Class06 HW

AUTHOR
Dylan Mullaney

I want function that will read files based on a letter and numerical input. If I put in s1, I want it to read back the results from “4AKE”. Same for corresponding code for s2 and s3.

Q1. Each read.pbd() function is reading a protein data bank file and returns a line that details that it is accessing an online.

```
#library(bio3d)
#s1 <- read.pdb("4AKE") # kinase with drug
#s2 <- read.pdb("1AKE") # kinase no drug
#s3 <- read.pdb("1E4Y") # kinase with drug

#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")
```

The trim.pdb() function pulls out the data of interest based on what we are asking it to trim. You are creating a new subset of items based on a larger structure.

```
#?trim.pdb()
```

Q6. Simplify reading files

```
library(bio3d)

read_kinase <- function(num)  {
  pbds <- c("4AKE", "1AKE", "1E4Y")
  pdb_data <- read.pdb (pbds[num])
  return(pdb_data)
}

read_kinase(1)
```

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

```
 Call:  read.pdb(file = pbds[num])

   Total Models#: 1
     Total Atoms#: 3459,  XYZs#: 10377  Chains#: 2  (values: A B)

     Protein Atoms#: 3312  (residues/Calpha atoms#: 428)
     Nucleic acid Atoms#: 0  (residues/phosphate atoms#: 0)

     Non-protein/nucleic Atoms#: 147  (residues: 147)
     Non-protein/nucleic resid values: [ HOH (147) ]

   Protein sequence:
      MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVT
      DELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRI
      VGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIG
      YYSKEAEAGNTKYAKVDGTKPVAEVRADLEKILGMRIILLGAPGA...<cut>...KILG

+ attr: atom, xyz, seqres, helix, sheet,
        calpha, remark, call
```

```
read_kinase (2)
```

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


 Call:  read.pdb(file = pbds[num])

   Total Models#: 1
     Total Atoms#: 3804,  XYZs#: 11412  Chains#: 2  (values: A B)

     Protein Atoms#: 3312  (residues/Calpha atoms#: 428)
     Nucleic acid Atoms#: 0  (residues/phosphate atoms#: 0)

     Non-protein/nucleic Atoms#: 492  (residues: 380)
     Non-protein/nucleic resid values: [ AP5 (2), HOH (378) ]

   Protein sequence:
      MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVT
      DELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRI
      VGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIG
      YYSKEAEAGNTKYAKVDGTKPVAEVRADLEKILGMRIILLGAPGA...<cut>...KILG

+ attr: atom, xyz, seqres, helix, sheet,
        calpha, remark, call
```

```
read_kinase (3)
```

```
 Note: Accessing on-line PDB file


 Call:  read.pdb(file = pbds[num])

   Total Models#: 1
     Total Atoms#: 3428,  XYZs#: 10284  Chains#: 2  (values: A B)

     Protein Atoms#: 3314  (residues/Calpha atoms#: 428)
     Nucleic acid Atoms#: 0  (residues/phosphate atoms#: 0)

     Non-protein/nucleic Atoms#: 114  (residues: 2)
     Non-protein/nucleic resid values: [ AP5 (2) ]

   Protein sequence:
      MRIILLGALVAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVT
      DELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRI
      VGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIG
      YYSKEAEAGNTKYAKVDGTKPVAEVRADLEKILGMRIILLGALVA...<cut>...KILG

+ attr: atom, xyz, seqres, helix, sheet,
        calpha, remark, call
```

Simplify trimming pdb files

```
trim_chain <- function (read_kinase) {

  trimmed_pdb <- trim.pdb (read_kinase, chain = "A", elety= "CA")
  return(trimmed_pdb)
}

trim_chain(read_kinase(1))
```

```
  Note: Accessing on-line PDB file

Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
/var/folders/k3/fzvl598s4x981fkzz132rkkr0000gn/T//RtmpM7To23/4AKE.pdb exists.
Skipping download


 Call:  trim.pdb(pdb = read_kinase, chain = "A", elety = "CA")

   Total Models#: 1
     Total Atoms#: 214,  XYZs#: 642  Chains#: 1  (values: A)

     Protein Atoms#: 214  (residues/Calpha atoms#: 214)
     Nucleic acid Atoms#: 0  (residues/phosphate atoms#: 0)

     Non-protein/nucleic Atoms#: 0  (residues: 0)
     Non-protein/nucleic resid values: [ none ]
```

```
    Protein sequence:
       MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVT
       DELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRI
       VGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIG
       YYSKEAEAGNTKYAKVDGTKPVAEVRADLEKILG

+ attr: atom, helix, sheet, seqres, xyz,
        calpha, call
```

```
trim_chain(read_kinase(2))
```

```
  Note: Accessing on-line PDB file

Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
/var/folders/k3/fzvl598s4x981fkzz132rkkr0000gn/T//RtmpM7To23/1AKE.pdb exists.
Skipping download

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


 Call:  trim.pdb(pdb = read_kinase, chain = "A", elety = "CA")

   Total Models#: 1
     Total Atoms#: 214,  XYZs#: 642  Chains#: 1  (values: A)

     Protein Atoms#: 214  (residues/Calpha atoms#: 214)
     Nucleic acid Atoms#: 0  (residues/phosphate atoms#: 0)

     Non-protein/nucleic Atoms#: 0  (residues: 0)
     Non-protein/nucleic resid values: [ none ]

   Protein sequence:
      MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVT
      DELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRI
      VGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIG
      YYSKEAEAGNTKYAKVDGTKPVAEVRADLEKILG

+ attr: atom, helix, sheet, seqres, xyz,
        calpha, call
```

```
trim_chain(read_kinase(3))
```

```
  Note: Accessing on-line PDB file

Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
/var/folders/k3/fzvl598s4x981fkzz132rkkr0000gn/T//RtmpM7To23/1E4Y.pdb exists.
Skipping download


 Call:  trim.pdb(pdb = read_kinase, chain = "A", elety = "CA")
```

```
   Total Models#: 1
     Total Atoms#: 214,  XYZs#: 642  Chains#: 1  (values: A)

     Protein Atoms#: 214  (residues/Calpha atoms#: 214)
     Nucleic acid Atoms#: 0  (residues/phosphate atoms#: 0)

     Non-protein/nucleic Atoms#: 0  (residues: 0)
     Non-protein/nucleic resid values: [ none ]

   Protein sequence:
      MRIILLGALVAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVT
      DELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRI
      VGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIG
      YYSKEAEAGNTKYAKVDGTKPVAEVRADLEKILG

+ attr: atom, helix, sheet, seqres, xyz,
        calpha, call
```

Simplify accessing a subset of this data and reassigning to a new value. Accessing chain A atom, B factor column w/in atom data.

```
extract_factors <- function (trim_chain)
{

return (trim_chain$chainA$atom$b)
}

extract_factors(read_kinase(1))
```

```
  Note: Accessing on-line PDB file

Warning in get.pdb(file, path = tempdir(), verbose = FALSE):
/var/folders/k3/fzvl598s4x981fkzz132rkkr0000gn/T//RtmpM7To23/4AKE.pdb exists.
Skipping download

NULL
```

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

Simplify plotting the data

```
# plot_new <- function(b_factors, sse, y_label = "Bfactor") {
#  plotb3(b_factors, sse = sse, typ = "l", ylab = y_label)
# }

#plot_new()
#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")
```