class10

```
, 6), Multiple_Methods = c(208, 8, 7, 14, 0, 1), Neutron = c(81, 2, 0, 3, 0, 0), Other = c
```

	Molecular_Type	X_ray	EM	NMR	Multiple_Methods	Neutron	Other
1	Protein (only)	169745	16880	12580	208	81	32
2	Protein/Oligosaccharide	9948	2843	34	8	2	0
3	Protein/NA	8803	5078	286	7	0	0
4	Nucleic acid (only)	2891	156	1521	14	3	1
5	Other	170	10	33	0	0	0
6	Oligosaccharide (only)	11	0	6	1	0	4

	Total
1	199526
2	12835
3	14174
4	4586
5	213
6	22

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

```
# Create the dataset
my_data <- data.frame(
  X_ray = c(169563, 9939, 8801, 2890, 170, 11),
  EM = c(16774, 2839, 5062, 151, 10, 0),
  Total = c(199236, 12822, 14156, 4580, 213, 22)
)

# Calculate the percentage of structures solved by X-ray and EM
xray_percentage <- (sum(my_data$X_ray) / sum(my_data$Total)) * 100
em_percentage <- (sum(my_data$EM) / sum(my_data$Total)) * 100

# Print the percentages
cat("Percentage of structures solved by X-ray: ", round(xray_percentage, 2), "%\n")
```

```
Percentage of structures solved by X-ray:  82.84 %
```

```
cat("Percentage of structures solved by Electron Microscopy: ", round(em_percentage, 2),
```

```
Percentage of structures solved by Electron Microscopy:  10.75 %
```

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

```
# Dataset with total structures
my_data <- data.frame(
  Molecular_Type = c("Protein (only)", "Protein/Oligosaccharide", "Protein/NA",
                     "Nucleic acid (only)", "Other", "Oligosaccharide (only)"),
```

```
    Total = c(199236, 12822, 14156, 4580, 213, 22)
)

# Calculate the proportion of protein-related structures
protein_proportion <- sum(my_data$Total[1:3]) / sum(my_data$Total)
cat("Proportion of structures that are protein:", round(protein_proportion, 4))
```

 Proportion of structures that are protein: 0.9792

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? 231,029 structures

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

We see water molecules instead of atoms.

Q5: There is a critical "conserved" water molecule in the binding site. Can you identify this water molecule? What residue number does this water molecule have?

308

Q6:Generate and save a figure clearly showing the two distinct chains of HIV-protease along with the ligand. Consider showing the catalytic residues ASP 25 in each chain and the critical water (we recommend "Ball & Stick" for these side-chains). Add this figure to your Quarto document.

Q7: How many amino acid residues are there in this pdb object?

```
library(bio3d)
pdb <- read.pdb("1hsg")
```

  Note: Accessing on-line PDB file

```
calpha_atoms <- pdb$atom[pdb$atom$resno > 0 & pdb$atom$elety == "CA", ]
num_residues <- length(calpha_atoms$resno)
print(num_residues)
```

 [1] 198

198 amino acid residues

Q8: Name one of the two non-protein residues? Water

Q9: How many protein chains are in this structure? 2