pIR: An *R* package for isoelectric point prediction based on amino acid sequences

Yasset Perez-Riverol and Enrique Audain

February 27, 2016

Abstract

Accurate estimation of the isoelectric point value (pI) based on the amino acid sequence becomes critical to perform proteomics experiments. Also, it is one of the most useful electrostatic properties to study peptides and proteins. Different methods has been proposed to compute the theoretical isoelectric point of peptides and proteins using several pK sets [1, 2]. This vignette provides a brief overview of the available interface and functionality as well as a short use case.

Keywords: proteomics, peptides, proteins, electrophoresis, mass spectrometry, isoelectric point, tutorial.

Contents

1 Compute isoelectric protein of peptides and proteins

1.1 Iterative Method

Isoelectric point can be defined as the point in a titration curve at which the net surface charge of a protein or peptide equals to zero. The called **"Iterative Method"** to predict the iseoelct point only considers the contribution of individual pKa values to the Henderson-Hasselbach equation. Keeping in main this we can use Henderson-Hasselbach equation to calculate protein charge in certain pH:

- for negative charged macromolecules:

$$pH = -1/(1 + 10^{(pK_n - pH)}) \text{ Equation}(1)$$

where pK_n is the acid dissociation constant of negatively charged amino acid

- **for positive charged macromolecules**:

$$pH = 1/(1 + 10^{(pH - pK_p)}) \text{ Equation(2)}$$

where pK_p is the acid dissociation constant of positively charged amino acid

The most import moment during isoelectric point determination is usage of appropriate pK values. Unfortunately, there is no agreement in this matter. Each source gives different pKs. Some of them are presented pK that are available for the Iterative method in **pIR** are:

Table 1: pK values for the Iterative method

	Amino acid								
pK value set	NH2	COOH	C	D	E	H	K	R	Y
emboss	8.6	3.6	8.5	3.9	4.1	6.5	10.8	12.5	10.1
DTASelect	8.0	3.1	8.5	4.4	4.4	6.5	10.0	12.0	10.0
solomon	9.6	2.4	8.3	3.9	4.3	6.0	10.5	12.5	10.1
sillero	8.2	3.2	9.0	4.0	4.5	6.4	10.4	12.0	10.0
rodwell	8.0	3.1	8.33	3.68	4.25	6.0	11.5	11.5	10.07
patrickios	11.2	4.2	-	4.2	4.2	-	11.2	11.2	-
lehninger	9.69	2.34	8.33	3.86	4.25	6.0	10.5	12.4	10.0
grimsley	7.7	3.3	6.8	3.5	4.2	6.6	10.5	12.04	10.3

1.1.1 Calculate isoelectric point

```
# compute the isoelectric point using solomon method
 library(pIR)
 seq <- "GLPRKILCAIAKKKGKCKGPLKLVCKC"
 pI.value <- pIIterative(sequence = seq, pkSetMethod = "solomon")
 pI.value
# 10.526

# ------------------------------------------------------------------ #

# compute the isoelectric point using  rodwell method
 library(pIR)
 seq <- "GLPRKILCAIAKKKGKCKGPLKLVCKC"
 pI.value <- pIIterative(sequence = seq, pkSetMethod = "rodwell")
 pI.value
# 11.404

# ------------------------------------------------------------------ #

# compute the isoelectric point using  emboss method
 library(pIR)
 seq <- "GLPRKILCAIAKKKGKCKGPLKLVCKC"
 pI.value <- pIIterative(sequence = seq, pkSetMethod = "emboss")
 pI.value
# 10.801

# ------------------------------------------------------------------ #
# compute the isoelectric point using  lehninger method
 library(pIR)
 seq <- "GLPRKILCAIAKKKGKCKGPLKLVCKC"
 pI.value <- pIIterative(sequence = seq, pkSetMethod = "lehninger")
 pI.value
# 10.530

# ------------------------------------------------------------------- #

# compute the isoelectric point using  grimsley method
 library(pIR)
 seq <- "GLPRKILCAIAKKKGKCKGPLKLVCKC"
 pI.value <- pIIterative(sequence = seq, pkSetMethod = "grimsley")
 pI.value
# 10.495

# ------------------------------------------------------------------- #

# compute the isoelectric point using  patrickios method
 library(pIR)
 seq <- "GLPRKILCAIAKKKGKCKGPLKLVCKC"
 pI.value <- pIIterative(sequence = seq, pkSetMethod = "patrickios")
 pI.value
# 11.200

# ------------------------------------------------------------------- #

# compute the isoelectric point using  DtaSelect method
 library(pIR)
 seq <- "GLPRKILCAIAKKKGKCKGPLKLVCKC"
 pI.value <- pIIterative(sequence = seq, pkSetMethod = "DtaSelect")
 pI.value
# 10.025
```

1.1.2 Compute pI using all the pk sets

```
# compute the isoelectric point using  all pK sets
library(pIR)
seq <- "GLPRKILCAIAKKKGKCKGPLKLVCKC"
pIValue <- computeAllIterativeValues(seq = seq)
pIValue
#       method values
# 1    solomon 10.526
# 2    rodwell 11.404
# 3     emboss 10.801
# 4  lehninger 10.530
# 5   grimsley 10.495
# 6 patrickios 11.200
# 7  DtaSelect 10.025
```

1.2 Bjellqvist Method

The Bjellquivst method [2] is based on the determination of the pK differences between closely related amino acids, and it was the first algorithm that proposed a different pKa value depending on the amino acid position in the sequence. The current version of the pIR supports four different pKsets: **bjell** [2], **calibrated** [3], **skoog** and **expasy** http://expasy.org/

1.2.1 Calculate isoelectric point

```
# -------------------------------------------------------------------- #

# compute the isoelectric point using  bjell pk set
 library(pIR)
 seq <- "AGAAPYVQAFDSLLAGPVAE"
 pI.value <- pIBell(sequence = seq, pkSetMethod = "bjell")
 pI.value
# 3.04

# -------------------------------------------------------------------- #

# compute the isoelectric point using  calibrated pk set
 library(pIR)
 seq <- "AGAAPYVQAFDSLLAGPVAE"
 pI.value <- pIBell(sequence = seq, pkSetMethod = "calibrated")
 pI.value
# 4.09

# -------------------------------------------------------------------- #

# compute the isoelectric point using  skoog pk Set
 library(pIR)
 seq <- "AGAAPYVQAFDSLLAGPVAE"
 pI.value <- pIBell(sequence = seq, pkSetMethod = "skoog")
 pI.value
# 2.93

# -------------------------------------------------------------------- #

# compute the isoelectric point using  expasy pK set
 library(pIR)
 seq <- "AGAAPYVQAFDSLLAGPVAE"
 pI.value <- pIBell(sequence = seq, pkSetMethod = "expasy")
 pI.value
# 4.09
```

1.2.2 Compute pI using all the pk sets

```
# compute the isoelectric point using all the pK sets
seq <- "AGAAPYVQAFDSLLAGPVAE"
pIValue <- computeAllBjellValues(seq = seq)
pIValue
# method values
```

```
#1     expasy   4.09
#2      skoog   2.93
#3 calibrated   4.09
#4      bjell   3.04
```

1.3 SVM Method

2 Session info

Here is the output of `sessionInfo` on the system on which this document was compiled:

```
> toLatex(sessionInfo())
```

- R version 3.2.3 (2015-12-10), `x86_64-apple-darwin13.4.0`
- Locale: `C/en_GB.UTF-8/en_GB.UTF-8/C/en_GB.UTF-8/en_GB.UTF-8`
- Base packages: base, datasets, grDevices, graphics, methods, stats, utils
- Other packages: pIR 0.99.0
- Loaded via a namespace (and not attached): BiocStyle 1.8.0, MASS 7.3-45, Matrix 1.2-3, MatrixModels 0.4-1, Rcpp 0.12.3, SparseM 1.7, car 2.1-1, caret 6.0-64, codetools 0.2-14, colorspace 1.2-6, corrplot 0.73, devtools 1.10.0, digest 0.6.9, foreach 1.4.3, ggplot2 2.0.0, grid 3.2.3, gtable 0.1.2, iterators 1.0.8, kernlab 0.9-23, knitr 1.12.3, lattice 0.20-33, lme4 1.1-11, lsr 0.5, magrittr 1.5, memoise 1.0.0, mgcv 1.8-11, minqa 1.2.4, munsell 0.4.3, nlme 3.1-124, nloptr 1.0.4, nnet 7.3-12, parallel 3.2.3, pbkrtest 0.4-6, plyr 1.8.3, quantreg 5.21, reshape2 1.4.1, roxygen2 5.0.1, scales 0.3.0, splines 3.2.3, stats4 3.2.3, stringi 1.0-1, stringr 1.0.0, tools 3.2.3, withr 1.0.1

References

[1] Yasset Perez-Riverol, Enrique Audain, Aleli Millan, Yassel Ramos, Aniel Sanchez, Juan Antonio Vizcaíno, Rui Wang, Markus Müller, Yoan J Machado, Lazaro H Betancourt, et al. Isoelectric point optimization using peptide descriptors and support vector machines. *Journal of proteomics*, 75(7):2269–2274, 2012.

[2] Bengt Bjellqvist, Graham J Hughes, Christian Pasquali, Nicole Paquet, Florence Ravier, Jean-Charles Sanchez, Séverine Frutiger, and Denis Hochstrasser. The focusing positions of polypeptides in immobilized ph gradients can be predicted from their amino acid sequences. *Electrophoresis*, 14(1):1023–1031, 1993.

[3] Sharon Gauci, Bas Van Breukelen, Simone M Lemeer, Jeroen Krijgsveld, and Albert JR Heck. A versatile peptide pi calculator for phosphorylated and n-terminal acetylated peptides experimentally tested using peptide isoelectric focusing. *Proteomics*, 8(23-24):4898–4906, 2008.