# HW Class 6

Alexandra Garcia (PID: A16278166)

## Install required packages

N.B. Run `install.packages("bio3d")` in console

```
# Can you improve this analysis code?
library(bio3d)
s1 <- read.pdb("4AKE") # kinase with drug
```

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

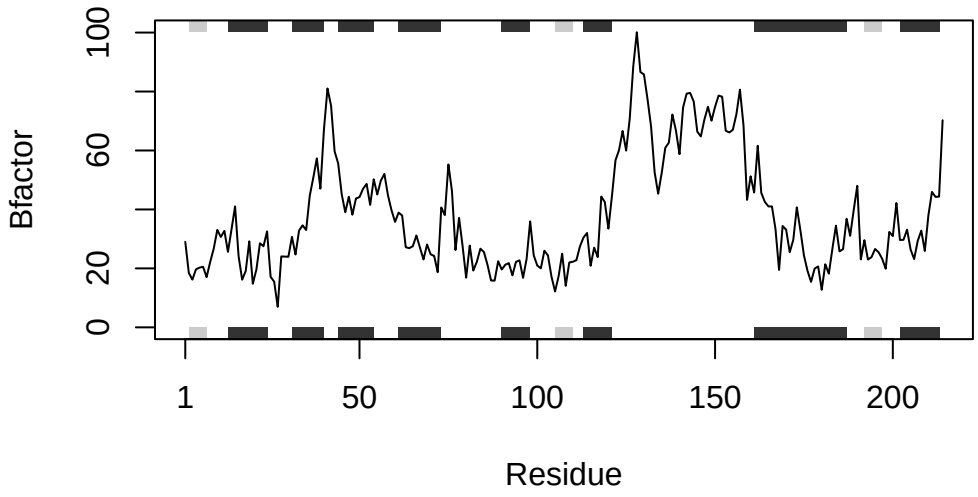
```
s2 <- read.pdb("1AKE") # kinase no drug
```

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

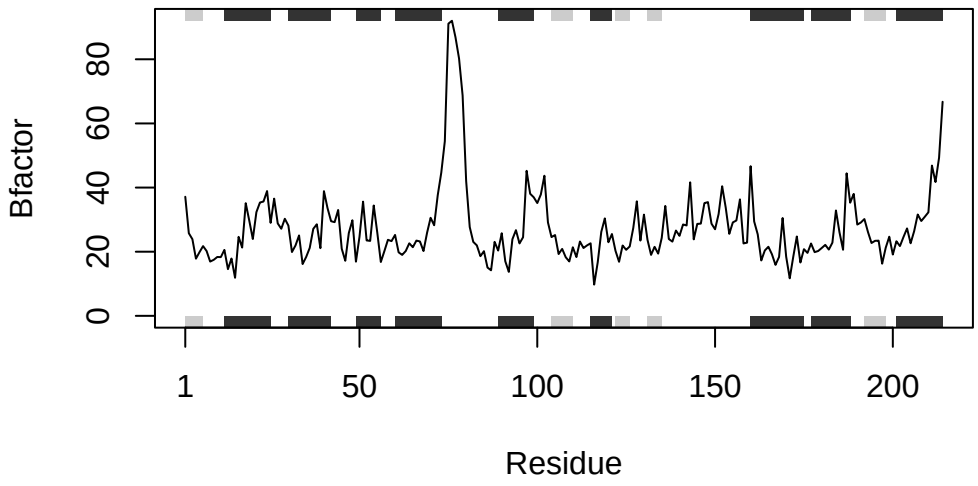
```
s3 <- read.pdb("1E4Y") # kinase with drug
```

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

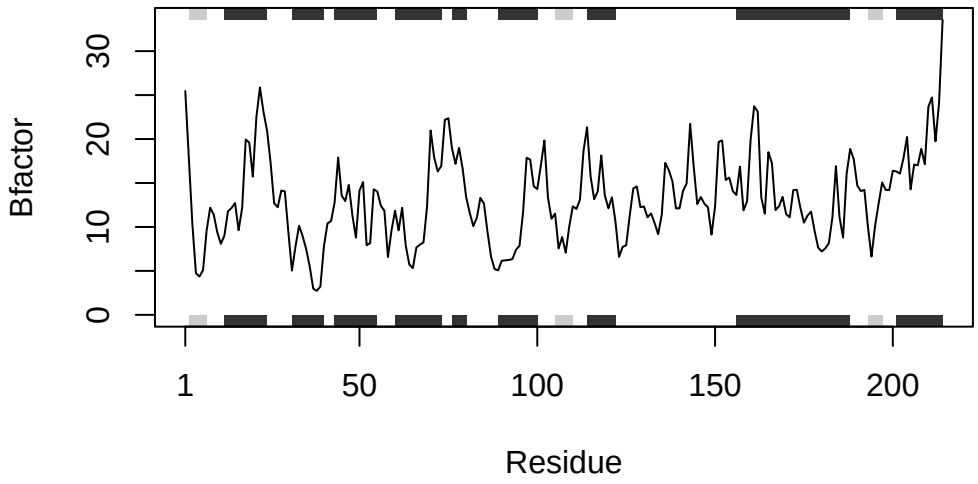
```
s1.chainA <- trim.pdb(s1, chain="A", elety="CA")
s2.chainA <- trim.pdb(s2, chain="A", elety="CA")
s3.chainA <- trim.pdb(s3, chain="A", elety="CA")
s1.b <- s1.chainA$atom$b
s2.b <- s2.chainA$atom$b
s3.b <- s3.chainA$atom$b
plotb3(s1.b, sse=s1.chainA, typ="l", ylab="Bfactor")
```




```
plotb3(s2.b, sse=s2.chainA, typ="l", ylab="Bfactor")
```



```
plotb3(s3.b, sse=s3.chainA, typ="l", ylab="Bfactor")
```




### Generalize the original code above to work with any set of input protein structures

> Replace the hardcoded parts that are doing the same thing but are being repeated over and over

Need to replace: **read.pdb**, which is manually inputting specific PDBs **trim.pdb**, which is trimming the PDBs to Chain A and Calcium atoms **s1.chainA***atom***b**, which is extracting the B factors **plotb3**, which is plotting each B factor separately

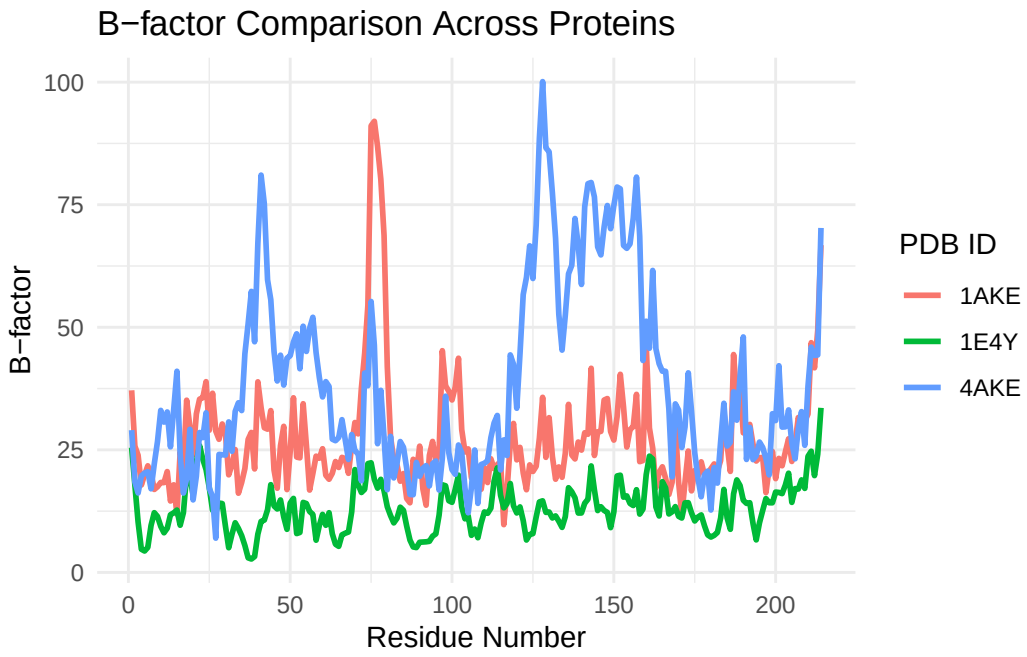
```
# Loads the required packages
library(bio3d)
library(ggplot2)

# Define the function
# The function is called `analyze_protein_drug_interaction()`
# pdb_ids: a vector of PDB IDs (protein structure codes)
# chain_id: which protein chain to look at (default is "A" in the original code)
# element_type: which atoms to include (default is "CA" in the original code)
analyze_protein_drug_interactions <- function(pdb_ids, chain_id = "A", element_type = "CA") {
```

```
# If statement checks that input is a vector of PDB IDs,
if (!is.character(pdb_ids)) {
  stop("Please provide a character vector of PDB IDs.")
}

# Initialize an empty data frame to hold all data (residue number, B-factor, which protein
all_data <- data.frame()

# Create two lists to store B-factors and trimmed structures
b_factors <- list()
trimmed_structures <- list()

# Starts a loop to go through every PDB ID in the input list one by one
for (pdb_id in pdb_ids) {
  # Prints a message in the console with which PDB is being read right now
  cat("Reading PDB:", pdb_id, "\n")
  # Reads the protein structure file for the current PDB ID and stores it in pdb
  pdb <- read.pdb(pdb_id)

  # Trim to specified chain and atom type
  pdb_trim <- trim.pdb(pdb, chain = chain_id, elety = element_type)

  # Extract B-factors
  b_factors <- pdb_trim$atom$b

  # Create a data frame with residue index, B-factor, and PDB ID
  df <- data.frame(
    Residue = seq_along(b_factors),  # residue index (1 to N, position of each residue)
    BFactor = b_factors,
    PDB = pdb_id                     # label for which protein this is
  )

  # Combine with the full dataset
  all_data <- rbind(all_data, df) # rbind() means "row-bind"
}

# Plot using ggplot2
# The output will be a plot with Residue on x-axis,
ggplot(all_data) +
  aes(x = Residue, y = BFactor, color = PDB) +
  geom_line(linewidth = 1) +
  labs(
```

```
      title = "B-factor Comparison Across Proteins",
      x = "Residue Number",
      y = "B-factor",
      color = "PDB ID"
    ) +
    theme_minimal()
}
```

```
# Example: comparing 3 kinases (some with drug, some without)
analyze_protein_drug_interactions(c("4AKE", "1AKE", "1E4Y"))
```

```
Reading PDB: 4AKE
  Note: Accessing on-line PDB file

Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
/var/folders/0v/f3csf_r56v17z4twjk64p92c0000gn/T//Rtmp7e1a6u/4AKE.pdb exists.
Skipping download

Reading PDB: 1AKE
  Note: Accessing on-line PDB file

Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
/var/folders/0v/f3csf_r56v17z4twjk64p92c0000gn/T//Rtmp7e1a6u/1AKE.pdb exists.
Skipping download

   PDB has ALT records, taking A only, rm.alt=TRUE
Reading PDB: 1E4Y
  Note: Accessing on-line PDB file

Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
/var/folders/0v/f3csf_r56v17z4twjk64p92c0000gn/T//Rtmp7e1a6u/1E4Y.pdb exists.
Skipping download
```


B−factor Comparison Across Proteins
B−factor
100
75
50
25
0
0
50
100
150
200
Residue Number
PDB ID
1AKE
1E4Y
4AKE