Journal of Chromatography B, 878 (2010) 836–840

Contents lists available at ScienceDirect

Journal of Chromatography B

journal homepage: www.elsevier.com/locate/chromb

Short communication

Even-numbered peptides from a papain hydrolysate of silk fibroin

Jaeho Jeong[a], Won Hur[b,*]

[a] The Center for Cell Signaling & Drug Discovery Research, Ewha Womans University, Seoul 120-750, Republic of Korea
[b] Department of BioEngineering and Technology, Kangwon National University, 192-1 Hyoja-dong, Chunchon, Kangwon 200-701, Republic of Korea

ARTICLE INFO

Article history:
Received 18 September 2009
Accepted 21 January 2010
Available online 29 January 2010

Keywords:
Fibroin
Hydrolysate
Peptide composition

ABSTRACT

A protease with broad substrate specificity usually produces a complex peptide mixture. However, even-numbered peptides were obtained at high proportion upon papain hydrolysis of fibroin composed of highly repetitive Ala- and Gly-rich blocks. MALDI-TOF and ESI mass spectrometric analysis revealed that the even-numbered peptides were in the forms of di-, tetra-, hexa-, and octa-peptides with repeating units in combination of Ala–Gly, Ser–Gly, Tyr–Gly, and Val–Gly. Application of tandem mass spectrometry identified the sequences of the tetra-peptides to be in the order of Ala–Gly–X–Gly (X = Tyr or Val). Therefore, the substrate specificity of papain and the unique repetitive sequence of fibroin generated the hydrolysate composed of even number of amino acids at a high percentage. In this work, fibroin hydrolysate was investigated as an example of an end product of protein hydrolysis, which provides a clue to understand the fate of peptides in a protein hydrolysate.

1. Introduction

There have been many reports on enzymatic hydrolysis of proteins available in large quantities but little information is available on the composition of the peptide mixtures. Recently, application of mass spectrometry identified 41 peptides with a sequence coverage of 75% in a peptic digest of bovine β-casein [1]. However, it is generally difficult to fully characterize complex mixtures, such as protein hydrolysates, that are composed of more than 100 different small peptides [2].

Silkworm fibroin, unlike most other proteins, is mainly composed of five amino acids (glycine 48%, alanine 32%, serine 11%, tyrosine 4.5%, valine 2%) which account for more than 97% of the total number of residues [3]. The DNA sequencing of fibroin gene [4] elucidated that its unique repetitive sequence in 12 low-complexity "crystalline" domains was made up of Gly–X repeats and covering 94% of the sequence; X is Ala in 65%, Ser in 23%, and Tyr in 9% of the repeats [5]. Because of its repetitive sequence, we can expect that fibroin can be proteolyzed into a discrete set of peptides even by a protease of broad substrate specificity. Papain has been known to hydrolyzes proteins with broad specificity for peptide bonds, but with a preference for amino acids bearing a large hydrophobic side chain at the P_2 position [6]. Thus, papain was chosen for this study to hydrolyze fibroin and to investigate the resulting hydrolysate.

Insoluble fibroin can be solubilized in a hot salt solution and subsequent dialysis [7]. The regenerated fibroin can be hydrolyzed by proteolytic enzymes, but the resulting peptide mixture has rarely been identified. Its hydrolysate has been reported to exhibit various biological functions such as blood pressure-depressing activity [8], and anti-genotoxic properties [9]. Two peptides from fibroin hydrolysate showed fibroblast growth-promoting activity [10].

In this study, papain was used to hydrolyze regenerated fibroin, and the resulting hydrolysate was analyzed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) and electrospray ionization mass spectrometry (ESI-MS). Peptides in the hydrolysate were identified by mass spectrometry on the basis of the known primary sequence of fibroin.

2. Materials and methods

2.1. Materials

Regenerated fibroin was prepared from degummed silk fibers of *Bombyx mori* by solubilization in a boiling 50% $CaCl_2 \cdot 2H_2O$ solution and subsequent dialysis. Acetonitrile and methanol of HPLC grade were purchased from Fluka (Milwaukee, WI, USA). Papain (crude papain at 1 U/mg), formic acid, trifluoroacetic acid (TFA) of analytical grade, and the matrix materials α-cyano-4-hydroxy cinnamic acid (CHCA) and 2,5-dihydroxybenzoic acid (DHB) were purchased from Sigma–Aldrich (Milwaukee, WI, USA).

2.2. Hydrolysis of fibroin by papain

Freeze-dried regenerated fibroin (2 g) was dissolved in 40 mL of reaction buffer (5 mM cysteine, 2 mM ethylenediaminetetraacetic acid, pH 7). The proteolytic reaction was begun by adding 0.1 g of

* Corresponding author. Tel.: +82 33 243 6350; fax: +82 33 250 6276.
E-mail address: wonhur@kangwon.ac.kr (W. Hur).

1570-0232/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2010.01.034

papain (1 U/mg). The reaction mixture was stirred at 60 °C for 1 h. The hydrolysis of fibroin was confirmed by tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis (tricine SDS-PAGE) [11].

2.3. Gel permeation chromatography of the hydrolyzed peptides

The molecular-weight distribution of the fibroin hydrolysate was investigated by size exclusion chromatography (SEC) using a TSK-GEL G4000SW column (7.5 mm × 300 mm, Ultrapac, Sweden) with fractionation ranges of 20–10,000 kDa. The mobile phase was 20 mM Tris (pH 7.0) at flow rates of 1.0 mL/min. Blue dextran (2000 kDa), thyroglobulin (669 kDa), apoferritin (443 kDa), β-amylase (200 kDa), alcohol dehydrogenase (150 kDa), albumin (66 kDa), carbonic anhydrase (29 kDa), cytochrome C (12.4 kDa), aprotinin (6512.4 Da), and Tyr–Ala (252.3 Da) were used as molecular-weight markers.

2.4. MALDI-TOF-MS analysis of fibroin peptide

MALDI-TOF-MS (AXIMA-CFR, Shimadzu, Japan) equipped with a 337-nm nitrogen laser and a 1.2-m drift tube, was applied to analyze the molecular weights of fibroin peptides. Samples were extracted with an equal volume of octanol and then the octanol phase (10 μL) was mixed with 10 μL of 50% (v/v) acetonitrile (ACN) with 0.1% (v/v) trifluoroacetic acid (TFA). Aliquots (0.5 μL) were applied onto the target and allowed to air dry. Subsequently, 0.5 μL of CHCA in 50% ACN with 0.1% TFA was applied to the dried sample and again allowed to dry. MALDI-TOF-MS spectra were acquired in positive-ion, with acceleration potential of 20 kV and reflection modes with external calibration.

2.5. HPLC–ESI-MS and tandem MS of fibroin peptide

To identify peptide compositions, samples were dissolved in 1% acetic acid for ionization and analyzed using HPLC–ESI-MS on a LCMS-2010 system (Shimadzu, Japan) consisting of a binary pump, a degasser, and a photodiode array (PDA) detector with an Alltima HP C_{18} column (150 mm × 2.1 mm, Alltech, USA). Mass spectra were acquired in positive-ion mode, scanning from 50 to 2000 m/z, with electrospray ionization. Mobile phase A was 0.1% acetic acid and mobile phase B was ACN. The flow rate was 0.2 mL/min and the injection volume was 5 μL and the column temperature was maintained at 40 °C.

The sequences of the fibroin peptides were determined using collision-induced dissociation (CID) with ESI-MS (Q-TOF2, Micromass, UK) and tandem MS experiments were performed with an ion-trap mass spectrometer Finnigan LCQ Classic (Thermo Electron, San Jose, CA) with an electrospray interface operated in positive-ion mode [1].

3. Results and discussion

3.1. Enzymatic hydrolysis of fibroin

A reaction mixture of papain hydrolysate of fibroin was prepared, insoluble precipitate in the mixture was removed by centrifugation, and the supernatant was examined by tricine SDS-PAGE and SEC using TSK-GEL G4000SW column to analyze the hydrolysis reaction. The profiles of the hydrolysates with varying levels of crude papain (1–8%) and with varying reaction times (10–70 min) were analyzed by tricine SDS-PAGE, which showed a significant degree of similarity in the hydrolysate profiles. To ensure sufficient hydrolysis of fibroin, hydrolysis reaction was carried out with 5% (w/w) crude papain for 60 min and the hydrolyzed sample was analyzed by SEC (Fig. 1). The SEC of the hydrolysate sample showed a broad peak in the range from 22 to 32 min of retention time. The peak was mostly located below the lower size-exclusion limit except the left shoulder in the range from 22 to 27 min. Chromatogram of fibroin showed a sharp narrow peak at the upper size-exclusion limit and a broad peak in the range from 15 to 28 min of retention time. The peak shift in the chromatogram verified hydrolysis of fibroin in the reaction mixture but the left shoulder of the hydrolysate peak indicated that the reaction mixture contains higher molecular weight peptides.

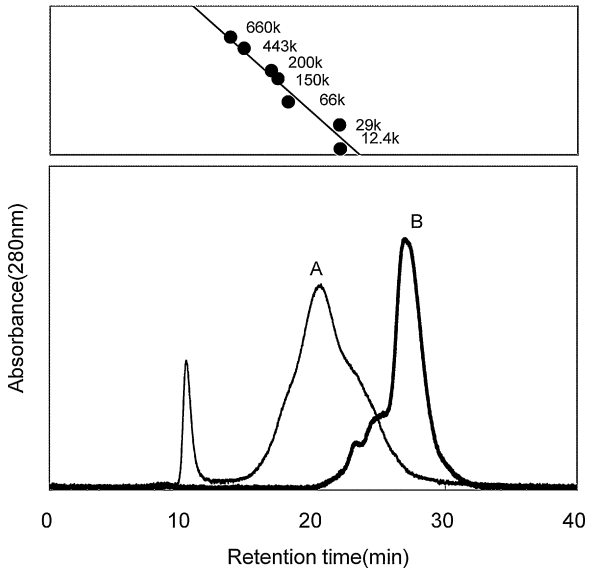

Fig. 1. Size exclusion chromatogram of (A) regenerated silk fibroin and (B) its hydrolysate and molecular weight standards using a TSK-GEL G4000SW column at flow rates of 1.0 mL/min of Tris buffer.

The molecular weight of the heavy chain of fibroin has been reported to be 370 kDa [4]. However, the broad distribution of molecular weight of regenerated fibroin indicates degradation of fibroin molecules during preparation as reported previously by other researchers [12]. The sharp satellite peak at the upper exclusion limit in Fig. 1 could be peptide aggregates routinely observed in regenerated fibroin solutions. Accordingly, it was difficult to determine the molecular weight of the hydrolyzed product using tricine SDS-PAGE or SEC. Thus, mass spectrometric methods were used to analyze the molecular weights of the hydrolyzed peptides in the mixture.

3.2. Mass spectrometric analysis of the peptide mixture

Samples for MALDI-TOF-MS analysis had been prepared by the conventional methods as described elsewhere [13] but failed to get sufficient ionization intensities. After repeated modification of the methods, it was found that octanol extraction of the peptide mixture gave a sufficient intensity (Fig. 2). The mass spectrum showed several discrete mass peaks of strong intensities with many background peaks, suggesting that the hydrolyzed product consisted of a limited number of identical peptides at large portion and many other peptides. Additionally, it was found that the major peaks differ by the mass-to-charge ratio (m/z) of 128 or 144, corresponding to the monoisotopic masses of dipeptides of glycine–alanyl and glycine–seryl units, 128.06 and 144.05, respectively. Thus, the result suggests that papain hydrolysate is composed of peptides those are hydrolyzed by the Gly–Ala and Gly–Ser units.

ESI-MS was also applied to analyze the reaction mixture to exclude the possibility of overlap with matrix peaks and preferential extraction by octanol in the MALDI-TOF-MS analysis. ESI-MS spectrum of the hydrolysate also produced a limited number of discrete peaks at large portion and many background

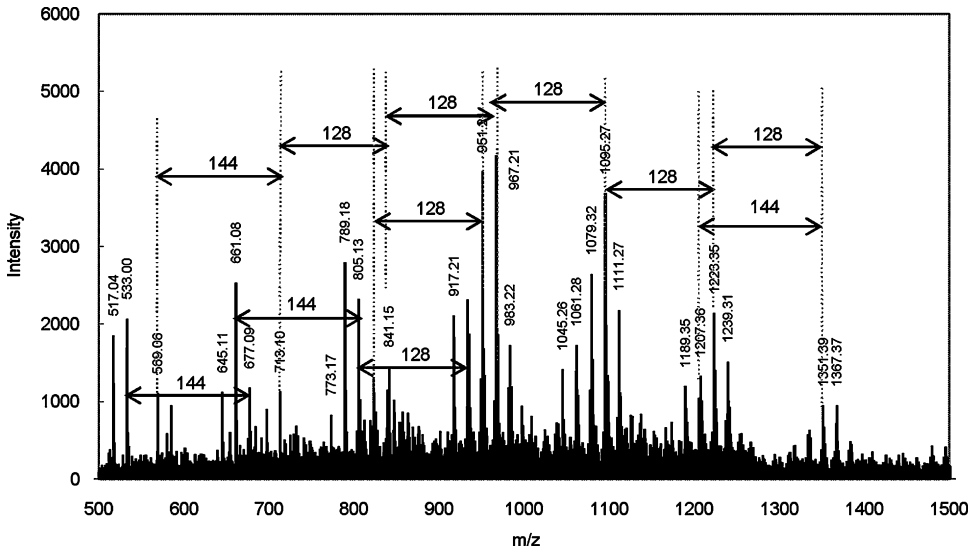

Fig. 2. Mass spectrum of papain hydrolysate of fibroin using MALDI-TOF MS.

peaks (Fig. 3). Background peaks were found in the range from 300 to 1300 *m*/*z*. Each major peaks were identified one by one using a mass-difference diagram with two axes of mass increases of 128.06 and 144.05. Starting from the peak at 275.15 *m*/*z*, later identified as AGAG by ESI-MS–MS, the peak of 403.15 *m*/*z* (=275.15 + 128.06) corresponds to AGAGAG and the peak of 447.20 *m*/*z* (=275.15 + 144.05) corresponds to **SG**AGAG or AGAG**SG**. Another major peak at 367.18 *m*/*z*, also identified as AGYG by ESI-MS–MS, elucidates that 495.24 *m*/*z* (=367.18 + 128.06) matches **AG**AGYG and AGYG**AG** and that 511.23 *m*/*z* (=367.18 + 144.05) matches AGYG**SG** and **SG**AGYG. Accordingly, each peak was identified to corresponding peptides in fibroin amino acid sequence (Table 1).

The highly repetitive sequence of fibroin accounts for the discrete population of even-numbered peptides, which can be the unique feature not found in other proteins. Collagen also has repetitive arrangements of amino acids in the pattern Gly–Pro–Y or Gly–X–Hyp (hydroxyproline) but, to our knowledge, there are no report on the formation of specific peptides at high percentage. Collagen hydrolysate seems to consist of more diverse peptides digested by proteases of broad substrate specificity, because X and Y may be any of various other amino acid residues.

3.3. Identification of peptides in the mixture

The mass spectrum of the hydrolysate enabled us to identify their amino acid compositions of the major peaks but does not provide information about their sequences. Thus, the four peaks of 275.15, 291.15, 303.18 and 367.18 *m*/*z* were chosen for collision-induced dissociation (CID) tandem mass spectrometry. The mass peak of 275 *m*/*z* corresponded to a peptide of two glycines and two alanines and its cleavage pattern included the y3 ion for the sequence AGAG at 204.09 *m*/*z*. Its intensity was twice that of the [M+H$^+$] at 275.13 *m*/*z*. However, no y3 or b3 ions for a sequence starting with G, such as GAGA, were found. This suggests that papain hydrolysis yielded AGAG, not GAGA, although both occur frequently in fibroin sequence. *De novo* sequencing of the peaks of 303.18 and 367.18 *m*/*z* revealed the sequences to be AGVG and AGYG, respectively. However, the peak of 291.15 *m*/*z* was found to be a mixture of two peptides, AGSG and SGAG.

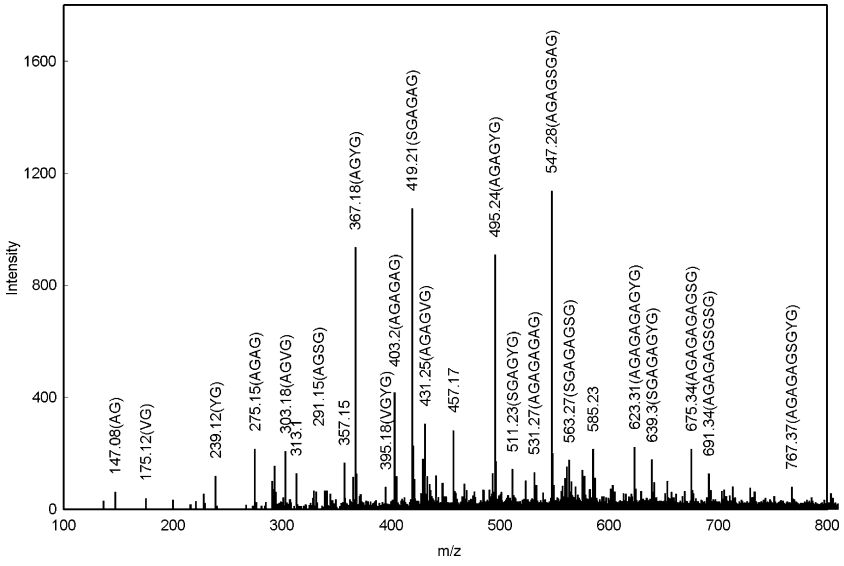

Fig. 3. Mass spectrum of papain hydrolysate of fibroin using ESI-MS.

Table 1
Identification of major peptides of papain hydrolysate of fibroin in ESI-MS spectrum and their occurrences in the reported sequence of fibroin.

Peptide	Theoretical $[M+H]^+$ (monoisotopic)[a]	Peak (m/z)	No. of occurrence in the sequence[b]	Sequence by tandem mass
AG	147.076	147.08	1471	–
SG	163.071	N.F[c]	565	–
VG	175.108	175.12	71	–
YG	239.103	239.12	223	–
AGAG	275.135	275.15	613	AGAG
AGVG	303.166	303.18	49(AGVG) 37(VGAG)	AGVG
AGYG	367.161	367.18	193(AGYG) 198(YGAG)	AGYG
AGSG	291.130	291.15	476(AGSG) 490(SGAG)	AGSG SGAG
AGAGAG	403.194	403.20	96	–
AGAGSG	419.189	419.21	431(AGAGSG) 394(AGSGAG) 442(SGAGAG)	–
AGAGVG	431.225	431.25	17(AGAGVG) 31(AGVGAG) 3(VGAGAG)	–
AGAGYG	495.220	495.24	120(AGAGYG) 140(AGYGAG) 129(YGAGAG)	–
AGAGAGSG	547.247	547.28	230(AGAGSGAG) 14(AGAGAGSG) 217(AGSGAGAG) 30(SGAGAGAG)	–

[a] Calculated by peptide mass calculator V3.2 (http://rna.rega.kuleuven.ac.be/masspec/pepcalc.htm).
[b] Number of occurrence was counted in the sequence of fibroin heavy and light chains (NCBI Reference Sequence: NP_001106733.1 and NP_001037488).
[c] Not found.

It can be concluded that the peptides have glycine at their C-terminal.

It agrees with the substrate specificity of papain, of which active site consists of seven subsites (S_1–S_4 and S'_1–S'_3) accommodating one amino acid residue of a substrate (P_1–P_4 and P'_1–P'_3) [14]. The specificity is controlled by the S_2 subsite, a hydrophobic pocket that preferentially accommodates hydrophobic or aromatic P_2 side chain of the substrate [15]. The sequences of the peptides suggest that papain cleaves polypeptide when alanine, serine, valine, and tyrosine were present at the P_2 position. Though, not all the major peaks were analyzed by tandem mass sequencing, it can be concluded that all the peptide corresponding to the major peaks have glycine at the C-terminal because of the substrate specificity of papain. It also should be noted that crude papain may contain chymopapain during its preparation [16]. The substrate specificity of papain and chymopapain shows a considerable degree of similarity but chymopapain showed strong P_2 specificity for the branched amino acids Val, Thr, and Ile for a model penta-peptide [17]. Therefore, fine composition of peptides of the hydrolysate can be affected by chymopapain if present in the crude papain.

3.4. Peptide composition

Therefore, the major peaks were identified one by one was listed in the form of peptide sequence ending with G in Table 1 with their occurrences in the amino acid sequences of heavy and light chains of fibroin. In spite of frequent occurrences of dipeptides of SG, AG, VG and YG in fibroin amino acid sequence, no peak corresponding to the protonated mass of SG were found and the peaks corresponding to AG, VG and YG had low intensities in the mass spectrum, suggesting that fibroin was hardly digested to dipeptide by papain. The frequency of occurrences of AGYG and YGAG were similar but only AGYG was found to exist in the hydrolysate. It suggests that papain preferentially cut YGAG segment in peptide chains rather than AGYG segment and hardly hydrolyzes the tetra-peptide AGYG. The ionization intensities of AGYG and AGAGYG (sequence not determined) were higher than those of di-peptide YG and octa-peptide containing YG sequence, which indicates that tetra-peptide AGYG and hexa-peptide AGAGYG were not readily digested by papain. Octa-peptide of AGAGAGSG(sequence not determined) and hexa-peptide of AGAGSG(sequence not determined) corresponding to major peaks of high ionization intensity can be the end products of fibroin hydrolysis by papain.

These results seem to be attributable to the substrate specificity of papain of which active site consists of seven subsites, since the small even-numbered peptides may have poor substrate affinity for papain or may act as reversible inhibitors of papain. The inhibitors for cysteine protease has been thoroughly studied [18] and papain has been used as a surrogate enzyme in a drug design effort to obtain potent and selective inhibitors of cathepsin K [19]. Inhibitors such as Z–Phe–Gly–aldehyde and Z–Gly–Phe–Gly–aldehyde (Z: benzyloxycarbonyl) [18] were struc-

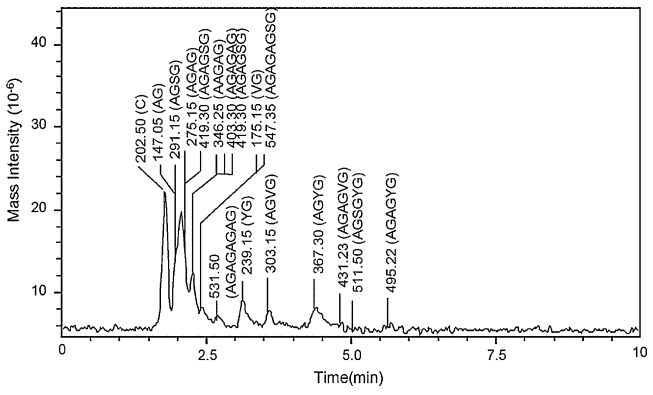

Fig. 4. Reverse-phase chromatogram of the fibroin peptide mixture using HPLC–ESI-MS.

turally similar to YG and AGYG in the hydrolysate. Thus, the peptide pool can be a potential inhibitor or a poor substrate for papain, which could be further investigated by fractionation of the mixture.

The sequences of peptide digested by crude papain is different from those hydrolyzed by other enzymes, which produced two peptides of Gly–Val–Gly–Tyr (GVGY) and Gly–Val–Gly–Ala–Gly–Tyr (GVGAGY) showing blood pressure-depressing activity [9] and chymotrypsin producing octa-peptide (NINDFDED) and decapeptide (VITTDSDGNE) enhancing proliferation of cultured human skin fibroblasts [11]. Thus, papain hydrolysate might not have the identical biological activities presented above but it seems worth to investigate a potential biological activity, for example, cathepsin inhibition from the papain hydrolysate in the present study.

3.5. RP-HPLC of fibroin hydrolysate

To investigate the possibility of fractionation of the peptide mixture based on hydrophobicity, the completely hydrolyzed peptides were tested by ESI-LC/MS using an analytical C18 column. To optimize the resolution of the chromatogram of the peptide mixture, the acetonitrile concentration in the mobile phase was tested at 2.5, 5, 10, 20, 30, and 50%. An isocratic mode of operation with 0.1% acetic acid and 10% acetonitrile was found to be optimal, and ion intensities were plotted against retention time. Peptide sequences were designated at the maximum intensity along with the retention time (Fig. 4). The peptide mixtures eluted in the order of AG, AGSG, AGAG, AGAGSG (sequence not determined), and AGAGAG over the first 2.5 min. The retention times of VG and AGAGSG (sequence not determined) were identical, YG and AGVG followed, after which AGYG, AGAGVG (sequence not determined), AGSGYG (sequence not determined), and AGAGYG (sequence not determined) finally eluted. Thus, the peptides carrying tyrosine residue could be resolved from the mixture by reverse-phase HPLC.

4. Conclusion

Mass spectrometry of a papain hydrolysate of regenerated fibroin revealed that the hydrolysate consists of a large portion of even-numbered peptides and many background peptides. The even-numbered peptides consist of di-, tetra-, hexa-, octa- and deca-peptides found in fibroin amino acid sequence. Tandem mass analysis identified the sequence of the tetra-peptides to be Ala–Gly–X–Gly (X = Tyr or Val) and the presence of glycine at c-terminal of the peptides. Therefore, the substrate specificity of papain and the unique repetitive sequence of fibroin generated the hydrolysate composed of even number of amino acids at a high percentage.

Acknowledgement

The authors thank for the support from Kangwon National University.

References

[1] C.E. Schmelzer, R. Schops, R. Ulbrich-Hofmann, R.H. Neubert, K. Raith, J. Chromatogr. A 1055 (2004) 87.
[2] K. Hakansson, H.J. Cooper, R.R. Hudgins, C.L. Nilsson, Curr. Org. Chem. 7 (2003) 1503.
[3] K. Shimura, A. Kikuchi, K. Ohtomo, Y. Katagata, A. Hyodo, J. Biochem. 80 (1976) 693.
[4] C.Z. Zhou, F. Confalonieri, N. Medina, Y. Zivanovic, C. Esnault, T. Yang, M. Jacquet, J. Janin, M. Duguet, R. Perasso, Z.G. Li, Nucleic Acids Res. 28 (2000) 2413.
[5] C.Z. Zhou, F. Confalonieri, M. Jacquet, R. Perasso, Z.G. Li, J. Janin, Proteins 44 (2001) 119.
[6] J. Lowbridge, J.S. Fruton, J. Biol. Chem. 249 (1974) 6754.
[7] Y. Miyaguchi, J. Hu, Food Sci. Technol. Res 11 (2005) 37.
[8] K. Igarashi, K. Yoshioka, K. Mizutani, M. Miyakoshi, T. Murakami, T. Akizawa, Biosci. Biotechnol. Biochem. 70 (2006) 517.
[9] K.J. Park, H.H. Jin, C.K. Hyun, Process Biochem. 38 (2002) 411.
[10] H. Yamada, Y. Igarashi, Y. Takasu, H. Saito, K. Tsubouchi, Biomaterials 25 (2004) 467.
[11] H. Schagger, G. von Jagow, Anal. Biochem. 166 (1987) 368.
[12] J.H. Yeo, K.G. Lee, Y.W. Lee, S.Y. Kim, Eur. Polym. J. 39 (2003) 1195.
[13] E. Nordhoff, M. Schurenberg, G. Thiele, C. Lubbert, K.D. Kloeppel, D. Theiss, H. Lehrach, J. Gobom, Int. J. Mass Spectrom. 226 (2003) 163.
[14] I. Schechter, A. Berger, Biochem. Biophys. Res. Commun. 27 (1967) 157.
[15] J.R. Kimmel, E.L. Smith, Adv. Enzymol. Relat. Subj. Biochem. 19 (1957) 267.
[16] B.S. Baines, K. Brocklehurst, Biochem. J. 173 (1978) 345.
[17] D.N. Gosalia, C.M. Salisbury, J.A. Ellman, S.L. Diamond, Mol. Cell. Proteomics 4 (2005) 626.
[18] H.H. Otto, T. Schirmeister, Chem. Rev. 97 (1997) 133.
[19] J.M. LaLonde, B. Zhao, W.W. Smith, C.A. Janson, R.L. DesJarlais, T.A. Tomaszek, T.J. Carr, S.K. Thompson, H.-J. Oh, D.S. Yamashita, D.F. Veber, S.S. Abdel-Meguid, J. Med. Chem. 41 (1998) 4567.