

Points isoélectriques théoriques avec le logiciel R

14 novembre 2006

Résumé

La fonction `computePI()` du module `seqinr` permet de calculer simplement le pI théorique d'une séquence d'acides aminés avec R.

Table des matières

1	Données concernant les acides α-aminés	1
2	Point isoélectrique d'une séquence d'acides aminés	2
2.1	Charge d'une fonction ionisable en fonction du pH	2
2.1.1	Fonction amine (basique)	2
2.1.2	Fonction carboxyle (acide)	2
2.2	Tracé des courbes $q^{+/-} = f(pH)$ avec R	2
3	Point isoélectrique (pI) d'une séquence d'acides aminés	3
3.1	Définition	3
3.2	Calcul avec la fonction <code>computePI()</code>	3

1 Données concernant les acides α -aminés

Pour des acides aminés engagés dans une séquence peptidique, les fonctions susceptibles d'être chargées (+ ou -) sont l'amine n -terminale, le carboxyle c -terminal ainsi que les fonctions ionisables des chaînes latérales de certains d'entre-eux.

TABLEAU 1 : Valeurs de pK à 25°C des acides aminés rencontrés dans les protéines (d'après [5]).

α -amino acide	$pK_{\alpha-COOH}$	$pK_{\alpha-NH_3^+}$	pK Chaîne latérale
Alanine (A)	2,35	9,87	-
Arginine (R)	1,82	8,99	12,48 (guanidino)
Asparagine (N)	2,1	8,84	-
Acide aspartique (D)	1,99	9,90	3,90 ($\beta - COOH$)
Cystéine (C)	1,92	10,78	8,33 (sulfhydryle)
Acide glutamique (E)	2,10	9,47	4,07 ($\gamma - COOH$)
Glutamine (Q)	2,17	9,13	-
Glycine (G)	2,35	9,78	-
Histidine (H)	1,80	9,33	6,04 (imidazole)
Isoleucine (I)	2,32	9,76	-
Leucine (L)	2,33	9,74	-
Lysine (K)	2,16	9,18	10,79 ($\epsilon - NH_3^+$)
Méthionine (M)	2,13	9,28	-
Phénylalanine (F)	2,16	9,18	-
Proline (P)	2,95	10,65	-
Sérine (S)	2,19	9,21	-
Thréonine (T)	2,09	9,10	-
Tryptophane (W)	2,43	9,44	-
Tyrosine (Y)	2,20	9,11	10,13 (phénol)
Valine (V)	2,29	9,74	-

2 Point isoélectrique d'une séquence d'acides aminés

2.1 Charge d'une fonction ionisable en fonction du pH

2.1.1 Fonction amine (basique)

La fraction chargée q^+ peut s'exprimer en fonction du pH :

Réaction d'ionisation : $NH_2 + H^+ \rightleftharpoons NH_3^+$

En appelant K la constante de dissociation et α l'avancement de la réaction :

$$K = \frac{[NH_3^+]}{[NH_2].[H^+]}$$

$$\frac{[NH_2]}{[NH_3^+]} = \frac{(1 - \alpha)}{\alpha} = 10^{pH - pK}$$

$$\alpha = q_{pH}^+ = \frac{1}{1 + 10^{pH - pK}}$$

2.1.2 Fonction carboxyle (acide)

De la même manière, la fraction chargée q^- est une fonction du pH :

Réaction d'ionisation : $COOH \rightleftharpoons COO^- + H^+$

$$K = \frac{[COO^-].[H^+]}{[COOH]}$$

$$\frac{[COO^-]}{[COOH]} = \frac{\alpha}{(1 - \alpha)} = 10^{pH - pK}$$

$$\alpha = q_{pH}^- = \frac{10^{pH - pK}}{1 + 10^{pH - pK}}$$

2.2 Tracé des courbes $q^{+/-} = f(pH)$ avec R

Les instructions en bleu sont à saisir dans la console de R, soit en mode interactif, soit en mode différé¹.

```
qplus <- function(pH, pK) {1/(1 + 10^(pH - pK))}
qmoins <- function(pH, pK) {10^(pH - pK)/(1 + 10^(pH - pK))}
pH <- seq(from = 0, to = 14, length = 140)
plot(x = pH, y = sapply(pH, qplus, 12.48), xlab = "pH", ylab = "Charge q",
     las = 1, type = "l", lty = 1, col = "red", main = "Charge en fonction du pH")
legend(0, 0.45, legend = c("Arginine pK = 12,48", "Histidine pK = 6,04",
    "Lysine pK = 10,79"), col = c("red", "red", "red"), lty = c(1,
    2, 3))
legend(0, 0.75, legend = c("Aspartate pK = 3,90", "Cysteine pK = 8,33",
    "Glutamate pK = 4,07", "Tyrosine pK = 10,13"), col = c("blue",
    "blue", "blue", "blue"), lty = c(1, 2, 3, 4))
lines(pH, sapply(pH, qplus, 6.04), lty = 2, col = "red")
lines(pH, sapply(pH, qplus, 10.79), lty = 3, col = "red")
lines(pH, sapply(pH, qmoins, 3.9), lty = 1, col = "blue")
lines(pH, sapply(pH, qmoins, 8.33), lty = 2, col = "blue")
lines(pH, sapply(pH, qmoins, 4.07), lty = 3, col = "blue")
lines(pH, sapply(pH, qmoins, 10.13), lty = 4, col = "blue")
```

¹L'aide des différentes fonctions s'obtient en tapant ?fonction, ?plot, ?lines ou encore ?legend dans la console.

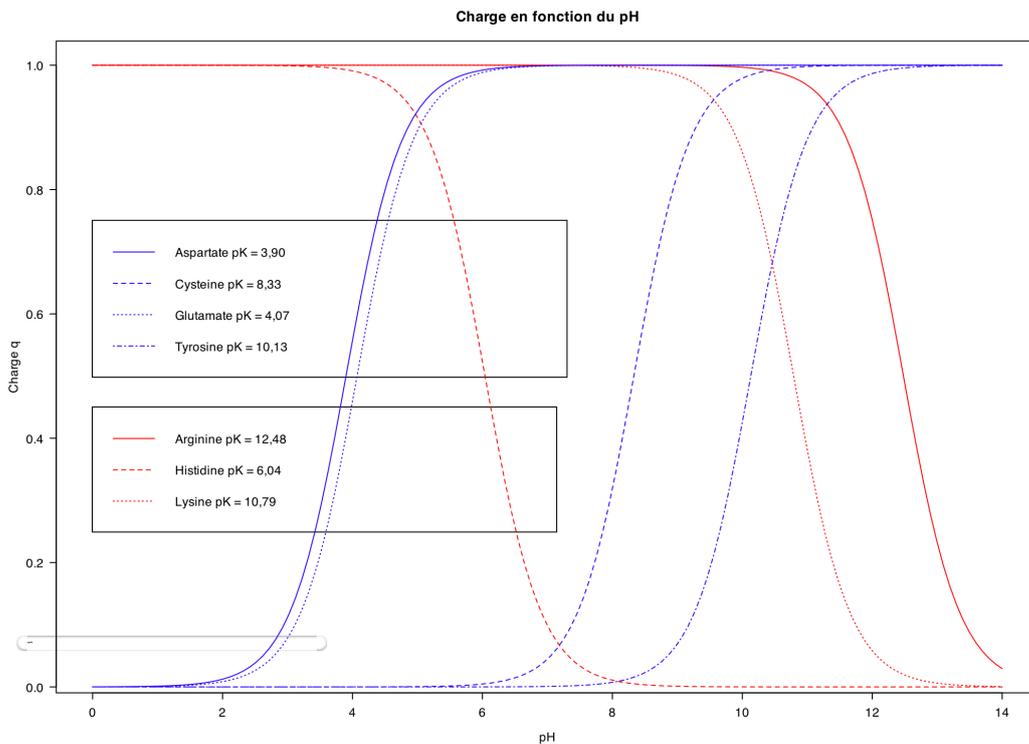


FIG. 1 – Tracé de la charge en fonction du pH avec R. Les valeurs de pK sont celles données par le TABLEAU 1, d'après [5]. Les courbes $q^+ = f(pH)$ sont des fonctions décroissantes du pH tandis les courbes $q^- = f(pH)$ sont des fonctions croissantes du pH. Quelle que soit la fonction prise en compte, la demi-charge ($q = \frac{1}{2}$), est obtenue au point d'inflexion, lorsque $pH = pK$. Dans cette zone (autour du pK), la charge varie très rapidement et doit donc être calculée avec une précision suffisante.

3 Point isoélectrique (pI) d'une séquence d'acides aminés

3.1 Définition

Le pI d'une protéine est le pH auquel la somme des charges positives égale les charges négatives. Si une protéine est constituée de n acides aminés, en appelant $q_i^+(pH)$ ou $q_i^-(pH)$ les charges individuelles éventuellement présentées par chaque acide aminé, le pI est la valeur de pH solution de l'équation :

$$\sum_{i+} q_i^+(pH) = \sum_{i-} q_i^-(pH)$$

3.2 Calcul avec la fonction `computePI()`

La fonction `computePI()` du module `seqinr` [2] permet de calculer simplement le pI théorique² d'une séquence d'acides aminés avec R³.

L'exemple suivant montre comment calculer le pI de séquence : *Met-Pro-Arg-Leu-Leu-Lys-Lys-Lys-Lys*.

```
library(seqinr)
ma_seq <- c("M", "P", "R", "L", "L", "K", "K", "K", "K")
ma_seq
```

```
[1] "M" "P" "R" "L" "L" "K" "K" "K" "K"
```

```
computePI(ma_seq)
```

```
[1] 11.33363
```

²En ne tenant pas compte d'éventuelles modifications post-traductionnelles.

³Les valeurs de pK utilisées par `computePI()` sont celles publiées par [1] sur la base de migrations électrophorétiques en gradients de pH à 15°C ou à 25°C. Elles sont accessibles depuis l'aide de la fonction `computePI()`

Références

- [1] B. et al. Bjellqvist. The focusing positions of polypeptides in immobilized ph gradients can be predicted from their amino acid sequences. *Electrophoresis*, 14 :1023–1031, 1993.
- [2] D. Charif and J.R. Lobry. Seqinr 1.0-2 : a contributed package to the r project for statistical computing devoted to biological sequences retrieval and analysis. In H.E. Roman U. Bastolla, M. Porto and M. Vendruscolo, editors, *Structural approaches to sequence evolution : Molecules, networks, populations*, Biological and Medical Physics, Biomedical Engineering. Springer Verlag, New York, 2006.
- [3] Lobry J. R. Mélanges et points isoélectriques de protéines. Fiche TD avec le logiciel R : tdr223, 09 2005.
- [4] R Development Core Team. *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, isbn 3-900051-07-0 edition, 2006.
- [5] Voet J. G. Voet D. *Biochemistry*, chapter 4. Wiley, 1990.