

class11

Kai Zhao (PID: A17599942)

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("BiocManager")` `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  MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMRLAAVKSSELGKQAKDIMDAGKLV
      1      .      .      .      .      .      .      60
      61      .      .      .      .      .      .      120
pdb|1AKE|A  DELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRI
      61      .      .      .      .      .      .      120
      121      .      .      .      .      .      .      180
pdb|1AKE|A  VGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTRKDDQEETVRKRLVEYHQMTAPLIG
      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)
```

Searching ... please wait (updates every 5 seconds) RID = MN9A6M4V013

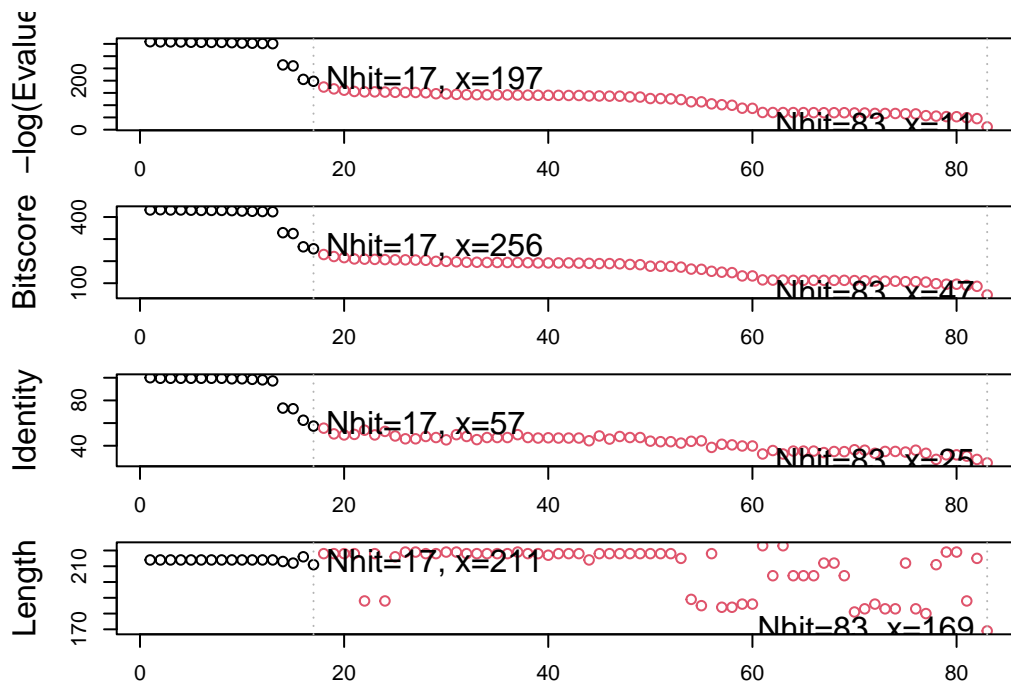
.....

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_81707	1AKE_A	100.000	214	0	0	1
2	Query_81707	8BQF_A	99.533	214	1	0	1
3	Query_81707	4X8M_A	99.533	214	1	0	1
4	Query_81707	6S36_A	99.533	214	1	0	1
5	Query_81707	6RZE_A	99.533	214	1	0	1
6	Query_81707	4X8H_A	99.533	214	1	0	1

	q.end	s.start	s.end	evaluate	bitscore	positives	mlog.evaluate	pdb.id	acc
1	214	1	214	1.45e-156	432	100.00	358.8317	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

These 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 they are, what species they are from, stuff about the experiment they were solved in etc.)

For this we can use `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	ligandId
1AKE_A	2.00	Adenylate kinase	Adenylate kinase (ADK)	AP5
8BQF_A	2.05	<NA>	Adenylate kinase (ADK)	AP5
4X8M_A	2.60	<NA>	Adenylate kinase (ADK)	<NA>
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)
4X8H_A	2.50	<NA>	Adenylate kinase (ADK)	<NA>

	ligandName	source
1AKE_A	BIS(ADENOSINE)-5'-PENTAPHOSPHATE	Escherichia coli
8BQF_A	BIS(ADENOSINE)-5'-PENTAPHOSPHATE	Escherichia coli
4X8M_A	<NA>	Escherichia coli

```

6S36_A CHLORIDE ION (3),SODIUM ION,MAGNESIUM ION (2) Escherichia coli
6RZE_A          SODIUM ION (3),CHLORIDE ION (2) Escherichia coli
4X8H_A                                <NA> Escherichia coli

```

```

1AKE_A STRUCTURE OF THE COMPLEX BETWEEN ADENYLATE KINASE FROM ESCHERICHIA COLI AND THE INHIB
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()` functions.

```

# 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/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

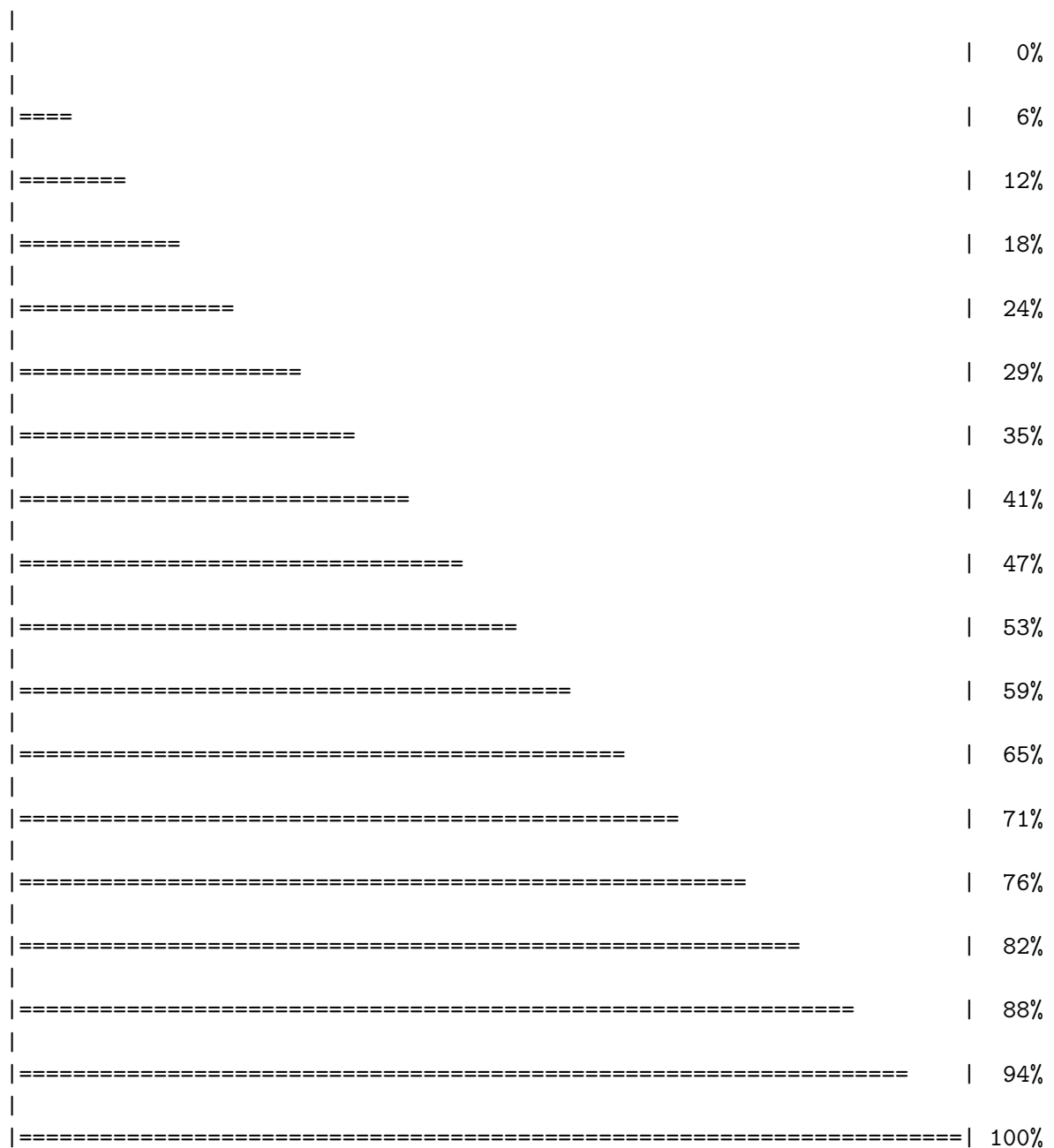
Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/6HAM.pdb.gz exists. Skipping download

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



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

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

Reading PDB files:

pdb/split_chain/1AKE_A.pdb
pdb/split_chain/8BQF_A.pdb
pdb/split_chain/4X8M_A.pdb
pdb/split_chain/6S36_A.pdb
pdb/split_chain/6RZE_A.pdb
pdb/split_chain/4X8H_A.pdb
pdb/split_chain/3HPR_A.pdb
pdb/split_chain/1E4V_A.pdb
pdb/split_chain/5EJE_A.pdb
pdb/split_chain/1E4Y_A.pdb
pdb/split_chain/3X2S_A.pdb
pdb/split_chain/6HAP_A.pdb
pdb/split_chain/6HAM_A.pdb
pdb/split_chain/4K46_A.pdb
pdb/split_chain/4NP6_A.pdb
pdb/split_chain/3GMT_A.pdb
pdb/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: pdb/split_chain/1AKE_A.pdb
 PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 2 name: pdb/split_chain/8BQF_A.pdb
 PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 3 name: pdb/split_chain/4X8M_A.pdb
pdb/seq: 4 name: pdb/split_chain/6S36_A.pdb
 PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 5 name: pdb/split_chain/6RZE_A.pdb
 PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 6 name: pdb/split_chain/4X8H_A.pdb
pdb/seq: 7 name: pdb/split_chain/3HPR_A.pdb
 PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 8 name: pdb/split_chain/1E4V_A.pdb


```

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

```

pdbc

```

[Truncated_Name:1] 1AKE_A.pdb      1          .          .          .          40
-----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
TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDLVIALVKE
[Truncated_Name:1] 1AKE_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDLVIALVKE
[Truncated_Name:2] 8BQF_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDLVIALVKE
[Truncated_Name:3] 4X8M_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDLVIALVKE
[Truncated_Name:4] 6S36_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDLVIALVKE

```

[Truncated_Name:5] 6RZE_A.pdb	TGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKE
[Truncated_Name:6] 4X8H_A.pdb	TGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKE
[Truncated_Name:7] 3HPR_A.pdb	TGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKE
[Truncated_Name:8] 1E4V_A.pdb	TGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKE
[Truncated_Name:9] 5EJE_A.pdb	TGDMRLRAAVKSGSELGKQAKDIMDACKLVTDDELVIALVKE
[Truncated_Name:10] 1E4Y_A.pdb	TGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKE
[Truncated_Name:11] 3X2S_A.pdb	TGDMRLRAAVKSGSELGKQAKDIMDCGKLVTDDELVIALVKE
[Truncated_Name:12] 6HAP_A.pdb	TGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVRE
[Truncated_Name:13] 6HAM_A.pdb	TGDMRLRAAIKSGSELGKQAKDIMDAGKLVTDDEIIIALVKE
[Truncated_Name:14] 4K46_A.pdb	TGDMRLRAAIKAGTELGKQAKSVIDAGQLVSDDIILGLVKE
[Truncated_Name:15] 4NP6_A.pdb	TGDMRLRAAIKAGTELGKQAKAVIDAGQLVSDDIILGLIKE
[Truncated_Name:16] 3GMT_A.pdb	TGDMRLRAAVKAGTPLGVEAKTYMDEGKLVPSDLIIIGLVKE
[Truncated_Name:17] 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] 6RZE_A.pdb	RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:6] 4X8H_A.pdb	RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:7] 3HPR_A.pdb	RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:8] 1E4V_A.pdb	RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:9] 5EJE_A.pdb	RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:10] 1E4Y_A.pdb	RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:11] 3X2S_A.pdb	RIAQEDSRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:12] 6HAP_A.pdb	RICQEDSRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:13] 6HAM_A.pdb	RICQEDSRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:14] 4K46_A.pdb	RIAQDDCAKGFLLDGFPR TIPQADGLKEVG VVDYVIEFD
[Truncated_Name:15] 4NP6_A.pdb	RIAQADCEKGFLLDGFPR TIPQADGLKEMGINVDYVIEFD
[Truncated_Name:16] 3GMT_A.pdb	RLKEADCANGYLF DGFPR TIPQADAMKEAGVAIDYVLEID
[Truncated_Name:17] 4PZL_A.pdb	RISKNDCNNGFLLDGVPR TIPQAQELDKLG VNI DYI VEVD
	*~ *~* ** ***** ** ^ *~ ^***^* *
	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] 6RZE_A.pdb	VPDELIVDAIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG

[Truncated_Name:6] 4X8H_A.pdb	VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKD DVTG
[Truncated_Name:7] 3HPR_A.pdb	VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKD DGTG
[Truncated_Name:8] 1E4V_A.pdb	VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKD DVTG
[Truncated_Name:9] 5EJE_A.pdb	VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKD DVTG
[Truncated_Name:10] 1E4Y_A.pdb	VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKD DVTG
[Truncated_Name:11] 3X2S_A.pdb	VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKD DVTG
[Truncated_Name:12] 6HAP_A.pdb	VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKD DVTG
[Truncated_Name:13] 6HAM_A.pdb	VPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKD DVTG
[Truncated_Name:14] 4K46_A.pdb	VADSVIVERMAGRRAHLASGRTYHNVPNPKVEGKD DVTG
[Truncated_Name:15] 4NP6_A.pdb	VADDVIVERMAGRRAHLPSGRTYHVVPNPKVEGKD DVTG
[Truncated_Name:16] 3GMT_A.pdb	VPFSEI IERMSGRRTHPASGRTYHVKNPPKVEGKD DVTG
[Truncated_Name:17] 4PZL_A.pdb	VADNLLIERITGRIHPASGRTYHTKFNNPKVADKDDVTG
	* ^^^ ^ *** * *** ** ^***** *** **
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] 6RZE_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYSKEAEAGN
[Truncated_Name:6] 4X8H_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAALIGYSKEAEAGN
[Truncated_Name:7] 3HPR_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYSKEAEAGN
[Truncated_Name:8] 1E4V_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYSKEAEAGN
[Truncated_Name:9] 5EJE_A.pdb	EELTTRKDDQEECVRKRLVEYHQMTAPLIGYSKEAEAGN
[Truncated_Name:10] 1E4Y_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYSKEAEAGN
[Truncated_Name:11] 3X2S_A.pdb	EELTTRKDDQEETVRKRLCEYHQMTAPLIGYSKEAEAGN
[Truncated_Name:12] 6HAP_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYSKEAEAGN
[Truncated_Name:13] 6HAM_A.pdb	EELTTRKDDQEETVRKRLVEYHQMTAPLIGYSKEAEAGN
[Truncated_Name:14] 4K46_A.pdb	EDLVIREDDKEETVLARLG VYHNQTAPLIA YYGKEAEAGN
[Truncated_Name:15] 4NP6_A.pdb	EDLVIREDDKEETVRARLN VYHTQTAPLIE YYGKEAAAGK
[Truncated_Name:16] 3GMT_A.pdb	EPLVQRDDDK EETVKKRLDVYE AQT KPLITYYGDWARRGA
[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--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:8] 1E4V_A.pdb      T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:9] 5EJE_A.pdb      T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:10] 1E4Y_A.pdb     T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:11] 3X2S_A.pdb     T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:12] 6HAP_A.pdb     T--KYAKVDGTPVCEVRADLEKILG-
[Truncated_Name:13] 6HAM_A.pdb     T--KYAKVDGTPVCEVRADLEKILG-
[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)
```

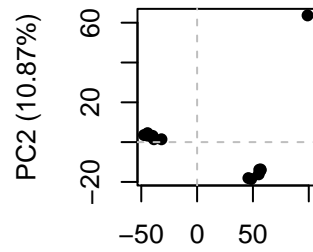
```
+ attr: xyz, resno, b, chain, id, ali, resid, sse, call
```

Principal Component Analysis

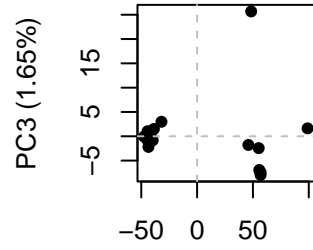
```

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

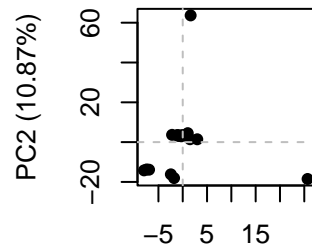
```



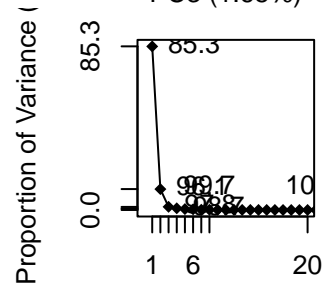
PC1 (85.27%)



PC1 (85.27%)



PC3 (1.65%)



Eigenvalue Rank

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

Visualization of the models and their estimated reliability

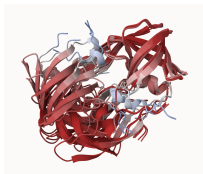
```
getwd()
```

```
[1] "/Users/zhaokai/Desktop/Rbimm/class11"
```

#Monomer



#Dimer



```
results_dir <- "hivprdimer_23119/"
```

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

pdb_files
```

```
[1] "hivprdimer_23119//hivpr_dimer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_1_s
[2] "hivprdimer_23119//hivpr_dimer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_s
[3] "hivprdimer_23119//hivpr_dimer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_s
[4] "hivprdimer_23119//hivpr_dimer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_2_s
[5] "hivprdimer_23119//hivpr_dimer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_3_s
```

```
library(bio3d)
```

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

Reading PDB files:

```
hivprdimer_23119//hivpr_dimer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_1_seed_0
hivprdimer_23119//hivpr_dimer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_0
hivprdimer_23119//hivpr_dimer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_0
hivprdimer_23119//hivpr_dimer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_2_seed_0
hivprdimer_23119//hivpr_dimer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_3_seed_0
.....
```

Extracting sequences

```
pdb/seq: 1    name: hivprdimer_23119//hivpr_dimer_23119_unrelaxed_rank_001_alphafold2_multimer
pdb/seq: 2    name: hivprdimer_23119//hivpr_dimer_23119_unrelaxed_rank_002_alphafold2_multimer
pdb/seq: 3    name: hivprdimer_23119//hivpr_dimer_23119_unrelaxed_rank_003_alphafold2_multimer
pdb/seq: 4    name: hivprdimer_23119//hivpr_dimer_23119_unrelaxed_rank_004_alphafold2_multimer
pdb/seq: 5    name: hivprdimer_23119//hivpr_dimer_23119_unrelaxed_rank_005_alphafold2_multimer
```

```
pdbs
```

```

1 . . . . 50
[Truncated_Name:1]hivpr_dime PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPMIGGI
[Truncated_Name:2]hivpr_dime PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPMIGGI
[Truncated_Name:3]hivpr_dime PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPMIGGI
[Truncated_Name:4]hivpr_dime PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPMIGGI
[Truncated_Name:5]hivpr_dime PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPMIGGI
*****
1 . . . . 50

51 . . . . 100
[Truncated_Name:1]hivpr_dime GGFIVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
[Truncated_Name:2]hivpr_dime GGFIVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
[Truncated_Name:3]hivpr_dime GGFIVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
[Truncated_Name:4]hivpr_dime GGFIVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
[Truncated_Name:5]hivpr_dime GGFIVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
*****
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:

```
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(pdb)
```

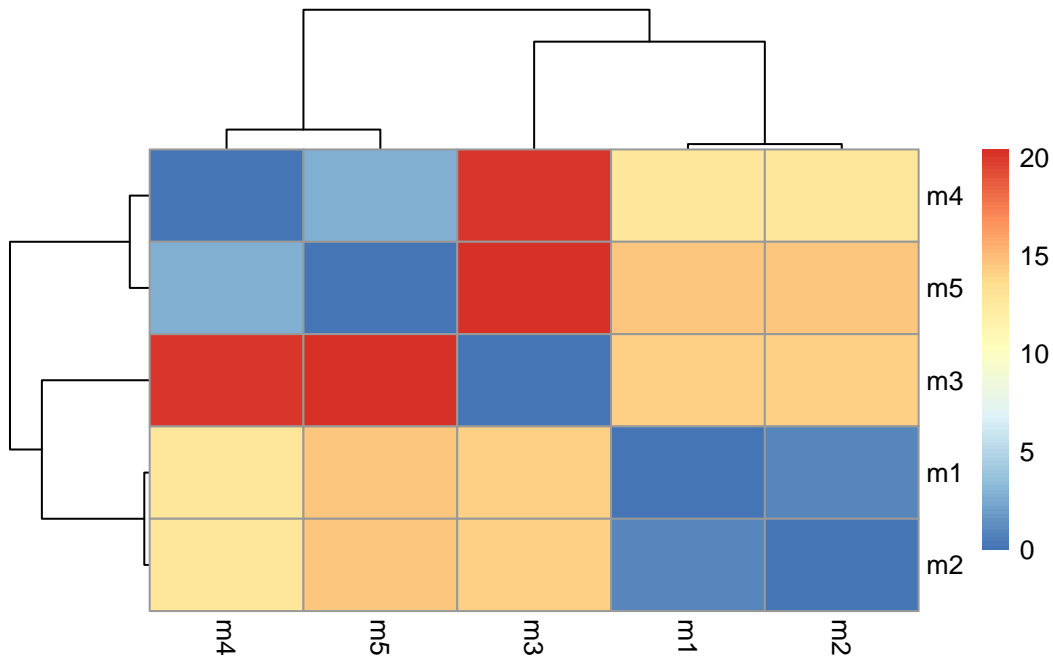
Warning in rmsd(pdb): No indices provided, using the 198 non NA positions

```
range(rd)
```

```
[1] 0.000 20.431
```

```
library(pheatmap)
```

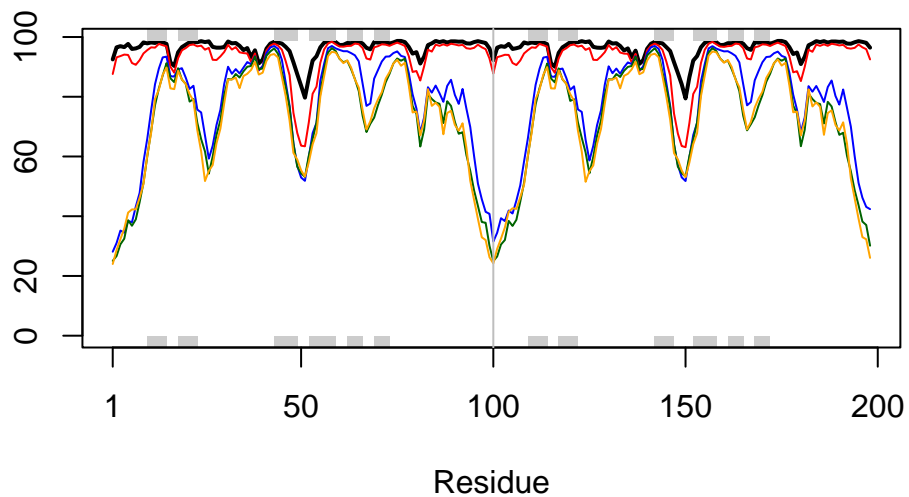
```
colnames(rd) <- paste0("m",1:5)  
rownames(rd) <- paste0("m",1:5)  
pheatmap(rd)
```




```
# 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$b)
```

```
core size 197 of 198  vol = 6154.839
core size 196 of 198  vol = 5399.676
core size 195 of 198  vol = 5074.795
core size 194 of 198  vol = 4802.518
core size 193 of 198  vol = 4520.256
```

core size 192 of 198	vol = 4305.362
core size 191 of 198	vol = 4089.792
core size 190 of 198	vol = 3886.145
core size 189 of 198	vol = 3758.321
core size 188 of 198	vol = 3620.18
core size 187 of 198	vol = 3496.698
core size 186 of 198	vol = 3389.985
core size 185 of 198	vol = 3320.114
core size 184 of 198	vol = 3258.683
core size 183 of 198	vol = 3208.591
core size 182 of 198	vol = 3156.736
core size 181 of 198	vol = 3141.668
core size 180 of 198	vol = 3136.574
core size 179 of 198	vol = 3155.52
core size 178 of 198	vol = 3185.362
core size 177 of 198	vol = 3204.487
core size 176 of 198	vol = 3211.978
core size 175 of 198	vol = 3234.993
core size 174 of 198	vol = 3244.062
core size 173 of 198	vol = 3237.845
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 = 2928.272
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 = 2636.516
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 = 1632.173

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 = 1367.706
core size 145 of 198	vol = 1302.596
core size 144 of 198	vol = 1251.985
core size 143 of 198	vol = 1207.976
core size 142 of 198	vol = 1167.112
core size 141 of 198	vol = 1118.27
core size 140 of 198	vol = 1081.664
core size 139 of 198	vol = 1029.75
core size 138 of 198	vol = 981.766
core size 137 of 198	vol = 944.446
core size 136 of 198	vol = 899.224
core size 135 of 198	vol = 859.402
core size 134 of 198	vol = 814.694
core size 133 of 198	vol = 771.862
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 = 324.738
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 = 128.924
core size 104 of 198  vol = 114.054
core size 103 of 198  vol = 100.936
core size 102 of 198  vol = 90.431
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 = 14.975
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

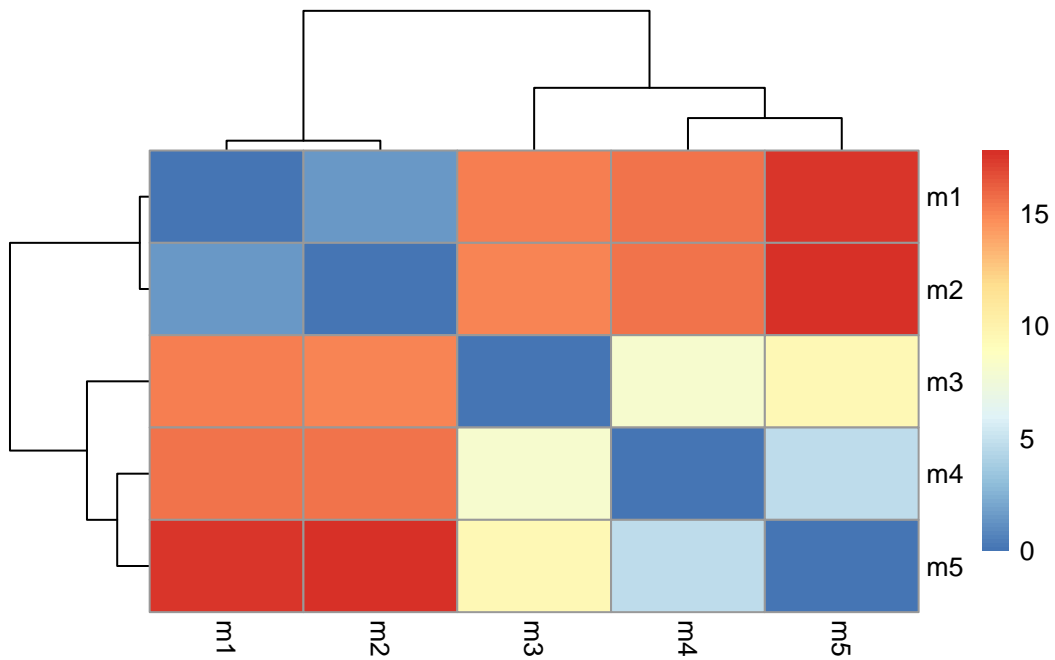
```

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

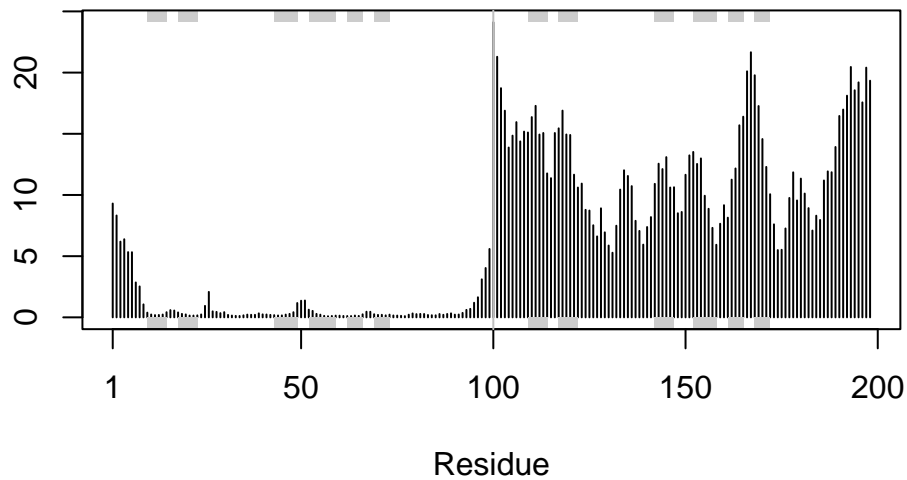
```
rd <- rmsd(xyz)
```

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")
```



```
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
```

The maximum PAE values - we can see that model 5 is much worse than model 1. The lower the better.

```
pae1$max_pae
```

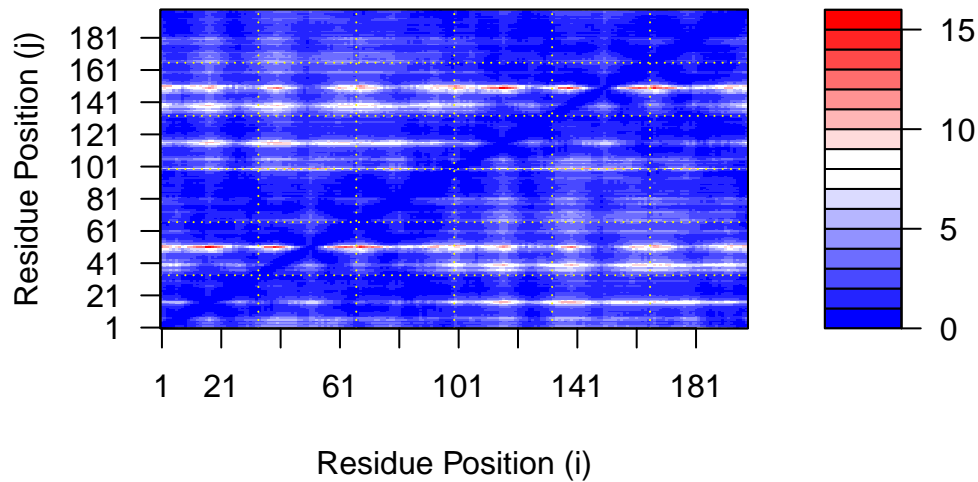
```
[1] 15.54688
```

```
pae5$max_pae
```

```
[1] 29.29688
```

We can plot these with ggplot or with functions from the Bio3D package:

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

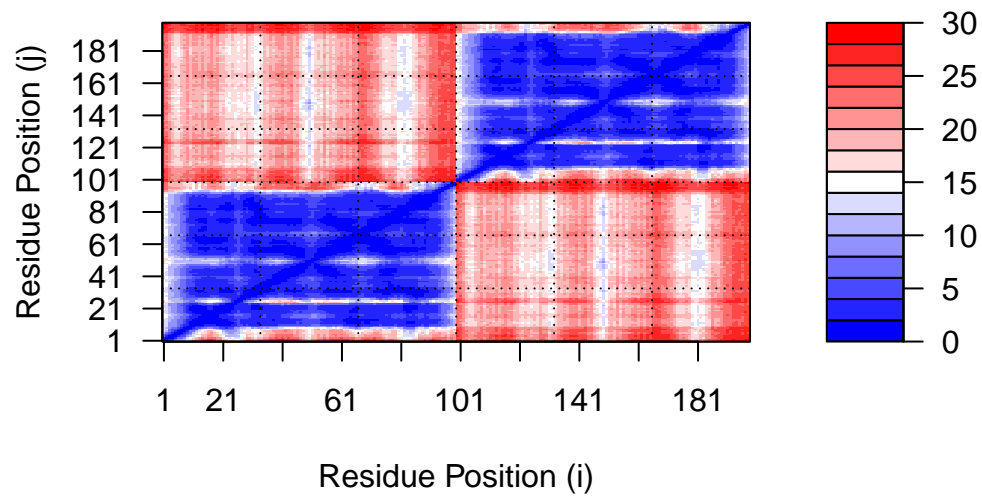


```
plot.dmat(pae5$pae,  
          xlab="Residue Position (i)",
```

```

ylab="Residue Position (j)",
grid.col = "black",
zlim=c(0,30))

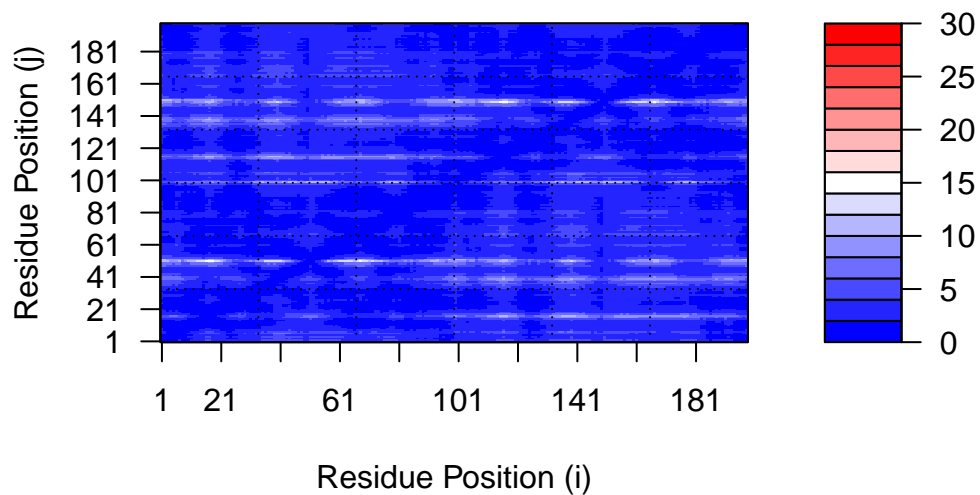
```



```

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

```

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

aln_file
```

```
[1] "hivprdimer_23119//hivpr_dimer_23119.a3m"
```

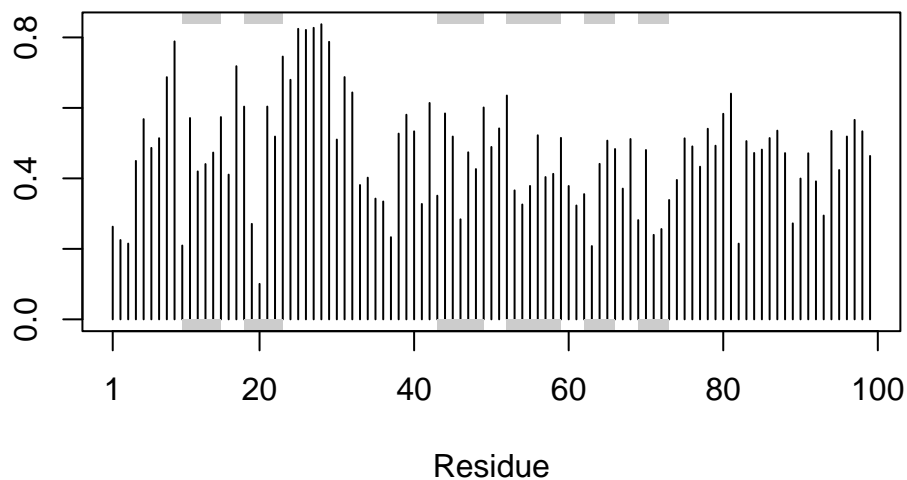
```
aln <- read.fasta(aln_file[1], to.upper = TRUE)
```

```
[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")
```

