

Class 11: Structural Bioinformatics Part 2

Olivia Baldwin

Today, before delving into structure prediction with AlphaFold we will finish off our previous Lab 10 comparative 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.

```
print(aa)
```

```
      1      .      .      .      .      .      .      60  
pdb|1AKE|A  MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMRLAAVKSGSELGKQAKDIMDAGKLV  
      1      .      .      .      .      .      .      60  
  
      61      .      .      .      .      .      .      120  
pdb|1AKE|A  DELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRI  
      61      .      .      .      .      .      .      120  
  
      121      .      .      .      .      .      .      180  
pdb|1AKE|A  VGRRVHAPSGRVYHVKNPPKVEGKDDVTGEELTRKDDQEETVRKRLVEYHQMTAPLIG  
      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
```

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

```
Searching ... please wait (updates every 5 seconds) RID = JXHC7HAP016
```

```
...
```

```
Reporting 85 hits
```

```
attributes(b)
```

```
$names
```

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

```
$class
```

```
[1] "blast"
```

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

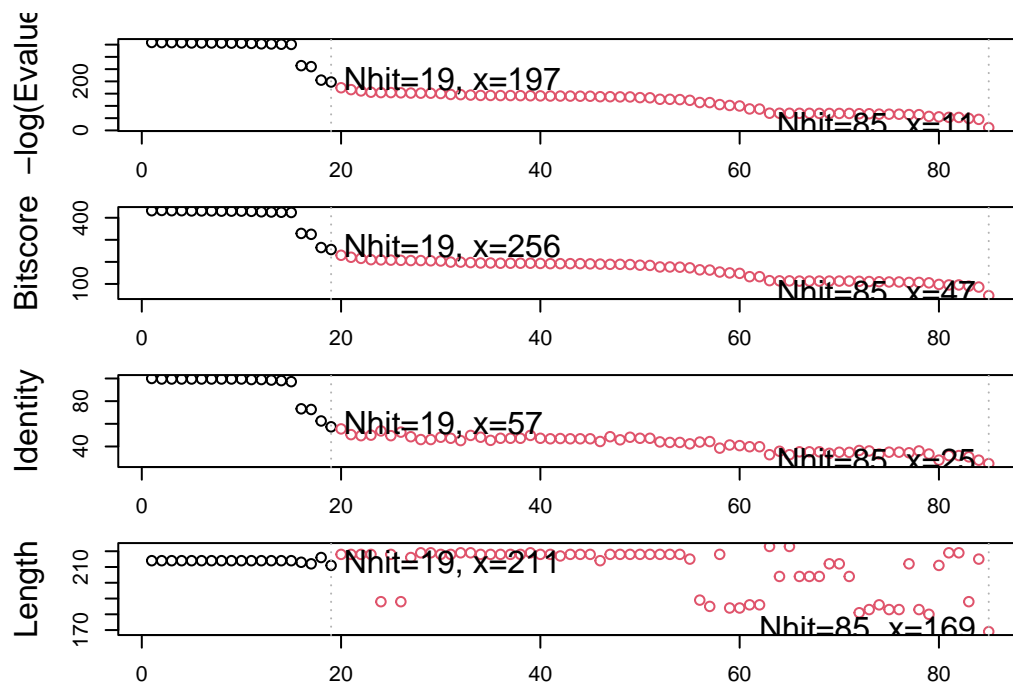
	queryid	subjectids	identity	alignmentlength	mismatches	gapopens	q.start
1	Query_7805245	1AKE_A	100.000	214	0	0	1
2	Query_7805245	8BQF_A	99.533	214	1	0	1
3	Query_7805245	4X8M_A	99.533	214	1	0	1
4	Query_7805245	6S36_A	99.533	214	1	0	1
5	Query_7805245	8Q2B_A	99.533	214	1	0	1
6	Query_7805245	8RJ9_A	99.533	214	1	0	1

	q.end	s.start	s.end	evaluate	bitscore	positives	mlog.evaluate	pdb.id	acc
1	214	1	214	1.58e-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

```
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"
```

```
#downloads the pdb files from our top hits in the blast result
```

```
files <- get.pdb(hits$ pdb.id, path="pdbs", plot=TRUE, gzip=TRUE)
```

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

I am going to install BiocManager package from CRAN. Then I can use `BiocManager::install()` to install any bioconductor package.

```
#align pdbs so they are superposed (all turned/put in the same view angle)  
pdbs <- pdbaln(files, fit = TRUE, exefile="msa")
```

Reading PDB files:

pdb/1AKE.pdb
pdb/8BQF.pdb
pdb/4X8M.pdb
pdb/6S36.pdb
pdb/8Q2B.pdb
pdb/8RJ9.pdb
pdb/6RZE.pdb
pdb/4X8H.pdb
pdb/3HPR.pdb
pdb/1E4V.pdb
pdb/5EJE.pdb
pdb/1E4Y.pdb
pdb/3X2S.pdb
pdb/6HAP.pdb
pdb/6HAM.pdb
pdb/4K46.pdb
pdb/4NP6.pdb
pdb/3GMT.pdb
pdb/4PZL.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
. 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/1AKE.pdb
 PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 2 name: pdb/8BQF.pdb
 PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 3 name: pdb/4X8M.pdb
pdb/seq: 4 name: pdb/6S36.pdb
 PDB has ALT records, taking A only, rm.alt=TRUE

pdb/seq: 5 name: pdbc/8Q2B.pdb
 PDB has ALT records, taking A only, rm.alt=TRUE
 pdb/seq: 6 name: pdbc/8RJ9.pdb
 PDB has ALT records, taking A only, rm.alt=TRUE
 pdb/seq: 7 name: pdbc/6RZE.pdb
 PDB has ALT records, taking A only, rm.alt=TRUE
 pdb/seq: 8 name: pdbc/4X8H.pdb
 pdb/seq: 9 name: pdbc/3HPR.pdb
 PDB has ALT records, taking A only, rm.alt=TRUE
 pdb/seq: 10 name: pdbc/1E4V.pdb
 pdb/seq: 11 name: pdbc/5EJE.pdb
 PDB has ALT records, taking A only, rm.alt=TRUE
 pdb/seq: 12 name: pdbc/1E4Y.pdb
 pdb/seq: 13 name: pdbc/3X2S.pdb
 pdb/seq: 14 name: pdbc/6HAP.pdb
 pdb/seq: 15 name: pdbc/6HAM.pdb
 PDB has ALT records, taking A only, rm.alt=TRUE
 pdb/seq: 16 name: pdbc/4K46.pdb
 PDB has ALT records, taking A only, rm.alt=TRUE
 pdb/seq: 17 name: pdbc/4NP6.pdb
 PDB has ALT records, taking A only, rm.alt=TRUE
 pdb/seq: 18 name: pdbc/3GMT.pdb
 pdb/seq: 19 name: pdbc/4PZL.pdb
 PDB has ALT records, taking A only, rm.alt=TRUE

pdbc

	1	60
pdbc/1AKE.pdb	-----						
pdbc/8BQF.pdb	MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLV						
pdbc/4X8M.pdb	-----						
pdbc/6S36.pdb	-----						
pdbc/8Q2B.pdb	-----						
pdbc/8RJ9.pdb	-----						
pdbc/6RZE.pdb	-----						
pdbc/4X8H.pdb	-----						
pdbc/3HPR.pdb	-----						
pdbc/1E4V.pdb	-----						
pdbc/5EJE.pdb	-----						
pdbc/1E4Y.pdb	-----						
pdbc/3X2S.pdb	-----						
pdbc/6HAP.pdb	-----						

pdb/6HAM.pdb	-----								
pdb/4K46.pdb	-----								
pdb/4NP6.pdb	-----								
pdb/3GMT.pdb	-----								
pdb/4PZL.pdb	-----								
		1	60
		61	120
pdb/1AKE.pdb	-----								
pdb/8BQF.pdb	DELVIALVKERIAQEGFLLDGFPRTIPQADAMKEAGINVDYVIEFDVPDELIVDRIVGRR								
pdb/4X8M.pdb	-----								
pdb/6S36.pdb	-----								
pdb/8Q2B.pdb	-----								
pdb/8RJ9.pdb	-----								
pdb/6RZE.pdb	-----								
pdb/4X8H.pdb	-----								
pdb/3HPR.pdb	-----								
pdb/1E4V.pdb	-----								
pdb/5EJE.pdb	-----								
pdb/1E4Y.pdb	-----								
pdb/3X2S.pdb	-----								
pdb/6HAP.pdb	-----								
pdb/6HAM.pdb	-----								
pdb/4K46.pdb	-----								
pdb/4NP6.pdb	-----								
pdb/3GMT.pdb	-----								
pdb/4PZL.pdb	-----								
		61	120
		121	180
pdb/1AKE.pdb	-----								
pdb/8BQF.pdb	VHAPSGRVYHVKNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSK								
pdb/4X8M.pdb	-----								
pdb/6S36.pdb	-----								
pdb/8Q2B.pdb	-----								
pdb/8RJ9.pdb	-----								
pdb/6RZE.pdb	-----								
pdb/4X8H.pdb	-----								
pdb/3HPR.pdb	-----								
pdb/1E4V.pdb	-----								
pdb/5EJE.pdb	-----								

pdb/1E4Y.pdb	-----						
pdb/3X2S.pdb	-----						
pdb/6HAP.pdb	-----						
pdb/6HAM.pdb	-----						
pdb/4K46.pdb	-----						
pdb/4NP6.pdb	-----						
pdb/3GMT.pdb	-----						
pdb/4PZL.pdb	-----						
		121	180
		181	240
pdb/1AKE.pdb	-----						
pdb/8BQF.pdb	EAEAGNTKYAKVDGTPVAEVRADLEKILMRIILLGAPGAGKGTQAQFIMEKYGIPQIST						
pdb/4X8M.pdb	-----						
pdb/6S36.pdb	-----						
pdb/8Q2B.pdb	-----						
pdb/8RJ9.pdb	-----						
pdb/6RZE.pdb	-----						
pdb/4X8H.pdb	-----						
pdb/3HPR.pdb	-----						
pdb/1E4V.pdb	-----						
pdb/5EJE.pdb	-----						
pdb/1E4Y.pdb	-----						
pdb/3X2S.pdb	-----						
pdb/6HAP.pdb	-----						
pdb/6HAM.pdb	-----						
pdb/4K46.pdb	-----						
pdb/4NP6.pdb	-----NAMRIILLGAPGAGKGTQAQFIMEKFGIPQIST						
pdb/3GMT.pdb	-----						
pdb/4PZL.pdb	-----TENLYFQSNAMRIILLGAPGAGKGTQAKIIEQYNIHIST						
		181	240
		241	300
pdb/1AKE.pdb	-----						
pdb/8BQF.pdb	GDMLRAAVKSGSELGKQAKDIMDAGKLVTDLVIALVKERIAQEDCRNGFLLDGFPRITP						
pdb/4X8M.pdb	-----						
pdb/6S36.pdb	-----						
pdb/8Q2B.pdb	-----						
pdb/8RJ9.pdb	-----						
pdb/6RZE.pdb	-----						
pdb/4X8H.pdb	-----						

pdb/3HPR.pdb	-----						
pdb/1E4V.pdb	-----						
pdb/5EJE.pdb	-----						
pdb/1E4Y.pdb	-----						
pdb/3X2S.pdb	-----						
pdb/6HAP.pdb	-----						
pdb/6HAM.pdb	-----						
pdb/4K46.pdb	-----						
pdb/4NP6.pdb	GDMLRAAIKAGTELGKQAKAVIDAGQLVSDDIILGLIKERIAQADCEKGFLLDGFPRTIP						
pdb/3GMT.pdb	-----						
pdb/4PZL.pdb	GDMIRETIKSGSALGQELKKVLDAGELVSDEFI IKIVKDRISKNDCNNGFLLDGVPRTIP						
		241	300
		301	360
pdb/1AKE.pdb	-----						
pdb/8BQF.pdb	QADAMKEAGINVDYVIEFDVPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGKDDVTGE						
pdb/4X8M.pdb	-----						
pdb/6S36.pdb	-----						
pdb/8Q2B.pdb	-----						
pdb/8RJ9.pdb	-----						
pdb/6RZE.pdb	-----						
pdb/4X8H.pdb	-----						
pdb/3HPR.pdb	-----						
pdb/1E4V.pdb	-----						
pdb/5EJE.pdb	-----						
pdb/1E4Y.pdb	-----						
pdb/3X2S.pdb	-----						
pdb/6HAP.pdb	-----						
pdb/6HAM.pdb	-----						
pdb/4K46.pdb	-----						
pdb/4NP6.pdb	QADGLKEMGINVDYVIEFDVADDVIVERMAGRRHLPSGRTYHVVNPPKVEGKDDVTGE						
pdb/3GMT.pdb	-----						
pdb/4PZL.pdb	QAQELDKLGVNIDYIVEVDVADNLLIERITGRRIH PASGRTYHTKFNPPKVADKDDVTGE						
		301	360
		361	420
pdb/1AKE.pdb	-----						
pdb/8BQF.pdb	ELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGNT--KYAKVDGTPVAEVRADL						
pdb/4X8M.pdb	-----						
pdb/6S36.pdb	-----						
pdb/8Q2B.pdb	-----						

pdb/8RJ9.pdb	-----	
pdb/6RZE.pdb	-----	
pdb/4X8H.pdb	-----	
pdb/3HPR.pdb	-----	
pdb/1E4V.pdb	-----	
pdb/5EJE.pdb	-----	
pdb/1E4Y.pdb	-----	
pdb/3X2S.pdb	-----	
pdb/6HAP.pdb	-----	
pdb/6HAM.pdb	-----	
pdb/4K46.pdb	-----	
pdb/4NP6.pdb	DLVIREDDKEETVRARLNVYHTQTAPLIEYYGKEAAAGKT--QYLKFDGKQVSEVSADI	
pdb/3GMT.pdb	-----	
pdb/4PZL.pdb	PLITRTDDNEDTVKQRLSVYHAQTAKLIDFYRNFSSNTKIPKYIKINGDQAVEKVSQDI	
	361	420
	421	480
pdb/1AKE.pdb	-----	
pdb/8BQF.pdb	EKILGMRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMRLAAVKSGSELGKQAKDIMDA	
pdb/4X8M.pdb	-----	
pdb/6S36.pdb	-----	
pdb/8Q2B.pdb	-----	
pdb/8RJ9.pdb	-----	
pdb/6RZE.pdb	-----	
pdb/4X8H.pdb	-----	
pdb/3HPR.pdb	-----	
pdb/1E4V.pdb	-----	
pdb/5EJE.pdb	-----	
pdb/1E4Y.pdb	-----	
pdb/3X2S.pdb	-----	
pdb/6HAP.pdb	-----	
pdb/6HAM.pdb	-----	
pdb/4K46.pdb	-----	
pdb/4NP6.pdb	AKALAA-----	
pdb/3GMT.pdb	-----	
pdb/4PZL.pdb	FDQLNKTENLYFQSNA-----	
	421	480
	481	540
pdb/1AKE.pdb	-----	
pdb/8BQF.pdb	GKLVTDLVIALVKERIAQGFLLDGFPRTIPQADAMKEAGINVDYVIEFDVPDELIVDRI	

pdbs/4X8M.pdb -----
 pdbs/6S36.pdb -----
 pdbs/8Q2B.pdb -----
 pdbs/8RJ9.pdb -----
 pdbs/6RZE.pdb -----
 pdbs/4X8H.pdb -----
 pdbs/3HPR.pdb -----
 pdbs/1E4V.pdb -----
 pdbs/5EJE.pdb -----
 pdbs/1E4Y.pdb -----
 pdbs/3X2S.pdb -----
 pdbs/6HAP.pdb -----
 pdbs/6HAM.pdb -----
 pdbs/4K46.pdb -----
 pdbs/4NP6.pdb -----
 pdbs/3GMT.pdb -----
 pdbs/4PZL.pdb -----

481 540

541 600

pdbs/1AKE.pdb -----
 pdbs/8BQF.pdb VGGRRVHAPSGRVYHVKNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIG
 pdbs/4X8M.pdb -----
 pdbs/6S36.pdb -----
 pdbs/8Q2B.pdb -----
 pdbs/8RJ9.pdb -----
 pdbs/6RZE.pdb -----
 pdbs/4X8H.pdb -----
 pdbs/3HPR.pdb -----
 pdbs/1E4V.pdb -----
 pdbs/5EJE.pdb -----
 pdbs/1E4Y.pdb -----
 pdbs/3X2S.pdb -----
 pdbs/6HAP.pdb -----
 pdbs/6HAM.pdb -----
 pdbs/4K46.pdb -----
 pdbs/4NP6.pdb -----
 pdbs/3GMT.pdb -----
 pdbs/4PZL.pdb -----

541 600

	601	660
pdb/1AKE.pdb		-----MRIILLGAPGAGKGTQAQFIMEKYG					
pdb/8BQF.pdb		YYSKEAEAGNTKYAKVDGTPVAEVRADLEKILGHMRIILLGAPGAGKGTQAQFIMEKYG					
pdb/4X8M.pdb		-----					
pdb/6S36.pdb		-----					
pdb/8Q2B.pdb		-----MRIILLGAPGAGKGTQAQFIMEKYG					
pdb/8RJ9.pdb		-----MRIILLGAPGAGKGTQAQFIMEKYG					
pdb/6RZE.pdb		-----					
pdb/4X8H.pdb		-----					
pdb/3HPR.pdb		-----MRIILLGAPGAGKGTQAQFIMEKYG					
pdb/1E4V.pdb		-----MRIILLGAPVAGKGTQAQFIMEKYG					
pdb/5EJE.pdb		-----MRIILLGAPGAGKGTQAQFIMEKYG					
pdb/1E4Y.pdb		-----MRIILLGALVAGKGTQAQFIMEKYG					
pdb/3X2S.pdb		-----MRIILLGAPGAGKGTQAQFIMEKYG					
pdb/6HAP.pdb		-----					
pdb/6HAM.pdb		-----					
pdb/4K46.pdb		-----					
pdb/4NP6.pdb		-----MRIILLGAPGAGKGTQAQFIMEKFG					
pdb/3GMT.pdb		-----MRLILLGAPGAGKGTQANFIKEKFG					
pdb/4PZL.pdb		-----MRIILLGAPGAGKGTQAKIIEQKYN					

	601	660
--	-----	---	---	---	---	---	-----

	661	720
pdb/1AKE.pdb		IPQISTGDMRLAAVKSGSELGKQAKDIMDAGKLVDELVIALVKERIAQEDCRNGFLLDG					
pdb/8BQF.pdb		IPQISTGDMRLAAVKSGSELGKQAKDIMDAGKLVDELVIALVKERIAQEDCRNGFLLDG					
pdb/4X8M.pdb		-----					
pdb/6S36.pdb		-----					
pdb/8Q2B.pdb		IPQISTGDMRLAAVKSGSELGKQAKDIMDAGKLVDELVIALVKERIAQEDCRNGFLLDG					
pdb/8RJ9.pdb		IPQISTGDMRLAAVKSGSELGKQAKDIMDAGKLVDELVIALVKERIAQEDCRNGFLLAG					
pdb/6RZE.pdb		-----					
pdb/4X8H.pdb		-----					
pdb/3HPR.pdb		IPQISTGDMRLAAVKSGSELGKQAKDIMDAGKLVDELVIALVKERIAQEDCRNGFLLDG					
pdb/1E4V.pdb		IPQISTGDMRLAAVKSGSELGKQAKDIMDAGKLVDELVIALVKERIAQEDCRNGFLLDG					
pdb/5EJE.pdb		IPQISTGDMRLAAVKSGSELGKQAKDIMDACKLVDELVIALVKERIAQEDCRNGFLLDG					
pdb/1E4Y.pdb		IPQISTGDMRLAAVKSGSELGKQAKDIMDAGKLVDELVIALVKERIAQEDCRNGFLLDG					
pdb/3X2S.pdb		IPQISTGDMRLAAVKSGSELGKQAKDIMDCGKLVDELVIALVKERIAQEDSRNGFLLDG					
pdb/6HAP.pdb		-----					
pdb/6HAM.pdb		-----					
pdb/4K46.pdb		-----					
pdb/4NP6.pdb		IPQISTGDMRLAAIKAGTELGKQAKAVIDAGQLVSDDIILGLIKERIAQADCEKGFLLDG					
pdb/3GMT.pdb		IPQISTGDMRLAAVKAGTPLGVEAKTYMDEGKLVPSLIIGLVKERLKEADCANGYLFDG					
pdb/4PZL.pdb		IAHISTGDMIRETIKSGSALGQELKKVLDAGELVSDEFIIVKDRISKNDCCNGFLLDG					

661	720
721	780
pdb/1AKE.pdb	FPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGK					
pdb/8BQF.pdb	FPRTIPQADAMKEAGINVDYVIEFDVPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGK					
pdb/4X8M.pdb	-----					
pdb/6S36.pdb	-----					
pdb/8Q2B.pdb	FPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGK					
pdb/8RJ9.pdb	FPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGK					
pdb/6RZE.pdb	-----					
pdb/4X8H.pdb	-----					
pdb/3HPR.pdb	FPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGK					
pdb/1E4V.pdb	FPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGK					
pdb/5EJE.pdb	FPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGK					
pdb/1E4Y.pdb	FPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGK					
pdb/3X2S.pdb	FPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHVKNPPKVEGK					
pdb/6HAP.pdb	-----					
pdb/6HAM.pdb	-----					
pdb/4K46.pdb	-----					
pdb/4NP6.pdb	FPRTIPQADGLKEMGINVDYVIEFDVADDVIVERMAGRRRAHLPSGRTYHVYNPPKVEGK					
pdb/3GMT.pdb	FPRTIAQADAMKEAGVAIDYVLEIDVPFSEI IERMSGRRTHPASGRTYHVKNPPKVEGK					
pdb/4PZL.pdb	VPRTPQAQELDKLGVNIDYIVEVDVADNLLIERITGRRIHPASGRTYHTKFNPPKVADK					
721	780
781	840
pdb/1AKE.pdb	DDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIGYY----SKEAEAGNT--KYAKVDGT					
pdb/8BQF.pdb	DDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIGYY----SKEAEAGNT--KYAKVDGT					
pdb/4X8M.pdb	-----					
pdb/6S36.pdb	-----					
pdb/8Q2B.pdb	DDVTGEELTTRKADQEETVRKRLVEYHQMTAPLIGYY----SKEAEAGNT--KYAKVDGT					
pdb/8RJ9.pdb	DDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIGYY----SKEAEAGNT--KYAKVDGT					
pdb/6RZE.pdb	-----					
pdb/4X8H.pdb	-----					
pdb/3HPR.pdb	DDGTGEELTTRKDDQEETVRKRLVEYHQMTAPLIGYY----SKEAEAGNT--KYAKVDGT					
pdb/1E4V.pdb	DDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIGYY----SKEAEAGNT--KYAKVDGT					
pdb/5EJE.pdb	DDVTGEELTTRKDDQEECVRKRLVEYHQMTAPLIGYY----SKEAEAGNT--KYAKVDGT					
pdb/1E4Y.pdb	DDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIGYY----SKEAEAGNT--KYAKVDGT					
pdb/3X2S.pdb	DDVTGEELTTRKDDQEETVRKRLCEYHQMTAPLIGYY----SKEAEAGNT--KYAKVDGT					
pdb/6HAP.pdb	-----					
pdb/6HAM.pdb	-----					
pdb/4K46.pdb	-----					

pdb/4NP6.pdb	DDVTGEDLVIREDDKEETVRARLNVYHTQTAPLIEYY----
pdb/3GMT.pdb	DDVTGEPLVQRDDDKKEETVKRRLDVYEAQTKPLITYYGDWARRGAENGLKAPAYRKISG-
pdb/4PZL.pdb	DDVTGEPLITRTDDNEDTVKQRLSVYHAQTAKLIDFY----RNFSSTNTKIPKYIKINGD
	781 840
	841 900
pdb/1AKE.pdb	KPVAEVRADLEKILG-----
pdb/8BQF.pdb	KPVAEVRADLEKILGMRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMMLRAAVKSGSEL
pdb/4X8M.pdb	-----
pdb/6S36.pdb	-----
pdb/8Q2B.pdb	KPVAEVRADLEKILG-----
pdb/8RJ9.pdb	KPVAEVRADLEKILG-----
pdb/6RZE.pdb	-----
pdb/4X8H.pdb	-----
pdb/3HPR.pdb	KPVAEVRADLEKILG-----
pdb/1E4V.pdb	KPVAEVRADLEKILG-----
pdb/5EJE.pdb	KPVAEVRADLEKILG-----
pdb/1E4Y.pdb	KPVAEVRADLEKILG-----
pdb/3X2S.pdb	KPVAEVRADLEKILG-----
pdb/6HAP.pdb	-----
pdb/6HAM.pdb	-----
pdb/4K46.pdb	-----
pdb/4NP6.pdb	KQVSEVSADIAKALANA-----
pdb/3GMT.pdb	-----
pdb/4PZL.pdb	QAVEKVSQDIFDQLNKRNA-----
	841 900
	901 960
pdb/1AKE.pdb	-----
pdb/8BQF.pdb	GKQAKDIMDAGKLVTDDELVIALVKERIAQNGFLLDGFPRTPQADAMKEAGINVDYVIEF
pdb/4X8M.pdb	-----
pdb/6S36.pdb	-----
pdb/8Q2B.pdb	-----
pdb/8RJ9.pdb	-----
pdb/6RZE.pdb	-----
pdb/4X8H.pdb	-----
pdb/3HPR.pdb	-----
pdb/1E4V.pdb	-----
pdb/5EJE.pdb	-----
pdb/1E4Y.pdb	-----
pdb/3X2S.pdb	-----

pdb/6HAP.pdb	-----								
pdb/6HAM.pdb	-----								
pdb/4K46.pdb	-----								
pdb/4NP6.pdb	-----								
pdb/3GMT.pdb	-----								
pdb/4PZL.pdb	-----								
		901	960
		961	1020
pdb/1AKE.pdb	-----								
pdb/8BQF.pdb	DVPDELIVDRIVGRRVHAPSGRVYHVKFNPKEGKDDVTGEELTTRKDDQEETVRKRLV								
pdb/4X8M.pdb	-----								
pdb/6S36.pdb	-----								
pdb/8Q2B.pdb	-----								
pdb/8RJ9.pdb	-----								
pdb/6RZE.pdb	-----								
pdb/4X8H.pdb	-----								
pdb/3HPR.pdb	-----								
pdb/1E4V.pdb	-----								
pdb/5EJE.pdb	-----								
pdb/1E4Y.pdb	-----								
pdb/3X2S.pdb	-----								
pdb/6HAP.pdb	-----								
pdb/6HAM.pdb	-----								
pdb/4K46.pdb	-----								
pdb/4NP6.pdb	-----								
pdb/3GMT.pdb	-----								
pdb/4PZL.pdb	-----								
		961	1020
		1021	1080
pdb/1AKE.pdb	-----MRIILLGAPGAGKGT								
pdb/8BQF.pdb	EYHQMTAPLIGYYSKEAEAGNTKYAKVDGTPVAEVRADLEKILHMRIILLGAPGAGKGT								
pdb/4X8M.pdb	-----MRIILLGAPGAGKGT								
pdb/6S36.pdb	-----MRIILLGAPGAGKGT								
pdb/8Q2B.pdb	-----MRIILLGAPGAGKGT								
pdb/8RJ9.pdb	-----MRIILLGAPGAGKGT								
pdb/6RZE.pdb	-----MRIILLGAPGAGKGT								
pdb/4X8H.pdb	-----MRIILLGAPGAGKGT								
pdb/3HPR.pdb	-----MRIILLGAPGAGKGT								
pdb/1E4V.pdb	-----MRIILLGAPVAGKGT								


```

pdbs/5EJE.pdb -----MRIILLGAPGAGKGT
pdbs/1E4Y.pdb -----MRIILLGALVAGKGT
pdbs/3X2S.pdb -----MRIILLGAPGAGKGT
pdbs/6HAP.pdb -----MRIILLGAPGAGKGT
pdbs/6HAM.pdb -----MRIILLGAPGAGKGT
pdbs/4K46.pdb -----MRIILLGAPGAGKGT
pdbs/4NP6.pdb -----MRIILLGAPGAGKGT
pdbs/3GMT.pdb -----MRLILLGAPGAGKGT
pdbs/4PZL.pdb -----MRIILLGAPGAGKGT
                                     **~*****  *****
1021 . . . . . 1080

1081 . . . . . 1140
pdbs/1AKE.pdb QAQFIMEKYGIPQISTGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKERIAQE
pdbs/8BQF.pdb QAQFIMEKYGIPQISTGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKERIAQE
pdbs/4X8M.pdb QAQFIMEKYGIPQISTGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKERIAQE
pdbs/6S36.pdb QAQFIMEKYGIPQISTGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKERIAQE
pdbs/8Q2B.pdb QAQFIMEKYGIPQISTGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKERIAQE
pdbs/8RJ9.pdb QAQFIMEKYGIPQISTGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKERIAQE
pdbs/6RZE.pdb QAQFIMEKYGIPQISTGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKERIAQE
pdbs/4X8H.pdb QAQFIMEKYGIPQISTGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKERIAQE
pdbs/3HPR.pdb QAQFIMEKYGIPQISTGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKERIAQE
pdbs/1E4V.pdb QAQFIMEKYGIPQISTGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKERIAQE
pdbs/5EJE.pdb QAQFIMEKYGIPQISTGDMRLRAAVKSGSELGKQAKDIMDACKLVTDDELVIALVKERIAQE
pdbs/1E4Y.pdb QAQFIMEKYGIPQISTGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVKERIAQE
pdbs/3X2S.pdb QAQFIMEKYGIPQISTGDMRLRAAVKSGSELGKQAKDIMDCGKLVTDDELVIALVKERIAQE
pdbs/6HAP.pdb QAQFIMEKYGIPQISTGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDDELVIALVRERICQE
pdbs/6HAM.pdb QAQFIMEKYGIPQISTGDMRLRAAIKSGSELGKQAKDIMDAGKLVTDDEIIIALVKERICQE
pdbs/4K46.pdb QAQFIMAKFGIPQISTGDMRLRAAIKAGTELGKQAKSVIDAGQLVSDDIILGLVKERIAQD
pdbs/4NP6.pdb QAQFIMEKFGIPQISTGDMRLRAAIKAGTELGKQAKAVIDAGQLVSDDIILGLIKERIAQA
pdbs/3GMT.pdb QANFIKEKFGIPQISTGDMRLRAAVKAGTPLGVEAKTYMDEGKLVPSLIIGLVKERLKEA
pdbs/4PZL.pdb QAKIIEQKYNIAHISTGDMIRETIKSGSALGQELKKVLDAGELVSDEFIIVKDRISK
** * *~* *****~* ~* *~** * ~* ** * ~^~~~~~^
1081 . . . . . 1140

1141 . . . . . 1200
pdbs/1AKE.pdb DCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHV
pdbs/8BQF.pdb DCRNGFLLDGFPR TIPQADAMKEAGINVDYVIEFDVPDELIVDRIVGRRVHAPSGRVYHV
pdbs/4X8M.pdb DCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHV
pdbs/6S36.pdb DCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFDVPDELIVDKIVGRRVHAPSGRVYHV
pdbs/8Q2B.pdb DCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHV
pdbs/8RJ9.pdb DCRNGFLLAGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRIVGRRVHAPSGRVYHV
pdbs/6RZE.pdb DCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFDVPDELIVDAIVGRRVHAPSGRVYHV

```


pdb/8Q2B.pdb	AKVDGTPVAEVRADL-EKILG-----
pdb/8RJ9.pdb	AKVDGTPVAEVRADL-EKILG-----
pdb/6RZE.pdb	AKVDGTPVAEVRADL-EKILG-----
pdb/4X8H.pdb	AKVDGTPVAEVRADL-EKILG-----
pdb/3HPR.pdb	AKVDGTPVAEVRADL-EKILG-----
pdb/1E4V.pdb	AKVDGTPVAEVRADL-EKILG-----
pdb/5EJE.pdb	AKVDGTPVAEVRADL-EKILG-----
pdb/1E4Y.pdb	AKVDGTPVAEVRADL-EKILG-----
pdb/3X2S.pdb	AKVDGTPVAEVRADL-EKILG-----
pdb/6HAP.pdb	AKVDGTPVCEVRADL-EKILG-----
pdb/6HAM.pdb	AKVDGTPVCEVRADL-EKILG-----
pdb/4K46.pdb	LKFDGTPKVAEVSDEL-EKALA-----
pdb/4NP6.pdb	LKFDGTPQVSEVSADI-AKALAAMRIILLGAPGAGKGTQAQFIMEKFGIPQISTGDMRLA
pdb/3GMT.pdb	RKIS-----
pdb/4PZL.pdb	IKINGDQAVEKVSQDIFDQLNKAMRIILLGAPGAGKGTQAKIIEQKYNIAHISTGDMIRE
	*
	1261 1320
	1321 1380
pdb/1AKE.pdb	-----
pdb/8BQF.pdb	-----
pdb/4X8M.pdb	-----
pdb/6S36.pdb	-----
pdb/8Q2B.pdb	-----
pdb/8RJ9.pdb	-----
pdb/6RZE.pdb	-----
pdb/4X8H.pdb	-----
pdb/3HPR.pdb	-----
pdb/1E4V.pdb	-----
pdb/5EJE.pdb	-----
pdb/1E4Y.pdb	-----
pdb/3X2S.pdb	-----
pdb/6HAP.pdb	-----
pdb/6HAM.pdb	-----
pdb/4K46.pdb	-----
pdb/4NP6.pdb	AIKAGTELKQAKAVIDAGQLVSDDIILGLIKERIAQADCEKGFLLDGVPRTIPQADGLK
pdb/3GMT.pdb	-----
pdb/4PZL.pdb	TIKSGSALGQELKKVLDAGELVSDEFIIVKDRISKNDNNGFLLDGVPRTIPQAQELD
	1321 1380
	1381 1440
pdb/1AKE.pdb	-----

pdb/8BQF.pdb	-----	
pdb/4X8M.pdb	-----	
pdb/6S36.pdb	-----	
pdb/8Q2B.pdb	-----	
pdb/8RJ9.pdb	-----	
pdb/6RZE.pdb	-----	
pdb/4X8H.pdb	-----	
pdb/3HPR.pdb	-----	
pdb/1E4V.pdb	-----	
pdb/5EJE.pdb	-----	
pdb/1E4Y.pdb	-----	
pdb/3X2S.pdb	-----	
pdb/6HAP.pdb	-----	
pdb/6HAM.pdb	-----	
pdb/4K46.pdb	-----	
pdb/4NP6.pdb	EMGINVDYVIEFDVADDVIVERMAGRAHLPSGRTYHVYNNPPKVEGKD----EDLVIRE	
pdb/3GMT.pdb	-----	
pdb/4PZL.pdb	KLGVNIDYIVEVDVADNLLIERITGRRHPASGRTYHTKFNPPKVADKDDVTGEPLITRT	
	1381	1440
	1441	1500
pdb/1AKE.pdb	-----	
pdb/8BQF.pdb	-----	
pdb/4X8M.pdb	-----	
pdb/6S36.pdb	-----	
pdb/8Q2B.pdb	-----	
pdb/8RJ9.pdb	-----	
pdb/6RZE.pdb	-----	
pdb/4X8H.pdb	-----	
pdb/3HPR.pdb	-----	
pdb/1E4V.pdb	-----	
pdb/5EJE.pdb	-----	
pdb/1E4Y.pdb	-----	
pdb/3X2S.pdb	-----	
pdb/6HAP.pdb	-----	
pdb/6HAM.pdb	-----	
pdb/4K46.pdb	-----	
pdb/4NP6.pdb	DDKEETVRARLNVYHTQTAPLIEYYGKEAAAGKT--QYLKFDGTKQVSEVSADIAKALA-	
pdb/3GMT.pdb	-----	
pdb/4PZL.pdb	DDNEDTVKQRLSVYHAQTAKLIDFYRNFSSNTNPKYIKINGDQAVEKVSQDIFDQLNK	
	1441	1500

Call:

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

Class:

```
pdb, fasta
```

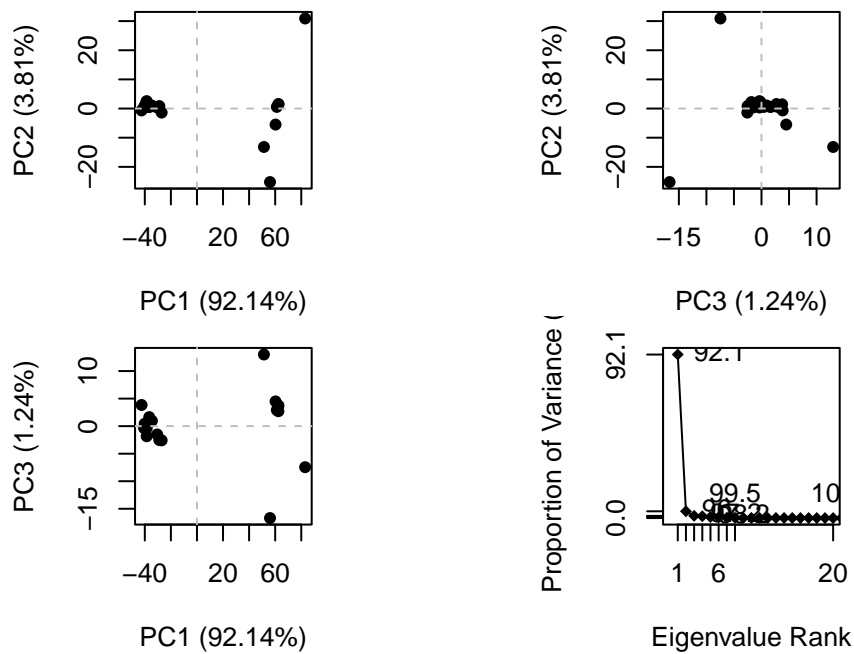
Alignment dimensions:

```
19 sequence rows; 1500 position columns (197 non-gap, 1303 gap)
```

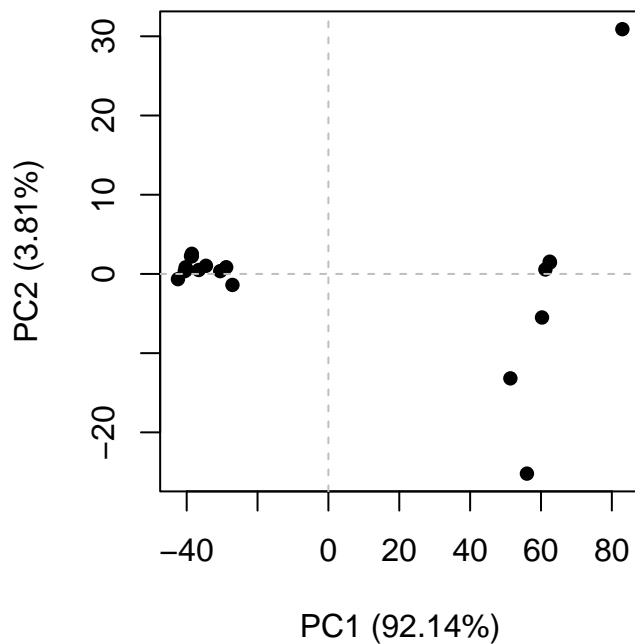
```
+ attr: xyz, resno, b, chain, id, ali, resid, sse, call
```

##Principal Component Analysis

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



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



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

```
#"make trajectory"

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

Analysis of Alpha Fold Results

```
results_dir <- "HIVdimertest_23119"

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

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

```
[1] "HIVdimertest_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_1_seed_000.pdb"
[2] "HIVdimertest_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000.pdb"
```

```
[3] "HIVdimertest_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000.pdb"
[4] "HIVdimertest_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_2_seed_000.pdb"
[5] "HIVdimertest_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_3_seed_000.pdb"
```

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

Reading PDB files:

```
HIVdimertest_23119/HIVdimertest_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_1_seed_000.pdb
HIVdimertest_23119/HIVdimertest_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000.pdb
HIVdimertest_23119/HIVdimertest_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000.pdb
HIVdimertest_23119/HIVdimertest_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_2_seed_000.pdb
HIVdimertest_23119/HIVdimertest_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_3_seed_000.pdb
.....
```

Extracting sequences

```
pdb/seq: 1 name: HIVdimertest_23119/HIVdimertest_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_1_seed_000.pdb
pdb/seq: 2 name: HIVdimertest_23119/HIVdimertest_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000.pdb
pdb/seq: 3 name: HIVdimertest_23119/HIVdimertest_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000.pdb
pdb/seq: 4 name: HIVdimertest_23119/HIVdimertest_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_2_seed_000.pdb
pdb/seq: 5 name: HIVdimertest_23119/HIVdimertest_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_3_seed_000.pdb
```

HIV_pdb<-

```

1 . . . . . 50
[Truncated_Name:1]HIVdimerte PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
[Truncated_Name:2]HIVdimerte PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
[Truncated_Name:3]HIVdimerte PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
[Truncated_Name:4]HIVdimerte PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
[Truncated_Name:5]HIVdimerte PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
*****
1 . . . . . 50

51 . . . . . 100
[Truncated_Name:1]HIVdimerte GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
[Truncated_Name:2]HIVdimerte GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
[Truncated_Name:3]HIVdimerte GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
[Truncated_Name:4]HIVdimerte GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
[Truncated_Name:5]HIVdimerte GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
*****
51 . . . . . 100
```

```

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

151          .          .          .          .          198
[Truncated_Name:1]HIVdimerte  GFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
[Truncated_Name:2]HIVdimerte  GFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
[Truncated_Name:3]HIVdimerte  GFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
[Truncated_Name:4]HIVdimerte  GFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
[Truncated_Name:5]HIVdimerte  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(HIV_pdb, fit=T)
```

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

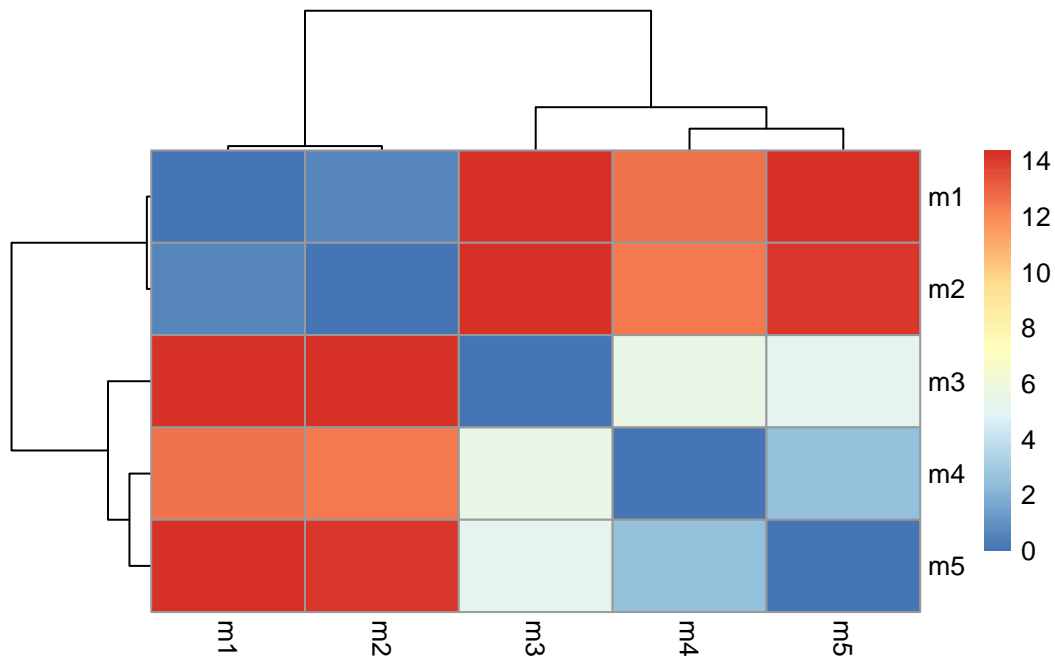
```
range(rd)
```

```
[1] 0.000 14.376
```

```
library(pheatmap)
```



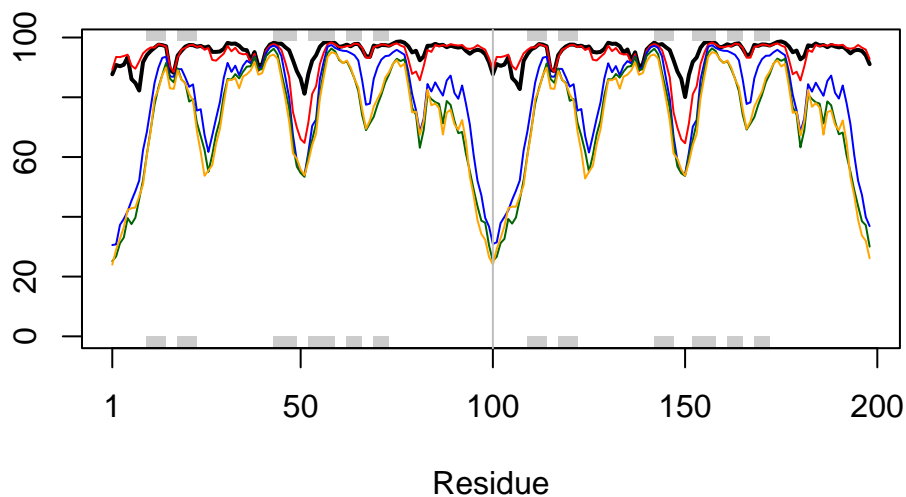
```
colnames(rd) <- paste0("m",1:5)
rownames(rd) <- paste0("m",1:5)
pheatmap(rd)
```



```
ref_pdb <- read.pdb("1hsg")
```

Note: Accessing on-line PDB file

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



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

```
core size 197 of 198 vol = 4916.702
core size 196 of 198 vol = 4311.481
core size 195 of 198 vol = 4101.445
core size 194 of 198 vol = 3907.124
core size 193 of 198 vol = 3711.925
core size 192 of 198 vol = 3546.511
core size 191 of 198 vol = 3440.437
core size 190 of 198 vol = 3317.571
core size 189 of 198 vol = 3220.079
core size 188 of 198 vol = 3142.057
core size 187 of 198 vol = 3066.79
core size 186 of 198 vol = 3015.892
core size 185 of 198 vol = 2959.969
core size 184 of 198 vol = 2913.74
core size 183 of 198 vol = 2880.923
core size 182 of 198 vol = 2848.081
core size 181 of 198 vol = 2857.001
core size 180 of 198 vol = 2871.24
core size 179 of 198 vol = 2905.696
core size 178 of 198 vol = 2953.776
```

core size 177 of 198	vol = 3020.847
core size 176 of 198	vol = 3087.22
core size 175 of 198	vol = 3109.99
core size 174 of 198	vol = 3129.601
core size 173 of 198	vol = 3135.085
core size 172 of 198	vol = 3092.283
core size 171 of 198	vol = 3036.012
core size 170 of 198	vol = 2947.995
core size 169 of 198	vol = 2886.897
core size 168 of 198	vol = 2829.355
core size 167 of 198	vol = 2746.377
core size 166 of 198	vol = 2671.189
core size 165 of 198	vol = 2600.848
core size 164 of 198	vol = 2534.651
core size 163 of 198	vol = 2464.3
core size 162 of 198	vol = 2390.171
core size 161 of 198	vol = 2322.47
core size 160 of 198	vol = 2236.698
core size 159 of 198	vol = 2160.475
core size 158 of 198	vol = 2077.281
core size 157 of 198	vol = 2003.596
core size 156 of 198	vol = 1939.94
core size 155 of 198	vol = 1859.188
core size 154 of 198	vol = 1781.083
core size 153 of 198	vol = 1699.1
core size 152 of 198	vol = 1622.558
core size 151 of 198	vol = 1546.319
core size 150 of 198	vol = 1473.01
core size 149 of 198	vol = 1414.087
core size 148 of 198	vol = 1352.547
core size 147 of 198	vol = 1295.278
core size 146 of 198	vol = 1246.999
core size 145 of 198	vol = 1203.962
core size 144 of 198	vol = 1163.009
core size 143 of 198	vol = 1110.955
core size 142 of 198	vol = 1064.672
core size 141 of 198	vol = 1028.458
core size 140 of 198	vol = 986.121
core size 139 of 198	vol = 944.003
core size 138 of 198	vol = 895.914
core size 137 of 198	vol = 853.508
core size 136 of 198	vol = 827.977
core size 135 of 198	vol = 796.874

core size 134 of 198	vol = 772.763
core size 133 of 198	vol = 743.108
core size 132 of 198	vol = 707.65
core size 131 of 198	vol = 669.172
core size 130 of 198	vol = 634.655
core size 129 of 198	vol = 594.035
core size 128 of 198	vol = 559.154
core size 127 of 198	vol = 525.971
core size 126 of 198	vol = 493.19
core size 125 of 198	vol = 466.473
core size 124 of 198	vol = 438.433
core size 123 of 198	vol = 410.725
core size 122 of 198	vol = 401.38
core size 121 of 198	vol = 391.76
core size 120 of 198	vol = 362.084
core size 119 of 198	vol = 338.183
core size 118 of 198	vol = 312.338
core size 117 of 198	vol = 282.176
core size 116 of 198	vol = 262.215
core size 115 of 198	vol = 241.577
core size 114 of 198	vol = 225.151
core size 113 of 198	vol = 204.137
core size 112 of 198	vol = 185.038
core size 111 of 198	vol = 162.728
core size 110 of 198	vol = 146.181
core size 109 of 198	vol = 133.352
core size 108 of 198	vol = 123.207
core size 107 of 198	vol = 109.228
core size 106 of 198	vol = 98.824
core size 105 of 198	vol = 89.735
core size 104 of 198	vol = 81.206
core size 103 of 198	vol = 74.188
core size 102 of 198	vol = 67.042
core size 101 of 198	vol = 62.043
core size 100 of 198	vol = 58.432
core size 99 of 198	vol = 55.149
core size 98 of 198	vol = 51.114
core size 97 of 198	vol = 45.798
core size 96 of 198	vol = 41.161
core size 95 of 198	vol = 35.619
core size 94 of 198	vol = 29.784
core size 93 of 198	vol = 23.233
core size 92 of 198	vol = 16.669

```

core size 91 of 198 vol = 9.459
core size 90 of 198 vol = 4.595
core size 89 of 198 vol = 3.161
core size 88 of 198 vol = 2.678
core size 87 of 198 vol = 2.293
core size 86 of 198 vol = 1.935
core size 85 of 198 vol = 1.619
core size 84 of 198 vol = 1.367
core size 83 of 198 vol = 1.09
core size 82 of 198 vol = 0.906
core size 81 of 198 vol = 0.764
core size 80 of 198 vol = 0.649
core size 79 of 198 vol = 0.596
core size 78 of 198 vol = 0.53
core size 77 of 198 vol = 0.486
FINISHED: Min vol ( 0.5 ) reached

```

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

```
# 78 positions (cumulative volume <= 0.5 Angstrom^3)
```

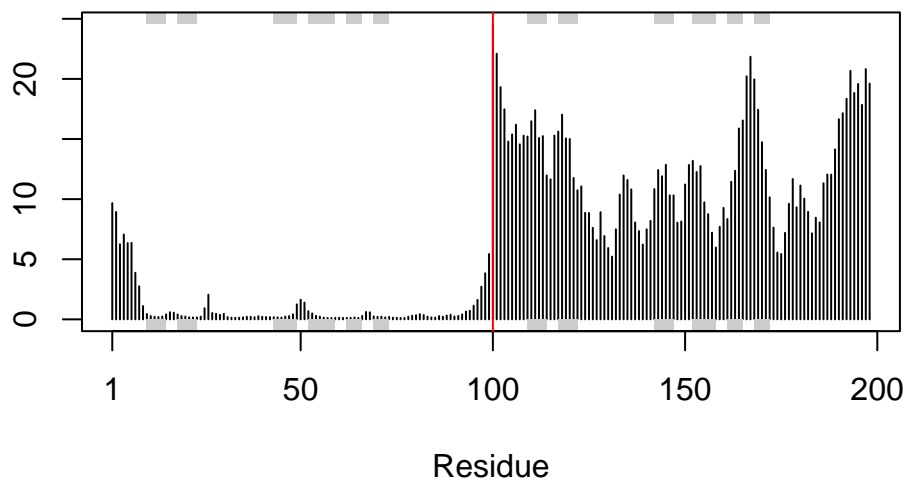
	start	end	length
1	10	25	16
2	28	48	21
3	53	93	41

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

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

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

```
abline(v=100, col="red", ylab="RMSF")
```



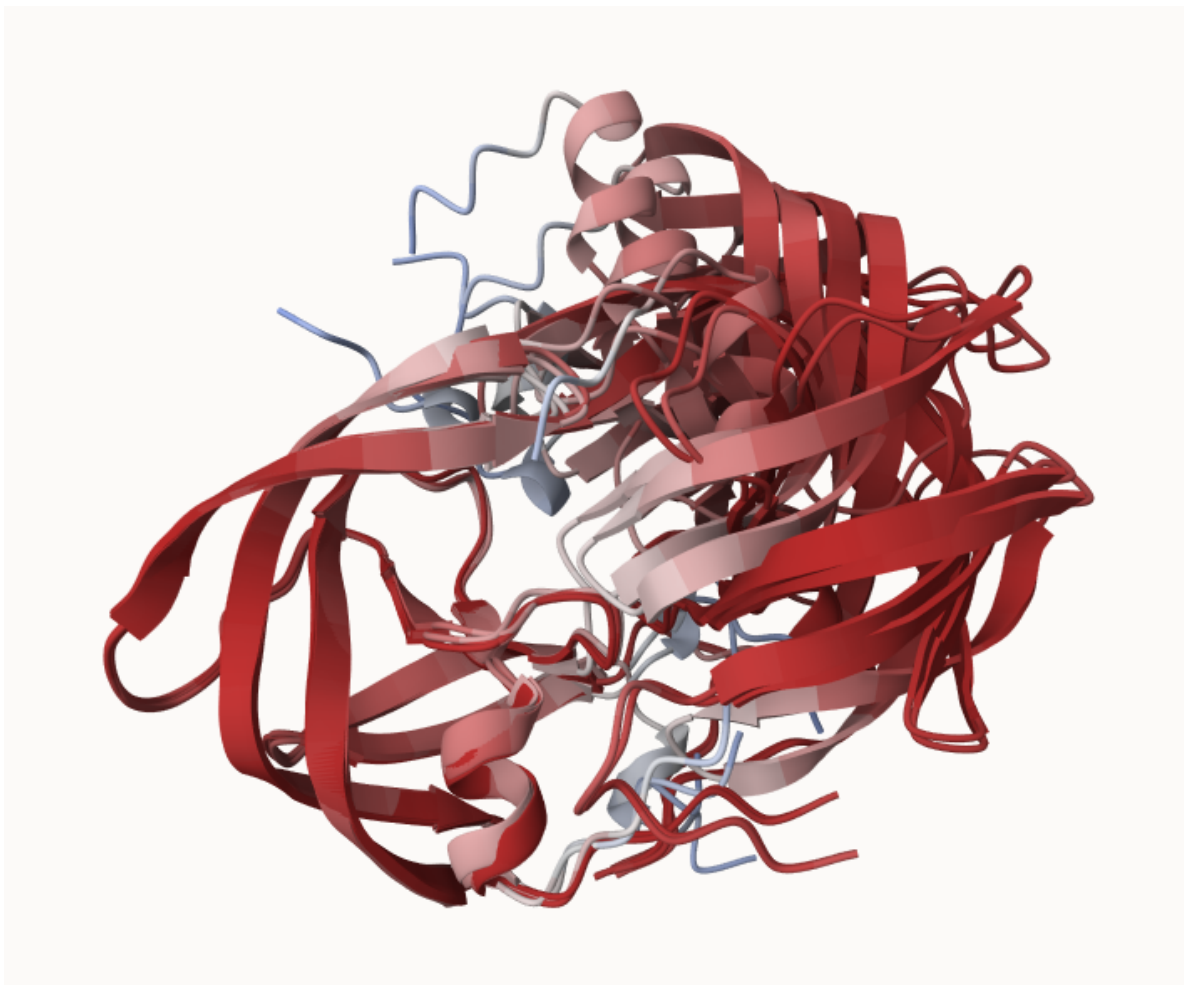


Figure 1: Colored by uncertainty

Predicting Alignment Error

```
library(jsonlite)
```

```
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"
```

```
head(pae1$plddt)
```

```
[1] 87.69 90.81 90.38 90.88 93.44 86.06
```

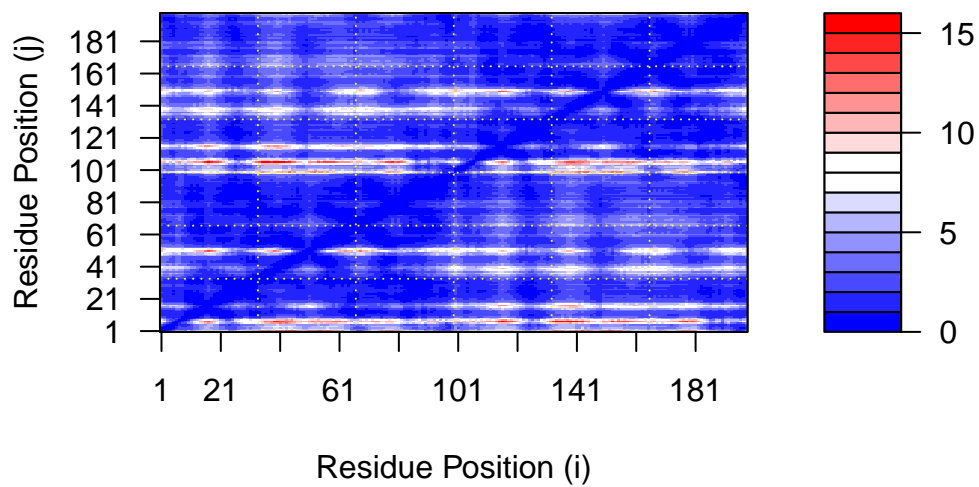
```
pae1$max_pae
```

```
[1] 15.47656
```

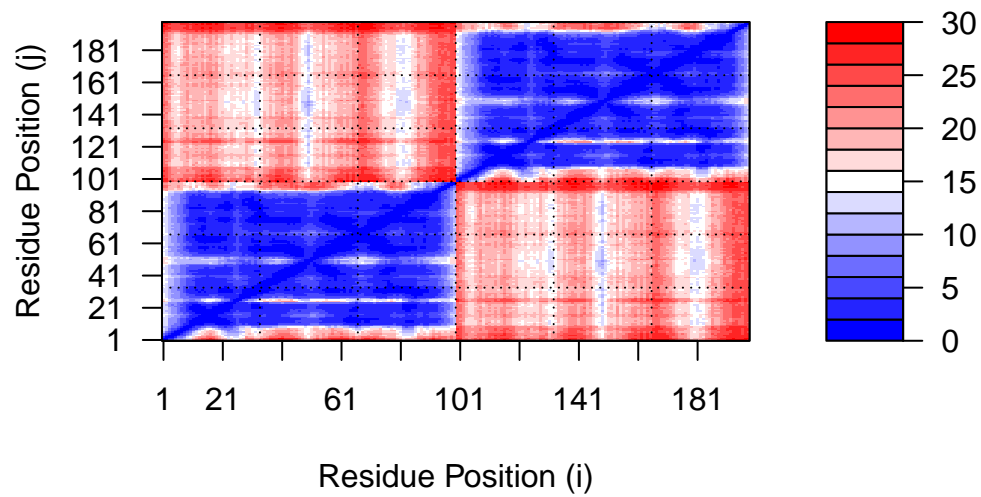
```
pae5$max_pae
```

```
[1] 29.32812
```

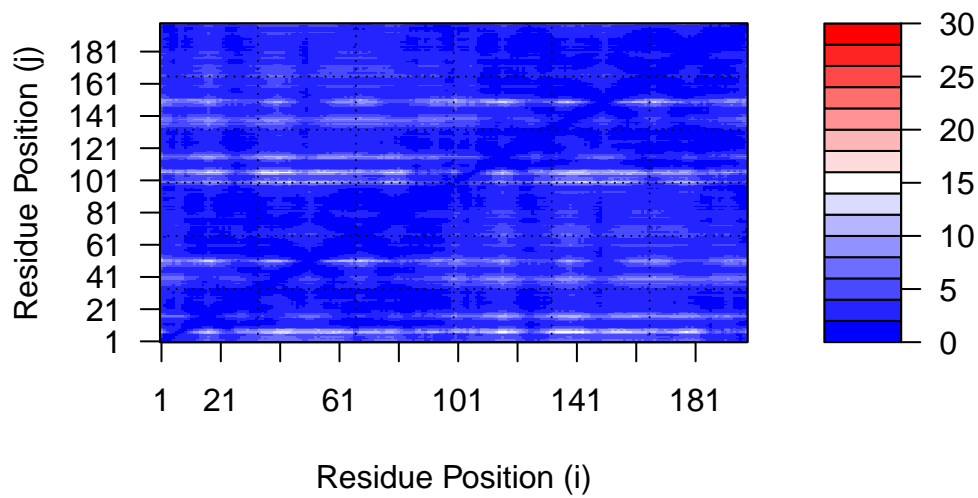
```
plot.dmat(pae1$pae,  
          xlab="Residue Position (i)",  
          ylab="Residue Position (j)")
```




```
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))
```



The pae1 seems to be better than the pae5.

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

```
[1] "HIVdimertest_23119/HIVdimertest_23119.a3m"
```

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

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

```
attributes(aln)
```

```
$names
[1] "id"   "ali"  "call"
```

```
$class
[1] "fasta"
```

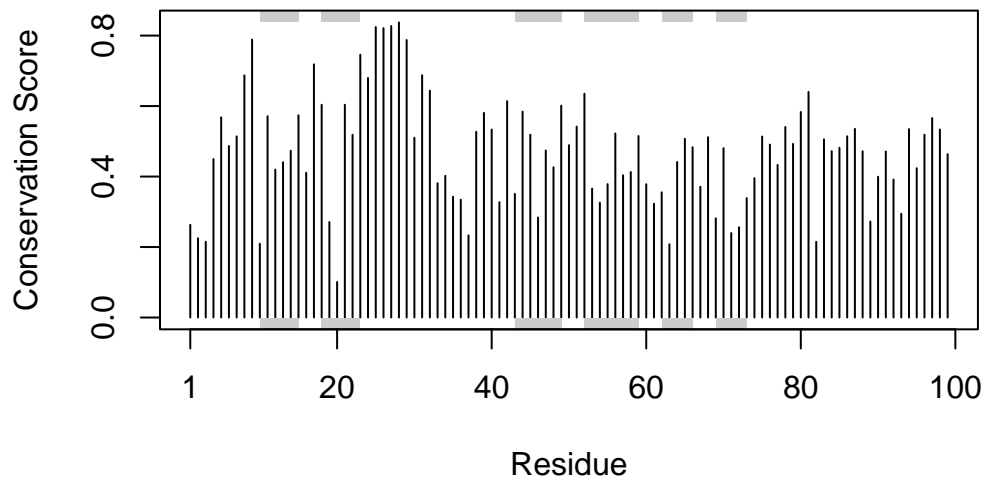
```
dim(aln$ali)
```

```
[1] 5378 132
```

Scores residue conservation:

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

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



Generate a consensus to highlight conserved residues.

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

```
[1] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"  
[19] "-" "-" "-" "-" "-" "-" "D" "T" "G" "A" "-" "-" "-" "-" "-" "-" "-"  
[37] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"  
[55] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"  
[73] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
```

```

[91] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[109] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[127] "-" "-" "-" "-" "-" "-"

```

Creates a file for final visualization:

```

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")

```

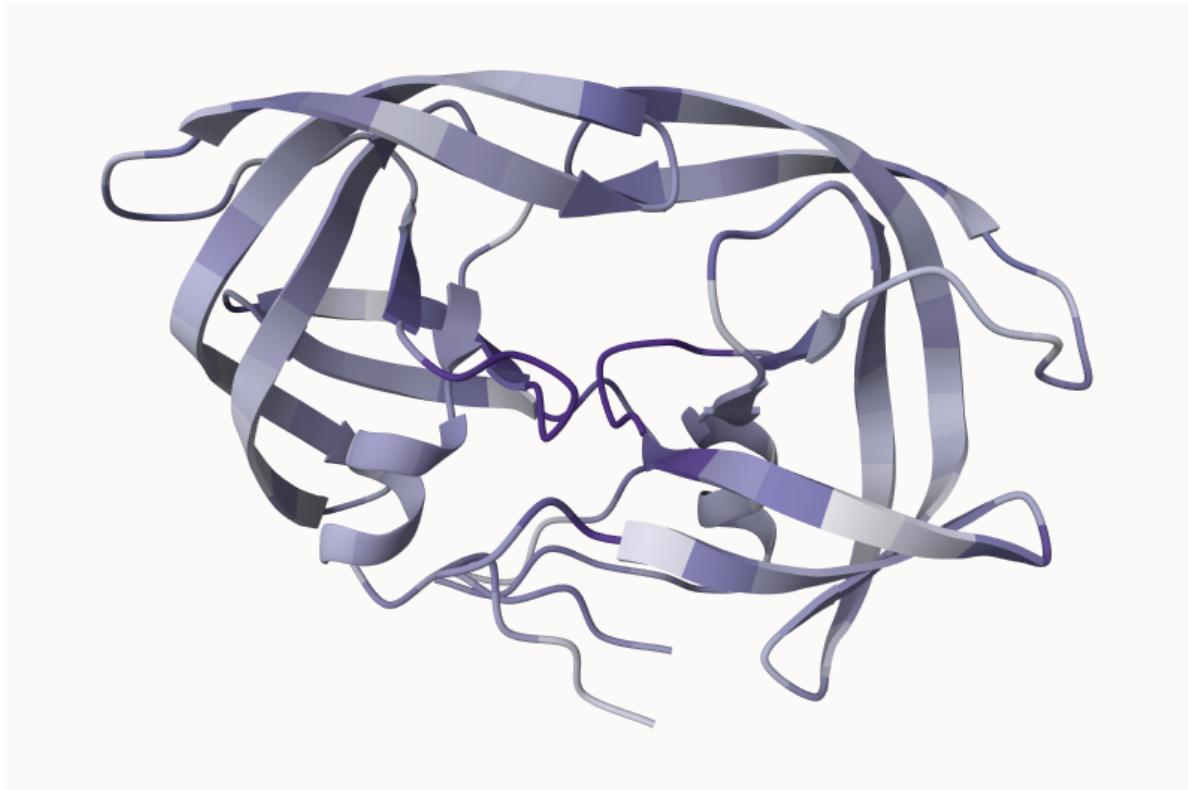


Figure 2: Colored by occupancy