Class 6: R functions HW

Canbin Cai (PID: A18087473)

Table of contents

Q.6 in the worksheet

Can you improve this analysis code?

Original 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
```

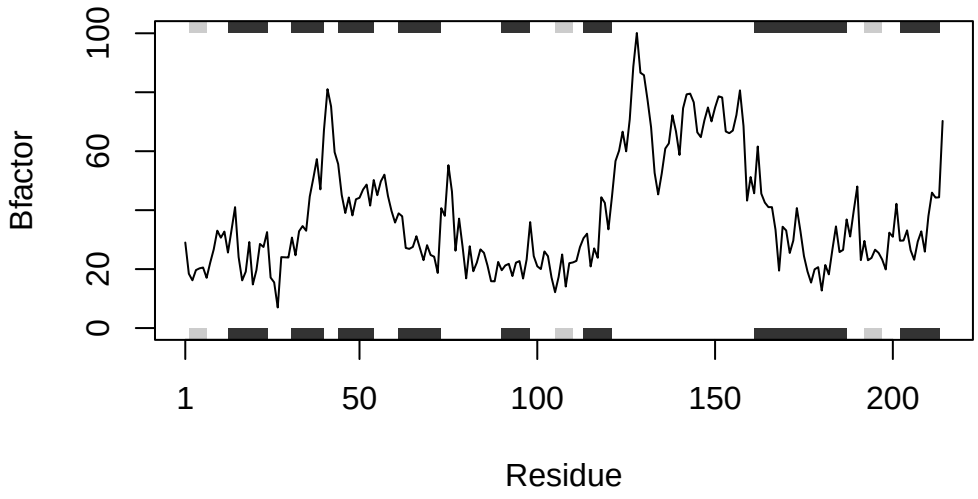
```
s3 <- read.pdb("1E4Y") # kinase with drug
```

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

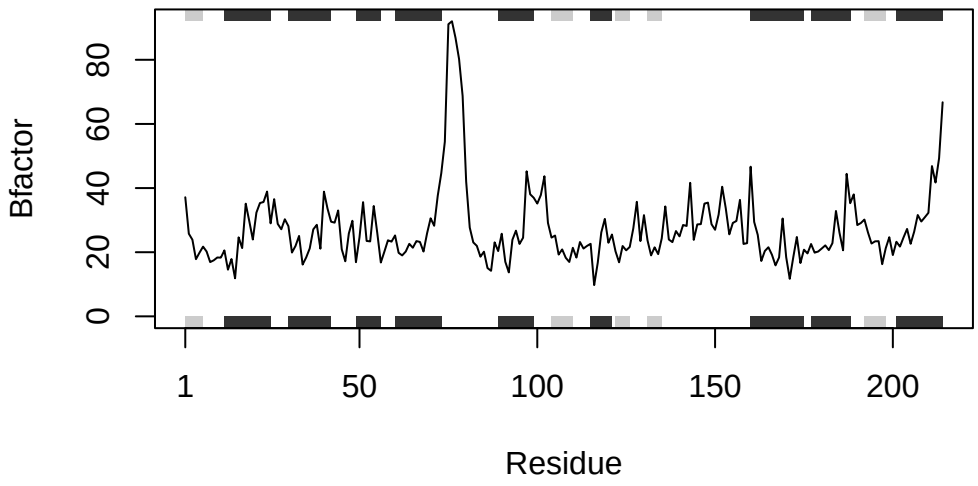
```
s1.chainA <- trim.pdb(s1, chain="A", elety="CA")
s2.chainA <- trim.pdb(s2, chain="A", elety="CA")
s3.chainA <- trim.pdb(s1, chain="A", elety="CA")
```

```
s1.b <- s1.chainA$atom$b
s2.b <- s2.chainA$atom$b
s3.b <- s3.chainA$atom$b
```

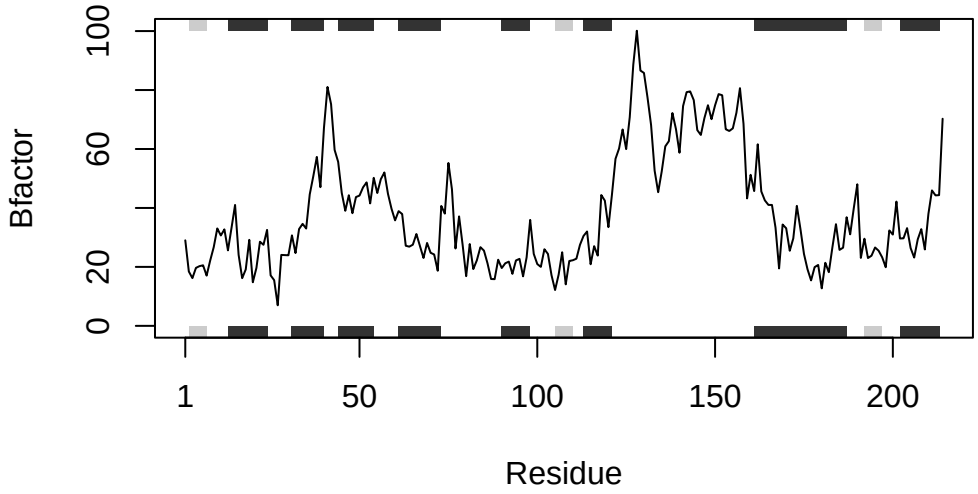
```
plotb3(s1.b, sse=s1.chainA, typ="l", ylab="Bfactor")
```




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




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



Improve the code above —> My Solution.

Steps on improving the code in Q.6:

1. I created the `function()` named **protein** with the argument of x, which will be the three kinase proteins *4AKE*, *1AKE*, and *1E4Y*.
2. By using the `read.pdb()` function, it will access the online PDB file of the *bio3d package* that is being installed and access the data for the argument of x, which will be the proteins *4AKE*, *1AKE*, and *1E4Y*. And I named this **s**.
3. By using the `trim.pdb()` function, it will access the specific protein chain and atom type we want to focus on. In this case, I want to only focus on the alpha carbon atoms in chain A of each protein structure in **s**, which will be proteins *4AKE*, *1AKE*, and *1E4Y*. And I named this **s.chain**.
4. Then, I want to access the data frame in each *atom type* and *b-factors* column. And I named this **s.b**.
5. Lastly, I want all the results that I generated above to be on graph plots using the `plotb3()` function.

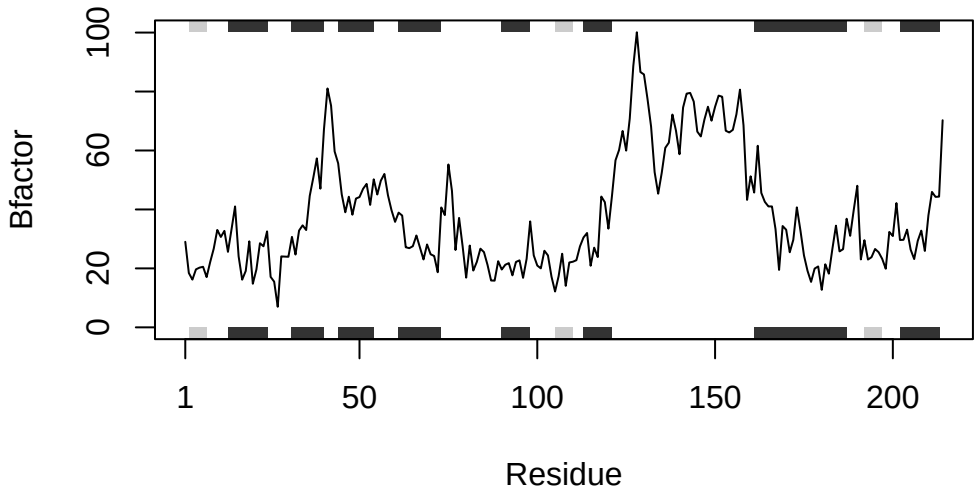
```
library(bio3d)
protein <- function(x){
  s <-read.pdb(x)
  s.chain  <- trim.pdb(s, chain="A", elety="CA")
  s.b <- s.chain$atom$b
  plotb3(s.b, sse=s.chain, typ="l", ylab="Bfactor")
}
```

Plot for kinase **4AKE**:

```
protein("4AKE")
```

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

```
Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
/var/folders/9z/lsyt5f2n0b582k6k3s2r24nw0000gn/T//Rtmpc2ekSN/4AKE.pdb exists.
Skipping download
```



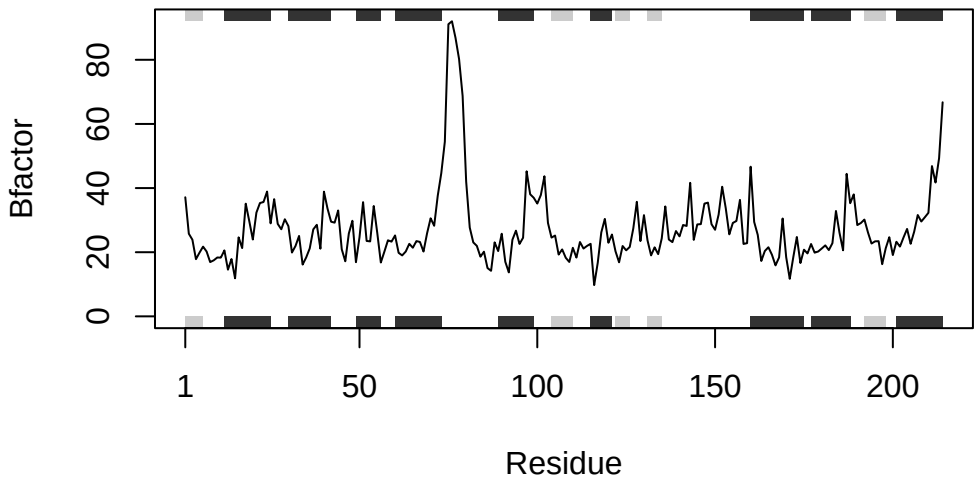
Plot for kinase **1AKE**:

```
protein("1AKE")
```

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

```
Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
/var/folders/9z/lsyt5f2n0b582k6k3s2r24nw0000gn/T//Rtmpc2ekSN/1AKE.pdb exists.
Skipping download
```

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

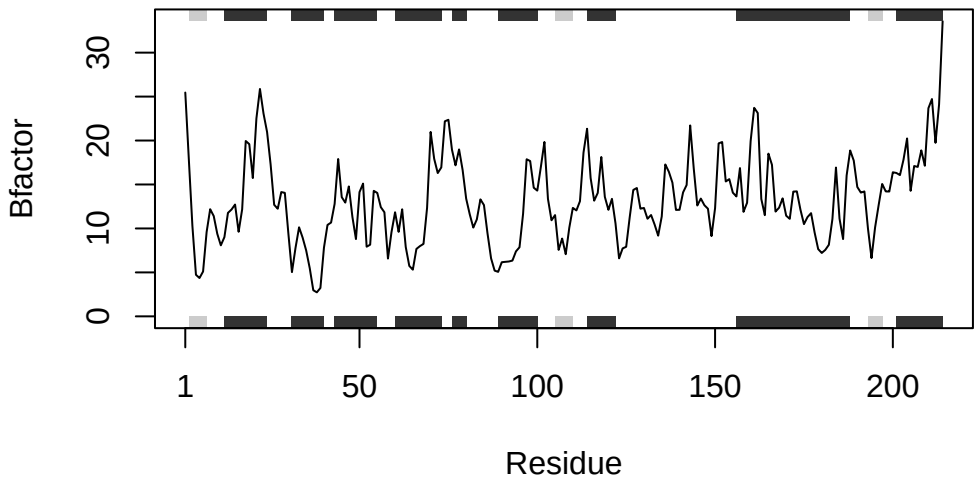


Plot for kinase **1E4Y**:

```
protein("1E4Y")
```

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

```
Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
/var/folders/9z/lsyt5f2n0b582k6k3s2r24nw0000gn/T//Rtmpc2ekSN/1E4Y.pdb exists.
Skipping download
```


Bfactor
0
10
20
30
1
50
100
150
200
Residue