

# Comparative Structure Analysis

Aileen Andrade

2025-02-11

Here we run through a complete “pipeline” of structure analysis that begins with a single sequence identifier and ends in a PCA analysis.

```
library(bio3d)
```

## Step 1. Get a sequence

Retrieve a sequence for the protein we are interested in. We will take ADK “1ake\_A”

```
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 MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGMLRAAVKSGSELGKQAKDIMDAGKLV  
      1      .      .      .      .      .      60  
  
     61      .      .      .      .      .      120  
pdb|1AKE|A DELVIALVKERIAQEDCRNGFLLDGFPRTPQADAMKEAGINVDYVLEFDVPDELIVDR  
     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
```

## Step 2. Blast search

Run a BLAST search of the PDB for all related sequences to our input `aa`

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

```
Searching ... please wait (updates every 5 seconds) RID = V5C1XSJ9016
```

```
.....
```

```
Reporting 87 hits
```

```
hits <- plot(blast)
```

```

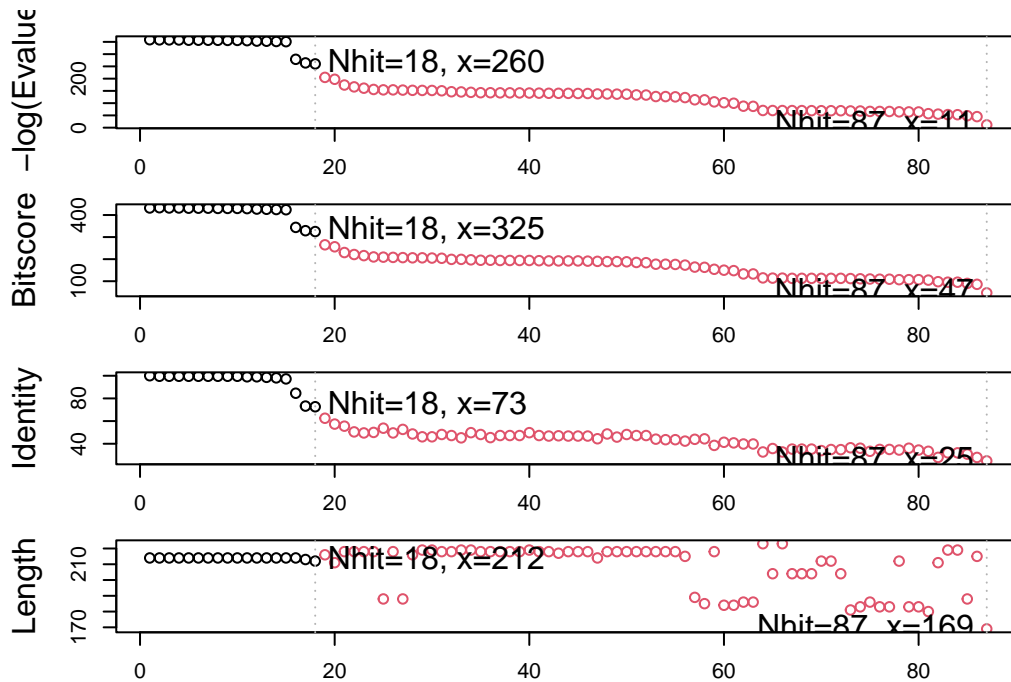
* Possible cutoff values: 260 11
  Yielding Nhits:       18 87

```

```

* Chosen cutoff value of: 260
  Yielding Nhits:       18

```



### Step 3. Download all structure

These are our “top hits” i.e. all the structures in the PDB database related to our input sequence.

```
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" "8PVW_A"
[17] "4K46_A" "4NP6_A"
```

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

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

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

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

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

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

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

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

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

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

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

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

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

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

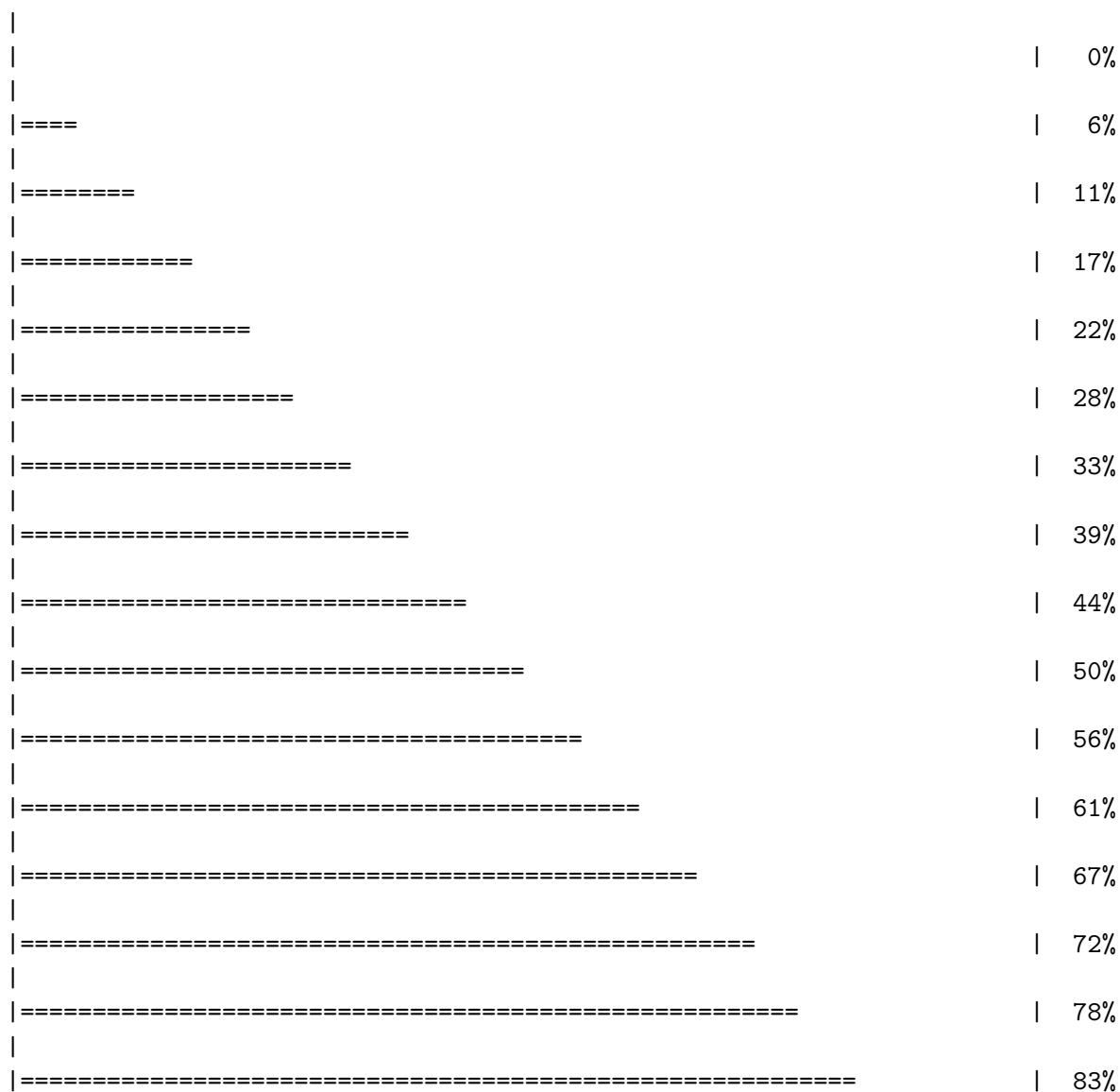
Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):  
pdbs/6HAP.pdb.gz exists. Skipping download

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

Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):  
pdbs/8PVW.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



	=====	89%
	=====	94%
	=====	100%

#### Step 4. Align and superpose