AlphaFold

Kelly Isbell (A59019188)

```
#0.073% of the sequences on Uniprot are have structures in the PDB
(183201/251600768)*100
```

```
[1] 0.07281417
```

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

```
                          X.ray     EM    NMR Multiple.methods Neutron Other
Protein (only)          158,844 11,759 12,296              197      73    32
Protein/Oligosaccharide   9,260  2,054     34                8       1     0
Protein/NA                8,307  3,667    284                7       0     0
Nucleic acid (only)       2,730    113  1,467               13       3     1
Other                       164      9     32                0       0     0
Oligosaccharide (only)       11      0      6                1       0     4
                          Total
Protein (only)          183,201
Protein/Oligosaccharide  11,357
Protein/NA               12,265
Nucleic acid (only)       4,327
Other                       205
Oligosaccharide (only)       22
```

```
string <- c("10", "100", 1, "1,000")
as.numeric(string) + 1
```

```
Warning: NAs introduced by coercion
```

```
[1]  11 101   2  NA
```

```
x <- string
remove.comma <- function(x){
  as.numeric(gsub(",", "", x))
  }
pdbstats <- apply(stats, 2, remove.comma)
pdbstats
```

```
     X.ray    EM   NMR Multiple.methods Neutron Other  Total
[1,] 158844 11759 12296              197      73    32 183201
[2,]   9260  2054    34                8       1     0  11357
[3,]   8307  3667   284                7       0     0  12265
[4,]   2730   113  1467               13       3     1   4327
[5,]    164     9    32                0       0     0    205
[6,]     11     0     6                1       0     4     22
```

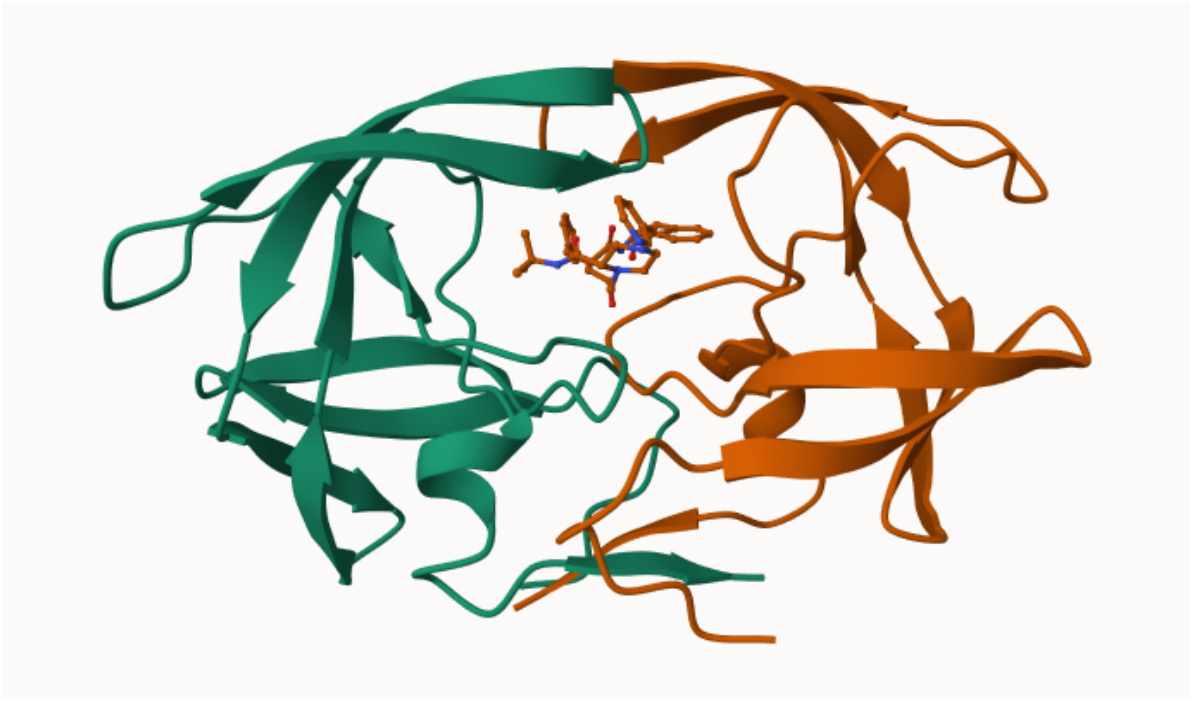
```
rownames(pdbstats) <- rownames(stats)
pdbstats
```

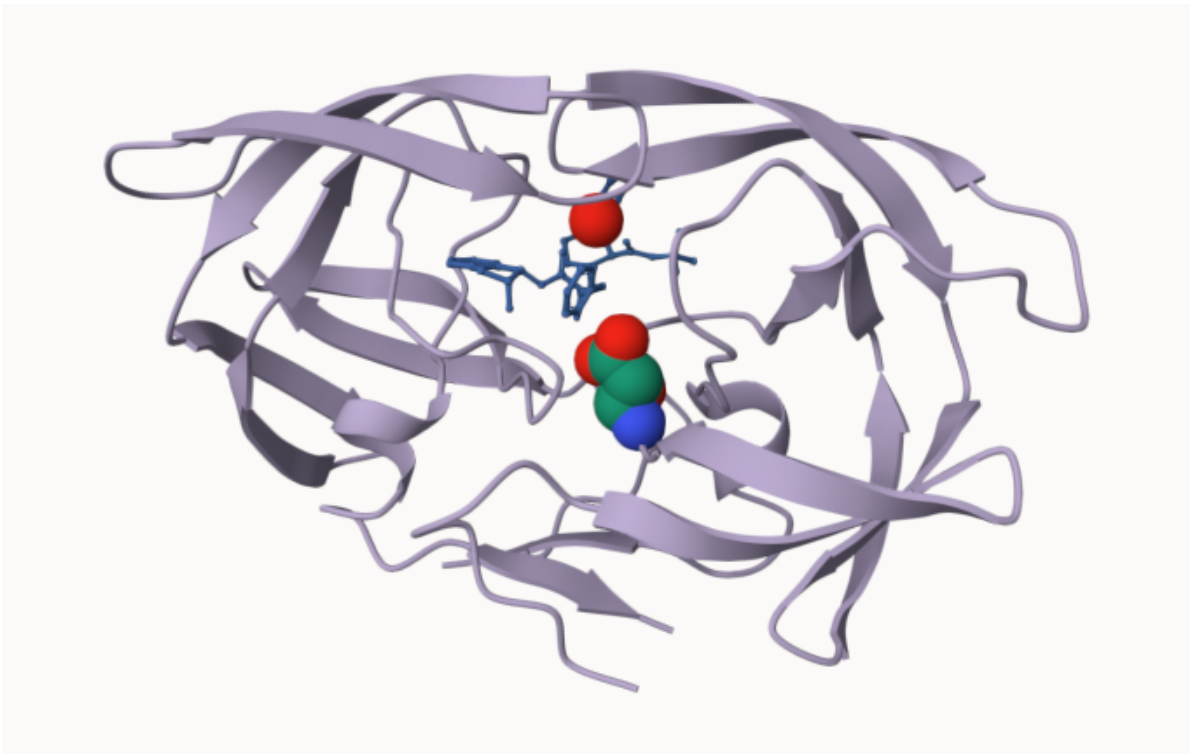
```
                        X.ray    EM   NMR Multiple.methods Neutron Other
Protein (only)         158844 11759 12296              197      73    32
Protein/Oligosaccharide  9260  2054    34                8       1     0
Protein/NA               8307  3667   284                7       0     0
Nucleic acid (only)      2730   113  1467               13       3     1
Other                     164     9    32                0       0     0
Oligosaccharide (only)     11     0     6                1       0     4
                        Total
Protein (only)         183201
Protein/Oligosaccharide 11357
Protein/NA              12265
Nucleic acid (only)      4327
Other                     205
Oligosaccharide (only)     22
```

```
totals <- apply(pdbstats, 2, sum)
round(totals/totals["Total"] *100, 2)
```

```
          X.ray              EM             NMR Multiple.methods
          84.83            8.33            6.68            0.11
        Neutron           Other           Total
           0.04            0.02          100.00
```

```
#Q1 84.83% are solved by X-ray crystallography, 8.33% are solved by EM
#Q2
#Q3
```





```
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)

    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
```

```
attributes (pdb)
```

```
$names
[1] "atom"   "xyz"    "seqres" "helix"  "sheet"  "calpha" "remark" "call"

$class
[1] "pdb" "sse"
```

```
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>
```

```
head(pdb$atom$resid)
```

```
[1] "PRO" "PRO" "PRO" "PRO" "PRO" "PRO"
```

```
aa321(pdb$atom$resid[pdb$calpha])
```

```
  [1] "P" "Q" "I" "T" "L" "W" "Q" "R" "P" "L" "V" "T" "I" "K" "I" "G" "G" "Q"
 [19] "L" "K" "E" "A" "L" "L" "D" "T" "G" "A" "D" "D" "T" "V" "L" "E" "E" "M"
 [37] "S" "L" "P" "G" "R" "W" "K" "P" "K" "M" "I" "G" "G" "I" "G" "G" "F" "I"
 [55] "K" "V" "R" "Q" "Y" "D" "Q" "I" "L" "I" "E" "I" "C" "G" "H" "K" "A" "I"
 [73] "G" "T" "V" "L" "V" "G" "P" "T" "P" "V" "N" "I" "I" "G" "R" "N" "L" "L"
 [91] "T" "Q" "I" "G" "C" "T" "L" "N" "F" "P" "Q" "I" "T" "L" "W" "Q" "R" "P"
[109] "L" "V" "T" "I" "K" "I" "G" "G" "Q" "L" "K" "E" "A" "L" "L" "D" "T" "G"
[127] "A" "D" "D" "T" "V" "L" "E" "E" "M" "S" "L" "P" "G" "R" "W" "K" "P" "K"
[145] "M" "I" "G" "G" "I" "G" "G" "F" "I" "K" "V" "R" "Q" "Y" "D" "Q" "I" "L"
[163] "I" "E" "I" "C" "G" "H" "K" "A" "I" "G" "T" "V" "L" "V" "G" "P" "T" "P"
[181] "V" "N" "I" "I" "G" "R" "N" "L" "L" "T" "Q" "I" "G" "C" "T" "L" "N" "F"
```

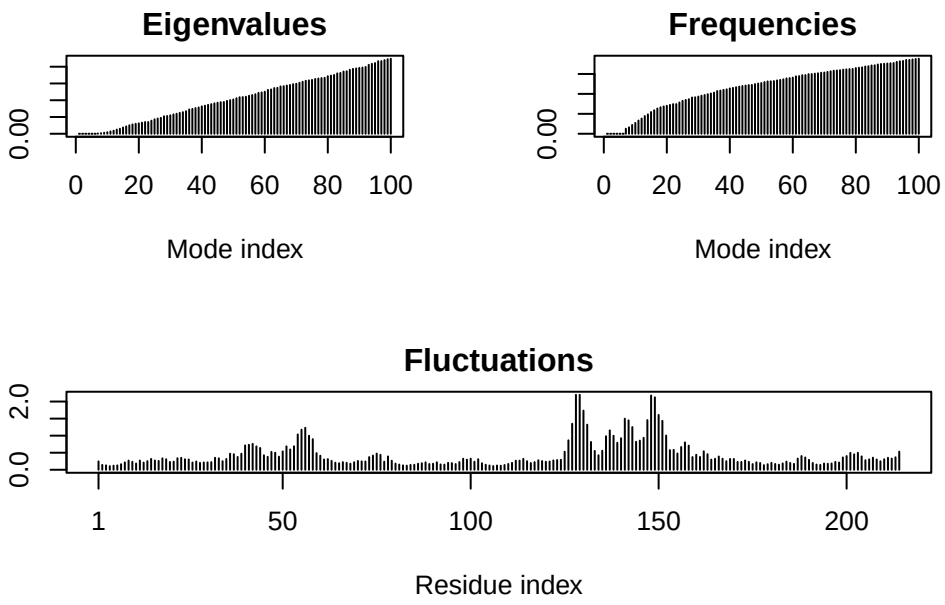
```
adk <- read.pdb("6s36")
```

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

```
modes <- nma(adk)
```

```
Building Hessian...        Done in 0.02 seconds.
Diagonalizing Hessian...   Done in 0.26 seconds.
```

```
plot(modes)
```




```
mktrj(modes, pdb=adk, file = "modes.pdb")
```