Analysis of modified sequences using pIR

Enrique Audain

29th February 2016

In this example, we show how to perform an analysis of a dataset containing modified sequences. At the moment, pIR supports the isolectric point prediction considering two modification: N-terminal acetilation and phosphorilation. The PTM notation used is: "p" to phosphorylation, and "n" to N-terminal acetylation.

To start, take us a view to the way in which pIR compute pI for a sinlge modified sequence.

```
library(pIR)


#sequence modified at N-terminal
seqA <- "nADTAVDTTTTSEITAK"

#computing pI using Iterative method with pK set lehninger.
pIa  <- pIIterative(sequence = seqA, pkSetMethod = "lehninger")

#the result will be
pIa

## [1] 2.874

#Eliminating the contribution of N-terminal acetylation
seqB <- "ADTAVDTTTTSEITAK"

pIb <- pIIterative(sequence = seqB, pkSetMethod = "lehninger")

#the result will be
pIb

## [1] 3.7026
```

Now, we will perform a similar analysis reading sequences from a file. The file must containt two column: sequence and pIExp.

```
#reading sequence with experimental pI associated from any source
dataAcetylated <-
read.table(file="E:/$WORK/Bioinformatics/Github/pIR/data/gauci_dataset_acetylated.txt", header=TRUE,
sep="\t")

#showing the data (truncated)
head(dataAcetylated, n=10)

##                  sequence pIExp
## 1   nAAAEDTNSNVTQNPSGSDAPK  3.99
## 2        nADTAVDTTTTSEITAK  3.99
```

```
## 3       nANLTDAASLQQFDELLK  3.99
## 4        nSDTSTGDVGSVEPGVR  3.99
## 5  nAEEGIAAGGVMDVNTALPEVLK  3.99
## 6           nADQAFVTLATTDK  3.99
## 7     nSSIGTGYDLSASTFSPDGR  3.99
## 8      nAAAMDVDTPSGANSGASK  3.99
## 9             nAAQVTESDQIK  3.99
## 10            nANGAEDVVFCR  3.99

#Estimating pI with Bjell method
dataBjell <- pIBjellMultipleSequences(sequences = dataAcetylated, pkSetMethod = "expasy")

#showing the data (truncated)
head(dataBjell, n=10)

##                   sequence pIExp expasy
## 1   nAAAEDTNSNVTQNPSGSDAPK  3.99 3.4926
## 2        nADTAVDTTTTSEITAK  3.99 3.4926
## 3       nANLTDAASLQQFDELLK  3.99 3.4926
## 4        nSDTSTGDVGSVEPGVR  3.99 3.4926
## 5  nAEEGIAAGGVMDVNTALPEVLK  3.99 3.5041
## 6           nADQAFVTLATTDK  3.99 3.5639
## 7     nSSIGTGYDLSASTFSPDGR  3.99 3.5639
## 8      nAAAMDVDTPSGANSGASK  3.99 3.5639
## 9             nAAQVTESDQIK  3.99 3.6662
## 10            nANGAEDVVFCR  3.99 3.6662
```

Using pIR it is possible also to evaluate the impact of a modification such as phosphorylation and N-terminal acetylation on the isoelectric point prediction using any method. Here, we show how to do such analysis from two files containing the sequences modified and without modification. Both files must have the same structure (number and name of columns, etc.)

```
#Comparing effect of N-terminal acetylation on pI prediction

#Reading file containing sequence whitout acetylation
data_non_acetylated <-
read.table(file="E:/$WORK/Bioinformatics/Github/pIR/data/gauci_dataset_non_acetylated.txt",
header=TRUE, sep="\t")

#Estimating pI with Bjell method
dat1 <- pIBjellMultipleSequences(sequences = data_non_acetylated, pkSetMethod = "expasy")

#showing the data (truncated)
head(dat1, n=5)

##                 sequence pIExp expasy
## 1  AAAEDTNSNVTQNPSGSDAPK  3.99 4.0281
## 2       ADTAVDTTTTSEITAK  3.99 4.0281
## 3      ANLTDAASLQQFDELLK  3.99 4.0281
## 4       SDTSTGDVGSVEPGVR  3.99 4.0276
## 5 AEEGIAAGGVMDVNTALPEVLK  3.99 4.0020

#removing sequence column
dat1 <- removeFirstColumn(dat1)

#Reading file containing sequence acetylated
data_acetylated <-
read.table(file="E:/$WORK/Bioinformatics/Github/pIR/data/gauci_dataset_acetylated.txt", header=TRUE,
sep="\t")

#Estimating pI with Bjell method
```

```
dat2 <- pIBjellMultipleSequences(sequences = data_acetylated, pkSetMethod = "expasy")

#showing the data (truncated)
head(dat2, n=5)

##                  sequence pIExp expasy
## 1  nAAAEDTNSNVTQNPSGSDAPK   3.99 3.4926
## 2       nADTAVDTTTTSEITAK   3.99 3.4926
## 3      nANLTDAASLQQFDELLK   3.99 3.4926
## 4       nSDTSTGDVGSVEPGVR   3.99 3.4926
## 5 nAEEGIAAGGVMDVNTALPEVLK   3.99 3.5041

#removing sequence colum
dat2 <- removeFirstColumn(dat2)

#getting plots. If there are several
plots <- plotPTMShiftOverallData(dat1, dat2)
```

Output graph visualyzing the correlation between predicted and experimental pI both acetylated and non-acetylated data.