Ann-Kristin Englund*
Gunnar Johansson
Per Lundahl

Department of Biochemistry, Biomedical Center, Uppsala University, Uppsala, Sweden

Calculation of the isoelectric points of native proteins with spreading of pK_a values

The isoelectric points (pI) of native proteins are important in several separation techniques. For estimating pI values the net charge of several proteins was calculated *versus* pH by use of the Henderson-Hasselbalch equation. Amino acid composition, pK_a values for amino acid side chains and for the *N*- and *C*-terminal groups, and the presence of other charged groups were taken into account. A set of pK_a values was chosen for amino acid residues with ionizable side chains. Each particular type of ionizable group was assumed to have pK_a values distributed around the chosen value, thereby simulating the situation in proteins and polypeptides. The calculated pI values showed reasonably good agreement with experimental ones for most of 16 native proteins over a wide pH range (3.4–11) when charge contributions of heme groups, sialic acid residues, *etc.*, were taken into account. The calculated pI for the human red cell glucose transporter (Glut1) with one sialic acid residue was decreased from 8.8 to 8.5 by introducing pK_a value spreading and became consistent with the experimental pI value of 8.4 ± 0.05 at 15°C determined in the presence of 6 M urea. The pI of the native Glut1 was lower, 8.0 ± 0.1, at 22°C. In general, the pI values for native proteins are affected by the three-dimensional structure of the proteins, which causes greater differences between calculated and experimental pI values than in the case of polypeptides for which pI values are determined in the presence of urea.

1 Introduction

The isoelectric point (pI) of an amphoteric molecule is defined as the pH at which the net charge is zero. The variation of net charge with pH is of importance in charge-dependent separation methods like electrophoresis, isoelectric focusing, chromatofocusing and ion-exchange chromatography. Calculation procedures for estimating pI values for proteins or polypeptides have been described earlier [1–8]. The present study describes a simple calculation procedure with spreading of pK_a values around the chosen values. The procedure was applicable to native proteins with good results.

2 Theory

For a polypeptide of known amino acid composition an approximate pI value can be calculated by use of the ionization constant pK_a for amino acid side chain groups or other types of ionizable groups that may occur. The charge for each such group at any given pH was calculated by use of the Henderson-Hasselbalch equation,

$$x_b = 1/[10^{(pK_a - pH)} + 1] \quad (1)$$

where x_b is the molar fraction of the base form of the ionizable group, by taking into account whether the charge of the base form is zero (as for $-NH_2$) or -1 (as for $-COO^-$). The total net charge at each given pH is obtained by summing up the charge for each type of ionizable group times the number of groups. In the present study, suitable average pK_a values were selected for the ionizable amino acid side chains, and for the terminal groups. The individual ionizable side chains of each type of amino acid were assumed to have the pK_a values distributed around the selected pK_a value, thereby simulating the situation in polypeptides and proteins where a given type of ionizable amino acid side chain often appears in several positions in the amino acid sequence and with various individual ionization constants, depending both on the adjacent side chains and on the three-dimensional environment in the protein. By assuming a distribution of pK_a values, the calculated titration curves will be smoothed out (Fig. 1). For the calculations reported here, the average $pK_a = pK_{a,i}$ was used for one third of the groups of a given type i, $pK_{a,i} + 1$ for another third, and $pK_{a,i} - 1$ for the last third. The calculations can be performed manually or by use of a computer program. In the present approach calculations were performed by a program made in GWBASIC on a PC-compatible computer.

3 Results

3.1 Sets of pK_a values

Three sets of pK_a values were considered (A–C in Table 1). A series of calculations were performed for five proteins, for which experimentally determined pI were available. For several proteins (Table 2), the best agreement between calculated and experimental pI values was obtained with the pK_a value set at C. The pK_a values for His, Cys, Tyr, Lys, and Arg in this set of values are the same as those used by Bjellqvist *et al.* [7], whereas the

Correspondence: Dr. Per Lundahl, Department of Biochemistry, Biomedical Center, Uppsala University, Box 576, S–751 23 Uppsala, Sweden (Tel: +46 18 17 44 59; Fax: +46 18 55 21 39)

Nonstandard abbreviation: Glut1, human red cell glucose transporter

Keywords: Isoelectric point / Native proteins / pK_a value / Titration curve

* Present address: Pharmacia AB, Biopharmaceuticals, S–112 87 Stockholm, Sweden

0173-0835/95/0808-1377 $5.00+.25/0

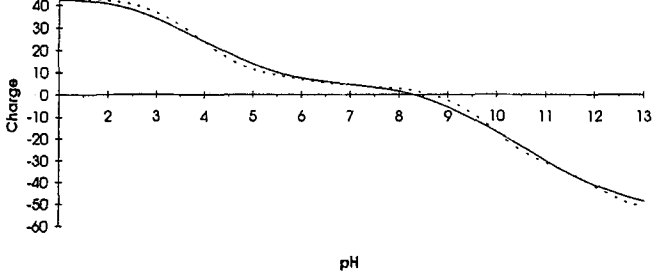

Figure 1. Titration curves for the human red cell glucose transporter, Glut1, with two sialic acid residues (from fresh red cells), with (solid line) and without (hatched line) spreading of the pK_a values in column C of Table 1 below. The calculated p*I* values were 8.29 and 8.67, respectively.

Table 1. pK_a values used for calculation of p*I* values

Amino acid residue or terminal group	A[a]	B[b]	C[c]
α-COOH	3.6	3.6	3
Asp	4.47	4	3.2
Glu	4.47	4.5	4.1
His	6.68	6.4	6
α-NH_2	7.6	8	8
Cys	9.5	9	9
Tyr	10	10	10
Lys	10	10.4	10
Arg	11.9	12	12

a) From [9], used for the p*I* calculations in [4]
b) From [10]
c) For Asp and Glu, the pK_a values were obtained from an NMR determination of all ionization constants for carboxyl groups in ribonuclease HI [11]. For α-COOH, His and Lys, the pK_a values from [9] were modified, whereby a better fit between experimental and calculated p*I* values was obtained.

Table 2. Comparison between the sets of pK_a values in Table 1

		p*I* value[a]				
Protein	Source	A[b]	B[b]	C[c]	Exp	
Heme domain (CDH)	*Phanerochaete chrysosporium*	4.21 (4.31)	3.94 (4.03)	3.34 (3.38)	3.42	[12]
Cellobiohydrolase I	*Trichoderma reesi*	4.75 (4.63)	4.51 (4.39)	3.89 (3.79)	3.9	[13]
β_2-Microglobulin	Guinea pig	7.35 (7.15)	7.22 (6.97)	6.88 (6.59)	6.6	[14]
Glut1	Human	8.58 (8.77)	8.59 (8.90)	8.46 (8.81)	8.0	[15]
Phospholipase A_2	Common viper	9.24 (9.55)	9.43 (9.72)	9.25 (9.55)	9.2	[16]

a) Calculated and experimental p*I* values are given. The pK_a value spreading was used, except for values given in parenthesis, which were calculated without spreading. For experimental p*I* values reference numbers are given within brackets.
b), c) See footnote in Table 1

pK_a values for the *C*-terminal carboxylate group and the Asp and Glu side-chain carboxylate groups are lower than the corresponding values used by these authors and by Matthew [10]. The B set of pK_a values gave good results for prediction of p*I* values (Fig. 2), but we chose to use the C set for further calculations.

3.2 Spreading of pK_a values

The native proteins studied showed reasonably good agreement between calculated and experimental values

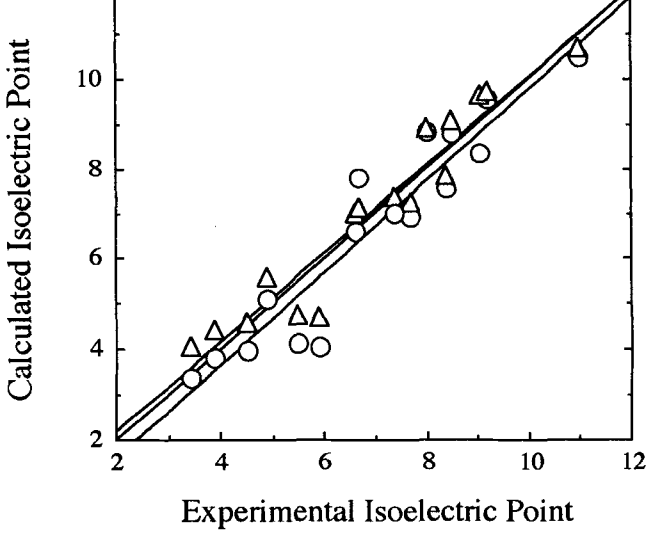

Figure 2. Comparison between calculated p*I* values for the native proteins listed in Table 2, using the B (Δ) and C (o) sets of pK_a values (Table 1), without spreading. The middle line represents the ideal correlation.

over a wide p*I* range. The linear regression was improved by the pK_a value-spreading procedure (from $r = 0.943$ to $r = 0.983$, Fig. 3) and 12 out of 16 calculated p*I* values showed better agreement with the experimental value after the spreading procedure (Table 3). An even better fit might be achieved by using slightly higher pK_a values for the alkaline side chains. The experimental p*I* values for the few urea-denatured proteins of high p*I* that were included for comparison were consistent with the calculated values (Table 3). This was expected, since hidden charges become exposed and modifications of pK_a values of individual ionizable groups due to their three-dimensional environment will largely be eliminated when the polypeptide unfolds.

4 Discussion

The transmembrane protein Glut1 and two cellobiohydrolases were studied in more detail. Membrane proteins may be difficult to handle in isoelectric focusing experiments due to their tendency to self-associate even in the presence of detergent. The p*I* of Glut1 in the presence of urea was determined to be 8.4 [4, 24]. Micropreparative free-zone isoelectric focusing of the native protein in complex with *n*-dodecyl octaoxyethylene ($C_{12}E_8$), in the absence of urea, showed a p*I* of 8.0 [15]. The calculated p*I* value 8.46 for Glut1 closely corresponds to the experimental value obtained in the presence of urea, but deviates moderately from the value for the native protein. The two exocellulases cellobiohydrolase CBH 58 (denoted CBH 1 in Ref. 27) and CBH 62 have p*I* values of 3.85 and 4.85, respectively [27]. Therefore the cDNA sequence [28] that gave a calculated p*I* value of 4.73 probably corresponds to CBH 62.

Procedures for calculation of p*I* values for polypeptides have recently been published [1–8]. Calculations of titra-

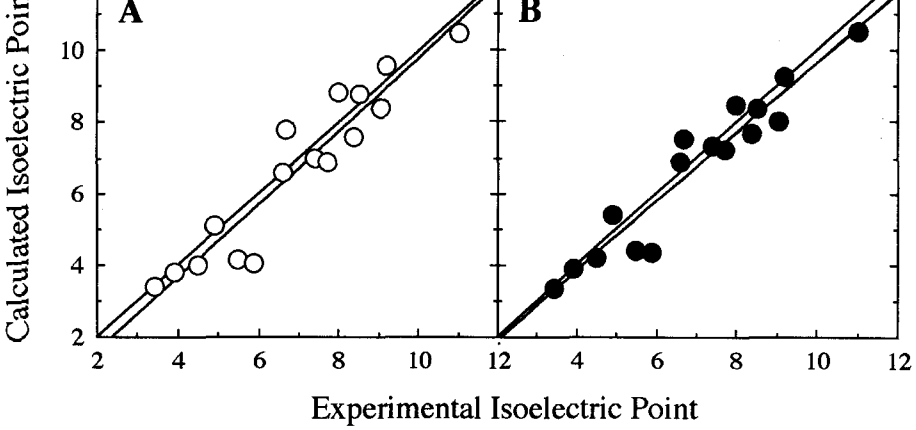

Figure 3. Calculated p*I* values *versus* experimental p*I* values for the native proteins listed in Table 3. (A) Without spreading of pK_a values. (B) With spreading of pK_a values. The pK_a values in column C of Table 1 were used in both panels. Linear regression lines (lower ones) and ideal lines (upper ones) are shown.

Table 3. Experimental p*I* values (with reference numbers within brackets) for (A) native proteins and (B) polypeptides obtained by denaturation of proteins with urea, and corresponding calculated p*I* values without and with spreading of the pK_a values in column C of Table 1

Protein	Source	Code	Experimental p*I* value	Calculated p*I* value	
				Without spreading	With spreading
A. Native proteins					
Heme domain (CDH)[a),b),c)]	*P. chrysosporium*		3.42 [12]	3.38	3.34
Cellobiohydrolase I[b),c)]	*T. reesei*		3.9 [13]	3.79	3.89
Endoglucanase I[b),c)]	*T. reesei*		4.5 [13]	3.97	4.19
Growth hormone[c)]	Human		4.9 [17]	5.1	5.4
Endoglucanase II[b),c),d)]	*T. reesei*		5.5 [13]	4.12	4.40
Cellobiohydrolase II[b),e)]	*T. reesei*		5.9 [13]	4.06	4.34
β-$_2$-microglobulin[c)]	Guinea pig	Sw:Bmg_Cavpo	6.6 [14]	6.59	6.88
Pyruvate kinase	*S. cerevisiae*	Sw:Kpyk_Yeast	6.7 [18]	7.79	7.53
Myoglobin[a)]	Horse heart	Sw:Myg_Horse	7.4 [19]	6.97	7.31
Myoglobin[a)]	Chicken	SW:Myg_Chick	7.7 [20]	6.88	7.22
Glut1[f)]	Human	Sw:Gtrl_Human	8.0 [15]	8.81	8.46
Myoglobin[a)]	Sperm whale	Sw:Myg_Phyca	8.4 [19]	7.55	7.69
Myoglobin[a)]	Emperor penguin	Sw:Myg_Aptfo	8.5 [21]	8.77	8.37
Colicin E	*E. coli* JC411	P:IKEC1	9.05 [22]	8.35	8.00
Phospholipase A_2[c)]	Common Viper	Sw:Pa_Vipbb	9.2 [16]	9.55	9.25
Lysozyme[c)]	Human		11 [23]	10.46	10.52
B. Polypeptides in urea					
Glut1[f)]	Human	Sw:Gtr1_Human	8.40 [24]	8.81	8.46
Cytochrome c	Horse heart	Sw:Cyc_Horse	9.4 [25]	9.59	9.35
Ribonuclease[c)]	Bovine pancreas	P:NRBO	9.58 [26]	9.37	8.81

a) Contains a heme group, pK_a values 4.0 and 4.8
b) *N*-terminus blocked
c) All cysteines form disulfides
d) Denoted endoglucanase III in [13]
e) Two free cysteines
f) From aged red cells, with one sialic acid residue [4], pK_a values 2.75

tion curves and p*I* values for polypeptides of known composition (sequence) are useful for understanding charge properties of the polypeptides or the native protein with or without additional charged groups. Our simple calculation procedure in most cases showed good results for proteins over a wide p*I* range. A small improvement was achieved by spreading the ionization constants. We want to emphasize that effects of the three-dimensional structure (hidden charges and abnormal pK_a values) may have caused discrepancies between calculated and experimental values, which could not be eliminated by the spreading procedure.

We used relatively low pK_a values for the Glu and Asp side chain carboxyl groups obtained by NMR measurements on ribonuclease HI [11]. This protein contains a large number of Lys and Arg and thus the protein will be positively charged at pH close to the pK_a values of the carboxyl acids. This will create local increase of the pH that may decrease the pK_a values of some Glu and Asp residues to become lower than the corresponding average values in proteins. Nevertheless, this set of pK_a values gave a better prediction of the p*I* than the other sets examined (Table 2).

Supported by grants to Göran Pettersson from the Swedish Natural Science Research Council and the Swedish Research Council for Engineering Sciences and by grants to Per Lundahl from the Swedish Natural Science Research Council and the O. E. and Edla Johansson Science Foundation. We are grateful to E. Greijer and E. Brekkan for valuable help and advice.

Received April 26, 1995

[1] Cameselle, J. C., Ribeiro, J. M., Sillero, A., *Biochem. Educ.* 1986, *14*, 131–136.
[2] Skoog, B., Wichman, A., *Trends Anal. Chem.* 1986, *5*, 82–83.
[3] Sillero, A., Ribeiro, J. M., *Anal. Biochem.* 1989, *179*, 319–325.
[4] Englund, A.-K., Lundahl, P., *Biochim. Biophys. Acta* 1991, *1065*, 185–194.
[5] Bjellqvist, B., Hughes, G. J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J.-C., Frutiger, S., Hochstrasser, D., *Electrophoresis* 1993, *14*, 1023–1031.
[6] Mosher, R. A., Gebauer, P., Thormann, W., *J. Chromatogr.* 1993, *638*, 155–164.
[7] Bjellqvist, B., Basse, B., Olsen, E., Celis, J. E., *Electrophoresis* 1994, *15*, 529–539.
[8] Watts, N. R. M., Singh, R. P., *Electrophoresis* 1995, *16*, 22–27.
[9] Tanford, C., *Adv. Protein Chem.* 1962, *17*, 69–165.
[10] Matthew, J. B., *Annu. Rev. Biophys. Biophys. Chem.* 1985, *14*, 387–417.
[11] Oda, Y., Yamazaki, T., Nagayama, K., Kanaya, S., Kuroda, Y., Nakamura, H., *Biochemistry* 1994, *33*, 5275–5284.
[12] Henrikson, G., *Thesis*, Uppsala University, Uppsala 1995.
[13] Ståhlberg, J., *Thesis*, Uppsala University, Uppsala 1991.
Berggård, I., *Biochemistry* 1978, *17*, 947–955.
[15] Englund, A.-K., Lundahl, P., Elenbring, K., Ericson, C., Hjertén, S., *J. Chromatogr.* 1995, in press.
[16] Boffa, G. A., Boffa, M.-C., Winchenne, J.-J., *Biochim. Biophys. Acta* 1976, *429*, 828–838.
[17] Li, C. H., *Mol. Cell. Biochem.*, 1982, *46*, 31–41.
[18] Aust, A. E., Suelter, C. H., *J. Biol. Chem.*, 1978, *253*, 7508–7512.
[19] Öjteg, G., Nygren, K., Wolgast, M., *Acta Physiol. Scand.* 1987, *129*, 277–286.
[20] Itoh, T., Satoh, H., Adachi, S., *Comp. Biochem. Physiol.* 1976, *55B*, 559–561.
[21] Weber, R. E., Hemmingsen, E. A., Johansen, K., *Comp. Biochem. Physiol.* 1974, *49B*, 197–214.
[22] Schwartz, S. A., Helinski, D. R., *J. Biol. Chem.* 1971, *246*, 6318–6327.
[23] Stryer, L., *Biochemistry,* W. H. Freeman New York 1981.
[24] Englund, A.-K., Lundahl, P., *Electrophoresis* 1993, *14*, 1307–1311.
[25] Heaney, A., Weller, D. L., *J. Chem. Educ.* 1970, *47*, 724–726.
[26] Ui, N., *Biochim. Biophys. Acta* 1971, *229*, 567–581.
[27] Uzcategui, E., Ruiz, A., Montesino, R., Johansson, G., Pettersson, G., *J. Biotechnol.* 1991, *19*, 271–286.
[28] Sims, P., James, C., Broda, P., *Gene* 1988, *74*, 411–422.