BIOCHEMISTRY AND MOLECULAR BIOLOGY EDUCATION
Vol. 38, No. 4, pp. 230–237, 2010

Article

Isoelectric Point, Electric Charge, and Nomenclature of the Acid–Base Residues of Proteins*

Received for publication, September 13, 2009, and in revised form, January 14, 2010; accepted March 8, 2010

Andrés A. Maldonado‡, João M. Ribeiro§, and Antonio Sillero‡¶

From the ‡Departamento de Bioquímica, Instituto de Investigaciones Biomédicas Alberto Sols UAM/CSIC, Facultad de Medicina, Madrid 28029, Spain, §Grupo de Enzimología, Departamento de Bioquímica y Biología Molecular y Genética, Facultad de Medicina, Universidad de Extremadura, Badajoz, Spain

The main object of this work is to present the pedagogical usefulness of the theoretical methods, developed in this laboratory, for the determination of the isoelectric point (p*I*) and the net electric charge of proteins together with some comments on the naming of the acid-base residues of proteins

Keywords: isoelectric point, electric charge of proteins, acid–base residues, PICAL.

In the recent years, we have approached the theoretical calculation of the isoelectric point and the net electric charge of proteins and other macromolecules [1–7] and we have used this study to explain these topics to students in regular academic courses of Biochemistry. The object of this publication is to present our pedagogical experience in this field, new applications and possibilities of using the Internet for these procedures and some comments on the naming of the acid–base residues of proteins, widely used in textbooks of Biochemistry.

An essential point for the development of this work was to rename the acid–base residues of proteins as type-N groups (from Neutral when undissociated) and type-P groups (from Positively charged when undissociated). A full classification of these acid–base residues as type-P or type-N groups (or subgroups), and their p*K* values is presented in Table I.

There are a plethora of online methods for the determination of the isoelectric point of proteins, some of which can be used freely from the Internet:

- http://www.scripps.edu/%7Ecdputnam/protcalc.html
- http://www.embl-heidelberg.de/cgi/pi-wrapper.pl
- http://www.bioinformatics.vg/sms/protein_iep.html
- http://www.expasy.ch/tools/pi_tool.html
- http://www.expasy.ch/tools/protparam.html

However, none of these applications is intended as a teaching tool to address the concept of protein p*I* or the relationships between the p*I* value of a protein and the p*K* values of its acid–base groups. An exception is a report recently made available at http://isoelectric.ovh.org, which pursues objectives somewhat similar to ours.

The following is an application of two useful methods developed by us for the determination of the p*I* value and electric charge of a protein. The Simplified Method, calculates the p*I* value with the help of three Tables included in this article, and without assistance of a computer. The other, Improved Oscillating Method (IOM), can be run with a program freely accessible from the Web. The IOM has been adapted for student use and has some special characteristics. The pK_a values of amino acids can be easily modified and the p*I* values compared with those obtained with the standard p*K* values of amino acids. In addition, a list of specially designed proteins with didactical purposes and with a predetermined amino acid composition is included. Finally, a graph with the charge of the protein *vs*. the pH values can be easily obtained.

DETERMINATION OF THE ISOELECTRIC POINT (p*I*) VALUE OF PROTEINS BY THE SIMPLIFIED METHOD

As previously reported [1–8], the electric charge of type-N and type-P groups at a given pH value can be calculated according to equations 1 and 2, both derived from the Henderson-Hasselbach equation:

$$\text{Charge of type–N groups} = \sum_{i=1}^{n} \frac{-1}{1 + 10^{pK_{Ni} - pH}} \qquad (1)$$

$$\text{Charge of type–P groups} = \sum_{i=1}^{p} \frac{1}{1 + 10^{pH - pK_{Pi}}} \qquad (2)$$

By application of these equations to a macromolecule containing a number of acid–base groups, of known p*K* values, its isoelectric point can be determined. In a previous work [2], three methods for the theoretical determi-

*This work is supported by: Dirección General de Investigación Científica y Técnica (BFU 2008-00666/BMC), Consejería de Economía, Comercio e Innovación, Junta de Extremadura, Spain (GRU09135), FSE, FEDER.

¶To whom correspondence should be addressed. E-mail: antonio.sillero@uam.es.

TABLE I
Amino acid residues potentially bearing electric charges

Group type	Conjugate acid	Conjugate base + H^+	p*K* Value
N_1	R-COOH (terminal carboxyl)	R-$(COO)^-$ (terminal carboxylate)	3.2
N_2	R-COOH (β-carboxyl of aspartate)	R-$(COO)^-$ (β-carboxylate of aspartic)	4.0
N_3	R-COOH (γ-carboxyl of glutamate)	R-$(COO)^-$ (γ -carboxylate of glutamic)	4.5
N_4	R-SH (thiol of cysteine)	R-S^- (thiolate of cysteine)	9.0
N_5	R-C_6H_4OH (phenyl of tyrosine)	R-$(C_6H_4O)^-$ (phenolate of tyrosine)	10
N_a	N_a= [N_1 or N_2 or N_3]		4.2
N_b	N_b = [N_4 or N_5]		9.5
P_1	R-$(C_3H_5N_2)^+$ (imidazolium of histidine)	R-$C_3H_4N_2$ (imidazole of histidine)	6.4
P_2	R-$(NH_3)^+$ (terminal ammonium)	R-NH_2 (terminal amino)	8.2
P_3	R-$(NH_3)^+$ (ε-ammonium of lysine)	R-NH_2 (ε-amino of lysine)	10.4
P_4	R-$(CH_6N_3)^+$ (guanidinium of arginine)	R-CH_5N_3 (guanidine of arginine)	12
P_a	P_a = [P_2 or P_3 or P_4]		11.2
P_b	P_b = P_1		6.4

The acid-base residues are classified as belonging to group-N or -P depending on whether they are neutral (N) or positively (P) charged when undissociated.

nation of the p*I* values of proteins were described: Simplified, Abridged and Comprehensive. The Abridged [2] and the Comprehensive [2, 3, 5] methods convey the use of complex mathematical calculations and will not be considered here. A short presentation of the Simplified Method [2] is given below, because of its Simplicity and potential use by students, without any need of computer help. This method requires only the handling of Tables II–IV. The rationale of this method is partially based on the following premises [2]: (i) due to the similarity in their p*K* values, N_1, N_2, and N_3 are set equal to N_a (pK_{Na} value of 4.2); N_4 and N_5 set equal to N_b (pK_{Nb} value of 9.5); P_2, P_3, and P_4 set equal to P_a (pK_{Pa} value of 11.2); by analogy, P_1 is renamed as P_b (Table I); the same simplification on the p*K* values was applied in the Abridged Method [2] and, in this context, the results obtained with the Simplified and Abridged methods are rather similar; (ii) the relationship between the number of type-P_a and type-N_a groups (P_a/N_a or N_a/P_a) is the main factor influencing the p*I* value of a protein (see also below); (iii) at some pH ranges (5–7), the charge values of the P_a and the N_a groups are considered with opposite charge, and the charge afforded by the N_b groups negligible. The theoretical basis of the Simplified Method is described in detail in [2]. Following the information given in Tables II–IV, students may calculate the ratios involving the number of N_a, N_b, P_a and P_b groups present in a protein and from those ratios its p*I* value can be easily deduced. Depending on the time available in the class room, the professor could invite the students to apply Tables II–IV for the determination of the p*I* value of a certain protein, go through the theoretical bases or encourage the students to peruse the original paper [2].

TABLE II
A simplified procedure to calculate approximate pI values of proteins (modified from [2])

Case	Procedure
C1	$N_a/(P_a + P_b) \geq 1.06$ → Table III
C2	$N_a/(P_a + P_b) < 1.06$ → Calculate $P_b/(N_a - P_a)$ → Table IV
C3	$P_a/(N_a + N_b) \geq 1.09$ → Table III
C4	$P_a/(N_a + N_b) < 1.09$ → Calculate $N_b/(P_a - N_a)$ → Table IV

N_a = 1 + Asp + Glu; P_a = 1 + Lys + Arg; N_b = Cys + Tyr; P_b = His. If $N_a > P_a$, apply c1, c2; if $P_a > N_a$, apply c3, c4.

DETERMINATION OF p*I* VALUE AND ELECTRIC CHARGES OF PROTEINS VERSUS pH USING THE IMPROVED OSCILLATING METHOD (IOM)

Theoretical Bases

The Improved Oscillating Method (IOM) is written in Visual Basic [6, 7] without the mathematical complications of the Abridged [2] and the Comprehensive

TABLE III
Correlation between the pI value of a protein and its content in aspartate, glutamate, lysine, and arginine (modificated from [2])

p*I*	N_a/P_a	P_a/N_a	p*I*
4.0	2.5849	2.5849	11.4
4.1	2.2589	2.2589	11.3
4.2	2.0000	2.0000	11.2
4.3	1.7943	1.7943	11.1
4.4	1.6310	1.6310	11.0
4.5	1.5012	1.5012	10.9
4.6	1.3981	1.3981	10.8
4.7	1.3162	1.3162	10.7
4.8	1.2512	1.2512	10.6
4.9	1.1995	1.1995	10.5
5.0	1.1585	1.1585	10.4
5.1	1.1259	1.1259	10.3
5.2	1.1000	1.1000	10.2
5.3	1.0794	1.0794	10.1
5.4	1.0631	1.0631	10.0
5.5	1.0501	1.0501	9.9
5.6	1.0398	1.0398	9.8
5.7	1.0316	1.0316	9.7
5.8	1.0251	1.0251	9.6
5.9	1.0199	1.0199	9.5
6.0	1.0158	1.0158	9.4
6.1	1.0126	1.0126	9.3
6.2	1.0100	1.0100	9.2
6.3	1.0079	1.0079	9.1
6.4	1.0063	1.0063	9.0
6.5	1.0050	1.0050	8.9
6.6	1.0040	1.0040	8.8
6.7	1.0031	1.0031	8.7
6.8	1.0025	1.0025	8.6
6.9	1.0019	1.0019	8.5
7.0	1.0015	1.0015	8.4
7.1	1.0012	1.0012	8.3
7.2	1.0009	1.0009	8.2
7.3	1.0007	1.0007	8.1
7.4	1.0005	1.0005	8.0
7.5	1.0003	1.0003	7.9
7.6	1.0001	1.0001	7.8
7.7	1.0000	1.0000	7.7

N_a = 1 + Asp + Glu; P_a = 1 + Lys + Arg.

TABLE IV
Correlation between pI values of proteins and some ratios of their acid-base groups (modified from [2])

pI	$P_b/(N_a - P_a)$	$N_b/(Pa - Na)$	pI
5.2	1.0631	–	–
5.3	1.0794	–	–
5.4	1.1000	–	–
5.5	1.1259	–	–
5.6	1.1585	–	–
5.7	1.1995	1.1995	10.2
5.8	1.2512	1.2512	10.1
5.9	1.3162	1.3162	10.0
6.0	1.3981	1.3981	9.9
6.1	1.5012	1.5012	9.8
6.2	1.6310	1.6310	9.7
6.3	1.7943	1.7943	9.6
6.4	2.0000	2.0000	9.5
6.5	2.2589	2.2589	9.4
6.6	2.5849	2.5849	9.3
6.7	2.9953	2.9953	9.2
6.8	3.5119	3.5119	9.1
6.9	4.1623	4.1623	9.0
7.0	4.9811	4.9811	8.9
7.1	6.0119	6.0119	8.8
7.2	7.3096	7.3096	8.7
7.3	8.9433	8.9433	8.6
7.4	11.0000	11.0000	8.5
7.5	13.5893	13.5893	8.4
7.6	16.8489	16.8489	8.3
7.7	20.9526	20.9526	8.2
7.8	26.1189	26.1189	8.1
7.9	32.6228	32.6228	8.0
8.0	40.8107	40.8107	7.9
8.1	51.1187	51.1187	7.8
8.2	64.0957	64.0957	7.7
8.3	80.4328	80.4328	7.6
8.4	101.000	101.0000	7.5

$N_a = 1 + Asp + Glu$; $P_a = 1 + Lys + Arg$; $N_b = Cys + Tyr$; $P_b = His$.

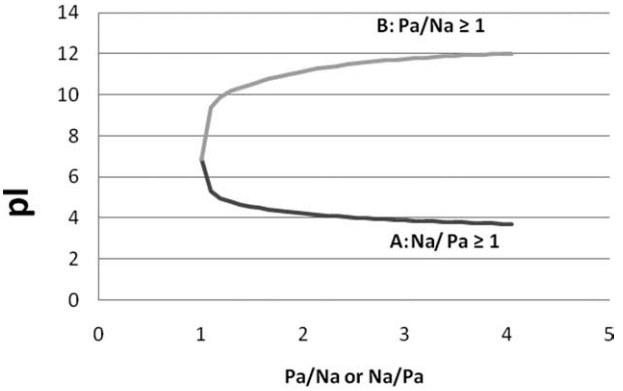
Methods [2, 3, 5]. It can be used not only to calculate the p*I* value of a protein, with the same accuracy as the Comprehensive Method, but also Evaluate the electric charge of a protein at different pH values. The Improved Oscillating Method (IOM) [6, 7], has been partially rewritten and adapted for pedagogical purposes and has been applied to the software "PICAL for BAMBED" (p*I* Calculation for Biochemistry and Molecular Biology Education).

The core of the program uses, by default, the standard p*K* values of each one of the acid–base groups of proteins (Table I). After introducing the number of each one of these groups, the program uses equations 1 and 2 to calculate the charge of each residue (N_1, N_2, N_3, N_4, N_5, P_1, P_2, P_3, P_4, Table I) and to estimate the net charge of the protein (NQ) at increasing pH values, starting from zero and with increments of 0.1 pH units. Note as all the natural proteins are positively charged (NQ $>$ 0) at very acid pH values, when the type-N and the type-P groups are neutral or positively charged, respectively; depending on the nature and content of its type-N groups, the charge of the protein becomes gradually less positive, as the pH value of the medium increases, until a value of net charge NQ $\leq$ 0 is reached. The pH value corresponding to the transition from NQ $>$ 0 to NQ $<$ 0 is called p*I*(a). In the very improbable case that NQ(a) is exactly zero, the p*I* for the protein [p*I* = p*I*(a)] is reported on the screen. When NQ(a) $<$ 0, the following situation arises: at pH = p*I*(a) the net charge of the protein is negative and at pH = p*I*(a) − 0.1 the charge is positive. Therefore p*I*(a) is the protein p*I* value with one significant decimal figure. For the calculation of the second decimal figure of p*I*, the loop starts at pH = p*I*(b) = p(a) − 0.1, with increments of 0.01 pH units. A pH value p*I*(c) is then reached at which NQ(c) $\leq$ 0. The process continues oscillating until the precision required is reached. The number of decimal figures of the p*I* value can be set up by the operator (see below).

The algorithm was first developed for handling only one protein (Oscillating Method) [6] and later adapted (Improved Oscillating Method) for the calculation of the net electric charge and p*I* values of the proteins contained in a file or data base [7]. Computer details of both procedures have been described in [6, 7]. What follows is a simple route to facilitate its operation together with some significant examples to be used by students in a regular course of Biochemistry and/or by researchers in basic or applied research on proteins.

Model Proteins

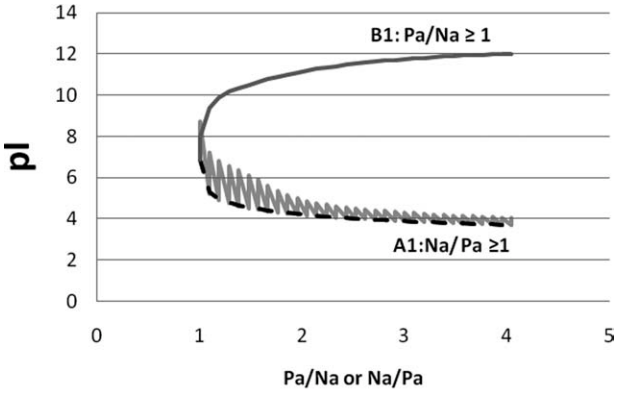
Several theoretical proteins were designed for student analysis, with compositions chosen to give an overview on the influence of each type of amino acid on the p*I* of



FILE A ($N_a / P_a \geq 1$)								
Protein	**pI**	**Asp**	**Glu**	**Cys**	**Tyr**	**His**	**Lys**	**Arg**
1	6.89	10	10	0	0	0	10	10
2	5.29	11	11	0	0	0	10	10
3	4.98	12	12	0	0	0	10	10
4	4.8	13	13	0	0	0	10	10
...	...	...	...	0	0	0	10	10
33	3.7	42	42	0	0	0	10	10

FILE B ($P_a / N_a \geq 1$)								
Protein	**pI**	**Asp**	**Glu**	**Cys**	**Tyr**	**His**	**Lys**	**Arg**
1	6.89	10	10	0	0	0	10	10
2	9.4	10	10	0	0	0	11	11
3	9.89	10	10	0	0	0	12	12
4	10.16	10	10	0	0	0	13	13
...	...	10	10	0	0	0	...	...
33	10.35	10	10	0	0	0	42	42

FIG. 1. Amino acid composition of files A and B, containing only type-N (Asp and Glu) and type-P residues (Lys and Arg) with different N_a and P_a ratios. The p*I* values obtained with the IOM method are related to the ratio N_a and P_a in the upper part of the Figure.



FILE A1 ($N_a/P_a \geq 1 \pm His$)								
Protein	**pI**	**Asp**	**Glu**	**Cys**	**Tyr**	**His**	**Lys**	**Arg**
1	6.89	10	10	0	0	0	10	10
2	7.31	10	10	0	0	1	10	10
3	7.67	10	10	0	0	5	10	10
4	7.84	10	10	0	0	10	10	10
5	7.95	10	10	0	0	15	10	10
6	5.29	11	11	0	0	0	10	10
7	5.56	11	11	0	0	1	10	10
...	...	...	...	0	0	...	10	10
165	4.08	42	42	0	0	15	10	10

FILE B1 ($P_a/N_a \geq 1 \pm His$)								
Protein	**pI**	**Asp**	**Glu**	**Cys**	**Tyr**	**His**	**Lys**	**Arg**
1	6.89	10	10	0	0	0	10	10
2	7.31	10	10	0	0	1	10	10
3	7.67	10	10	0	0	5	10	10
4	7.84	10	10	0	0	10	10	10
5	7.95	10	10	0	0	15	10	10
6	9.4	10	10	0	0	0	11	11
7	9.4	10	10	0	0	1	11	11
...	...	10	10	0	0	...	...	...
165	12.02	10	10	0	0	15	42	42

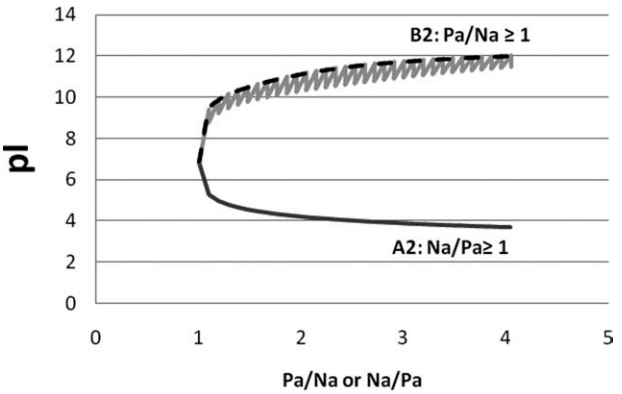
FIG. 2. Amino acid composition of files A1 and B1, containing type-N (Asp and Glu), and type-P residues (Lys and Arg) with different N_a and P_a ratios and different amounts of His, as indicated. The p*I* values obtained with the IOM method are refered to the ratio N_a and P_a in the upper part of the Figure.

a protein. They are listed in www.estudiantesmedicina.com/bambed.htm and in an abbreviated form in Figs. 1–4. These proteins have been grouped in several databases (A, A_1, A_2, A_3, B, B_1, B_2, and B_3) with the characteristics and purposes described below.

Proteins with Only Type-N_a and Type-P_a Acid–Base Groups (Contained in Files A and B)—The proteins containing both, but only, N_a groups (aspartates + glutamates) and P_a groups (lysines + arginines) have been grouped in files A and B. File A contains proteins with a variable number of aspartates (10–42) and glutamates (10–42) and a fixed number of lysines (10) and arginines (10), which makes a total of 33 proteins, all of them with a ratio $N_a/P_a \geq 1$ (Fig. 1, file A; see previous page). File B contains proteins (with a ratio $P_a/N_a \geq 1$) with a variable number of lysines (10–42) and arginines (10–42) and a fixed number of aspartates (10) and glutamates (10), also making a total of 33 proteins (Fig. 1, file B).

Influence of the Incorporation of Histidines, Cysteines, or Tyrosines to Proteins from File A and File B—The influence of these amino acids on the p*I* value of proteins containing only aspartates, glutamates, lysines and arginines was approached in a similar way by adding histidines, cysteines or tyrosines residues. Each one of the 33 proteins present in file A and file B were supplemented with series of 0, 1, 5, 10, or 15 histidines (Fig. 2, files A_1 and B_1); 0, 1, 5, 10, or 15 tyrosines (Fig. 3, files A_2, B_2); 0, 1, 5, 10, or 15 cysteines (Fig. 4, files A_3, B_3). Each file A_1, A_2, A_3, B_1, B_2, B_3, contains $33 \times 5 = 165$ proteins.

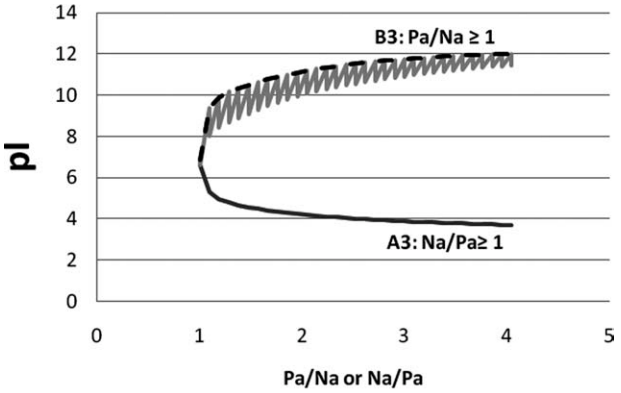
Calculation of the Net Electric Charge and pI Value of a Protein—The first screen of the program (Fig. 5) presents three choices: to calculate the isoelectric point and charge of one protein, to load a database or to create a new database. If the first option is selected the second screen "Calculation of the isoelectric point of one protein" appears (Fig. 6). The name of the protein (optional) and the number of each one of the amino acids (potentially) bearing charge is then requested. The program assumes, by default, the occurrence of one terminal carboxylic and one terminal ammonium group in the protein. The number of decimal figures to be reported for the isoelectric point may also be set. After pressing the "Calcu-



FILE A2 ($N_a/P_a \geq 1 \pm Tyr$)								
Protein	**pI**	**Asp**	**Glu**	**Cys**	**Tyr**	**His**	**Lys**	**Arg**
1	6.89	10	10	0	0	0	10	10
2	6.89	10	10	0	1	0	10	10
3	6.87	10	10	0	5	0	10	10
4	6.86	10	10	0	10	0	10	10
5	6.84	10	10	0	15	0	10	10
6	5.29	11	11	0	0	0	10	10
7	5.29	11	11	0	1	0	10	10
...	...	...	...	0	...	0	10	10
165	3.7	42	42	0	15	0	10	10

FILE B2 ($P_a/N_a \geq 1 \pm Tyr$)								
Protein	**pI**	**Asp**	**Glu**	**Cys**	**Tyr**	**His**	**Lys**	**Arg**
1	6.86	10	10	0	0	0	10	10
2	6.84	10	10	0	1	0	10	10
3	9.4	10	10	0	5	0	10	10
4	9.32	10	10	0	10	0	10	10
5	9.09	10	10	0	15	0	10	10
6	8.92	10	10	0	0	0	11	11
7	8.8	10	10	0	1	0	11	11
...	...	10	10	0	...	0	...	...
165	11.46	10	10	0	15	0	42	42

FIG. 3. Amino acid composition of files A2 and B2, containing type-N (Asp and Glu), and type-P residues (Lys and Arg) with different N_a and P_a ratios and different amounts of Tyr, as indicated. The p*I* values obtained with the IOM method are refered to the ratio N_a and P_a in the upper part of the Figure.



FILE A3 ($N_a/P_a \geq 1 \pm Cys$)

Protein	pI	Asp	Glu	Cys	Tyr	His	Lys	Arg
1	6.89	10	10	0	0	0	10	10
2	6.86	10	10	1	0	0	10	10
3	6.77	10	10	5	0	0	10	10
4	6.69	10	10	10	0	0	10	10
5	6.63	10	10	15	0	0	10	10
6	5.29	11	11	0	0	0	10	10
7	5.29	11	11	1	0	0	10	10
...	...	...	...	...	0	0	10	10
165	3.7	42	42	15	0	0	10	10

FILE B3 ($P_a/N_a \geq 1 \pm Cys$)

Protein	pI	Asp	Glu	Cys	Tyr	His	Lys	Arg
1	6.89	10	10	0	0	0	10	10
2	6.86	10	10	1	0	0	10	10
3	6.77	10	10	5	0	0	10	10
4	6.69	10	10	10	0	0	10	10
5	6.63	10	10	15	0	0	10	10
6	9.4	10	10	0	0	0	11	11
7	9.09	10	10	1	0	0	11	11
...	...	10	10	...	0	0	...	...
165	11.45	10	10	15	0	0	42	42

FIG. 4. Amino acid composition of files A3 and B3, containing type-N (Asp and Glu), and type-P residues (Lys and Arg) with different N_a and P_a ratios and different amounts of Cys, as indicated. The p*I* values obtained with the IOM method are refered to the ratio N_a and P_a in the upper part of the Figure.

lation" button, the isoelectric point value and a graph showing the net electric charge of the protein *vs*. pH appears on the screen. The following possibilities are now open: i) to produce a Table of the electric charge of the protein at each pH value; ii) to save the values obtained for this protein; (iii) to load a new protein (human hemoglobin α chain is incorporated in the program as an example); (iv) to change the p*K* value of any amino acid and to proceed to a new calculation; in this case the influence of changing the p*K* values of one or several acid–base residues on the electric charge and the p*I* value of one protein can be graphically visualized. The original p*K* values of the amino acids can be restored with the buttons "change p*K* values/reset values"; the original p*K* values of the amino acids are restored every time the program is initiated. With this modality, the parameters of each protein can be determined sequentially.

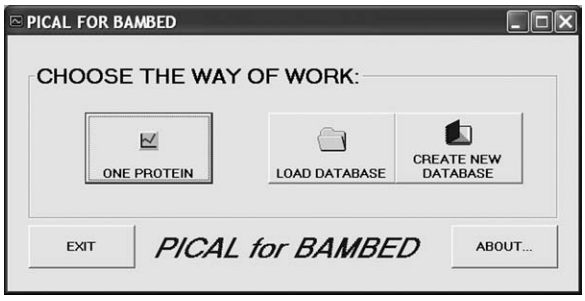


FIG. 5. The first screen of PICAL invites to follow three options: work with one protein, with a stored database or to create a new database (modified from [7]).

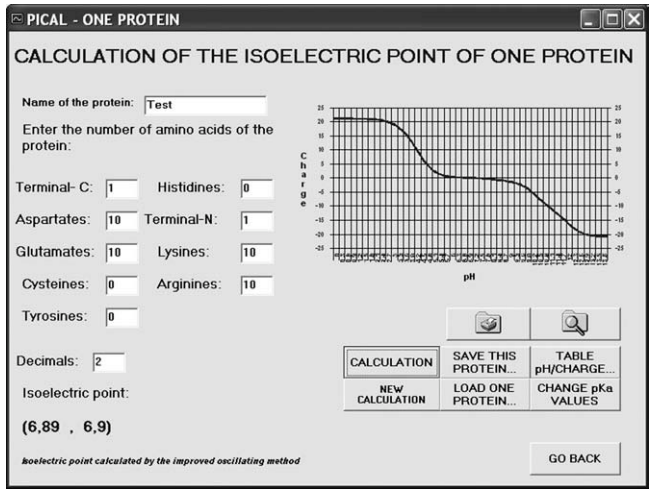


FIG. 6. Second main screen of PICAL: calculation of the electric charge and pI value of one protein. For more details see the text (modified from [7]).

*Calculation of the Net Electric Charge and p*I *Values of Several Proteins or Proteins from a Data Bank*—After pressing the "Load database" button in screen number 1 (Fig. 5), the files stored by default appear. These files (A, A_1, A_2, A_3, B, B_1, B_2, and B_3) correspond to the pool of proteins described above. The abbreviated amino acid composition of these files is in Figs. 1–4, and the complete composition can be found in www.estudiantesmedicina.com/bambed.htm. After selecting any of these files, screen number 3 (Fig. 7) is displayed offering multiple options. By default, the first protein of the chosen file appears, with the number of amino acids and p*I* already calculated. The net electric charge vs. pH chart of this particular protein can be seen on the screen by pressing the corresponding button. By pressing the First, Back, Next, and Last buttons, several proteins from the selected file can be handled.

Additionally, the following operations can be executed jointly with the pool of proteins contained in a file: determination of the p*I* value of all the proteins of the file; ordering of the proteins according to their names (by default), p*I* value and data input and, once ordered, a new global p*I* chart and list of the proteins can be obtained; change of a file by deletion or addition of new proteins; loading and creation of new data base are also possible. This is made directly from the program (using the buttons "Load Database" and "Create New Database", Fig. 5), which use the file format *.mdb (Microsoft

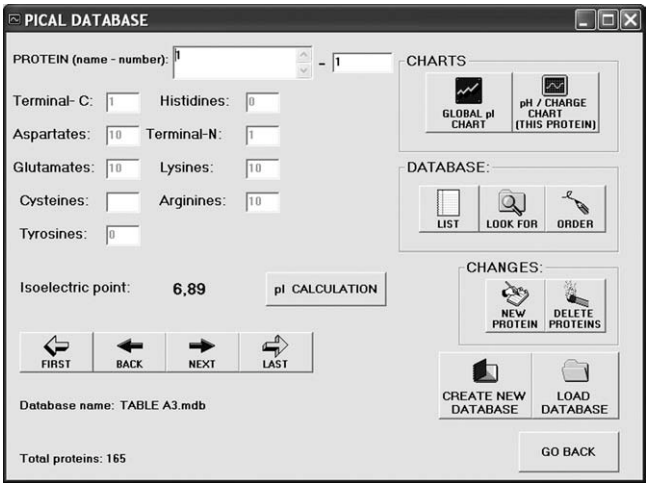


FIG. 7. Third main screen of PICAL: calculation of the electric charge and pI value from a database. For more details see the text (modified from [7]).

Access). Other uses of the program, including "global pI charge", "pH/charge chart (for one protein)", "go back", "exit" or "print" buttons, are self-evident.

GENERAL CONSIDERATIONS ON THE p*I* VALUES OF THE PROPOSED PROTEINS

By application of the Improved Oscillating Method to the proteins of the files described above, Figs. 1–4 were obtained. As shown in Fig. 1 the p*I* value of a protein is greatly influenced by the ratio of its content in aspartates and glutamates and of lysines and arginines, i.e. the ratios N_a/P_a and P_a/N_a: that one giving a value of ≥ 1 is used in the graphs. A ratio of 1 corresponds to a p*I* value of 6.89. The ratios between P_a and N_a have been calculated considering also the presence of one terminal carboxyl and one terminal ammonium group in a protein. The p*I* value of a protein increases, as the ratio P_a/N_a increases and decreases as the ratio N_a/P_a increases (Fig. 1).

The addition of histidines to proteins with a ratio $N_a/P_a \geq 1$ tends to increase their p*I* values (Fig. 2, A_1) and have almost no influence on proteins with a ratio $P_a/N_a \geq 1$ (Fig. 2, B_1). Contrary to histidine, the addition of tyrosines (Fig. 3, A_2, B_2) or cysteines (Fig. 4, A_3, B_3) has more influence on the p*I* values of proteins with a $P_a/N_a \geq 1$). For a discussion on these points see [2].

To better appreciate the influence of the addition of histidines, tyrosines or cysteines on the p*I* value of the standard proteins composed only by P_a and N_a groups, the curves of the p*I* values of the basal proteins (Fig. 1, A and B) have been superimposed, when appropriate, as dashed lines in the graphs in Figs. 2–4.

Comparison of the Simplified and the Improved Oscillating Methods

As stated above the Simplified (and Abridged) Method is based on the main premise that a mean p*K* value was assumed for some groups of residues: [N_1 or N_2 or N_3] = N_a (pK_{Na} value = 4.2); [N_4 or N_5] = N_b (pK_{Nb} = 9.5); [P_2 or P_3 or P_4] = P_a (pK_{Pa} = 11.2). The possibility offered by the Improved Oscillating Method of changing the standard p*K* value of any acid–base residue of a protein allows an examination of the deviation from the p*I* value and net electric charge of proteins introduced by the Simplified (or Abridged) Method [2] in comparison with IOM. This was made by running the IOM, with the p*K* values (Table I) used in the Simplified Method. As an example, the charts of the p*I* value of proteins contained in files A, B (the basal proteins) and A_1, B_1 (the basal proteins supplemented with histidine) are shown in Fig. 8. The results presented in Fig. 8 show that the p*I* values obtained by the Simplified Method using the Tables II–IV are very similar to those obtained with IOM.

Other Pedagogical Suggestions

The Improved Oscillating Method, here adapted for its use in regular courses of Biochemistry, offers many alternatives to existing methods. Its use may stimulate the students to investigate the electric charge and p*I* values of proteins, including those devised by themselves, to measure the influence of changing p*K* values, or to test the effect of covalent modifications of proteins (such as phosphorylations) on both parameters. The deviation of

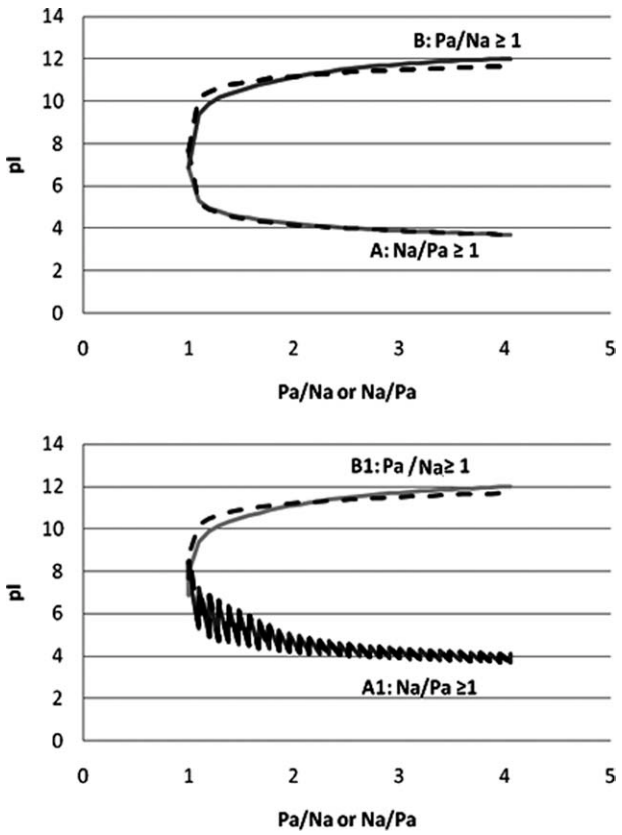


FIG. 8. Representation of the proteins from files A, B, A1 and B1, with the IOM (Improved Oscillating Method), using the standard pK values (continuous line) or the pK values used in the Simplified/Abridged Method [2, 7] (discontinuous line). The amino acid composition of those files are in Figs. 1 and 2. For more details see the text.

the experimental value from the theoretical p*I* value of a protein may indicate changes in the p*K* values of some acid–base residues in the native protein. The IOM may serve as a predictive tool to determine the best chromatographic conditions (type of resins and pH to be used) or the best pH for the precipitation of a protein (pH = p*I*). The IOM method (used by PICAL program) can be easily extended for use in the determination of the electric charge of nucleosides, nucleotides and polynucleotides [1] and to devise and understand a chromatographic method for their separation. In our view the theoretical methods here described for the determination of the electric charge and p*I* value of macromolecules offer many alternatives for their use in the classroom, or in the laboratory, to stimulate students' participation. In this sense, we developed a written contest around this methodology, with strong participation, among the students of regular courses of Biochemistry. The students were invited to analyze the pI value and electric charge of model proteins designed by them, with a brief discussion on the results obtained. At the end of the contest three diplomas (first, second and third) were awarded in a public ceremony.

ON THE NOMENCLATURE OF THE ACID–BASE RESIDUES USED IN SOME TEXTBOOKS OF BIOCHEMISTRY

For the sake of simplicity we shall consider firstly the carboxyl and ammonium groups. At a first glance, the R-COOH (carboxyl) group of any amino acid or protein is an acidic group, able to donate protons, once the pH is above 3–4. Considering a physiological pH value (around 7.4) this group is in the form of $R\text{-}COO^-$ (carboxylate), a conjugate base able to take protons from the medium. By the same token, the $R\text{-}NH_2$ (amino) residue is a basic group able to take protons. However in a very wide range of pH values (from 0 to 9) this group is in the form of $R\text{-}NH_3^+$ (ammonium), an acid able to donate protons to the medium. At physiological pH values, the $R\text{-}COO^-$ and $R\text{-}NH_3^+$ groups could be theoretically regarded as basic and acidic residues, respectively, in contrast to their usual treatment in many textbooks. Similar reasoning could be applied to other acid–base groups of amino acids and proteins (see Table I). Because of the difficulty in qualifying those groups as acidic or basic residues, we started considering all these groups in the undissociated form, as belonging to type-N, type-P, or to any of their subgroups (see above; Table I).

Certain confusion may then arise in the naming of the acid–base residues of amino acids. The plethora of excellent textbooks for Biochemistry and the fact that the same textbook frequently applies distinct names to those residues depending on the framework in which they are discussed open so many possibilities that it is out of the scope of this article to present a survey of them. Nevertheless, as these relate the electric charge of proteins, a small summary is below presented.

In general, there is a tendency to consider aspartic and glutamic acid as acidic amino acids carrying a negative charge and to consider lysine, arginine and histidine as basic amino acids carrying a positive charge [9–14]. In some cases amino acids are mainly classified as uncharged or charged [14, 15]. Finally the presence of protonated species (HA and $R\text{-}NH_3^+$) and unprotonated species (A^- and $R\text{-}NH_2$) are clearly stated in some texts [14, 16].

In our view, it would be convenient to state that each one of those residues, potentially bearing charge, behaves as an acid–base pair, in the Brönsted-Lowry theory, in which depending on the pH, the acid–conjugate or the basic-conjugate forms predominate (Table I).

1. At acidic pH values (<3) all the acid–base residues of the amino acids are in the acidic form and have either a neutral charge (N-groups) or a positive charge (P groups) (Table I).
2. At a physiological pH value (7.4), the acid–base residues are: a) in their conjugate base form, with a negative charge ($R\text{-}COO^-$), mainly uncharged (imidazole of histidine), or b) in their acidic form, bearing no charge (phenyl of tyrosine and thiol of cysteine) or a positive charge (ammonium of lysine and guanidinium of arginine) (Table I).
3. At more basic pH values (>12) all the amino acids are predominantly in their conjugate base form.

It seems to us that given the diversity of acid–base forms present at physiological pH it could be more convenient to pivot the nomenclature at acid pH values. Disregarding other acid–base residues, the amino acids could be classified as monoammonium monocarboxylic (1:1 ammonium:carboxylate); diammonium monocarboxylic (2:1 ammonium:carboxylate) and monoammonium dicarboxylic (1:2 ammonium:carboxylate) depending on the relative content of these two types of groups, so determinants in the charge of the proteins.

Acknowledgments—The authors thank Dr. María Antonia Günther for critical reading of the manuscript. Part of this investigation was carried out during the stay of A. Maldonado in the Plastic Surgery Department, Unfallkrankenhaus Berlin (UKB), Germany, and they also thank Prof. Dr. Markus Küntscher for his encouragement and support. A. Maldonado is actually a MIR (Medical Resident) in the Plastic Surgery Department of the Hospital Universitario de Getafe (Madrid, Spain).

REFERENCES

[1] J. Cameselle, J. Ribeiro, A. Sillero (1986) Derivation and use of a formula to calculate the net charge of acid–base compounds. Its application to amino acids, proteins and nucleotides, *Biochem. Educ.* **14**, 131–136.
[2] A. Sillero, J.M. Ribeiro (1989) Isoelectric points of proteins: Theoretical determination, *Anal. Biochem.* **179**, 319–325.
[3] J. M. Ribeiro, A. Sillero (1990) 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**, 235–242.
[4] J. M. Ribeiro, A. Ruiz, M. A. Sillero, A. Sillero (1990) Theoretical isoelectric points and electric charges of mutated human hemoglobin subunits, *Clin. Chim. Acta* **190**, 189–197.
[5] J. M. Ribeiro, A. Sillero (1991) A program to calculate the isoelectric point of macromolecules, *Comput. Biol. Med.* **21**, 131–141.
[6] A. Sillero, A. Maldonado (2006) Isoelectric point determination of proteins and other macromolecules: Oscillating method, *Comput. Biol. Med.* **36**, 157–166.
[7] A. Maldonado, F. Vara, A. Sillero (2008) Improved application of the oscillating method for the isoelectric point determination of protein:

Potential connection with protein data banks, *Comput. Biol. Med.* **38**, 23–30.
[8] D. S. Moore (1985) Amino acid and peptide net charges: A simple calculation procedure, *Biochem. Educ.* **13**, 10–11.
[9] J. M. Berg, J. L. Tymoczko, L. Stryer (2002) Biochemistry, 5th ed., W.H. Freeman and Company, New York.
[10] R. Montgomery, T. W. Conway, B. H. Ginsberg, A. A. Spector (1990) Biochemistry: A Case-Oriented Approach, 5th ed., Mosby, St. Louis.
[11] T. McKee, J. R. McKee (2009) Biochemistry: The Molecular Basis of Life, 4th ed., Oxford University Press, New York.
[12] D. E. Metzler (2003) Biochemistry the Chemical Reactions of Living Cells, 2nd ed., Harcourt Academic Press, San Diego, California.
[13] C. K. Mathews, K. E. Van Holde (1990) Biochemistry, The Benjamin/Cummings, Redwood City, California.
[14] A. L. Lehninger (1982) Principles of Biochemistry, Worth Publishers, New York.
[15] D. Voet, J. G. Voet (2004) Biochemistry, 3rd ed., J. Wiley & Sons, New York.
[16] R. K. Murray, D. K. Graner, V. W. Rodwell (2006) Harper's Illustrated Biochemistry, 27th ed., McGraw-Hill, New York.