ELSEVIER

Journal of Chromatography A, 924 (2001) 291–306

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Peptide mapping by capillary zone electrophoresis: How close is theoretical simulation to experimental determination

George M. Janini*, Climaco J. Metral, Haleem J. Issaq

SAIC Frederick, National Cancer Institute at Frederick, P.O. Box B, Frederick, MD 21702, USA

Abstract

A multi-variable computer model is presented for the prediction of the electrophoretic mobilities of peptides at pH 2.5 from known physico–chemical constants of their amino acid residues. The model is empirical and does not claim any theoretical dependencies; however, the results suggest that, at least at this pH, peptides may be theoretically represented as classical polymers of freely joined amino acid residues of unequal sizes. The model assumes that the electrophoretic mobility can be represented by a product of three functions that return the contributions of peptide charge, length and width, respectively to the mobility. The model relies on accurate experimental determination of the electrophoretic mobilities of a diverse set of peptides, by capillary zone electrophoresis (CZE), at 22°C, with a 50 m*M* phosphate buffer, at pH 2.5. The electrophoretic mobilities of a basis set of 102 peptides that varied in charge from 0.65 to 16 and in size from two to 42 amino acid residues were accurately measured at these fixed experimental conditions using a stable 10% linear polyacrylamide-coated column. Data from this basis set was used to derive the peptide charge, length, and width functions respectively. The main purpose of this endeavor is to use the model for the prediction of peptide mobilities at pH 2.5, and for simulation of CZE peptide maps of protein digests. Excellent agreement was obtained between predicted and experimental electrophoretic mobilities for all categories of peptides, including the highly charged and the hydrophobic. To illustrate the utility of this model in protein studies it was used to simulate theoretical peptide maps of the digests of glucagon and horse cytochrome *c*. The resulting maps were compared and contrasted with their experimental counterparts. The potential of this approach and its limitations are discussed. Published by Elsevier Science B.V.

Keywords: Peptide mapping; Electrophoretic mobility; Computer simulation; Glucagon; Cytochromes; Proteins

1. Introduction

Peptide mapping by capillary zone electrophoresis (CZE) [1–7], and capillary electrophoresis–mass spectrometry (CZE–MS) [1,8,9], is increasingly being utilized as a complement, if not a viable substitute, for the already established technique of high-performance liquid chromatography (HPLC). Capillary zone electrophoresis has several advantages over HPLC, including speed, higher resolution and smaller sample size requirement. Peptide mapping by HPLC is typically conducted on reversed-phase columns using gradient elution. Although recent advances in column technology have improved the reproducibility of peptide maps, the separation column, guard column and solvent filters are still subject to deterioration with repeated use. The HPLC method is time consuming because columns need to be conditioned and equilibrated. In contrast, peptide mapping in CZE, using polyacrylamide coated columns and acidic buffers, is fast, simple and rugged. While HPLC samples have

*Corresponding author. Tel.: +1-301-846-7189; fax: +1-301-846-6037.

E-mail address: janini@ncifcrf.gov (G.M. Janini).

0021-9673/01/$ – see front matter Published by Elsevier Science B.V.
PII: S0021-9673(01)00919-0

to be run one at a time, multiple samples can be run simultaneously using the recently introduced multiplexing CZE instruments [6]. Moreover, capillary electrophoresis offers a straight-forward correlation of migration time with physical properties, making it possible to theoretically predict the electrophoretic mobility of a peptide from knowledge of its amino acid content.

In peptide mapping, a protein is fragmented into a pool of smaller peptides by specific chemical reactions or proteolytic enzymes. The pool of fragments is typically separated by HPLC or CZE yielding a peptide map that serves as a fingerprint for protein identification. In addition, protein sequence information can be obtained if MS is used in tandem with HPLC or CZE. Peptide mapping is being used to establish the fidelity of protein translation in biological system, and to provide information on protein variants and disulfide linkage. It is also used to identify post translational modifications and chemical degradation.

Peptides are relatively simple structures that are efficiently separated by CZE, and their electrophoretic migration characteristics as a function of pH are amenable to prediction by theoretical models [10–23]. The subject has been extensively reviewed [24–27]. All these models, in one way or another, are derived from Stoke's law for the motion of ions in an electric field. Stoke's law states that when a charged particle is placed in an electric field, it experiences a force that is proportional to its effective charge (Q) and the electric field strength (E). The translational movement of the particle is opposed by a viscous drag force that is proportional to the particle's velocity (v), its hydrodynamic radius (r) and the viscosity of the medium (η). When the two forces are counterbalanced, the particle moves with a steady state velocity [28].

$$v_{ef} = \mu_{ef} E = \mu_{ef} \cdot \frac{V}{L} \quad (1)$$

where V = applied voltage, L = column length and μ_{ef} = the electrophoretic mobility, which is given by:

$$\mu_{ef} = \frac{Q}{4\pi r \eta} \quad (2)$$

Strictly speaking, Stoke's law is only valid for rigid spherical molecules in low ionic strength buffers. The situation is not as simple for molecules that are large or even small, but not spherical. For example, for rod-shaped particles, the aspects ratio and the orientation with respect to the direction of motion enter into play [11].

Several semi-empirical models have been advanced to explain the dependence of peptide electrophoretic mobility on charge and size [24–27]. Most models estimate the charge from the ionization constants of the amino acids, and relate (r) to the molar mass (M) or the number of amino acid residues (N). In summary, the various models suggest a direct dependence of μ_{ef} on Q and a size dependency ranging from $1/r$ for small molecules to $1/r^2$ for large molecules. Assuming that the molar mass is proportional to the molar volume, and that a molecule can be represented by a sphere of volume = $(4/3)\pi r^3$, then r can be replaced by $M^{1/3}$ and r^2 by $M^{2/3}$. An interesting discussion of the assumptions involved in the derivation of the different models and the conditions under which these assumptions are valid was presented by Cross and Cao [22]. Correlation of peptide electrophoretic mobility with $1/r^2$ ($M^{2/3}$) rather than $1/r$ ($M^{1/3}$), as stipulated by the Stoke's model, is largely attributed to Offord [29] who theorized that large non-spherical ions moving through a conducting medium experience a retarding force that is proportional to their surface areas and not their radii. The Stoke and Offord models were combined by Compton [15] who argued that the dependence of mobility on molar mass is a continuous function ranging from $1/M^{1/3}$ to $1/M^{2/3}$. Small molecules in low ionic strength buffer are more closely correlated with $1/M^{1/3}$ while large molecules in high ionic strength buffer are better correlated with $1/M^{2/3}$. Molecules of intermediate-size in medium-strength buffers show dependence on $1/M^{1/2}$ [15]. Taking a different approach, Grossman et al. [13] considered the peptide as a classical linear polymer with N amino acid residues and arrived at an equation that correlates μ_{ef} with ln $[(Q+1)/N^{0.43}]$. The logarithmic dependence of Q was introduced to compensate for charge suppression due to mutual electrostatic interactions of the charged groups, which becomes increasingly significant for highly charged peptides [13,27]. A modification to the classical linear model of Grossman et al. [13], was reported by Cifuentes and Poppe [10], who

retained the logarithmic dependence of mobility on charge but substituted N for M for the size dependence.

Successful correlations using any of the models described above require an accurate determination of charge. The fundamental equation that is invariably used for the calculation of the net charge on a peptide is the Henderson–Hasselbalch equation [30]. The use of this equation requires an accurate knowledge of the ionization constants of the amino acid residues. Since these values are not accurately known, most researchers resort to the use of a standard set of pK values [14], assuming that the ionization of each amino acid residue is not affected by electrostatic and steric interactions with neighboring residues [27]. Another source of error in these correlations is the use of the molar mass of the peptide to represent its hydrodynamic radius in solution. Different peptides of similar molar masses might have different hydrodynamic radii depending on their conformation in solution.

Recently Janini et al. [31] determined the electrophoretic mobility of 58 peptides ranging in size from two to 39 amino acids and in charge from 0.65 to 7.82. The electrophoretic data was used to test existing theoretical models that correlate electrophoretic mobility with charge and size. The results showed that the Offord model gave the best overall mobility. The model, however, failed when applied to hydrophobic and highly-charged peptides because it does not account for peptide hydrophobicity and the phenomenum of shielding of charge in highly charged peptides. The deficiency in the Offord model and the other existing models led us to conclude that peptide electrophoretic mobility cannot be successfully predicted with reasonable degree of accuracy for all different categories of peptides by relying on two parameter models, namely Q and M or N. This prompted us to investigate a computer model that gives peptide electrophoretic mobility as a product of several functions that contain the different variables that influence the motion of peptides in an electric field [32]. The model gave excellent correlation between predicted and measured electrophoretic mobilities of 62 peptides of various charges, sizes, and categories.

This work presents an improved version of our previous model. The predictive ability of the model was enhanced by expanding the basis set from 62 to 102 peptides and by incorporating a closest-neighbor algorithm in the model. The electrophoretic mobility of an unknown peptide is calculated as follows. First, the unknown peptide is matched to its closest neighbor in the basis set using a closest-neighbor algorithm. Second, the experimental electrophoretic mobility of the closest-neighbor is used as a first order approximation, and corrected applying the multi-variable equation on the difference between the unknown and its closest-neighbor.

The model was used to calculate the electrophoretic mobilities of all theoretical fragments of a polypeptide and a protein of defined fragmentation patterns. The electrophoretic mobilities were converted to migration times and the CZE electropherograms were simulated with a Gaussian function where the peak maximum is the calculated migration time and the peak area at the detection wavelength of 200 nm is assumed to be proportional to $N-1$, where N is the number of amino acids in the peptide [33,34].

2. Materials and methods

2.1. Materials

The peptides used in this study are listed in Table 1, together with their physical parameters, N, M, Q, and W, where W is the average residue mass and N, M and Q are as defined earlier. The net charge of the peptide at pH 2.5, was calculated as the algebraic sum of all charged amino acid residues and carboxyl- and amino-terminals. Each arginine (R), lysine (K), histidine (H) and the amino terminal contribute a positive charge of $+1$. Each aspartic acid (D), glutamic acid (E) and the carboxyl terminal contribute a partial negative charge, calculated using the Henderson–Hasselbach relation [30]. The pK_a values used were 3.20 for the carboxyl terminal, 3.50 for D residues and 4.50 for E residues [14]. Despite the criticism that the ionization constants of the carboxy groups may vary depending on their local environment, most investigators use a standard set [27]. According to Rickard et al. [14] good agreement is to be expected between calculated and actual charge, since at this pH, the carboxy groups are nearly

Table 1
List of peptides used in this study together with their relevant parameters

Peptide sequence	*N*	*W*	*Q*	*M*
AA	2	15.00	0.83	160.15
AAA	3	15.00	0.83	231.19
AAAA	4	15.00	0.83	302.24
AAAAA	5	15.00	0.83	373.28
AAANLVPMVATV	12	38.67	0.83	1154.59
AAGIGILTV	9	32.22	0.83	812.46
ACHGRDRRT	9	60.78	4.74	1069.46
ACLGRDRRTEE	11	60.82	3.72	1303.54
ACPGKDRRTGGGN	13	41.54	3.74	1286.63
ACPGRNRRTEEENL	14	60.00	3.80	1642.68
ACPGTDRRTGGGN	13	39.38	2.74	1258.63
ACSGRDRRTEE	11	58.45	3.72	1277.54
AFLPWHRLF	9	73.67	2.83	1185.46
ANSK	4	44.00	1.83	418.24
CRHHRRRHRRGC	12	78.17	9.83	1628.59
CRHRRRHRRGC	11	77.91	8.83	1491.54
DAEKSDICTDEY	12	58.00	1.54	1386.59
DD	2	59.00	0.65	248.15
DGLAPPQHRIRVEGNLR	17	56.24	4.73	1926.81
DRVIEVVQGAYRAIRHIPRRIRQGLERRIHIGPGRAFYTTKN	42	61.74	12.72	4964.91
EE	2	73.00	0.81	276.15
EPPEVGSDYHHPLQLHV	17	57.76	3.72	1952.81
FA	2	53.00	0.83	236.15
FD	2	75.00	0.74	280.15
FF	2	91.00	0.83	312.15
FFF	3	91.00	0.83	459.19
FG	2	46.00	0.83	222.15
FIGITEAAANLVPMVATV	18	43.72	0.82	1813.85
FL	2	74.00	0.83	278.15
FLTPKKLQCVDLHVISNDVCAQVHPQKVTK	30	56.20	6.65	3385.38
FV	2	67.00	0.83	264.15
GG	2	1.00	0.83	132.15
GIGAVLK	7	35.14	1.83	656.37
GSDCTTIHCNYM	12	54.25	1.74	1341.59
HG	2	41.00	1.83	212.15
HMTE	4	68.25	1.82	515.24
HMTEVVRHCPHHER	14	68.79	6.81	1765.68
HMTEVVRRYPHHER	14	74.42	6.81	1844.56
HRSCRRRKRRSCRHR	15	79.33	11.83	2048.72
IITLEDSSNLLGRNSF	16	53.88	1.73	1776.77
KKK	3	72.00	3.83	402.19
KKKK	4	72.00	4.83	530.24
KKKKK	5	72.00	5.83	658.28
KLVVVGAAGV	10	33.30	1.83	911.50
KLVVVGADGV	10	37.70	1.74	955.50
KLVVVGAGDVGKSALTI	17	38.47	2.74	1624.81
KQIINMWQEVGKAMYAPPISGQIRRIHIGPGRAFYTTKN	39	58.03	7.82	4466.95
KSSQYIKANSKFIGITE	17	55.35	3.82	1911.81
KSSQYIKANSKFIGITEAAANLVPMVATV	29	48.45	3.82	3048.34

Table 1. Continued

Peptide sequence	*N*	*W*	*Q*	*M*
LAKTCPVRLWVDSTPP	16	54.06	2.74	1779.77
LAKTYPVQLWVDS	13	59.31	1.74	1517.63
LAPPQHLIQVEGNLRV	16	54.25	2.82	1782.77
LDDRNTFRRSVVVPYE	16	65.56	3.64	1963.77
LGRNSFEVCVCACPGRD	17	50.24	2.73	1824.81
LL	2	57.00	0.83	244.15
LLGRNSFEMRV	11	62.36	2.82	1320.55
MGGMNWRPILTIIT	14	56.93	1.83	1599.68
MLDLQPETT	9	58.00	0.73	1044.46
MM	2	75.00	0.83	280.15
NHQLLSPAKTGWRIFHP	17	60.53	4.83	1999.81
NSFCMGGMNRR	11	57.91	2.83	1271.54
NTFRHSVVEPYEPPEVG	17	57.88	2.80	1954.81
PARR	4	64.00	2.83	498.24
PG	2	21.00	0.83	172.15
PHRERCSDSDGL	12	56.67	3.64	1370.59
PPPGTRVRVMAIYKQSQ	17	56.18	3.83	1925.81
RK	2	86.00	2.83	302.15
RPKPQQFFGLM	11	64.82	2.83	1347.54
RPPGF	5	54.80	1.83	572.28
RPPGFSPFR	9	59.67	2.83	1059.46
RQQ	3	81.33	1.83	430.19
RTHCQSHYRRRHCSR	15	74.80	8.83	1980.72
RTHGQSHYRRRHCSRRRLHRIHRRQ	25	76.96	15.83	3343.16
SPALNKMFCELAKT	14	53.43	2.82	1550.68
SSCMGGMNQRPILTIIT	17	49.88	1.83	1818.81
SSQYIK	6	61.67	1.83	724.33
SSS	3	31.00	0.83	279.19
TPPPGTRVQQSQHMTEV	17	54.00	2.82	1888.81
TTIHYNYICNSS	12	60.17	1.83	1412.59
TYSPALNRMFCQLAKT	16	57.88	2.83	1840.77
VISNDVCAQV	10	46.80	0.74	1046.50
VLQELNVTV	9	54.44	0.82	1012.46
VLTTGLPALISWIK	14	50.43	1.83	1508.68
VPYEPPEVGSVYHHPLQLHV	20	57.85	3.81	2295.94
VV	2	43.00	0.83	216.15
VVRRCPHQRCSDSDGL	16	57.00	4.65	1826.77
VVRRYPHHE	9	74.33	4.82	1191.46
WW	2	130.00	0.83	390.15
YAEGDVHATSK	11	49.18	2.73	1175.55
YAEGDVHATSKPARR	15	53.13	4.73	1655.72
YGGFL	5	51.40	0.83	555.28
YGGFM	5	55.00	0.83	573.28
YKLVVVGAAGVGKSALT	17	38.82	2.83	1630.81
YKLVVVGACGVKGSALT	17	40.71	2.83	1662.81
YKLVVVGANGVGKSALT	17	41.35	2.83	1673.81
YKLVVVGARGVGKSALT	17	43.82	3.83	1715.81
YKLVVVGAVGVGKSALT	17	40.47	2.83	1658.81
YLSGADLNL	9	49.11	0.74	964.46
YMDGTMSQV	9	56.33	0.74	1029.46
YNYMCNSSGMGGMNRRP	17	56.82	2.83	1936.81
YSPALNKMCCQLAKT	15	54.00	2.83	1668.72
YY	2	107.00	0.83	344.15

N = number of amino acids; *Q* = charge at pH = 2.5; *W* = average residue width; *M* = molar mass in g/mol.

completely protonated. Exceptions to this assertion are to be expected for highly hydrophobic peptides (charge <1), where variations in the ionization constants of D and E due to their local environment might be significant, and for highly charged peptides, where the calculated charge might deviate from actual charge due to mutual electrostatic interactions of charged groups in proximity of each other. Table 1 includes 18 dipeptides, 32 peptides with five or less amino acid amino acid residues and 72 peptides with six or more amino acid residues. The peptides range in size from two to 42 amino acid residues and in calculated charge from 0.65 to 16. The dipeptides and most of the small peptides (five amino acid residues or less), endoproteinases, Lys-C and Glu-C, horse cytochrome *c*, and glucagon were purchased from Sigma (St. Louis, MO, USA). The rest were custom synthesized by commercial laboratories for Dr. J.A. Berzofsky (NCI, NIH, USA). The buffer components were purchased from Fisher Scientific (Pittsburgh, PA, USA). Only one buffer system was used throughout. The buffer was made up of 50 m*M* phosphoric acid that was adjusted to pH 2.5 with triethylamine (TEA). All reagents were used as received. The fused-silica columns were purchased from Polymicro Technologies (Phoenix, AZ, USA). Proteins were digested by incubating a 100 μg amount of each with 10 μg of the appropriate endoproteinase for 37°C in 100 μl of 25 m*M* Tris–HCl, pH 8.5.

2.2. Apparatus and procedures

A Beckman CZE Model P/ACE 5510 equipped with a UV detector, an automatic injector, a fluid-cooled column cartridge, and a System Gold data station were used in this study. All runs were performed at 200 nm and 22°C. The buffers were prepared fresh daily, passed through 0.2 μm nylon filters, and degassed. Injections were made using the pressure mode at 0.5 p.s.i. (1 p.s.i. = 6894.76 Pa). The capillary inlet and outlet vials were replenished after every ten injections. The columns were coated with a dense layer of 10% polyacrylamide as described in detail elsewhere [35]. The polyacrylamide-coated columns provided stability and migration time reproducibility throughout the experiments to within 1% RSD. This was established by periodically monitoring the electroosmotic mobility and measuring the migration time of a set of reference solutes. Because of the very small electroosmotic mobility exhibited by these columns an indirect procedure was used for its measurement as described elsewhere [31]. Compound 4-dimethylaminopyridine (Sigma, St. Louis, MO, USA) was selected as a reference to be run with every peptide as a check on column stability and migration time reproducibility, and also to offset minor day-to-day variations in electroosmotic mobility, if present [31]. The electrophoretic mobilities of peptides were measured as follows:

For each peptide, a 0.25–0.5 mg/ml solution was made, injected, and the migration time [t_m(pep)] recorded. Simultaneously a sample of the reference was introduced as a second injection and its migration time recorded, t_m(ref). The apparent electrophoretic mobility of each peptide (μ_{app}(pep)) was determined from the equation:

$$\mu_{app}(\text{pep}) = \mu_{app}(\text{ref}) \cdot \frac{t_m(\text{ref})}{t_m(\text{pep})} \quad (3)$$

Finally, the electrophoretic mobility of each peptide (μ_{ef}) was determined from the equation:

$$\mu_{ef}(\text{pep}) = \mu_{app}(\text{pep}) - \mu_{eo} \quad (4)$$

where:

$$\mu_{eo} = \mu_{app}(\text{ref}) - 30.02 \cdot 10^{-5}\ \text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1} \quad (5)$$

μ_{app}(ref) was simply calculated from the equation:

$$\mu_{app}(\text{ref}) = \frac{Ll}{Vt_m(\text{ref})} \quad (6)$$

where L = total column length in cm, l = injector-to-detector column length in cm, V = applied voltage in volts and t_m(ref) = the migration time of the reference in seconds. The quantity, $30.02 \cdot 10^{-5}$ cm^2 V^{-1} s^{-1} is the electrophoretic mobility of the reference that was accurately measured at the beginning of the experiment using the average value of several determinations with two independent columns. This procedure ensures the accuracy and consistency of peptide electrophoretic mobility values as μ_{eo} is accounted for with every data point.

The electrophoretic mobility of six peptides and the reference standard were measured with each of two columns using two different buffer preparations

Table 2
Comparison of μ_{ef} data measured with two independent experimental setups[a]

Peptide	$\mu_{ef} \times 10^5$			
	Column 1	Column 2	Average±Std.	RSD (%)
AA	19.00	19.53	19.27±0.27	1.37
AAA	15.80	15.00	15.43±0.40	2.60
YKLVVVGACGVGKSALT	13.82	14.86	14.34±0.52	3.62
VPYEPPEVGSVYHHPLQLHV	15.13	15.51	15.32±0.19	1.24
DGLAPPQHRIRVEGNLR	20.00	18.98	19.49±0.51	2.62
VLQELNVTV	6.33	6.97	6.65±0.32	4.81
4-dimethyl amino pyridine (ref.)	30.06	29.98	30.02±0.04	0.13
Average				2.34

μ_{ef} values in $cm^2\ V^{-1}\ s^{-1}$.
[a] Column 1=47 cm and Column 2=37 cm. Each data point in columns 1 and 2 is an average of three determinations to within ±1% RSD.

in separate days. The results are presented in Table 2. Excellent repeatability was obtained with each column (±1%); however, the RSD of the averages from the two columns varied from 0.13 to 4.81 with an average of 2.34% for the set. This, we believe, represents the level of uncertainty of our experimental determination of electrophoretic mobilities.

3. Results and discussion

The main objective of this study is to arrive at a model that accurately predicts the electrophoretic mobilities of peptides at pH 2.5 from knowledge of physico–chemical parameters of their amino acid constituents. Once identified, the model is to be used for the practical goal of simulation of peptide maps of protein digests. The approach selected is purely phenomenological and empirical. It assumes that the electrophoretic mobility is fully contained in three functions representing charge, number of amino acids and average residue width respectively, but it does not imply any theoretical dependencies. The success of this approach, as measured by how close the simulation is to experimental values, is dependent on the accurate measurement of electrophoretic mobilities of a large number of peptides with wide range of charge and molar mass. For this purpose a diverse set of peptides was assembled (Table 1) and their electrophoretic mobilities were accurately measured, under fixed experimental conditions, as described in the experimental section and listed in Table 3. A pH of 2.5 was selected for this application because all peptides are positively charged at this pH and move in the same direction towards the cathode where the detector is placed. Also, at this pH, one would expect a very good agreement between calculated and actual charge, which enhance the predictive ability of the computer model. The 10% polyacrylamide-coated columns are particularly suited for this application because (a) at pH 2.5 the coating is chemically stable and resistant to hydrolytic degradation, (b) it is hydrophilic, (c) it provides a thick layer of surface coverage that effectively eliminates electroosmotic flow and shields the peptides from direct contact with silanol groups on the silica surface, (d) it does not interact with the peptides by adsorption or otherwise, thus ensuring high column efficiency, and (e) the coating procedure yields reproducible columns with consistent batch-to-batch migration times (Table 2). In a recent publication [31] we tested the existing theoretical models that correlate electrophoretic mobility with charge and size parameters and concluded that the Offord model that correlates electrophoretic mobility with $Q/M^{2/3}$ is superior to the other models tested. Based on a data set of 58 peptides we arrived at the equation: $\mu_{ef} = 10^{-5}(2.44 + 581.85\ \ Q/M^{2/3})$ $cm^2\ V^{-1}\ s^{-1}$. The equation offered the best correlation between calculated and experimental μ_{ef} values for most peptides in the data set, with the exception of the highly charged and the hydrophobic. This is because the Offord model and the other two-parameter models cannot account for differences in shape,

Table 3
Comparison of experimental electrophoretic mobilities at pH = 2.5 with theoretical calculations according to our multi-variable model and Offord's model

Peptide sequence	$\mu_{ef} \times 10^5$ Multi-var.	$\mu_{ef} \times 10^5$		$\mu_{ef} \times 10^5$	
		Exp.	% Dev.	Offord's	% Dev.
AA	19.37	18.77	−3.20	18.89	+ 0.63
AAA	15.77	14.96	−5.39	15.32	−2.39
AAAA	13.22	13.87	4.67	13.21	4.73
AAAAA	12.60	12.34	−2.12	11.80	4.37
AAANLVPMVATV	6.10	6.15	0.78	6.85	−11.32
AAGIGILTV	6.53	6.50	−0.46	8.01	−23.19
ACHGRDRRT	28.54	26.54	−7.55	28.80	−8.50
ACLGRDRRTEE	21.69	20.97	−3.42	20.58	1.88
ACPGKDRRTGGGN	19.94	19.11	−4.34	20.83	−9.01
ACPGRNRRTEEENL	19.23	19.40	0.88	18.32	5.54
ACPGTDRRTGGGN	15.28	15.08	−1.31	16.11	−6.84
ACSGRDRRTEE	21.18	21.91	3.31	20.82	4.97
AFLPWHRLF	18.16	16.55	−9.73	17.15	−3.65
ANSK	21.37	20.91	−2.22	21.51	−2.87
CRHHRRRHRRGC	29.61	29.68	0.24	43.74	−47.38
CRHRRRHRRGC	29.75	29.68	−0.24	41.78	−40.78
DAEKSDICTDEY	10.02	9.91	−1.07	9.65	2.68
DD	10.33	10.31	−0.19	12.05	−16.84
DGLAPPQHRIRVEGNLR	20.49	18.98	−7.95	20.21	−6.51
DRVIEVVQGAYRAIRHIPRRIRQGLERRIHIGPGRAFYTTKN	21.79	20.83	−4.61	27.85	−33.72
EE	13.16	12.52	−5.14	13.61	−8.68
EPPEVGSDYHHPLQLHV	17.74	16.91	−4.89	16.30	3.62
FA	14.83	14.86	0.23	15.14	−1.87
FD	12.85	13.00	1.18	12.54	3.57
FF	13.18	12.81	−2.91	12.98	−1.36
FFF	10.76	10.38	−3.67	10.59	−2.05
FG	15.13	15.16	0.19	15.67	−3.35
FIGITEAAANLVPMVATV	4.97	4.73	−5.13	5.66	−19.72
FL	13.91	13.33	−4.33	13.82	−3.70
FLTPKKLQCVDLHVISNDVCAQVHPQKVTK	19.83	18.68	−6.14	19.59	−4.87
FV	13.67	13.90	1.66	14.22	−2.32
GG	21.35	21.70	1.61	21.14	2.58
GIGAVLK	16.24	15.50	−4.75	16.56	−6.85
GSDCTTIHCNYM	12.39	12.41	0.16	10.77	13.25
HG	27.05	27.04	−0.03	32.42	−19.90
HMTE	18.61	18.91	1.56	18.92	−0.06
HMTEVVRHCPHHER	25.07	26.41	5.06	29.55	−11.89
HMTEVVRRYPHHER	27.11	26.42	−2.62	30.32	−14.77
HRSCRRRKRRSCRHR	31.41	30.27	−3.76	45.09	−48.97
IITLEDSSNLLGRNSF	11.58	11.33	−2.21	9.31	17.83
KKK	32.54	33.03	1.49	43.35	−31.25
KKKK	33.53	33.03	−1.51	45.34	−37.27
KKKKK	33.18	33.03	−0.46	47.26	−43.09
KLVVVGAAGV	14.01	14.10	0.61	13.79	2.23
KLVVVGADGV	13.21	13.13	−0.61	12.89	1.83
KLVVVGAGDVGKSALTI	13.94	13.69	−1.80	13.98	−2.11
KQIINMWQEVGKAMYAPPISGQIRRIHIGPGRAFYTTKN	17.71	17.78	0.39	19.21	−8.04
KSSQYIKANSKFIGITE	18.24	17.05	−6.98	16.87	1.04
KSSQYIKANSKFIGITEAAANLVPMVATV	13.52	14.21	4.85	13.01	8.41

Table 3. Continued

Peptide sequence	$\mu_{ef} \times 10^5$ Multi-var.	$\mu_{ef} \times 10^5$		$\mu_{ef} \times 10^5$	
		Exp.	% Dev.	Offord's	% Dev.
LAKTCPVRLWVDSTPP	14.25	15.13	5.78	13.29	12.12
LAKTYPVQLWVDS	9.91	10.51	5.75	10.11	3.77
LAPPQHLIQVEGNLRV	14.66	15.01	2.33	13.61	9.33
LDDRNTFRRSVVVPYE	17.47	18.30	4.54	15.94	12.88
LGRNSFEVCVCACPGRD	14.12	13.66	−3.34	13.09	4.21
LL	14.58	14.55	−0.23	14.85	−2.09
LLGRNSFEMRV	17.11	17.02	−0.51	16.09	5.49
MGGMNWRPILTIIT	10.18	10.20	0.22	10.23	−0.31
MLDLQPETT	6.51	6.33	−2.90	6.58	−3.93
MM	13.28	13.86	4.15	13.77	0.69
NHQLLSPAKTGWRIFHP	18.85	19.42	2.95	20.14	−3.72
NSFCMGGMNRR	17.39	18.30	4.98	16.48	9.93
NTFRHSVVEPYEPPEVG	14.12	13.55	−4.20	12.87	5.03
PARR	26.21	27.65	5.23	28.66	−3.65
PG	17.86	18.43	3.10	18.11	1.71
PHRERCSDSDGL	20.98	19.33	−8.53	19.61	−1.43
PPPGTRVRVMAIYKQSQ	17.01	18.20	6.52	16.84	7.48
RK	30.41	32.00	4.98	39.03	−21.98
RPKPQQFFGLM	16.89	16.98	0.51	15.95	6.07
RPPGF	18.63	18.36	−1.47	17.91	2.43
RPPGFSPFR	19.17	19.71	2.74	18.30	7.15
RQQ	22.55	24.00	6.04	21.15	11.86
RTHCQSHYRRRHCSR	26.39	28.96	8.86	34.99	−20.84
RTHGQSHYRRRHCSRRRLHRIHRRQ	28.32	29.01	2.36	43.60	−50.31
SPALNKMFCELAKT	15.35	15.71	2.27	14.69	6.47
SSCMGGMNQRPILTIIT	10.20	10.66	4.32	9.59	10.00
SSQYIK	16.47	16.71	1.45	15.66	6.26
SSS	13.20	13.22	0.12	13.80	−4.36
TPPPGTRVQQSQHMTEV	14.44	14.17	−1.93	13.18	7.01
TTIHYNYICNSS	11.24	10.59	−6.10	10.90	−2.94
TYSPALNRMFCQLAKT	14.01	14.77	5.15	13.41	9.24
VISNDVCAQV	6.03	5.83	−3.32	6.63	−13.72
VLQELNVTV	7.15	6.97	−2.56	7.19	−3.18
VLTTGLPALISWIK	10.52	10.50	−0.22	10.54	−0.40
VPYEPPEVGSVYHHPLQLHV	15.10	15.13	0.20	15.18	−0.35
VV	15.42	15.39	−0.19	15.90	−3.34
VVRRCPHQRCSDSDGL	19.22	20.75	7.36	20.54	1.00
VVRRYPHHE	25.46	27.38	7.02	27.40	−0.06
WW	10.91	11.05	1.27	11.52	−4.28
YAEGDVHATSK	18.62	17.40	−7.00	16.70	4.01
YAEGDVHATSKPARR	21.94	21.38	−2.62	22.10	−3.36
YGGFL	9.70	9.75	0.54	9.62	1.32
YGGFM	9.58	9.53	−0.54	9.47	0.63
YKLVVVGAAGVGKSALT	15.21	14.22	−6.98	14.33	−0.78
YKLVVVGACGVKGSALT	15.04	14.33	−4.95	14.18	1.06
YKLVVVGANGVGKSALT	14.27	14.36	0.59	14.13	1.62
YKLVVVGARGVGKSALT	18.13	17.80	−1.84	17.99	−1.07
YKLVVVGAVGVGKSALT	14.35	15.06	4.71	14.20	5.73
YLSGADLNL	6.25	6.23	−0.32	6.87	−10.25
YMDGTMSQV	6.43	6.62	2.82	6.68	−0.87
YNYMCNSSGMGGMNRRP	13.71	14.29	4.03	13.05	8.69
YSPALNKMCCQLAKT	15.25	14.90	−2.32	14.15	5.03
YY	12.18	12.10	−0.64	12.32	−1.78

μ_{ef} values in $cm^2\ V^{-1}\ s^{-1}$ units; Offord's values are calculated according to the equation $\mu_{ef} = [2.44 + 581.85\ q/M^{2/3}] \cdot 10^{-5}$ (from Ref. [31]; % De$v = (\mu_{ef}(\text{Exp}) - \mu_{ef}(\text{Theo})/\mu_{ef}(\text{Exp}) \cdot 100$

hydrophobicity and for the mutual electrostatic interactions between ionogenic groups in highly charged peptides.

Realizing that the electrophoretic mobility of peptides cannot be accurately represented by a single analytical equation with two variables, we devised a computer model that fits the experimental mobilities of a basis set of peptides to a product of several functions representing the main variables that affect the motion of peptides in an electric field. An early version of this approach that contains a basis set of 64 peptides was published previously [32]. This work presents an improved version of the previous model. More peptides were added to the basis set to better populate the fields of the parameters, and a computer algorithm was added that matches the physical parameters of unknown peptides to their closest-neighbor in the peptide set for more accurate estimation of their mobilities. According to this model the electrophoretic mobility of a peptide, under the same conditions of pH and other experimental parameters, can be represented by a product of three functions:

$$\mu_{ef} = n(N)q(Q)w(W) \quad (7)$$

where $n(N)$ is a peptide length function; $q(Q)$ is a peptide charge function and $w(W)$ is a peptide average residue width function. This representation is conveniently amenable to computer manipulation where the three physical parameters (N, Q and W) are easily calculated and the product of the three functions converge the calculated mobilities to their experimental values. It is important to note that, in contrast to the two-parameter models) accurate values of charge are not essential in this approach as long as a consistent method is used for their determination.

Peptide average residue width function $w(W)$

The first function to be calculated is $w(W)$. For this purpose we selected (from Table 1) the subset of dipeptides with neutral residues (GG, AA, PG, VV, FG, FA, LL, FV, FL, MM, FF, YY, Trp–Trp). The residue width for each amino acid was assumed to be proportional the the relative mass of the residue, being 1 for glycine and 130 for tryptophane. Note that the one-letter symbols are used to represent amino acids throughout, except for tryptophan which is represented in the text by the three-letter symbol so as not to confuse it with our symbol for average residue width (W). The function $w(W)$ was obtained by plotting the experimental mobility for each of these dipeptides (from Table 3) versus its average residue width (from Table 1), and fitting the resulting curve to a polynomial. The resulting function is:

$$w(W) = -6.76 \cdot 10^{-6} W^3 + 1.90 \cdot 10^{-3} W^2 - 21.6 \cdot 10^{-2} W + 21.86 \quad (8)$$

This equation was used to calculate the average residue width contribution to the electrophoretic mobility.

3.1. Peptide length function n(N)

In our earlier version of this computer model we arrived at the peptide length function, $n(N)$, by plotting peptide mobility versus peptide length, as measured by the number of amino acid residues, for a homologous series of neutral peptides [Table 2 in Ref. [32]). A power series fit of the data yielded the expression: $n(N) = 30.90\ N^{-0.4892}$, ($R^2 = 0.99$), which suggests that peptide mobility is proportional to $N^{-0.4892}$. It is worth noting that this comes extremely close to the theoretical dependence ($N^{-0.5}$) derived from the classical polymer theory considering the peptide as a polymer of freely joined chain of amino acids of equal size [13]. Grossman et al. [13] used this model to predict peptide mobilities, but, in order to fit the model to the experimental data, they adjusted the power of N from -0.5 to -0.43. This correction was deemed necessary to account for differences in the sizes of the amino acid residues. In our multi-variable model, the dependence of mobility on peptide length was kept as derived from the experimental data with the power of N rounded to -0.5 to match the theoretical value. Differences in amino acid residue sizes were accounted for by the average residue width function.

Accordingly, the peptide length function was represented by:

$$n(N) = N^{-0.5} \quad (9)$$

3.2. Peptide charge function q(Q)

The peptide charge function $q(Q)$ was obtained by selecting (from Table 1) a representative set of

Table 4
List of peptides used to derive the $q(Q)$ function, together with their relevant parameters

Sequence	$\mu_{ef} \times 10^5$, Exp.	N	W	Q	$w(W)$	$w(W).n(N)$	$q(Q)$
DD	10.31	2	59.00	0.65	14.35	10.15	1.02
AA	18.77	2	15.00	0.83	19.04	13.46	1.39
AAAA	13.87	4	15.00	0.83	19.04	9.52	1.46
RQQ	24.00	3	81.33	1.83	13.23	7.64	3.14
RPPGF	18.36	5	54.80	1.83	14.63	6.54	2.81
TYSPALNRMFCQLAKT	14.77	16	57.88	2.83	14.42	3.61	4.10
PHRERCSDSDGL	19.33	12	56.67	3.64	14.50	4.19	4.62
KSSQYIKANSKFIGITE	17.05	17	55.35	3.82	14.59	3.54	4.82
VVRRCPHQRCSDSDGL	20.75	16	57.00	4.65	14.48	3.62	5.73
KKKK	33.03	4	72.00	4.83	13.64	6.82	4.84
KKKKK	33.03	5	72.00	5.83	13.64	6.10	5.41
KQIINMWQEVGKAMYAPPISGQIRRIHIGPGRAFYTTKN	17.78	39	58.03	7.82	14.42	2.31	7.70
HRSCRRRKRRSCRHR	30.27	15	79.33	11.83	13.31	3.44	8.81
DRVIEVVQGAYRAIRHIPRRIRQGLERRIHIGPGRAFYTTKN	20.83	42	61.74	12.72	14.19	2.19	9.52
RTHGQSHYRRRHCSRRRLHRIHRRQ	29.01	25	76.96	15.83	13.42	2.68	10.81

μ_{ef} values in $cm^2\ V^{-1}\ s^{-1}$.

peptides that cover the range of charges of the whole basis set and presented them in Table 4. $q(Q)$ was calculated for each peptide in Table 4 by dividing the experimental mobility by the product of $w(W)$ and $n(N)$. $q(Q)$ was, then, plotted versus Q and the resulting curve was subjected to a polynomial fit yielding the expression:

$$q(Q) = 2.9 \cdot 10^{-3} Q^3 - 9.59 \cdot 10^{-2} Q^2 + 1.45Q + 0.373 \quad (10)$$

3.3. The closest-neighbor algorithm and the calculation of μ_{ef}

The electrophoretic mobility of any peptide under the experimental conditions specified in the experimental section can now be calculated using Eq. (7), by substituting the values of W, N, and Q in their respective equations. However, better accuracy is achieved by utilizing a computer algorithm that matches an unknown peptide to its closest-neighbor in the basis set as a starting point in the calculation of its mobility. This is done by calculating N, W, and Q of the unknown peptide and submitting the values to a database that comprise all the peptides in Table 1. The database is queried to compare the submitted parameters with their counterparts for each peptide in the list and identify the peptide that results in a minimum in the sum of the differences ΔW, ΔN, and ΔQ as the closest-neighbor. Once the closest-neighbor is identified, then the electrophoretic mobility of the unknown peptide is calculated as follows:

$$\mu_{ef}(\text{pep}) = \mu_{ef}(\text{cn}) \cdot \frac{[w(W)n(N)q(Q)]_{\text{pep}}}{[w(W)n(N)q(Q)]_{\text{cn}}} \quad (11)$$

where μ_{ef} (cn) is the experimental mobility of the closest-neighbor.

To test the predictive ability of this approach, the electrophoretic mobility of each of the peptides in the basis set was calculated by considering that peptide as an unknown and applying the closest-neighbor algorithm to the basis set after removing the peptide of interest from it. The results for all peptides are recorded in Table 3 together with the experimental values and the values obtained according to the Offord model using the equation $\mu_{ef} = 10^{-5}\ (2.44 + 581.85\ Q/M^{2/3})\ cm^2\ V^{-1}\ s^{-1}$ [31]. Fig. 1A shows the correlation between experimental mobilities and those calculated according to the Offord model and Fig. 1B shows a corresponding correlation between experimental mobilities and those calculated according to our multi-variable model. Inspection of the entries in Table 3 and the degree of correlation in Fig. 1A and B reveals that our multi-variable computer model offers a significant improvement compared to the Offord model in fitting the experimental data for all categories of

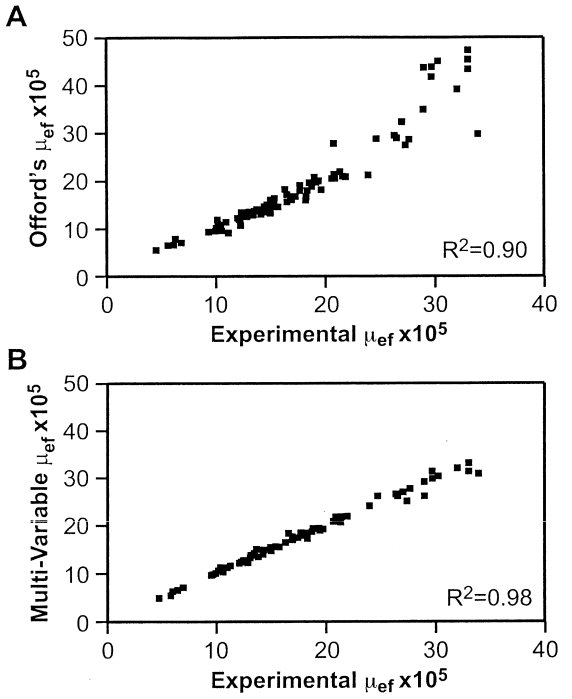

Fig. 1. (A) Correlation of predicted peptide mobilities based on the Offord's model versus experimental; (B) Correlation of predicted peptide mobilities based on the multi-variable model versus experimental. μ_{ef} units are in $cm^2\ V^{-1}\ s^{-1}$.

peptides. Table 5 highlights the results for some of the highly charged and hydrophobic peptides. Inspection of the entries in Table 5 demonstrates the excellent predictive power of the multi-variable model in contrast to the excessive deviations between experimental values and those predicted by Offord's model.

3.4. Peptide mapping by CZE: Experimental versus simulation

The potential of using the multi-variable computer model to simulate peptide maps of protein digests was first explored with glucagon, a small, single chain polypeptide of 29 amino acid residues. According to the supplier (Sigma), glucagon is highly purified and used as a control for proteolytic digestion, sequencing and amino acid analysis. The polypeptide was digested with endoproteinase Glu-C with specificity of cleavage at the C-terminal of aspartic (D) and glutamic (E) acid residues. Glucagon has three D residues and no E residues, and therefore, complete digestion is expected to yield four fragments. Complete digestion was ensured by monitoring the disappearance of the parent glucagon peak. The digest was directly injected into the CZE system without any further sample manipulation and the resulting electropherogram is shown in Fig. 2. The figure also shows the simulated electropherogram for comparative purposes. The simulation was conducted in two steps. First, the electrophoretic mobility of each fragment was calculated using the multi-variable model and subsequently converted to migration time using the same experimental parameters of column length and applied voltage as were used to generate the experimental electropherogram. After obtaining the migration time for each theoretical fragment the peaks were simulated with a Gaussian functions assuming that peak area (optical density) for each peak is proportional to $N-1$, where N is the number of amino acid residues in each fragment. This is based on the observation that the optical absorption of peptides at 200 nm (which is the wavelength used in this work to generate the experimental electropherograms) is largely attributed to the peptide bonds, and to a first degree of approximation, the contribution of amino acid residues to absorption at 200 nm can be ignored [33,34]. The percent error introduced as a result of this approximation is not expected to be of material value in this application. However, in future refinements of this approach the contribution of amino acid residues have to be accounted for. Gaussian peak variance that controls peak width was arbitrarily adjusted to approximately match the simulated peak width with the experimental counterpart. Also the peak area of the largest fragment was considered 100% and the peak areas of the other fragments were normalized accordingly. Inspection of Fig. 2 reveals that the experimental electropherogram of glucagon have a striking similarity to the simulated electropherogram. Each of the simulated peaks has a counterpart in the experimental electropherogram with excellent correspondence in migration times and reasonable correspondence in peak areas. However, the experimental electropherogram shows several minor peaks that have no counterparts in the simulated electropherogram. These are most likely due to nonspecific digestion or autolysis of the endoproteinase. The

Table 5
Comparison of the multi-variable model with Offord's for highly charged and hydrophobic peptides

Peptide	$\mu_{ef} \times 10^5$, Multi-var	$\mu_{ef} \times 10^5$		$\mu_{ef} \times 10^5$	
		Exp.	% Dev.	Offord's	% Dev.
IITLEDSSNLLGRNSF	11.58	11.33	−2.21	9.31	17.83
GSDCTTIHCNYM	12.39	12.41	0.16	10.77	13.25
LDDRNTFRRSVVVPYE	17.47	18.30	4.54	15.94	12.88
LAKTCPVRLWVDSTPP	14.25	15.13	5.78	13.29	12.12
RQQ	22.55	24.00	6.04	21.15	11.86
SSCMGGMNQRPILTIIT	10.20	10.66	4.32	9.59	10.00
YLSGADLNL	6.25	6.23	−0.32	6.87	−10.25
AAANLVPMVATV	6.10	6.15	0.78	6.85	−11.32
HMTEVVRHCPHHER	25.07	26.41	5.06	29.55	−11.89
VISNDVCAQV	6.03	5.83	−3.32	6.63	−13.72
HMTEVVRRYPHHER	27.11	26.42	−2.62	30.32	−14.77
DD	10.33	10.31	−0.19	12.05	−16.84
FIGITEAAANLVPMVATV	4.97	4.73	−5.13	5.66	−19.72
HG	27.05	27.04	−0.03	32.42	−19.90
RTHCQSHYRRRHCSR	26.39	28.96	8.86	34.99	−20.84
RK	30.41	32.00	4.98	39.03	−21.98
AAGIGILTV	6.53	6.50	−0.46	8.01	−23.19
KKK	32.54	33.03	1.49	43.35	−31.25
DRVIEVVQGAYRAIRHIPRRIRQGLERRIHIGPGRAFYTTKN	21.79	20.83	−4.61	27.85	−33.72
KKKK	33.53	33.03	−1.51	45.34	−37.27
CRHRRRHRRGC	29.75	29.68	−0.24	41.78	−40.78
KKKKK	33.18	33.03	−0.46	47.26	−43.09
CRHHRRRHRRGC	29.61	29.68	0.24	43.74	−47.38
HRSCRRRKRRSCRHR	31.41	30.27	−3.76	45.09	−48.97
RTHGQSHYRRRHCSRRRLHRIHRRQ	28.32	29.01	2.36	43.60	−50.31

μ_{ef} values in $cm^2\ V^{-1}\ s^{-1}$.

simulated electropherogram represents the ideal peptide map had the digestion been perfect, and any deviations from ideality expose shortcomings in the digestion process.

Next, horse cytochrome *c*, a more complex protein with 104 amino acid residues was tackled. The protein was digested with endoproteinase Lys-C with specificity of cleavage at the C-terminal of lysine residues, and the digest was directly injected into the CZE system without any further sample manipulation. The resulting electropherogram is presented in Fig. 3. The simulation of the theoretical peptide map was carried as described for glucagon. Table 6 gives a list of the theoretical fragments together with their predicted migration times, and the simulated electropherogram is shown in Fig. 3. Only 13 peaks were obtained for the 15 theoretical fragments. The pairs of fragments (8, 11), (2, 9), and (1, 10) co-migrated because of the closeness of their migration times (see Table 6). Note that peak variance and control peak width can be adjusted to produce narrower peaks and separate all the fragments, however this exercise produces an unrealistic electropherogram with column efficiency far exceeding the efficiency of the experimental system. The electropherogram produced will not be as useful for comparison purposes as that produced with column efficiency that matches the experimental. Inspection of Fig. 3 reveals some interesting similarities between the experimental and theoretical electropherograms, but it also exposes problems in matching experimental peaks with their theoretical counterparts. On the positive side, the two peaks at around 23 min agree in position and intensity within experimental error, and the experimental map does not show any peaks in the time range of 15 to 22 min in agreement with the theoretical prediction. On the negative side, the peaks clustered in the time range of 10 to 15 min cannot be unambiguously matched with corresponding peaks in the theoretical map.

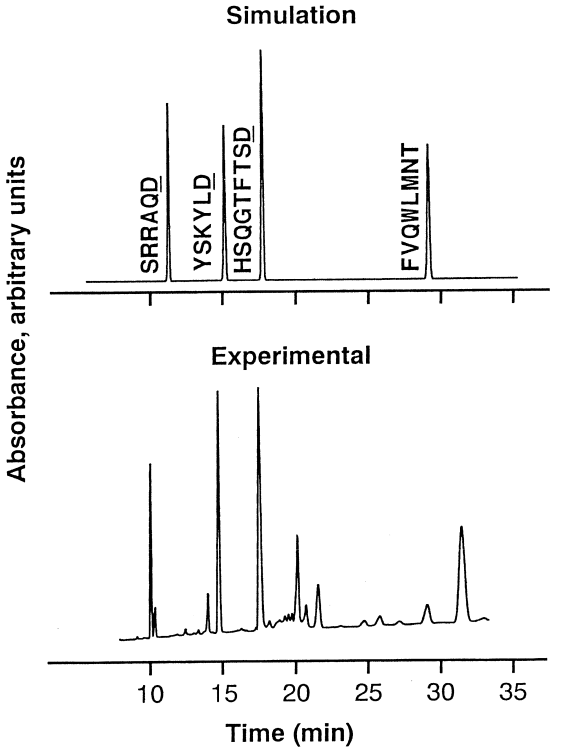

Fig. 2. Experimental and simulated electropherograms of the endoproteinase Glu-C digest of glucagon. Column: 10% *T* polyacrylamide-coated fused-silica [*T*=(g acrylamide+g *N,N′*-methylenebisacrylamide)/100 ml solution]; Column dimensions: 37 cm (effective length 30 cm)×50 μm I.D; Instrument, Beckman model P/ACE system 5500; voltage, 8 kV; current, 18 μA; injection, 5 s at 0.5 p.s.i.; buffer, 50 m*M* phosphoric acid adjusted to pH 2.5 with TEA; temperature, 22°C; detection, UV at 200 nm. Digestion: 100 μg of the protein + 10 μg of the proteanase in 100 μl of 25 m*M* Tris–HCl, pH 8.5 for 18 h at 37°C.

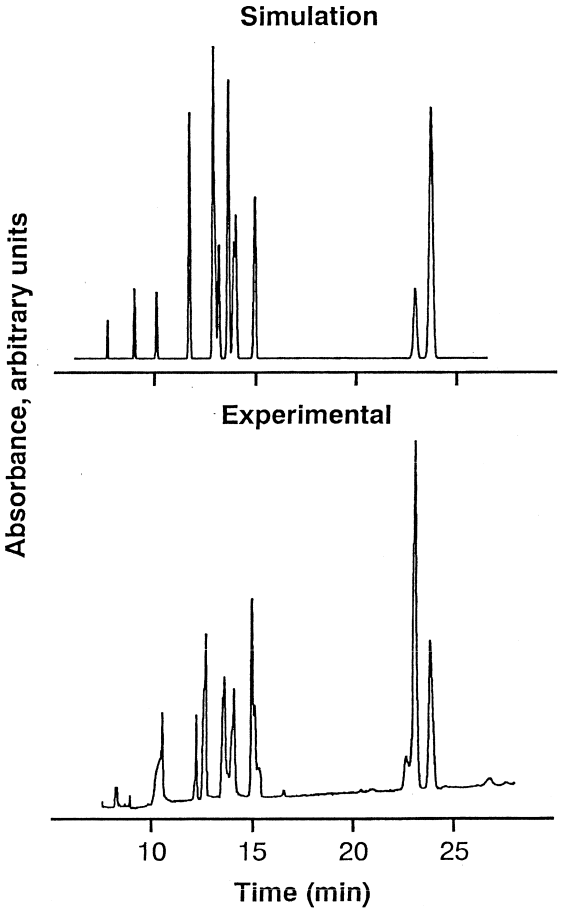

Fig. 3. Experimental and simulated electropherograms of the endorproteinase Lys-C digest of horse cytochrome *c*. Experimental conditions: As in Fig. 2.

4. Conclusion

A multi-variable computer model that predicts the mobilities of peptides at pH 2.5, and simulates CZE peptide maps at this pH is developed. The long range, ultimate goal of this study is to construct a database of CZE theoretical peptide maps of known proteins, each theoretically fragmented with different proteinases, and use it as an aid in the identification of unknown proteins by submitting the experimental peptide maps to the database and interrogate it for the closest match in terms of the correspondence of the major peaks in their migration times and relative areas. At this stage of the development there are many problems that need to be addressed before this dream gets closer to reality. These problems are mostly in the experimental part. Peptide maps of protein digests are encumbered with the presence of peaks that are not theoretically expected had the digestion process been perfect. The science and art of enzymatic digestion of proteins has come a long way since the discovery and purification of enzymes of high specificity for particular types of peptide bonds, but there are still shortcomings that result in less than perfect digestion. In practice the digestion process has many artifacts and potential problems that encumber the experimental peptide maps. These include, but not limited to, autolysis of the digestion proteinase, incomplete digestion and nonspecific

Table 6
Theoretical product of endoproteinase Lys-C digest of horse cytochrome *c*

Peptide number	Peptide sequence	Predicted migration time (min)
1	GDVEK	13.0
2	GK	8.3
3	K	–
4	IFVQK	13.5
5	CAQCHTVEK	12.0
6	GGK	9.6
7	HK	7.0
8	TGPNLHGLFGRK	11.9
9	TGQAPGFTYTDANK	23.0
10	NK	8.6
11	GITWK	13.2
12	EETLMEYLENPK	23.1
13	K	–
14	YIPGTK	13.7
15	MIFAGIK	14.6
16	K	–
17	K	–
18	TEREDLIAYLK	15.1
19	K	–
20	ATNE	22.2

digestion. All these processes, if present, will result in experimental maps that deviate from the ideal theoretical maps. Extraneous peaks in the peptide map could also be caused by impurities in the protein or the proteinase.

A more realistic goal of this endeavor is to use it as an aid in studies of single known proteins, such as the identification of the location of disulfide linkages and protein modifications. Theoretical simulation yields the ideal peptide map had the digestion process been perfect. Thus the fidelity of the digestion process can be monitored by comparing the theoretical maps with the experimental under different conditions.

Acknowledgements

This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract No. NO1-CO-56000.

By acceptance of this article, the publisher or recipient acknowledges the right of the US Government to retain a nonexclusive, royalty-free license and to any copyright covering the article. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

References

[1] T. Van de Goor, A. Apffel, J. Chakel, W. Hancock, in: J.P. Landers (Ed.), Handbook of Capillary Electrophoresis, 2nd ed, CRC Press, Boca Raton, FL, 1997, pp. 213–258.
[2] M. Dong, R.P. Oda, M.A. Strausbauch, P.J. Wettstein, J.P. Landers, L.J. Miller, Electrophoresis 18 (1997) 1767.
[3] N. Bihoreau, C. Ramon, R. Vincentelli, J.-P. Levillain, D.V. Troalen, J. Capillary Electrophor. 2 (1995) 197.
[4] M. Castagnola, L. Cassiano, R. Rabino, D.V. Rossetti, F.A. Bassi, J. Chromatogr. 572 (1991) 51.
[5] R.S. Rush, P.L. Derby, T.W. Strickland, M.F. Rohde, Anal. Chem. 65 (1993) 1834.
[6] S.H. Kang, X. Gong, E.S. Yeung, Anal. Chem. 72 (2000) 3014.
[7] M. Stromquist, J. Chromatogr. A 667 (1994) 304.
[8] M.A. Winkler, S. Kundu, T.E. Robey, W.G. Robey, J. Chromatogr. A 744 (1996) 177.
[9] P. Cao, M. Moini, Rapid Commun. Mass Spectrom. 12 (1998) 864.
[10] A. Cifuentes, H. Poppe, J. Chromatogr. A 680 (1994) 321.
[11] P.D. Grossman, D.S. Soane, Anal. Chem. 62 (1990) 1592.

[12] P.D. Grossman, R.J. Wilson, G. Petrie, H.J. Lauer, Anal. Biochem. 173 (1988) 265.
[13] P.D. Grossman, J.C. Colburn, H.H. Lauer, Anal. Biochem. 179 (1989) 28.
[14] E.C. Rickard, M.M. Strohl, R.G. Nielsen, Anal. Biochem. 197 (1991) 197.
[15] B.J. Compton, J. Chromatogr. 559 (1991) 357.
[16] M.A. Survay, D.M. Goodall, S.A.C. Wren, R.C. Rowe, J. Chromatogr. A 741 (1996) 99.
[17] H.-J. Gaus, A.G. Beck-Sickinger, E. Bayer, Anal. Chem. 65 (1993) 1399.
[18] V.J. Hilser Jr., G.D. Worosila, S.E. Rudnick, J. Chromatogr. 630 (1993) 329.
[19] N. Chen, L. Wang, Y.K. Zhang, Chromatographia 37 (1993) 429.
[20] M.A. Survay, D.M. Goodall, S.A.C. Wren, R.C. Rowe, J. Chromatogr. 636 (1993) 81.
[21] S.K. Basak, M.R. Ladisch, Anal. Biochem. 226 (1995) 51.
[22] R.F. Cross, J. Cao, J. Chromatogr. A 786 (1997) 171.
[23] M. Castagnola, L. Cassiano, I. Messana, G. Nocca, R. Rabino, D.V. Rossetti, B. Giardina, J. Chromatogr. B 656 (1994) 87.
[24] N.J. Adamson, E.C. Reynolds, J. Chromatogr. B 699 (1997) 133.
[25] I. Messana, D.V. Rossetti, L. Cassiano, F. Misiti, B. Giardina, M. Castagnola, J. Chromatogr. B 699 (1997) 149.
[26] V. Kasicka, Electrophoresis 20 (1999) 3084.
[27] A. Cifuentas, H. Poppe, Electrophoresis 18 (1997) 2362.
[28] R.J. Wieme, in: E. Heftmann (Ed.), Chromatography – A Laboratory Handbook of Chromatographic and Electrophoretic Methods, 3rd ed, Van Nostrand Reinhold, New York, 1975, Chapter 10.
[29] R.E. Offord, Nature 211 (1966) 591.
[30] B. Skoog, A. Wichman, Trends Anal. Chem. 5 (1986) 82.
[31] G.M. Janini, C.J. Metral, H.J. Issaq, G.M. Muschik, J. Chromatogr. A 848 (1999) 417.
[32] C.J. Metral, G.M. Janini, G.M. Muschik, H.J. Issaq, J. High Resolut. Chromatogr. 22 (1999) 373.
[33] R.K. Scopes, Anal. Biochem. 59 (1974) 277.
[34] M. Herold, G.A. Ross, R. Grimm, D.N. Heiger, in: K.D. Altria (Ed.), Capillary Electrophoresis Guidebook, Principles, Operation, and Application, Methods in Molecular Biology, Vol. 52, Humana Press, Totowa, NJ, 1996, pp. 285–308.
[35] G.M. Janini, K.C. Chan, G.M. Muschik, H.J. Issaq, J. Chromatogr. B 657 (1994) 419.