HW_Q6

Nicole Lin (PID: A17826971)

Table of contents

1. First, we load the bio3d package using library(bio3d). Then, we have R read the PDB files using `read.pdb()`.

```
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
```

2. We trim each protein into only chain A and alpha carbon atoms using `trim.pbd()`. Then, we extract B-factor from alpha carbon.

```
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
```

3. Then, we combine the B-factor vectors into a protein matrix using `rbind()`.

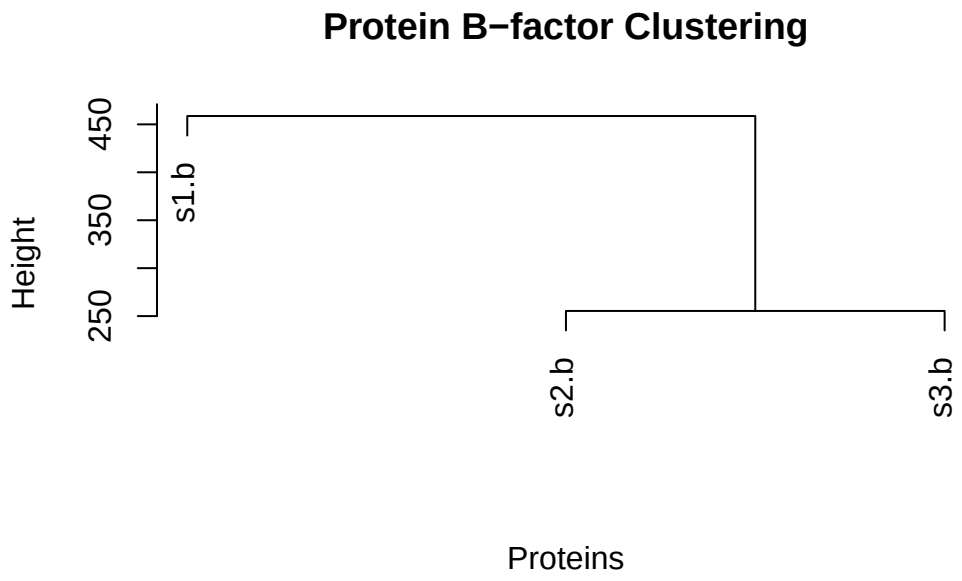
```
protein_matrix <- rbind(s1.b, s2.b, s3.b)
```

4. Make a protein cluster using `hclust()`.

```
protein_dist <- dist(protein_matrix)
hc <- hclust(protein_dist)
```

5. Then, we plot the cluster diagram using `plot()`.

```
plot(hc,
     main = "Protein B-factor Clustering",
     xlab = "Proteins",
     sub = "")
```



6. To generalize this process, we wrap it into a function called `analyze_proteins()`. We use `lapply()` to apply the trimming and B-factor extraction to each protein. `do.call()` combines the results into a matrix.

```
analyze_proteins <- function(pdb_objects) {
  bfactor_list <- lapply(pdb_objects, function(pdb) {
    chainA <- trim.pdb(pdb, chain="A", elety="CA")
    chainA$atom$b
  })
```

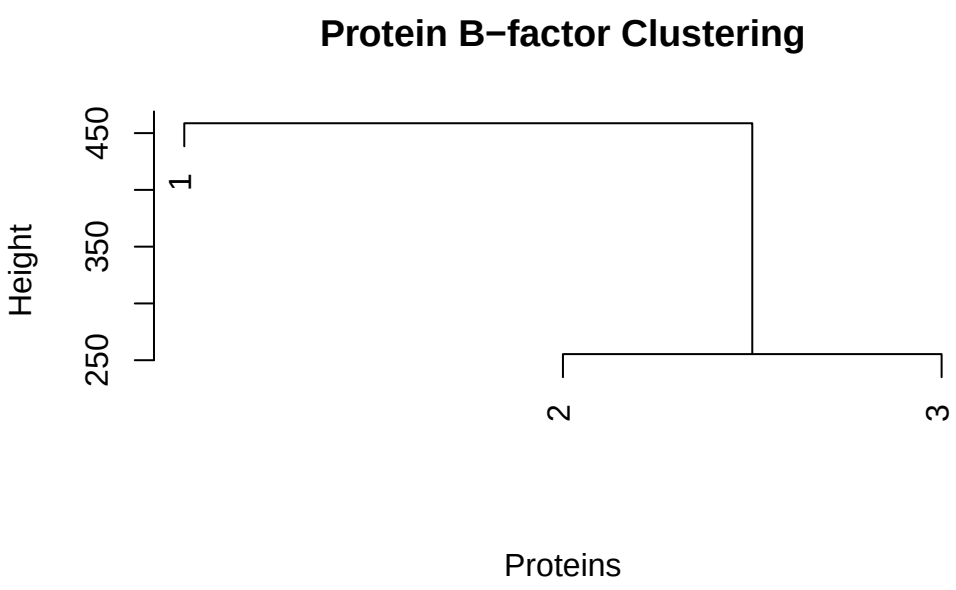
```
  protein_matrix <- do.call(rbind, bfactor_list)
  rownames(protein_matrix) <- names(bfactor_list)

  protein_dist <- dist(protein_matrix)
  hc <- hclust(protein_dist)

  plot(hc,
       main = "Protein B-factor Clustering",
       xlab = "Proteins",
       sub = "")
}
```

7. To run the function, we first create a list of protein objects like `protein_list <- list(s1, s2, s3)` and then call `analyze_proteins(protein_list)`. The function will automatically produce a dendrogram showing how similar the proteins are based on their B-factors.

```
protein_list <- list(s1, s2, s3)
analyze_proteins(protein_list)
```


Protein B−factor Clustering
Height
450
350
250
1
2
3
Proteins