Class9

Hetian Su

Checking PDB database statistics

Q1.

Approximately 92.8% structures are solved by X-ray and EM.

Q2

Approximately 87% are protein only.

Q3

There are 1264 structures of HIV protease.

Viewing PDB Structure with Molstar

Reading and working with structures in R

```r
# use the bio3d package
# install.packages('bio3d')

library(bio3d)

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

Note: Accessing on-line PDB file

```r
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
```

```r
# the atom attribute contains important information of a pdb file
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>
```

```
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
```
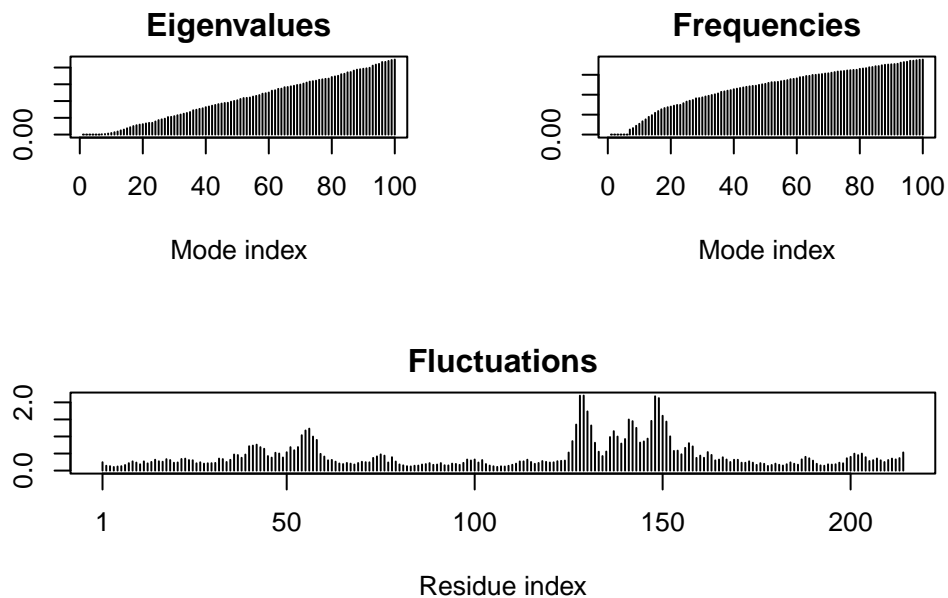
Predicting functional motions using Normal Mode Analysis (NMA), revealing the flexible parts of the structure.

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

```
Building Hessian...      Done in 0.02 seconds.
Diagonalizing Hessian... Done in 0.41 seconds.
```

```
plot(m)
```

Eigenvalues

Frequencies

Fluctuations

```
mktrj(m, file='adk_nma.pdb')
```

How to analyze variant forms of the structure:

retrieve the sequence, search against the pdb database, retrieve the sequences of
the hits, load files in Molstar.

```
aa <- get.seq('1ake_a')
```

```
Warning in get.seq("1ake_a"): Removing existing file: seqs.fasta
```

```
Fetching... Please wait. Done.
```

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

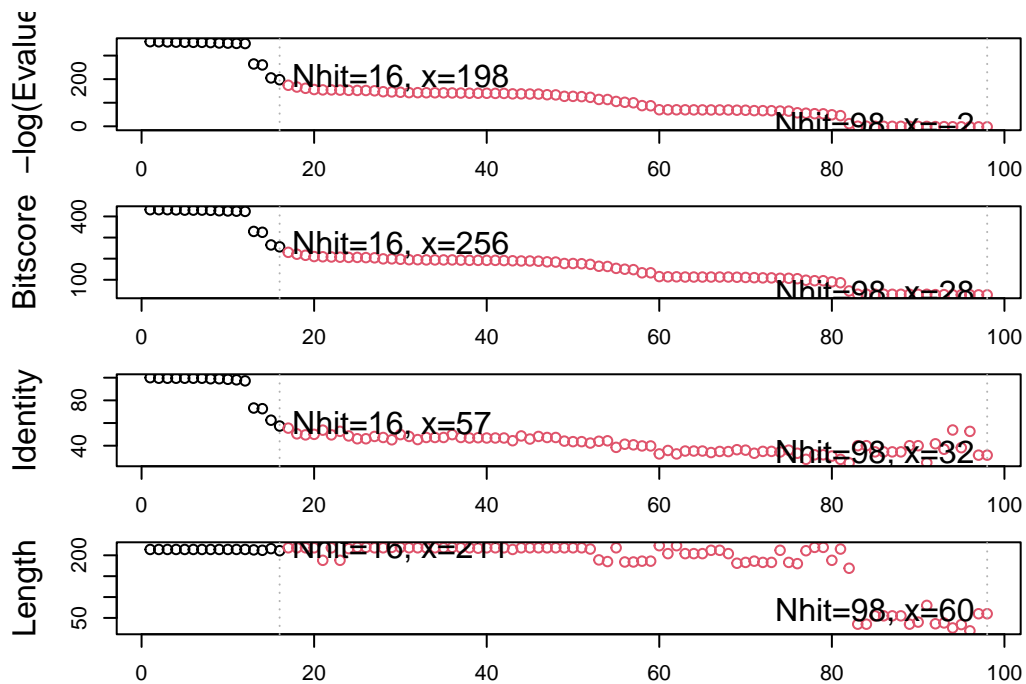Searching ... please wait (updates every 5 seconds) RID = NKFZCEA5016
.
Reporting 98 hits

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

```
* Possible cutoff values:    197 -3
         Yielding Nhits:    16 98

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



```
# insepct the hits with low E-values
pdb.annotate(hits$pdb.id)
```

      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
1AKE_A       2.000 Adenylate kinase Adenylate kinase, active site lid (ADK_lid)
4X8M_A       2.600              <NA> Adenylate kinase, active site lid (ADK_lid)
6S36_A       1.600              <NA> Adenylate kinase, active site lid (ADK_lid)
6RZE_A       1.690              <NA> Adenylate kinase, active site lid (ADK_lid)
4X8H_A       2.500              <NA> Adenylate kinase, active site lid (ADK_lid)
3HPR_A       2.000              <NA> Adenylate kinase, active site lid (ADK_lid)
1E4V_A       1.850 Adenylate kinase Adenylate kinase, active site lid (ADK_lid)
5EJE_A       1.900              <NA> Adenylate kinase, active site lid (ADK_lid)
1E4Y_A       1.850 Adenylate kinase Adenylate kinase, active site lid (ADK_lid)
3X2S_A       2.800              <NA> Adenylate kinase, active site lid (ADK_lid)
6HAP_A       2.700              <NA> Adenylate kinase, active site lid (ADK_lid)
6HAM_A       2.550              <NA> Adenylate kinase, active site lid (ADK_lid)
4K46_A       2.010              <NA> Adenylate kinase, active site lid (ADK_lid)
4NP6_A       2.004              <NA> Adenylate kinase, active site lid (ADK_lid)
3GMT_A       2.100              <NA> Adenylate kinase, active site lid (ADK_lid)
4PZL_A       2.100              <NA> Adenylate kinase, active site lid (ADK_lid)
            ligandId
1AKE_A            AP5
4X8M_A           <NA>
6S36_A CL (3),NA,MG (2)
6RZE_A    NA (3),CL (2)
4X8H_A           <NA>
3HPR_A            AP5
1E4V_A            AP5
5EJE_A         AP5,CO
1E4Y_A            AP5
```

```
3X2S_A    JPY (2),AP5,MG
6HAP_A             AP5
6HAM_A             AP5
4K46_A     ADP,AMP,PO4
4NP6_A            <NA>
3GMT_A         SO4 (2)
4PZL_A      CA,FMT,GOL
```

```
                                                              ligandName
1AKE_A                                 BIS(ADENOSINE)-5'-PENTAPHOSPHATE
4X8M_A                                                             <NA>
6S36_A                  CHLORIDE ION (3),SODIUM ION,MAGNESIUM ION (2)
6RZE_A                            SODIUM ION (3),CHLORIDE ION (2)
4X8H_A                                                             <NA>
3HPR_A                                 BIS(ADENOSINE)-5'-PENTAPHOSPHATE
1E4V_A                                 BIS(ADENOSINE)-5'-PENTAPHOSPHATE
5EJE_A              BIS(ADENOSINE)-5'-PENTAPHOSPHATE,COBALT (II) ION
1E4Y_A                                 BIS(ADENOSINE)-5'-PENTAPHOSPHATE
3X2S_A N-(pyren-1-ylmethyl)acetamide (2),BIS(ADENOSINE)-5'-PENTAPHOSPHATE,MAGNESIUM ION
6HAP_A                                 BIS(ADENOSINE)-5'-PENTAPHOSPHATE
6HAM_A                                 BIS(ADENOSINE)-5'-PENTAPHOSPHATE
4K46_A          ADENOSINE-5'-DIPHOSPHATE,ADENOSINE MONOPHOSPHATE,PHOSPHATE ION
4NP6_A                                                             <NA>
3GMT_A                                                SULFATE ION (2)
4PZL_A                                CALCIUM ION,FORMIC ACID,GLYCEROL
```

```
                                       source
1AKE_A                         Escherichia coli
4X8M_A                         Escherichia coli
6S36_A                         Escherichia coli
6RZE_A                         Escherichia coli
4X8H_A                         Escherichia coli
3HPR_A                    Escherichia coli K-12
1E4V_A                         Escherichia coli
5EJE_A         Escherichia coli O139:H28 str. E24377A
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
```

```r
files <- get.pdb(hits$pdb.id, path = 'pdbs', split = T, gzip = T)
```

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

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

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

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

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

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

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

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

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

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

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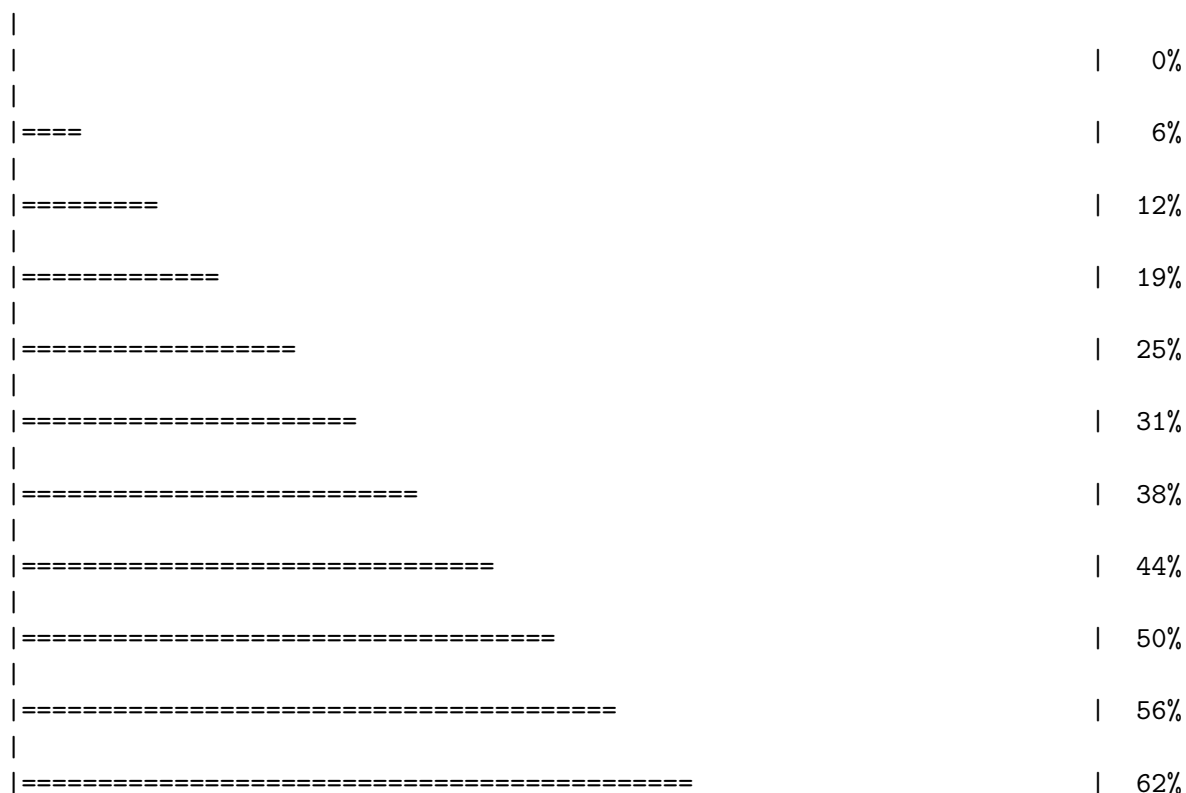6HAP.pdb exists. Skipping download


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


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


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


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


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


  |
  |                                                            |   0%
  |
  |====                                                        |   6%
  |
  |========                                                    |  12%
  |
  |============                                                |  19%
  |
  |=================                                           |  25%
  |
  |=====================                                       |  31%
  |
  |=========================                                   |  38%
  |
  |=============================                               |  44%
  |
  |=================================                           |  50%
  |
  |=====================================                       |  56%
  |
  |=========================================                   |  62%
```

```
  |
  |==============================================     |  69%
  |
  |=================================================   |  75%
  |
  |======================================================  |  81%
  |
  |=========================================================  |  88%
  |
  |==============================================================  |  94%
  |
  |================================================================| 100%
```

Multiple sequence align and superpose these structures

```
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
```

12

```
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
```

pdbs

```
                                  1         .         .         .         40
[Truncated_Name:1]1AKE_A.pdb      ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:2]4X8M_A.pdb      ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:3]6S36_A.pdb      ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:4]6RZE_A.pdb      ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:5]4X8H_A.pdb      ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:6]3HPR_A.pdb      ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:7]1E4V_A.pdb      ----------MRIILLGAPVAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:8]5EJE_A.pdb      ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:9]1E4Y_A.pdb      ----------MRIILLGALVAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:10]3X2S_A.pdb     ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:11]6HAP_A.pdb     ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:12]6HAM_A.pdb     ----------MRIILLGAPGAGKGTQAQFIMEKYGIPQIS
[Truncated_Name:13]4K46_A.pdb     ----------MRIILLGAPGAGKGTQAQFIMAKFGIPQIS
```

```
[Truncated_Name:14]4NP6_A.pdb    --------NAMRIILLGAPGAGKGTQAQFIMEKFGIPQIS
[Truncated_Name:15]3GMT_A.pdb    ----------MRLILLGAPGAGKGTQANFIKEKFGIPQIS
[Truncated_Name:16]4PZL_A.pdb    TENLYFQSNAMRIILLGAPGAGKGTQAKIIEQKYNIAHIS
                                         **^*****  *******  *  *^ *  **
                                 1         .         .         .        40


                                 41        .         .         .        80
[Truncated_Name:1]1AKE_A.pdb     TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:2]4X8M_A.pdb     TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:3]6S36_A.pdb     TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:4]6RZE_A.pdb     TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:5]4X8H_A.pdb     TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:6]3HPR_A.pdb     TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:7]1E4V_A.pdb     TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:8]5EJE_A.pdb     TGDMLRAAVKSGSELGKQAKDIMDACKLVTDELVIALVKE
[Truncated_Name:9]1E4Y_A.pdb     TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:10]3X2S_A.pdb    TGDMLRAAVKSGSELGKQAKDIMDCGKLVTDELVIALVKE
[Truncated_Name:11]6HAP_A.pdb    TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVRE
[Truncated_Name:12]6HAM_A.pdb    TGDMLRAAIKSGSELGKQAKDIMDAGKLVTDEIIIALVKE
[Truncated_Name:13]4K46_A.pdb    TGDMLRAAIKAGTELGKQAKSVIDAGQLVSDDIILGLVKE
[Truncated_Name:14]4NP6_A.pdb    TGDMLRAAIKAGTELGKQAKAVIDAGQLVSDDIILGLIKE
[Truncated_Name:15]3GMT_A.pdb    TGDMLRAAVKAGTPLGVEAKTYMDEGKLVPDSLIIGLVKE
[Truncated_Name:16]4PZL_A.pdb    TGDMIRETIKSGSALGQELKKVLDAGELVSDEFIIKIVKD
                                 ****^*   ^* *^ **    *   ^*    ** *   ^^  ^^^^
                                 41        .         .         .        80


                                 81        .         .         .       120
[Truncated_Name:1]1AKE_A.pdb     RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:2]4X8M_A.pdb     RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:3]6S36_A.pdb     RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:4]6RZE_A.pdb     RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:5]4X8H_A.pdb     RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:6]3HPR_A.pdb     RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:7]1E4V_A.pdb     RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:8]5EJE_A.pdb     RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:9]1E4Y_A.pdb     RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:10]3X2S_A.pdb    RIAQEDSRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:11]6HAP_A.pdb    RICQEDSRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:12]6HAM_A.pdb    RICQEDSRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
[Truncated_Name:13]4K46_A.pdb    RIAQDDCAKGFLLDGFPRTIPQADGLKEVGVVVDYVIEFD
[Truncated_Name:14]4NP6_A.pdb    RIAQADCEKGFLLDGFPRTIPQADGLKEMGINVDYVIEFD
[Truncated_Name:15]3GMT_A.pdb    RLKEADCANGYLFDGFPRTIAQADAMKEAGVAIDYVLEID
[Truncated_Name:16]4PZL_A.pdb    RISKNDCNNGFLLDGVPRTIPQAQELDKLGVNIDYIVEVD
```

```
                                    *^     *    *^* ** **** **   ^    *^ ^**^^* *
                                   81          .          .          .          120

                                   121         .          .          .          160
[Truncated_Name:1]1AKE_A.pdb       VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:2]4X8M_A.pdb       VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:3]6S36_A.pdb       VPDELIVDKIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:4]6RZE_A.pdb       VPDELIVDAIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:5]4X8H_A.pdb       VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:6]3HPR_A.pdb       VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDGTG
[Truncated_Name:7]1E4V_A.pdb       VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:8]5EJE_A.pdb       VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:9]1E4Y_A.pdb       VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:10]3X2S_A.pdb      VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:11]6HAP_A.pdb      VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:12]6HAM_A.pdb      VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:13]4K46_A.pdb      VADSVIVERMAGRRAHLASGRTYHNVYNPPKVEGKDDVTG
[Truncated_Name:14]4NP6_A.pdb      VADDVIVERMAGRRAHLPSGRTYHVVYNPPKVEGKDDVTG
[Truncated_Name:15]3GMT_A.pdb      VPFSEIIERMSGRRTHPASGRTYHVKFNPPKVEGKDDVTG
[Truncated_Name:16]4PZL_A.pdb      VADNLLIERITGRRIHPASGRTYHTKFNPPKVADKDDVTG
                                   *     ^^^ ^ *** *  *** **  ^*****  *** **
                                   121         .          .          .          160

                                   161         .          .          .          200
[Truncated_Name:1]1AKE_A.pdb       EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:2]4X8M_A.pdb       EELTTRKDDQEETVRKRLVEWHQMTAPLIGYYSKEAEAGN
[Truncated_Name:3]6S36_A.pdb       EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:4]6RZE_A.pdb       EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:5]4X8H_A.pdb       EELTTRKDDQEETVRKRLVEYHQMTAALIGYYSKEAEAGN
[Truncated_Name:6]3HPR_A.pdb       EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:7]1E4V_A.pdb       EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:8]5EJE_A.pdb       EELTTRKDDQEECVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:9]1E4Y_A.pdb       EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:10]3X2S_A.pdb      EELTTRKDDQEETVRKRLCEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:11]6HAP_A.pdb      EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:12]6HAM_A.pdb      EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN
[Truncated_Name:13]4K46_A.pdb      EDLVIREDDKEETVLARLGVYHNQTAPLIAYYGKEAEAGN
[Truncated_Name:14]4NP6_A.pdb      EDLVIREDDKEETVRARLNVYHTQTAPLIEYYGKEAAAGK
[Truncated_Name:15]3GMT_A.pdb      EPLVQRDDDKEETVKKRLDVYEAQTKPLITYYGDWARRGA
[Truncated_Name:16]4PZL_A.pdb      EPLITRTDDNEDTVKQRLSVYHAQTAKLIDFYRNFSSTNT
                                   * *   * ** *^ *  **   ^    *  ** ^*
                                   161         .          .          .          200



                                            15
```

```
                                      201          .          .         227
[Truncated_Name:1]1AKE_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:2]4X8M_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:3]6S36_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:4]6RZE_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:5]4X8H_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:6]3HPR_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:7]1E4V_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:8]5EJE_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:9]1E4Y_A.pdb    T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:10]3X2S_A.pdb   T--KYAKVDGTKPVAEVRADLEKILG-
[Truncated_Name:11]6HAP_A.pdb   T--KYAKVDGTKPVCEVRADLEKILG-
[Truncated_Name:12]6HAM_A.pdb   T--KYAKVDGTKPVCEVRADLEKILG-
[Truncated_Name:13]4K46_A.pdb   T--QYLKFDGTKAVAEVSAELEKALA-
[Truncated_Name:14]4NP6_A.pdb   T--QYLKFDGTKQVSEVSADIAKALA-
[Truncated_Name:15]3GMT_A.pdb   E-------NGLKAPA-----YRKISG-
[Truncated_Name:16]4PZL_A.pdb   KIPKYIKINGDQAVEKVSQDIFDQLNK
                                              *
                                      201          .          .         227

Call:
  pdbaln(files = files, fit = T, exefile = "msa")
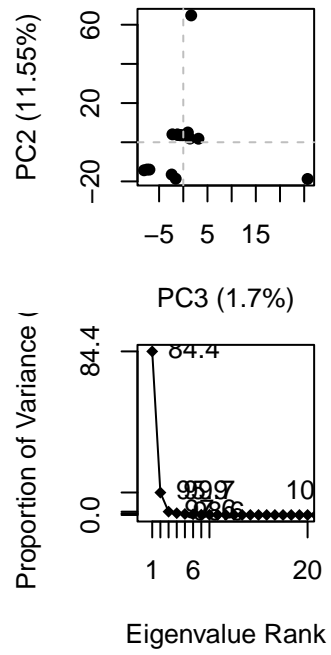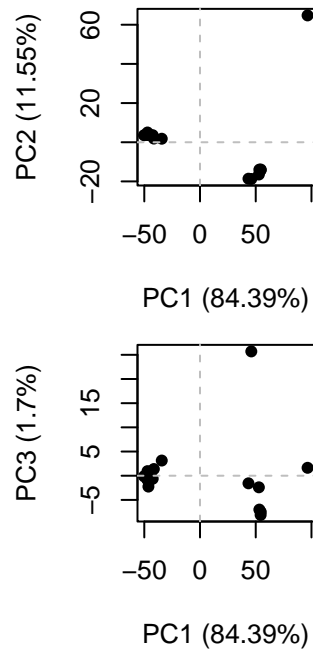
Class:
  pdbs, fasta

Alignment dimensions:
  16 sequence rows; 227 position columns (204 non-gap, 23 gap)

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

PCA the structures

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

```
mktrj(pca.xray, pc=1, file='pc_1.pdb')
```