

# Class11

Camryn McCann (PID: A15437387)

11/2/2021

#Introduction to the RCSB Protein Data Bank (PDB)

Downloaded the following CSV file from the PDB site.

```
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.**

```
round(sum(db$X-ray)/sum(db$Total)*100,2)
```

```
## [1] 87.55
```

```
round(sum(db$EM)/sum(db$Total)*100,2)
```

```
## [1] 4.92
```

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

```
round(db$Total[1]/sum(db$Total)*100,2)
```

```
## [1] 87.36
```

**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?**

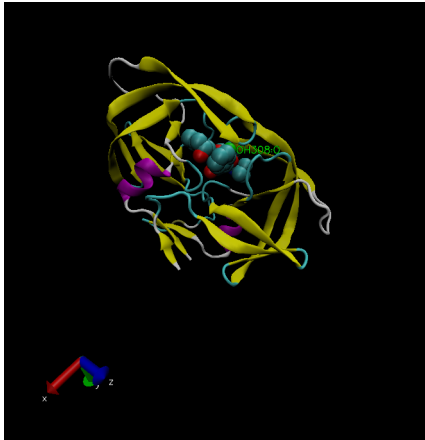
```
183790
```

```
#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?**

These water molecules only have one atom, which is oxygen, because the 2 hydrogen atoms present 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 water molecule is conserved at residue 308.

**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?**

After using the analysis of the sequence viewer on VMD, I think the secondary structure elements that only form in the homodimer rather than the monomer are the pink regions shown in the photo above. These pink regions are helices, have nearly identical protein composition, and are formed where the two monomers could interact in the middle.

