

Combining Peptide Modeling and Capillary Electrophoresis–Mass Spectrometry for Characterization of Enzymes Cleavage Patterns: Recombinant versus Natural Bovine Pepsin A

Carolina Simó,[†] Ramón González,[†] Coral Barbas,[‡] and Alejandro Cifuentes^{*†}

Institute of Industrial Fermentations (CSIC), Juan de la Cierva 3, 28006 Madrid, Spain, and Department of Chemistry, Faculty of Experimental and Health Sciences, University San Pablo-CEU, Urbanización Montepríncipe, 28668 Boadilla del Monte, Madrid, Spain

Nowadays there is an increasing number of recombinant enzymes made available to industry. Before replacing the use of natural enzymes with their cognate recombinant counterparts, one important issue to address is their actual equivalence. For a given recombinant proteolytic enzyme, its equivalence can be investigated by comparing its cleavage specificity with that obtained from the natural enzyme. This is mostly done by analyzing the fragments (i.e., peptidic map) attained after enzymatic digestion of a given protein used as substrate. The peptidic maps obtained are typically characterized using separation techniques together with MS and MS/MS systems. However, these procedures are known to be difficult and labor-intensive. In this work, the combined use of a theoretical model that relates electrophoretic behavior of peptides to their sequence together with capillary electrophoresis–mass spectrometry (CE–MS) is proposed to characterize in a very fast and simple way the cleavage specificity of new recombinant enzymes. Namely, the effectiveness of this procedure is demonstrated by analyzing in few minutes the fragments obtained from a protein hydrolyzed using recombinant and natural pepsin A. The usefulness of this strategy is further corroborated by CE–MS/MS. The proposed procedure is applicable in many other proteomic studies involving CE–MS of peptides.

For many industrial applications, recombinant enzymes constitute a convenient alternative to enzymes purified from natural sources in such fields as biomedicine, biotechnology, or the agro-food industry. Depending on the specific application and industrial requirements, these advantages could reside among others on purity (including reduced risk of transmission of human or zoonotic pathogens or the absence of unwanted contaminants), availability (when natural sources are in short supply), or ethical or religious issues (when species of origin or slaughtering procedure is a concern).¹ The possibility to specifically modify

enzyme properties through site-directed mutagenesis constitutes an additional advantage of recombinant enzymes.¹ However, to replace natural enzymes, their recombinant counterparts must be tested for equivalence in key parameters, including specificity, stability, or kinetic properties. The output of this comparison will help in decision making concerning the use of new recombinant enzymes. Moreover, characterization of sequence specificity for new recombinant proteases is especially important since in many cases, it will define their usefulness.

Among the different analytical strategies proposed to characterize proteins and enzymes (such as, e.g., SDS–PAGE, Western blot, N-terminal sequencing, HPLC, MS, etc), capillary electrophoresis–electrospray-mass spectrometry (CE–ESI-MS) has been demonstrated to be a powerful technique in this field of research.^{2–12} The combination of the high efficiencies provided by CE with the compound identification capability of mass spectrometry makes available a powerful analytical tool for the investigation of these biological compounds.^{12,13} It is noteworthy that this type of tool is in great demand by the emerging field of proteomics.

Comparative analysis by CE–MS of the peptidic maps obtained from a given protein after its hydrolysis with both a recombinant and its corresponding wild-type enzyme could provide interesting information on their similar (or different) behaviors. However, this procedure presents a difficult challenge to the analytical chemist due to the wide range of peptides that can come out with

- (2) Wu, J. T.; Qian, M. G.; Liu, L.; Lubman, D. M. *Anal. Chem.* **1996**, *68*, 3388–3396.
- (3) Yin, H. F.; Templin, C. K.; McManigill, D. *J. Chromatogr., A* **1996**, *744*, 45–54.
- (4) Winkler, M. A.; Kundu, S.; Robey, T. E.; Robey, W. G. *J. Chromatogr., A* **1996**, *744*, 177–185.
- (5) Zhang, H. Y.; Caprioli, R. M. *J. Mass Spectrom.* **1996**, *31*, 1039–1046.
- (6) Smith, R. D.; Udseth, J. H.; Wahl, J. H.; Goodlett, D. R.; Hofstadler, S. A. *Methods Enzymol.* **1996**, *271*, 448–486.
- (7) Valaskovic, G. A.; Kelleher, N. L.; McLafferty, W. *Science* **1996**, *273*, 1199–1202.
- (8) Banks, J. F. *Electrophoresis* **1997**, *18*, 2255–2266.
- (9) Broke, A.; Nicholson, G.; Bayer, E. *Electrophoresis* **2001**, *22*, 1215–1266.
- (10) Tomer, K. B. *Chem. Rev.* **2001**, *101*, 297–328.
- (11) Simó, C.; Elvira, C.; González, N.; San Román, J.; Barbas, C.; Cifuentes, A. *Electrophoresis* **2004**, *25*, 2056–2064.
- (12) Hernández-Borges, J.; Neusüss, C.; Cifuentes, A.; Pelzing, M. *Electrophoresis* **2004**, *25*, 2257–2281.
- (13) Schmitt-Kopplin, P.; Frommberger, M. *Electrophoresis* **2003**, *24*, 3837–3867.

* Corresponding author. Fax: 34-91-5644853. E-mail: acifuentes@ifi.csic.es.

[†] CSIC.

[‡] University San Pablo-CEU.

(1) Kirk, O.; Borchert, T. V.; Fuglsang, C. C. *Curr. Opin. Biotechnol.* **2002**, *13*, 345–351.

different pIs, hydrophobicities, and molecular masses.¹⁴ As a result, the adequate CE separation of peptides in terms of peak shape and resolution is usually labor-intensive. Moreover, the buffer has to bring about an adequate CE separation of peptides without disturbing the ESI-MS ionization and signal. In general, resolution and sensitivity of CE-MS become the overriding factors, often determining the success or failure of a particular electrophoretic method, buffer system, or application. Once adequate CE and ESI-MS conditions are achieved, a tentative characterization of peptides is first performed on the basis of the molecular weights obtained by CE-MS. This tentative assignment of peptides is further corroborated by obtaining their sequence using CE-MS/MS. Clearly, this process is tedious and labor-intensive. Moreover, an additional limitation that cannot be discarded is that the digestion of a known protein by a new recombinant enzyme can give rise to unexpected or unknown peaks that make even more complex the determination and assignment of the peptides obtained by CE-MS. These peaks can be associated, e.g., with enzyme autolysis products,¹⁵ including the products released upon autoactivation of the enzyme or contaminant proteins or peptides from the enzyme preparation.

A solution that can help to overcome these limitations is the development of new peptide modeling¹⁶⁻²³ that is capable of predicting the CE migration time and peak shape of peptides in different pHs and buffers. The use of these models should provide useful information on the nature of peptides in a fast and simple way.

The objective of this work is to study the possibilities of a peptide modeling developed at our lab called "system for prediction of peptide migration" (SPPM) together with CE-MS to characterize in a fast way the cleavage activity of a recombinant enzyme versus its corresponding natural type. Namely, the SPPM CE-MS procedure is employed to characterize the activity of a new recombinant bovine pepsin A obtained by expressing the bovine cDNA in the yeast *Saccharomyces cerevisiae*, as compared with its corresponding wild-type. In addition, in this work, new data are presented about the cleavage specificity of natural bovine pepsin A itself, an enzyme poorly characterized, as compared to porcine pepsin, despite its relevance for the food industry.^{24,25}

EXPERIMENTAL SECTION

Chemicals and Samples. All chemicals were of analytical reagent grade and were used as received. Ammonia (30%) from Panreac (Barcelona, Spain), methanol (HPLC grade) from Schar-

lau (Barcelona, Spain), and formic acid from Riedel-de Hën (Seelze, Germany) were used for the CE running buffers and sheath liquids. Water was deionized by using a Milli-Q system (Millipore, Bedford, MA).

The protein used as substrate, cytochrome *c* from bovine heart (Cyt_c), was obtained from Sigma (St. Louis, MO); commercial pepsin A from adult bovine rennet (IDF Standard 157) was obtained from Chr. Hansen (Denmark), and recombinant pepsinogen A was obtained as indicated below.

Recombinant Bovine Pepsinogen Obtention. Recombinant bovine pepsinogen was obtained by concentration of culture media after 72 h of induction on YPGal medium from *S. cerevisiae* strain BY4741 transformed with plasmid pBP05 as described elsewhere.²⁶

Pepsin Activity Determination. The overall proteolytic activities of natural pepsin or recombinant pepsinogen were quantified against bovine hemoglobin at pH 2 as described for pepsinogen.²⁷ Activation of bovine pepsinogen to pepsin under these reaction conditions is almost immediate,²⁸ allowing for direct comparison of the results obtained for both enzymes. These data were used for standardization of the proteolytic activities used in the hydrolysis experiments.

Hydrolysis Experiments. Cytochrome *c* was dissolved in a solution containing 10 mM hydrochloric acid and 100 mM sodium chloride (pH 2), at a concentration of 10 mg/mL. Next, commercial pepsin and recombinant pepsin were dissolved in the same solution at concentrations of 0.2 and 6 mg/mL, respectively. The concentration of the enzymes was chosen to give equivalent overall proteolytic activities, as described above. The digestion of the protein was carried out mixing 30 μ L of each cytochrome *c* and pepsin solution. All hydrolysis experiments were carried out in duplicate. Digest was allowed to proceed for 16 h (or the time indicated in each case, i.e., 24, 48, or 90 h) at 37 °C. The enzymatic digestion was stopped by increasing the temperature to 85 °C for 15 min. The protein digest was stored at -4 °C.

Capillary Electrophoresis-Electrospray-Mass Spectrometry (CE-ESI-MS). CE-ESI-MS analyses were carried out in a PACE/5500 CE apparatus (Beckman, Fullerton, CA) coupled to the MS detector using commercial coaxial sheath-flow interface (see below). Bare, fused-silica capillaries with 50- μ m i.d. were purchased from Composite Metal Services (Worcester, England). The capillary detection and total length was 90 cm. Between analyses, the capillary was rinsed with water for 3 min and with running buffer for 2 min. Running voltage was +20 kV, and injections were made in triplicate at the anodic end using N₂ pressure of 0.5 psi (1 psi = 6894.76 Pa) for 45 s. For reproducibility studies, injections were repeated five times using a cytochrome *c* sample hydrolyzed with recombinant pepsin. The instrument was controlled by a PC running the System Gold Software from Beckman.

Before their first use, the fused-silica capillaries were washed with 0.1 M sodium hydroxide for 20 min and Milli-Q water for 20 min (all rinses were done using N₂ pressure at 20 psi).

- (14) Wahl, J. H.; Gale, D. C.; Smith, R. D. *J. Chromatogr.* **1994**, *659*, 217-222.
 (15) Wang, C.; Oleschuk, R.; Ouchen, F.; Li, J.; Thibault, P.; Harrison, D. J. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1377-1383.
 (16) Cifuentes, A.; Poppe, H. *J. Chromatogr., A* **1994**, *680*, 321-340.
 (17) Cifuentes, A.; Poppe, H. *Electrophoresis* **1995**, *16*, 516-524.
 (18) Cifuentes, A.; Poppe, H. *Electrophoresis* **1997**, *18*, 2362-2376.
 (19) Janini, G. M.; Metral, C. J.; Issaq, H. J.; Muschik, G. M. *J. Chromatogr., A* **1999**, *848*, 417-433.
 (20) Metral, C. J.; Janini, G. M.; Muschik, G. M. *J. High Resolut. Chromatogr.* **1999**, *22*, 373-378.
 (21) Janini, G. M.; Metral, C. J.; Issaq, H. J. *J. Chromatogr., A* **2001**, *924*, 291-306.
 (22) Castagnola, M.; Rossetti, D. V.; Corda, M.; Pellegrini, M.; Misiti, F.; Olianias, A.; Gardina, B.; Messana, I. *Electrophoresis* **1998**, *19*, 1728-1732.
 (23) Castagnola, M.; Rossetti, D. V.; Corda, M.; Pellegrini, M.; Misiti, F.; Olianias, A.; Gardina, B.; Messana, I. *Electrophoresis* **1998**, *19*, 2273-2277.
 (24) Lang H. M.; Kassell, B. *Biochemistry* **1971**, *10*, 2296-2301.
 (25) Dunn, B. M.; Hung, S. H. *Biochim. Biophys. Acta* **2000**, *1477*, 231-240.

- (26) Muñoz, R.; García, J. L.; Carrascosa, A. V.; González, R. *Appl. Environ. Microbiol.* **2004**, *70*, 2588-2595.
 (27) Kasell, B.; Meitner, P. A. *Methods Enzymol.* **1970**, *19*, 337-347.
 (28) Harboe, M.; Andersen, P. M.; Foltmann, B.; Kay, J.; Kassell, B. *J. Biol. Chem.* **1974**, *249*, 4487-4494.

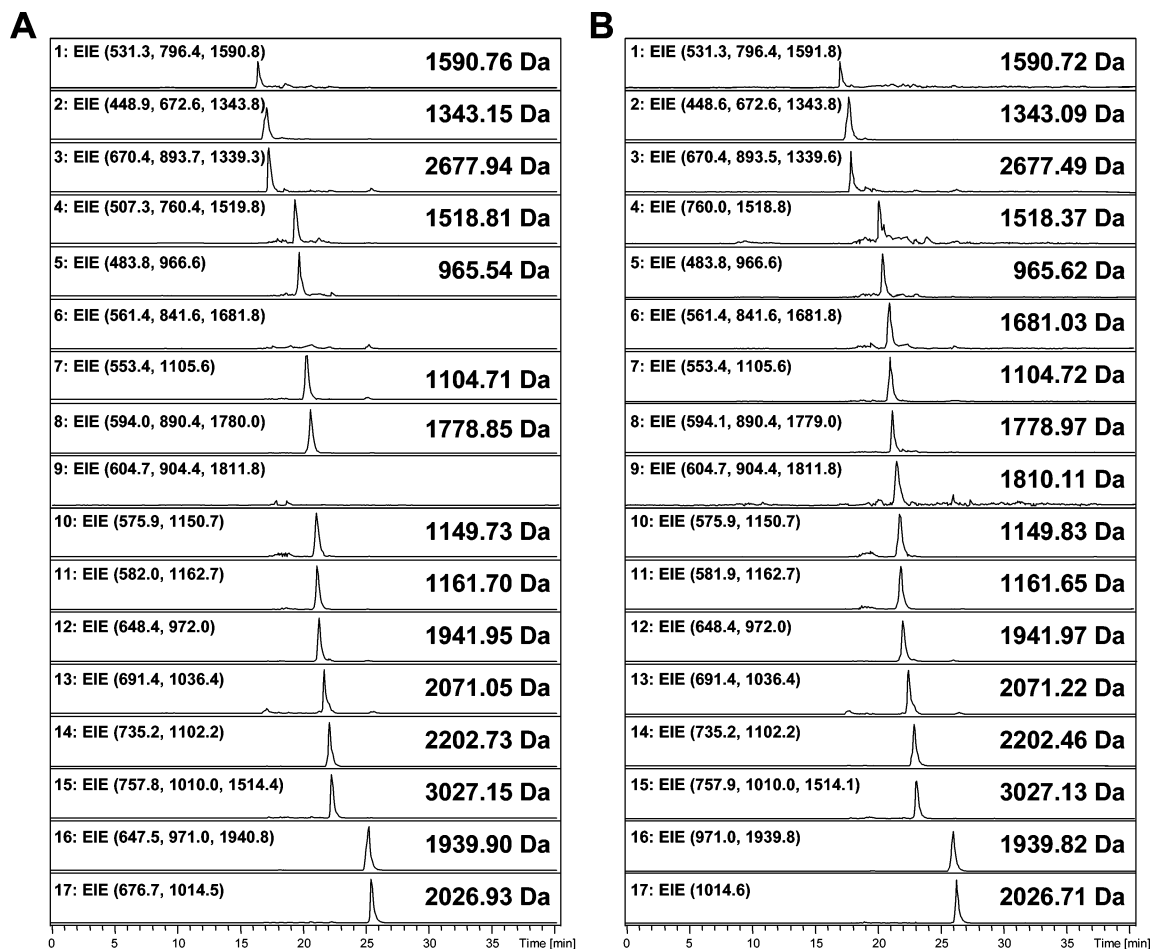


Figure 1. Extracted ion electropherograms (EIEs) of the peptide mixture obtained from hydrolysis of CytC with (A) natural pepsin and (B) recombinant pepsin. Running buffer: 0.9 M formic acid at pH 2; injection at 0.5 psi for 45 s. Fused-silica capillary with 50- μ m i.d. and 90-cm total length. Running voltage: +20 kV. Temperature: 25 °C. MS conditions: sheath liquid, methanol–water (50:50, v/v) containing 0.05% (v/v) formic acid; sheath flow, 4 μ L/min; nebulizer gas, 4 psi, 8 L/min at a temperature of 120 °C. MS scan range, m/z 400–2000. The numbers between parenthesis correspond to the major ions used to obtain the extracted ion electropherograms.

MS experiments were performed on a Bruker Daltonik Esquire 2000 ion-trap mass spectrometer (Bruker Daltonik, Bremen, Germany) equipped with an orthogonal electrospray interface (model 61607A from Agilent Technologies, Palo Alto, CA). Electrical contact at the electrospray needle tip was established via a sheath liquid delivered by a 74900-00-05 Cole Palmer syringe pump (Vernon Hills, IL). The mass spectrometer was operated in the positive ion mode. The spectrometer was scanned at 400–2000 m/z range at 13 000 u/s during separation and detection. ESI parameters were nebulizer pressure, 4 psi; dry gas flow equal to 8 L/h; dry gas temperature, 120 °C; and a sheath liquid made of methanol–water (50:50, v/v) with 0.05% formic acid at 4 μ L/min flow-rate. MS/MS experiments were performed with an isolation width of 4 m/z and a fragmentation amplitude of 1 V. The instrument was controlled by a PC running the Esquire 5.0 software from Bruker Daltonics.

Computer Programs. The SPPM used to simulate the peptide migration in CE has been described elsewhere.¹⁶ This program includes a computer program for calculating peptide pK_a values, an equation that relates the peptide sequence to its electrophoretic mobility, and a coupled program for prediction of electropherograms.¹⁶

The Biotoools program from Bruker Daltonics was used to carry out searches of probable peptide sequences within CytC that

match the molecular mass determined by CE–MS after hydrolysis with recombinant or natural pepsin.

RESULTS AND DISCUSSION

Optimizing CE–MS Conditions To Analyze Peptidic Fragments Obtained Using Recombinant and Natural Pepsin.

Optimization of the CE–ESI–MS parameters was carried out testing different CE separation buffers and ESI–MS conditions injecting the group of peptides obtained after enzymatic hydrolysis of CytC with natural pepsin. Namely, several buffers consisting of acetic acid, formic acid, and ammonium hydroxide at different pHs were tested. The best results in terms of sensitivity and resolution were obtained with 0.9 M formic acid adjusted with ammonium hydroxide to pH 2. On the other hand, after testing different sheath liquid compositions, the best MS signals were obtained using a methanol–water (50:50, v/v) dissolution with 0.05% formic acid. Figure 1A shows the extracted electropherograms obtained under the selected CE–MS conditions for the peptidic fragments from CytC produced using natural pepsin. As can be seen in Figure 1A, the CE–MS conditions permit the simultaneous analysis of 15 peptides in a single run with high efficiency and sensitivity in analysis times lower than 30 min. Also from Figure 1, it can be deduced that the CE–MS conditions allow the unambiguous identification of the molecular masses of these

Table 1. CE–MS Experimental Mass of the Peptidic Fragments after Hydrolysis of CytC Using Natural Pepsin (M_{nat}) and Recombinant Pepsin (M_{rec}) Together with Probable Sequences from CytC Matching the M_{rec} Values

no.	M_{nat}^a	M_{rec}^b	range	M , theor	probable sequence ^c		
1	1590.76	1590.72	6–19	1590.92	EK	GKKIFVQKCAQCHT	VE
			22–36	1590.83	VE	KGGKHKGTGPNLHGLF	GR
2	1343.15	1343.09	39–51	1342.41	GR	KTGQAPGFSYTD	NK
			83–94	1343.53	IF	AGIKKKGEREDL	IA
3	2677.64	2677.49	22–47	2678.02	VE	KGGKHKGTGPNLHGLFGRKTGQAPGFS	YT
4	1518.61	1518.37	68–80	1518.82	EY	LENPKKYIPGTKM	IF
			69–81	1518.82	YL	ENPKKYIPGTKMI	FA
5	965.54	965.62	58–65	966.07	GI	TWGEETLM	EY
			97–104	966.09	IA	YLKATNE	
6		1681.03	67–80	1681.00	ME	YLENPKKYIPGTKM	IF
			60–73	1680.88	TW	GEETLMEYLENPKK	YI
7	1104.71	1104.72	37–47	1105.21	LF	GRKTGQAPGFS	YT
8	1778.85	1778.97	68–82	1779.16	EY	LENPKKYIPGTKMIF	AG
			70–85	1778.17	LE	NPKKYIPGTKMIFAGI	KK
9		1810.11	65–79	1811.10	TL	MEYLENPKKYIPGTK	MI
			66–80	1811.10	LM	EYLENPKKYIPGTKM	IF
10	1149.73	1149.83	95–104	1150.33	DL	IAYLKKATNE	
			1–10	1162.34		Acetyl-GDVEKGGKIF	VQ
11	1161.70	1161.65	2–11	1162.38	G	DVEKGGKIFV	QK
			53–62	1161.27	AN	KNKGITWGEE	TL
			71–80	1162.45	EN	PKKYIPGTKM	IF
			76–86	1162.45	YI	PGTKMIFAGIK	KK
			29–47	1941.16	KT	GPNLHGLFGRKTGQAPGFS	YT
12	1941.95	1941.97	67–82	1942.33	ME	YLENPKKYIPGTKMIF	AG
			65–80	1942.31	TL	MEYLENPKKYIPGTKM	IF
			48–65	2071.27	FS	YTDANKNKGITWGEEETLM	EY
13	2071.05	2071.22	66–82	2071.44	LM	EYLENPKKYIPGTKMIF	AG
			67–84	2070.46	ME	YLENPKKYIPGTKMIFAGI	IK
			14–34	2202.48	QK	CAQCHTVEKGGKHKGTGPNLHG	LF
14	2202.73	2202.46	65–82	2202.64	TL	MEYLENPKKYIPGTKMIF	AG
			2–28	3027.53	G	DVEKGGKIFVQKCAQCHTVEKGGKHKHT	GP
15	3027.15	3027.13	37–64	3027.27	LF	GRKTGQAPGFSYTDANKNKGITWGEEETL	ME
			12–29	1940.22	FV	QKCAQCHTVEKGGKHKHTG	PN
16	1939.90	1939.82	33–50	1939.10	NL	HGLFGRKTGQAPGFSYTD	AN
			48–64	1940.08	FS	YTDANKNKGITWGEEETL	ME
			47–64	2027.15	GF	SYTDANKNKGITWGEEETL	ME
17	2026.93	2026.71	52–68	2026.28	DA	NKNKGITWGEEETLMEYL	EN
			54–70	2027.22	NK	NKGITWGEEETLMEYLEN	PK
			62–78	2026.32	GE	ETLMEYLENPKKYIPGT	KM

^a Molecular mass of the peptidic fragments obtained by CE–MS using natural pepsin. ^b Molecular mass of the peptidic fragments obtained by CE–MS using recombinant pepsin. ^c Theoretical CytC sequence determined using Biotoools, M_{rec} values and considering an error <1 Da.

fragments. As a comparison, Figure 1B shows the extracted electropherograms obtained under optimum conditions for the peptidic fragments from CytC produced using recombinant pepsin. At first sight, comparison among the peptidic fragments obtained for both enzymes (Figure 1A and B) shows that the use of the recombinant pepsin gives rise to 17 fragments, whereas the natural enzyme gives rise to 15 fragments, with peptides 6 and 9 only produced by the recombinant pepsin (see Figure 1B and Table 1). Moreover, the other 15 fragments seem to be similar for the two enzymes (recombinant and natural), as can be deduced from the same molecular weight found for these compounds. Reproducibility of CE–MS analysis was confirmed by calculating the relative standard deviations (%RSD) for migration times and peak areas of the 17 peptides of Figure 1B after five consecutive injections. Thus, as can be seen in Table 2, the %RSD values ranged from 1.1 to 1.9% for migration times and from 3.2 to 4.9% for peak areas, corroborating the reproducibility of our procedure.

A First Approach To Characterize the Peptidic Fragments Using Their Mass Values and a Sequence Search Program.

Table 2. Analysis Time and Peak Area Reproducibility of the CE–MS Analysis of the 17 Peptides Obtained Using Recombinant Pepsin

peptide no.	t (min)	%RSD _{t} ^a	A ($\times 10^5$ arb. units)	%RSD _{A} ^b
1	17.0	1.1	0.4	3.2
2	17.8	1.2	7.8	4.0
3	17.9	1.2	1.8	3.5
4	20.3	1.4	0.7	3.3
5	20.6	1.4	1.6	3.5
6	21.2	1.5	1.7	3.6
7	21.2	1.5	3.0	3.7
8	21.4	1.5	5.0	3.8
9	21.8	1.6	1.4	3.6
10	22.0	1.6	5.4	4.1
11	22.1	1.7	5.2	4.0
12	22.3	1.6	17.5	4.4
13	22.7	1.6	4.0	3.9
14	23.1	1.7	17.7	4.8
15	23.2	1.7	5.9	4.3
16	26.4	1.9	19.7	4.9
17	26.5	1.9	3.0	4.1

^a Relative standard deviation for analysis time ($n = 5$) of the 17 peptidic fragments. ^b Relative standard deviation for peak areas ($n = 5$).

Since the sequencing of all peptides from Figure 1 is a difficult and labor-intensive task, employing common MS computer tools (e.g., Biotoools from Bruker, MassLynx from Micromass) can provide interesting information about the nature of the peptides detected in Figure 1 and, as result, on the similar (or different) cleavage specificity of the recombinant and natural pepsin studied. Thus, Table 1 shows the experimental mass values obtained by CE-MS of the peptidic fragments after hydrolysis of CytC with natural pepsin (M_{nat}) and recombinant pepsin (M_{rec}). Also in Table 1 are given some probable sequences from CytC found using the Biotoools program and the M_{rec} values as input considering an error <1 Da (i.e., a mass error lower than 0.1% is then assumed for practically all the fragments). As can be seen in Table 1, this approach could help to characterize some peptides (see, e.g., peptides 7 and 10) for which only a possible sequence comes out. However, for the rest of the peptides, a minimum of two probable sequences was found, increasing to four or five possible sequences for some cases (see peptides 11 and 17 in Table 1), which would increase the experimental time necessary to correctly identify these peptides. Logically, this limitation could be overcome by reducing the error of the experimental mass determined using, for example, more accurate instruments, such as TOF or FT-MS. However, in this case, the very high price of these instruments is an important limitation to take into account. An alternative, and cheaper, strategy can be the use of CE peptide modeling for these proteomic studies.

Characterization of the Peptidic Fragments Using Their Mass Values, a Sequence Search Program, and a CE Model for Peptides (SPPM). The usefulness of different separation modeling programs in CE has already been demonstrated.^{16,29–32} Among these, SPPM has been demonstrated to provide accurate simulated CE electropherogram of peptide separations,^{33,34} requiring as input data values on pK_a and electrophoretic mobilities of the different buffer ions,^{35–38} capillary dimensions, injection time and voltage (or injected plug length), separation voltage, electroosmotic mobility, and sample concentration. These parameters can be directly taken from an experiment if a given separation has to be simulated or freely chosen if different running conditions want to be tested. The program also needs as input the pK_a and electrophoretic mobilities of the different sample ions. This is the main complication, since pK_a and mobility values are available for only very few peptides in the literature.^{39,40,41} To calculate the pK_a and mobility values of each peptide, we have written a computer program,¹⁶ which allows us to know these values on the basis of

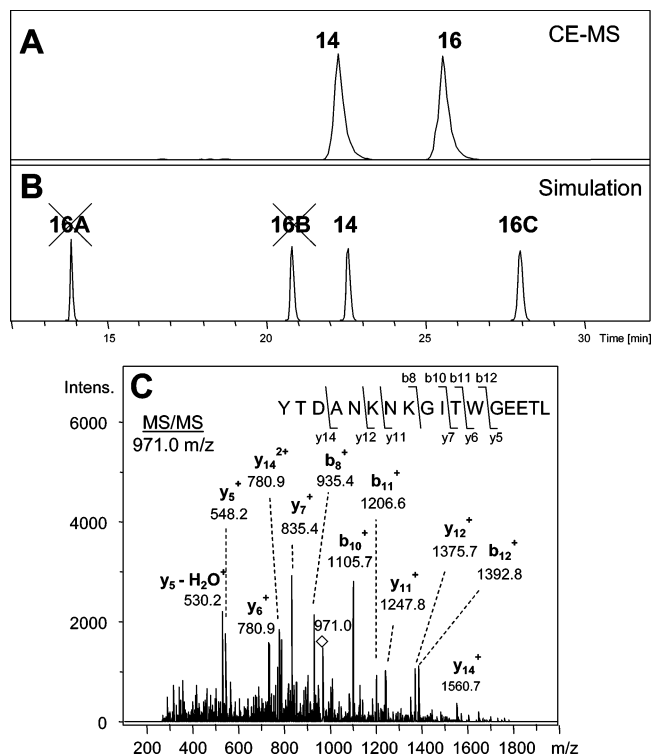


Figure 2. (A) CE-MS extracted ion electropherogram of peptides 14 and 16; (B) simulated electropherogram of peptides 14, 16A, 16B, and 16C; and (C) MS/MS spectra of peptide 16. Separation conditions as in Figure 1.

the structure of the peptide with just the amino acid sequence of the peptide as the input. Thus far, it works only for linear, nonbranched, noncyclic peptides, whereas the number of acid-base active groups that can be handled is as large as 14.

The SPPM can help to elucidate the most probable peptidic sequence (and therefore, to characterize the pepsin cleavage) among the different possibilities given in Table 1. For instance, as indicated in Table 1, the peptide number 16 with a M_{rec} value equal to 1939.8 is obtained using both the natural and recombinant pepsin, and it has three compatible sequences according to the Biotoools search program (i.e., QKCAQCHTVEKGGKHKKTG, HGLFGRKTGQAPGFSYTD, or YTDANKNKGITWGEETL). The SPPM can be an important help to easily elucidate the most probable sequence of this peptide. Thus, Figure 2A shows the experimental CE-MS electropherogram in which for clarity only peptides 14 (as reference) and 16 have been included. Figure 2B shows the simulated electropherogram obtained using the SPPM under the same experimental conditions, assuming that sequence of peptide 16 can be QKCAQCHTVEKGGKHKKTG (peptide 16A), HGLFGRKTGQAPGFSYTD (peptide 16B), or YTDANKNKGITWGEETL (peptide 16C). Thus, according to the simulated electropherogram of Figure 2B, the sequences 16A and 16B do not seem to be correct, since they migrate faster than peptide 14. Sequence 16C is the most probable one, as can be deduced from the good agreement in the migration times and order obtained for this sequence between simulation (Figure 2B) and real separation (Figure 2A). It can also be observed by comparing peptides 16C and 16 in Figure 2 that an exact match between the times obtained in the experimental and simulated electropherograms is not achieved in this case. It should be taken into account that during

- (29) Jaros, M.; Vceláková, K.; Zusková, I.; Gas, B. *Electrophoresis* **2002**, *23*, 2667–2677.
- (30) Mosher, R. A.; Saville, D. A.; Thormann, W. *The Dynamics of Electrophoresis*; VCH: Weinheim; 1992.
- (31) Beckers, J. L. *J. Chromatogr., A* **1997**, *764*, 111–126.
- (32) Gebauer, P.; Bocek, P. *Anal. Chem.* **1997**, *69*, 1557–1563.
- (33) Simó, C.; López Soto-Yárritu, P.; Cifuentes, A. *Electrophoresis* **2002**, *23*, 2288–2295.
- (34) Simó, C.; Cifuentes, A. *Electrophoresis* **2003**, *24*, 834–842.
- (35) Poppe, H. *Anal. Chem.* **1992**, *64*, 1908–1919.
- (36) Pospichal, J.; Gebauer, P.; Bocek, P. *Chem. Rev.* **1989**, *89*, 419–430.
- (37) Hirokawa, T.; Nishino, M.; Kiso, Y. *J. Chromatogr.* **1982**, *252*, 49–65.
- (38) Pospichal, J.; Deml, M.; Bocek, P. *J. Chromatogr.* **1987**, *390*, 17–26.
- (39) Hirokawa, T.; Kiso, Y.; Gas, B.; Zuskova, I.; Vacik, J. *J. Chromatogr.* **1993**, *628*, 283–308.
- (40) Wronski, M. *J. Chromatogr.* **1993**, *657*, 165–173.
- (41) Jalali-Heravi, M.; Shen, Y.; Hassanisadi, M.; Khaledi, M. G. *Electrophoresis* **2005**, *26*, 1874–1885.

Table 3. Tentative Sequence Assignment of the Peptidic Fragments Given in Table 1 after Using the Peptide Modeling

no.	M_{nat}^a	M_{rec}^b	range	MW (theor)	sequence given by SPPM		
1	1590.76	1590.72	22–36	1590.83	VE	KGGKHKTPGNLHGLF	GR
2	1343.15	1343.09	83–94	1343.53	IF	AGIKKKGEREDL	IA
3	2677.64	2677.49	22–47	2678.02	VE	KGGKHKTPGNLHGLFGRKTGQAPGFS	YT
4	1518.61	1518.37	68–80	1518.82	EY	LENPKKYIPGTKM	IF
			69–81	1518.82	YL	ENPKKYIPGTKMI	FA
5	965.54	965.62	97–104	966.09	IA	YLKATNE	
6		1681.03	67–80	1681.00	ME	YLENPKKYIPGTKM	IF
7	1104.71	1104.72	37–47	1105.21	LF	GRKTGQAPGFS	YT
8	1778.85	1778.97	68–82	1779.16	EY	LENPKKYIPGTKMIF	AG
			65–79	1811.10	TL	MEYLENPKKYIPGTK	MI
9		1810.11	66–80	1811.10	LM	EYLENPKKYIPGTKM	IF
10	1149.73	1149.83	95–104	1150.33	DL	IAYLKKATNE	
11	1161.70	1161.65	1–10	1162.34		Acetyl-GDVEKGGKIF ????	VQ
12	1941.95	1941.97	67–82	1942.33	ME	YLENPKKYIPGTKMIF	AG
13	2071.05	2071.22	66–82	2071.44	LM	EYLENPKKYIPGTKMIF	AG
14	2202.73	2202.46	65–82	2202.64	TL	MEYLENPKKYIPGTKMIF	AG
15	3027.15	3027.13	37–64	3027.27	LF	GRKTGQAPGFSYTDANKNKGITWGEETL	ME
16	1939.90	1939.82	48–64	1940.08	FS	YTDANKNKGITWGEETL	ME
17	2026.93	2026.71	47–64	2027.15	GF	SYTDANKNKGITWGEETL	ME

^a Molecular mass of the peptidic fragments obtained by CE–MS using natural pepsin. ^b Molecular mass of the peptidic fragments obtained by CE–MS using recombinant pepsin.

the development of any simulation program, several approximations have to be included. For instance, the adsorption phenomenon was not incorporated into the SPPM program, and the ionic strength effect, although included in the program for obtaining the electropherograms, was not included in the program for pK_a calculations. Obviously, these approximations can affect in some cases the agreement between real electropherograms and simulations. In addition, some other effects, such as the formation of moving ionic boundaries inside the capillary due to migration of liquid sheath counterions into the separation capillary, can also play a role in some cases.⁴²

It is interesting to note that the achievement of the simulation of Figure 2B needed a few seconds in a PC. Moreover, this tentative assignment was further corroborated by carrying out a CE–MS/MS analysis of this peptide. Figure 2C shows the mass spectrum obtained after CE–MS/MS for peptide 16. From the y_i and b_i fragments shown in Figure 2C, it could be corroborated that the sequence of this peptide is YTDANKNKGITWGEETL (peptide 16C).

The same procedure was used to tentatively assign the other sequences. For instance, peptide 6 is only obtained using the recombinant pepsin. According to its M_{rec} value (1681.0), this fragment can be due to two possible CytC fragments, namely, the 67–80 fragment having as its sequence YLENPKKYIPGTKM (peptide 6A) or the 60–73 fragment having as its sequence GEETLMEYLENPKK. As above, Figure 3A shows the experimental CE–MS electropherogram, and Figure 3B, its corresponding simulation obtained in seconds, considering these two possible sequences under identical separation conditions. The most probable sequence seems to be the peptide 6A, as can be deduced from the good agreement between the simulation (Figure 3B) and the real separation (Figure 3A) observed for the analysis time and migration order for peptide 6A. As above, this tentative assignment was further corroborated by CE–MS/MS (see MS/MS spectrum of the peptide given in Figure 3C), thus demonstrating the good possibilities of our approach.

(42) Foret, F.; Thompson, T. J.; Vouros, P.; Karger, B. L.; Gebauer, P.; Bocek, P. *Anal. Chem.* **1994**, *64*, 4450–4456.

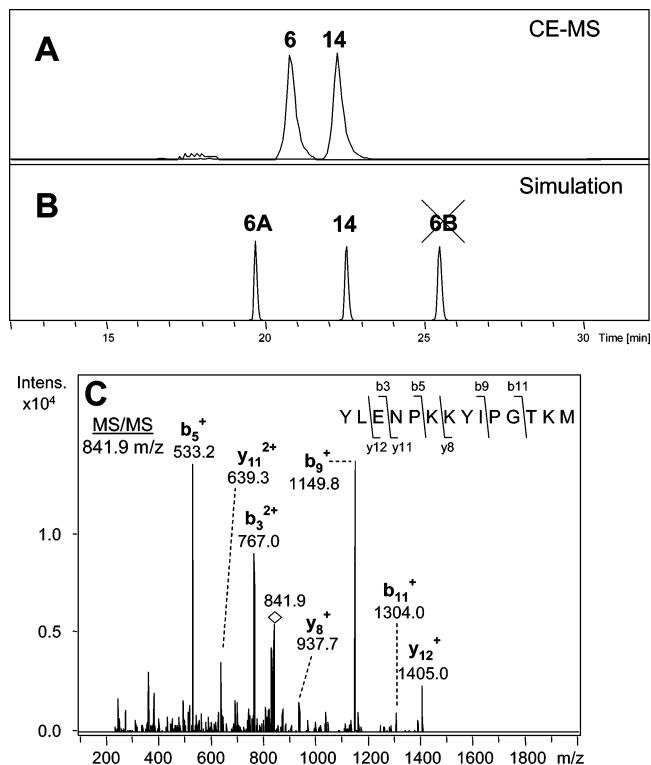


Figure 3. (A) CE–MS–extracted ion electropherogram of peptides 6 and 14; (B) simulated electropherogram of peptides 6A, 6B and 14; and (C) MS/MS spectra of peptide 6. Separation conditions as in Figure 1.

A similar procedure was followed for all the peptides from Table 1, giving rise in a few minutes to the tentative sequences shown in Table 3 that are next discussed. As can be seen in Table 3, in only four cases could the SPPM modeling not elucidate the most probable sequence, namely, for peptides 3, 4, 9, and 11. However, in one of these four cases (peptide 3) the SPPM provided interesting information, since it allowed discarding the only probable peptidic sequence given in Table 1 for this peptide 3, which was later corroborated by MS/MS (see below). For peptides 4 and 9, the SPPM provided more than one possible

Table 4. Peptide Sequences Corroborated by MS/MS Fragmentation

no.	M_{nat}^a	M_{rec}^b	range	MW (theor)	sequence after MS/MS		
1	1590.76	1590.72	22–36	1590.83	VE	KGGKHKTGPNLHGLF	GR
2	1343.15	1343.09	83–94	1343.53	IF	AGIKKKGREDL	IA
3	2677.64	2677.49	22–47	2678.02	VE	KGGKHKTGPNLHGLFGRKTGQAPGFS	YT
4	1518.61	1518.37	68–80	1518.82	EY	LENPKKYIPGTKM	IF
5	965.54	965.62	97–104	966.09	IA	YLKKTATNE	
6		1681.03	67–80	1681.00	ME	YLENPKKYIPGTKM	IF
7	1104.71	1104.72	37–47	1105.21	LF	GRKTGQAPGFS	YT
8	1778.85	1778.97	68–82	1779.16	EY	LENPKKYIPGTKMIF	AG
9		1810.11	66–80	1811.10	LM	EYLENPKKYIPGTKM	IF
10	1149.73	1149.83	95–104	1150.33	DL	IAYLKKATNE	
11	1161.70	1161.65	1–10	1162.34		Acetyl-GDVEKGGKIF	VQ
12	1941.95	1941.97	67–82	1942.33	ME	YLENPKKYIPGTKMIF	AG
13	2071.05	2071.22	66–82	2071.44	LM	EYLENPKKYIPGTKMIF	AG
14	2202.73	2202.46	65–82	2202.64	TL	MEYLENPKKYIPGTKMIF	AG
15	3027.15	3027.13	37–64	3027.27	LF	GRKTGQAPGFSYTDANKNKGITWGEETL	ME
16	1939.90	1939.82	48–64	1940.08	FS	YTDANKNKGITWGEETL	ME
17	2026.93	2026.71	47–64	2027.15	GF	SYTDANKNKGITWGEETL	ME

^a Molecular mass of the peptidic fragments obtained by CE–MS using natural pepsin. ^b Molecular mass of the peptidic fragments obtained by CE–MS using recombinant pepsin.

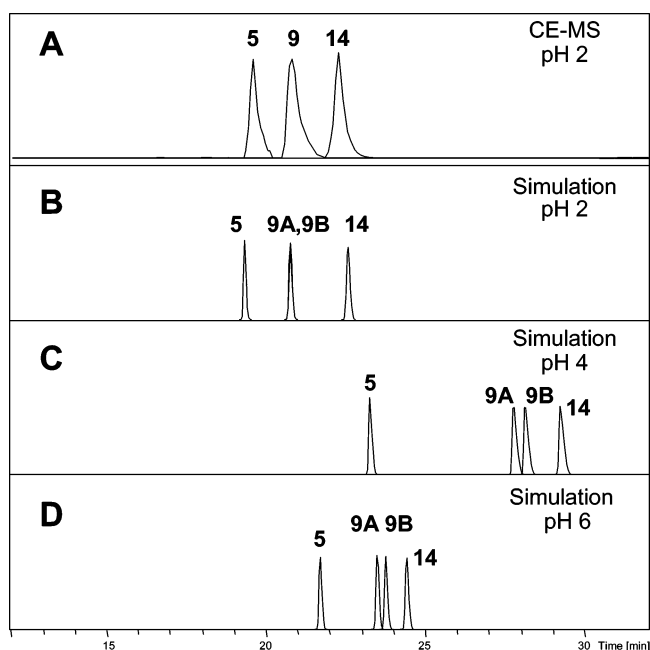


Figure 4. (A) CE–MS extracted ion electropherogram of peptides 5, 9, and 14 at pH 2 (all conditions as in Figure 1). Simulated electropherogram of peptides 5, 9A, 9B, and 14 using an electrolyte consisting of (B) 0.9 M formic acid at pH 2, (C) 50 mM formic acid at pH 4, or (D) 40 mM formic acid at pH 6.

sequence, making the tentative assignment not possible, whereas for peptide 11, the simulation could not be completed due to one of the probable peptides' being acetylated, and this form is not included at the moment in the SPPM program. However, the SPPM also made it possible in this case to discard the other four possible sequences provided by the Biotoools search program for peptide 11 (i.e., DVEKGGKIFV, KNKGITWGEE, PKKYIPGTKM, or PGTKMIFAGIK; see Table 1) since these sequences were not compatible with the CE migration experimentally observed for this peptide.

As an example of the negative results mentioned above, the SPPM simulation obtained for peptide 9 is given in Figure 4. Thus, the two possible sequences for this peptide, MEYLENPKKYIPGTK (peptide 9A) or EYLENPKKYIPGTKM (peptide 9B), comigrate in CE at the three separation pHs tested (namely, 2, 4,

and 6), which is in good agreement with the simulation (see Figure 4A–D), thus making its tentative identification by the SPPM model not possible. Clearly, this is due to the very similar nature of these two sequences (their only difference is that Met is at the amine or carboxylic end), which makes their separation by CE a very difficult task.

Table 4 shows the final peptidic sequences corroborated after CE–MS/MS experiments. As can be seen by comparing the sequences given in Tables 3 and 4, there was good agreement in all the cases, corroborating the usefulness of our approach and confirming the great possibilities of the combined use of peptide modeling and CE–MS to characterize in a fast and simple way the cleavage specificity of new enzymes.

Comparative Study of the Cleavage from Recombinant and Commercial Pepsin. Figure 5 summarizes all the cleavage sites detected on CytC using native and recombinant bovine pepsin A (as indicated above, peptides 6 and 9 were only obtained with the latter enzyme). A clear preference for Phe at position P1 (carboxyl terminal side) is observed, and the peptides corresponding to cleavage on the carboxyl terminal side of all Phe residues in the primary sequence of CytC have been found. This is similar to what has been described for porcine pepsin.⁴³ However, preference for Leu at the same position seems to be less marked using bovine pepsin than for the porcine enzyme. This is in contrast to previous studies of the cleavage pattern of bovine pepsin A, using β -casein as substrate, where no activity was detected at positions defined by Phe at P1.⁴⁴ In that study, Leu, Met, and Thr were shown to be preferred at position P1. This is also in contrast with the present study, because no peptides corresponding to cleavage on the carboxyl terminal side of Thr residues and only 2 out of 6 Leu positions have been cut in our conditions. The main agreement with previously published results on natural bovine pepsin A is on Met specificity because both Met positions were cut. These differences might be due to the use of different substrates (CytC or β -casein) and digestion conditions (pH 3 for the previous work and pH 2 for the present

(43) Neurath, H. In *Proteolytic Enzymes, A Practical Approach*; Beynon R. J., Bond, J. S., Eds.; IRL Press: Oxford, 1990; pp 1–240.

(44) Guillou, H.; Miranda, G.; Pelissier, J. P. *Int. J. Pept. Protein Res.* **1991**, *37*, 494–501.

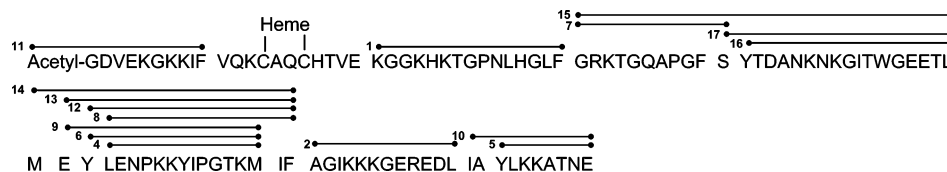


Figure 5. Cleavage pattern of the recombinant pepsin on CytC. Peptide identification number is shown to the left of each peptide.

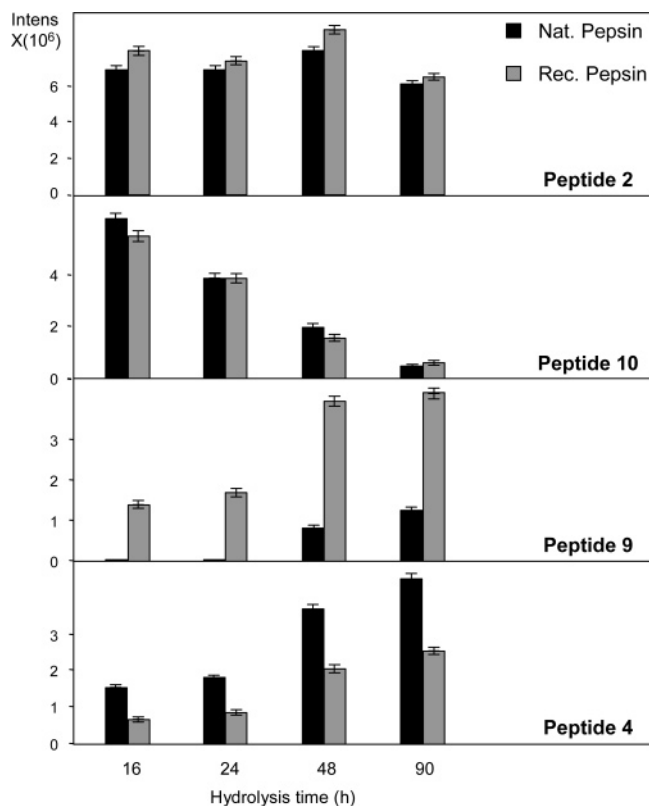


Figure 6. Hydrolysis kinetics for peptides 2, 4, 9, and 10 obtained using natural and recombinant pepsin. Confidence intervals were calculated for each sample injected in triplicate at the 95% confidence level.

study). A similar important effect of the pH on specificity has also been described for porcine pepsin.⁴⁵

CE-MS of the hydrolysis experiments also allows for a comparison between natural and recombinant bovine pepsin A. Thus, the accumulation kinetics of several of the peptides detected was similar for both enzymes, including amino- and carboxyl-terminal peptides, 5, 10 and 11, as well as the internal peptides 2, 7, 14, and 15. Results for peptides 2 and 10 are shown in Figure 6 for illustration. This is consistent with the fact that similar overall proteolytic activities were used. In contrast, changes in concentration of other peptides with digestion time followed different patterns for both enzymes. This is the case for peptides 6, 9, and 12, which accumulate to higher levels in the reaction catalyzed

(45) Keil, B. *Specificity of Proteolysis*; Springer-Verlag: Berlin-Heidelberg-New York, 1992.

(46) Meitner, P. A.; Kassell, B. *Biochem. J.* **1971**, *121*, 249–256.

by the recombinant enzyme as compared with the natural one. As an example, results for peptide 9 are shown in Figure 6. Conversely, peptide 4, which can be considered as an end-product because it is the smallest peptide of a series consisting of 9, 6, and 4 (see Figure 5), is accumulated at a higher rate by the natural enzyme (results for peptide 4 are shown in Figure 6). This is also the case for peptides 1 and 8. The trivial explanation, that is, that one of the enzyme preparations was more active, is not supported by the similar accumulation kinetics observed for other peptides. One alternative explanation would be differences in specificity for both enzymes. These differences would consist of a lower efficiency of the recombinant enzymes to remove one or two amino acid residues from the N-terminal end of peptides 9 and 6, or a different specificity for the primary digestion of CytC, leading to increased amounts of short peptides in the case of the natural enzyme. This different behavior might indicate differences in posttranslational modifications of the enzymes. However, it is probably not due to differential glycosylation, since the enzymes do not show consensus sequences for N or O glycosylation, and the natural enzyme has been characterized as nonglycosylated.⁴⁴ Differential phosphorylation or other posttranslational modifications might account for the different specificities detected.⁴⁶ A deeper characterization of the posttranslational differences between natural and recombinant bovine pepsin will allow understanding of the basis of their different specificities. This work is now underway at our laboratory.

CONCLUSIONS

In this work, we have demonstrated that the combined use of peptide modeling in CE and CE-MS is a powerful tool for characterizing the cleavage pattern of new enzymes in a fast and simple way. Peptidic maps obtained from a recombinant and natural pepsin A could be identified in a few minutes by using this procedure. Moreover, new information was provided about the cleavage pattern obtained for these enzymes. This procedure can be very useful in many other proteomic approaches involving CE-MS of peptides.

ACKNOWLEDGMENT

C.S. expresses thanks to the Comunidad de Madrid for a fellowship.

Received for review June 17, 2005. Accepted September 16, 2005.

AC051067D