ELSEVIER

Journal of Chromatography A, 680 (1994) 321–340

JOURNAL OF CHROMATOGRAPHY A

Simulation and optimization of peptide separation by capillary electrophoresis

Alejandro Cifuentes, Hans Poppe*

Laboratory for Analytical Chemistry, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, Netherlands

Abstract

The different mobility equations that have appeared in the literature for predicting peptide mobility were compared. A modified equation that relates the mobility of individual proteolytic species of a peptide to its composition has been obtained: $\mu = 1758 \log(1 + 0.297q)/M^{0.411}$, where μ is the electrophoretic mobility in 10^{-9} $m^2/s \cdot V$, q is the integral value of the charge of the species and M its molecular mass. Also, a rough estimation of the set of pK_a values for a peptide was developed. The usefulness of this equation together with a computer program for predicting separations of compounds by capillary zone electrophoresis is demonstrated, employing real electropherograms of peptides from the literature or from experiments.

1. Introduction

The separation of peptides and proteins by capillary electrophoresis (CE) is one of the most important application fields of this technique [1–4]. The CE analysis of these biopolymers provides valuable information about the identity, purity and structural changes of the peptides themselves and the proteins they constitute [2,5–7]. The high efficiencies normally obtained in these separations and the short analysis times (less than 30 min) have made CE a major laboratory tool for the separation, analysis and characterization of such biomolecules [2,8].

Nevertheless, the analysis and identification of peptides from real samples and the interpretation of peptidic maps resulting from protein hydrolysis are still time consuming [2,5,9]. There is a necessity to develop new approaches that can shorten the long time normally needed to identify peptides, and at the same time to facilitate the improvement of the quality, in terms of efficiency and resolution, of CE peptide separations.

An ab initio method that could predict residence time from structure would be valuable. Correlations of structure with mobility are therefore important. This paper presents a study comparing the different mobility equations that have appeared in the literature for predicting the mobility of peptides, in order to obtain an expression that can be applied in general terms, and consequently able to predict the electrophoretic mobilities of peptides under different separation conditions. Such an equation requires a knowledge of the charge of the species, i.e., a knowledge of pK values. Part of the work was therefore devoted to the development of an a priori estimation of pK values.

* Corresponding author.

0021-9673/94/$07.00
SSDI 0021-9673(94)00363-E

We also describe an approach that allows the prediction of migration times and peak shapes of peptides in different buffer and at different pH values. Predictions of electropherograms are carried out employing a computer program [10] in combination with one of the above-mentioned models that best relates the peptide mobility to its amino acid sequence. The predictions are compared with real separations obtained in our laboratory or from the literature.

2. Experimental

2.1. Instrumentation

Separations were carried out using a laboratory-made electrophoretic system. The apparatus included a Hivolt Model V.C.S 303/1 power supply (Wallis Electronics Worthing, UK) and a Spectroflow 757 variable-wavelength UV–Vis detector (ABI, Ramsey, NJ, USA) with a in-house modified flow cell, operated at 210 nm. Cooling of the capillary to room temperature was achieved with a liquid-thermostated air stream propelled by a fan. During electrophoresis, the current through the capillary was measured using a Metex M-3800 multimeter. A fused-silica capillary (Polymicro Technologies, Phoenix, AZ, USA) of 50 μm I.D. and 360 μm O.D. was used; the total length of the capillary was 68.4 cm and the effective length (from the injection point to the detector) was 43.8 cm. Injection was carried out at the anodic side using electromigration. The computer programs employed were Quattro Pro (Borland International, Scotts Valley, CA, USA) for the optimization of equations, one written at our laboratory for the simulation of the electrophoretic separations and whose characteristics have been described [10] and another custom program to carry out the pK_a calculations. All the programs were used on an 80386SX microprocessor-based PC (Laser 386SX).

2.2. Samples and chemicals

Peptides AGG, GGG and LGF were purchased from Nutritional Biochemicals (Cleveland, OH, USA) and peptides GGP and LGG from Fluka (Buchs, Switzerland). All the long peptides (Table 5) were purchased from Bachem Feinchemikalien (Bubendorf, Switzerland) and used as received. The peptides were dissolved in water, previously purified by passage through a PSC filter assembly (Barnstead, Boston, MA, USA), at the concentrations indicated in each instance. The samples were stored at −20°C and heated to room temperature before use. Ethanolamine, formic acid and acetic acid (E. Merck, Darmstadt, Germany), N-[tris(hydroxymethyl)methyl]glycine (Tricine) and 3-cyclohexylamino-1-propanesulphonic acid (CAPS) (Aldrich, Brunel, Netherlands) were used in the different running buffers. The pH of these solutions was adjusted using sodium hydroxide solution (1 mol/l). The buffers were stored at 4°C and heated to room temperature before use.

3. Results and discussion

3.1. General

The utility of computer programs to predict the electrophoretic behaviour of small molecules and to obtain optimum capillary zone electrophoretic (CZE) separations has been already shown [10–15]. These programs, briefly, describe substance mobility in CZE in terms of fundamental constants of each solute [pK_a and mobility of the dissociated forms (μ_{A^-})] and buffer characteristics [pH, concentration, pK_a and mobility (μ_{B^-}) of the different substances that form the buffer]. However, the possibilities of employing these programs decreases the higher is the analyte complexity in terms of the number of charged groups [16] (e.g., peptides and proteins), as there is a lack of data on the required parameters (pK_a and μ values) for such complex biopolymers.

The computer program that we have employed has been treated in more detail elsewhere [10]. It produces the complete electropherogram, and therefore data on capillary dimensions, injection

time and voltage (or injected plug length), separation voltage, electroosmotic flow and sample concentration are required.

3.2. Prediction of pK_a values of peptides

The parameters for the running buffer are usually easily obtained from the literature [17–19]. Electroosmotic flow is obtained experimentally or taken from the literature. For only very few peptides are pK_a values available in the literature. However, for a fundamental understanding of electrophoretic behaviour this knowledge is essential. Most workers resort to a set of average pK_a values for charged amino acid residues contained in peptides and a set of N-terminal and C-terminal pK_a values. One such a set is given in Ref. [6]. Such sets of values have been shown to be effective, e.g., in calculating the isoelectric point (*IP*) [20,21]. However, it is not likely that such value will lead to accurate predictions of peptide charges at pH values very different from the *IP*, where the charge is larger than unity, for two reasons. First, assigning equal pK_a values to the same type of groups (terminal COOH and NH_2, side-chain groups of the same type) and considering these as pK_a values of the entire molecule neglects the statistical effect that occurs when more than one protolytic group is present [22]. This occurs independent of any (additional) electrostatic or conformational effect of one group on the other, as will be considered later. For example, in a dicarboxylic acid with a large distance between the two carboxyls to exclude these latter effects, each having a "local" pK_a value of 4.7, the compound $COOH\cdots COO^-$ is indistinguishable from $COO^-\cdots COOH$. Therefore, this form is twice as likely to occur, on statistical groups; overall K_{a_1} and K_{a_2} values, as observed with, e.g., titrations or when estimating the charge from the mean mobility, are twice as large and twice as small, respectively, as those of an isolated COOH group; that is, the pK_a values are shifted by log 2 to 4.4 and 5.0, respectively.

In simple peptides, having only a few protolytic groups, this statistical effect may often be smaller than given in the above example because in general the "local" pK_a values differ substantially. However, with larger, multiply charged peptides it may also be greater as so many groups are involved and the likelihood that some are very similar increases. We therefore decided to follow a "brute force" approach and calculated this effect in full detail. The procedure was as follows. Assuming that "local" group pK_a values are available, the formation constant of each individual species (thus distinguishing the two forms given above) can be found as the product of K values of those groups that have given up a proton. The overall $K_{a,n}$ values can then be found by adding all such K values that pertain to a given value of n, the number of protons released.

This calculation turned out to be time consuming, even on a fast PC with a numeric processor, and while truncating the addition given above for forms that contribute very little to the sum. For instance, with eight dissociating groups already a total of $2^8 = 256$ individual forms have to be considered. Nevertheless, it was feasible to carry out the whole calculation is less than 1 min for up to fourteen dissociating groups.

A second complication stems from the mutual electrostatic interaction of charged groups: if the COOH groups in the above example are not far apart, the K value of a given group is smaller when the other group is already ionized. This effect was taken into account in the following crude manner [23]:

$$\Delta pK_{a,i} = Cq_j/D_{i,j}$$

where C is a universal constant, q_j is the charge on the influencing group (0 or 1), and $D_{i,j}$ is the distance of the reaction centres of groups i and j. The C value reflects fundamental constants, such as the elemental charge and the permittivity of the medium. Here it is very empirical in nature. The interaction will take place partly though the solution and partly through the "medium" of the molecule itself. Depending on the nature of the molecule and, e.g., the ionic strength of the solution, these contributions may differ in relative importance. Also, this approach only takes

the enthalpic effect into account; any entropy effect is neglected.

The value to be taken for the distance $D_{i,j}$ is of a similar empirical nature and is difficult to decide for each combination of functional groups in a given molecule. We found the following expression (measuring distances, D, N, Y, in units of one atom–atom bond, of the order of 0.15 nm) to be the most effective in the correlations:

$$(D_{i,j})^2 = 3^2 N_{i,j}/(3^2 + Y_i^2 + Y_j^2)$$

where $N_{i,j}$ is the (contour) distance between the anchoring points of the groups in the main chain of peptide bonds and Y_i and Y_j are the distances from the reaction centres to the anchoring points of the two groups.

This choice reflects the idea that the peptide chain forms a random chain, with random flight segments of length three times the atom–atom distance (= one peptide unit), and that summing of distances should be done quadratically because of the random orientation in space of all the length segments. The choice can be criticized in numerous ways. For example, the randomness (stiffness) of the chain may not be as described, and may not be a constant but rather depend on the charges present and the ionic strength. Also, specific conformations of the chain may be preferred, because of electrostatic and hydrophobic effects, effects that in long chains ultimately lead to folding. We nevertheless worked with these schemes, as this approach allows one to obtain estimates of pK_a values that are in our opinion better than those with the fixed set of group pK_as, and obtaining the necessary data requires only geometrically analysing the chemical formula of the peptide.

Thus, the local pK_a values were corrected with a number of terms such as given by $\Delta pK_{a,i}$, depending on the charge distribution in the form considered, and next the summation described above was carried out.

This approach requires the following data: constant C, pK_a values for twenty terminal NH groups, pK_a values for twenty COOH groups and pK_a values for seven side-chain groups (Arg, Asp, Cys, Glu, His, Lys, Tyr), a total of 48 values. (This large number of parameters may seem excessive. Indeed, a satisfactory correlation can also be obtained by assigning one value to all terminal NH_2 groups, one to proline and one to all terminal COOH-groups. With the seven side-chain values, this would lead to only ten parameters. We decided to use the 48, arguing that we were after a prediction as good as possible for a situation where the structure of the peptide is entirely known. As a result of the large number of parameters there is some triviality in the applied fitting procedure, e.g., some of the amino acids occurring in the training set do not occur in any of the peptides in the training set. As a result for the adaptation of its group pK_a values the only reference point is the amino acid itself: the training pK_a value is just reproduced in the group pK_a value, with a number of degrees of freedom of zero. We did not see this as a disadvantage, through.)

The 48 parameters were obtained by minimizing the sum of squares between predicted pK_a values and those of 43 amino acids and peptides from the literature [24–26] ranging from two to five amino acids, 98 values in total. Literature values for pK_as were corrected for zero ionic strength.

The residual sum of squared deviations (SSQ) was about a factor of 3 smaller (9.35 vs. 24.74) than can be found by just neglecting statistical and electrostatic interaction, e.g., in the way described by Rickard et al. [6].

Some indication of the validity of the procedure can also be derived from the C value obtained, 3.0. An ab initio estimate of C can be found as follows: neglecting all entropy effects, the change in a pK_a value due to a charge e at a distance D (measured in atom–atom distance, $1.5 \cdot 10^{-10}$ m) can be calculated as

$$\Delta pK_a = 0.43 e^2/(4\pi\varepsilon_0\varepsilon_r D \cdot 1.5 \cdot 10^{-10})/kT$$

where e is the elementary charge, ε_0 is the permittivity of vacuum, ε_r is the relative permittivity of the medium, k is the Boltzmann constant and T is the absolute temperature. This expression at $T = 300$ K, with ε_r set to that of water, corresponds to a C value of

Table 1
Values used in CZE mobility calculations

Peptide	M	n	q		μ_{exp} [a]	μ_i [b]	μ_a [c]
			This work	From literature			
20 mmol/l sodium citrate, pH 2.5, 30°C [36]							
LEMY	554.7	4	0.89	0.82	13.1	13.2	12.4
GFY	385.4	3	0.90	0.83	15.8	15.6	14.6
GPETLCGAELVDAL	1443.6	14	0.86	0.71	9.0	8.8	7.3
QF	1293.3	2	0.84	0.83	17.4	16.3	16.3
GFYF	532.6	4	0.92	0.83	14.1	13.9	12.7
GPETLCGAELVDAL-QFVCGDR	2305.6	21	1.65	1.61	11.8	12.7	12.4
VCGDR	604.6	5	1.68	1.73	22.5	22.9	22.8
SCDLR	648.7	5	1.67	1.73	22.3	22.1	22.1
LEMYCAPLK	1123.4	9	1.82	1.82	19.1	18.7	18.4
RLEMY	710.9	5	1.79	1.82	21.4	21.8	22.2
PAK	314.4	3	1.78	1.83	26.0	30.3	31.2
CAPLK	586.7	5	1.84	1.83	18.6	25.2	24.1
FNLPTGY	825.9	7	0.95	1.83	19.7	12.0	21.0
RAPQTGIVDECCFR	1706.9	14	2.62	2.73	19.4	21.1	21.3
RLEMYCAPLK	1279.6	10	2.68	2.82	23.1	24.0	24.5
LEMYCAPLKPAKSA	1577.9	14	2.65	2.82	20.3	21.7	22.5
FNLPTGYGSSSR	1300.4	12	1.93	2.83	22.1	18.2	24.4
RLEMYCAPLKPAK	1576	13	3.48	3.82	24.6	26.6	28.1
39 mmol/l ethanolamine, pH 10, 25°C [37]							
AA	160.2	2	−0.97	−0.98	−24.2	−24.0	−24.2
AAA	231.3	3	−0.98	−0.98	−19.8	−20.9	−20.8
AN	203.2	2	−0.97	−0.98	−22.8	−21.8	−18.7
AG	146.1	2	−0.97	−0.98	−25.9	−24.9	−25.1
AGG	203.2	3	−0.98	−0.98	−22.4	−22.0	−22.0
AH	226.2	2	−0.97	−0.98	−19.2	−20.8	−21.0
GGGGG	303.3	5	−0.99	−0.98	−18.9	−18.8	−18.6
AL	207.3	2	−0.97	−0.98	−21.3	−21.8	−22.0
ALG	259.3	3	−0.98	−0.98	−19.0	−19.9	−19.9
AV	188.2	2	−0.97	−0.98	−22.5	−22.5	−22.7
GGGG	246.2	4	−0.99	−0.98	−23.3	−20.4	−20.3
GGGGGG	360.3	6	−0.99	−0.98	−17.2	−17.5	−17.4
20 mmol/l citric acid, pH 2.5, 30°C [27]							
AFDDING	750.8	7	0.75	0.33	10.3	9.9	4.7
KKKKKKK	915.5	7	7.02	7.33	50.7	52.3	53.5
AKKKKKK	858.2	7	6.05	6.33	49.5	49.0	50.3
SYSMEHFRWGKPV	1624	13	3.43	2.98	22.0	25.7	23.2
ILPWKWPWWPWRR	1907.5	13	3.63	3.32	26.7	25.1	23.5
GRTGRRNSIHDIL	1495.6	13	4.12	4.38	32.1	30.2	31.5
AFKAKNG	734.9	7	2.63	2.41	31.3	29.2	27.4
AFKIKNG	777	7	2.63	2.41	30.4	28.5	26.7
GFLRRIRPKLK	1383.9	11	5.19	5.33	37.8	36.5	37.1
AFKADNG	721.8	7	1.68	1.37	21.7	20.5	17.4
AGCKNFFWKTFTSC	1659.8	13	2.79	2.14	21.4	22.0	17.8
YAGFM	587.8	5	0.95	0.38	12.0	13.8	5.9
YVNWLLAQKGKKN-DWKHNITQ	2586.1	21	4.84	5.28	28.1	26.9	28.5

(Continued on p. 326)

Table 1 *(continued)*

Peptide	M	n	q		μ_{exp} [a]	μ_i [b]	μ_a [c]
			This work	From literature			
GGFMTSEKSQTPLVT-LFKNAIIKNAYKKGE	3304.9	30	5.01	5.30	22.9	24.9	25.8
HFRWGKPVGKKRRP-VKVYP	2336.1	19	8.00	8.24	36.8	38.4	39.0
SYSMEHFRWGKPVG-KKRRPVKVYP	2933.9	24	7.29	8.22	33.1	33.1	35.5
RKRSRKE	959.2	7	4.87	5.31	43.8	40.7	43.0
AFKKKKK	877.2	7	5.11	5.33	45.8	43.6	44.7

[a] Experimental mobility in 10^{-9} $m^2/s \cdot V$ obtained from literature.
[b] Theoretical mobility in 10^{-9} $m^2/s \cdot V$ obtained considering the charge of isolated species and Eq. 5.
[c] Theoretical mobility in 10^{-9} $m^2/s \cdot V$ obtained considering an average charge value and Eq. 5.

$$C = 0.43e^2/(4\pi \cdot 8.8 \cdot 10^{-12} \cdot 80 \cdot 1.5 \cdot 10^{-10})/(1.38 \cdot 10^{-23} \cdot 300) = 2.00$$

The value actually found was 2.96. Although this does not coincide exactly with the ab initio estimate, the fact that it is of the right order of magnitude and bearing in mind the uncertainties connected with the choice of the distance expression and the ε_r value (that of the pure water, 80) provides some confidence that the procedure partially reflects physical reality.

The final *SSQ* value, 9.35, corresponds to a

Table 2
Results from optimization for the different equations

Equation No.	Model	$\sigma = \sqrt{\frac{\sum_{i=48pep} (\mu_{iexp} - \mu_{ipre})^2}{48 - n_{parameters}}}$ (10^{-9} $m^2/s \cdot V$)		$\sigma = \sqrt{\frac{\sum_{i=25pep} (\mu_{iexp} - \mu_{ipre})^2}{25}}$ (10^{-9} $m^2/s \cdot V$)
		σ value using q from this work	σ value using average q from literature	
1	$\frac{94.58 \log(1+q)}{n^{0.312}}$	3.13	3.32	2.61
2	$\frac{839.6\ q}{M^{2/3}}$	4.24	4.74	2.51
3	$\frac{486.4\ q}{M^{0.584}}$	3.94	4.64	2.00
4	$\frac{1222\ q}{3.01M^{1/3} + 1.12M^{2/3}}$	4.02	4.70	2.02
5	$\frac{1758 \log(1 + 0.297\ q)}{M^{0.411}}$	2.23	2.49	1.72

Table 3
Statistical F test of the five equations from Table 3

	$\sigma_1 = 3.13$	$\sigma_2 = 4.24$	$\sigma_3 = 3.94$	$\sigma_4 = 4.02$	$\sigma_5 = 2.23$
$\sigma_1 = 3.13$		−	−	−	+
$\sigma_2 = 4.24$	−		−	−	+
$\sigma_3 = 3.94$	−	−		−	+
$\sigma_4 = 4.02$	−	−	−		+
$\sigma_5 = 2.23$	+	+	+	+	

Significance level 5%; degrees of freedom 45–48; average F value 1.64. Minus signs indicate no significant difference ($\sigma_i^2/\sigma_j^2 < 1.64$) between the equations; plus signs indicate significant difference ($\sigma_i^2/\sigma_j^2 > 1.64$).

root mean squared expected error of $\sqrt{9.35/(98\text{-}48)} = 0.43$. This requires the following comment: this precision is certainly not good enough for the prediction of special pH-induced selectivity effects. One must bear in mind that in the highly efficient CE system substances having a relative overall mobility of only 1.02 are easily separated, whereas of course a value of 1.00 leads to no separation whatsoever. The value 1.02 corresponds to a pH shift (in the middle of a titration branch) of $2 \cdot 0.43 \cdot 0.02 = 0.02$. Hence, the accuracy of the prediction is at least one order of magnitude worse than would be required to predict whether two randomly chosen peptides could be separated due to different pK_a values at some pH.

Table 4
pK_a values and electrophoretic mobilities (μ) used for the simulated electorpherograms of the five peptides in Fig. 4B

No.	Peptide	pK	μ (10^9 m^2/s·V)
1	LGF	8.06	−18.20
			0
		3.47	18.20
2	GGP	8.27	−21.27
			0
		3.21	21.27
3	GGG	8.27	−23.02
			0
		3.30	23.02
4	LGG	8.06	−20.69
			0
		3.30	20.69
5	AGG	8.24	−22.35
			0
		3.30	22.35

Nevertheless, on might hope that for structurally related peptides the prediction of differences in pK_a values is better and the prediction might

Table 5
pK_a values and electrophoretic mobilities (μ) used for the simulated electropherograms of the five peptides in Figs. 5B, 6B and 7B

No.	Peptide	pK	μ (10^9 m^2/s·V)
1	LQAAPALDKL	11.21	−20.49
		7.77	−11.43
			0
		4.16	11.43
		2.97	20.49
2	WAGGDASGE	8.58	−30.43
		5.45	−22.27
		4.23	−12.42
			0
		3.24	12.42
3	FHPKRPWIL	12.42	−10.80
			0
		10.18	10.80
		6.68	19.36
		5.27	26.45
		2.78	32.51
4	SYSMEHPRWG	13.21	−25.96
		10.30	−19.00
		7.36	−10.60
			0
		6.30	10.60
		3.74	19.00
		2.61	25.96
5	ELAGAPPEPA	8.52	−29.04
		5.18	−21.25
		4.22	−11.85
			0
		3.37	11.85

have some limited use in this respect. Also, at pH values where large peptides are multiply charged, an estimate of the total charge q is of great importance. Below we present results to that effect.

3.3. Prediction of mobilities

Next, we need an expression that permits us to obtain the electrophoretic mobilities of individual charged species. Several equations for the electrophoretic mobility of peptides have been published [6,27–30]. These semi-empirical equations relate the electrophoretic mobility (μ) of peptides with structural parameters, such as the charge (q), molecular mass (M) and amino acid numbers (n) of peptides. The equations are as follows:

Grossman's equation [27] (as also used in Ref. [28]);

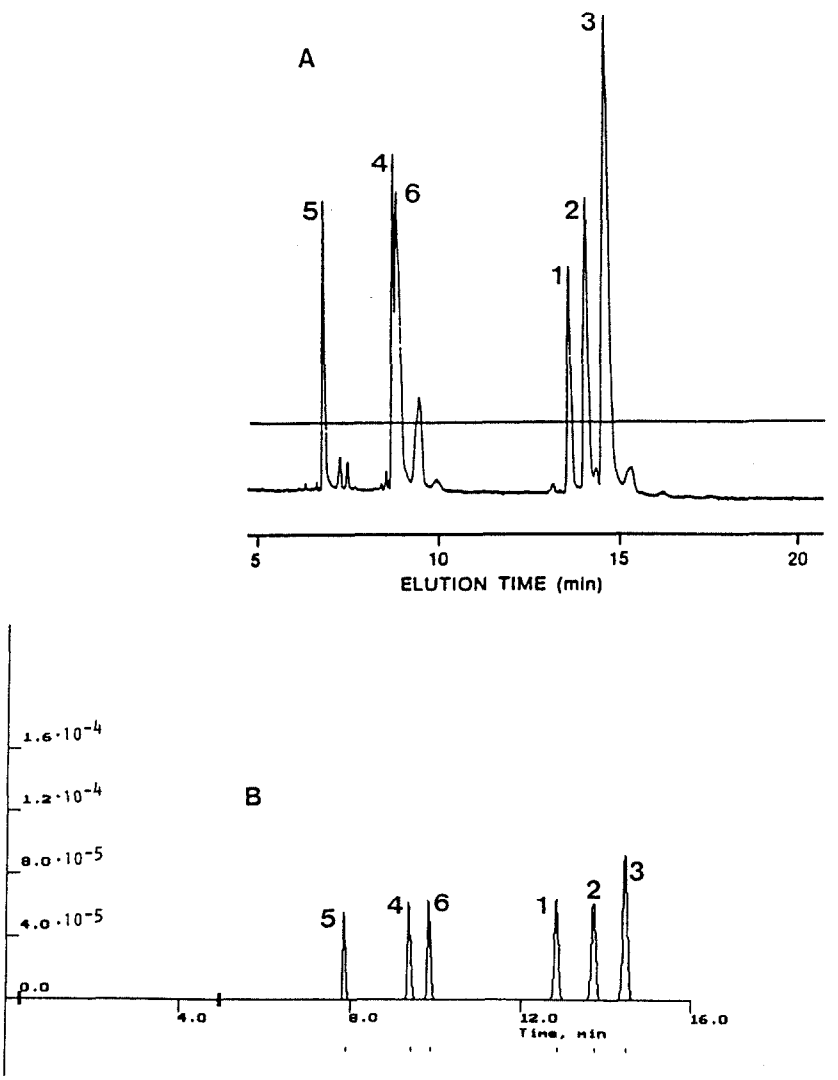

Fig. 1. Electropherogram of six model peptides: 1 = AFAAING; 2 = AFDAING, 3 = AFDDING; 4 = AFKAING; 5 = AFKKING; 6 = AFKADNG. (A) Capillary 65 cm (45 cm to detector) × 50 μm I.D. × 320 μm O.D.; electric field, 277 V/cm; buffer, 20 mM citric acid (pH 2.50); UV detection at 200 nm. Redrawn from Ref. [8]. (B) Simulated electropherogram: injection length, 2 mm; sample concentration, $1.2 \cdot 10^{-4}$–$1.8 \cdot 10^{-4}$ M; electroosmotic flow adjusted to $\mu_{eo} = 9 \cdot 10^{-9}$ m^2/s·V.

$$\mu = \frac{A \log(q+1)}{n^B} \quad (1)$$

Offord's equation [29] (as also used in Refs. [6,30–32]):

$$\mu = \frac{Aq}{M^{2/3}} \quad (2)$$

Compton's equation [33,34]:

$$\mu = \frac{Aq}{M^m} \quad (3)$$

$$\mu = \frac{Aq}{BM^{1/3} + CM^{2/3}} \quad (4)$$

In these equations A, B and C are constants depending on the solvent system used and m

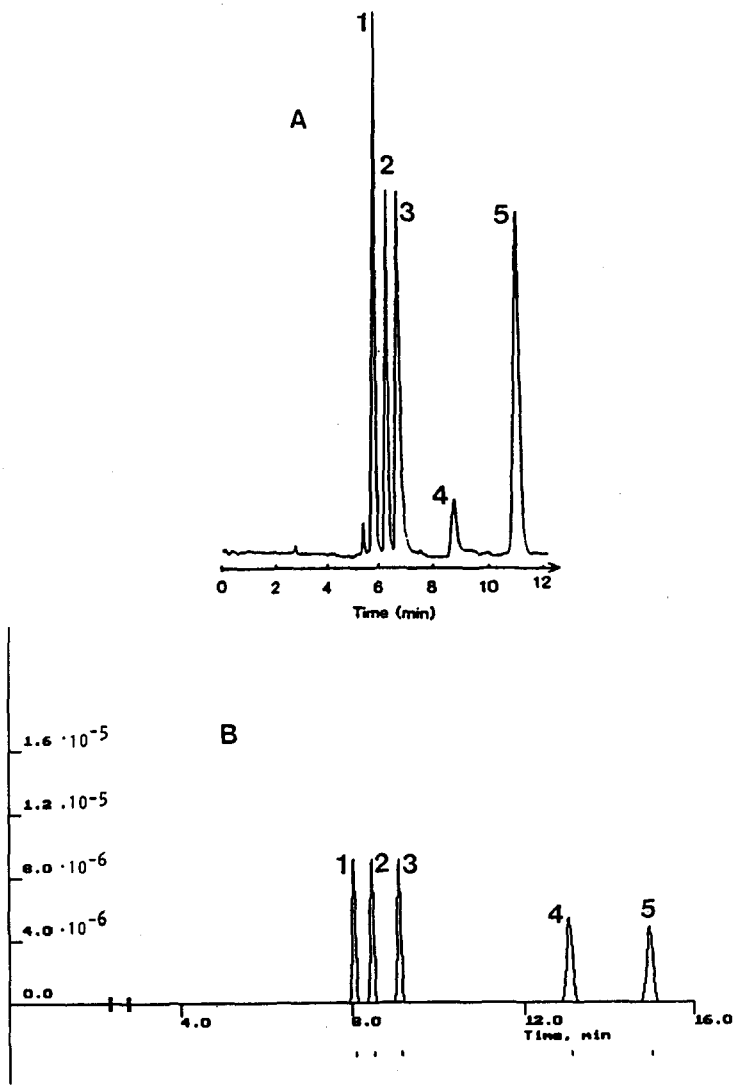

Fig. 2. Electropherogram of substance P degraded by an endopeptidase for 180 min. (A) Non-crosslinked polyacrylamide-coated capillary ($\mu_{eo} = 0$ m^2/s·V), 100 μm I.D., length from injection to detection 20 cm; run voltage, 3000 V; buffer, 30 mM phosphoric acid (pH 2.6); UV detection at 200 nm. Redrawn from Ref. [38]. (B) Simulated electropherogram: injection at 500 V for 30 s; sample concentration considered to be 10^{-5} M.

varies between 1/3 and 2/3 depending on the system and M [34,35]. These equations were used in our work, but Eq. 1 was used in a different way from that used by the originators. They (and others extending their work) inserted average (charge) q values, mostly derived from the application of the Henderson–Hasselbach equation with pK_a values from Ref. [6]. We think that this procedure is inconsistent with a first principle: the average mobility of a peptide occurring in different protolytic forms should be the weighted average of the mobilities of all forms. For mobility correlations that are non-linear in q, such as Eq. 1, this is clearly not the case. For instance, a peptide in two forms, with charges 0 and +1 in equal amounts, thus having $q = 0.5$, would give for the logarithmic factor in Eq. 1 $\log q = \log(1 + 0.5) = 0.176$; weighing of

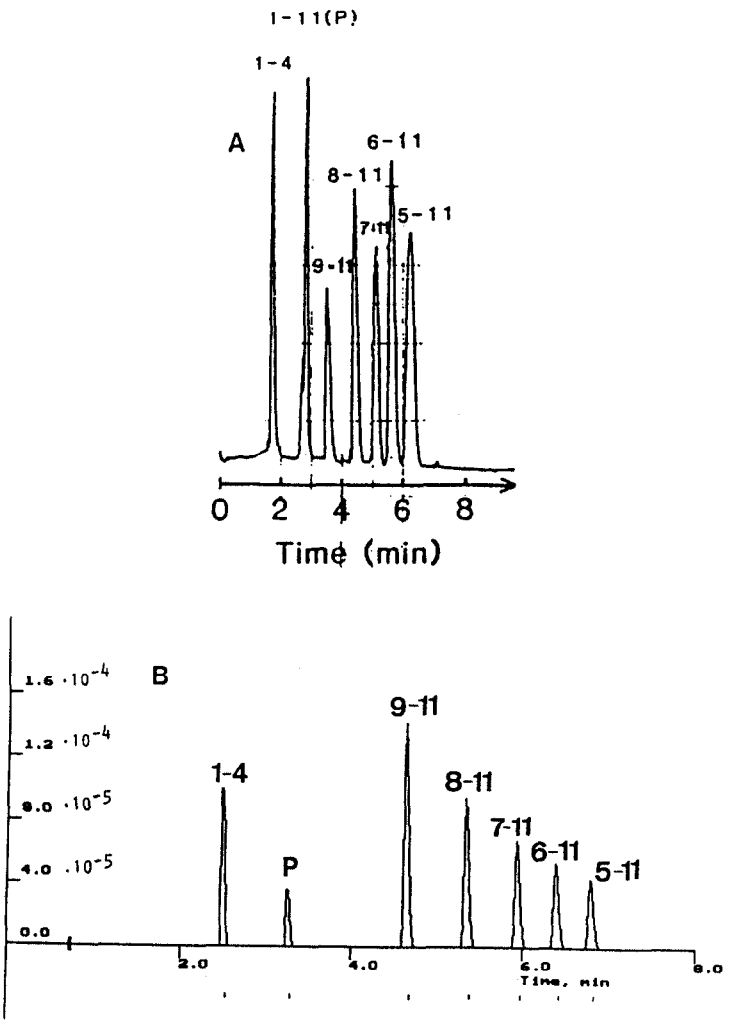

Fig. 3. Electropherogram of an artificial mixture of substance P and its (1–4), (9–11), (8–11), (7–11), (6–11) and (5–11) fragments. Sample concentration, 50 μg/ml of each peptide. (A) Non-crosslinked polyacrylamide-coated capillary ($\mu_{eo} = 0$ $m^2/s \cdot V$), 50 μm I.D., length from injection to detection 16 cm; separation voltage, 6000 V; buffer, 20 mM phosphoric acid (pH 2.6); UV detection at 200 nm. Redrawn from Ref. [38]. (B) Simulated electropherogram: injection at 2000 V for 10 s; sample concentration, 50 μg/ml.

the logarithmic factor for the individual forms yields 0.5 log (1 + 0) + 0.5 log (1 + 1) = 0.151. We remedied this by applying Eq. 1 and our modification Eq. 5 (see below) to the *individual* forms, using integer charge (qn) values instead of q, and next performing the averaging of the mobility with the Henderson–Hasselbach equation. This is numerically more complicated, but, when data are stored in a computer the expense is still negligible compared with that of an experiment.

To test these equations we used electropho-

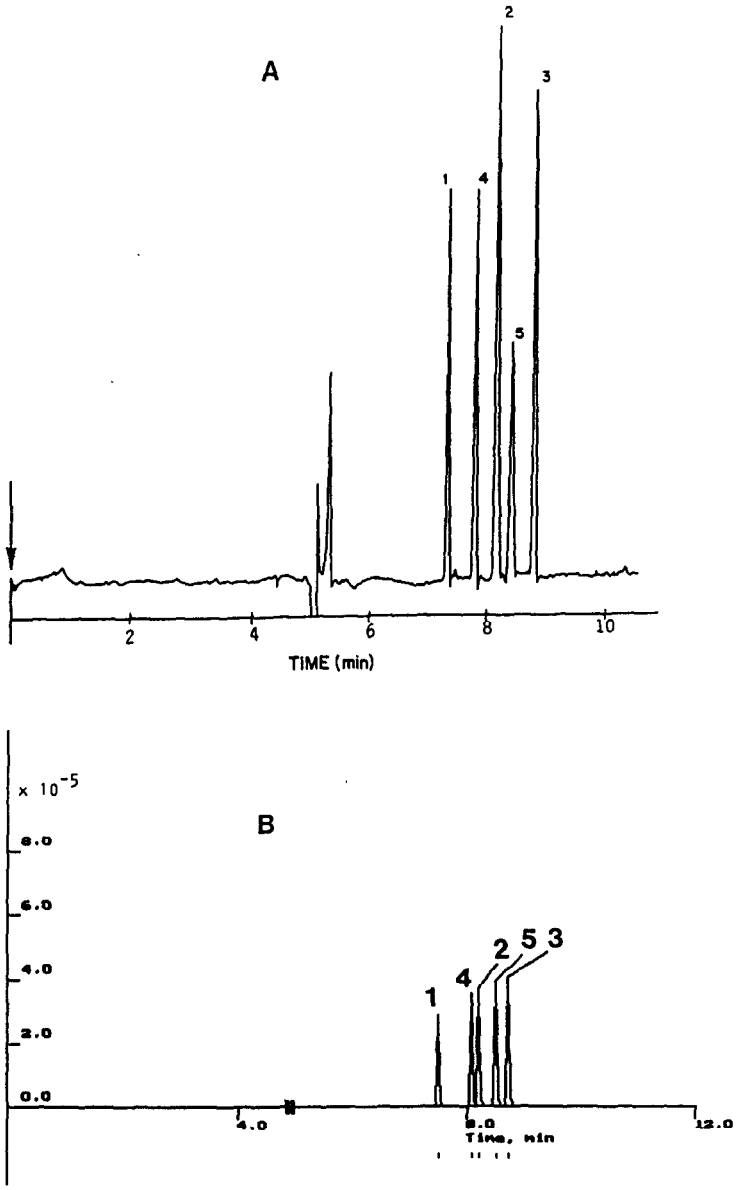

Fig. 4. Electropherograms of peptides from Table 4. Sample concentration, 0.4 mg/ml for each peptide. (A) Capillary, 68.4 cm (43.8 cm to detector) × 50 μm I.D. × 360 μm O.D.; injection at 300 V for 4 s; run voltage, 20 kV; buffer, 0.1 *M* CAPS (pH 10.4) ($\mu_{eo} = 50.8 \cdot 10^{-9}$ m^2/s·V); UV detection at 210 nm. (B) Simulated electropherogram.

retic mobility, charge and molecular mass (M or n) data from different peptides with different pH, ionic strength and buffers. These values, obtained from the literature [27,36,37], are given in Table 1. In addition we considered a modification of Eq. 1:

$$\mu = \frac{A \log(1 + Bq)}{M^C} \quad (5)$$

A logarithmic relationship between charge and mobility was chosen by the authors of Eq. 1, while considering that as the total charge on the peptide increases, the effect of other additional charges on its mobility should decrease [27]. However, the curvature described by $\log(1 + q)$ is entirely arbitrary, and we preferred to have an adjustable (via the parameter B) curvature. Also, we considered that the dependence of mobility on molecular mass could be more precise than the classical dependence on the number of amino acids in a polymer model.

The charge data from the literature given in Table 1 had been calculated employing the Henderson–Hasselbach equation and Rickard et al.'s pK values [6], and they were obtained

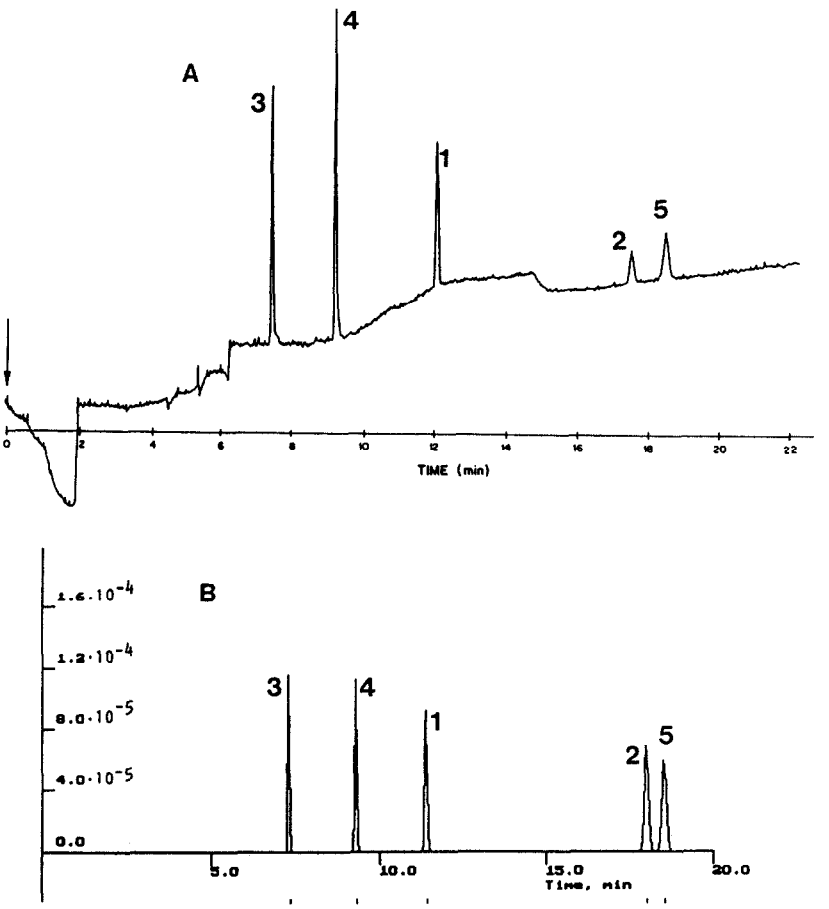

Fig. 5. Electropherograms of peptides from Table 5. (A) Conditions. as in Fig. 4, except buffer, 1.54 M acetic acid–0.66 M formic acid (pH 1.9) ($\mu_{eo} = 0$ m^2/s·V), and injection at 4500 V for 6 s (B) Simulated electropherogram.

directly from Refs. [27] and [36]. Charge data from Ref. [37] were not directly available, so they were calculated employing the same procedure as used in Refs. [27] and [36]. The presently obtained charge values ("q this work" in Table 1) were deduced employing the pK_a values obtained with our program.

The optimization of the parameters A, B, C, etc in Eqs. 1–5 was done by calculating the SSQ values between the experimental electrophoretic

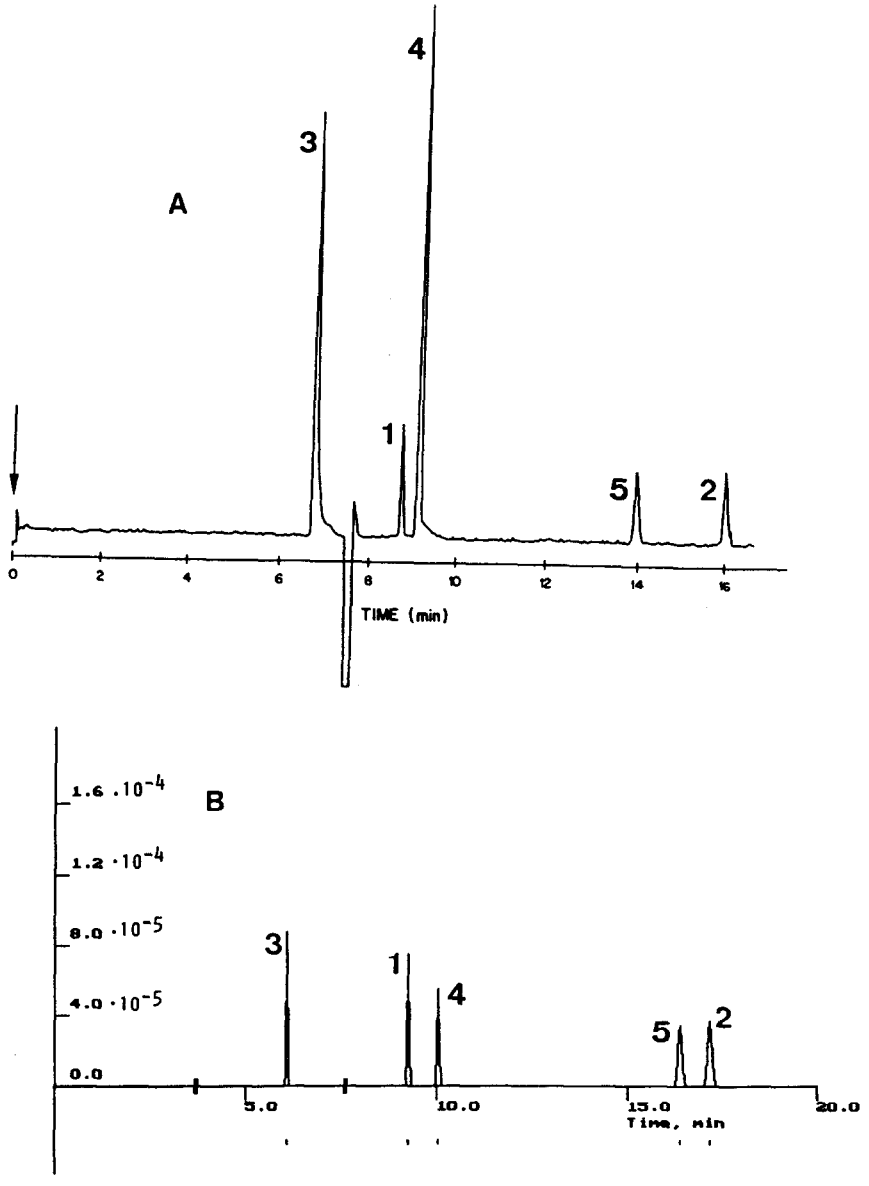

Fig. 6. Electropherograms of peptides from Table 5. (A) Conditions as in Fig. 4, except run voltage, 15 kV, buffer, 0.1 M Tricine–20 mM ethanolamine (pH 8.1) ($\mu_{eo} = 43.7 \cdot 10^{-9}$ m^2/s·V) and injection at 2000 V for 5 s. (B) Simulated electropherogram.

mobility of the 48 peptides and their respective predicted values, and adjusting the parameters until a minimum *SSQ* was obtained. The results with this optimization for the five equations were tested by calculating the electrophoretic mobility for two groups of five different peptides. The first group was measured at two values of pH (1.9 and 10.4) and the other at three values (1.9, 8.1 and 11.5). Then, 25 predicted mobility values from each equation were compared with the experimental data.

The results are shown in Table 2 in terms of σ, equal to SSO/α, where α is the number of data minus the number of parameters. The results show lower σ values for all the models when the charge values obtained in this work were employed instead of the average charges from the literature, so all later calculations were carried out employing the first values. To compare the five different equations a statistical F test was applied to theses σ values (Table 3). Considering a significance level of 5%, and a number of degrees of freedom ranging from 45 to 47 depending on the tested equation, we obtained

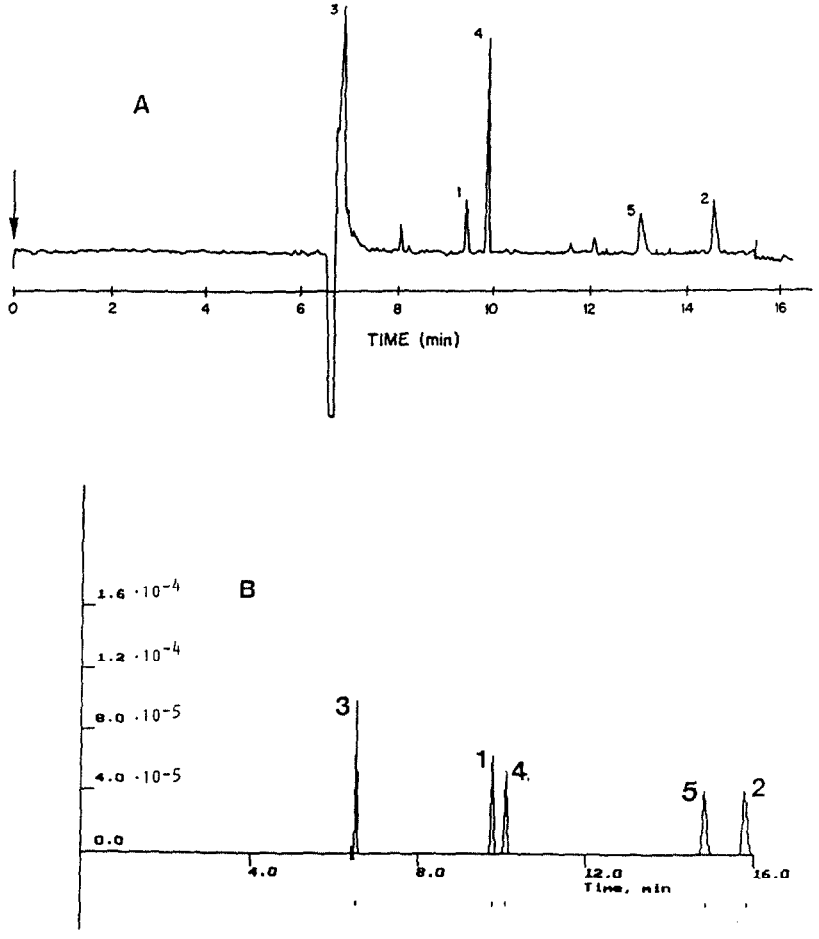

Fig. 7. Electropherograms of peptides from Table 5. (A) Conditions as in Fig. 4, except run voltage, 15 kV, buffer, 40 m*M* CAPS (pH 11.5) ($\mu_{eo} = 51.4 \cdot 10^{-9}$ $m^2/s \cdot V$) and injection at 2000 V for 5 s. (B) Simulated electropherogram.

from the statistical tables an average F value of 1.64. Employing this F value no significant differences ($\sigma_i^2/\sigma_j^2 < 1.64$) were observed between the Grossman, Offord and Compton equations (this result is marked with minus signs Table 3). Therefore, the ability of these equations to predict mobility values can be considered to be similar. However, when the Eq. 5 was compared with the other models significant differences were found ($\sigma_i^2/\sigma_j^2 > 1.64$). This result was corroborated when the σ values were calculated for the 25 mobility values from our experiments. As can be seen in Table 2, the best result ($\sigma = 1.72$) was also obtained employing Eq. 5.

3.4. Rapid identification of peptides knowing their amino acid sequence

The utility of employing the described mobility Eq. 5 and the corrected pK_a values for predicting peptides separations from its amino acid sequence can be seen in the simulated separation of six different peptides at pH 2.5. The correct migration order of peptides and their peak shape in the simulation (Fig. 1B) agrees fairly well with the real electropherogram [8] (Fig. 1A). In this instance, and also for Fig. 12 and 9, it was impossible to obtain the electroosmotic flow values from the literature, so these were adjusted in order to obtain similar migration times. It should be noted, however, that this is the only parameter adjusted; all other necessary data were derived at this stage from the amino acid sequence of the peptides. This also holds in Section 5, 6 and 7.

The rapid identification of peptides can be shown using the CE separation of five peptides obtained by Nyberg et al. [38] (Fig. 2A). They mixed the sample from an enzymatic reaction with each standard peptide in order to identify the different peaks. Using the computer program and the amino acid sequence of each peptide, their identification is obtained immediately (Fig. 2B). Identical results were achieved employing other separations from the same work (Fig. 3A), showing again that with the present scheme the correct migration order is predicted (Fig. 3B).

3.5. Choosing the best buffer and pH for peptide analysis in order to improve the separation efficiency and resolution

The application was carried out by simulating the separation of five short peptides with very similar size and pK values (Table 4) and five larger peptides (Table 5) ranging from a highly basic (e.g., peptide 3) to a highly acidic (e.g., peptide 2) with about 30 different running buffers with different pH values (data not shown). As the necessary time for each simulation is only a few seconds, the best buffers in terms of resolution and speed can be found quickly. The results (Figs. 4A, 5A, 6A and 7A) show that the experimentally obtained electropherograms agree fairly well with those predicted theoretically (Figs. 4B, 5B, 6B and 7B, respectively).

We also tested for these simulations the pK_a values given in Ref. [6] instead the pK_a values that we had obtained (given in Tables 4 and 5). Employing those pK values and for the simulation at pH 11.5 an incorrect migration order appeared (data not shown) between peaks 1 and 4. In order to obtain good agreement it was necessary to shift the pK value of peptide 1 from

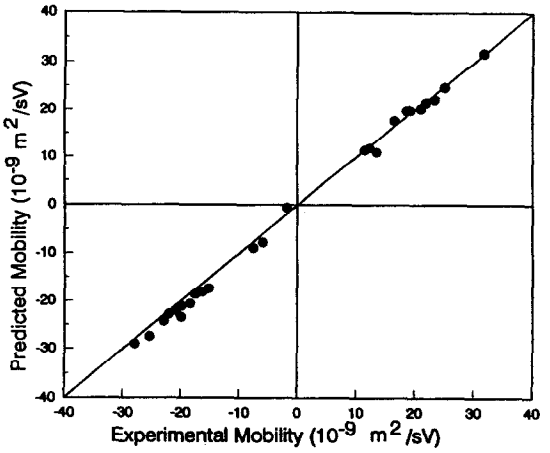

Fig. 8. Correlation between predicted and experimental mobility for the five peptides in Table 4 (at pH 1.9 and 10.4) and the five peptides in Table 5 (at pH 1.9, 8.1 and 11.5). The slope is 1.04 with a correlation coefficient of 0.9993 determined by linear regression.

10.3 to 11.8. This effect is probably the result of electrostatic interactions within the molecule, because in this peptide the side-chain charge on the amino acid lysine (K) is very close to the terminal charge on leucine (L). However, the program that we have developed considers this effect and such shifts of pK values were not necessary.

When the predicted mobility was plotted against the experimental mobility (Fig. 8) for those peptides we obtained good agreement, although five different buffer systems were used. The slope was 1.04 with a correlation factor of 0.9993 determined by linear regression.

The computer program also permits us to study the separation conditions in terms of sample capacity. Thus, buffer containing 0.1 M Tricine permitted injections ten times larger than that shown in Fig. 6A without a noticeable decrease in efficiency. This possibility should be useful in the future for micropreparative applications of CE.

3.6. Checking impurities assigned to peaks

As an example of this application we can consider the separation obtained by Nielsen and

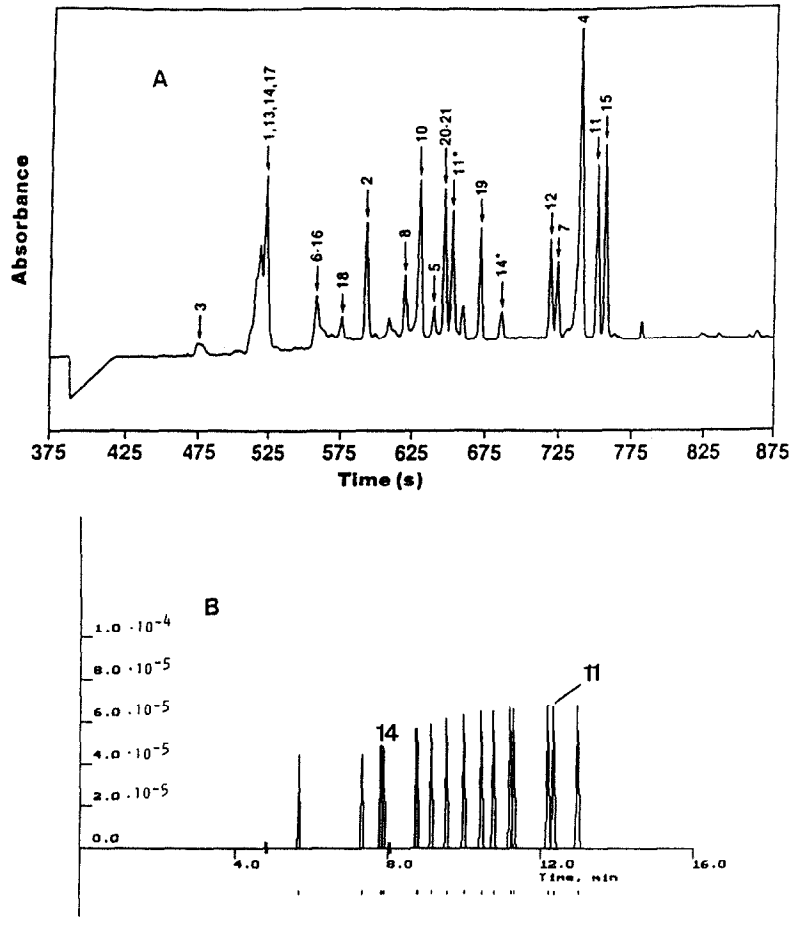

Fig. 9. Electropherogram of hGH digest. Sample concentration, 90 μM for each peptide. (A) Capillary, 100 cm (80 cm to detector) × 50 μm I.D. × 360 μm O.D.; separation voltage, 30 kV; buffer, 0.1 M Tricine–20 mM morpholine (pH 8.1); UV detection at 200 nm. Redrawn from Ref. [9]. (B) Simulated electropherogram: Injection length, 3 mm; electroosmotic flow considered to be $\mu_{eo} = 55 \cdot 10^{-9}$ m^2/s·V.

Rickard [9] (Fig. 9A). They isolated the peptides by RP-HPLC or anion-exchange chromatography, and their sequence was confirmed by spiking individual fragments for RP-HPLC or amino acid analysis. However, they had difficulty in distinguishing impurities 14* and 11* from the peptides 14 and 11, respectively; they showed [5] an initial error in the peak assignment [2] between peaks 14 and 14* because of their very similar composition. Comparing the predicted separation (Fig. 9B) with the real one (Fig. 9A), these peptides could be easily distinguished from their respective impurities taking into account the very different migration times for the proper peptides and their impurities.

Nevertheless, several problems appeared in the simulation of this separation. First, we could not include the peptides with double chain (numbers 6–16 and 20–21) in the simulated electropherogram because the computer program cannot predict the corrected pK_a values for these structures. Second, several peptides showed strange behaviour in the simulation when it was compared with the real electropherogram; a summary of these inconsistencies is given in Fig. 10. Namely, peptides 4, 5, 10, 15 and 17 show an incorrect order in the simulated electropherogram. This effect is probably due to the charge variation that several amino acids suffer around the separation pH employed (8.1), where sufficiently accurate predictions are much more difficult to carry out than at extreme values. In order to achieve a better simulation, in addition to the electrostatic effect [39,40] other effects that can influence the peptide mobility should also be considered, e.g., hydrophobic effects [7] and the influence of the different orientation of the side-chains [7] that can appear in long peptides. The determination of detailed long peptide surface charge distributions directly from the sequence is an unsolved problem [6,27,39,40].

3.7. Understanding the electrophoretic behaviour of various peptides

We employed the separation of six peptides at pH 11 obtained by Grossman et al. [8] (Fig. 11A). In their work they explained that "the poor peak shape seen in peak 3 is probably due to a perturbation of the electrical field caused by an increased conductivity in the sample band relative to the buffer". Looking at the simulated electropherogram (Fig. 11B) we can see that this explanation seems correct for peak 3. Nevertheless, they did not give any explanation about the poor shape of peak 5; if we compare the theoretically obtained peak 5 with that in the real electropherogram, we can see that, in this in-

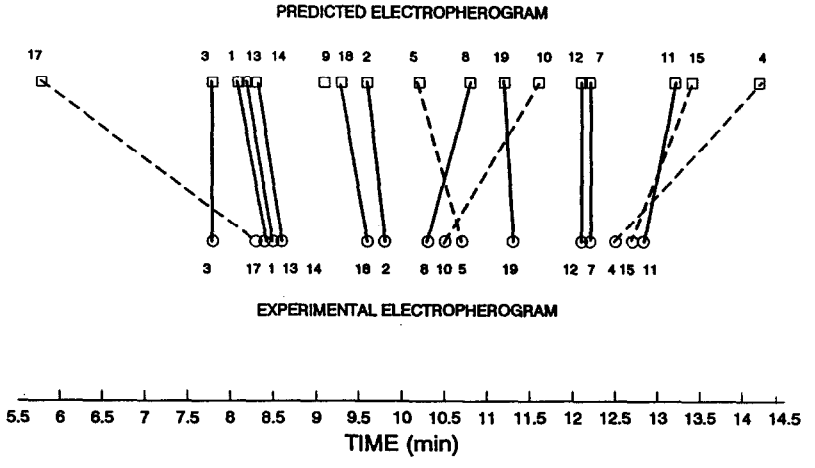

Fig. 10. Summary of the inconsistencies (dashed lines) observed in Fig. 9 between the real and predicted electropherogram.

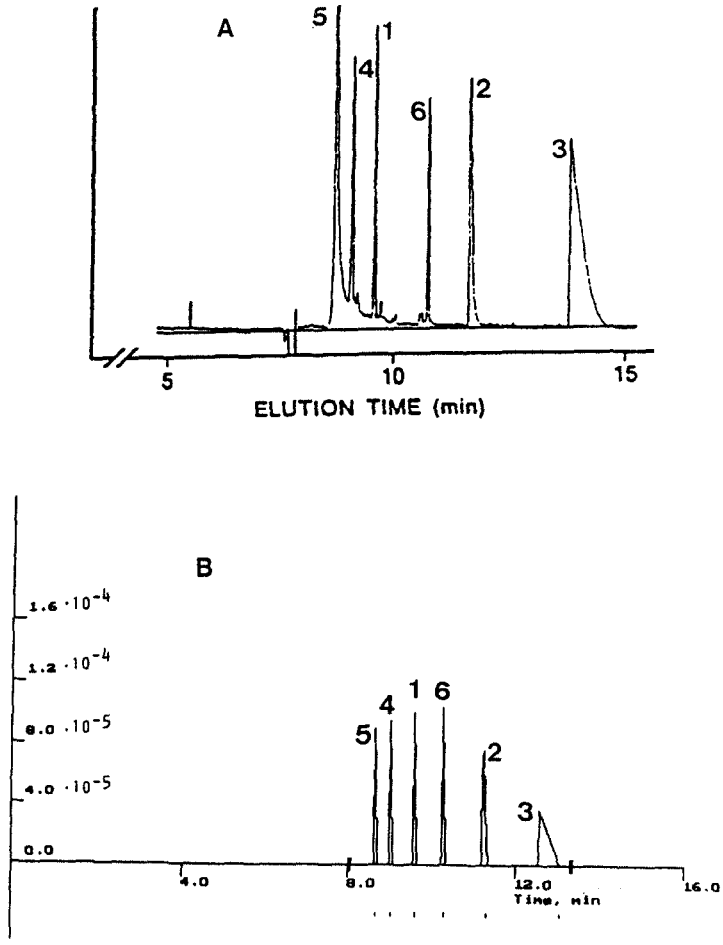

Fig. 11. Electropherogram of the six model peptides in Fig. 1A. (A) Capillary, 120 cm (100 cm to detector) × 50 μm I.D. × 320 μm O.D.; electric field, 250 V/cm ($\mu_{eo} = 83 \cdot 10^{-9}$ $m^2/s \cdot V$, calculated from the figure); buffer, 20 m*M* CAPS (pH 11); UV detection at 200 nm. Redrawn from Ref. [8]. (B) Simulated electropherogram: injection length, 4 mm; sample concentration, $1.2 \cdot 10^{-4}$–$1.8 \cdot 10^{-4}$ *M*.

stance, the explanation cannot be the conductivity difference. In our opinion the peak broadening more likely to be due to solute–capillary wall interactions, because of the strongly basic character of this peptide. A similar effect can be observed when comparing the peak shape obtained experimentally for peptide 3 at pH 8.1 (Fig. 6A) with the peak shape for the same peptide in the simulated electropherogram (Fig. 6B). Actually, the computer program can predict peak shapes, fronts or tailings, due to conductivity effects, injection and axial diffusion, fairly well but it cannot predict analyte–capillary wall adsorption effects. On the other hand, comparing the theoretical (Fig. 12A) and experimental (Fig. 12B) electropherograms at pH 4, we observed that two more peaks appear after 60 min of analysis.

Acknowledgements

We thank Hans Boelens and Raivo Kilg for assistance with the statistical calculations and data optimization. This work was supported by the Commission of the European Communities

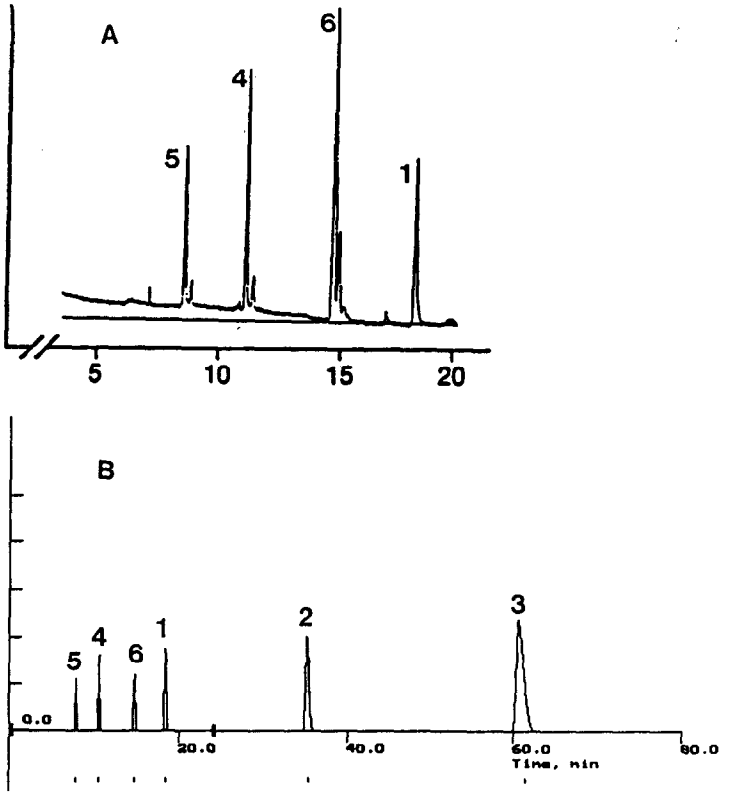

Fig. 12. (A) Conditions as in Fig. 1A except buffer pH, 4. Redrawn from Ref. [8]. Electropherogram simulated (B): injection length, 2 mm; sample concentration, $1.2 \cdot 10^{-4}$–$1.8 \cdot 10^{-4}$ M; electroosmotic flow, considered to be $\mu_{eo} = 11 \cdot 10^{-9}$ $m^2/s \cdot V$.

(Human Capital and Mobility Programme, bursary No. ERB4001GT920989).

References

[1] M. Novotny, K.A. Cobb and J. Liu, *Electrophoresis*, 11 (1990) 735.
[2] P.D. Grossman, J.C. Colburn, H.H. Lauer, R.G. Nielsen, R.M. Riggin, G.S. Sittampalam and E.C. Rickard, *Anal. Chem.*, 61 (1989) 1186.
[3] K.A. Cobb and M. Novotny, *Anal. Chem.*, 61 (1989) 2226.
[4] W.G. Kuhr and C.A. Monning, *Anal. Chem.*, 64 (1992) 389R.
[5] R.G. Nielsen, R.M. Riggin and E.C. Rickard, *J. Chromatogr.*, 480 (1989) 393.
[6] E.C. Rickard, M.M. Strohl and R.G. Nielsen, *Anal. Biochem.*, 197 (1991) 197.
[7] H.J. Gaus, A.G. Beck-Sickinger and E. Bayer, *Anal. Chem.*, 65 (1993) 1399.
[8] P.D. Grossman, K.J. Wilson, G. Petrie and H.H. Lauer, *Anal. Biochem.*, 173 (1988) 265.
[9] R.G. Nielsen and E.C. Rickard, *J. Chromatogr.*, 516 (1990) 99.
[10] H. Poppe, *Anal. Chem.*, 64 (1992) 1908.
[11] S.C. Smith and M.G. Khaledi, *Anal. Chem.*, 65 (1993) 193.
[12] T.A.A.M. van de Goor, P.S.L. Janssen, J.W. van Nispen, M.J.M. van Zeeland and F.M. Everaerts, *J. Chromatogr.*, 545 (1991) 379.
[13] J.C. Reijenga and E. Kenndler, *J. Chromatogr. A*, 659 (1994) 403.
[14] J.C. Reijenga and E. Kenndler, *J. Chromatogr. A*, 659 (1994) 417.
[15] E.V. Dose and G.A. Guiochon, *Anal. Chem.*, 63 (1991) 1063.
[16] R.A. Mosher, P. Gebauer and W. Thormann, *J. Chromatogr.*, 638 (1993) 155.
[17] J. Pospichal, P. Gebauer and P. Bocek, *Chem. Rev.*, 89 (1989) 419.
[18] T. Hirokawa, M. Nishino and Y. Kiso, *J. Chromatogr.*, 252 (1982) 49.

[19] J. Pospichal, M. Deml and P. Bocek, *J. Chromatogr.*, 390 (1987) 23.
[20] B. Skoog and A. Wichman, *Trends Anal. Chem.*, 5 (1986) 82. .
[21] A. Sillero and J.M. Ribeiro, *Anal Biochem.*, 179 (1989) 319.
[22] E.Q. Adams, *J. Am. Chem. Soc.*, 38 (1916) 1503.
[23] N. Bjerrum, *Z. Phys. Chem.*, 106 (1923) 219.
[24] A.E. Martell and R.M. Smith, *Critical Stability Constants: Amino Acids*, Vol. 1, Plenum Press, London, 1974.
[25] D.D. Perrin, *Dissociation Constants of Organic Bases in Aqueous Solution*, Butterworth, London, 1972.
[26] E.P. Serjeant and B. Dempsey, *Ionization Constants of Organic Acids in Aqueous Solution (IUPAC Chemical Data Series*, No. 23), Pergamon Press, Oxford, 1979.
[27] P.D. Grossman, J.C. Colburn and H.H. Lauer, *Anal. Biochem.*, 179 (1989) 28.
[28] J. Bongers, T. Lambros, A.M. Felix and E.P. Heimer, *J. Liq. Chromatogr.*, 15 (1992) 1115.
[29] R.E. Offord, *Nature*, 211 (1966) 591.
[30] H.J. Issaq, G.M. Janini, I.Z. Atamna, G.M. Muschik and J. Lukszo, *J. Liq. Chromatogr.*, 15 (1992) 1129.
[31] J. Frenz, S.L. Wu and W.S. Hancock, *J. Chromatogr.*, 480 (1989) 379.
[32] Z. Deyl, V. Rohlicek and M. Adam, *J. Chromatogr.*, 480 (189) 371.
[33] B.J. Compton, *J. Chromatogr.*, 559 (1991) 357.
[34] B.J. Compton and E.A. O'Grady, *Anal. Chem.*, 63 (1991) 2597.
[35] N. Chen, L. Wang and Y.K. Zhang, *Chromatographia*, 37 (1993) 429.
[36] V.J. Hilser, G.D. Worosila and S.E. Rudnick, *J. Chromatogr.*, 630 (1993) 329.
[37] T. Hirokawa, Y. Kiso, B. Gas, I. Zuskova and J. Vacik, *J. Chromatogr.*, 628 (1993) 283.
[38] F. Nyberg, M.D. Zhu, J.L. Liao and S. Hjerten, in C. Schafer-Nielsen (Editor), *Electrophoresis '88*, VCH, Weinheim, 1988, pp. 141–150.
[39] C. Tanford and J.G. Kirkwood, *J. Am. Chem. Soc.*, 79 (1957) 5333.
[40] J.B. Matthew, *Annu. Rev. Biophys. Biophys. Chem.*, 14 (1985) 387.