Class 11 Pt1

Natasha (PID: A15393874)

11/02/2021

A quick look at the PDB

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

```
##                           X.ray   NMR   EM Multiple.methods Neutron Other  Total
## Protein (only)          142795 11825 6141              177      70    32 161040
## Protein/Oligosaccharide   8454    31 1010                5       0     0   9500
## Protein/NA                7513   274 2039                3       0     0   9829
## Nucleic acid (only)       2380  1382   60                8       2     1   3833
## 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.

```
methods.sums <- colSums(db)
round((methods.sums/methods.sums["Total"])*100,2)
```

```
##            X.ray              NMR               EM Multiple.methods
##            87.47             7.35             5.02             0.11
##          Neutron            Other            Total
##             0.04             0.02           100.00
```

87.53% for X-Ray and 7.36% for NMR > Q2: What proportion of structures in the PDB are protein?

```
round(db$Total/methods.sums["Total"]*100,2)
```

```
## [1] 87.33  5.15  5.33  2.08  0.10  0.01
```

87.35% > 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 1828 HIV-1 protease structures in PDB.

2. 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? In these water molecules we can only see one atom which is the oxygen atom because the hydrogens are too small to see.

   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)? The watermolecule is HOH308:O. It is conserved at residue 308.

```
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
```

Q6: As you have hopefully observed HIV protease is a homodimer (i.e. it is composed of two identical chains). With the aid of the graphic display and the sequence viewer extension can you identify secondary structure elements that are likely to only form in the dimer rather than the monomer? Extended beta and helix are likely to only form in the dimer rather thatn the monomer. They need at least two chains to form structures.