HW Class 6

Chloe Do

October 16, 2022

Section 1: Improving analysis code by writing function

First, install the package into RStudio

```
#RStudio > Tools > Install Packages
#install.packages("bio3d)
```

Then, run the code from pdf to check if it works

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

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

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

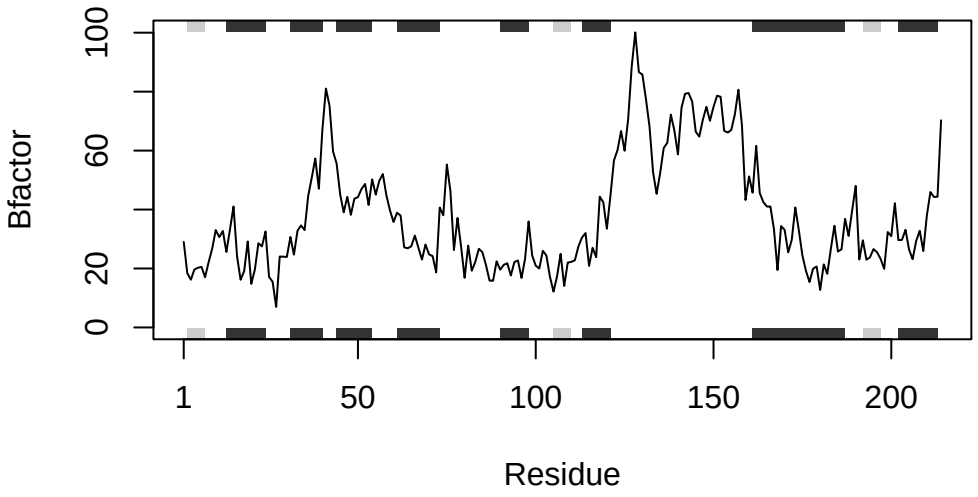
```
Note: Accessing on-line PDB file
 PDB has ALT records, taking A only, rm.alt=TRUE
```

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

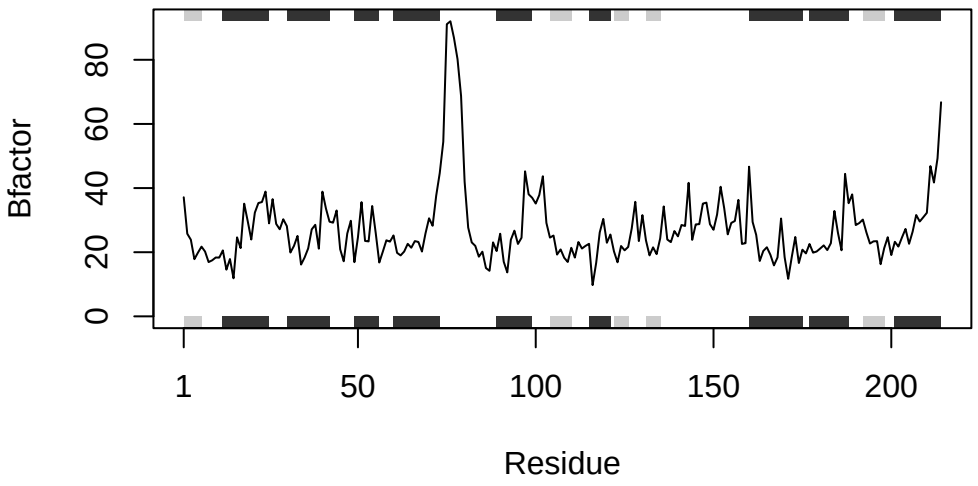
```
Note: Accessing on-line PDB file
```

```
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")
```

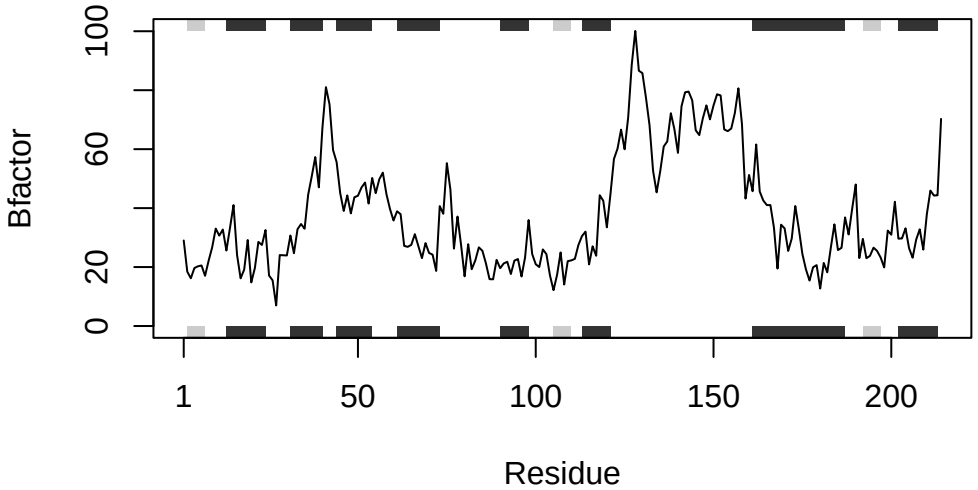
```
s1.b <- s1.chainA$atom$b
s2.b <- s2.chainA$atom$b
s3.b <- s3.chainA$atom$b
plotb3(s1.b, sse=s1.chainA, typ="l", ylab="Bfactor")
```




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




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



Review the code errors

```
#We can improve this by remove the error at s3.chainA by changing "s1" to "s3"
library(bio3d)
s1 <- read.pdb("4AKE")  # kinase with drug
```

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

```
Warning in get.pdb(file, path = tempdir(), verbose = FALSE): /var/folders/72/
rxcwp5tn57j3p8cc302fpp_r0000gn/T//RtmpYs2oeT/4AKE.pdb exists. Skipping download
```

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

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

```
Warning in get.pdb(file, path = tempdir(), verbose = FALSE): /var/folders/72/
rxcwp5tn57j3p8cc302fpp_r0000gn/T//RtmpYs2oeT/1AKE.pdb exists. Skipping download
```

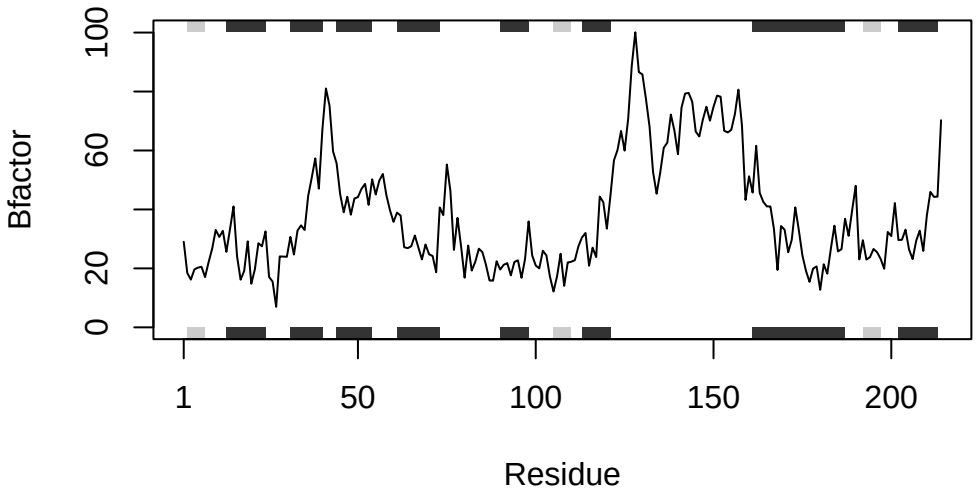
```
   PDB has ALT records, taking A only, rm.alt=TRUE
```

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

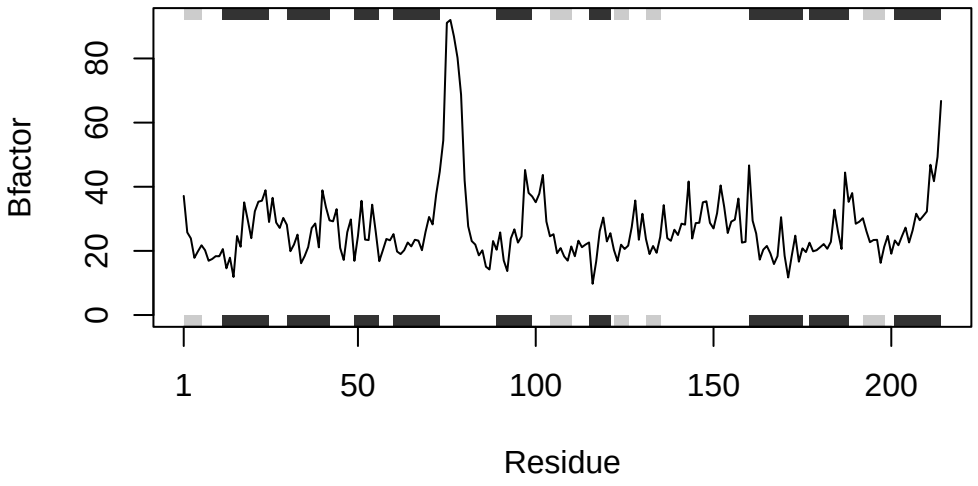
```
  Note: Accessing on-line PDB file
```

```
Warning in get.pdb(file, path = tempdir(), verbose = FALSE): /var/folders/72/
rxcwp5tn57j3p8cc302fpp_r0000gn/T//RtmpYs2oeT/1E4Y.pdb exists. Skipping download
```

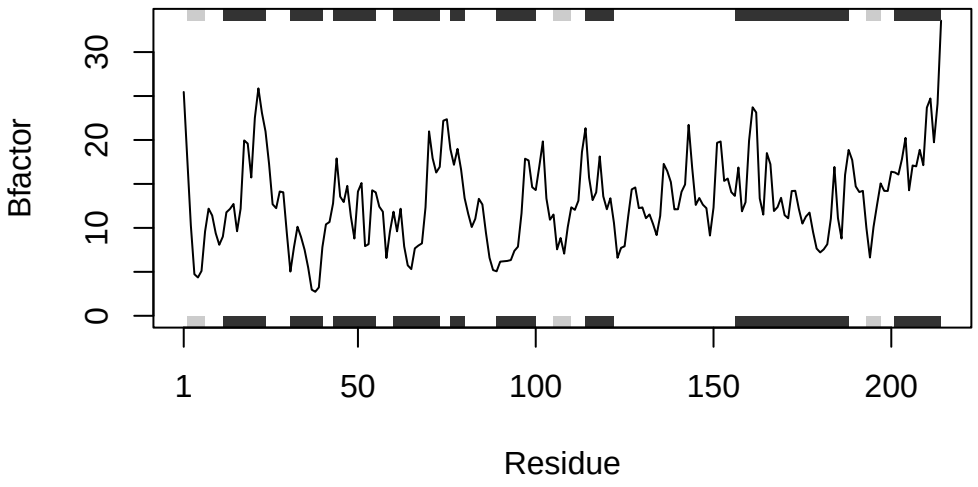
```
s1.chainA <- trim.pdb(s1, chain="A", elety="CA")
s2.chainA <- trim.pdb(s2, chain="A", elety="CA")
s3.chainA <- trim.pdb(s3, chain="A", elety="CA") #Change s1 to s3
s1.b <- s1.chainA$atom$b
s2.b <- s2.chainA$atom$b
s3.b <- s3.chainA$atom$b
plotb3(s1.b, sse=s1.chainA, typ="l", ylab="Bfactor")
```




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



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



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

What are the inputs to this function?

The inputs to this function are the PDB files "4AKE", "1AKE", and "1E4Y"

```
#Below are the inputs from the original code from the pdf
#s1 <- read.pdb("4AKE")  # kinase with drug
#s2 <- read.pdb("1AKE")  # kinase no drug
#s3 <- read.pdb("1E4Y")  # kinase with drug
```

What the function do and how to use it?

The function will take the PDB data files and provide the effect of drug binding as output

```
#Inputs are the name of PDB files

#Naming the function, I will name it as `plot.protein`
plot.protein <- function(x) {
#To use "bio3d", add it to library
```

```
  library(bio3d)
#Open and trim the input file and pay attention to the specific information
  s <- read.pdb(x)
  s.chainA <- trim.pdb(s, chain="A", elety="CA")
  s.b <- s.chainA$atom$b
#Plot the data to see the effect on drug binding
  plotb3(s.b, sse=s.chainA, typ="l", ylab="Bfactor")
}
```

What is the output of the function?

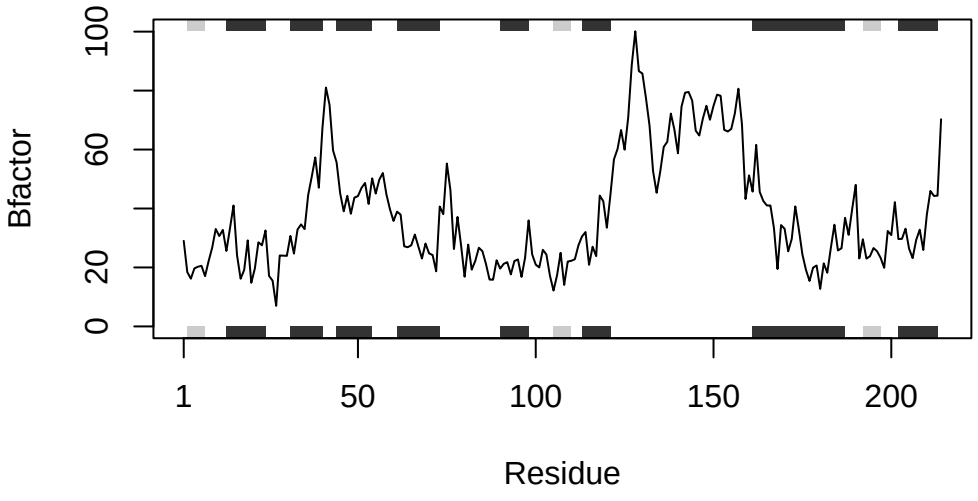
The output of the function is the plot of the effect of drug binding

Let test the function with the three given input to see if the function is working.

```
plot.protein("4AKE")
```

```
  Note: Accessing on-line PDB file

Warning in get.pdb(file, path = tempdir(), verbose = FALSE): /var/folders/72/
rxcwp5tn57j3p8cc302fpp_r0000gn/T//RtmpYs2oeT/4AKE.pdb exists. Skipping download
```

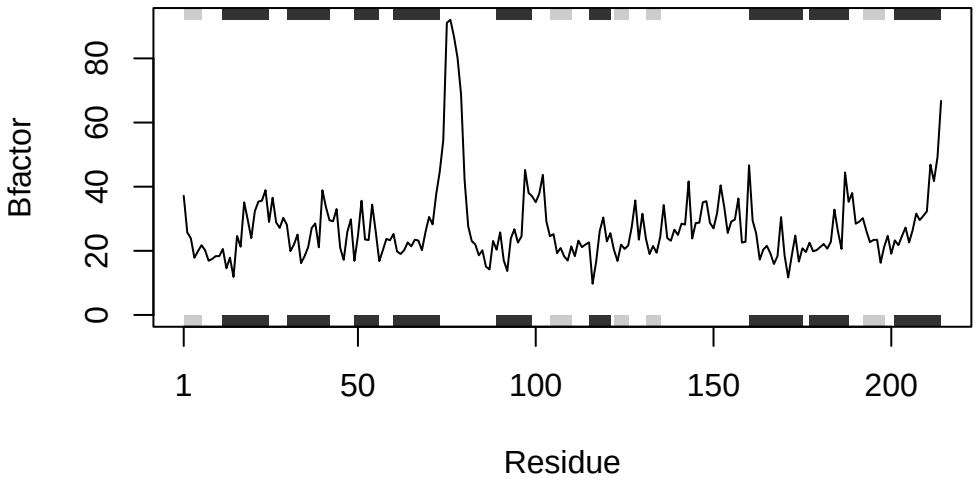


```
plot.protein("1AKE")
```

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

```
Warning in get.pdb(file, path = tempdir(), verbose = FALSE): /var/folders/72/
rxcwp5tn57j3p8cc302fpp_r0000gn/T//RtmpYs2oeT/1AKE.pdb exists. Skipping download
```

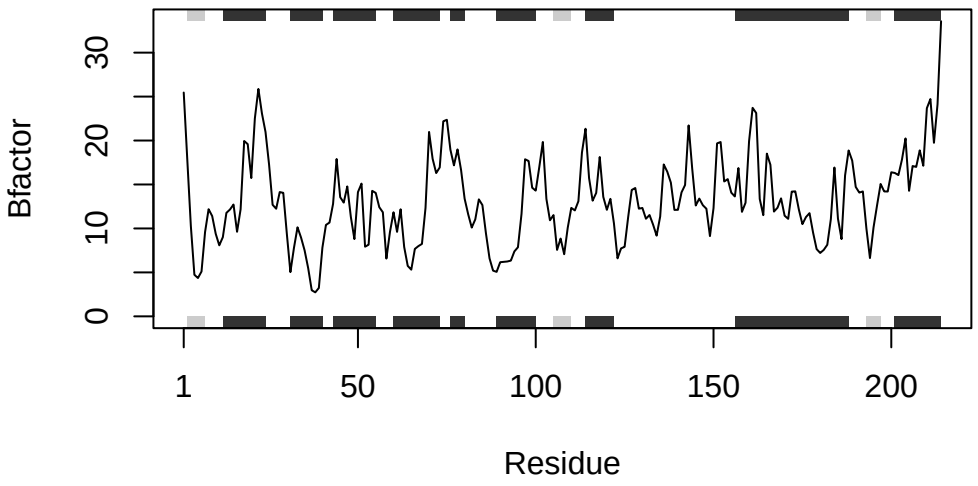
```
   PDB has ALT records, taking A only, rm.alt=TRUE
```




```
plot.protein("1E4Y")
```

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

```
Warning in get.pdb(file, path = tempdir(), verbose = FALSE): /var/folders/72/
rxcwp5tn57j3p8cc302fpp_r0000gn/T//RtmpYs2oeT/1E4Y.pdb exists. Skipping download
```



Below is the final function that would work for any set of protein structure

```
plot.protein <- function(x){
  library(bio3d)
  s <- read.pdb(x)
  s.chainA <- trim.pdb(s, chain="A", elety="CA")
  s.b <- s.chainA$atom$b
  plotb3(s.b, sse=s.chainA, type="l", ylab="Bfactor")
}
```