class11.Rmd

Anu Chaparala

11/2/2021

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

```
##                           X.ray   NMR   EM Multiple.methods Neutron Other  Total
## Protein (only)          142303 11804 5999              177      70    32 160385
## Protein/Oligosaccharide   8414    31  979                5       0     0   9429
## Protein/NA                7491   274 1986                3       0     0   9754
## Nucleic acid (only)       2368  1372   60                8       2     1   3811
## Other                      149    31    3                0       0     0    183
## Oligosaccharide (only)      11     6    0                1       0     4     22
```

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

87.6% structures in PDB solved by X-ray and 4.92% by EM.

```
method.sums <- colSums(db)
method.sums
```

```
##            X.ray              NMR               EM Multiple.methods
##           160736            13518             9027              194
##          Neutron            Other            Total
##               72               37           183584
```

```
XRay <- sum(db$X.ray)
EM <- sum(db$EM)
total <- sum(db$Total)

(XRay/total) *100
```

```
## [1] 87.55447
```

```
(EM/total) *100
```

```
## [1] 4.917095
```

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

87.4% are protein only structures in the PDB.

```
type.sums <- rowSums(db)
type.sums
```

```
##          Protein (only) Protein/Oligosaccharide              Protein/NA
##                  320770                   18858                   19508
##     Nucleic acid (only)                   Other  Oligosaccharide (only)
##                    7622                     366                      44
```

```
total1 <- sum(type.sums)
(type.sums["Protein (only)"]/total1)*100
```

```
## Protein (only)
##       87.36328
```

> 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?

There are currently 1828 HIV-1 protease structures in the current PDB.

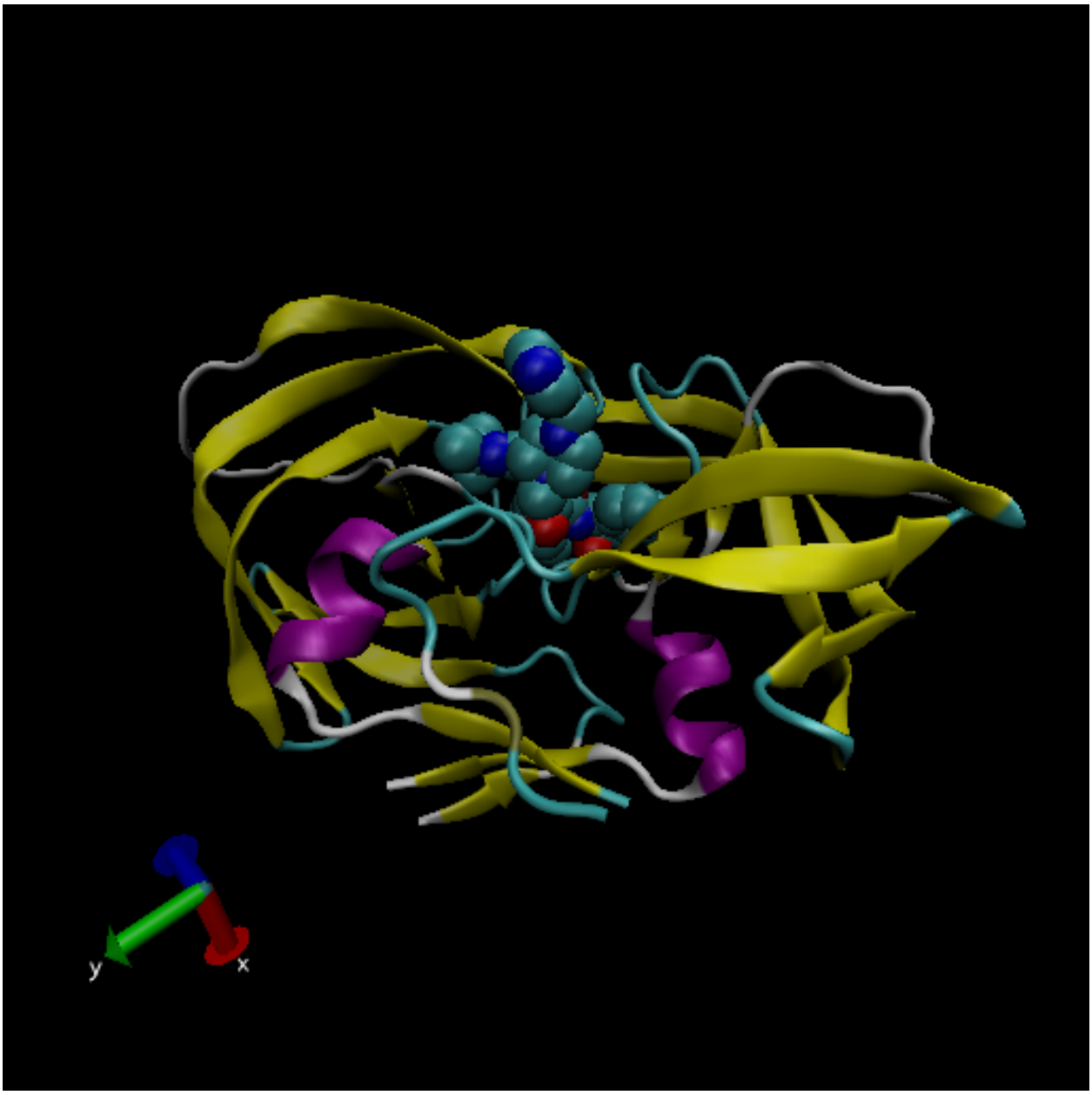
> Q4: Water molecules normally have 3 atoms. Why do we see just one atom per water molecule in this structure?

The hydrogen atoms per water molecule (2) are not shown because they are very small atoms compared to the oxygen (the one acutally shown) of the water molecules and the other atoms of the structure.

> Q5: There is a conserved water molecule in the binding site. Can you identify this water molecule? What residue number does this water molecule have (see note below)?

residue number MK1902:O4

VMD structure visualization image



Just write text but you have some **formatting** options.

I need to load library

```
library(bio3d)

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

```
##   Note: Accessing on-line PDB file
```

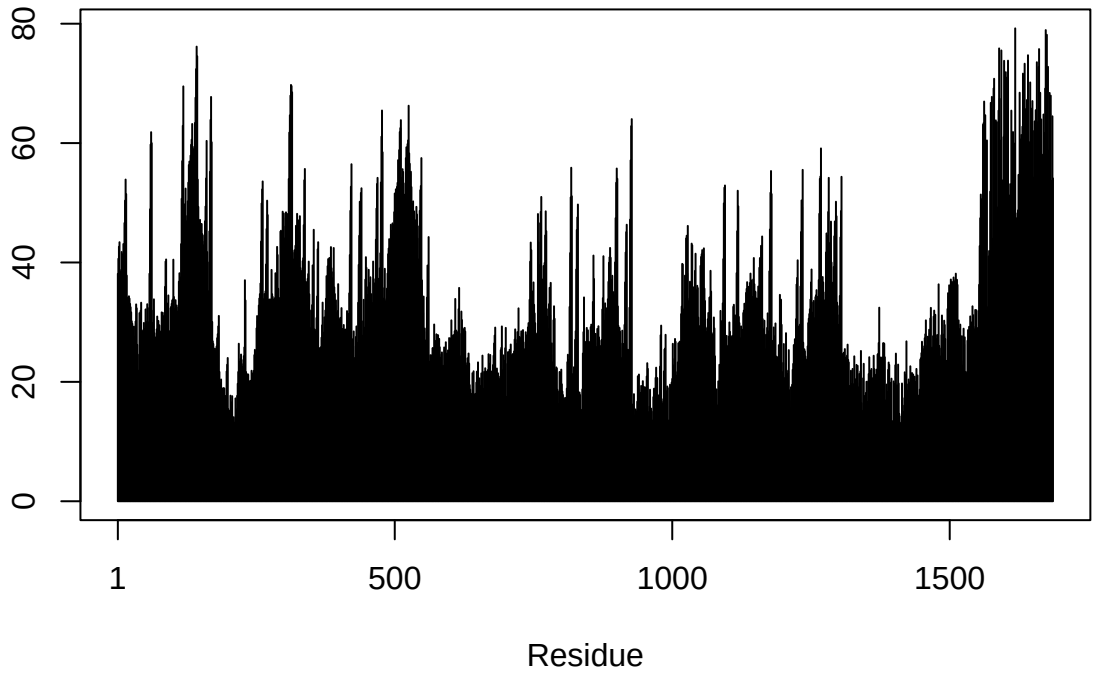
```
pdb
```

```
##
##  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
```

Plot B Factor

```
plot.bio3d(pdb$atom$b, sse = pdb)
```

```
## Warning in plotb3(...): Length of input 'sse' does not equal the length of input
## 'x'; Ignoring 'sse'
```



The ATOM records

```
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>
```