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

Yasset Perez-Riverol

June 17, 2015

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, 3]. 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
```

1.2 Bjellqvist Method

1.3 Co-factor Method

1.4 SVM Method

2 Session info

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

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

- R version 3.1.3 (2015-03-09), `x86_64-apple-darwin10.8.0`
- Locale: `en_GB.UTF-8/en_GB.UTF-8/en_GB.UTF-8/C/en_GB.UTF-8/en_GB.UTF-8`
- Base packages: base, datasets, graphics, grDevices, methods, stats, utils
- Loaded via a namespace (and not attached): BiocStyle 1.2.0, tools 3.1.3

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] Benjamin J Cargile, Joel R Sevinsky, Amal S Essader, Jerry P Eu, and James L Stephenson. Calculation of the isoelectric point of tryptic peptides in the ph 3.5–4.5 range based on adjacent amino acid effects. *Electrophoresis*, 29(13):2768–2778, 2008.

[3] 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.