIKE
[Truncated_Name:18] 3GMT_A.pdb	TGDMRLRAAVKAGTPLGVEAKTYMDEGKLVPDSLIIIGLVKE
[Truncated_Name:19] 4PZL_A.pdb	TGDMIRETIKSGSALGQELKKVLDAGELVSDEFIIKIVKD
	****~* ~* *~ ** * ~* ** * ~ ~ ~ ~
	41 . . . 80
	81 . . . 120
[Truncated_Name:1] 1AKE_A.pdb	RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:2] 8BQF_A.pdb	RIAQE---GFLLDGFPR TIPQADAMKEAGINVDYVIEFD
[Truncated_Name:3] 4X8M_A.pdb	RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:4] 6S36_A.pdb	RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:5] 8Q2B_A.pdb	RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:6] 8RJ9_A.pdb	RIAQEDCRNGFLLAGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:7] 6RZE_A.pdb	RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:8] 4X8H_A.pdb	RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:9] 3HPR_A.pdb	RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:10] 1E4V_A.pdb	RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:11] 5EJE_A.pdb	RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:12] 1E4Y_A.pdb	RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:13] 3X2S_A.pdb	RIAQEDSRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:14] 6HAP_A.pdb	RICQEDSRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:15] 6HAM_A.pdb	RICQEDSRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:16] 4K46_A.pdb	RIAQDDCAKGFLLDGFPR TIPQADGLKEVG VVVVDYVIEFD
[Truncated_Name:17] 4NP6_A.pdb	RIAQADCEKGFLLDGFPR TIPQADGLKEMGINVDYVIEFD
[Truncated_Name:18] 3GMT_A.pdb	RLKEADCANGYLF DGFPR TIPQADAMKEAGVAIDYVLEID
[Truncated_Name:19] 4PZL_A.pdb	RISKNDCNNGFLLDGVPR TIPQAQELDKLGVNIDYIVEVD
	*~ *~* * **** ** ^ *~ ~*~*~* *
	81 . . . 120
	121 . . . 160
[Truncated_Name:1] 1AKE_A.pdb	VPDELIVDRIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG
[Truncated_Name:2] 8BQF_A.pdb	VPDELIVDRIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG
[Truncated_Name:3] 4X8M_A.pdb	VPDELIVDRIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG
[Truncated_Name:4] 6S36_A.pdb	VPDELIVDKIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG
[Truncated_Name:5] 8Q2B_A.pdb	VPDELIVDRIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG
[Truncated_Name:6] 8RJ9_A.pdb	VPDELIVDRIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG
[Truncated_Name:7] 6RZE_A.pdb	VPDELIVDAIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG
[Truncated_Name:8] 4X8H_A.pdb	VPDELIVDRIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG
[Truncated_Name:9] 3HPR_A.pdb	VPDELIVDRIVGRRVHAPSGRVYHV KFNPPKVEGKDDGTG
[Truncated_Name:10] 1E4V_A.pdb	VPDELIVDRIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG

[Truncated_Name:11] 5EJE_A.pdb	VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG			
[Truncated_Name:12] 1E4Y_A.pdb	VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG			
[Truncated_Name:13] 3X2S_A.pdb	VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG			
[Truncated_Name:14] 6HAP_A.pdb	VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG			
[Truncated_Name:15] 6HAM_A.pdb	VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTG			
[Truncated_Name:16] 4K46_A.pdb	VADSVIVERMAGRRAHLASGRTYHNVPNPPKVEGKDDVTG			
[Truncated_Name:17] 4NP6_A.pdb	VADDVIVERMAGRRAHLPSGRTYHVYNPPKVEGKDDVTG			
[Truncated_Name:18] 3GMT_A.pdb	VPFSEIIERMSSGRRTHPASGRTYHVKNPPKVEGKDDVTG			
[Truncated_Name:19] 4PZL_A.pdb	VADNLLIERITGRRIH PASGRTYHTKFNNPKVADKDDVTG			
	* ^^^ ^ *** * *** ** ^***** *** **			
121	.	.	.	160
	161	.	.	200
[Truncated_Name:1] 1AKE_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYSKEAEAGN			
[Truncated_Name:2] 8BQF_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYSKEAEAGN			
[Truncated_Name:3] 4X8M_A.pdb	EELTTRKDDQEETVRKRLVEWHQMTAPLIGYSKEAEAGN			
[Truncated_Name:4] 6S36_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYSKEAEAGN			
[Truncated_Name:5] 8Q2B_A.pdb	EELTTRKADQEETVRKRLVEYHQMTAPLIGYSKEAEAGN			
[Truncated_Name:6] 8RJ9_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYSKEAEAGN			
[Truncated_Name:7] 6RZE_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYSKEAEAGN			
[Truncated_Name:8] 4X8H_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAALIGYSKEAEAGN			
[Truncated_Name:9] 3HPR_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYSKEAEAGN			
[Truncated_Name:10] 1E4V_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYSKEAEAGN			
[Truncated_Name:11] 5EJE_A.pdb	EELTTRKDDQEECVRKRLVEYHQMTAPLIGYSKEAEAGN			
[Truncated_Name:12] 1E4Y_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYSKEAEAGN			
[Truncated_Name:13] 3X2S_A.pdb	EELTTRKDDQEETVRKRLCEYHQMTAPLIGYSKEAEAGN			
[Truncated_Name:14] 6HAP_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYSKEAEAGN			
[Truncated_Name:15] 6HAM_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYSKEAEAGN			
[Truncated_Name:16] 4K46_A.pdb	EDLVIREDDKEETVLARLG VYHNQTAPLIA YYGKEAEAGN			
[Truncated_Name:17] 4NP6_A.pdb	EDLVIREDDKEETVRARLN VYHTQTAPLIEYYGKEAAAGK			
[Truncated_Name:18] 3GMT_A.pdb	EPLVQRDDDKEETVKKRLDVYE AQT KPLITYYGDWARRGA			
[Truncated_Name:19] 4PZL_A.pdb	EPLITRTDDNEDTVKQRSVYHAQTAKLIDFYRNFSSTNT			
	* * * * * ^ * ** ^ * ** ^*			
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] 8Q2B_A.pdb	T--KYAKVDGTKPVAEVRADLEKILG-			
[Truncated_Name:6] 8RJ9_A.pdb	T--KYAKVDGTKPVAEVRADLEKILG-			
[Truncated_Name:7] 6RZE_A.pdb	T--KYAKVDGTKPVAEVRADLEKILG-			

```

[Truncated_Name:8]4X8H_A.pdb      T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:9]3HPR_A.pdb      T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:10]1E4V_A.pdb     T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:11]5EJE_A.pdb     T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:12]1E4Y_A.pdb     T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:13]3X2S_A.pdb     T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:14]6HAP_A.pdb     T--KYAKVDGTPVCEVRADLEKILG-
[Truncated_Name:15]6HAM_A.pdb     T--KYAKVDGTPVCEVRADLEKILG-
[Truncated_Name:16]4K46_A.pdb     T--QYLKFDGTKAVAEVSAELEKALA-
[Truncated_Name:17]4NP6_A.pdb     T--QYLKFDGTKQVSEVSADIAKALA-
[Truncated_Name:18]3GMT_A.pdb     E-----NGLKAPA-----YRKISG-
[Truncated_Name:19]4PZL_A.pdb     KIPKYIKINGDQAVEKVSQDIFDQLNK
                                     *
                                201      .      .      227

```

Call:

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

Class:

```
pdbs, fasta
```

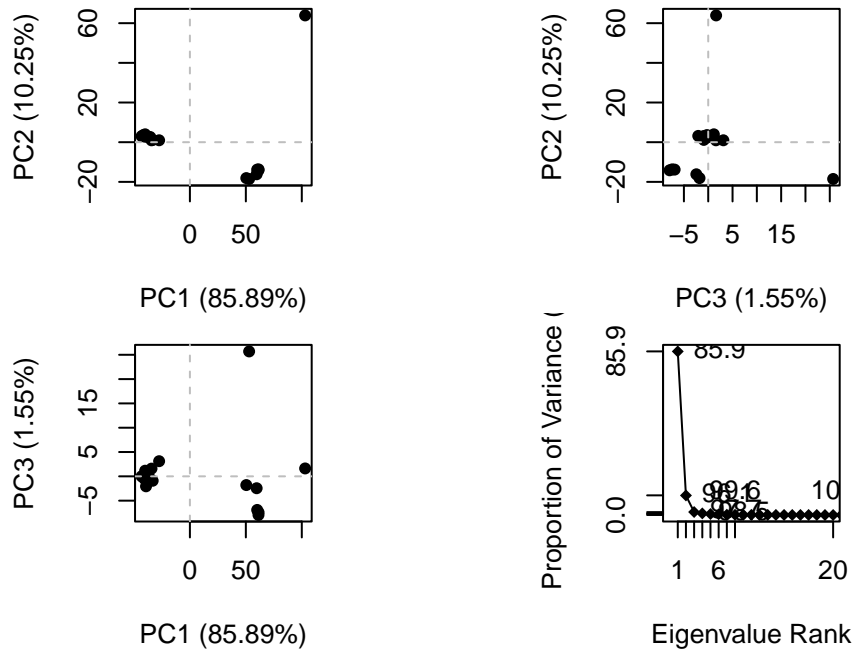
Alignment dimensions:

```
19 sequence rows; 227 position columns (199 non-gap, 28 gap)
```

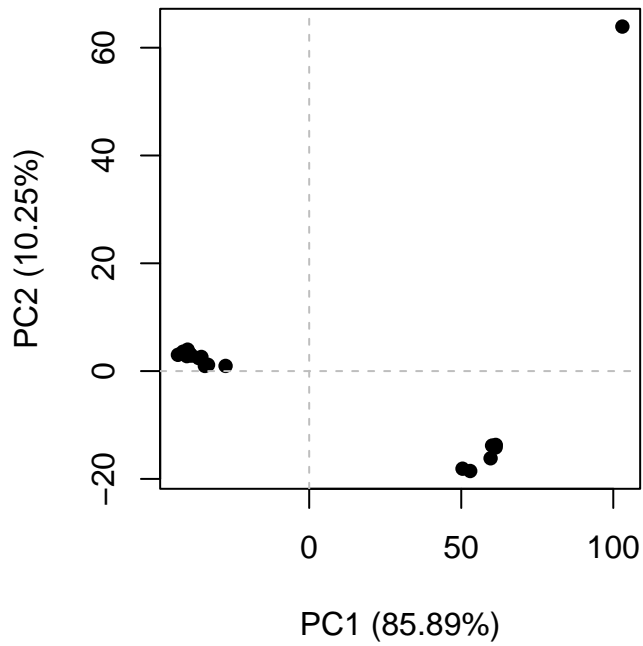
```
+ attr: xyz, resno, b, chain, id, ali, resid, sse, call
```

Principle Component Analysis

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



```
plot(pc, pc.axes = c(1:2))
```



To examine in more detail what PC1 (or any PC2) is capturing here we can plot the loading or make a wee (small) movie (trajectory) of moving along PC1.

```
#create a pdb file
mktrj(pc, pc = 1, file = "pc1.pdb")
```

Custom analysis of resulting models

```
results_dir <- "/Users/muditg19/Downloads/BGGN213/Class 11/dimer_test_82aef"

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

# Check if PDB files were found
if (length(pdb_files) == 0) {
  stop("No PDB files found in the specified directory.")
}

# Print file names
print(basename(pdb_files))
```

```
[1] "dimer_test_82aef_unrelaxed_rank_001_alphafold2_ptm_model_3_seed_000.pdb"
[2] "dimer_test_82aef_unrelaxed_rank_002_alphafold2_ptm_model_4_seed_000.pdb"
[3] "dimer_test_82aef_unrelaxed_rank_003_alphafold2_ptm_model_5_seed_000.pdb"
[4] "dimer_test_82aef_unrelaxed_rank_004_alphafold2_ptm_model_1_seed_000.pdb"
[5] "dimer_test_82aef_unrelaxed_rank_005_alphafold2_ptm_model_2_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:

```
/Users/muditg19/Downloads/BGGN213/Class 11/dimer_test_82aef/dimer_test_82aef_unrelaxed_rank_0
/Users/muditg19/Downloads/BGGN213/Class 11/dimer_test_82aef/dimer_test_82aef_unrelaxed_rank_0
/Users/muditg19/Downloads/BGGN213/Class 11/dimer_test_82aef/dimer_test_82aef_unrelaxed_rank_0
/Users/muditg19/Downloads/BGGN213/Class 11/dimer_test_82aef/dimer_test_82aef_unrelaxed_rank_0
/Users/muditg19/Downloads/BGGN213/Class 11/dimer_test_82aef/dimer_test_82aef_unrelaxed_rank_0
.....
```

Extracting sequences

```

pdb/seq: 1   name: /Users/muditg19/Downloads/BGGN213/Class 11/dimer_test_82aef/dimer_test_82
pdb/seq: 2   name: /Users/muditg19/Downloads/BGGN213/Class 11/dimer_test_82aef/dimer_test_82
pdb/seq: 3   name: /Users/muditg19/Downloads/BGGN213/Class 11/dimer_test_82aef/dimer_test_82
pdb/seq: 4   name: /Users/muditg19/Downloads/BGGN213/Class 11/dimer_test_82aef/dimer_test_82
pdb/seq: 5   name: /Users/muditg19/Downloads/BGGN213/Class 11/dimer_test_82aef/dimer_test_82

```

pdbs

```

1           .           .           .           .           50
[Truncated_Name:1]dimer_test AKSDCVVDNFHVKENFNKTRYTGKWFYAFKKDPRGLFLETDIRADFRIN
[Truncated_Name:2]dimer_test AKSDCVVDNFHVKENFNKTRYTGKWFYAFKKDPRGLFLETDIRADFRIN
[Truncated_Name:3]dimer_test AKSDCVVDNFHVKENFNKTRYTGKWFYAFKKDPRGLFLETDIRADFRIN
[Truncated_Name:4]dimer_test AKSDCVVDNFHVKENFNKTRYTGKWFYAFKKDPRGLFLETDIRADFRIN
[Truncated_Name:5]dimer_test AKSDCVVDNFHVKENFNKTRYTGKWFYAFKKDPRGLFLETDIRADFRIN
*****
1           .           .           .           .           50

51          .           .           .           .           100
[Truncated_Name:1]dimer_test GTMIAKAKGRVTLLPDWVCAEMMGTFNDTENPAKFQLKYWGAAEYLQKG
[Truncated_Name:2]dimer_test GTMIAKAKGRVTLLPDWVCAEMMGTFNDTENPAKFQLKYWGAAEYLQKG
[Truncated_Name:3]dimer_test GTMIAKAKGRVTLLPDWVCAEMMGTFNDTENPAKFQLKYWGAAEYLQKG
[Truncated_Name:4]dimer_test GTMIAKAKGRVTLLPDWVCAEMMGTFNDTENPAKFQLKYWGAAEYLQKG
[Truncated_Name:5]dimer_test GTMIAKAKGRVTLLPDWVCAEMMGTFNDTENPAKFQLKYWGAAEYLQKG
*****
51          .           .           .           .           100

101         .           .           .           .           150
[Truncated_Name:1]dimer_test NDDHWIIDTDYDTYAITYTCRKLYENRTCADSYFVFSRNPQGLNPETRQ
[Truncated_Name:2]dimer_test NDDHWIIDTDYDTYAITYTCRKLYENRTCADSYFVFSRNPQGLNPETRQ
[Truncated_Name:3]dimer_test NDDHWIIDTDYDTYAITYTCRKLYENRTCADSYFVFSRNPQGLNPETRQ
[Truncated_Name:4]dimer_test NDDHWIIDTDYDTYAITYTCRKLYENRTCADSYFVFSRNPQGLNPETRQ
[Truncated_Name:5]dimer_test NDDHWIIDTDYDTYAITYTCRKLYENRTCADSYFVFSRNPQGLNPETRQ
*****
101         .           .           .           .           150

151         .           .           174
[Truncated_Name:1]dimer_test VVRKWQDHICLAFKYKRVLQSDSC
[Truncated_Name:2]dimer_test VVRKWQDHICLAFKYKRVLQSDSC
[Truncated_Name:3]dimer_test VVRKWQDHICLAFKYKRVLQSDSC
[Truncated_Name:4]dimer_test VVRKWQDHICLAFKYKRVLQSDSC
[Truncated_Name:5]dimer_test VVRKWQDHICLAFKYKRVLQSDSC

```

```
*****  
151      .      174
```

Call:

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

Class:

```
pdb, fasta
```

Alignment dimensions:

```
5 sequence rows; 174 position columns (174 non-gap, 0 gap)
```

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

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

Warning in rmsd(pdb, fit = T): No indices provided, using the 174 non NA positions

```
range(rd)
```

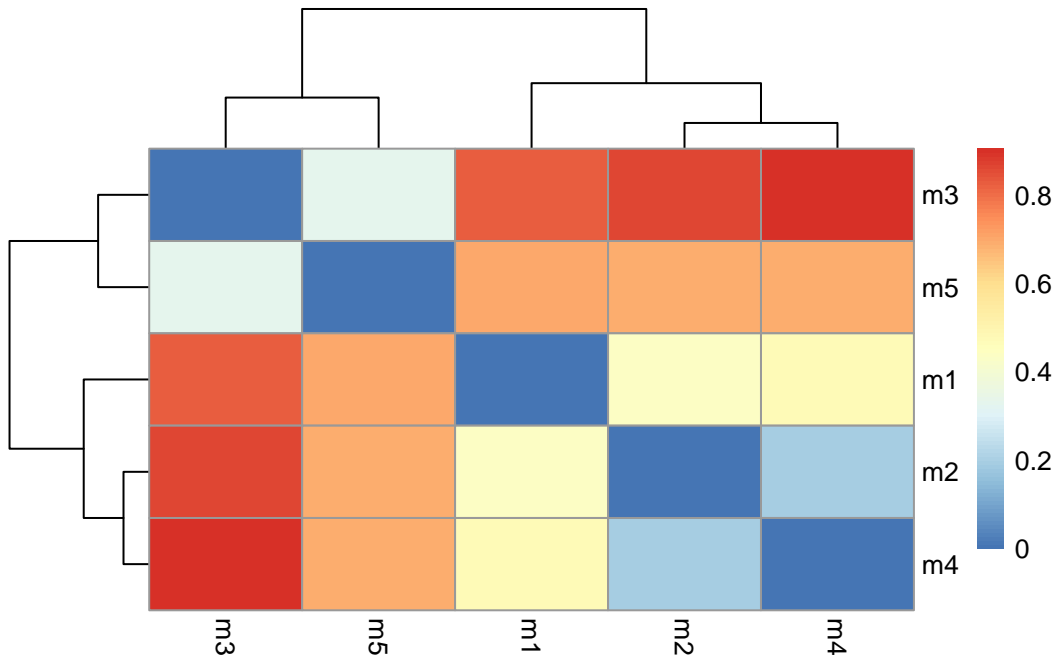
```
[1] 0.000 0.905
```

```
library(pheatmap)
```

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

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

```
pheatmap(rd)
```

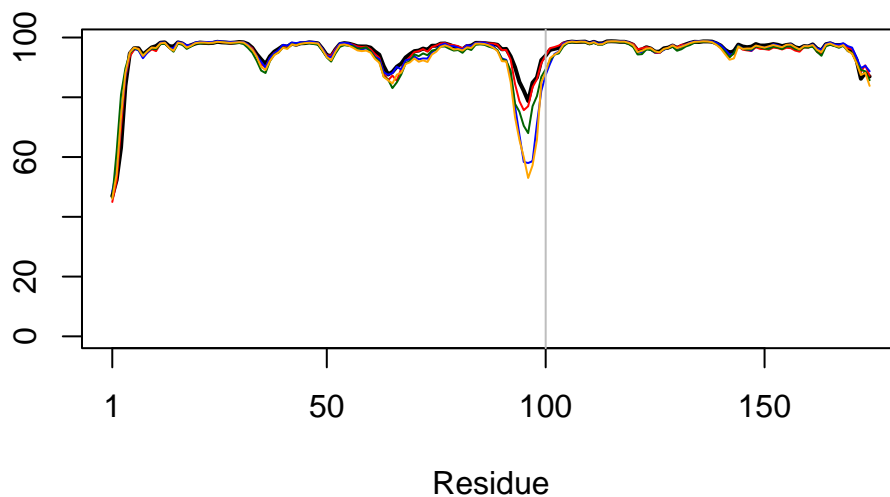
```
# Read a reference PDB structure
pdb <- read.pdb("1hsg")
```

Note: Accessing on-line PDB file

```
plotb3(pdb$b[1,], typ="l", lwd=2, sse=pdb)
```

Warning in plotb3(pdb\$b[1,], typ = "l", lwd = 2, sse = pdb): Length of input 'sse' does not equal the length of input 'x'; Ignoring 'sse'

```
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 173 of 174  vol = 0.817
core size 172 of 174  vol = 0.615
core size 171 of 174  vol = 0.432
FINISHED: Min vol ( 0.5 ) reached
```

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

```
# 172 positions (cumulative volume <= 0.5 Angstrom^3)
  start end length
1     1  94     94
2    96  96      1
3    98 174     77
```

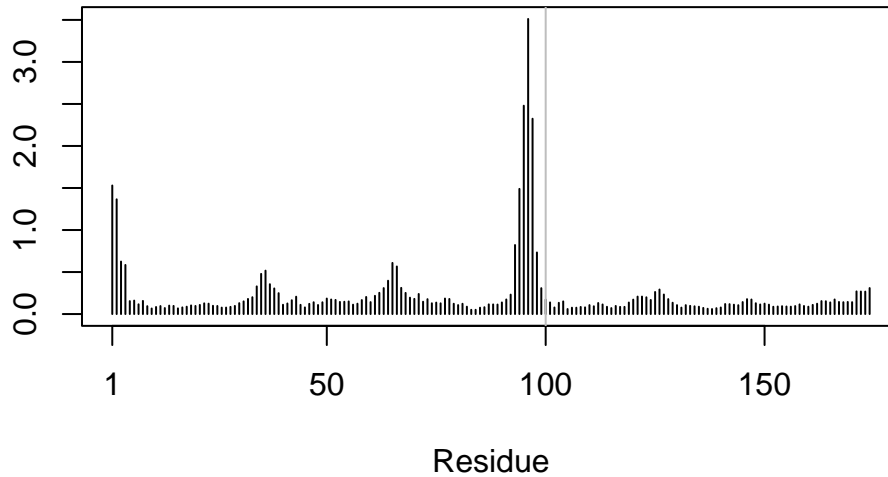
```
xyz <- pdbfit(pdb, core.inds, outpath="corefit_structures")
```

```
rf <- rmsf(xyz)
```

```
plotb3(rf, sse=pdb)
```

Warning in plotb3(rf, sse = pdb): Length of input 'sse' does not equal the length of input 'x'; Ignoring 'sse'

```
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"
```

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

```
[1] 46.91 52.34 63.16 83.81 94.69 96.44
```

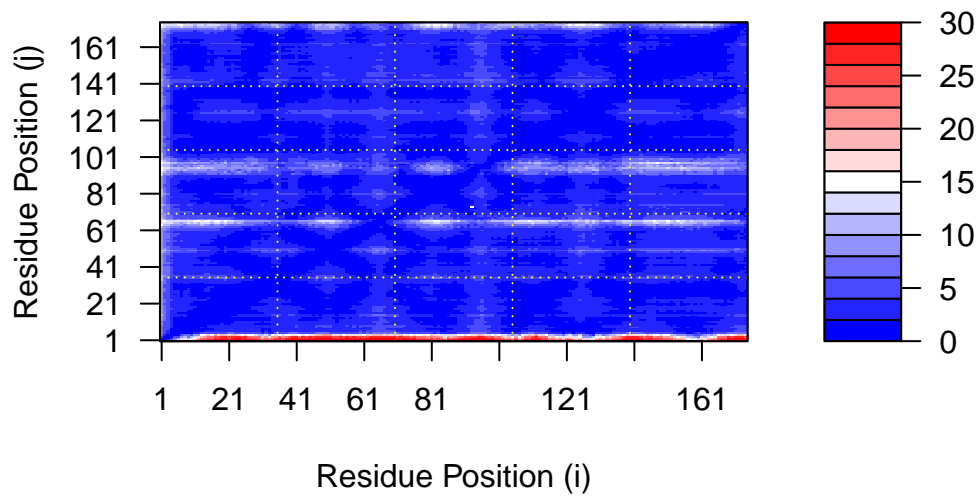
```
pae1$max_pae
```

```
[1] 29.78125
```

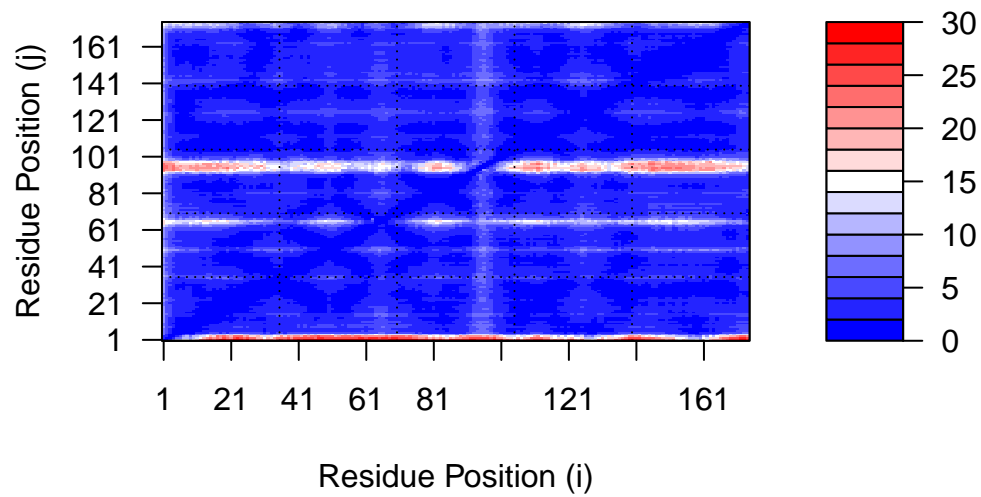
```
pae5$max_pae
```

```
[1] 29.67188
```

```
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))
```

