Computing the feature of a protein sequence

A sequence can provide certain information about the protein. Its properties include the amino acid composition, hydrophobicity, and so on. To get the features from the sequence data, perform the following steps:

1. Install and load the **protr** and **bio3d** packages

```
> install.packages("protr")
> install.packages("bio3d")
> library(bio3d)
> library(protr)
```

2. To retrieve the sequence from PDB, type the following commands::

```
> pdb1 <- read.pdb("1BG2")
> s1 <- aa321(pdb1$seqres)
```

3. The protR package needs the sequence as a sequence string. Therefore, first collapse a vector into a string

```
> s1 <- paste(s1, sep="",collapse="")
```

   Alternatively, use the following commands:

```
> library(seqinr)
> s1 <- c2s(s1)
```

4. To obtain the amino acid composition, using the following *extractAAC* function:

```
> extractAAC(s1)
```

   Q1: What is the sequence composition of 1BG2?

Handling the PDB file

Perform the following steps to work with the PDB file

1. Load the protr and bio3d libraries

```
> library(protr)
> library(bio3d)
```

2. Read a PDB file for the human kinesin domain with the ID as 1BG2 from PDB

```
> pdb <- read.pdb("1BG2")
```

3. Check the different parts and components of the created PDB object

```
> class(pdb)
> attributes(pdb)
> head(pdb)
> head(pdb$atom[, c("x","y","z")])
```

4. To get the C-alpha coordinates in the protein molecule, simply access the corresponding record of the PDB object by typing the following command:

```
> head(pdb$atom[pdb$calpha, c("resid", "elety", "x","y","z")])
```

5. To access the sequence from the PDB object, access the sequence record as follows:

```
> aa321(pdb$seqres)
```

6. Use the *write.pdb* function to write a PDB file to a file. The *read.pdb* function can be used to read the file into the R session again

```
> write.pdb(pdb, file="myPDBfile.pdb")
> read.pdb("myPDBfile.pdb")
```

Ramachandran plot

The Ramachandran plot is a way to visualize the dihedral angle ϕ against ψ of amino acid residues in the protein structure. In principle, it shows the possible conformations of the and angles for amino acid residue in a protein due to steric constrains. It also serves the purpose of structure validation
Visualize the Ramachandran plot for the protein 1BG2

1. Load the **bio3d** library into the R session

```
> library(bio3d)
```

2. Read the PDB file

```
> pdb <- read.pdb("1BG2")
```

3. Extract the torsion angles, ϕ and ψ , with the following function:

```
> tor <- torsion.pdb(pdb)
```

4. Plot the components of the torsion angles with a plot function :

```
> plot(tor$phi, tor$psi, main="(A) Ramachandran plot 1BG2")
```

5. Now, enhance the visualization by slightly sophisticated code. First extract the angles

```
> scatter_psi <- tor$psi
> scatter_phi <- tor$phi
```

6. Then create the color of the contours of the new Ramachandran plot

```
> library(RColorBrewer) # load RColourBrewer package
> k <- 10 # define number of colours
> my.cols <- rev(brewer.pal(k, "RdYlBu")) # Brew color pallette
```

7. With these colors and torsion angles in place, create a better looking Ramachandran plot using the following command:

```
> smoothScatter(x=scatter_phi, y=scatter_psi,
      colramp=colorRampPalette(my.cols), xlim=c(-180,180), ylim=c(-
      180,180),xlab="Phi", ylab="Psi", main="(B) Ramachandran plot
      1BG2", pch=19, cex=0.00)
```

Q2 Save the two Ramachandran plots. Base on the plots, can you predict the type of 1BG2 (Alpha protein, Beta protein or Alpha and Beta protein)? Why?