

Class 10: Structural Bioinformatics pt1

Alana (PID: A16738319)

What is in the PDB database

The main repository of bio molecular structure info is the PDB < www.rcsb.org >.

Let's see what this database contains:

```
pdb = "classs10_pdb_stats.csv"
stats <- read.csv(pdb, row.names=1)
head(stats)
```

	X.ray	EM	NMR	Multiple.methods	Neutron	Other
Protein (only)	161,663	12,592	12,337	200	74	32
Protein/Oligosaccharide	9,348	2,167	34	8	2	0
Protein/NA	8,404	3,924	286	7	0	0
Nucleic acid (only)	2,758	125	1,477	14	3	1
Other	164	9	33	0	0	0
Oligosaccharide (only)	11	0	6	1	0	4
Total						
Protein (only)	186,898					
Protein/Oligosaccharide	11,559					
Protein/NA	12,621					
Nucleic acid (only)	4,378					
Other	206					
Oligosaccharide (only)	22					

Q1: What percentage of structures in the PDB are solved by X-Ray and Electron Microscopy.

X.ray	EM
0.8454405519	0.0872433746

```
#as.numeric(stats$X.ray)
```

We got to get rid of commas (to use as.numeric). Can you find a function to get rid of the commas?

```
as.numeric(c(10, "1000"))
```

```
[1] 10 1000
```

```
stats$X.ray <- as.numeric(gsub(",", "", stats$X.ray))
stats$X.ray
```

```
[1] 161663 9348 8404 2758 164 11
```

```
x <- stats$X.ray
sum(as.numeric(gsub(",", "", x)))
```

```
[1] 182348
```

I am going to turn this into a function and then use `apply()` to work on the entire table of data.

```
#sumcomma <- function(x) {
#x <- stats$X.ray
# sum(as.numeric(gsub(",", "", x)))
#}
```

```
#sumcomma(stats$X.ray)
```

This is the function:

```
sumcomma <- function(x) {
  x <- gsub(",", "", x) # Remove commas
  sum(as.numeric(x))    # Convert to numeric and sum
}
```

```
x
```

```
[1] 161663    9348    8404    2758    164     11
```

```
n.total <- sumcomma(stats$Total)
n.total
```

```
[1] 215684
```

```
sumcomma(stats$EM)
```

```
[1] 18817
```

```
apply(stats, 2, sumcomma)
```

X.ray	EM	NMR	Multiple.methods
182348	18817	14173	230
Neutron	Other	Total	
79	37	215684	

```
apply(stats, 2, sumcomma) / sumcomma(stats$Total)
```

X.ray	EM	NMR	Multiple.methods
0.8454405519	0.0872433746	0.0657118748	0.0010663749
Neutron	Other	Total	
0.0003662766	0.0001715473	1.0000000000	

X.ray	EM
0.8454405519	0.0872433746

Q2: What proportion of structures in the PDB are protein?

```
protein_count <- sum(tolower(stats$Type) == "protein")
total_count <- nrow(stats)
protein_proportion <- protein_count / total_count

print(protein_proportion)
```

```
[1] 0
```

Q3: Type HIV in the PDB website search box on the home page and determine how many HIV-1 protease structures are in the current PDB?

In UniProt there are 248,805,733 entries which compared to PDB protein entries (186,898) means there are only ~7% of known sequences with a known structure

248,805,733(uniprot) 186,898(rcsb)

248,805,733 - 186,898

```
186898/248805733 * 100
```

```
[1] 0.07511804
```

Visualizing the HIV-1 protease structure

Mol*("mol-star") viewer is now everywhere. The homepage is here: <https://molstar.org/viewer/>

I want to insert my image from Mol* here.

Working with bio3d package

```
library(bio3d)
```

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

Note: Accessing on-line PDB file

```
pdb
```

```
Call: read.pdb(file = "1hsg")
```

```
Total Models#: 1
```

```
Total Atoms#: 1686, XYZs#: 5058 Chains#: 2 (values: A B)
```

```
Protein Atoms#: 1514 (residues/Calpha atoms#: 198)
```

```
Nucleic acid Atoms#: 0 (residues/phosphate atoms#: 0)
```

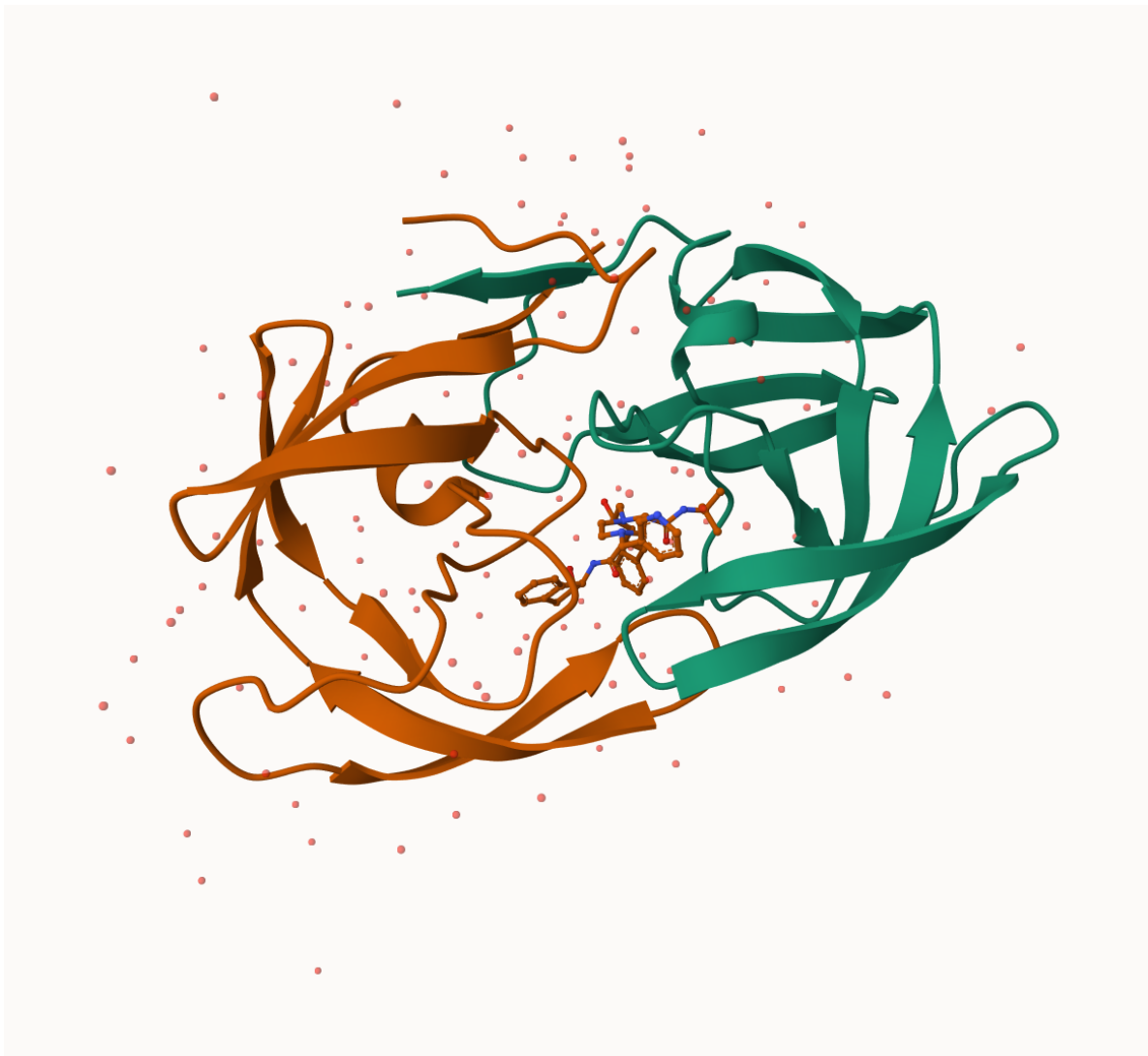


Figure 1: My first

Non-protein/nucleic Atoms#: 172 (residues: 128)
 Non-protein/nucleic resid values: [HOH (127), MK1 (1)]

Protein sequence:

PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIGGFIKVRQYD
 QILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFPQITLWQRPLVTIKIGGQLKE
 ALLDTGADDTVLEEMSLPGRWKPKMIGGIGGFIKVRQYDQILIEICGHKAIGTVLVGPTP
 VNIIGRNLLTQIGCTLNF

+ attr: atom, xyz, seqres, helix, sheet,
 calpha, remark, call

`head(pdb$atom)`

	type	eleno	elety	alt	resid	chain	resno	insert	x	y	z	o	b
1	ATOM	1	N	<NA>	PRO	A	1	<NA>	29.361	39.686	5.862	1	38.10
2	ATOM	2	CA	<NA>	PRO	A	1	<NA>	30.307	38.663	5.319	1	40.62
3	ATOM	3	C	<NA>	PRO	A	1	<NA>	29.760	38.071	4.022	1	42.64
4	ATOM	4	O	<NA>	PRO	A	1	<NA>	28.600	38.302	3.676	1	43.40
5	ATOM	5	CB	<NA>	PRO	A	1	<NA>	30.508	37.541	6.342	1	37.87
6	ATOM	6	CG	<NA>	PRO	A	1	<NA>	29.296	37.591	7.162	1	38.40

	segid	elesy	charge
1	<NA>	N	<NA>
2	<NA>	C	<NA>
3	<NA>	C	<NA>
4	<NA>	O	<NA>
5	<NA>	C	<NA>
6	<NA>	C	<NA>

`pdbseq(pdb)[25]`

25
 "D"
 4.
 5. HOH 308

Predicting functional motions of a single structure

We can do a bioinformatics prediction of functional motions (i.e. flexibility/dynamics):

```
pdb <- read.pdb("6s36")
```

Note: Accessing on-line PDB file

PDB has ALT records, taking A only, rm.alt=TRUE

```
pdb
```

```
Call: read.pdb(file = "6s36")
```

```
Total Models#: 1
```

```
Total Atoms#: 1898, XYZs#: 5694 Chains#: 1 (values: A)
```

```
Protein Atoms#: 1654 (residues/Calpha atoms#: 214)
```

```
Nucleic acid Atoms#: 0 (residues/phosphate atoms#: 0)
```

```
Non-protein/nucleic Atoms#: 244 (residues: 244)
```

```
Non-protein/nucleic resid values: [ CL (3), HOH (238), MG (2), NA (1) ]
```

```
Protein sequence:
```

```
MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMRLRAAVKSGSELGKQAKDIMDAGKLV  
TDELVIALVKERIAQEDCRNGFLLDGFPRTPQADAMKEAGINVDYVLEFDVPDELIVDKI  
VGRRVHAPSGRVYHVKNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIG  
YYSKEAEAGNTKYAKVDGTPVAEVRADLEKILG
```

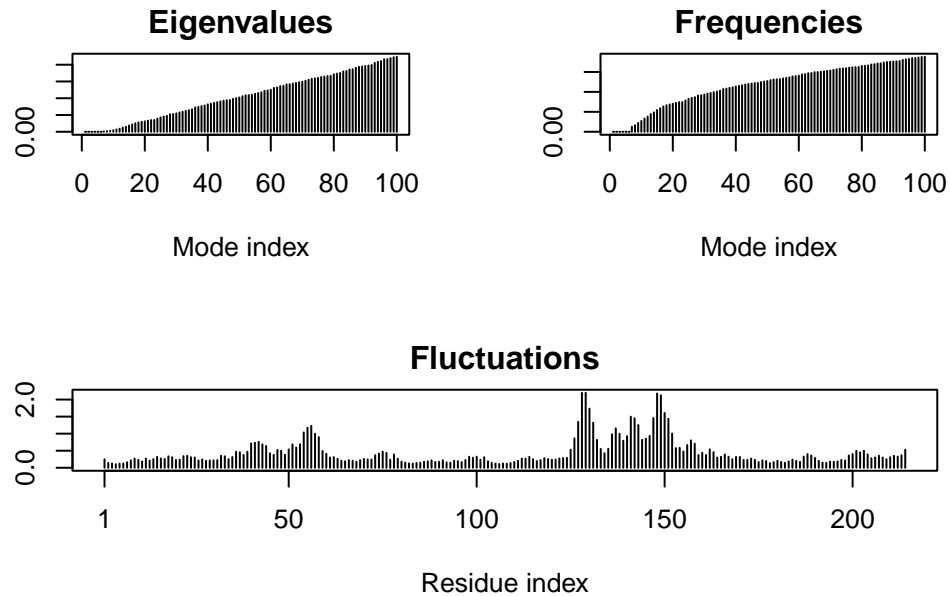
```
+ attr: atom, xyz, seqres, helix, sheet,  
      calpha, remark, call
```

```
m <- nma(pdb)
```

```
Building Hessian... Done in 0.034 seconds.
```

```
Diagonalizing Hessian... Done in 0.355 seconds.
```

```
plot(m)
```



```
mktrj(m, file="adk_m7.pdb")
```

Comparative Analysis of Protein Structures

Here we will work through a complete pipeline of analysis for a given protein family starting from one single database ID for a member of this family all the way to a deeper understanding of structure, dynamics and function.

```
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  MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLV
      1      .      .      .      .      .      .      60

      61      .      .      .      .      .      .      120
pdb|1AKE|A  DELVIALVKERIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFDVPDELIVDRI
      61      .      .      .      .      .      .      120

     121      .      .      .      .      .      .      180
pdb|1AKE|A  VGRRVHAPSGRVYHV KFNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQM TAPLIG
     121      .      .      .      .      .      .      180

     181      .      .      .      214
pdb|1AKE|A  YYSKEAEAGNTKYAKVDG TKPVAEVRADLEKILG
     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

```
# Blast or hmmer search
b <- blast.pdb(aa)
```

Searching ... please wait (updates every 5 seconds) RID = 3V0HP6VC013

.....

Reporting 83 hits

b

\$hit.tbl

	queryid	subjectids	identity	alignmentlength	mismatches	gapopens
1	Query_3576997	1AKE_A	100.000	214	0	0
2	Query_3576997	8BQF_A	99.533	214	1	0
3	Query_3576997	4X8M_A	99.533	214	1	0
4	Query_3576997	6S36_A	99.533	214	1	0
5	Query_3576997	6RZE_A	99.533	214	1	0
6	Query_3576997	4X8H_A	99.533	214	1	0
7	Query_3576997	3HPR_A	99.533	214	1	0
8	Query_3576997	1E4V_A	99.533	214	1	0
9	Query_3576997	5EJE_A	99.065	214	2	0
10	Query_3576997	1E4Y_A	99.065	214	2	0
11	Query_3576997	3X2S_A	98.598	214	3	0
12	Query_3576997	6HAP_A	98.131	214	4	0
13	Query_3576997	6HAM_A	97.196	214	6	0
14	Query_3576997	4K46_A	73.239	213	57	0
15	Query_3576997	4NP6_A	72.642	212	58	0
16	Query_3576997	3GMT_A	62.500	216	75	1
17	Query_3576997	4PZL_A	57.346	211	86	2
18	Query_3576997	5G3Y_A	55.505	218	88	2
19	Query_3576997	5G3Z_A	50.459	218	99	2
20	Query_3576997	5G40_A	49.541	218	101	2
21	Query_3576997	5X6J_A	50.000	218	98	3
22	Query_3576997	2C9Y_A	53.723	188	83	1
23	Query_3576997	1S3G_A	49.541	218	99	3
24	Query_3576997	1AK2_A	52.660	188	85	1
25	Query_3576997	3BE4_A	48.611	216	102	3
26	Query_3576997	1AKY_A	46.119	219	108	3
27	Query_3576997	3AKY_A	46.119	219	108	3
28	Query_3576997	3FB4_A	48.165	218	104	2
29	Query_3576997	4QBI_A	47.248	218	106	2
30	Query_3576997	1DVR_A	45.205	219	110	3
31	Query_3576997	3DKV_A	49.772	219	99	3
32	Query_3576997	3DLO_A	48.165	218	104	2
33	Query_3576997	1ZIN_A	45.413	218	110	2
34	Query_3576997	2P3S_A	47.248	218	106	2
35	Query_3576997	2EU8_A	47.248	218	106	2
36	Query_3576997	1P3J_A	47.248	218	106	2
37	Query_3576997	4QBF_A	49.772	219	99	3
38	Query_3576997	2ORI_A	47.248	218	106	2
39	Query_3576997	5X6I_A	46.789	218	107	2
40	Query_3576997	2QAJ_A	47.005	217	106	2
41	Query_3576997	2007_A	46.789	218	107	2
42	Query_3576997	20SB_A	46.789	218	107	2

43	Query_3576997	4MKF_A	46.789	218	107	2
44	Query_3576997	3TLX_A	44.393	214	106	3
45	Query_3576997	4MKH_A	48.624	218	101	3
46	Query_3576997	4QBH_A	45.872	218	109	2
47	Query_3576997	4TYQ_A	48.165	218	102	3
48	Query_3576997	4QBG_B	47.248	218	104	3
49	Query_3576997	4TYP_A	47.248	218	104	3
50	Query_3576997	4JKY_A	44.037	218	103	5
51	Query_3576997	2RGX_A	43.578	218	104	4
52	Query_3576997	4JLO_A	43.578	218	104	5
53	Query_3576997	1ZAK_A	42.326	215	112	3
54	Query_3576997	1ZD8_A	43.915	189	96	3
55	Query_3576997	2AK3_A	44.324	185	101	2
56	Query_3576997	4NTZ_A	38.532	218	119	4
57	Query_3576997	2AR7_A	41.304	184	102	3
58	Query_3576997	3NDP_A	40.761	184	103	3
59	Query_3576997	1P4S_A	39.785	186	77	2
60	Query_3576997	2CDN_A	39.785	186	77	2
61	Query_3576997	3LOP_A	32.735	223	131	7
62	Query_3576997	5X6L_A	35.784	204	98	3
63	Query_3576997	2XB4_A	32.735	223	131	7
64	Query_3576997	5XRU_A	35.294	204	99	3
65	Query_3576997	5YCC_A	35.294	204	99	3
66	Query_3576997	5X6K_A	35.294	204	99	3
67	Query_3576997	5XZ2_A	33.962	212	107	3
68	Query_3576997	5YCF_A	34.906	212	105	3
69	Query_3576997	5YCB_A	34.804	204	100	3
70	Query_3576997	5YCD_A	36.464	181	87	2
71	Query_3576997	3ADK_A	36.066	183	89	3
72	Query_3576997	3UMF_A	33.333	186	92	3
73	Query_3576997	1Z83_A	34.973	183	91	3
74	Query_3576997	7X7S_A	34.973	183	91	3
75	Query_3576997	3CM0_A	34.434	212	106	5
76	Query_3576997	7DE3_A	36.066	183	89	5
77	Query_3576997	7N6G_6M	33.333	180	108	4
78	Query_3576997	1UKY_A	27.962	211	117	5
79	Query_3576997	1TEV_A	31.963	219	109	7
80	Query_3576997	7E9V_A	31.963	219	109	7
81	Query_3576997	2BWJ_A	30.851	188	98	4
82	Query_3576997	1QF9_A	27.907	215	117	5
83	Query_3576997	7N6G_6A	24.852	169	84	6
q.start q.end s.start s.end evalue bitscore positives mlog.evalue pdb.id						
1	1	214	1	214	1.51e-156	432.0 100.00 358.79116 1AKE_A

2	1	214	21	234	2.47e-156	433.0	100.00	358.29906	8BQF_A
3	1	214	1	214	2.71e-156	432.0	100.00	358.20633	4X8M_A
4	1	214	1	214	3.97e-156	432.0	100.00	357.82451	6S36_A
5	1	214	1	214	1.14e-155	431.0	99.53	356.76966	6RZE_A
6	1	214	1	214	1.50e-155	430.0	99.53	356.49522	4X8H_A
7	1	214	1	214	2.13e-155	430.0	99.53	356.14457	3HPR_A
8	1	214	1	214	2.25e-155	430.0	99.53	356.08976	1E4V_A
9	1	214	1	214	6.74e-155	429.0	99.07	354.99263	5EJE_A
10	1	214	1	214	3.90e-154	427.0	99.07	353.23713	1E4Y_A
11	1	214	1	214	6.53e-154	426.0	98.60	352.72170	3X2S_A
12	1	214	1	214	1.93e-153	425.0	98.60	351.63800	6HAP_A
13	1	214	1	214	3.95e-153	424.0	98.60	350.92180	6HAM_A
14	1	213	1	213	1.75e-115	329.0	84.98	264.23767	4K46_A
15	2	213	5	216	9.73e-114	325.0	84.43	260.21949	4NP6_A
16	2	211	10	225	7.61e-90	265.0	71.30	205.20320	3GMT_A
17	2	209	26	235	1.82e-86	256.0	74.41	197.42348	4PZL_A
18	1	214	1	213	2.69e-76	230.0	68.81	174.00693	5G3Y_A
19	1	214	1	213	5.15e-73	221.0	69.27	166.44972	5G3Z_A
20	1	214	1	213	1.05e-70	216.0	68.35	161.13217	5G40_A
21	1	213	1	212	1.92e-68	210.0	65.60	155.92346	5X6J_A
22	1	184	17	204	9.71e-68	209.0	69.68	154.30263	2C9Y_A
23	1	213	1	212	9.99e-68	208.0	65.14	154.27420	1S3G_A
24	1	184	17	204	2.51e-67	207.0	70.21	153.35292	1AK2_A
25	2	213	7	217	6.73e-67	206.0	68.06	152.36663	3BE4_A
26	1	214	5	218	9.51e-67	206.0	65.75	152.02086	1AKY_A
27	1	214	5	218	1.16e-66	205.0	65.30	151.82220	3AKY_A
28	1	214	1	213	4.25e-66	204.0	65.14	150.52370	3FB4_A
29	1	214	1	213	2.06e-64	199.0	65.60	146.64274	4QBI_A
30	1	214	5	218	4.44e-64	199.0	64.84	145.87479	1DVR_A
31	1	214	1	213	2.39e-63	197.0	66.67	144.19157	3DKV_A
32	1	214	1	213	1.05e-62	195.0	66.97	142.71149	3DLO_A
33	1	214	1	213	1.17e-62	195.0	65.60	142.60327	1ZIN_A
34	1	214	1	213	1.77e-62	194.0	66.97	142.18930	2P3S_A
35	1	214	1	213	1.85e-62	194.0	66.97	142.14509	2EU8_A
36	1	214	1	213	2.51e-62	194.0	66.97	141.83999	1P3J_A
37	1	214	1	213	4.01e-62	194.0	66.21	141.37148	4QBF_A
38	1	214	1	213	6.28e-62	193.0	66.97	140.92291	2ORI_A
39	1	214	1	213	1.19e-61	192.0	66.51	140.28374	5X6I_A
40	1	213	1	212	1.29e-61	192.0	66.82	140.20305	2QAJ_A
41	1	214	1	213	1.50e-61	192.0	66.51	140.05223	2007_A
42	1	214	1	213	2.56e-61	192.0	66.51	139.51768	2OSB_A
43	1	214	1	213	2.85e-61	191.0	66.51	139.41037	4MKF_A
44	2	211	31	235	1.80e-60	190.0	64.95	137.56732	3TLX_A

45	1	213	3	214	1.84e-60	189.0	65.60	137.54534	4MKH_A
46	1	214	1	213	3.68e-60	189.0	64.68	136.85219	4QBH_A
47	1	213	1	212	4.53e-60	188.0	65.60	136.64438	4TYQ_A
48	1	213	1	212	7.92e-59	185.0	65.60	133.78313	4QBG_B
49	1	213	1	212	1.49e-58	184.0	65.14	133.15116	4TYP_A
50	1	214	1	203	6.84e-56	177.0	66.51	127.02198	4JKY_A
51	1	214	1	203	7.84e-56	177.0	65.60	126.88553	2RGX_A
52	1	214	1	203	2.34e-55	176.0	66.51	125.79203	4JLO_A
53	1	214	6	209	4.71e-54	173.0	63.72	122.78991	1ZAK_A
54	1	185	8	190	3.66e-50	164.0	64.55	113.83179	1ZD8_A
55	1	185	7	189	5.03e-50	163.0	65.41	113.51383	2AK3_A
56	1	213	6	213	1.80e-46	154.0	62.39	105.33113	4NTZ_A
57	1	182	28	207	6.34e-45	150.0	64.13	101.76945	2AR7_A
58	1	182	6	185	5.30e-44	148.0	63.59	99.64604	3NDP_A
59	1	182	1	155	8.80e-39	133.0	56.99	87.62607	1P4S_A
60	1	182	21	175	1.58e-38	133.0	56.99	87.04081	2CDN_A
61	1	209	1	218	3.38e-31	115.0	54.26	70.16226	3LOP_A
62	3	205	13	184	4.13e-31	114.0	52.45	69.96186	5X6L_A
63	1	209	1	218	4.47e-31	114.0	54.26	69.88275	2XB4_A
64	3	205	11	182	4.96e-31	113.0	52.45	69.77873	5XRU_A
65	3	205	11	182	5.83e-31	113.0	52.45	69.61712	5YCC_A
66	3	205	13	184	6.91e-31	113.0	52.45	69.44717	5X6K_A
67	3	213	13	192	8.15e-31	113.0	52.83	69.28212	5XZ2_A
68	3	213	11	190	8.30e-31	113.0	51.89	69.26388	5YCF_A
69	3	205	11	182	8.95e-31	113.0	52.45	69.18848	5YCB_A
70	3	182	11	164	2.37e-30	112.0	54.70	68.21466	5YCD_A
71	3	184	12	167	2.84e-30	111.0	54.10	68.03375	3ADK_A
72	3	185	32	188	1.50e-29	110.0	56.45	66.36950	3UMF_A
73	3	184	12	167	1.76e-29	109.0	54.64	66.20965	1Z83_A
74	3	184	16	171	2.09e-29	109.0	54.64	66.03780	7X7S_A
75	3	214	7	185	8.50e-29	107.0	50.94	64.63490	3CMO_A
76	3	184	11	166	1.12e-28	107.0	56.28	64.35905	7DE3_A
77	1	168	1239	1418	1.46e-25	105.0	53.33	57.18619	7N6G_6
78	3	210	18	196	9.19e-25	97.8	53.55	55.34651	1UKY_A
79	3	213	6	192	6.51e-24	95.5	49.77	53.38870	1TEV_A
80	3	213	24	210	1.01e-23	95.5	49.77	52.94951	7E9V_A
81	3	187	15	173	6.73e-22	90.1	49.47	48.75030	2BWJ_A
82	3	213	9	189	2.62e-20	85.9	47.44	45.08853	1QF9_A
83	76	207	760	922	1.22e-05	47.0	42.01	11.31407	7N6G_6

acc
1 1AKE_A
2 8BQF_A
3 4X8M_A

4 6S36_A
5 6RZE_A
6 4X8H_A
7 3HPR_A
8 1E4V_A
9 5EJE_A
10 1E4Y_A
11 3X2S_A
12 6HAP_A
13 6HAM_A
14 4K46_A
15 4NP6_A
16 3GMT_A
17 4PZL_A
18 5G3Y_A
19 5G3Z_A
20 5G40_A
21 5X6J_A
22 2C9Y_A
23 1S3G_A
24 1AK2_A
25 3BE4_A
26 1AKY_A
27 3AKY_A
28 3FB4_A
29 4QBI_A
30 1DVR_A
31 3DKV_A
32 3DLO_A
33 1ZIN_A
34 2P3S_A
35 2EU8_A
36 1P3J_A
37 4QBF_A
38 2ORI_A
39 5X6I_A
40 2QAJ_A
41 2007_A
42 2OSB_A
43 4MKF_A
44 3TLX_A
45 4MKH_A
46 4QBH_A

47 4TYQ_A
 48 4QBG_B
 49 4TYP_A
 50 4JKY_A
 51 2RGX_A
 52 4JLO_A
 53 1ZAK_A
 54 1ZD8_A
 55 2AK3_A
 56 4NTZ_A
 57 2AR7_A
 58 3NDP_A
 59 1P4S_A
 60 2CDN_A
 61 3LOP_A
 62 5X6L_A
 63 2XB4_A
 64 5XRU_A
 65 5YCC_A
 66 5X6K_A
 67 5XZ2_A
 68 5YCF_A
 69 5YCB_A
 70 5YCD_A
 71 3ADK_A
 72 3UMF_A
 73 1Z83_A
 74 7X7S_A
 75 3CMO_A
 76 7DE3_A
 77 7N6G_6M
 78 1UKY_A
 79 1TEV_A
 80 7E9V_A
 81 2BWJ_A
 82 1QF9_A
 83 7N6G_6A

\$raw

	queryid	subjectids	identity	alignmentlength	mismatches	gapopens
1	Query_3576997	1AKE_A	100.000	214	0	0
2	Query_3576997	8BQF_A	99.533	214	1	0
3	Query_3576997	4X8M_A	99.533	214	1	0

4	Query_3576997	6S36_A	99.533	214	1	0
5	Query_3576997	6RZE_A	99.533	214	1	0
6	Query_3576997	4X8H_A	99.533	214	1	0
7	Query_3576997	3HPR_A	99.533	214	1	0
8	Query_3576997	1E4V_A	99.533	214	1	0
9	Query_3576997	5EJE_A	99.065	214	2	0
10	Query_3576997	1E4Y_A	99.065	214	2	0
11	Query_3576997	3X2S_A	98.598	214	3	0
12	Query_3576997	6HAP_A	98.131	214	4	0
13	Query_3576997	6HAM_A	97.196	214	6	0
14	Query_3576997	4K46_A	73.239	213	57	0
15	Query_3576997	4NP6_A	72.642	212	58	0
16	Query_3576997	3GMT_A	62.500	216	75	1
17	Query_3576997	4PZL_A	57.346	211	86	2
18	Query_3576997	5G3Y_A	55.505	218	88	2
19	Query_3576997	5G3Z_A	50.459	218	99	2
20	Query_3576997	5G40_A	49.541	218	101	2
21	Query_3576997	5X6J_A	50.000	218	98	3
22	Query_3576997	2C9Y_A	53.723	188	83	1
23	Query_3576997	1S3G_A	49.541	218	99	3
24	Query_3576997	1AK2_A	52.660	188	85	1
25	Query_3576997	3BE4_A	48.611	216	102	3
26	Query_3576997	1AKY_A	46.119	219	108	3
27	Query_3576997	3AKY_A	46.119	219	108	3
28	Query_3576997	3FB4_A	48.165	218	104	2
29	Query_3576997	4QBI_A	47.248	218	106	2
30	Query_3576997	1DVR_A	45.205	219	110	3
31	Query_3576997	3DKV_A	49.772	219	99	3
32	Query_3576997	3DLO_A	48.165	218	104	2
33	Query_3576997	1ZIN_A	45.413	218	110	2
34	Query_3576997	2P3S_A	47.248	218	106	2
35	Query_3576997	2EU8_A	47.248	218	106	2
36	Query_3576997	1P3J_A	47.248	218	106	2
37	Query_3576997	4QBF_A	49.772	219	99	3
38	Query_3576997	2ORI_A	47.248	218	106	2
39	Query_3576997	5X6I_A	46.789	218	107	2
40	Query_3576997	2QAJ_A	47.005	217	106	2
41	Query_3576997	2007_A	46.789	218	107	2
42	Query_3576997	2OSB_A	46.789	218	107	2
43	Query_3576997	4MKF_A	46.789	218	107	2
44	Query_3576997	3TLX_A	44.393	214	106	3
45	Query_3576997	4MKH_A	48.624	218	101	3
46	Query_3576997	4QBH_A	45.872	218	109	2

47	Query_3576997	4TYQ_A	48.165	218	102	3
48	Query_3576997	4QBG_B	47.248	218	104	3
49	Query_3576997	4TYP_A	47.248	218	104	3
50	Query_3576997	4JKY_A	44.037	218	103	5
51	Query_3576997	2RGX_A	43.578	218	104	4
52	Query_3576997	4JLO_A	43.578	218	104	5
53	Query_3576997	1ZAK_A	42.326	215	112	3
54	Query_3576997	1ZD8_A	43.915	189	96	3
55	Query_3576997	2AK3_A	44.324	185	101	2
56	Query_3576997	4NTZ_A	38.532	218	119	4
57	Query_3576997	2AR7_A	41.304	184	102	3
58	Query_3576997	3NDP_A	40.761	184	103	3
59	Query_3576997	1P4S_A	39.785	186	77	2
60	Query_3576997	2CDN_A	39.785	186	77	2
61	Query_3576997	3LOP_A	32.735	223	131	7
62	Query_3576997	5X6L_A	35.784	204	98	3
63	Query_3576997	2XB4_A	32.735	223	131	7
64	Query_3576997	5XRU_A	35.294	204	99	3
65	Query_3576997	5YCC_A	35.294	204	99	3
66	Query_3576997	5X6K_A	35.294	204	99	3
67	Query_3576997	5XZ2_A	33.962	212	107	3
68	Query_3576997	5YCF_A	34.906	212	105	3
69	Query_3576997	5YCB_A	34.804	204	100	3
70	Query_3576997	5YCD_A	36.464	181	87	2
71	Query_3576997	3ADK_A	36.066	183	89	3
72	Query_3576997	3UMF_A	33.333	186	92	3
73	Query_3576997	1Z83_A	34.973	183	91	3
74	Query_3576997	7X7S_A	34.973	183	91	3
75	Query_3576997	3CM0_A	34.434	212	106	5
76	Query_3576997	7DE3_A	36.066	183	89	5
77	Query_3576997	7N6G_6M	33.333	180	108	4
78	Query_3576997	1UKY_A	27.962	211	117	5
79	Query_3576997	1TEV_A	31.963	219	109	7
80	Query_3576997	7E9V_A	31.963	219	109	7
81	Query_3576997	2BWJ_A	30.851	188	98	4
82	Query_3576997	1QF9_A	27.907	215	117	5
83	Query_3576997	7N6G_6A	24.852	169	84	6

	q.start	q.end	s.start	s.end	eval	bitscore	positives
1	1	214	1	214	1.51e-156	432.0	100.00
2	1	214	21	234	2.47e-156	433.0	100.00
3	1	214	1	214	2.71e-156	432.0	100.00
4	1	214	1	214	3.97e-156	432.0	100.00
5	1	214	1	214	1.14e-155	431.0	99.53

6	1	214	1	214	1.50e-155	430.0	99.53
7	1	214	1	214	2.13e-155	430.0	99.53
8	1	214	1	214	2.25e-155	430.0	99.53
9	1	214	1	214	6.74e-155	429.0	99.07
10	1	214	1	214	3.90e-154	427.0	99.07
11	1	214	1	214	6.53e-154	426.0	98.60
12	1	214	1	214	1.93e-153	425.0	98.60
13	1	214	1	214	3.95e-153	424.0	98.60
14	1	213	1	213	1.75e-115	329.0	84.98
15	2	213	5	216	9.73e-114	325.0	84.43
16	2	211	10	225	7.61e-90	265.0	71.30
17	2	209	26	235	1.82e-86	256.0	74.41
18	1	214	1	213	2.69e-76	230.0	68.81
19	1	214	1	213	5.15e-73	221.0	69.27
20	1	214	1	213	1.05e-70	216.0	68.35
21	1	213	1	212	1.92e-68	210.0	65.60
22	1	184	17	204	9.71e-68	209.0	69.68
23	1	213	1	212	9.99e-68	208.0	65.14
24	1	184	17	204	2.51e-67	207.0	70.21
25	2	213	7	217	6.73e-67	206.0	68.06
26	1	214	5	218	9.51e-67	206.0	65.75
27	1	214	5	218	1.16e-66	205.0	65.30
28	1	214	1	213	4.25e-66	204.0	65.14
29	1	214	1	213	2.06e-64	199.0	65.60
30	1	214	5	218	4.44e-64	199.0	64.84
31	1	214	1	213	2.39e-63	197.0	66.67
32	1	214	1	213	1.05e-62	195.0	66.97
33	1	214	1	213	1.17e-62	195.0	65.60
34	1	214	1	213	1.77e-62	194.0	66.97
35	1	214	1	213	1.85e-62	194.0	66.97
36	1	214	1	213	2.51e-62	194.0	66.97
37	1	214	1	213	4.01e-62	194.0	66.21
38	1	214	1	213	6.28e-62	193.0	66.97
39	1	214	1	213	1.19e-61	192.0	66.51
40	1	213	1	212	1.29e-61	192.0	66.82
41	1	214	1	213	1.50e-61	192.0	66.51
42	1	214	1	213	2.56e-61	192.0	66.51
43	1	214	1	213	2.85e-61	191.0	66.51
44	2	211	31	235	1.80e-60	190.0	64.95
45	1	213	3	214	1.84e-60	189.0	65.60
46	1	214	1	213	3.68e-60	189.0	64.68
47	1	213	1	212	4.53e-60	188.0	65.60
48	1	213	1	212	7.92e-59	185.0	65.60

49	1	213	1	212	1.49e-58	184.0	65.14
50	1	214	1	203	6.84e-56	177.0	66.51
51	1	214	1	203	7.84e-56	177.0	65.60
52	1	214	1	203	2.34e-55	176.0	66.51
53	1	214	6	209	4.71e-54	173.0	63.72
54	1	185	8	190	3.66e-50	164.0	64.55
55	1	185	7	189	5.03e-50	163.0	65.41
56	1	213	6	213	1.80e-46	154.0	62.39
57	1	182	28	207	6.34e-45	150.0	64.13
58	1	182	6	185	5.30e-44	148.0	63.59
59	1	182	1	155	8.80e-39	133.0	56.99
60	1	182	21	175	1.58e-38	133.0	56.99
61	1	209	1	218	3.38e-31	115.0	54.26
62	3	205	13	184	4.13e-31	114.0	52.45
63	1	209	1	218	4.47e-31	114.0	54.26
64	3	205	11	182	4.96e-31	113.0	52.45
65	3	205	11	182	5.83e-31	113.0	52.45
66	3	205	13	184	6.91e-31	113.0	52.45
67	3	213	13	192	8.15e-31	113.0	52.83
68	3	213	11	190	8.30e-31	113.0	51.89
69	3	205	11	182	8.95e-31	113.0	52.45
70	3	182	11	164	2.37e-30	112.0	54.70
71	3	184	12	167	2.84e-30	111.0	54.10
72	3	185	32	188	1.50e-29	110.0	56.45
73	3	184	12	167	1.76e-29	109.0	54.64
74	3	184	16	171	2.09e-29	109.0	54.64
75	3	214	7	185	8.50e-29	107.0	50.94
76	3	184	11	166	1.12e-28	107.0	56.28
77	1	168	1239	1418	1.46e-25	105.0	53.33
78	3	210	18	196	9.19e-25	97.8	53.55
79	3	213	6	192	6.51e-24	95.5	49.77
80	3	213	24	210	1.01e-23	95.5	49.77
81	3	187	15	173	6.73e-22	90.1	49.47
82	3	213	9	189	2.62e-20	85.9	47.44
83	76	207	760	922	1.22e-05	47.0	42.01

\$url

"https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Get&FORMAT_OBJECT=Alignment&ALIGNMENT_VIEW=Tab

attr(", "class")

[1] "blast"

```

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

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

```

```

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

```

```

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

```

```

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

```

```

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

```

```

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

```

```

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

```

```

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

```

```

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

```

```

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

```

```

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

```

```

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

```

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

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

	0%
=====	8%
=====	15%
=====	23%
=====	31%
=====	38%
=====	46%
=====	54%
=====	62%
=====	69%
=====	77%
=====	85%
=====	92%
=====	100%

```
#install.packages("BiocManager")
```

```
#BiocManager::install("msa")
```

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

Reading PDB files:

```
pdbs/split_chain/1AKE_A.pdb
pdbs/split_chain/6S36_A.pdb
pdbs/split_chain/6RZE_A.pdb
pdbs/split_chain/3HPR_A.pdb
pdbs/split_chain/1E4V_A.pdb
pdbs/split_chain/5EJE_A.pdb
pdbs/split_chain/1E4Y_A.pdb
pdbs/split_chain/3X2S_A.pdb
pdbs/split_chain/6HAP_A.pdb
pdbs/split_chain/6HAM_A.pdb
pdbs/split_chain/4K46_A.pdb
pdbs/split_chain/3GMT_A.pdb
pdbs/split_chain/4PZL_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
.   PDB has ALT records, taking A only, rm.alt=TRUE
.   PDB has ALT records, taking A only, rm.alt=TRUE
.   PDB has ALT records, taking A only, rm.alt=TRUE
..  PDB has ALT records, taking A only, rm.alt=TRUE
.... PDB has ALT records, taking A only, rm.alt=TRUE
.   PDB has ALT records, taking A only, rm.alt=TRUE
...
```

Extracting sequences

```
pdb/seq: 1  name: pdbs/split_chain/1AKE_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 2  name: pdbs/split_chain/6S36_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 3  name: pdbs/split_chain/6RZE_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 4  name: pdbs/split_chain/3HPR_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 5  name: pdbs/split_chain/1E4V_A.pdb
pdb/seq: 6  name: pdbs/split_chain/5EJE_A.pdb
  PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 7  name: pdbs/split_chain/1E4Y_A.pdb
pdb/seq: 8  name: pdbs/split_chain/3X2S_A.pdb
pdb/seq: 9  name: pdbs/split_chain/6HAP_A.pdb
```

```

pdb/seq: 10   name: pdbs/split_chain/6HAM_A.pdb
      PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 11   name: pdbs/split_chain/4K46_A.pdb
      PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 12   name: pdbs/split_chain/3GMT_A.pdb
pdb/seq: 13   name: pdbs/split_chain/4PZL_A.pdb

```

Print out the alignment

pdbs

```

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

41                                     .               .               .               80
[Truncated_Name:1] 1AKE_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDLVIALVKE
[Truncated_Name:2] 6S36_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDLVIALVKE
[Truncated_Name:3] 6RZE_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDLVIALVKE
[Truncated_Name:4] 3HPR_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDLVIALVKE
[Truncated_Name:5] 1E4V_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDLVIALVKE
[Truncated_Name:6] 5EJE_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDACKLVTDLVIALVKE
[Truncated_Name:7] 1E4Y_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDLVIALVKE
[Truncated_Name:8] 3X2S_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDCGKLVTDLVIALVKE
[Truncated_Name:9] 6HAP_A.pdb      TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDLVIALVRE
[Truncated_Name:10] 6HAM_A.pdb      TGDMLRAAIKSGSELGKQAKDIMDAGKLVTDIIIALVKE
[Truncated_Name:11] 4K46_A.pdb      TGDMLRAAIKAGTELGKQAKSVIDAGQLVSDDIILGLVKE
[Truncated_Name:12] 3GMT_A.pdb      TGDMLRAAVKAGTPLGVEAKTYMDEGKLVPSLIIGLVKE
[Truncated_Name:13] 4PZL_A.pdb      TGDMIRETIKSGSALGQELKKVLDAGELVSDEFIIKIVKD

```

```

****~*  ~* *~ **  *  ~*  ** *  ^^ ~*~^
41      .      .      .      80

81      .      .      .      120
[Truncated_Name:1] 1AKE_A.pdb  RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:2] 6S36_A.pdb  RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:3] 6RZE_A.pdb  RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:4] 3HPR_A.pdb  RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:5] 1E4V_A.pdb  RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:6] 5EJE_A.pdb  RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:7] 1E4Y_A.pdb  RIAQEDCRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:8] 3X2S_A.pdb  RIAQEDSRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:9] 6HAP_A.pdb  RICQEDSRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:10] 6HAM_A.pdb  RICQEDSRNGFLLDGFPR TIPQADAMKEAGINVDYVLEFD
[Truncated_Name:11] 4K46_A.pdb  RIAQDDCAKGFLLDGFPR TIPQADGLKEVGVVVDYVIEFD
[Truncated_Name:12] 3GMT_A.pdb  RLKEADCANGYLFDFGPR TIPQADAMKEAGVAIDYVLEID
[Truncated_Name:13] 4PZL_A.pdb  RISKNDCNNGFLLDGVPR TIPQAQELDKLGVNIDYIVEVD
*~  *  *~* ** ***** **  ^  *~ ~**~*~* *
81      .      .      .      120

121     .      .      .      160
[Truncated_Name:1] 1AKE_A.pdb  VPDELIVDRIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG
[Truncated_Name:2] 6S36_A.pdb  VPDELIVDKIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG
[Truncated_Name:3] 6RZE_A.pdb  VPDELIVDAIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG
[Truncated_Name:4] 3HPR_A.pdb  VPDELIVDRIVGRRVHAPSGRVYHV KFNPPKVEGKDDGTG
[Truncated_Name:5] 1E4V_A.pdb  VPDELIVDRIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG
[Truncated_Name:6] 5EJE_A.pdb  VPDELIVDRIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG
[Truncated_Name:7] 1E4Y_A.pdb  VPDELIVDRIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG
[Truncated_Name:8] 3X2S_A.pdb  VPDELIVDRIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG
[Truncated_Name:9] 6HAP_A.pdb  VPDELIVDRIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG
[Truncated_Name:10] 6HAM_A.pdb  VPDELIVDRIVGRRVHAPSGRVYHV KFNPPKVEGKDDVTG
[Truncated_Name:11] 4K46_A.pdb  VADSVIVERMAGRRASGR TYHNVNPPKVEGKDDVTG
[Truncated_Name:12] 3GMT_A.pdb  VPFSEIIERMSGRRTHPASGR TYHV KFNPPKVEGKDDVTG
[Truncated_Name:13] 4PZL_A.pdb  VADNLLIERITGRRIHPASGR TYHTKFNPPKVADKDDVTG
*  ^^^ ^ *** *  *** **  ^***** *** **
121     .      .      .      160

161     .      .      .      200
[Truncated_Name:1] 1AKE_A.pdb  EELTTRKDDQEETVRKRLVEYHQMTAP LIGYYSKEAEAGN
[Truncated_Name:2] 6S36_A.pdb  EELTTRKDDQEETVRKRLVEYHQMTAP LIGYYSKEAEAGN
[Truncated_Name:3] 6RZE_A.pdb  EELTTRKDDQEETVRKRLVEYHQMTAP LIGYYSKEAEAGN
[Truncated_Name:4] 3HPR_A.pdb  EELTTRKDDQEETVRKRLVEYHQMTAP LIGYYSKEAEAGN
[Truncated_Name:5] 1E4V_A.pdb  EELTTRKDDQEETVRKRLVEYHQMTAP LIGYYSKEAEAGN

```



```

[Truncated_Name:6] 5EJE_A.pdb      EELTTRKDDQEECVRKRLVEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:7] 1E4Y_A.pdb      EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:8] 3X2S_A.pdb      EELTTRKDDQEETVRKRLCEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:9] 6HAP_A.pdb      EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:10] 6HAM_A.pdb      EELTTRKDDQEETVRKRLVEYHQM TAPLIGYYSKEAEAGN
[Truncated_Name:11] 4K46_A.pdb      EDLVIREDDKEETVLARLG VYHNQTAPLIAYYGKEAEAGN
[Truncated_Name:12] 3GMT_A.pdb      EPLVQRDDDKKEETVKKRLDVYEAQTKPLITYYGDWARRGA
[Truncated_Name:13] 4PZL_A.pdb      EPLITRTDDNEDTVKQRLSVYHAQTAKLIDFYRNFSSNTNT
                                     * * * * * ^ * * * * * ^ *
                                     161           .           .           .           200

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

```

Call:

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

Class:

```
pdbs, fasta
```

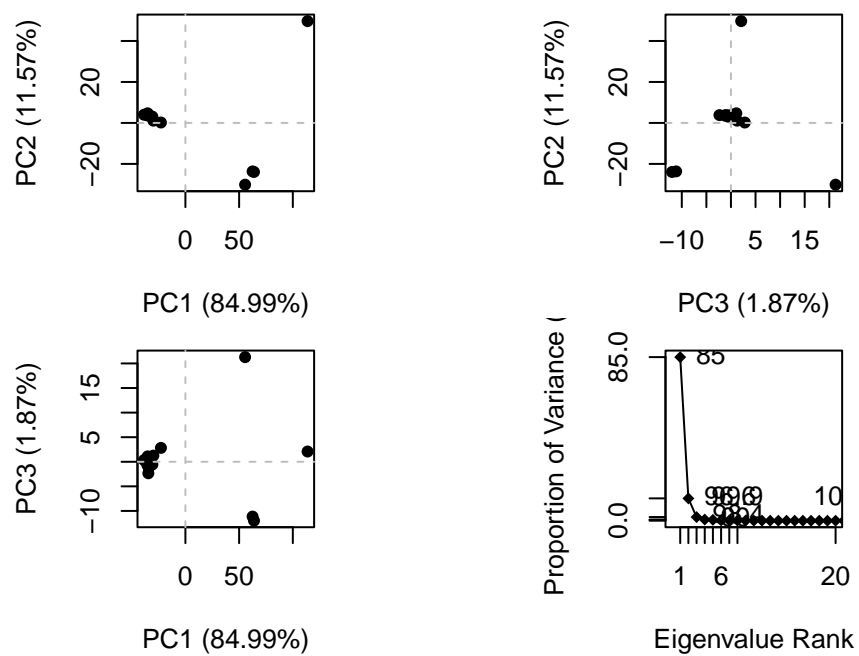
Alignment dimensions:

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

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

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

```
plot(pc.xray)
```



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

```
Total Frames#: 34
Total XYZs#: 612, (Atoms#: 204)
```

```
[1] 26.787 52.261 40.414 <...> 15.653 53.622 42.018 [20808]
```

```
+ attr: Matrix DIM = 34 x 612
```