Class 09 Structural Bioinformatics

Jessica Diaz-Vigil

5/7/23

Introduction to the RCSB Protein Data Bank (PDB)

PDB Statistics

Exporting the data:

```
pdb.data <- "Data Export Summary.csv"
```

```
pdb.df <- read.csv("Data Export Summary.csv", row.names = 1)
head(pdb.df)
```

	X.ray	EM	NMR	Multiple.methods	Neutron	Other
Protein (only)	154,766	10,155	12,187	191	72	32
Protein/Oligosaccharide	9,083	1,802	32	7	1	0
Protein/NA	8,110	3,176	283	6	0	0
Nucleic acid (only)	2,664	94	1,450	12	2	1
Other	163	9	32	0	0	0
Oligosaccharide (only)	11	0	6	1	0	4

	Total
Protein (only)	177,403
Protein/Oligosaccharide	10,925
Protein/NA	11,575
Nucleic acid (only)	4,223
Other	204
Oligosaccharide (only)	22

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

To do this, first we need to remove the commas, then we need to convert into numbers using `as.numeric`:

```
pdb.df$X.ray <- as.numeric(gsub(",", "", pdb.df$X.ray))
pdb.df$EM <- as.numeric(gsub(",", "", pdb.df$EM))
pdb.df$Total <- as.numeric(gsub(",", "", pdb.df$Total))
```

Secondly, we need to take the sum of the new `X.ray` values:

```
n_xray <- sum(pdb.df$X.ray)
n_em <- sum(pdb.df$EM)
n_total <- sum(pdb.df$Total)
```

Finally, we need to take the percentage using the sum divided by the total:

```
sum_xray_em <- n_xray + n_em
sum_xray_em/n_total*100
```

```
[1] 92.99297
```

The percentage of PDB structures solved by X-Ray and EM is 92.99%

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

  To do this, we will first need to see how many proteins there are, then divide by the total:

  ```
  n_total_protein <- sum(pdb.df$Total[1:3])
  n_proportion_protein <- n_total_protein / n_total
  n_proportion_protein*100
  ```

  ```
  [1] 97.82287
  ```

  It was found that 97.82% of the structures in PDB are proteins

- **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?

  I was able to see 5 protease structures after searching HIV-1 in the PDB

PDB Format

Downloading the PDB File:

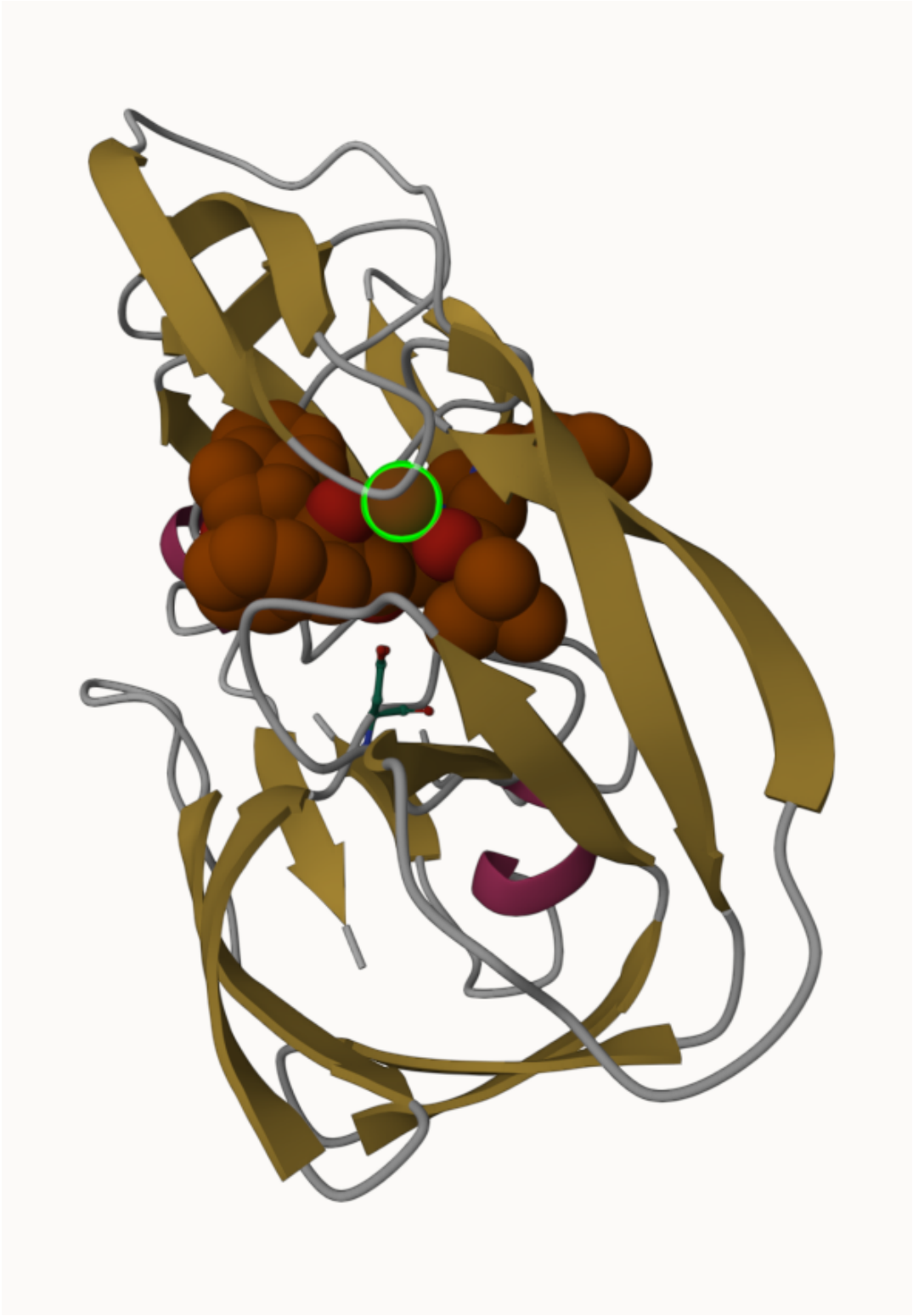
Visualizing the HIV-1 Protease Structure

- **Q4**: Water molecules normally have 3 atoms. Why do we see just one atom per water molecule in this structure?

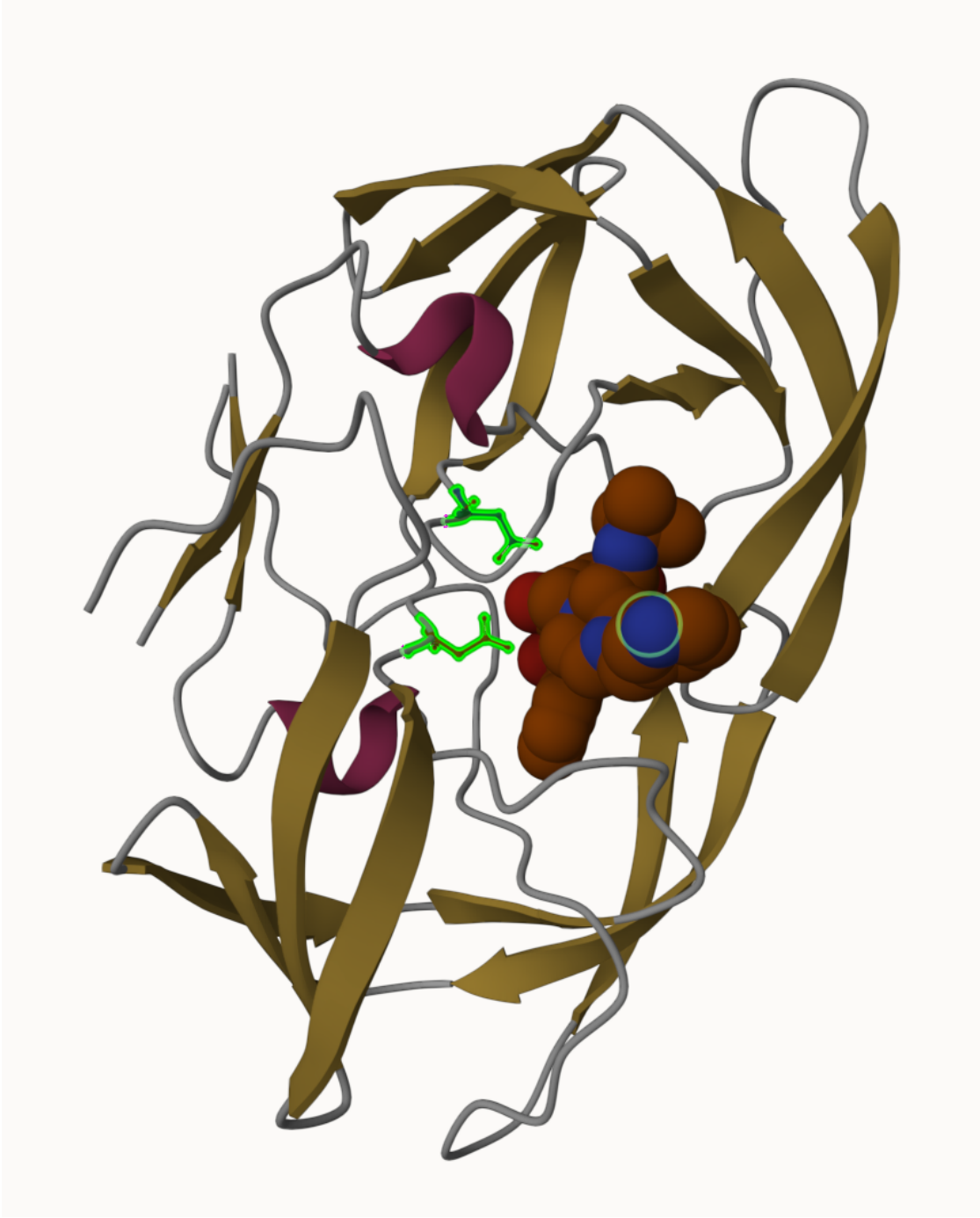
  Hydrogen is very small so we cannot see them (the resolution is not strong enough)

- **Q5**: There is a critical "conserved" water molecule in the binding site. Can you identify this water molecule? What residue number does this water molecule have

  This water molecule is HOH 308



- **Q6**: Generate and save a figure clearly showing the two distinct chains of HIV-protease along with the ligand. You might also consider showing the catalytic residues ASP 25 in each chain and the critical water (we recommend *"Ball & Stick"* for these side-chains). Add this figure to your Quarto document.



- **Q7:** How many amino acid residues are there in this pdb object?

  There are 99 amino acid residues in the PDB object which was found by looking at the sequence

Introduction to Bio3D

We first need to download Bio3D:

```
library(bio3d)
```

Calling PDB of HIV-1

```
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)

     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
```

- **Q7:** How many amino acid residues are there in this pdb object?

  198

- **Q8:** Name one of the two non-protein residues?

  HOH (127)

- **Q9:** How many protein chains are in this structure?

  2

Predicting Functional Motions of a Single Structure by NMA

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

```
 Note: Accessing on-line PDB file
  PDB has ALT records, taking A only, rm.alt=TRUE
```

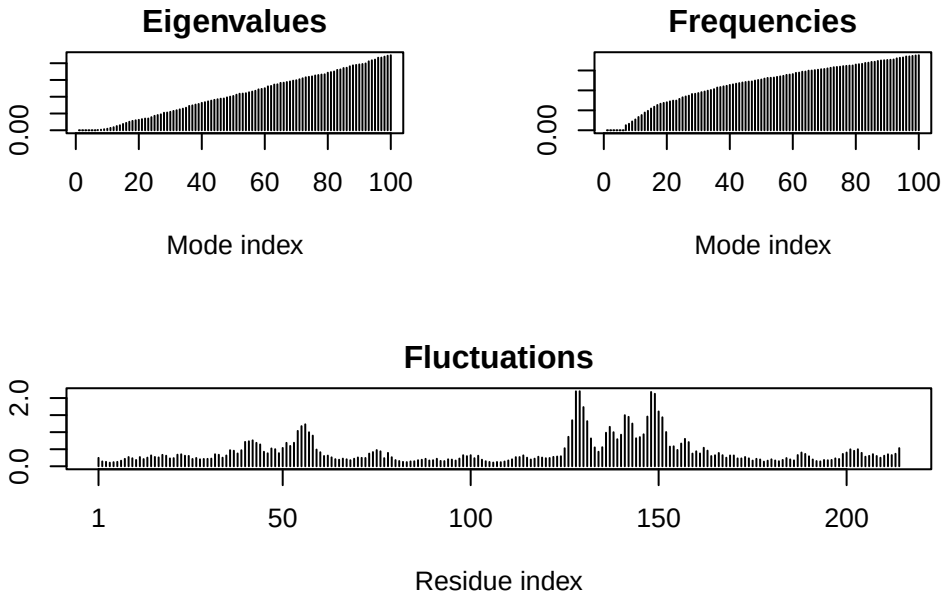
```
adk
```

```
 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:
      MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVT
      DELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDKI
      VGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIG
      YYSKEAEAGNTKYAKVDGTKPVAEVRADLEKILG

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

```
m <- nma(adk)
```

```
Building Hessian...        Done in 0.054 seconds.
Diagonalizing Hessian...   Done in 0.402 seconds.
```

```
plot(m)
```




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

Comparative Structure Analysis of Adenylate Kinase

```
#install.packages("bio3d")
#install.packages("devtools")
#install.packages("BiocManager")

#BiocManager::install("msa")
#devtools::install_bitbucket("Grantlab/bio3d-view")
```

- **Q10.** Which of the packages above is found only on BioConductor and not CRAN?

  msa

- **Q11.** Which of the above packages is not found on BioConductor or CRAN?:

  bio3d-view

- **Q12.** True or False? Functions from the devtools package can be used to install packages from GitHub and BitBucket?

  True!

Search and Retrieve ADK Structures

```
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   MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVT
             1        .         .         .         .         .         60

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

           121        .         .         .         .         .        180
pdb|1AKE|A   VGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIG
           121        .         .         .         .         .        180

           181        .         .         .   214
pdb|1AKE|A   YYSKEAEAGNTKYAKVDGTKPVAEVRADLEKILG
           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
```

- **Q13.** How many amino acids are in this sequence, i.e. how long is this sequence?

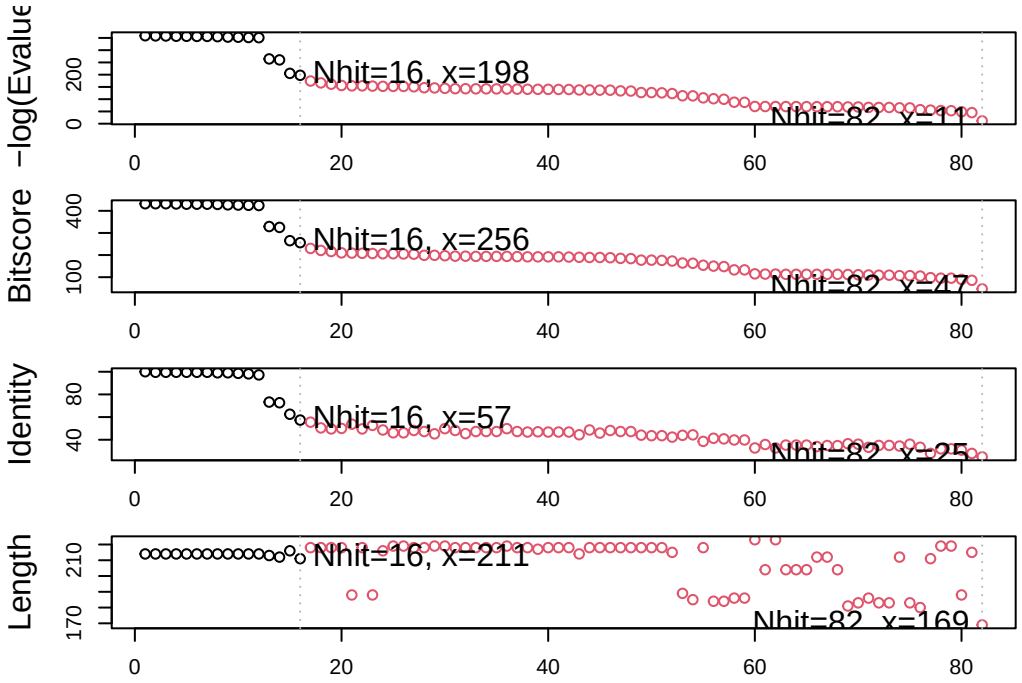
  There are 214 positions and therefore there are 214 amino acids in the sequence.

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

```
 Searching ... please wait (updates every 5 seconds) RID = 5FNR830X013
 ......
 Reporting 82 hits
```

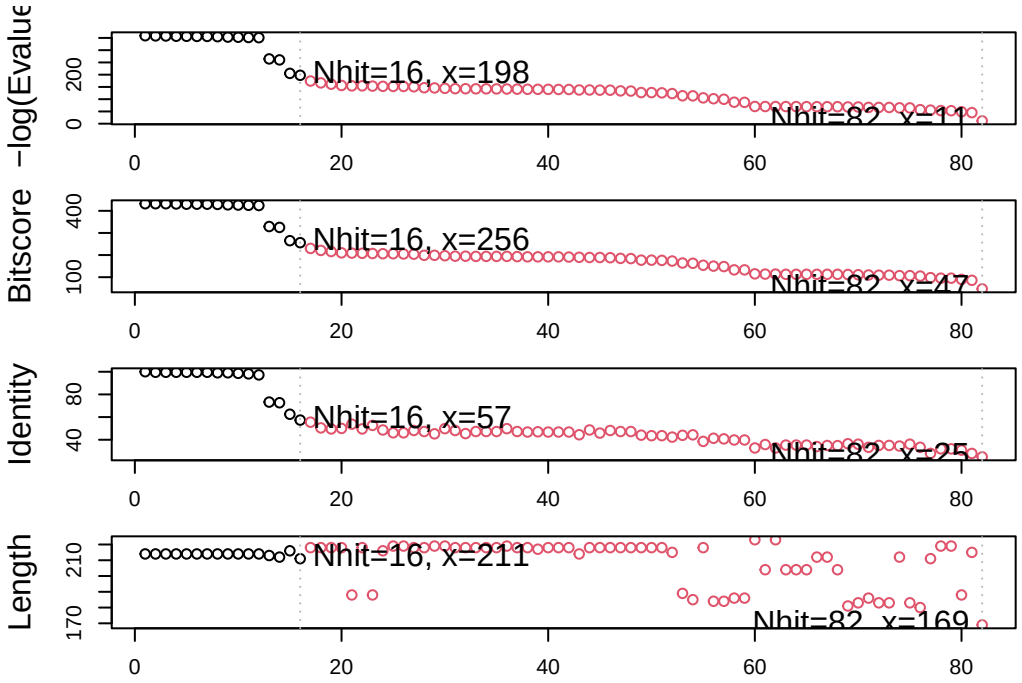
```
plot.blast(b)
```

```
  * Possible cutoff values:    197 11
            Yielding Nhits:    16 82

  * Chosen cutoff value of:    197
            Yielding Nhits:    16
```




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

```
* Possible cutoff values:    197 11
          Yielding Nhits:    16 82

* Chosen cutoff value of:    197
          Yielding Nhits:    16
```




```
head(hits$pdb.id)
```

```
[1] "1AKE_A" "4X8M_A" "6S36_A" "6RZE_A" "4X8H_A" "3HPR_A"
```

```
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/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


  |
  |                                                                      |   0%
  |
  |====                                                                  |   6%
  |
```

```
  |=========                                                             |  12%
  |
  |=============                                                         |  19%
  |
  |==================                                                    |  25%
  |
  |======================                                                |  31%
  |
  |==========================                                            |  38%
  |
  |==============================                                        |  44%
  |
  |==================================                                    |  50%
  |
  |======================================                                |  56%
  |
  |==========================================                            |  62%
  |
  |==============================================                        |  69%
  |
  |==================================================                    |  75%
  |
  |======================================================                |  81%
  |
  |==========================================================            |  88%
  |
  |==============================================================        |  94%
  |
  |======================================================================| 100%
```

Align and Superpose Structures

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

```
Reading PDB files:
pdbs/split_chain/1AKE_A.pdb
pdbs/split_chain/4X8M_A.pdb
pdbs/split_chain/6S36_A.pdb
pdbs/split_chain/6RZE_A.pdb
pdbs/split_chain/4X8H_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/4NP6_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/4X8M_A.pdb
pdb/seq: 3   name: pdbs/split_chain/6S36_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 4   name: pdbs/split_chain/6RZE_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 5   name: pdbs/split_chain/4X8H_A.pdb
pdb/seq: 6   name: pdbs/split_chain/3HPR_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 7   name: pdbs/split_chain/1E4V_A.pdb
pdb/seq: 8   name: pdbs/split_chain/5EJE_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 9   name: pdbs/split_chain/1E4Y_A.pdb
pdb/seq: 10   name: pdbs/split_chain/3X2S_A.pdb
pdb/seq: 11   name: pdbs/split_chain/6HAP_A.pdb
pdb/seq: 12   name: pdbs/split_chain/6HAM_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 13   name: pdbs/split_chain/4K46_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 14   name: pdbs/split_chain/4NP6_A.pdb
pdb/seq: 15   name: pdbs/split_chain/3GMT_A.pdb
```

```
pdb/seq: 16   name: pdbs/split_chain/4PZL_A.pdb
```

```
ids <- basename.pdb(pdbs$id)
#plot(pdbs, labels=ids)
```

Annotate Collected PDB Structures

```
anno <- pdb.annotate(ids)
unique(anno$source)
```

```
[1] "Escherichia coli"
[2] "Escherichia coli K-12"
[3] "Escherichia coli O139:H28 str. E24377A"
[4] "Escherichia coli str. K-12 substr. MDS42"
[5] "Photobacterium profundum"
[6] "Vibrio cholerae O1 biovar El Tor str. N16961"
[7] "Burkholderia pseudomallei 1710b"
[8] "Francisella tularensis subsp. tularensis SCHU S4"
```

```
anno
```

```
       structureId chainId macromoleculeType chainLength experimentalTechnique
1AKE_A        1AKE       A           Protein         214                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
3HPR_A        3HPR       A           Protein         214                X-ray
1E4V_A        1E4V       A           Protein         214                X-ray
5EJE_A        5EJE       A           Protein         214                X-ray
1E4Y_A        1E4Y       A           Protein         214                X-ray
3X2S_A        3X2S       A           Protein         214                X-ray
6HAP_A        6HAP       A           Protein         214                X-ray
6HAM_A        6HAM       A           Protein         214                X-ray
4K46_A        4K46       A           Protein         214                X-ray
4NP6_A        4NP6       A           Protein         217                X-ray
3GMT_A        3GMT       A           Protein         230                X-ray
4PZL_A        4PZL       A           Protein         242                X-ray
```

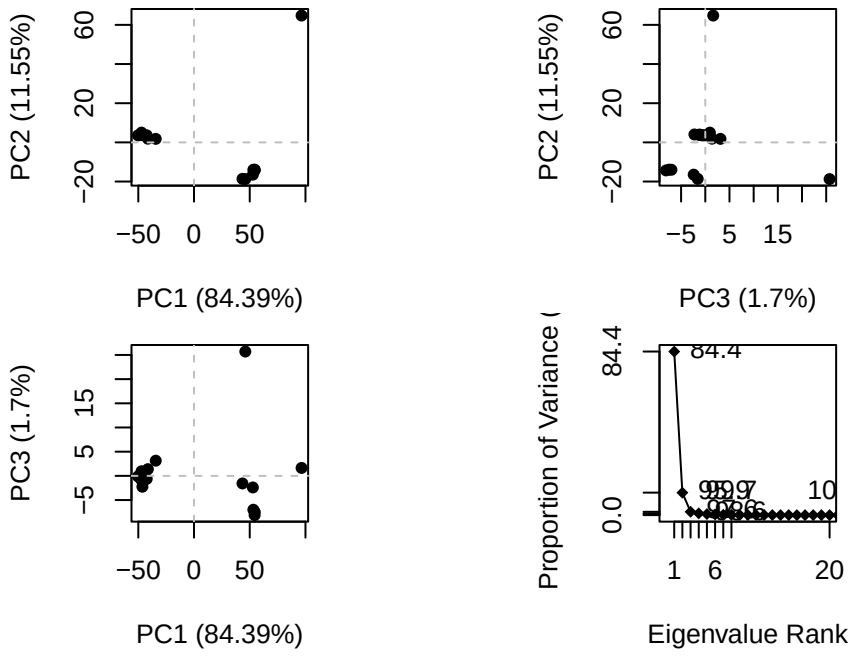
	resolution	scopDomain	pfam	ligandId	ligandName	source
1AKE_A	2.000	Adenylate kinase	Adenylate kinase (ADK)	AP5	BIS(ADENOSINE)-5'-PENTAPHOSPHATE	Escherichia coli
4X8M_A	2.600	<NA>	Adenylate kinase (ADK)	<NA>	<NA>	Escherichia coli
6S36_A	1.600	<NA>	Adenylate kinase (ADK)	MG (2),CL (3),NA	MAGNESIUM ION (2),CHLORIDE ION (3),SODIUM ION	Escherichia coli
6RZE_A	1.690	<NA>	Adenylate kinase (ADK)	NA (3),CL (2)	SODIUM ION (3),CHLORIDE ION (2)	Escherichia coli
4X8H_A	2.500	<NA>	Adenylate kinase (ADK)	<NA>	<NA>	Escherichia coli
3HPR_A	2.000	<NA>	Adenylate kinase (ADK)	AP5	BIS(ADENOSINE)-5'-PENTAPHOSPHATE	Escherichia coli K-12
1E4V_A	1.850	Adenylate kinase	Adenylate kinase (ADK)	AP5	BIS(ADENOSINE)-5'-PENTAPHOSPHATE	Escherichia coli
5EJE_A	1.900	<NA>	Adenylate kinase (ADK)	AP5,CO	BIS(ADENOSINE)-5'-PENTAPHOSPHATE,COBALT (II) ION	Escherichia coli O139:H28 str. E24377A
1E4Y_A	1.850	Adenylate kinase	Adenylate kinase (ADK)	AP5	BIS(ADENOSINE)-5'-PENTAPHOSPHATE	Escherichia coli
3X2S_A	2.800	<NA>	Adenylate kinase (ADK)	JPY (2),AP5,MG	N-(pyren-1-ylmethyl)acetamide (2),BIS(ADENOSINE)-5'-PENTAPHOSPHATE,MAGNESIUM ION	Escherichia coli
6HAP_A	2.700	<NA>	Adenylate kinase (ADK)	AP5	BIS(ADENOSINE)-5'-PENTAPHOSPHATE	Escherichia coli
6HAM_A	2.550	<NA>	Adenylate kinase (ADK)	AP5	BIS(ADENOSINE)-5'-PENTAPHOSPHATE	Escherichia coli
4K46_A	2.010	<NA>	Adenylate kinase (ADK)	AMP,PO4,ADP	ADENOSINE MONOPHOSPHATE,PHOSPHATE ION,ADENOSINE-5'-DIPHOSPHATE	Escherichia coli
4NP6_A	2.004	<NA>	Adenylate kinase (ADK)	<NA>	<NA>	Escherichia coli
3GMT_A	2.100	<NA>	Adenylate kinase (ADK)	SO4 (2)	SULFATE ION (2)	Escherichia coli
4PZL_A	2.100	<NA>	Adenylate kinase (ADK)	CA,FMT,GOL	CALCIUM ION,FORMIC ACID,GLYCEROL	Escherichia coli

```
1E4Y_A                                  Escherichia coli
3X2S_A          Escherichia coli str. K-12 substr. MDS42
6HAP_A            Escherichia coli O139:H28 str. E24377A
6HAM_A                              Escherichia coli K-12
4K46_A                           Photobacterium profundum
4NP6_A      Vibrio cholerae O1 biovar El Tor str. N16961
3GMT_A                    Burkholderia pseudomallei 1710b
4PZL_A Francisella tularensis subsp. tularensis SCHU S4

1AKE_A STRUCTURE OF THE COMPLEX BETWEEN ADENYLATE KINASE FROM ESCHERICHIA COLI AND THE INHIBI
4X8M_A
6S36_A
6RZE_A
4X8H_A
3HPR_A
1E4V_A
5EJE_A                                                                          Cryst
1E4Y_A
3X2S_A
6HAP_A
6HAM_A
4K46_A
4NP6_A
3GMT_A
4PZL_A                                                                      The cryst
                                                    citation rObserved   rFree
1AKE_A                 Muller, C.W., et al. J Mol Biol (1992)   0.19600      NA
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
3HPR_A  Schrank, T.P., et al. Proc Natl Acad Sci U S A (2009)   0.21000 0.24320
1E4V_A                   Muller, C.W., et al. Proteins (1993)   0.19600      NA
5EJE_A  Kovermann, M., et al. Proc Natl Acad Sci U S A (2017)   0.18890 0.23580
1E4Y_A                   Muller, C.W., et al. Proteins (1993)   0.17800      NA
3X2S_A                Fujii, A., et al. Bioconjug Chem (2015)   0.20700 0.25600
6HAP_A               Kantaev, R., et al. J Phys Chem B (2018)   0.22630 0.27760
6HAM_A               Kantaev, R., et al. J Phys Chem B (2018)   0.20511 0.24325
4K46_A                    Cho, Y.-J., et al. To be published   0.17000 0.22290
4NP6_A                       Kim, Y., et al. To be published   0.18800 0.22200
3GMT_A Buchko, G.W., et al. Biochem Biophys Res Commun (2010)   0.23800 0.29500
4PZL_A                       Tan, K., et al. To be published   0.19360 0.23680
        rWork spaceGroup
```

```
1AKE_A 0.19600  P 21 2 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
3HPR_A 0.20620  P 21 21 2
1E4V_A 0.19600  P 21 2 21
5EJE_A 0.18630  P 21 2 21
1E4Y_A 0.17800   P 1 21 1
3X2S_A 0.20700 P 21 21 21
6HAP_A 0.22370    I 2 2 2
6HAM_A 0.20311       P 43
4K46_A 0.16730 P 21 21 21
4NP6_A 0.18600       P 43
3GMT_A 0.23500   P 1 21 1
4PZL_A 0.19130       P 32
```

Principal Component Analysis

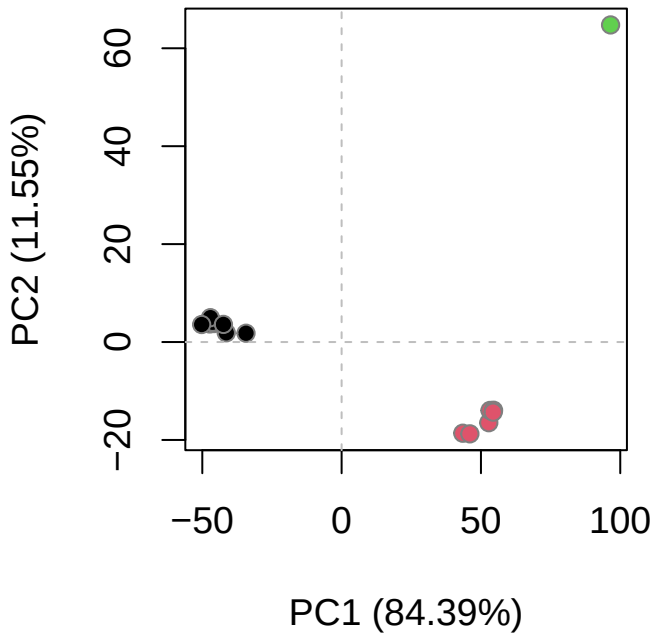
```
pc.xray <- pca(pdbs)
plot(pc.xray)
```



```
rd <- rmsd(pdbs)
```

```
Warning in rmsd(pdbs): No indices provided, using the 204 non NA positions
```

```
hc.rd <- hclust(dist(rd))
grps.rd <- cutree(hc.rd, k=3)
plot(pc.xray, 1:2, col="grey50", bg=grps.rd, pch=21, cex=1)
```

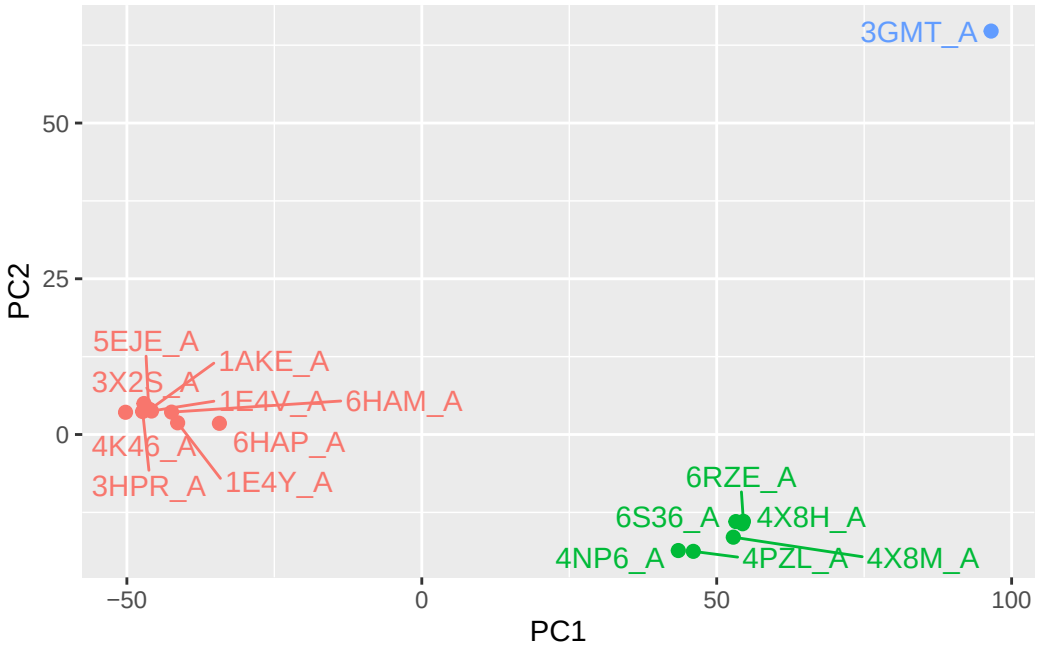


Optional Further Visualization

```
pc1 <- mktrj(pc.xray, pc=1, file="pc_1.pdb")
```

```
library(ggplot2)
library(ggrepel)

df <- data.frame(PC1=pc.xray$z[,1],
                 PC2=pc.xray$z[,2],
                 col=as.factor(grps.rd),
```

```
                  ids=ids)

p <- ggplot(df) +
  aes(PC1, PC2, col=col, label=ids) +
  geom_point(size=2) +
  geom_text_repel(max.overlaps = 20) +
  theme(legend.position = "none")
p
```



Normal Mode Analysis

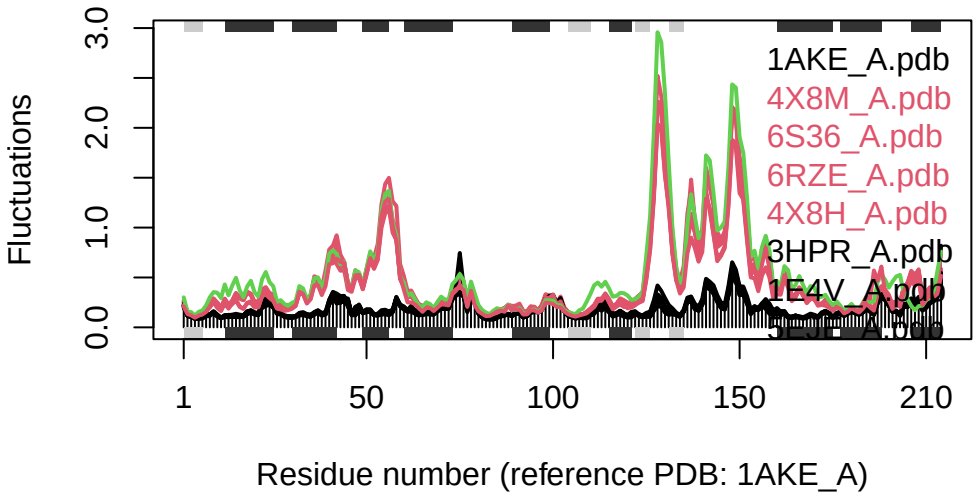
```
modes <- nma(pdbs)
```

```
Details of Scheduled Calculation:
  ... 16 input structures
  ... storing 606 eigenvectors for each structure
  ... dimension of x$U.subspace: ( 612x606x16 )
  ... coordinate superposition prior to NM calculation
```

```
  ... aligned eigenvectors (gap containing positions removed)
  ... estimated memory usage of final 'eNMA' object: 45.4 Mb


  |
  |                                                                      |   0%
  |
  |====                                                                  |   6%
  |
  |=========                                                             |  12%
  |
  |=============                                                         |  19%
  |
  |==================                                                    |  25%
  |
  |======================                                                |  31%
  |
  |==========================                                            |  38%
  |
  |==============================                                        |  44%
  |
  |==================================                                    |  50%
  |
  |======================================                                |  56%
  |
  |==========================================                            |  62%
  |
  |==============================================                        |  69%
  |
  |==================================================                    |  75%
  |
  |======================================================                |  81%
  |
  |==========================================================            |  88%
  |
  |================================================================      |  94%
  |
  |======================================================================| 100%
```

```
plot(modes, pdbs, col=grps.rd)
```

```
Extracting SSE from pdbs$sse attribute
```



- **Q14.** What do you note about this plot? Are the black and colored lines similar or different? Where do you think they differ most and why?

  The colored and black lines are very different in this plot. They most likely differ since they have more fluctuations in those proteins compared to the reference PDB