Lab 6 Homework

Kevin Tan (PID: A16774162)

Section 1

```
analysis <- function(x){
  x <- (x - min(x)) / (max(x) - min(x))
}
```

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

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

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

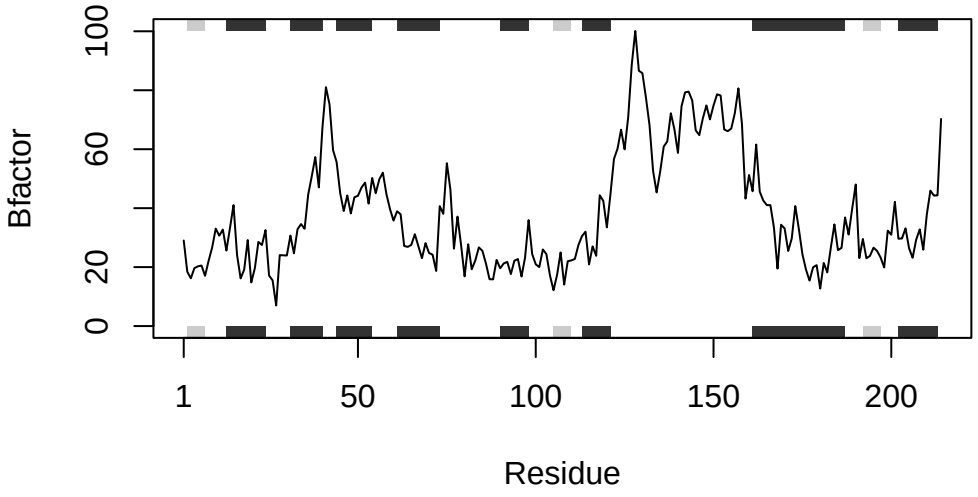
```
Note: Accessing on-line PDB file
 PDB has ALT records, taking A only, rm.alt=TRUE
```

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

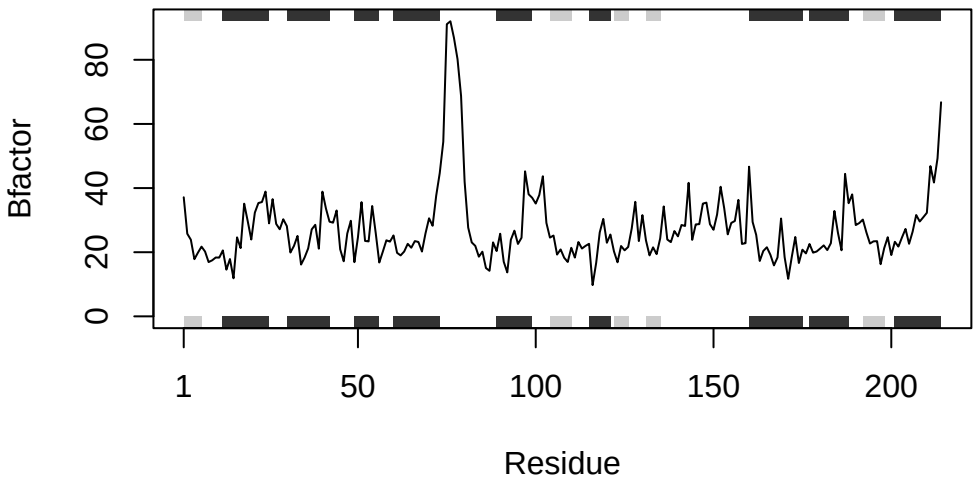
```
Note: Accessing on-line PDB file
```

```
s1.chainA <- trim.pdb(s1, chain="A", elety="CA")
s2.chainA <- trim.pdb(s2, chain="A", elety="CA")
s3.chainA <- trim.pdb(s1, chain="A", elety="CA")
s1.b <- s1.chainA$atom$b
s2.b <- s2.chainA$atom$b
s3.b <- s3.chainA$atom$b
```

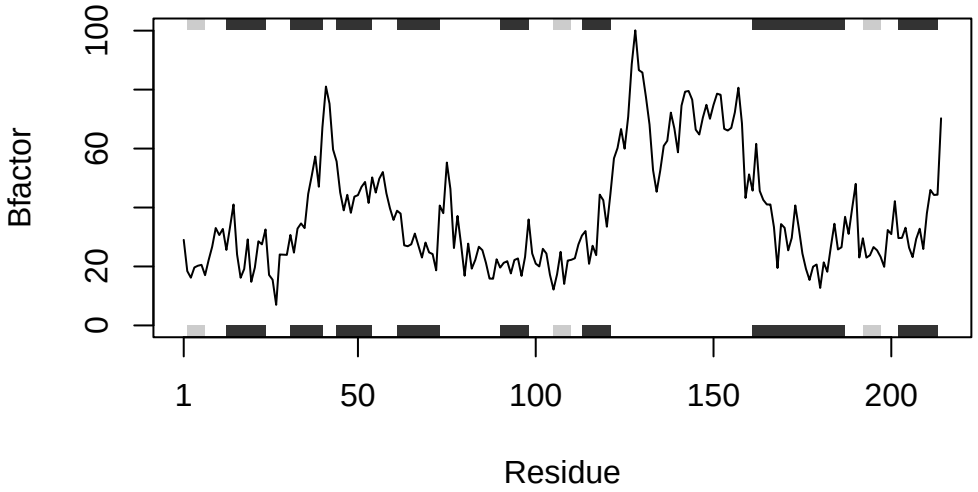
```
plotb3(s1.b, sse=s1.chainA, typ="l", ylab="Bfactor")
```




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

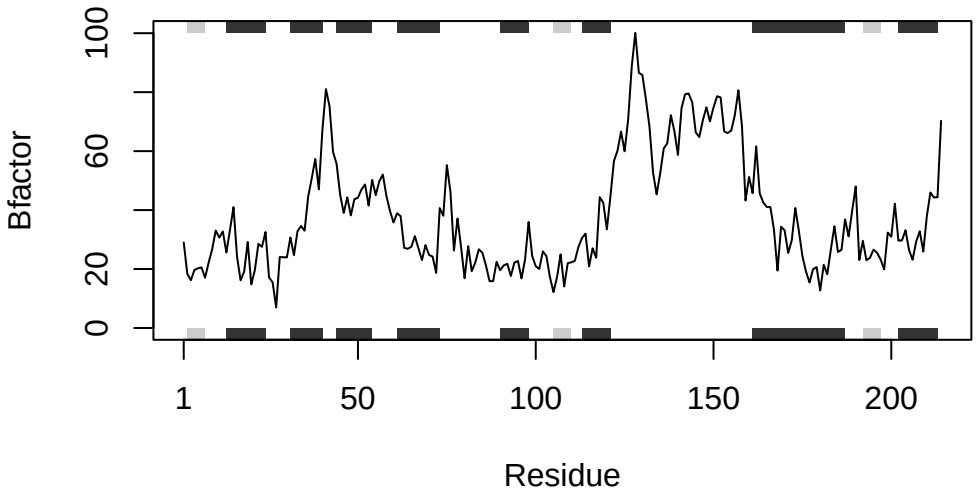



```
plotb3(s3.b, sse=s3.chainA, typ="l", ylab="Bfactor")
```



```
analysis2 <- function(x){
  x.chainA <- trim.pdb(x, chain="A", elety="CA")
  x.b <- x.chainA$atom$b
  plotb3(x.b, sse=s1.chainA, typ="l", ylab="Bfactor")
}
```

```
analysis2(s1)
```



Q1. What type of object is returned from the read.pdb() function?

```
mode(s1)
```

```
[1] "list"
```

Q2. What does the trim.pdb() function do?

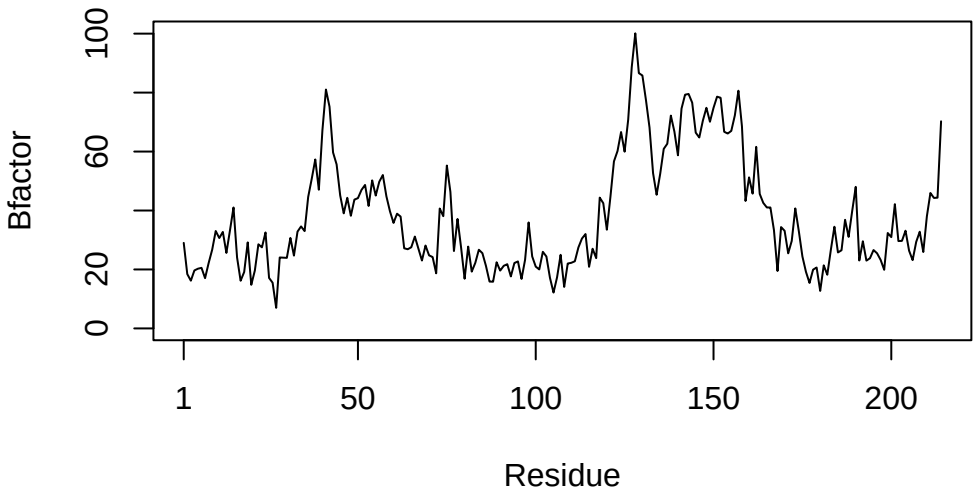
```
?trim.pdb
```

```
starting httpd help server ... done
```

it filters out specific structures from a PDB object, such as the chain

Q3. What input parameter would turn off the marginal black and grey rectangles in the plots and what do they represent in this case?

```
plotb3(s1.b, sse=NULL, typ="l", ylab="Bfactor")
```

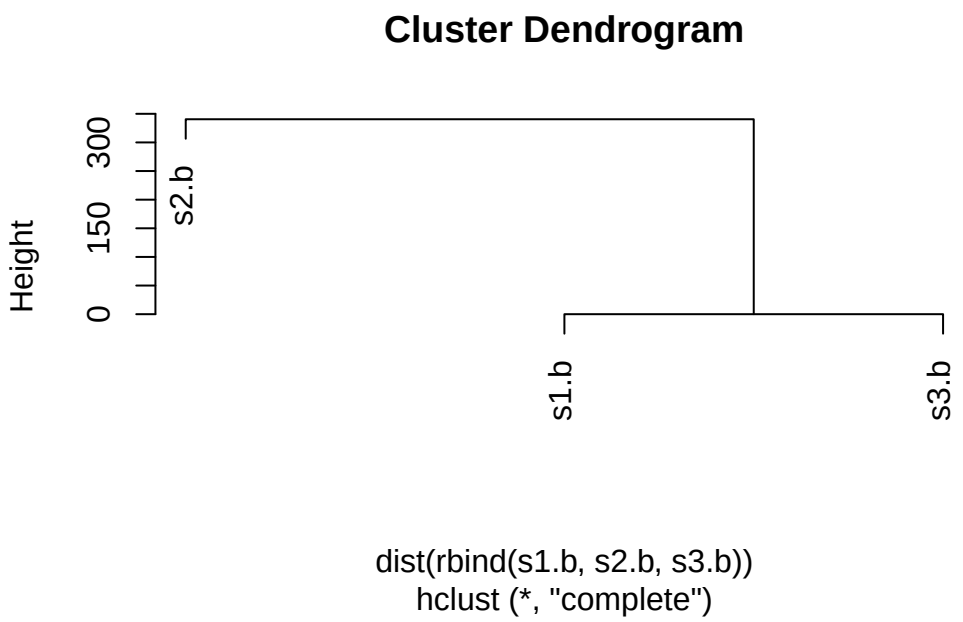


they are indicating a secondary structure object

Q4. What would be a better plot to compare across the different proteins?

Q5. Which proteins are more similar to each other in their B-factor trends. How could you quantify this? HINT: try the rbind(), dist() and hclust() functions together with a resulting dendrogram plot. Look up the documentation to see what each of these functions does.

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



4AKE and 1E4Y are the most similar to each other based on the Cluster dendrogram. We achieve this through combining the dataset for all 3 proteins, performing a distance matrix computation which shows how dissimilar each protein is to another, and then run a hierarchical clustering, which groups our proteins based on similarity. This finally plotted.

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

```
#This function takes in a protein code and returns a Bfactor plot
drug_analysis <- function(x){
  #input protein name for x and reads data from pdb
  s <- read.pdb(x)
  # Take chain A data
  s.chainA <- trim.pdb(s, chain="A", elety="CA")
  # Take chain A atom B data
  s.b <- s1.chainA$atom$b
  #Plot Data - Output
  plotb3(s.b, sse=s.chainA, typ="l", ylab="Bfactor")
}
```

Testing the function:

```
drug_analysis("4AKE")
```

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

```
Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
C:\Users\sirmo\AppData\Local\Temp\Rtmpiek7pk/4AKE.pdb exists. Skipping download
```