Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 1055 (2004) 87–92
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Mass spectrometric characterization of peptides derived by peptic cleavage of bovine β-casein

Christian E.H. Schmelzer[a], Regina Schöps[b], Renate Ulbrich-Hofmann[b], Reinhard H.H. Neubert[a], Klaus Raith[a,*]

[a] *Institute of Pharmaceutics and Biopharmaceutics, Martin Luther University Halle-Wittenberg, Wolfgang-Langenbeck-Strasse 4, 06120 Halle (Saale), Germany*
[b] *Institute of Biotechnology, Martin Luther University Halle-Wittenberg, Kurt-Mothes-Strasse 3, 06120 Halle (Saale), Germany*

Received 16 July 2004; accepted 1 September 2004

Abstract

This study investigated the digestion of the milk protein β-casein with pepsin under gastro-analogous conditions. Peptide sequences were identified using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry with post-source decay as well as liquid chromatography–tandem mass spectrometry by means of database searching. The new software tool, Mascot Distiller, improved the identification rate remarkably. In the case of small peptides, such as di- and tri-peptides, which are promising candidates for intestinal absorption and possible biological effects, identification was possible only after spectrum simulation and manual matching. A list of 41 identified peptides having 2–36 amino acids is given, and unexpected cleavage sites for pepsin are reported. Sequence coverage was 75%.

Keywords: Casein; Pepsin; Mass spectrometry; MALDI; LC/MS; Tandem MS; Enzymatic hydrolysis; Peptide sequencing; Cleavage sites

1. Introduction

Proteins, together with fats and carbohydrates, are the main nutrients in human food. Nutritional research deals with the effects of those nutrients and the products of their digestion on the functions of the organism. Particularly proteins show an amazing structural variety, and among their breakdown products, the peptides, which are mainly formed by hydrolytic reactions by digestive enzymes, are numerous compounds with a variety of biological effects [1–8]. A prerequisite for this kind of investigation is the analysis of peptide patterns resulting from certain conditions of enzymatic protein digestion.

Among the proteases responsible for the digestion of dietary proteins in humans, the endopeptidases pepsin, trypsin and chymotrypsin are of utmost importance. Whereas pepsin is an aspartate protease occuring in the stomach with an optimum pH 2–7, both trypsin and chymotrypsin are serine proteases from the intestinum with an optimum pH 7–9. In contrast to most other enzymes, these endopeptidases are not specific to a certain substrate, but are capable of cleaving almost any protein inside the peptide chain.

Literature has shown that often such peptides were investigated that were derived from hydrolysis with site-specific pancreatic proteases such as trypsin [7,9]. However, before proteins reach the intestinum, they undergo a much less specific digestion with pepsin, an important digestive enzyme of the gastric juice. It is known from the literature that pepsin preferentially cleaves C-terminal to aromatic amino acids and leucine [10].

In this study, β-casein from bovine milk was chosen as a model protein to characterize the peptide spectrum obtained by peptic digestion under gastro-analogous conditions. Caseins are by far the most abundant milk proteins (82%) and

* Corresponding author. Tel.: +49 345 5525215; fax: +49 345 5527292.
E-mail address: raith@pharmazie.uni-halle.de (K. Raith).

0021-9673/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2004.09.003

are divided into subclasses, β-casein being the second most abundant among them.

Very small peptides, such as di- and tri-peptides, show a high intestinal absorption rate because they are subject to the transmembrane transport by means of the well-characterized PepT1 carrier [11–13]. Therefore, they are promising candidates for possible in vivo effects. However, larger peptides, that are formed initially during digestion, may also cross the membrane barrier under certain circumstances, such as inflammatory diseases. Taking these factors into consideration, there is a need to characterize the full spectrum of peptides derived from a dietary protein under the conditions of a gastro-analogous digestion.

Mass spectrometry, namely electrospray tandem mass spectrometry (ESI-MS-MS) and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) with post-source decay (PSD) in conjunction with database search, has become the most powerful analytical technique for analyzing peptides [14–18]. Factors determining the success of peptide sequence identification by database searching include the quality of the peak lists obtained from mass spectrometric raw data and the optimal specification of the search conditions (e.g., mass accuracy, taxonomy, and enzyme). The interface between data set and database is usually semi-automatic peak detection with the help of which the peak list is prepared from the MS–MS or PSD data set. This peak detection is a critical step and is often the bottleneck in the identification procedure. The main problems of the peak detection process in fragment ion spectra are distinguishing low intensity signal peaks from noise and possible selection of the wrong peak or all of the peaks of an isotopic cluster. New additional software (Mascot Distiller, Matrix Science, London, UK) addresses these and other problems and detects peaks by attempting to fit an average isotope distribution of peptides to experimental data. Comparable algorithms have been described by Berndt et al. [19] and Gras et al. [20].

To our knowledge, this is the first systematic approach to elucidate the composition of a peptic β-casein digest using modern analytical methods, such as ESI-MS-MS and MALDI-TOF-MS with PSD.

2. Experimental

2.1. Materials

β-Casein, derived from bovine milk, and, pepsin derived from porcine stomach mucosa (471 U/mg), both were obtained from Sigma-Aldrich (Taufkirchen, Germany).

Acetonitrile of HPLC grade was purchased from Merck (Darmstadt, Germany), and water was doubly distilled. Formic acid and trifluoroacetic acid (TFA) of analytical grade were purchased from Merck and Fluka (Buchs, Switzerland), respectively. The matrices α-cyano-4-hydroxycinnamic acid (CHCA), 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid), and 2,5-dihydroxybenzoic acid (DHB) were purchased from Sigma-Aldrich. All substances were used without further purification.

2.2. Digestion of β-casein with pepsin

Bovine β-casein (0.2 g) was dissolved in 8 mL of water (25 mg/mL), adjusted to pH 2 with 1 M HCl and digested with pepsin (1 mg/mL) at 37 °C for 3 h under shaking. The whole pepsin digest was immediately fractionated at 4 °C by ultrafiltration through membrane disc filters (Omega-Series, PALL, Ann Arbor, MI, USA) with molecular mass cut-offs of 5000 and 3000. The resultant peptide fractions (M_r < 3000, 3000–5000, and >5000) were lyophilized before further use. Sample mixtures were stored at −30 °C.

2.3. MALDI-TOF-MS

MALDI experiments were carried out using a delayed extraction time-of-flight (TOF) mass spectrometer Voyager-DE PRO (Applied Biosystems, Weiterstadt, Germany) equipped with a pulsed nitrogen laser (λ = 337 nm, 3 ns pulse width, 20 Hz repetition rate). Three matrix systems were used: a solution of 10 mg/mL of CHCA in acetonitrile–0.1% TFA (1:1, v/v); 20 mg/mL of sinapinic acid in acetonitrile–0.1% TFA (1:1, v/v) and 10 mg/mL of DHB in acetonitrile–0.1% TFA (1:4, v/v). The peptide solution (100 μg/mL in acetonitrile–0.1% TFA) was mixed with the matrix solution 1:1 (v/v). 0.5 μL of the resulting mixture was applied to each spot of the sample plate. The dried droplet method was favored due to the enhanced reproducibility compared with the thin layer method. The samples were dried in a gentle stream of air at ambient temperature (24 °C). Measurements were performed operating in the positive-ion reflector mode at an acceleration voltage of 20 kV, grid voltage set to 76%, 0.002% guide wire voltage and an extraction delay of between 100 and 350 ns depending on experimental conditions. A low mass gate was set to 600 u to prevent detector saturation from matrix peaks. Survey spectra of the peptide mixtures were obtained by accumulating data from 500 to 2000 laser shots.

2.4. MALDI-PSD/collision-induced dissociation (CID) analysis

Post-source decay experiments were carried out by performing 10–15 steps of the reflectron voltage, and for each individual step the voltage was reduced to ∼75% of the previous step. Using the timed ion selector (TIS), different precursor ions were selected from the peptide mixtures and subjected to fragmentation. The TIS resolution was specified with R = 80 (full width at half maximum). After acquisition all spectra were stitched together by the data system. Each segment spectrum obtained was the mean of at least 1000 laser shots. For CID experiments [21], the collision cell was filled with

air so that the pressure in the source chamber reached ~800 × 10^{-6} Pa.

2.5. HPLC-ESI-MS and tandem MS

The system used for HPLC-ESI-MS consisted of a Spectra System P 4000 pump, equipped with an auto sampler AS 3000 and a controller SN 4000 (Thermo Electron, San José, CA, USA). The MS and tandem MS experiments were performed on an ion trap mass spectrometer Finnigan LCQ Classic (Thermo Electron) with electrospray interface operated in the positive ion mode. Ten microliters of each sample solution (500 μg/mL in 0.1% formic acid) were loaded onto a Nucleosil 120-5 C_{18} column (125 mm × 2 mm i.d., Macherey–Nagel, Düren, Germany) and peptides eluted using a linear gradient: 5–50% acetonitrile in 0.1% formic acid over 60 min. The column was maintained at 30 °C and the flow rate was 0.2 mL/min. Electrospray conditions were as follows: capillary temperature, 200 °C; sheath gas flow rate, 1.2 L/min; ESI voltage, +4.5 kV; capillary voltage, 46 V; and tube lens offset, 40 V.

Full scans were performed between *m*/*z* 50 and 2000. Peptides of interest were selected manually for further tandem MS experiments using CID. The mass isolation window for

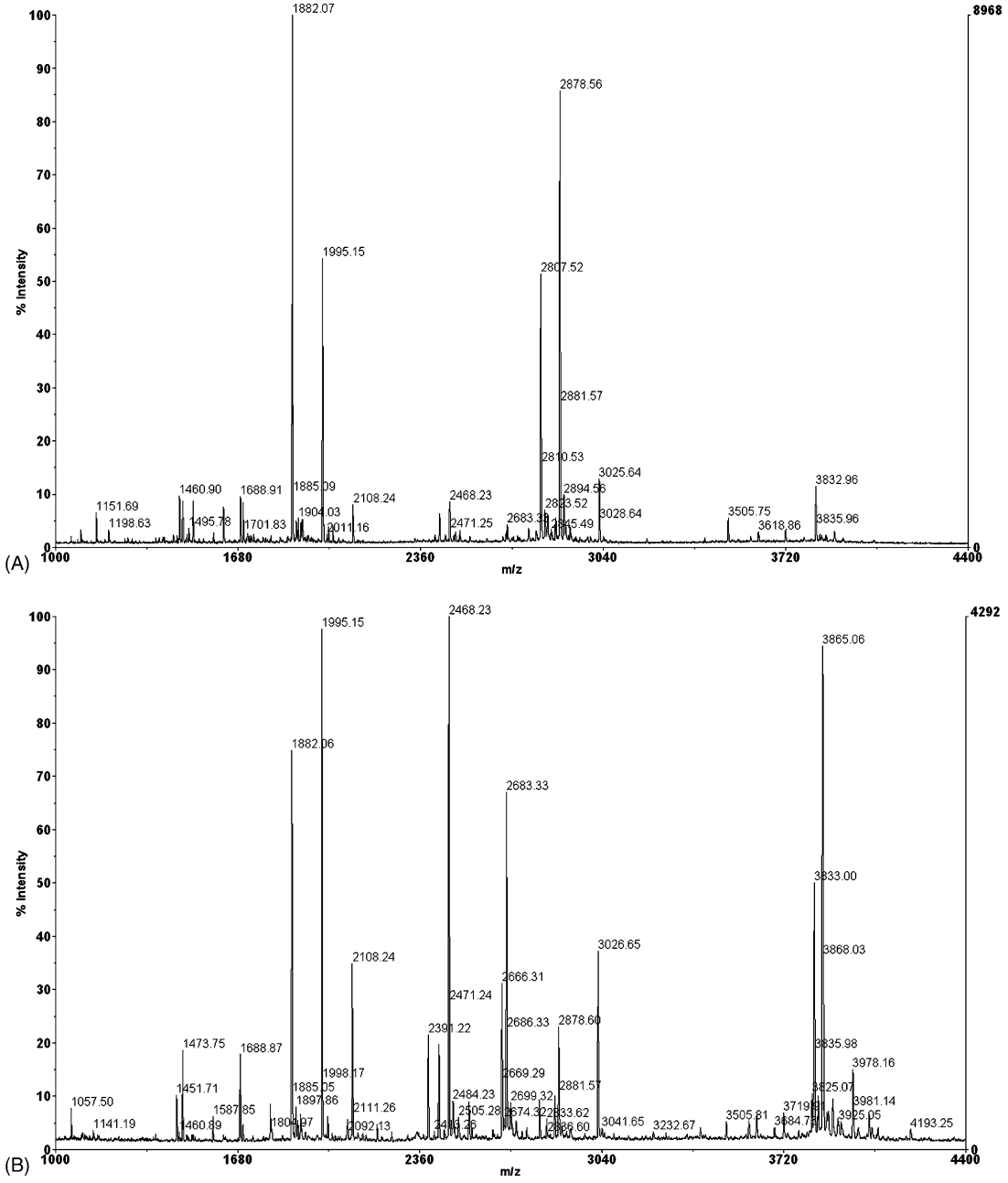

Fig. 1. Positive ion reflectron MALDI-TOF spectra of the peptic β-casein digest. The samples were prepared, using the dried droplet method, by mixing the sample with (A) α-cyano-4-hydroxycinnamic acid and (B) sinapinic acid, respectively.

CID was set between 1 and 3 u depending on experimental conditions. Fragmentation was carried out varying the relative collision energy between 18 and 38% to achieve optimal results for $[M + H]^+$ ions.

2.6. Database search

The fragment ion spectra of MALDI-PSD and tandem MS obtained were processed using Mascot Distiller and analyzed by searching the sequence databases Swiss-Prot and MSDB with Mascot (version 1.8, Matrix Science, London, UK) [22]. The searches were taxonomically restricted to "other mammalia". All peptide masses were taken as monoisotopic masses. The enzyme entry was set to "none" and possible phosphorylation of serine and threonine was considered. Ambiguous results especially from very small peptides were confirmed by comparison of fragment spectra with simulated tandem mass spectra using the software tool GPMAW (Lighthouse data, Odense M, Denmark) [23].

3. Results and discussion

The samples obtained from pepsin digestion of β-casein and following membrane fractionation were subjected to mass spectrometry in order to characterize the peptide pattern as comprehensively as possible. MALDI-TOF-MS was used as the preferable method because of its robustness in complex samples and the prevalence of singly charged peaks. ESI-MS-MS was used as a complementary approach, particularly for smaller peptides.

An important part of method development was the comparison of different MALDI matrices and preparation techniques. Sinapinic acid and α-cyano-4-hydroxycinnamic acid (CHCA) have been found to be suitable matrices for the peptides under investigation. In particular, the concerted use of both matrices in MALDI-MS made it possible to detect and identify a large number of peptides. Some of the peptides could be detected exclusively with CHCA or with sinapinic acid. Examples are the β-casein residues 74–108 and 158–171, which were detected by use of sinapinic acid or CHCA only. Fig. 1 illustrates the influence of the matrix by comparing mass spectra of the same peptide sample obtained with CHCA and sinapinic acid, respectively. A good overview, dealing with the influence of the different matrices in peptide analysis, has been given recently by Gonnet et al. [24] and Karas and Krüger [25].

The poor selectivity of the timed ion selector turned out to be an important limitation for the peptide identification in such complex mixtures. When the resolution is insufficient, more than one peptide is subjected to PSD analysis, resulting in overlapping fragment patterns. These mixtures cannot be processed with the software tools currently available.

In the case of the residue 109–142, the suspected sequence could be confirmed by comparison of their PSD spectra to theoretical tandem MS spectra using the software GPMAW.

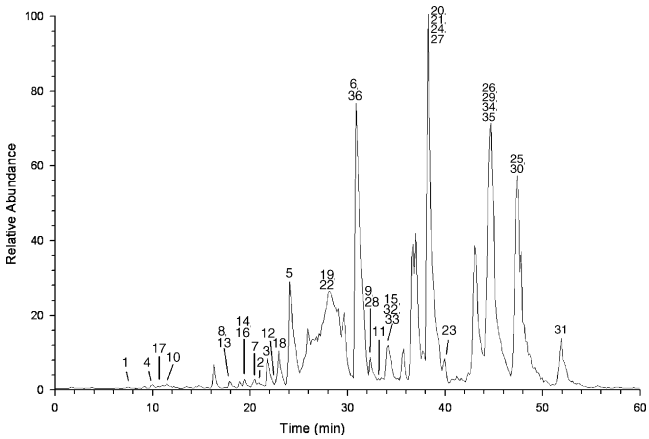

Fig. 2. Reversed-phase LC-ESI-MS base peak chromatogram of a peptic digest of β-casein.

About 20 more peptide candidates were detected but could not be identified with either tandem MS or PSD because of low intensity or insufficient resolution in PSD (see above). A sequence assignment based exclusively on *m*/*z* determination is hampered by the low specificity of pepsin and possible post-translational modifications, such as phosphorylation and glycosylation.

The presence of intensive matrix peaks in the mass range below 600 u makes the identification of peptide peaks in MALDI-MS almost impossible. Hence, LC-ESI-MS and tandem MS were used instead of it for the identification of smaller peptides. These experiments were carried out using an ion trap mass spectrometer. Fig. 2 shows a typical base peak chromatogram. However, considering the potential of this technique for database searching one has to bear in mind that unit mass resolution and a relatively small number of fragment peaks may result in ambiguous sequence information, so that the correct hit may not always be distinguishable from false positives.

For the identification of very small peptides, such as di- and tri-peptides, the situation is even more challenging. Whereas usually the choice of the correct hit from a given set of results is greatly facilitated by the available protein information, small peptides can occur in countless different proteins. In order to circumvent this problem, the database search should be either restricted to a limited set of proteins or the recorded fragment spectra can be compared with simulated tandem MS spectra of suspects. Using the second possibility, all potential sequences for an observed *m*/*z* value were calculated; corresponding tandem MS spectra were simulated and matched to the experimental MS–MS spectra. This was again carried out using GPMAW.

The software Mascot Distiller was used to process MS–MS and MALDI-PSD raw data to create peak lists before database searching. This iteratively working software tool detects signal peaks based on the average isotope distribution of peptides, thereby subtracting detected peaks from the remaining spectrum. Thus, the time needed

to create each single peak list was reduced dramatically from several minutes to approximately 5 s. Furthermore, the mechanism is capable of using information hidden in peaks of low intensity that otherwise could not be distinguished from noise. In this way, the identification rate and sequence coverage could be improved considerably. This is one of the first biomedical applications of this newly developed software.

Table 1
Results of HPLC-ESI-MS and MALDI-TOF-MS of the peptic digest of bovine β-casein

No.[a]	Sequence identified[b]	Calculated $[M+H]^+$	β-Casein residues[c]	MALDI-PSD	ESI-MS-MS	Database	Expected[d]
1	<**SL**>[e]	219.13		-	□	○	Yes
2	A<**FL**>L[e]	279.17	205-206	-	□	○	No
3	S<**LTL**>T	346.23	140-142	-	□	●	No
4	<**QSL**>[e]	347.19		-	□	○	No
5	Q<**AFL**>L	350.21	204-206	-	□	●	Yes
6	A<**FLL**>Y	392.25	205-207	-	□	●	Yes
7	L<**QSW**>M	420.19	156-158	-	□	○	Yes
8	S<**LTLT**>D	447.28	140-143	-	□	○	No
9	Q<**AFLL**>Y	463.29	204-207	-	□	●	Yes
10	L<**LYQE**>P	552.27	207-210	-	□	○	Yes
11	A<**FLLY**>Q	555.32	205-208	-	□	●	Yes
12	S<**QSLTL**>T	561.32	138-142	-	□	●	No
13	S<**LTLTD**>V	562.31	140-144	-	□	●	No
14	F<**PPQSVL**>S	640.37	173-178	-	□	●	Yes
15	I<**PPLTQT**>P	656.36	90-95	-	□	●	No
16	F<**LLYQE**>P	665.35	206-210	-	□	●	Yes
17	**RELEE**>L	675.33	16-20	-	□	●	Yes
18	L<**YQEPVL**>G	748.39	208-213	-	□	●	Yes
19	M<**FPPQSVL**>S	787.43	172-178	-	□	●	No
20	V<**PPFLQPEVM**>G	1057.54	100-108	■	□	●	No
21	G<**PVRGPFPIIV**	1094.67	215-224	■	□	●	No
22	E<**PFTESQSLTL**>T	1122.57	133-142	-	□	●	Yes
23	L<**GPVRGPFPIIV**	1151.69	214-224	■	□	●	Yes
24	T<**PVVVPPFLQPEVM**>G	1451.80	96-108	■	□	●	No
25	E<**PVLGPVRGPFPIIV**	1460.90	211-224	■	-	●	Yes
26	L<**TDVENLHLPLPLL**>Q	1473.83	143-155	■	□	●	Yes
27	L<**TDVENLHLPLPLLQS**>W	1688.92	143-157	■	□	●	No
28	S<**WMHQPHQPLPPTVM**>F	1698.82	158-171	■	□	●	No
29	L<**YQEPVLGPVRGPFPIIV**	1881.06	208-224	■	□	●	Yes
30	L<**LYQEPVLGPVRGPFPIIV**	1994.15	207-224	■	□	●	Yes
31	F<**LLYQEPVLGPVRGPFPIIV**	2107.23	206-224	■	-	●	Yes
32	L<**VYPFPGPIPNSLPQNIPPLTQT**>P	2390.28	74-95	■	-	●	No
33	L<**VYPFPGPIHNSLPQNIPPLTQT**>P	2430.28	74-95	■	-	●	No
34	S<**WMHQPHQPLPPTVMFPPQSVL**>S	2467.24	158-178	■	-	●	No
35	L<**QSWMHQPHQPLPPTVMFPPQSVL**>S	2682.33	156-178	■	-	●	Yes
36	L<**SLSQSKVLPVPQKAVPYPQRDMPIQ**>A	2806.53	179-203	■	-	●	Yes
37	L<**SLSQSKVLPVPQKAVPYPQRDMPIQA**>F	2877.57	179-204	■	-	●	Yes
38	L<**SLSQSKVLPVPQKAVPYPQRDMPIQAF**>L	3024.63	179-205	■	-	●	Yes
39	M<**GVSKVKEAMAPKHKEMPFPKYPVEPFTESQSLTL**>T	3830.99	109-142	■	-	○	No
40	L<**VYPFPGPIHNSLPQNIPPLTQTPVVVPPFLQPEVM**>G	3863.06	74-108	■	-	●	No
41	S<**LVYPFPGPIHNSLPQNIPPLTQTPVVVPPFLQPEVM**>G	3976.14	73-108	■	-	●	No

(■) Identification by MALDI-PSD fragment spectra; (□) identification by ESI tandem MS fragment spectra; (-) no fragmentation spectrum was created using the respective technique; (●) successful identification by database searching; and (○) successful identification by comparison of experimental and theoretical tandem MS spectra (see text).

[a] Numbers 1-36 refer to LC peaks as labeled in Fig. 2.

[b] Peptide sequences identified with adjacent amino acids.

[c] Corresponding site in the bovine β-casein sequence.

[d] Based on cleavage C-terminal to F, L, W, Y, A, E, and Q.

[e] Peptide with several possible positions in the protein sequence.

About 75% of the protein sequence of bovine β-casein was covered by identifying 41 peptides analyzing after their fragment ion spectra. One peptide from κ-casein was identified (data not shown). The genetic variant of the protein could not be determined unambiguously. For example, two different peptide sequences for the same protein segment (74–95) were found. However, most of the residues could be assigned to the genetic variant A^1 [26].

The peptides had lengths of between 2 and 36 amino acids and originated predominantly from the C-terminus of the protein. A complete list of all confirmed peptides is shown in Table 1. Expected cleavages for pepsin were assigned with the help of the cleavage rules of the program PEPTIDEMASS [27]. In addition to the cleavage sites given there (C-terminal to F, L, W, Y, A, E, and Q), some other rather unexpected cleavage sites were found. A cleavage C-terminal to S, T, and M was frequently observed, C-terminal to V, I, G, and D only occasionally.

4. Conclusion

A peptic digest of β-casein, was characterized in a comprehensive manner using two complementary mass spectrometric approaches, one based on MALDI-TOF with PSD/CID, the other one based on LC-ESI-MS with tandem MS. Only this combination made a high sequence coverage possible, which was 75% in this case, including peptides from 2 to 36 amino acids. The growing importance of powerful software tools was demonstrated with the successful application of Mascot Distiller, which greatly facilitates and speeds up the creation of peaklists from MS raw data. The data presented adds some new aspects concerning the possible cleavage sites of the relatively unspecific enzyme pepsin. The in-depth characterization of gastro-analogous digests opens the access to the investigation of biological effects resulting from peptide residues.

Acknowledgement

The authors wish to thank the German Federal Ministry of Education and Research (BMBF) for the financial support within the scope of the research project 0312750A.

References

[1] H. Korhonen, A. Pihlanto, Curr. Pharm. Des. 9 (2003) 1297.
[2] J. Leonil, V. Gagnaire, D. Molle, S. Pezennec, S. Bouhallab, J. Chromatogr. A 881 (2000) 1.
[3] W.R. Aimutis, J. Nutr. 134 (2004) 989S.
[4] N. Yamamoto, M. Ejiri, S. Mizuno, Curr. Pharm. Des. 9 (2003) 1345.
[5] A. Pihlanto, H. Korhonen, Adv. Food Nutr. Res. 47 (2003) 175.
[6] D.A. Clare, H.E. Swaisgood, J. Dairy Sci. 83 (2000) 1187.
[7] S.G. Rival, S. Fornaroli, C.G. Boeriu, H.J. Wichers, J. Agric. Food Chem. 49 (2001) 287.
[8] K.A. Tailford, C.L. Berry, A.C. Thomas, J.H. Campbell, Atherosclerosis 170 (2003) 13.
[9] J. Tauzin, L. Miclo, S. Roth, D. Molle, J.L. Gaillard, Int. Dairy J. 13 (2003) 15.
[10] W. Konigsberg, J. Goldstein, R.J. Hill, J. Biol. Chem. 238 (1963) 2028.
[11] M. Satake, M. Enjoh, Y. Nakamura, T. Takano, Y. Kawamura, S. Arai, M. Shimizu, Biosci. Biotechnol. Biochem. 66 (2002) 378.
[12] B. Brodin, C.U. Nielsen, B. Steffansen, S. Frokjaer, Pharmacol. Toxicol. 90 (2002) 285.
[13] H. Daniel, Annu. Rev. Physiol. 66 (2004) 361.
[14] H.F. Alomirah, I. Alli, Y. Konishi, J. Chromatogr. A 893 (2000) 1.
[15] G. Baggerman, P. Verleyen, E. Clynen, J. Huybrechts, A. De Loof, L. Schoofs, J. Chromatogr. B 803 (2004) 3.
[16] H. John, M. Walden, S. Schäfer, S. Genz, W.G. Forssmann, Anal. Bioanal. Chem. 378 (2004) 883.
[17] M. Mann, R.C. Hendrickson, A. Pandey, Annu. Rev. Biochem. 70 (2001) 437.
[18] J.R. Yates III, Electrophoresis 19 (1998) 893.
[19] P. Berndt, U. Hobohm, H. Langen, Electrophoresis 20 (1999) 3521.
[20] R. Gras, M. Müller, E. Gasteiger, S. Gay, P.A. Binz, W. Bienvenut, C. Hoogland, J.C. Sanchez, A. Bairoch, D.F. Hochstrasser, R.D. Appel, Electrophoresis 20 (1999) 3535.
[21] L. Li, R.W. Garden, E.V. Romanova, J.V. Sweedler, Anal. Chem. 71 (1999) 5451.
[22] D.N. Perkins, D.J. Pappin, D.M. Creasy, J.S. Cottrell, Electrophoresis 20 (1999) 3551.
[23] S. Peri, H. Steen, A. Pandey, Trends Biochem. Sci. 26 (2001) 687.
[24] F. Gonnet, W.G.G. Lemaitre, J. Tortjada, Proteome Sci. (2003).
[25] M. Karas, R. Krüger, Chem. Rev. 103 (2003) 427.
[26] J. Bonsing, J.M. Ring, A.F. Stewart, A.G. Mackinlay, Aust. J. Biol. Sci. 41 (1988) 527.
[27] M.R. Wilkins, I. Lindskog, E. Gasteiger, A. Bairoch, J.C. Sanchez, D.F. Hochstrasser, R.D. Appel, Electrophoresis 18 (1997) 403.