lab 10: structural bioinformatics

Elsa Chen (A16632961)

Table of contents

Introduction to the RCSB Protein Data Bank (PDB)

PDB is the main repository of biomolecular structure info <www.rcsb.org>

PDB statistics

```
PDBstats <- read.csv("Data Export Summary.csv")
PDBstats
```

```
           Molecular.Type   X.ray     EM    NMR Multiple.methods Neutron Other
1          Protein (only) 163,468 13,582 12,390              204      74    32
2 Protein/Oligosaccharide   9,437  2,287     34                8       2     0
3              Protein/NA   8,482  4,181    286                7       0     0
4     Nucleic acid (only)   2,800    132  1,488               14       3     1
5                   Other     164      9     33                0       0     0
6  Oligosaccharide (only)      11      0      6                1       0     4
    Total
1 189,750
2  11,768
3  12,956
```

```
4   4,438
5     206
6      22
```

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

need to get rid of commas in columns

```
decomma <- function(x){
  as.numeric(gsub(",", "", x))
}
```

```
PDBstats[-1] <- apply(PDBstats[-1], 2, decomma)
PDBstats
```

```
           Molecular.Type  X.ray    EM   NMR Multiple.methods Neutron Other
1          Protein (only) 163468 13582 12390               204      74    32
2 Protein/Oligosaccharide   9437  2287    34                 8       2     0
3              Protein/NA   8482  4181   286                 7       0     0
4     Nucleic acid (only)   2800   132  1488                14       3     1
5                   Other    164     9    33                 0       0     0
6  Oligosaccharide (only)     11     0     6                 1       0     4
   Total
1 189750
2  11768
3  12956
4   4438
5    206
6     22
```

```
(sum(PDBstats$X.ray))/(sum(PDBstats$Total))*100
```

```
[1] 84.12978
```

```
(sum(PDBstats$EM))/(sum(PDBstats$Total))*100
```

```
[1] 9.213745
```

84.12978% of structures in the PDB are solved by X-Ray, 9.2137446% of structures in the PDB are solved by EM

Visualizing the HIV-1 protease structure

Using Mol* <molstar.org/viewer/> to get to know HIV-Pr

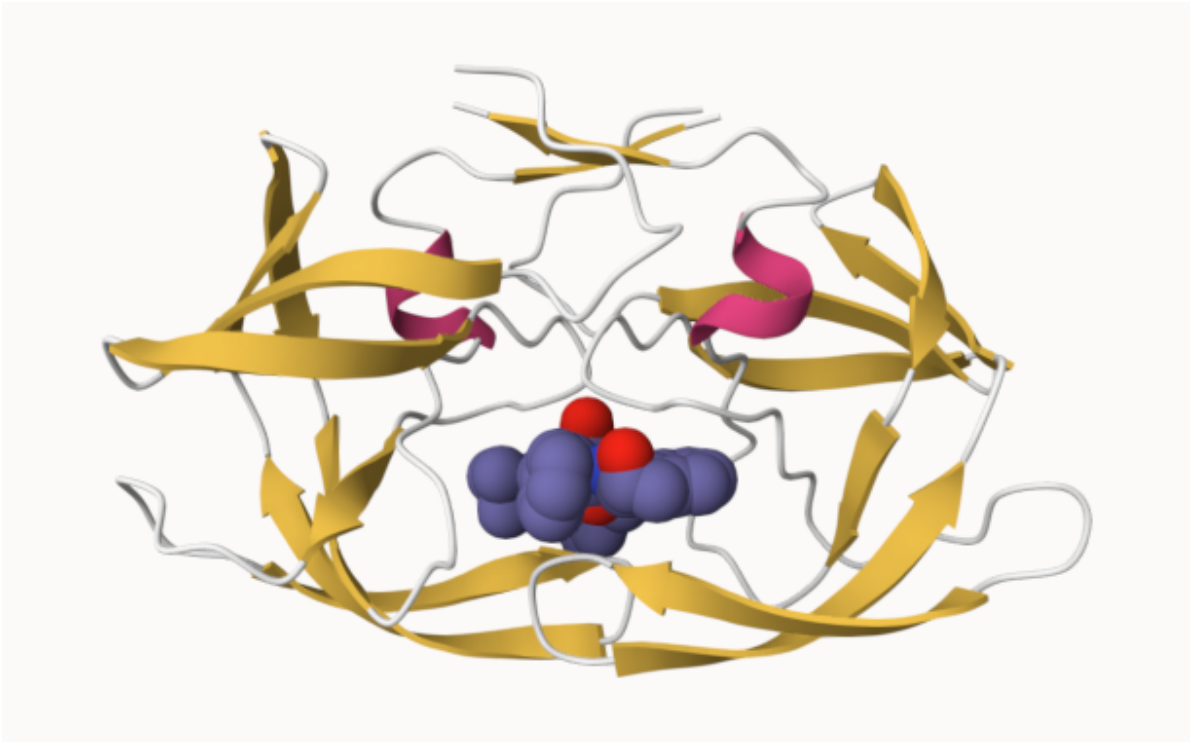


Figure 1: HIV-Pr molecular image

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

Because the resolution of this experiment isn't high enough to see the very small hydrogens

Bio3D in R

```
library(bio3d)
```

```
Warning: package 'bio3d' was built under R version 4.3.3
```

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

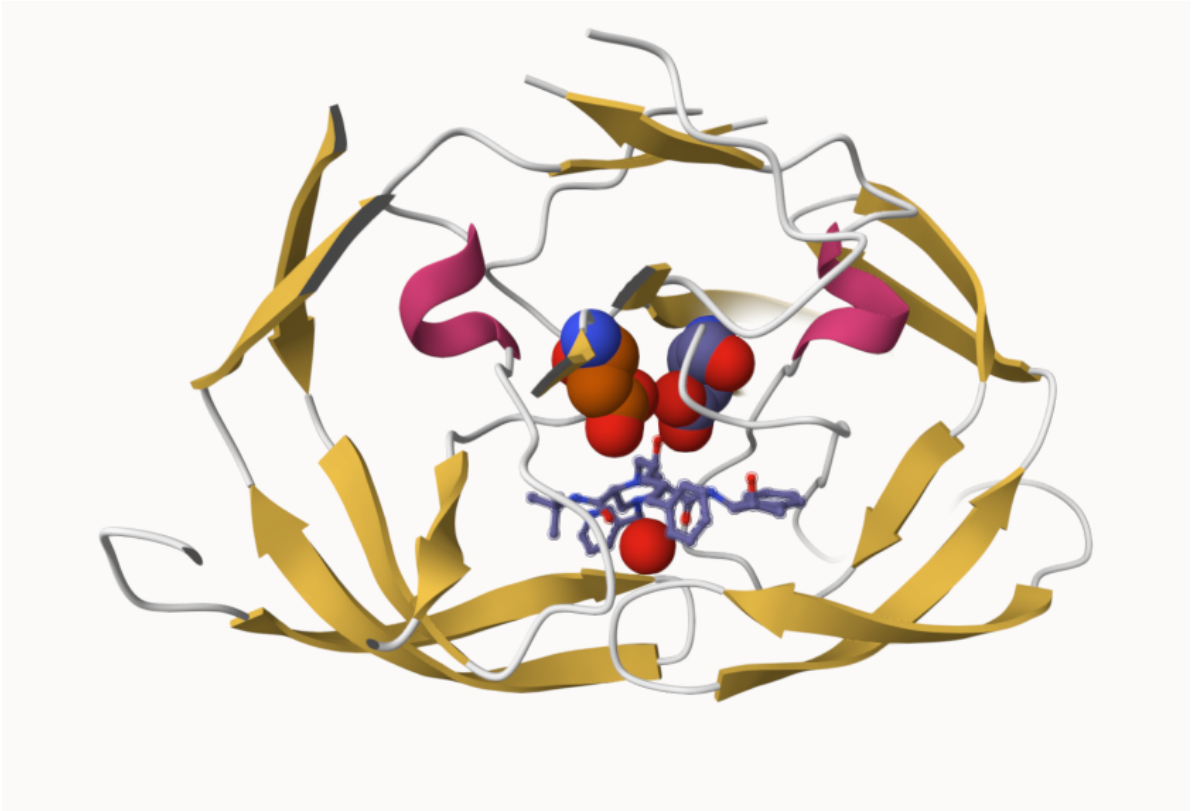


Figure 2: HIV-Pr highlighting ligand, D25, and H2O 308

```
  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
```

```
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"
```

Predicting functional motions of a single structure

we can predict the dynamics of a molecule

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

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

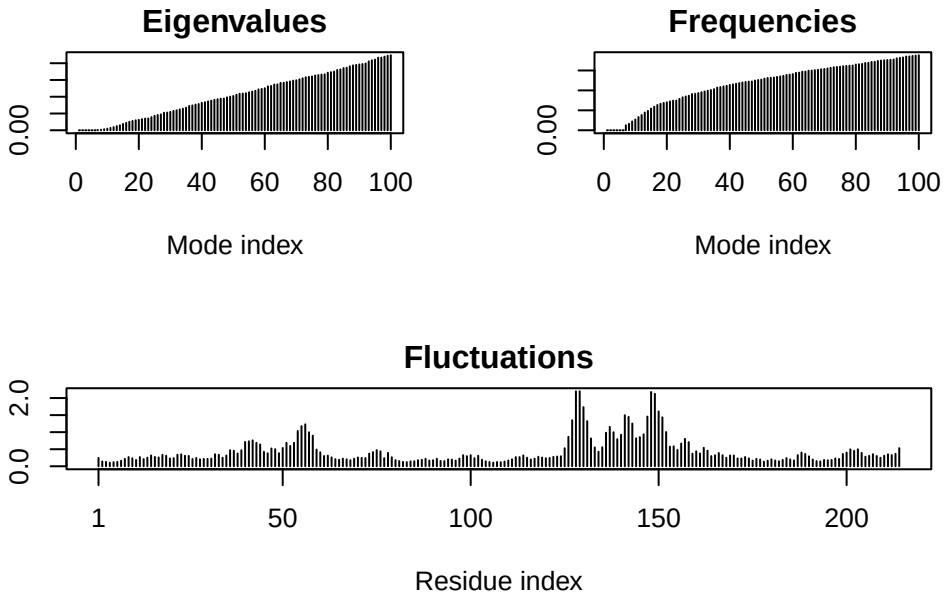
```
adk
```

```
 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) # normal mode analysis
```

```
Building Hessian...         Done in 0.07 seconds.
Diagonalizing Hessian...    Done in 0.56 seconds.
```

```
plot(m)
```




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