Lab 6 HW: New Function

Ian Gurholt (PID: A16767484)

Can you improve this analysis code?

```
library(bio3d)
```

```
Warning: package 'bio3d' was built under R version 4.3.3
```

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

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

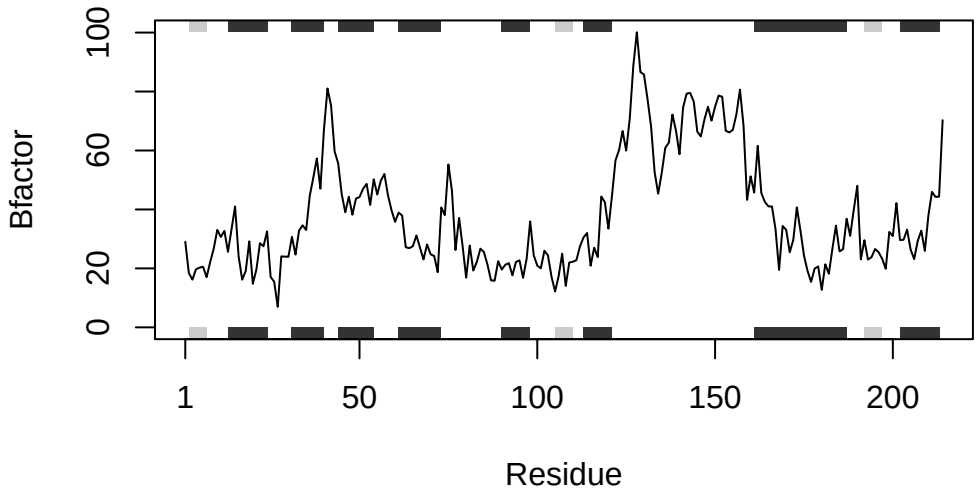
```
s2 <- read.pdb("1AKE") # kinase no drug
```

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

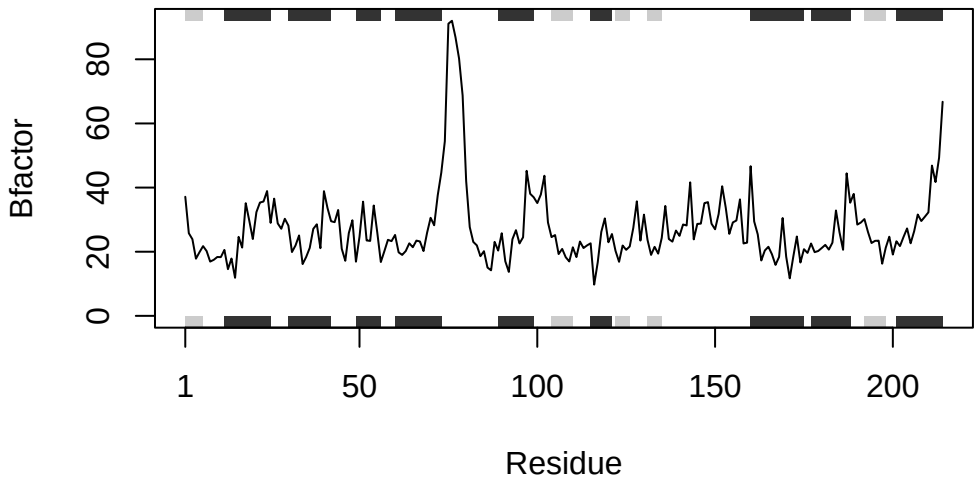
```
s3 <- read.pdb("1E4Y") # kinase with drug
```

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

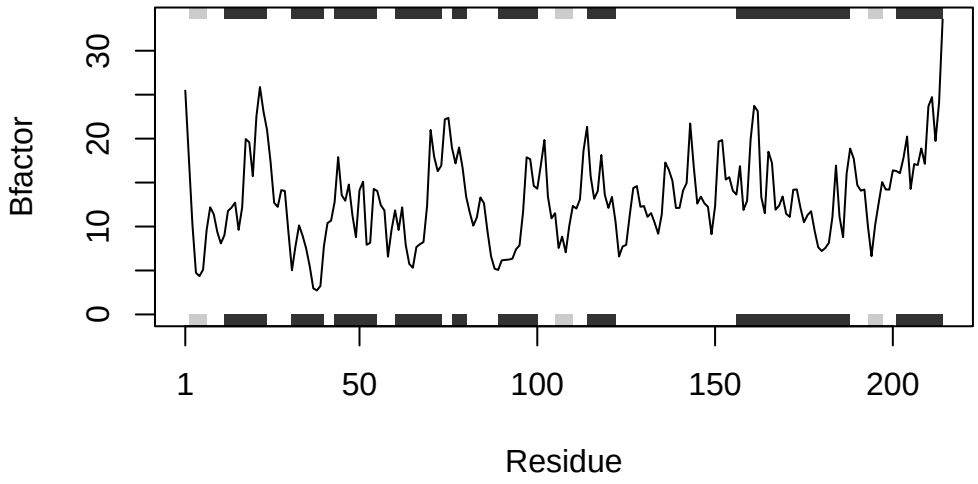
```
s1.chainA <- trim.pdb(s1, chain="A", elety="CA")
s2.chainA <- trim.pdb(s2, chain="A", elety="CA")
s3.chainA <- trim.pdb(s3, chain="A", elety="CA")
s1.b <- s1.chainA$atom$b
s2.b <- s2.chainA$atom$b
s3.b <- s3.chainA$atom$b
plotb3(s1.b, sse=s1.chainA, typ="l", ylab="Bfactor")
```




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



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



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

Breakdown of the new function

```
library(bio3d)

## Variable "x" to be substituted for any given protein PBD ID
## to be analyzed for protein drug interactions via the function

drug_prot_int<-function(x) {

  ## Extracts and reads the specified protein code ("x") from
  ## the PDB dataset to be used in the function

  rdpdb<- read.pdb(x)

  ## Takes the protein structure that was specified in the previous line of code
  ## and trims/segments the larger structure and
  ## various chain into its more essential functional regions as defined (elety="")
```

```
  segmprot<- trim.pdb(rdpdb, chain="A", elety="CA")

  ## This code identifies the B factors along the given chain of the protein
  ## that can then be related to the residues in downstream lines of code.

  bfac<- segmprot$atom$b

  ## Constructs a graphical representation of B factors versus
  ## residues along the entirety of the protein structure

  plotb3(bfac, sse=segmprot, typ="l", ylab="Bfactor")
}
```

Summary of New Function

- Overall, by utilizing my new simplified function 'drug_prot_int()' and inputting a specific PDB ID for a given protein of interest ("4AKE", "1AKE", or "1E4Y"), one is able to extract the protein data from the PDB database and subsequently compute/analyze the relationship between B Factors and residues along the specified protein structure. The outputted graph provides an efficient visual representation of how both B Factors and Residues change in relation to each other, and by altering the inputted protein PDB ID in the function, one can quickly analyze another protein of interest without laborous copying of code.

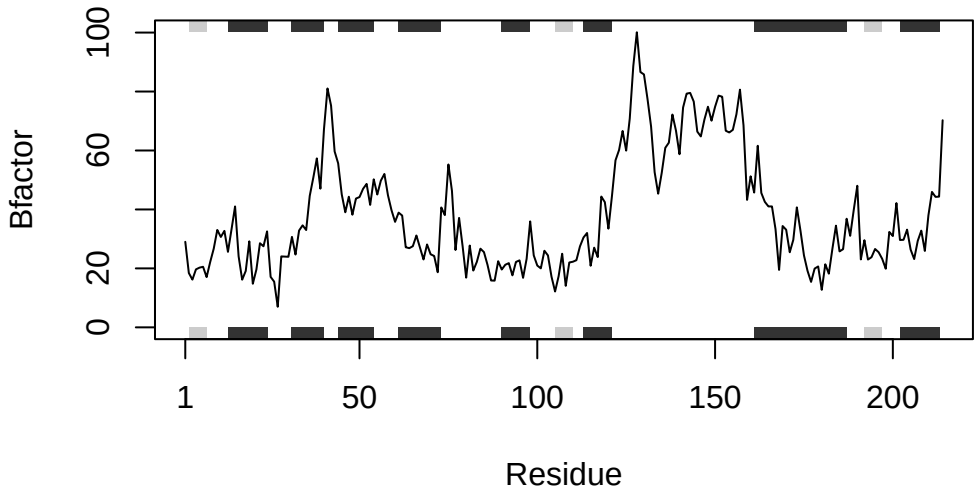
Evidence of Properly Working Function

-Graph of 4AKE with New Function

```
drug_prot_int("4AKE")
```

```
  Note: Accessing on-line PDB file

Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
C:\Users\iangu\AppData\Local\Temp\RtmpgD8RMb/4AKE.pdb exists. Skipping download
```



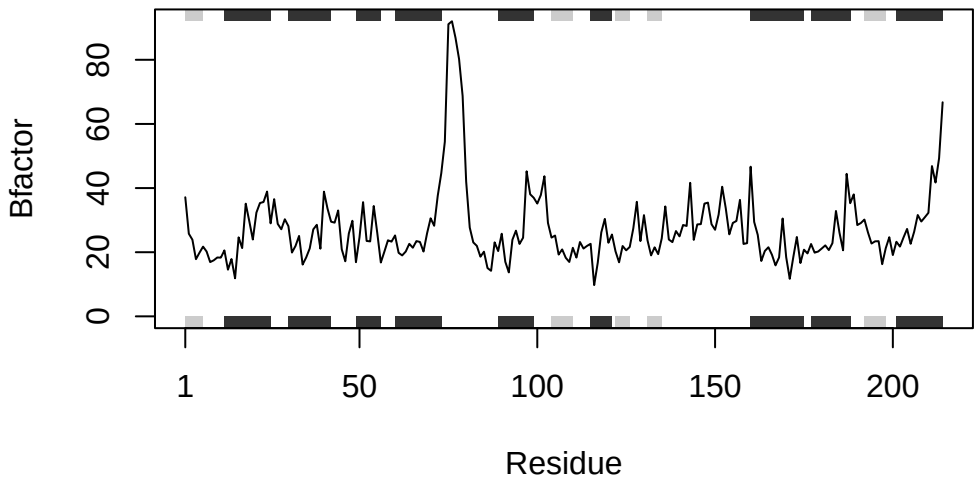
-Graph of lAKE with New Function

```
drug_prot_int("1AKE")
```

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

```
Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
C:\Users\iangu\AppData\Local\Temp\RtmpgD8RMb/1AKE.pdb exists. Skipping download
```

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

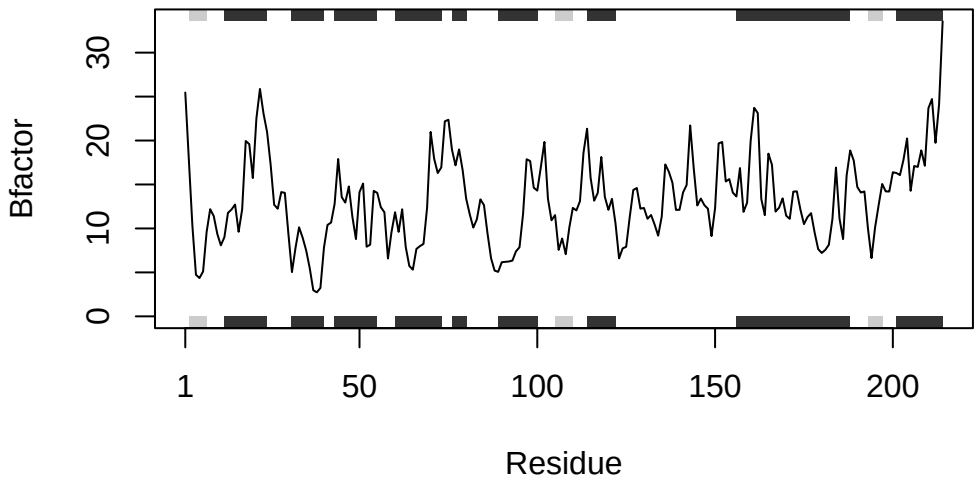


-Graph of 1E4Y with New Function

```
drug_prot_int("1E4Y")
```

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

```
Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
C:\Users\iangu\AppData\Local\Temp\RtmpgD8RMb/1E4Y.pdb exists. Skipping download
```


Bfactor
0
10
20
30
1
50
100
150
200
Residue