HW class 6

Charlize Molitor (PID: A18515740)

> Q6. How would you generalize the original code above to work with any set of input protein structures?

Scoring Rubric: Total 10 points assigned as follows: Documentation: 1 pt - comments on what are the inputs to the function. 1 pt - what the function does and how to use it. 1 pt - what is the output of the function. Code: 2 pt - function behaves as desired, producing the correct output and follows assignment specifications. 2 pt - the code is efficient meaning it uses best practices such as limiting calculation duplication. 2 pt - code is readable, meaning best practices are used including proper indentation and whitespace used, relevant variable names, and organized in a logical manner. 1 pt - function code and call executes and is working properly.

```
library(bio3d)
s1 <- read.pdb("4AKE") # kinase with drug
```

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

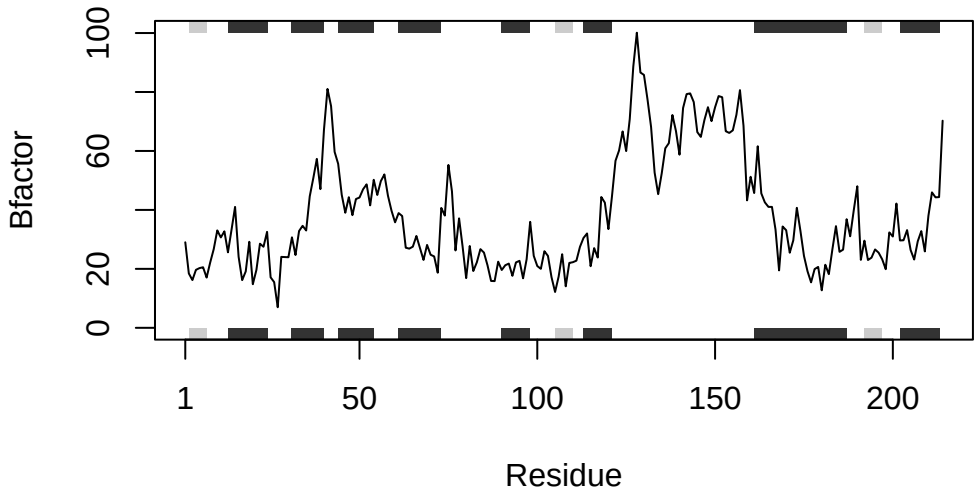
```
s2 <- read.pdb("1AKE") # kinase no drug
```

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

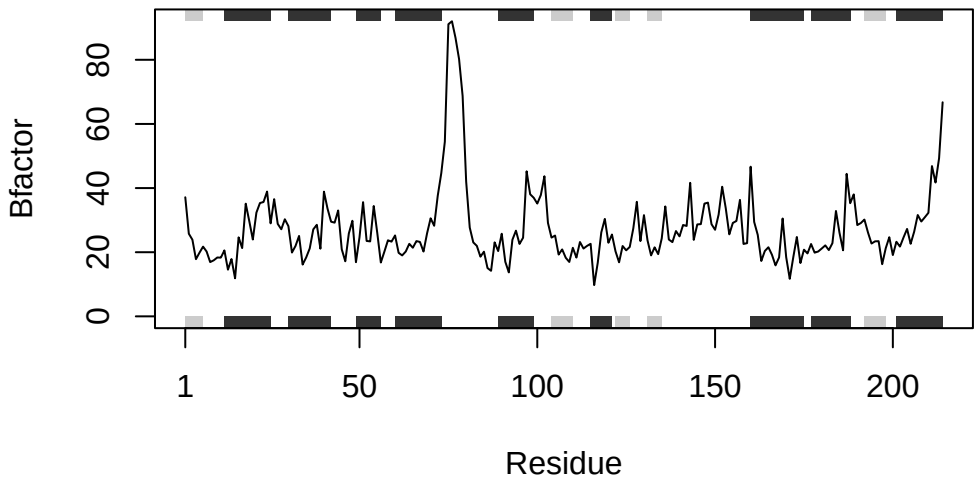
```
s3 <- read.pdb("1E4Y") # kinase with drug
```

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

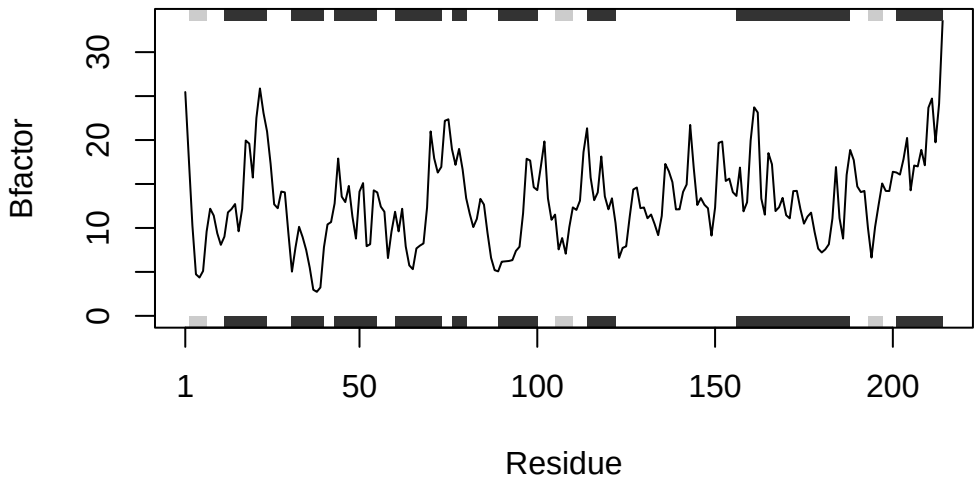
```
s1.chainA <- trim.pdb(s1, chain="A", elety="CA")
s2.chainA <- trim.pdb(s2, chain="A", elety="CA")
s3.chainA <- trim.pdb(s3, chain="A", elety="CA")
s1.b <- s1.chainA$atom$b
s2.b <- s2.chainA$atom$b
s3.b <- s3.chainA$atom$b
plotb3(s1.b, sse=s1.chainA, typ="l", ylab="Bfactor")
```



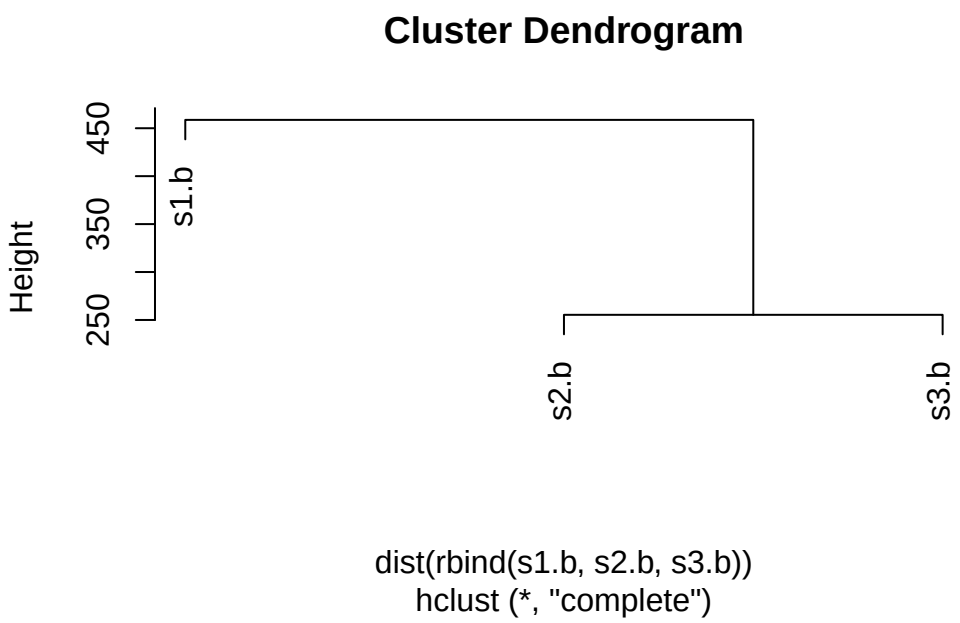

```
plotb3(s2.b, sse=s2.chainA, typ="l", ylab="Bfactor")
```




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



```
hc <- hclust( dist( rbind(s1.b, s2.b, s3.b) ) )
plot(hc)
```




```
 analyze_protein_clusters <- function(pdb_ids, chain="A", elety="CA") {
  library(bio3d)
  chains <- lapply(pdb_ids, function(id) trim.pdb(read.pdb(id), chain=chain, elety=elety))
  bfactors <- lapply(chains, function(x) x$atom$b)
  hc <- hclust(dist(do.call(rbind, bfactors)))
  plot(hc)
}
```

1. The inputs of the function are PDB ids, the chains and the elety.
2. The function is supposed to allow us to work with any set of input protein structures.
3. The output of the function is the protein structures.