ELSEVIER

Computers in Biology and Medicine 36 (2006) 157–166

Computers in Biology and Medicine

www.intl.elsevierhealth.com/journals/cobm

Isoelectric point determination of proteins and other macromolecules: Oscillating method

Antonio Sillero*, Andres Maldonado

Facultad de Medicina, Departamento de Bioquímica, Instituto de Investigaciones Biomédicas Alberto Sols, CSIC/UAM, Arzobispo Morcillo 4, 28029 Madrid, Spain

Received 5 April 2004; accepted 7 September 2004

Abstract

A program written in Visual Basic has been developed to calculate the isoelectric point of proteins and other macromolecules bearing acid–basic residues. The p*I* value can be theoretically calculated with the precision required. The computer automatically supplies a representation of the charge of the protein versus pH values. The corresponding values can also be obtained, on command, in the form of table.

Keywords: Electric charge; pH; Proteins; p*I* Theoretical determination; Visual Basic; Acid base residues

1. Introduction

Part of our investigation has been focused on the theoretical determination of the isoelectric point of proteins and other macromolecules, a point at which negative and positive charges of their acid–base residues are equal. Acid–base residues have been classified as belonging to the P or N types, depending on whether they are protonated or neutral, respectively, when not dissociated (Table 1) [1–5]. As deduced from the Henderson–Hasselbach equation, the charge of a specific acid–base group depends on the pH

* Corresponding author. Tel.: +34 91 4975413; fax: +34 91 5854401.
E-mail address: antonio.sillero@uam.es (A. Sillero).

0010-4825/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiomed.2004.09.006

of the medium and its pK value, according to the following equations:

$$\text{Charge of type-N groups} = \sum_{i=1}^{n} \frac{-1}{1 + 10^{\mathrm{p}K_{\mathrm{N}i} - \mathrm{pH}}} \tag{1}$$

$$\text{Charge of type-P groups} = \sum_{i=1}^{n} \frac{-1}{1 + 10^{\mathrm{pH} - \mathrm{p}K_{\mathrm{P}i}}}. \tag{2}$$

From these equations, net charge of a protein can be calculated and represented as a function of pH values. Point of intersection of calibration curve with x-axis is pI value of protein [1 and elsewhere]. Beside this usual graphical approach, three more methods have been reported in this laboratory for direct determination of isoelectric point of a protein: Comprehensive, Simplified and Abridged (not considered here).

In the Comprehensive method, pK value of each acid–base residue present in a protein is considered (Table 1). Starting from Eqs. (1) and (2) (see above), sum of the terms describing positive or negative charges of macromolecule are equaled, in which case pH = pI. After adequate operations, pI value can be expressed as a polynomial in $10^{\mathrm{p}I}$ [4]. The degree of the polynomial is equal to the number of residues with different K values present in the macromolecule, and coefficients can be calculated from the number of acid–base residues present in the molecule and their K values [3,4]. Once coefficients have been calculated, application of any polynomial solving program affords the theoretical isoelectric point of any protein or macromolecule bearing acid–base residues [3,4].

In the Simplified procedure only 2-P types (Pa and Pb) and 2-N types (Na and Nb) of acid–base groups were considered, with the characteristics and pK values described in Table 1. The pI value of the protein is related to specific calculated ratios among the number of those residues, and can be manually determined with the help of appropriate tables [2]. Other simplified procedures for the determination of pI values have also been reported [6,7].

Table 1
Types and number of groups, chemical structure and pK values of the acid–base residues considered to be present in a protein

Type	Chemical residue	PK value	Number
N_1	terminal carboxyl	$\mathrm{p}K_{\mathrm{N1}} = 3.2$	N1
N_2	β-carboxyl of aspartate	$\mathrm{p}K_{\mathrm{N2}} = 4.0$	N2
N_3	δ-carboxyl of glutamate	$\mathrm{p}K_{\mathrm{N3}} = 4.5$	N3
N_4	thiol of cysteine	$\mathrm{p}K_{\mathrm{N4}} = 9.0$	N4
N_5	phenol of tyrosine	$\mathrm{p}K_{\mathrm{N5}} = 10.0$	N5
Na	$Na = N_1 = N_2 = N_3$	$\mathrm{p}K_{\mathrm{Na}} = 4.2$	Na
Nb	$N_b = N_3 = N_4$	$\mathrm{p}K_{\mathrm{Nb}} = 9.5$	Nb
P_1	imidazolium of histidine	$\mathrm{p}K_{\mathrm{P1}} = 6.4$	P1
P_2	terminal ammonium	$\mathrm{p}K_{\mathrm{P2}} = 8.2$	P2
P_3	ε-ammonium of lysine	$\mathrm{p}K_{\mathrm{P3}} = 10.4$	P3
P_4	guanidinium of arginine	$\mathrm{p}K_{\mathrm{P4}} = 12.0$	P4
Pa	$P_a = P_2 = P_3 = P_4$	$\mathrm{p}K\mathrm{pa} = 9.5$	Pa
Pb	$Pb = P_1$	$\mathrm{p}K\mathrm{pb} = 6.4$	Pb

Here we have focused on the problem of the determination of the pI value of a macromolecule from a different view and a method (Oscillating method) that allows both the theoretical representation of the charge of the protein as a function of pH and the determination of its pI value, with the precision required, has been developed. The program written in Visual Basic can be run in any of the usual PC.

2. Result and discussion

2.1. General aspect of the Oscillating method

The program calculates the charge of each acid–base residue (N_1, N_2, N_3, N_4, N_5, P_1, P_2, P_3, P_4) at increasing pH values, starting from zero and with increment of 0.1 pH units (Table 1 and Fig. 1). At each pH value the net charge of the protein (NQ) is consecutively calculated until a value of $\mathrm{NQ}(a) \leqslant 0$ is obtained. The pH value corresponding to this charge is called pI (a). In the very improbable case that NQ(a) is exactly zero, the pI for the protein (pI = pI (a)) is reported on the screen. If $\mathrm{NQ}(a) < 0$, the following situation arises: at pH = pI (a) the net charge of the protein is negative and at pH = pI (a) − 0.1 the charge is positive. Therefore, pI (a) represents the first significant decimal figure of the theoretical pI value of the protein (Fig. 1). For the calculation of the second decimal figure of pI, the loop starts at pH = pI (b) = $p(a) - 0.1$, with increments of 0.01 pH units (Fig. 2). A pH value pI (c) is then reached at which $\mathrm{NQ}(c) \leqslant 0$. The process can continue oscillating until precision required is reached (it will increase a decimal figure in each loop). The program described in the Appendix has been implemented as to give three significant decimal figures.

For the above calculation, the pK values of Table 1 and Formulas (1) and (2) are used. If needed, both the pK values and the required precision can be easily modified, provided the Visual Basic program is available. This method allows for calculation of pI of protein or of any macromolecule bearing acid–base groups.

The program is also suitable for determining the isoelectric point and titration curve of

(i) a neutral side chain aminoacid (glycine, alanine, valine, leucine, isoleucine, proline, phenylalanine, tryptophan, methionione, serine, threonine), (ii) a charged side chain aminoacid (aspartate, glutamate, cysteine, tyrosine, histidine, lysine, arginine), (iii) a polypeptide chain composed only of neutral side chains aminoacids. In the cases (i) and (iii), only boxes corresponding to the terminal-C and terminal-N are filled with the number 1. In the case (ii), box for proper aminoacid and those for terminal-C and terminal-N are filled with number 1.

2.2. Other characteristics of the program

The main screen (Fig. 3) presents an intuitive interface in which the name of the protein and number of its charged residues could be introduced. After pressing the calculation button, the computer automatically supplies both the pI value of the protein with three decimal figures and a graphic representation of the charge of the protein vs. the pH value. In addition, the user may ask for (i) a hard copy of the main screen, (ii) a table of the charge of the protein vs. pH values or (iii) a new calculation. Both the Oscillating and

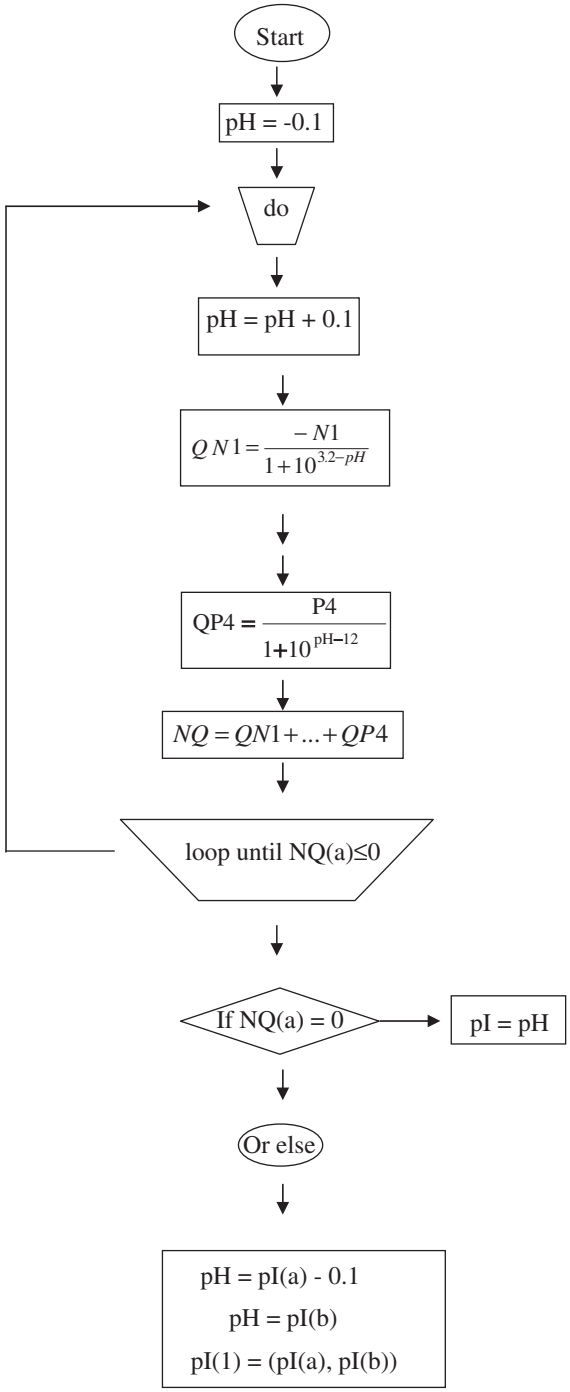

Fig. 1. Flow chart for calculation of the first decimal figure of isoelectric point of a protein pI (a). See text for details.

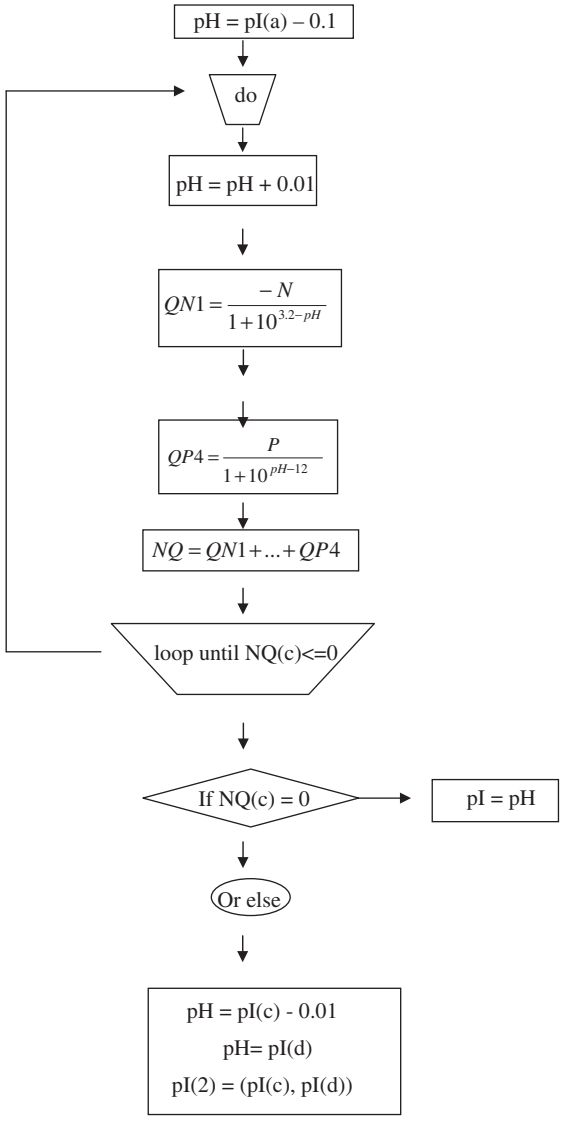

Fig. 2. Flow chart for calculation of the second decimal figure of the isoelectric point of a protein pI (c). See text for details.

the Comprehensive methods rendered identical results when tested with proteins of very different amino acid composition.

3. Summary

We have previously described several methods for theoretical determination of isoelectric point (pI) of proteins and other macromolecules. One of the methods, Comprehensive method [4], takes into account all the acid–base residues and potential chemical modifications affecting charge of aminoacids. The program,

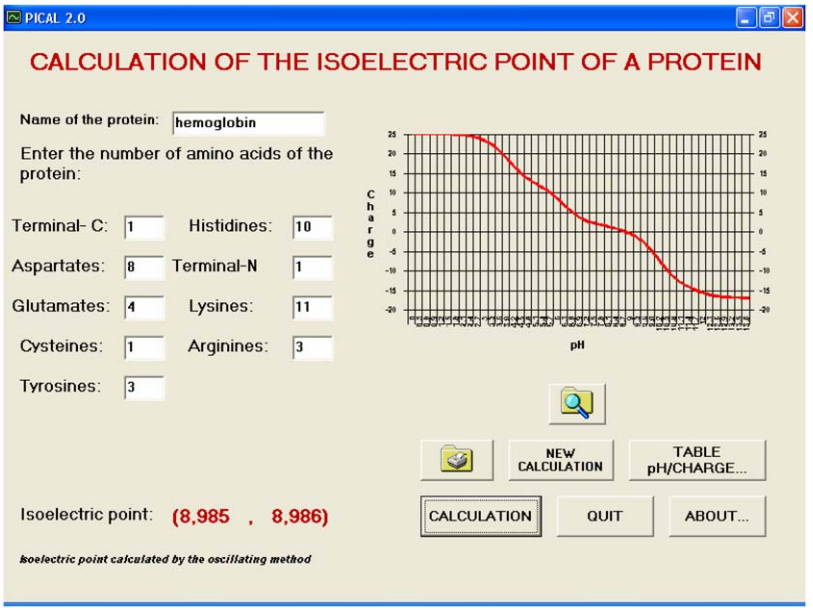

Fig. 3. Main screen of the Oscillating method. The name of the protein and the number of the different acid–base residues can be introduced. The pI value and a representation of the charge of the protein vs. pH values are automatically supplied by the computer. In addition, an amplification of that chart and a table with the same data can be requested.

written in BASIC, requires solution of complex polynomials. In the Simplified method [2], approximate pI values of proteins can be calculated with the help of tables relating pI values with aminoacid composition of protein. In this paper, a new method for pI determination is described. The program has been written in Visual Basic, takes into account all aminoacid residues of protein, has the accuracy of Comprehensive method and uses a much simpler approach to obtain pI value of any macromolecule. The program has been developed so as to give three significant figures for pI value.

An important part of modern proteomic deal with simultaneous analysis of mixtures of proteins contained in biological samples [8–13]. Development of new techniques for the detection of minimal amounts of proteins, together with the use of two-dimensional gel analysis combining separation procedures based on the different molecular weights and isoelectric points of proteins have been relevant landmarks in these types of studies. In this regard, the development of theoretical methods to evaluate and predict pI values of macromolecules are very important tools to rationalize the results experimentally obtained.

Acknowledgements

This investigation was supported by grants from Direccion General de Investigacion Cientifica y Tecnica (BMC-2002-00866), Comunidad de Madrid (08/0021.1/2001), and Instituto de Salud Carlos III, RMN (C03/08), Madrid, Spain. We thank Hugo Osorio for help in the preparation of this manuscript.

Appendix

The main lines of the program are presented below.

```
Module:
Function QT(ByVal N1, N2, N3, N4, N5, P1, P2, P3, P4, pH As Variant)
Q1 = (−N1)/(1 + (10^(3.2 − pH)))
Q2 = (−N2)/(1 + (10^(4 − pH)))
Q3 = (−N3)/(1 + (10^(4.5 − pH)))
Q4 = (−N4)/(1 + (10^(9 − pH)))
Q5 = (−N5)/(1 + (10^(10 − pH)))
Q6 = (P1)/(1 + (10^(pH − 6.4)))
Q7 = (P2)/(1 + (10^(pH − 8.2)))
Q8 = (P3)/(1 + (10^(pH − 10.4)))
Q9 = (P4)/(1 + (10^(pH − 12)))
QT = Round(Q1 + Q2 + Q3 + Q4 + Q5 + Q6 + Q7 + Q8 + Q9, 5)
End Function

Form1:
Option Explicit 'they are the variables
Dim N1, N2, N3, N4, N5, P1, P2, P3, P4
Dim Na, Pa, Nb, Pb
Dim pI
Dim QN1, QN2, QN3, QN4, QN5, QP1, QP2, QP3, QP4
Dim NQ 'it is the net charge of the protein
Dim pH
Dim pIa, pIb 'remember that pI1 = (pIa, pIb)
Dim pIc, pId 'remember that pI2 = (pIc, pId)
Dim pIe, pIf 'remember that pI3 = (pIe, pIf)
Dim printer, counter
Private Sub Calculationbutton_Click()
pI = ""
counter = 0
For counter = 0 To 8
If Text1(counter).Text = "" Then Text1(counter) = "0"
Next
N1 = Val(Text1(0).Text)
N2 = Val(Text1(1).Text)
N3 = Val(Text1(2).Text)
N4 = Val(Text1(3).Text)
N5 = Val(Text1(4).Text)
P1 = Val(Text1(5).Text)
```

```
P2 = Val(Text1(6).Text)
P3 = Val(Text1(7).Text)
P4 = Val(Text1(8).Text)
'- - - - - - - - - - - - - - - - -OSCILLATING METHOD- - - - - - - - - - - - - - - - -
pH = -0.1
Do
pH = 0.1 + pH 'increasing the pH value 0.1
QT = Module1.QT(N1, N2, N3, N4, N5, P1, P2, P3, P4, pH)
Loop Until NQ <= 0
If NQ = 0 Then
pI = pH
LabelIPA.Caption = pI
GoTo 10 'if we have found the pI with one decimal, the Oscillating method is over
else
pIa = pH
pIb = pH -0.1
End if
pH = pIa-0.1 'looking for another decimal
Do
pH = pH + 0.01
QT = Module1.QT(N1, N2, N3, N4, N5, P1, P2, P3, P4, pH)
Loop Until NQ <= 0
If NQ = 0 Then
pI = pH
LabelIPA.Caption = pI
GoTo 10
Else
pIc = pH
pId = pH -0.01
End If
pH = pIc-0.01 'looking for another decimal
Do
pH = pH + 0.001
QT = Module1.QT(N1, N2, N3, N4, N5, P1, P2, P3, P4, pH)
Loop Until NQ <= 0
If NQ = 0 Then
pI = pH
LabelIPA.Caption = pI
GoTo 10
Else
pIe = pH
```

```
pIf = pH −0.001
End If
LabelIPA.Caption = "(" & pIe & " , " & pIf & ")"

'- - - - - - - - - - - - - - -CHART REPRESENTATION- - - - - - - - - - - - - - - - - -
10 Command4.Enabled = True
MSChart1.Visible = True
Commandnewcalculation.Visible = True
Commandprint.Visible = True
Commandzoom.Visible = True
Commandtable.Visible = True
With MSChart1
.ColumnCount = 1
.RowCount = 141
.AutoIncrement = True
End With
pH = −0.1 'The oscillating method is used again
Do
pH = 0.1 + pH
QT = Module1.QT(N1, N2, N3, N4, N5, P1, P2, P3, P4, pH)
Form1.MSChart1.RowLabel = pH
Form1.MSChart1.Data = NQ
Loop While pH < 14
Form1.MSChart1.Repaint = True


'- - - - - - - - - - - - - - -CHARGE vs pH TABLE- - - - - - - - - - - - - - - - - -
Private Sub Tablebotton_Click()
counter = 0
For counter = 0 To 8
If Text1(counter).Text = "" Then Text1(counter) = "0"
Next
N1 = Val(Text1(0).Text)
N2 = Val(Text1(1).Text)
N3 = Val(Text1(2).Text)
N4 = Val(Text1(3).Text)
N5 = Val(Text1(4).Text)
P1 = Val(Text1(5).Text)
P2 = Val(Text1(6).Text)
P3 = Val(Text1(7).Text)
P4 = Val(Text1(8).Text)
pH = −0.5
counter = 0
```

```
Do
pH = 0.5 + pH
QT = Module1.QT(N1, N2, N3, N4, N5, P1, P2, P3, P4, pH)
Form_table.Text1(counter).Text = QT
Loop While pH <= 14
Form_table.Show
End sub
```

References

[1] J.C. Cameselle, J.M. Ribeiro, A. Sillero, Derivation and use of a formula to calculate the net charge of acid–base compounds, Its application to aminoacids, proteins and nucleotides, Biochem. Educ. 14 (1986) 131–136.
[2] A. Sillero, J.M. Ribeiro, Isoelectric points of proteins: theoretical determination, Anal. Biochem. 179 (1989) 319–325.
[3] J.M. Ribeiro, A. Sillero, An algorithm for the computer calculation of the coefficients of a polynomial that allows determination of isoelectric points of proteins and other macromolecules, Comput. Biol. Med. 20 (1990) 235–242.
[4] J.M. Ribeiro, A. Sillero, A program to calculate the isoelectric point of macromolecules, Comput. Biol. Med. 21 (1991) 131–141.
[5] J.M. Ribeiro, A. Ruiz, M.A. Sillero, A. Sillero, Theoretical isoelectric points and electric charges of mutated human haemoglobin subunits, Clin. Chim. Acta 190 (1990) 189–197.
[6] D.S. Moore, Amino acid and peptide net charges: a simple calculational procedure, Biochem. Educ. 13 (1985) 10–11.
[7] C.S. Patrickios, E.N. Yamasaki, Polypeptide amino acid composition and isoelectric point. II. Comparison between experiment and theory, Anal. Biochem. 231 (1995) 82–91.
[8] R. Gras, M. Muller, E. Gasteiger, S. Gay, P.A. Binz, W. Bienvenut, C. Hoogland, J.C. Sanchez, A. Bairoch, D.F. Hochstrasser, R.D. Appel, Improving protein identification from peptide mass fingerprinting through a parameterized multi-level scoring algorithm and an optimized peak detection, Electrophoresis 20 (1999) 3535–3550.
[9] G. Henriksson, A.K. Englund, G. Johansson, P. Lundahl, Calculation of the isoelectric points of native proteins with spreading of pKa values, Electrophoresis 16 (1995) 1377–1380.
[10] K. Hiller, M. Schobert, C. Hundertmark, D. Jahn, R. Munch, JVirGel, Calculation of virtual two-dimensional protein gels, Nucleic Acids Res. 31 (2003) 3862–3865.
[11] P.M. Larsen, L. Storgaard, S.J. Fey, Proteins present in bovine papillomavirus particles, J. Virol. 61 (1987) 3596–3601.
[12] M.W. Reeves, M.C. Drummond, M. Tager, Partial purification and characterization of the multiple molecular forms of staphylococcal clotting activity (coagulase), J. Bacteriol. 148 (1981) 861–868.
[13] A. Santagostino, G. Giagnoni, P. Fumagalli, D. Pavesi, E. Torretta, Isoelectric point determination of human and camel beta-endorphin, alpha-endorphin and enkephalins, Biochem. Biophys. Res. Commun. 104 (1982) 577–582.

Antonio Sillero received both the Ph.D. degree in Chemistry (Universidad Complutense de Madrid) and the M.D. Ph.D degree (Universidad de Granada) in 1966. He was a postdoctoral Fellow of both the U.S. National Institutes of Health and the Jane Coffin Childs Memorial Fund at the Department of Biochemistry of the New York University Medical Center (1967–1970). His research interests include control of the irreversible enzymes of glycolysis and gluconeogenesis in liver, metabolism of fructose, RNA bacteriophages, development of Artemia, proteases, purine nucleotide metabolism, degradation, synthesis and function of dinucleoside polyphosphates, theoretical determination of electric charges of macromolecules, enzyme kinetics. Lately he studies the response of yeast to oxidative stress. At present he is Professor of Biochemistry and Molecular Biology in the Medical School of Universidad Autonoma of Madrid.

Andres Maldonado, born in 1984 is a second year medical student whose hobbies are computer programming and chess.