Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography B, 792 (2003) 33–48
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

Determination of endogenous peptides in the porcine brain: possible construction of Peptidome, a fact database for endogenous peptides

Naoto Minamino[a,*], Junko Tanaka[a], Hiromiki Kuwahara[a], Takahiro Kihara[a], Yoshinori Satomi[b], Masami Matsubae[a,b], Toshifumi Takao[b]

[a]*Department of Pharmacology, National Cardiovascular Center Research Institute, 5-7-1 Fujishirodai, Suita, Osaka 565-8565, Japan*
[b]*Proteomics Research Center, Institute for Protein Research, Osaka University, Yamadaoka, Suita, Osaka 565-0871, Japan*

Received 8 November 2002; received in revised form 25 January 2003; accepted 17 March 2003

Abstract

Peptides play crucial roles in many physiological events. However, a database for endogenous peptides has not yet been developed, because the peptides are easily degraded by proteolytic enzymes during extraction and purification. In this study, we demonstrated that the data for endogenous peptides could be collected by minimizing the proteolytic degradation. We separated porcine brain peptides into 5250 fractions by 2-dimensional chromatography (first ion-exchange and second reversed-phase high-performance liquid chromatography), and 75 fractions of average peptide contents were analyzed in detail by mass spectrometers and a protein sequencer. Based on the analysis data obtained in this study, more than 10 000 peptides were deduced to be detected, and more than 1000 peptides to be identified starting from 2 g of brain tissue. Thus, we deduce that it is possible to construct a database for endogenous peptides starting from a gram level of tissue by using 2-dimensional high-performance liquid chromatography coupled with a mass spectrometer.

Keywords: Peptidome; Peptides, endogenous

1. Introduction

The draft sequences of the human genome were published in 2001 [1,2], and the complete sequences of the human genome will be released in the spring of 2003. As the next stage of genome research, the data of proteins that are encoded by the genome are rapidly being accumulated in the database in these years. Proteomic analysis was started by identification of proteins expressed in the target cells and tissues, but has now been extended to differential display and interaction in addition to the structural biology of the proteins [3–5]. On the other hand, analysis of the peptides with a relative molecular mass (M_r) ≤10000 can not be covered by the proteomic research, in spite that biologically active peptides are participating in many physiological events and play crucial roles in the regulation of metabolism, circulation, behavior and so on [6,7].

*Corresponding author. Tel.: +81-6-6833-5012x2507; fax: +81-6-6835-5349.
E-mail address: minamino@ri.ncvc.go.jp (N. Minamino).

1570-0232/03/$ – see front matter
doi:10.1016/S1570-0232(03)00280-0

The number of biologically active peptides identified in one species is not more than 100, even in the case of humans [8]. Based on the human genome sequences determined to date, many so-called orphan receptors have been deduced, and a portion of them are assumed to be the receptors for peptidergic ligands [9,10]. These findings suggest that unidentified biologically active peptides are present and function in human bodies. Isolation and identification of these peptides are recognized to be the key sources for drug discovery, development of diagnostic and therapeutic methods as well as for elucidating physiological and biological events [11–13].

Since peptides are generated from precursor proteins and then exert their biological effects, peptides must be cleaved off from the precursors at specific sites. The cleavage sites of the precursors are often different in each tissue, and distinct peptides are produced from the same precursor in the different tissues [14–16]. Modification of the peptides, such as amidation, acetylation and pyroglutamination, is essential in addition to the specific cleavage [17,18]. In the case of a recently identified peptide ghrelin, the new modification of octanoylation is required for eliciting its effects [19]. The processing including cleavage and modification is the most important feature of biologically active peptides, which is often difficult to deduce from the genome sequences. Thus, fact data for endogenous peptides, which are defined as the data obtained by direct measurement and analysis for the peptides originally present in the cells, tissues and living organisms, must be accumulated in order to advance peptide research.

Despite their physiological and biological importance, construction of the database for endogenous peptides has not yet succeeded. The main reason for the difficulties in the analysis of endogenous peptides is that both peptides and proteins are degraded by proteolytic enzymes during extraction and purification. By minimizing the artificial degradation of proteins and peptides, however, we demonstrated that endogenous peptides can be detected at substantial levels. By separating peptides by 2-dimensional (2-D) chromatography composed of ion-exchange and reversed-phase high-performance liquid chromatography (HPLC) and analyzing with mass spectrometers, more than 10 000 peptides were deduced to be detected. During this separation step, we are able to obtain the data for the physicochemical properties of peptides, such as hydrophobicity and net charge. These results are expected to provide the solid basement for constructing the fact database for endogenous peptides, designated ''Peptidome''.

As the first step, we analyzed the peptides with $M_r \leq 2500$ in the present study. Since adrenomedullin with 52 residues and insulin with 51 residues belong to the longest peptide group, the target molecular range of the peptides will be extended up to 6000 in order to cover all biologically active peptides.

2. Experimental

2.1. Small scale preparation of porcine brain

Porcine brain (14 pieces, 1150 g) was collected in the local slaughterhouse soon after killing and kept in ice in a plastic bag before extraction. After removing hematoma and leptomeninges (pia mater and arachnoid) and rinsing with saline, porcine brain was minced, added portion-wise to 5 volumes (v/w) of boiling water, and boiled for 10 min after addition of the final portion. After cooling, glacial acetic acid was added to make a final concentration of 1 *M* for facilitating extraction of peptides, and tissue was homogenized with a Polytron homogenizer. The homogenate was centrifuged, and the supernatant was condensed to 15% volume by evaporation. The condensate was submitted to acetone precipitation at a final concentration of 66% at 4 °C. After removing precipitates by centrifugation, the resulting supernatant was evaporated in a vacuum, and was then loaded onto a C_{18} column (120 ml, LC-SORB SPW-C-ODS, Chemco, Osaka, Japan). The column was washed with 2 column volumes each of 0.1% trifluoroacetic acid (TFA) and 10% acetonitrile containing 0.1% TFA. The materials adsorbed on the column were eluted with 60% acetonitrile containing 0.1% TFA. The column eluate was evaporated, and the resulting solution was subjected to batch-wise chromatography on an SP-Sephadex C-25 column (H^+-form; 2.4×4 cm, Amersham Bioscience, Buckinghamshire, UK). Successive elutions with 2 column volumes of 1 *M* acetic acid, 2 *M* pyridine and 2 *M* pyridine-acetate (pH 5.0) yielded three respective fractions of SP-I, SP-II and SP-III. The SP-III

fraction was used for purification and characterization.

2.2. Separation of peptides in SP-III fraction of the small scale preparation

SP-III fraction (500 g equivalents) was separated by gel permeation on a Sephadex G-50 column (fine, 1.8×135 cm, Amersham Bioscience), and an aliquot of each fraction was submitted to radioimmunoassay (RIA) for C-type natriuretic peptide (CNP). RIA for CNP was performed as reported previously [20], and two major endogenous forms of CNP have been identified as CNP-22 and CNP-53 of M_r 2197.6 and 5796.7 [21,22]. Two-fifth of the fractions (200 g equivalents) corresponding to $M_r \leq 6000$ were pooled and subjected to the second gel permeation chromatography on a Sephadex G-25 column (fine, 1.8×135 cm). Aliquots from every successive 5 fractions starting from fraction 26 were pooled, and each pooled fraction was separated by reversed-phase HPLC on a Chemcosorb 5ODS-H column (300A, 4.6×250 mm, Chemco) with a linear gradient elution of acetonitrile from 10% to 60% in 0.1% TFA at a flow-rate of 1 ml/min over 60 min.

2.3. Large scale preparation of porcine brain

A maximum of 4 kg of porcine brain were treated at one time. Extracts were prepared as described above, and proteins were removed by filtering through a Pellicon cassette PLTK (0.5 m^2, Millipore, Bedford, MA, USA). The filtrates were then condensed with a Pellicon cassette PLAC (0.5 m^2). The condensate was treated with a C_{18} column (1 l), as described above, and the fraction eluted with 60% acetonitrile containing 0.1% TFA was evaporated and condensed. In total, 20 kg equivalents of porcine brain extracts were pooled at this stage. The pooled condensate was then subjected to batch-wise chromatography on an SP-Sephadex C-25 column (H^+-form, 5×25 cm). Adsorbed materials on the column were successively eluted as described above. Since major components of the SP-II and SP-III fractions were peptides, these two fractions were combined and named ''crude peptide fraction''.

2.4. Separation of the crude peptide fraction

After lyophilization, the crude peptide fraction (1 kg equivalent) was separated by gel permeation chromatography on a Sephadex G-50 column (fine, 5×145 cm). Sephadex G-50 was repeated 20 times, and the eluates of 20 chromatographies were collected into the same tubes. An aliquot of each fraction was submitted to RIA for CNP, and the column eluates were divided into two fractions, $M_r \leq$ 2500 and M_r 2500–6000, based on the elution positions of CNP-53 and CNP-22. In this study, separation and analysis were performed only for the fraction of $M_r \leq 2500$, named ''small peptide fraction''.

2.5. Two-dimensional HPLC of the small peptide fraction

The small peptide fraction (50 g equivalents) was first separated by SP-ion-exchange HPLC (TSK gel SP-2SW, 4.6×250 mm, Tosoh, Tokyo, Japan) eluting with a linear gradient elution of ammonium formate (pH 3.8) from 10 m*M* to 1 *M* in the presence of 10% acetonitrile at a flow-rate of 1 ml/min over 120 min, and fractions were collected every 2 min (total 70 fractions). This ion-exchange HPLC was repeated twice, and the eluates were collected into the same tubes. All fractions obtained from SP-ion-exchange HPLC were then each subjected to reversed-phase HPLC on a C_{18} column (Symmetry 300A 5 μm C_{18}, 4.6×250 mm, Waters, Milford, MA, USA) eluting with a linear gradient elution of acetonitrile from 10% to 60% in the presence of 0.1% TFA at a flow-rate of 1 ml/min over 60 min, and fractions were collected every 1 min (total 75 fractions).

2.6. Mass spectrometric (MS) analysis

In the case of the small scale preparation, an aliquot of each fraction (2–3 μl/fraction) in reversed-phase HPLC was evaporated to dryness, and dissolved in 50% methanol containing 1% acetic acid. Electrospray ionization (ESI)-tandem mass spectrometric (MS/MS) analysis was carried out on a hybrid quadrupole orthogonal acceleration tandem mass spectrometer (Q-TOF II, Micromass, Manchester, UK). The peptide solutions were loaded into a

borosilicate nanoflow tip (Micromass), and set into an ESI source. MS/MS data were processed by a maximum entropy data enhancement program, MaxEnt 3™ (Micromass), which is capable of deconvoluting a spectrum where peaks in a variety of charge states are present, thus producing a simplified spectrum which consists of only mono-isotopic peaks in a single charge state. The resultant spectra were interpreted by SeqMS, a software aid for de novo sequencing by MS/MS (http://www.protein.osaka-u.ac.jp/rcsfp/profiling) [23].

In the case of large scale preparations, an aliquot of each fraction (0.5 μl/fraction, 50 mg equivalents) was spotted on the plate with an equal volume of a matrix solution (CHCA: α-cyano-4-hydroxycinnamic acid, Sigma–Aldrich, Milwaukee, WI, USA), and submitted to matrix-assisted laser desorption ionization (MALDI) time-of-flight (TOF) MS analysis with Voyager-DE Pro (Applied Biosystems, Foster City, CA, USA). Another aliquot of each fraction (20 μl/fraction, 2 g equivalents) was evaporated and analyzed with ESI-Q-TOF II as described above. The acquired MS/MS spectra were interpreted by SeqMS software, and database searching was performed with an in-house database server using Mascot software (Matrix Science, London, UK).

2.7. Amino acid sequence analysis with protein sequencer

Peptides in 75 fractions in the 2-D HPLC were further purified by reversed-phase HPLC on a diphenyl column (219TP5215, 1.0×150 mm; Vydac, Hesperia, CA, USA) using a linear gradient elution of acetonitrile from 5% to 60% in 0.1% TFA at a flow-rate of 40 μl/min. Amino acid sequence of the peptides obtaining more than 50% purity were analyzed with a Procise cLC protein sequencer (Applied Biosystems).

3. Results

3.1. Small scale preparation of peptides in porcine brain extracts

In the case of small scale preparations starting from 1150 g of porcine brain, about 350 mg of dried materials were recovered after extraction, acetone precipitation and reversed-phase C_{18} column treatment. By the successive batch-wise treatment with SP-Sephadex, we obtained the SP-III mainly composed of basic peptides (dry weight: 145 mg) for further characterization of peptides. After removing proteins in the SP-III fraction by Sephadex G-50 gel permeation chromatography, the peptide fraction of $M_r \leq 6000$ (dry weight: 10.5 mg/200 g equivalents), which included peptides up to CNP-53, was re-subjected to gel permeation on a Sephadex G-25 column. Peptides were substantially eluted from fraction 27 to fraction 55, and every successive 5 fractions starting from fraction 26 were each pooled as one group, and each group was separated by reversed-phase HPLC. Fig. 1 shows the elution profile of reversed-phase HPLC of fractions 36–40. Other pooled fractions obtained from gel permeation chromatography were also separated by reversed-phase HPLC in a similar manner (data not shown).

In addition to the SP-III fraction, the SP-II fraction was also separated by gel permeation chromatography and the peptide fraction of $M_r \leq 6000$ was recovered. Since greater than 90% of the peptides

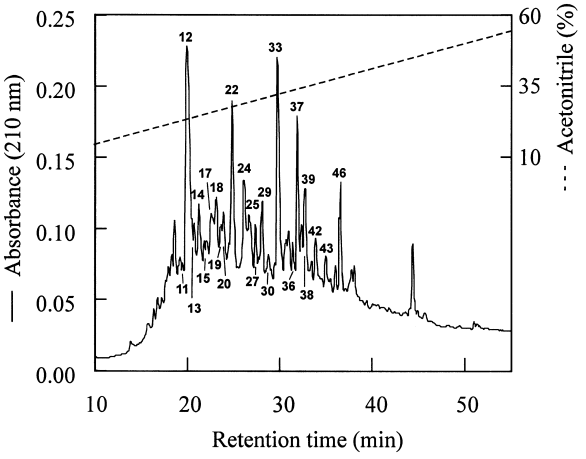

Fig. 1. Reversed-phase HPLC of fractions 36–40 in Sephadex G-25 gel permeation of the basic peptides of $M_r \leq 6000$ Da prepared from porcine brain extracts. Sample: one-fourth of fractions 36–40 in Sephadex G-25 gel permeation chromatography (50 g equivalents). Column: 4.6×250 mm, Chemcosorb 300 5ODS-H. Solvent system: linear gradient elution of acetonitrile from 10% to 60% in 0.1% TFA over 60 min. Flow-rate: 1 ml/min. Absorbance was measured at 210 nm and 280 nm, and the data for 210 nm is shown in the figure.

were confirmed to be recovered in the SP-II and SP-III fractions, total weight of the peptides of $M_r \leq$ 6000 was estimated to be about 140 mg starting from 1150 g of porcine brain tissue. As the protein content in the porcine brain was estimated at about 10% of the wet tissue weight, the peptide was deduced to be present at the concentration about 0.1% that of the protein. This ratio was reproducible in the small scale preparations, while the peptide weight was slightly higher in the case of large scale preparations.

3.2. MS analysis of peptides in the small scale preparation

Aliquots of collected fractions were submitted to MS and MS/MS analyses with ESI-Q-TOF II. Fig. 2 shows a typical mass spectrum of peak 24 in Fig. 1. The MS/MS spectra of major peaks in Fig. 2 were acquired for de novo sequencing of peptides, and only the peak of m/z 732.86 was identified to be GNSRTFTVAIGLTQ. Based on the EST database searching, this peptide was found to be identical to a portion of the transcript of human EST, F025111. Among the 66 peaks clearly observed in the mass spectra, we were able to unambiguously determine 20 amino acid sequences using the SeqMS software, as listed in Table 1. Three amino acid sequences preceding and succeeding to the identified peptides in the precursors are also shown as references even if their precursor sequences have not been in the pig, and the identified peptide sequence is underlined and flanked by marks in the precursor. In these 20 peptides, we found α-neo-endorphin, substance P and its methionine sulfoxide (M(O)) form, neurokinin A and its M(O) form, and rimorphin in addition to one fragment of chromogranin A, which were cleaved off from the precursors at the marked

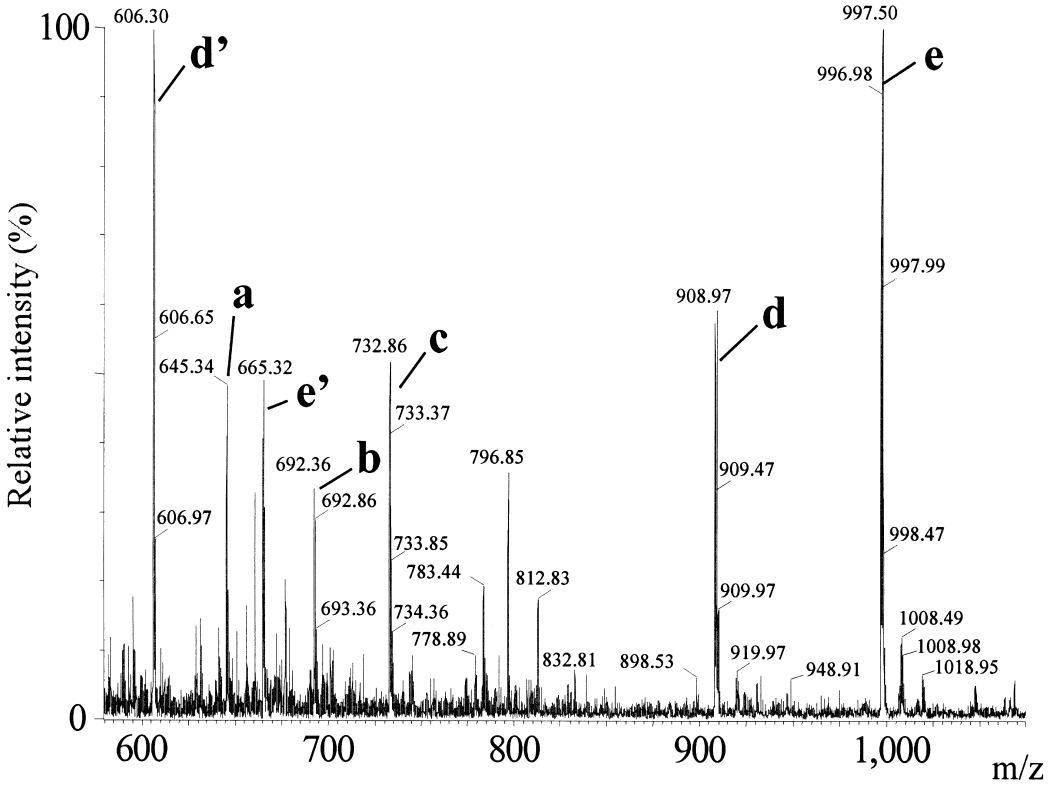

Fig. 2. Mass spectrum of fraction 24 in the reversed-phase HPLC of the small scale preparation of porcine brain extract measured with ESI-Q-TOF II. (a–e) Doubly charged ions, and d′ and e′ were triply charged ions. $(M+H)^+$ were calculated as follows; a, 1992.5; b, 1383.71; c, 1464.71; d, 1815.92; e, 1992.92.

Table 1
Mass spectrometric analysis of peptides in the reversed-phase HPLC of fractions 36–40 in Sephadex G-25 gel permeation chromatography of the basic peptides prepared from porcine brain extracts (small scale preparation)

Fr. no.	Mass	Sequence	Peptide name	Precursor/Protein	Database/Reg. no.	Species	Other ob. mass	Note
11	1279.61							
12	1210.65	RIV::**APPGGRANITSLG**		Dihydropyrimidinase-related prot-2	Swiss-Prot/O02675	Cow	1153.63, 1139.59, 1254.59, 1348.59, 1840.79	C-term
13	1134.66	GVG::**GIKVERQTFG**:: EAT		Cathepsin D	Swiss-Prot/P00795	Pig		
14	914.49						1399.67	
15	694.55							
17	845.61	GRN::**FRNPLAK**		α-Enolase	Swiss-Prot/Q9XSJ4	Cow	1649.98	C-term
18	1146.68	HKR::**HKTDSFVGLM(O)-a**::KRA	Neurokinin A	Preprotachykinin 1	Swiss-Prot/P01289	Cow		
	1267.81	FLA::**NVSTVLTSKYR**		Hemoglobin α chain	Swiss-Prot/P01965	Pig		C-term
19	1346.84	TWD::**DRRAEGTFPGKI**		Cytochrome C oxidase polypeptide VIb	Swiss-Prot/P00429	Cow	1263.81	C-term
20	1145.63	**APMSHFFVP(I/L)**					1150.57, 1164.61, 1204.67	
22	1228.69	VKR::**YGGFLRKYPK**::RSS	α-Neo-endorphin	β-Neo-endorphin-dynorphin prec	Swiss-Prot/P01214	Pig	811.47, 1564.75	
	1276.63	KEP::**DVPVFFPSGQLA**::ESV			Gene Bank/AL135960	Human		
24	1464.71	RQT::**GNSRTFTVAIGLTQ**::QNA			Gene Bank/F02511	Human	1289.67, 1383.71, 1815.92, 1992.95	
25	1420.88	LHS::**LGHPPGPSAVPVPVP**::GGG		Homeobox prot six2	Swiss-Prot/Q9NPC8	Human		
27	943.57						960.57, 1268.71, 1583.85,	
29	987.59						1720.85	
30	1424.75	WDD::**YVPKLYEQLSGK**		Phosphatidylethanolamine binding prot	Swiss-Prot/P13696	Cow	1031.59, 1696.85, 2045.01, 2294.11	C-term
33	1260.65	GSG::**FTNTMRVVPVP**		Pyruvate kinase, M1 isozyme	Swiss-Prot/P14618	Human	1199.63	C-term
	1363.71	IAR::**RPKPQQFFGLM(O)-a**::KRD	Substance P	Preprotachykinin 1	Swiss-Prot/P01289	Cow	1521.61	
36	1133.55	HKR::**HKTDSFVGLM-a**::KRA	Neurokinin A	Preprotachykinin 1	Swiss-Prot/P01289	Cow	1566.81	
37	1570.91	QKR::**YGGFLRRQFKVVT**::RSQ	Rimorphin	β-Neo-endorphin-dynorphin prec	Swiss-Prot/P01214	Pig	1365.71, 1404.69	
38	1483.82	TYN::**FAVLKLM(O)GRGTKF**		Myelin proteolipid prot	Swiss-Prot/P04116	Cow	1295.73, 1404.69, 1704.86,	C-term
39	1012.63						1548.77, 1672.87	
42	1346.69	SFR::**APAYGFRGPGLQL**::RRG		Chromogranin A prec	Swiss-Prot/P04404	Pig	1120.67, 1641.83	
43	1003.61	EGY::**ALPHAILRL**::DLA		β-Actin	Swiss-Prot/Q8SPK6	Pig	966.59	
46	1347.71	IAR::**RPKPQQFFGLM-a**::KRD	Substance P	Preprotachykinin 1	Swiss-Prot/P01289	Cow	1199.65, 1736.90	

Mass indicates $(M+H)^+$. Peptides with bold sequences were identified by the present study, and either end of the identified peptide is flanked by a mark :: in the precursor sequence. Abbreviations used in the table: Fr. no., fraction number in the reversed-phase HPLC; Reg. no., registration number; ob., observed; -a, amide structure at the C-terminus (an amide-donating Gly was deleted from the precursor instead); prot, protein; prec, precursor; (I/L), isoleucine or leucine; M(O), methionine sulfoxide; C-term, C-terminal fragment of protein. Species and registration number in the database is chosen in the order of cow, and human, if the precursor/protein has not been identified in the pig.

Table 2
Amino acid sequences of identified peptides in the reversed-phase HPLC of fraction 20 in SP-ion-exchange HPLC of the small peptide fraction prepared from porcine brain extracts (large scale preparation)

Fr. no.	Mass	Ion count	Sequence	Precursor/Protein	Database/Reg. no.	Species	Note
16	1221.65	15000	EGQ::**KPGFGYGGRAPD**::YKP	Myelin basic prot	Swiss/P81558	Pig	
	1624.99	4710	GGD::**GATKYITKSVTVTQK**::VEE	Neurofilament triplet M prot	Swiss/O77788	Cow	
	1777.96	9200	PAD::**SAVPGAQEEEAHRRQL**::RAV	Procholecystokinin	Swiss/P01356	Pig	
	1793.07	9400	PAS::**SAKTSPAKQQAPPVRNL**::HQS	Dihydropyrimidinase-related prot-2	Swiss/O02675	Cow	
17	1546.91	5898	YSN::**RVVDLM(O)VHM(O)ASKE**	Glyceraldehyde 3-phosphate dehyrogenase	Swiss/P00355	Pig	C-term
	1898.04	5204	**PHSYPALSAEQKKELSD**::IAL	Fructose-biphosphate aldolase C	Swiss/P09972	Human	N-term
	1903.05	8827	**TPEVSGYSYEKTER**				
	1955.10	6946	FSW::**GAEGQKPGFGYGGRAPDYK**::PAH	Myelin basic prot	Swiss/P81558	Pig	
18	1190.77	4745	DEP::**ILSNRSGDHRG**::KFK	Phosphatidylethanolamine-binding prot	Swiss/P13696	Cow	
	1633.03	9700	**SGSAKVAFSAIRSTNH***	Cerebellin	Swiss/P23436	Pig	RDB
19	1627.90	1371	**VLSAADKANVKAAWGK**::VGG	Hemoglobin α chain	Swiss/P01965	Pig	N-term
	1896.16	2803	LSG::**AQIDDNIPRRTTQRIV**::APP	Dihydropyrimidinase-related prot-2	Swiss/O02675	Cow	
20	1164.73	3332	SVP::**TEEELSPTEEE**::QKA	Protein-tyrosine phosphatase 1B	Swiss/P20417	Rat	
	1530.96	1736	YSN::**RVVDLM(O)VHMASKE**	Glyceraldehyde 3-phosphate dehyrogenase	Swiss/P00355	Pig	C-term
	1787.11	3137	KGF::**SPQHKITSFEEAKGLD**::RIN	Ser/Thr prot phosphatase 2B catalytic subunit α isoform	Swiss/P48452	Cow	
	2060.38	4942	GVG::**GIKVERQTFGEATKQPGLT**::FIA	Cathepsin D	Swiss/P00795	Pig	
21	1785.21	6127	**VLSAADKANVKAAWGKVG**::GQA	Hemoglobin α chain	Swiss/P01965	Pig	N-term
	1885.15	14000	LDD::**PGANAAAAKIQASFRGHM(O)A**::RKK	Neurogranin	Swiss/P35722	Cow	
22	1450.81	906	KPD::**M(O)AEIEKFDKSKL**::KKT	Thymosin β4	Swiss/P01253	Human	
	1583.99	6926	AGN::**SNTRFGIAAKYQID**::PDA	Voltage-dependent anion-selective channel prot-1	Swiss/P45879	Cow	
	1785.19	8484	**VLSAADKANVKAAWGKVG**::GQA	Hemoglobin α chain	Swiss/P01965	Pig	N-term
23	1885.14	3166	LDD::**PGANAAAAKIQASFRGHM(O)A**::RKK	Neurogranin	Swiss/P35722	Cow	
25	1301.72	650	NEA::**AGNKYVPRAILV**::DLE	Tubulin β chain	Swiss/P02554	Pig	
	1319.72	2192	AQD::**AQGTLSKIFKLGG**::RDS	Myelin basic prot	Swiss/P81558	Pig	
	1567.56	4060	**PPYTITYFPVRGR**::CEA	Glutathione S-transferase P	Swiss/P80031	Pig	N-term
	1764.11	10000	QLR::**AVQKDGESRAHLGALL**::ARY	Procholecystokinin	Swiss/P01356	Pig	
	1975.33	1945	M::**PFSNSHNTLKLRFPAED**::EFP	Creatine kinase, B chain	Swiss/P05124	Dog	
27	1397.78	32000	NVN::**AGGHKLGLGLEFQA**	Voltage-dependent anion-selective channel prot-1	Swiss/P45879	Cow	C-term

Mass indicates $(M+H)^+$. Ion count was measured by a MALDI-TOF mass spectrometer. Peptides with bold sequences were identified by the present study, and either end of the identified peptide is flanked by a mark :: in the precursor sequence. Abbreviations used in the table: Fr. no, fraction number in the reversed-phase HPLC; Reg. no., registration number; Swiss, Swiss-Prot Protein Knowledge Database; prot, protein; M(O), methionine sulfoxide; RDB, registered in DB; C-term, C-terminal fragment of protein; N-term, N-terminal fragment of protein. Species and registration number in the database is chosen in the order of cow, human, dog, and rat, if the precursor/protein has not been identified in the pig. *This peptide was identified in the pig brain extracts and registered in the database, but its precursor sequence has not been determined in the pig. The precursor sequence is omitted in this case.

sites. Three peptides were fragments of the abundant proteins (myelin proteolipid protein, β-actin and contaminated hemoglobin), and seven other peptides were cleaved off from enzymes, binding proteins or a homeobox protein. In the three other peptides, two were found to be portions of putative proteins registered in the EST or genome database (F02511 and AL135960), while another peptide of fraction 20 was sequenced but its identical sequence was not found by the database searching. In addition, seven peptides were derived from the C-termini of proteins, which were probably cleaved off from the proteins during the extraction step.

In addition to the peptides listed in Table 1, we also identified the following biologically active peptides in fractions 41–55; Met-enkephalin-Arg-Phe, Met-enkephalin-Arg-Gly-Leu, somatostatin, PH-8P (dynorphin A[1–8]), β-neo-endorphin, substance P[3–11] and their related peptides, such as M(O) forms. These peptides were observed as major or main peaks on the chromatograms (data not shown). In Tables 1 and 2, we selected the species in the database in the order of pig, cow, human, dog, and rat. Thus, the same sequence may be identified in the protein of other species.

3.3. Large scale preparations of peptides in porcine brain extracts

We extracted peptides from large amounts of porcine brain (20 kg) for the source of systematic analysis, because we planned to use the same preparation for extensive analysis as well as for distribution to other researchers. One more important point which is required for the construction of the database is that the methods should be generalized in order to be reproduced when performed by other researchers. For example, we employed a diaflow membrane filter for condensation, after confirming that the PLAC membrane retained well the peptides of M_r more than 500–600. By combining the simple and repeatable procedures, we were able to obtain definite amounts of dried materials in the six different batches compared with the starting tissue weight.

Although we separated peptides into two fractions, SP-II of weakly basic peptides and SP-III of basic peptides, in the small scale study, we combined these two fractions and analyzed in the large scale preparation. We separated the peptides by gel permeation chromatography into two fractions, $M_r \leq 2500$ (fractions 42–55, small peptide fraction) and 2500–6000 (fractions 34–41, large peptide fraction), based on the elution positions of CNP-22 and CNP-53 (Fig. 3). Only the small peptide fraction was analyzed in this study, but we are extending the target of this approach for the large peptide fraction.

3.4. Two-dimensional HPLC of the small peptide fraction

A portion (100 g equivalents) of the pooled small peptide fraction (fractions 42–55) was lyophilized, dissolved in the starting solvent, and separated into 70 fractions by SP-ion-exchange HPLC with a linear gradient elution of ammonium formate (Fig. 4). The major peaks of absorbance at 280 nm were observed at approximately 20 min in the case of the small peptide fraction, since the chromatographic conditions were set for the separation of both small and large peptide fractions. All fractions eluted from SP-ion-exchange HPLC were lyophilized, dissolved in 0.5 ml of the starting solvent, and then each fraction was separated into 75 fractions by reversed-phase HPLC on a C_{18} column with a linear gradient

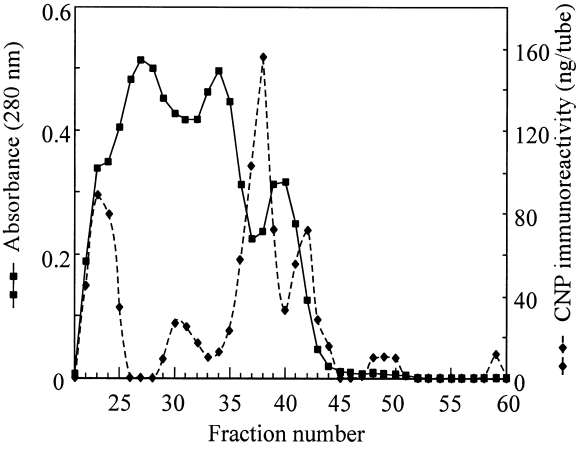

Fig. 3. Sephadex G-50 gel permeation chromatography of the crude peptide fraction of the large scale preparation of porcine brain extracts. Sample: One-twentieth of the porcine brain crude peptide fraction (dry weight: about 1 g, 1 kg equivalent). Column: 5×145 cm. Solvent: 1 *M* acetic acid. Fraction size: 60 ml/tube. Flow-rate: 60 ml/min. V_o and V_t were fractions 23 and 50, respectively.

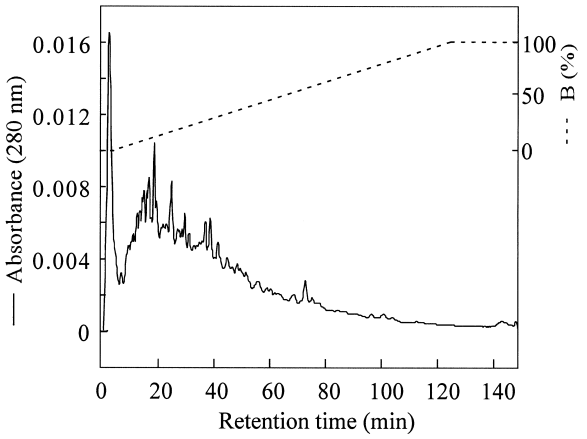

Fig. 4. SP-ion-exchange HPLC of the small peptide fraction of the large scale preparation of porcine brain extracts. Sample: One-twentieth of fractions 42–55 in Sephadex G-50 gel permeation chromatography (50 g equivalents). Column: TSK gel SP-2SW, 4.6×250 mm. Solvent system: linear gradient elution of from (A) to (B) over 60 min. (A) 10 m*M* ammonium formate (pH 3.8):acetonitrile=9:1 (V/V), (B) 1 *M* ammonium formate (pH 3.8):acetonitrile=9:1. Flow-rate: 1 ml/min. Fraction was collected every 2 min (total 70 fractions). This HPLC was performed twice, and column effluents were collected into the same tubes.

elution of acetonitrile (Fig. 5). The 70 chromatograms were serially aligned according to the fraction number of the SP-ion-exchange HPLC, combined as one figure, and expressed as a 2-D diagram using the density graduation for indicating absorbance at 210 nm (Fig. 6). This 2-D chart clearly demonstrates the distribution of peptides on the chromatogram based on their positive charges at pH 3.8 and the degree of hydrophobicity indicated by retention time on the ion-exchange and reversed-phase HPLC.

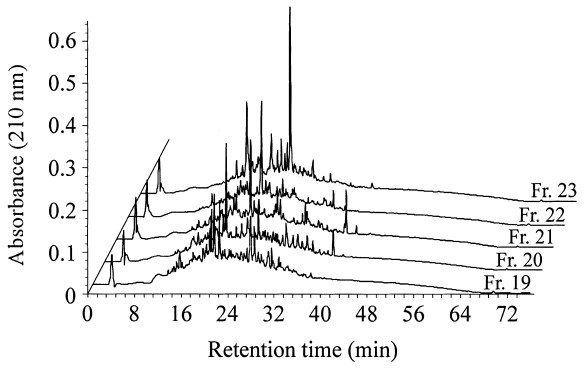

Fig. 5. Reversed-phase HPLC of fractions 18, 19, 20, 21, and 22 in the SP-ion-exchange HPLC. Sample: fractions 18, 19, 20, 21, and 22 (50 g equivalents) in Fig. 4 were each evaporated and injected. Column: Symmetry 300A 5 μm C_{18}, 4.6×250 mm. Solvent system: linear gradient elution of acetonitrile from 10% to 60% in 0.1% TFA over 60 min. Flow-rate: 1 ml/min. Absorbance was measured at 210 nm and 280 nm, and the data for 210 nm is shown in the figure. Fractions were collected every 1 min (total 75 fractions).

By the combination of two different chromatographies, small peptides of porcine brain were separated into 5250 fractions. In the present separation of the small peptide fraction, about half of these fractions did not contain detectable levels of peptides. In the case of the large peptide fraction, we confirmed that peptides were widely distributed on the 2-D diagram (data not shown).

3.5. MS analysis of fractions in 2-D HPLC

We have preliminarily analyzed aliquots of all 5250 fractions by MALDI-TOF MS. By rough evaluation, about 5000 peptides were detected by this analysis, although the same peptide may have been detected in more than two fractions (data not shown). Fraction 20 in SP-ion-exchange HPLC was analyzed in detail with an MALDI-TOF mass spectrometer, and the results were represented by a 2-D diagram of molecular masses and retention times on reversed-phase HPLC in Fig. 7. Ion counts observed in the MS analysis were shown as the degree of density graduation, and peptides in the porcine brain were visualized on the 2-D diagram in a manner similar to the 2-D gel profile of proteomic analysis. Quantitative measurements of peptide contents is well known to be difficult by MS analysis. However, a rough estimation of the peptide contents could be performed based on the ion counts, even though ionization efficiency, stability and other properties are different for each peptide. By analyzing all 70 fractions with the same procedures and accumulating the results, porcine brain peptides of $M_r \leq 2500$ Da will be detected and visualized in the virtual 2-D or 3-D diagram.

3.6. Data analysis of peptides identified in the present study

Table 2 summarized the peptide data so far obtained for fraction 20 in SP-ion-exchange HPLC,

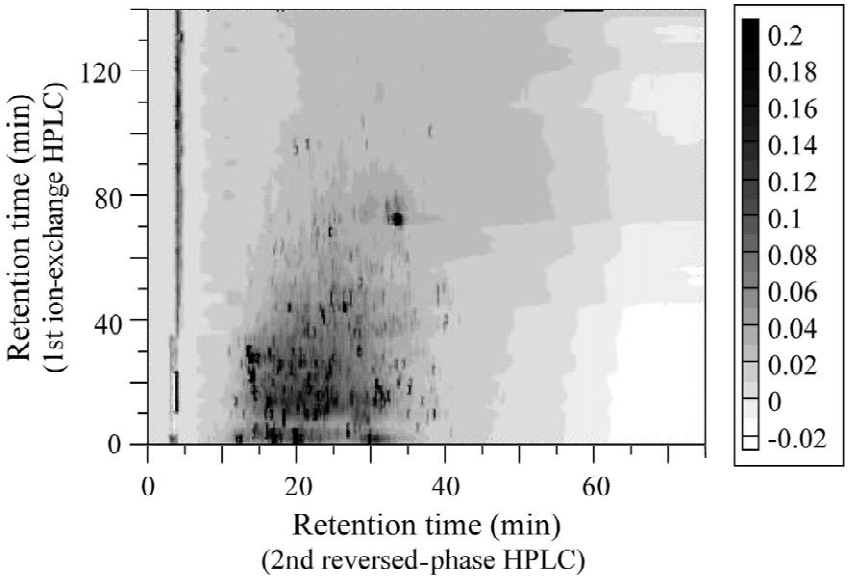

Fig. 6. Two-dimensional representation of absorbance at 210 nm of the small peptide fraction of porcine brain extracts. Absorbance at 210 nm observed in Fig. 5 (2nd reversed-phase HPLC) were serially aligned according to the fraction number (1–70) in Fig. 4 (1st ion-exchange HPLC) and combined. Intensity of absorbance was expressed as the degree of density shown in the contour legend.

which was separated by reversed-phase HPLC and then analyzed by mass spectrometers. In this table, the amino acid sequences of 28 peptides that were identified by de novo sequencing of the MS/MS analysis and confirmed with a protein sequencer are listed. Three amino acid sequences preceding and succeeding to the identified peptides as well as marks flanking the identified peptides and the precursors are shown in the Table, even if the precursor sequences have not been in the pig. In fraction 20 of the ion-exchange HPLC, biologically active peptides have not been identified, but two fragments of cholecystokinin precursor were identified, which suggested the presence of a biologically active unit in another fraction in the 2-D HPLC. As a peptide that has been registered in the database, cerebellin was identified in fraction 18. As a fragment of known peptides, a 12-residue fragment of thymosin β4 was detected in fraction 22. In addition to these peptides, fragments of chromogranin B and serotonin receptor were identified in fractions 25 and 27 with the protein sequencer (data not shown).

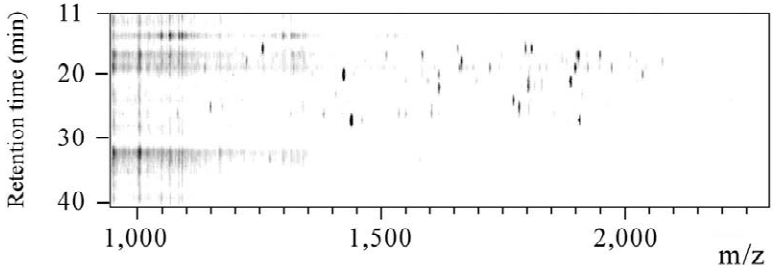

Fig. 7. Two-dimensional representation of reversed-phase HPLC and MS analysis data. Fraction 20 of SP-ion-exchange HPLC was separated by reversed-phase HPLC and analyzed by an MALDI-TOF mass spectrometer, and the molecular mass and retention time on reversed-phase HPLC were plotted in a 2-D diagram. Ion counts observed in the MS analysis are shown as the density graduation in this figure.

In Table 2, most of them are fragments of major cellular proteins required for cell organization and metabolism. Fragments of myelin basic protein (three fragments), glyceraldehyde 3-phosphate dehydrogenase (two fragments), neurofilament triplet M protein (one fragment), dihydropyrimidinase related protein-2 (two fragments), and tubulin β chain (one fragment) were detected in addition to hemoglobin fragments (three fragments). On the other hand, fragments of enzymes and channels that were assumed to be less abundant, such as protein-tyrosine phosphatase, ser/thr protein phosphatase, and voltage-dependent anion selective channel, were identified in the present analysis. One fragment of neurogranin was identified in fractions 21 and 23, which may be generated by oxidation of methionine between the first and second HPLCs.

The amino acid sequence of one peptide observed in fraction 17 was unambiguously determined by the MS/MS analysis and by the Edman degradation, but its precursor protein was not identified by the database searching. This is probably due to the fact that the number of porcine proteins, cDNAs and ESTs registered in the database is far less than that of human and mouse [24], as easily deduced from species names shown in Tables 1 and 2, where more than half of the sequences were identified in other species.

Among the peptides listed in Table 2, five and three peptides were derived from N- or C-termini of the parent proteins, respectively. These peptides are probably generated by proteolytic cleavage during the extraction process. On the other hand, one end of two fragments was deduced to be produced by acid hydrolysis of the Asp-Pro bond. In seven other cases, the N-terminal side of the cleavage site is Asp. Since Asp-X bond is relatively labile compared with other peptide bonds under the acidic conditions [25,26], it is necessary to examine the hydrolysis of the peptide bond during separation and storage.

In Table 3, molecular masses and ion counts of peptides that were detected with a MALDI-TOF mass spectrometer but could not be identified by the MS/MS analysis were listed. In addition to 28 peptides listed in Table 2, we were able to detected 184 peptides in fraction 20 in SP-ion-exchange HPLC using 50 mg equivalents of brain tissue extracts. Based on the absorbance at 280 nm observed in the SP-ion-exchange HPLC (Fig. 4), fraction 20 corresponded to about 4.0% of the total peptides of $M_r \leq 2500$. This result suggests that about 5300 peptides and 700 peptides could be detected and identified, if we extend the MALDI-TOF MS analysis as well as the MS/MS analysis to all fractions in the SP-ion-exchange HPLC on the assumption that peptides can be detected or identified in correlation with the dry weight of the fraction. Since the dry weight of the small peptide fraction of $M_r \leq 2500$ was less than 25% of the total peptide fraction of $M_r \leq 6000$, about 20 000 peptides and 2500 peptides could be detected and identified, if we further extend our assumption.

4. Discussion

Although the fact data for endogenous proteins have been accumulated by proteomic analysis [27], no database for endogenous peptides is available at present. The main reason why the peptide fact database has not been undertaken is that the majority of peptides in the tissue extracts prepared by conventional procedures are degradation products of proteins. As the cellular metabolism, a small portion of proteins in the tissue or cells are continuously degraded depending on their stability and effective duration. The protein content was shown to be 1000-fold greater than that of the peptide in this study. Even if we could freeze the metabolism of proteins, substantial amounts of peptides derived from the degradation of proteins are present in addition to a trace amount of peptides specifically processed from their precursor proteins. Furthermore, relatively large amounts of peptides are artificially generated during homogenization and extraction, and the endogenous peptides are degraded at the same time. As shown in the former half of this study, however, we demonstrated that peptides endogenously present in the tissue or cells can be detected by minimizing the artificial degradation of proteins.

We have isolated a series of biologically active peptides, including natriuretic peptides (BNP and CNP), neuromedin B, C, K, L, N and U, α- and β-neo-endorphin [21,22,28–32]. Fortunately, these peptides were later shown to be the molecules that are endogenously present in the tissue. Furthermore,

Table 3
Masses and ion counts of peptides in the reversed-phase HPLC of fraction 20 in SP-ion-exchange HPLC of the small peptide fraction prepared from porcine brain extracts (large scale preparation)

Fr. no.	Mass	Ion count	Fr. no.	Mass	Ion count	Fr. no.	Mass	Ion count	Fr. no.	Mass	Ion count	Fr. no.	Mass	Ion count	Fr. no.	Mass	Ion count
16	1152.69	1184	21	934.62	1550	22	1202.81	5623	24	1394.88	13000	25	1366.87	1225	27	1295.72	4619
	1170.68	1419		998.60	2959		1429.93	2944		1403.89	3768		1381.81	1139		1326.73	4342
	1243.62	1455		1096.69	8494		1466.98	2035		1418.81	3100		1394.87	1894		1341.79	5869
	1296.73	2281		1112.70	1769		1707.02	2112		1430.90	5034		1403.87	1017		1355.81	2363
17	877.08	3374		1153.82	7332		1801.19	5157		1436.90	2849		1409.88	1083		1380.59	1780
	1279.75	1984		1164.72	2679		1817.17	1785		1452.87	2099		1423.84	1311		1419.76	6204
	1314.82	1646		1216.81	1318		1885.20	1714		1493.96	3000		1436.87	3026		1429.84	1702
	1396.92	1235		1221.72	1905	23	1202.80	3042		1507.89	2782		1463.86	1009		1441.76	1299
	1469.95	1403		1256.35	1674		1305.20	2534		1583.94	3745		1482.90	1066		1457.73	1166
	2029.23	1513		1381.86	4467		1350.92	1809		1595.00	2724		1487.98	2185		1476.85	1472
18	1134.71	4329		1429.95	1341		1365.85	6681		1607.95	2726		1496.87	2204		1496.84	1686
	1156.80	2627		1472.96	1201		1452.92	1664		1632.02	2205		1512.88	2432		1566.89	2484
	1349.88	2070		1494.91	1378		1489.95	1526		1646.05	2766		1520.86	1354		1653.91	1534
	1426.88	3439		1512.91	2258		1524.00	2066		1666.97	3389		1527.95	1231		1685.97	2565
	1873.19	1372		1560.02	6996		1583.99	7854		1719.03	2588		1566.97	10000		1749.96	1337
19	976.63	1511		1566.97	2610		1612.00	2158		1750.08	17000		1589.96	2739		1759.07	1394
	1028.71	2041		1573.95	2294		1824.23	2394		1772.05	3111		1602.01	1069		1763.98	1527
	1112.69	3733		1583.01	2130		1852.30	1638		1792.08	7069		1606.98	2507		1776.92	1201
	1264.75	1633		1599.01	1085		1913.14	1789		1798.06	6667		1638.03	1541		1784.92	2782
	1281.86	3717		1633.07	1553		2556.58	1104		1824.19	3423		1666.02	1061		1792.91	1889
	1573.92	1721		1667.04	3756	24	988.57	13000		1852.24	2124		1719.04	3730		1798.93	1270
	1645.11	1556		1674.07	2158		1125.64	7169		1881.20	2817		1729.98	1400		1865.01	2259
20	893.10	3726		1682.04	1313		1165.47	4114		1886.07	2332		1734.07	1160		1875.08	3494
	1025.65	5104		1801.19	3514		1182.68	12000		1914.15	3259		1746.10	3328		1887.02	3844
	1041.65	3201		1817.19	2541		1190.64	2836		1917.17	2743		1750.07	5724		1905.00	21000
	1153.73	6922		1824.23	2036		1210.88	2614		2286.39	2700		1792.06	1549		1926.06	1544
	1321.32	5453		2029.30	1134		1253.73	4477	25	1182.66	1134		1806.10	1708		1974.04	1453
	1381.85	24000		2061.35	1137		1273.84	2647		1221.81	1576		1955.14	1812		2047.16	1824
	1583.00	7811	22	1096.68	5476		1302.80	5015		1253.75	1061	26	1064.00	1492		2067.09	2118
	1675.07	1580		1110.69	2642		1367.85	7104		1276.72	1183		1124.00	1765			
21	913.56	1926		1176.82	2632		1378.62	2609		1300.87	6763	27	952.99	2484			

Mass indicates $(M+H)^+$, and Fr. no. is the abbreviation of fraction number. Ion count was measured by a MALDI-TOF mass spectrometer. Each fraction was also submitted to the MS/MS analysis as shown in Table 2, but no definite sequence information was obtained for the peptides listed in this Table.

in the isolation of neuromedin N, two other peptides, neurotensin and neurokinin A were purified together. Three peaks of biological activity were observed on the chromatogram, which corresponded to the absorbance peaks of these peptides (see Fig. 2 of Ref. [28]). These findings suggest that peptides present in the controlled extracts are not a simple mixture of degradation products but reflect the endogenous peptides to some extent. To further confirm these previous findings, we compared water and acetic acid extraction methods with or without heat treatment (data not shown). The extraction under the 1 *M* acetic acid condition showed a higher extraction efficiency of peptides and proteins than that of the water extraction. However, the peptide fraction obtained from the acetic acid extraction without heat treatment gave a higher peptide content and a higher baseline on the reversed-phase HPLC, indicating the degradation of proteins. The heat treatment itself reduced the amounts of peptides and proteins extracted from the tissue, but addition of acetic acid increased the recovery of peptides and proteins. In

the case of brain tissue, several different conditions for the heat treatment were examined, and 10 min heat treatment at more than 90 °C was confirmed to be sufficient to inactivate intrinsic proteolytic enzymes based on the measurement of several peptide hormones such as CNP (data not shown). Thus, the extracts prepared by the heat treatment followed by acidification with acetic acid and homogenization were used for the present analysis.

From the crude extracts thus obtained, a basic peptide fraction was efficiently condensed by the acetone precipitation, batch-wise chromatographies on reversed-phase and SP-ion-exchange columns. The basic peptide fraction was subjected to gel permeation chromatography to remove proteins and roughly separate peptides. As shown in Fig. 1 and Table 1, biologically active peptides were clearly observed and identified after separation with reversed-phase HPLC. Although most of the peptides detected in this analysis are deduced to be degradation products of proteins, these facts mentioned above suggest that some of them are endogenously present as degradation products of proteins.

In addition to the efficient preparation method of the peptide fraction developed in this study, we are now able to detect and analyze peptides by using a tandem mass spectrometer with high sensitivity and resolution. It is attractive that amino acid sequences of peptides can be determined or deduced by the MS/MS analysis even though peptides were not purified in a homogenous state. Taking account of our experiences and previous findings reporting isolation of new peptides [33–36], we set up a 2-D chromatography system composed of ion-exchange HPLC and C_{18} reversed-phase HPLC for efficient and reproducible separation of the whole peptide fraction. By combining these systems, the fact data for endogenous peptides in the tissue and cells could be collected.

In this study, we only analyzed in detail peptides in fraction 20 eluted between 38 and 40 min in the SP-ion-exchange HPLC of the large scale preparation. By reversed-phase HPLC and MS analysis, we detected 212 peptides in fraction 20. Among these 212 peptides, 28 peptides have been identified by de novo sequencing of the MS/MS analysis and confirmed by the Edman degradation, as shown in Table 2. The rest of the peptides were detected with the MALDI-TOF mass spectrometer, but could not be identified by the MS/MS analysis. For the de novo sequencing, we used about 2 g equivalents of brain extracts, while only 50 mg equivalents of the extracts were consumed for the detection of 212 peptides. It may be speculative but about 20 000 peptides and 2500 peptides could be detected and identified, if we extend our deduction up to the peptides of $M_r \leq 6000$ based on the analysis data of 1.4% of the peptide fractions of $M_r \leq 2500$. As we used pig brain as a source of endogenous peptides, 50% of the peptide sequences listed in Table 2 were not found in the databases of the nucleotide and amino acid sequences for the pig. If we used human or mouse tissue as a target, more peptide sequences could be identified by combining the MS/MS analysis and the database searching starting from the same amount of the tissue extracts. This is due to the lower number of registered nucleotide and amino acid sequences for the pig than that for the human or for the mouse [1,24,37].

The number of peptides so far registered in the database for human, rat, pig and cow are only 200–400 in each species (personal communication, M. Isoyama, Protein Research Foundation, Osaka, Japan). Even though most peptides detected by the present procedures are degradation products of proteins, accumulation of data and information of the endogenous peptides in the fact database, designated ‘‘Peptidome’’, will provide a useful intellectual infrastructure for the peptide research.

As for proteomic analysis and profiling, 2-D gel electrophoresis is used for the common platform for data accumulation and comparison. In the case of peptides, there was neither a database nor a common platform. The findings of the present study suggest that 2-D HPLC composed of an ion-exchange HPLC and reversed-phase HPLC can be used as a common platform comparable to the 2-D gel electrophoresis in the proteomic analysis. Since the chromatographies need more detailed settings for columns, buffers and elution conditions, the conditions for 2D-HPLC analysis should be standardized and the data should be normalized by using standard peptides. On the other hand, the physicochemical properties and molecular masses of peptides obtained

during the separation and analysis process can also be used as parameters in the Peptidome database. In this database, therefore, the peptide information can be visualized in the virtual 3-D space (net positive charge, hydrophobicity and molecular mass) and searched with these parameters in addition to the peptide name and sequence.

Once the common platform is set up, the peptide fact data can be accumulated in the Peptidome database. This database is expected to be utilized in many research fields. For example, this database will be a target for searching for unidentified biologically active peptides and peptidergic ligands for orphan receptors. Since we have information for endogenous biologically active peptides in the database, the Peptidome database will certainly help excluding the known peptides and identifying new peptides. Peptide differential display using a common platform and this database will provide a new approach for identifying peptides related to physiological regulations and diseases, which in turn will be applied to the etiology and diagnostic methods. Since the Peptidome database has a potential to be applied to many other research purposes, we would like many researchers to participate in not only utilizing the database but also registering the peptide information to this database.

In the present study, we noticed several problems. (1) Proteolytic degradation occurred during the extraction and purification process, although much less than the conventional procedures. (2) Acid hydrolysis of labile peptide bonds such as Asp-Pro was often observed. (3) Oxidation of methionine was observed at a relatively high frequency. (4) Contamination of hemoglobin and its fragments was observed especially in the large scale preparations. The first problem may be derived from the large scale extraction, because it took a longer time for extraction and other treatments. It could be minimized by quick and small scale tissue extraction. The second point was observed at a frequency greater than we deduced. About 5.4% of the cleavage of peptide bonds listed in Table 2 were generated by acid hydrolysis of Asp-Pro bond during storage and separation. We have to examine which peptide bonds are weak for acid hydrolysis under the conditions such as 1 *M* acetic acid. The third problem is not serious, but methionine oxidation makes the analysis more complex. Anti-oxidation scavengers may be useful as long as they do not interfere with the following separation and analysis. The extraction and separation should be performed as quickly as possible, and peptides should be stored in the dry conditions without oxygen. As for the fourth problem, it is practically impossible to prevent contamination of fragments of blood- and vasculature-derived proteins when extraction is started from animal tissue. Perfusion of tissue soon after killing minimizes the contamination of blood proteins, but blood vessels and other vascular tissues can not be excluded.

Schulz-Knappe et al. and Seiler et al. recently proposed and started the ‘‘peptide bank’’ project for screening for and isolation of biologically active peptides [38,39]. They tried to develop this bank for human plasma and porcine brain, but no endogenous form of biologically active peptide has been identified. This idea and project were further reinforced and advanced by the BioVisioN group, who intended to identify peptides and small proteins of $M_r \leq 20\,000$ especially in the biological fluid, including plasma, urine, cerebrospinal fluid, tears and synovial fluid [40–42]. In the extensive analysis, however, a very limited number of biologically active peptides has been identified in human and mammalian tissues and plasma even by making the most use of mass spectrometers [43]. In insects, such as the fruit fly and locust, as well as in the venom of spiders and snakes, however, several groups reported the efficient identification of biologically active or toxic peptides [44–48]. These differences are suggested to be derived from the differences in the tissue concentrations, storage and extraction conditions, and proteolytic activity in the cells, tissues, plasma and so on. As peptides are easily degraded in a short period, the most suitable conditions must be found for each target tissue or cells.

We started to construct the Peptidome database for porcine brain peptides. Although porcine brain tissue is abundant and freely available, the analysis of porcine peptides is not favorable for database searching. Since the number of registered genome, cDNA and protein data for humans and mice is far more than that for pigs [24], it is reasonable to collect the fact data of peptides from human and mouse tissue as a next target of the Peptidome database.

Although the collected peptide data for the Pep-

tidome database is very limited, these data have clarified the possible construction of a peptide database by minimizing degradation and identifying peptides by using 2-D HPLC coupled with the MS and MS/MS analyses. The accumulation of the fact data of peptides will provide a new and solid intellectual infrastructure for developing drugs, diagnostic and therapeutic methods as well as for elucidating the peptide function in the biological systems.

5. Conclusion

The present study provides the basement and strategy for constructing the fact database for endogenous peptides, which has not yet been constructed, by minimizing proteolytic degradation and using 2-D HPLC coupled with MS analysis. More refinement of the procedures are required for Peptidome, but integration of the fact data for endogenous peptides and the knowledge stored in the databases on the common platform provided by the 2-D HPLC will facilitate the accumulation of peptide information and advance peptide research.

Acknowledgements

This study was supported by the Special Coordination Fund for the Promotion of Science and Technology (Intellectual Infrastructure Research and Development for Its Comprehensive Promotion System) from the Ministry of Education, Culture, Sports, Science and Technology of Japan. The authors are grateful for Drs K. Kangawa and H. Matsuo of this Institute, Dr M. Isoyama of the Protein Research Foundation for valuable discussion and suggestions. The authors also thank Ms Y. Matsui, M. Nakatani and M. Asaji of this Institute for their technical assistance.

References

[1] International Human Genome Sequencing Consortium, Nature 409 (2001) 860.
[2] J.C. Ventor, M.D. Adam, E.W. Myers et al., Science 291 (2001) 1304.
[3] M.R. Wilkins, K.L. Williams, R.D. Appel, D.F. Hochstrasser (Eds.), Proteome Research: New Frontiers in Functional Genomics, Springer, Berlin, 1997.
[4] A. Pandey, M. Mann, Nature 405 (2000) 837.
[5] A.C. Gavin, M. Bosche, R. Krause et al., Nature 415 (2002) 141.
[6] F.L. Strand, Neuropeptides: Regulation of Physiological Processes, MIT Press, Boston, 1999.
[7] C. Sandman, F. Strand, B. Beckwith, B. Chronwall, W. Flynn, R.J. Nachman (Eds.), Neuropeptides: Structure and Function in Biology and Behavior, New York Academy of Sciences, New York, 1999.
[8] C.H.V. Hoyle, Neuropeptides: Essential Data, Wiley, Chichester, 1996.
[9] S. Takeda, S. Kadowaki, T. Haga, H. Takaesu, S. Mitaku, FEBS Lett. 520 (2002) 97.
[10] GPCRDB: Information system for G protein-coupled receptors (GPCRs), http://www.gpcr.org/7tm/.
[11] J.M. Standel, S. Wilson, D.J. Bergsma, Trends Pharmacol. Sci. 18 (1997) 430.
[12] A.D. Howard, G. McAllister, S.D. Feighner, Q. Liu, R.P. Nargund, L.H. Van der Ploeg, A.A. Patchett, Trends Pharmacol. Sci. 22 (2001) 132.
[13] J. Bockaert, S. Claeysen, C. Becamel, S. Pinloche, A. Dumuis, Int. Rev. Cytol. 212 (2002) 63.
[14] K. Nakayama, Biochem. J. 327 (1997) 625.
[15] N.G. Seidah, R. Day, M. Marcinkiewicz, M. Chretien, Ann. N.Y. Acad. Sci. 839 (1998) 9.
[16] S. Solomon, Ann. N.Y. Acad. Sci. 885 (1999) 22.
[17] R.B. Harris, Arch. Biochem. Biophys. 275 (1989) 315.
[18] B.A. Eipper, D.A. Stoffers, R.E. Mains, Annu. Rev. Neurosci. 15 (1992) 57.
[19] M. Kojima, H. Hosoda, Y. Date, M. Nakazato, H. Matsuo, K. Kangawa, Nature 402 (1999) 656.
[20] S. Ueda, N. Minamino, M. Aburaya, K. Kangawa, S. Matsukura, H. Matsuo, Biochem. Biophys. Res. Commun. 175 (1991) 759.
[21] T. Sudoh, N. Minamino, K. Kangawa, H. Matsuo, Biochem. Biophys. Res. Commun. 168 (1990) 863.
[22] N. Minamino, K. Kangawa, H. Matsuo, Biochem. Biophys. Res. Commun. 170 (1990) 973.
[23] J.F. Cossio, J. Gonzalez, L. Betancourt, V. Besada, G. Padron, Y. Shimonishi, T. Takao, Rapid. Commun. Mass Spectrom. 12 (1998) 1867.
[24] Top 1000 organisms (1–100), http://www.ddbj.nig.ac.jp/ddbjnew/org1000/top100-e.html
[25] J. Schultz, H. Allison, M. Grice, Biochemistry 1 (1962) 694.
[26] T.F. Spande, B. Witkop, Y. Degani, A. Patchornik, Adv. Protein Chem. 24 (1970) 97.
[27] P. Gromov, M. Ostergaard, I. Gromova, J. Celis, Prog. Biophys. Mol. Biol. 80 (2002) 3.
[28] N. Minamino, K. Kangawa, H. Matsuo, Biochem. Biophys. Res. Commun. 122 (1984) 542.
[29] N. Minamino, T. Sudoh, K. Kangawa, H. Matsuo, Peptides 6 (Suppl. 3) (1985) 245.

[30] T. Sudoh, K. Kangawa, N. Minamino, H. Matsuo, Nature 332 (1988) 78.
[31] N. Minamino, K. Kangawa, N. Chino, S. Sakakibara, H. Matsuo, Biochem. Biophys. Res. Commun. 99 (1981) 864.
[32] K. Kangawa, N. Minamino, N. Chino, S. Sakakibara, H. Matsuo, Biochem. Biophys. Res. Commun. 99 (1981) 871.
[33] J.M. Vaughan, W.H. Fischer, C. Hoeger, J. Rivier, W. Vale, Endocrinology 125 (1989) 1660.
[34] T. Sakurai, A. Amemiya, M. Ishii, I. Matsuzaki, R.M. Chemelli, H. Tanaka, S.C. Williams, J.A. Richardson, G.P. Kozlowski, S. Wilson, J.R. Arch, R.E. Buckingham, A.C. Haynes, S.A. Carr, R.S. Annan, D.E. McNulty, W.S. Liu, J.A. Terrett, N.A. Elshourbagy, D.J. Bergsma, M. Yanagisawa, Cell 92 (1998) 573.
[35] S. Hinuma, Y. Habata, R. Fujii, Y. Kawamata, M. Hosoya, S. Fukusumi, C. Kitada, Y. Masuo, T. Asano, H. Matsumoto, M. Sekiguchi, T. Kurokawa, O. Nishimura, H. Onda, M. Fujino, Nature 393 (1998) 272.
[36] Y. Saito, H.P. Nothacker, Z. Wang, S.H. Lin, F. Leslie, O. Civelli, Nature 400 (1999) 265.
[37] Mouse Genome Sequencing Consortium, Nature 420 (2002) 520.
[38] P. Schulz-Knappe, M. Schrader, L. Ständker, R. Richter, R. Hess, M. Jürges, W.G. Forssmann, J. Chromatogr. A 776 (1997) 125.
[39] P. Seiler, L. Ständker, S. Mark, W. Hahn, W.G. Forssmann, M. Meyer, J. Chromatogr. A 852 (1999) 273.
[40] M. Schrader, P. Schulz-Knappe, Trends Biotechnol. 19 (2001) S55.
[41] P. Schulz-Knappe, H.D. Zucht, G. Heine, M. Jürges, R. Hess, M. Schrader, Comb. Chem. High Throughput Screen. 4 (2002) 201.
[42] M. Jürgens, M. Schrader, Cur. Opin. Mol. Ther. 4 (2002) 236.
[43] J. Gobom, K.O. Kraeuter, R. Persson, H. Steen, P. Roepstorff, R. Ekman, Anal. Chem. 72 (2000) 3320.
[44] S. Uttenweiler-Joseph, M. Moiatte, M. Lagueux, A. Van Dorsselae, J.A. Hofmann, P. Bulet, Proc. Natl. Acad. Sci. USA 95 (1998) 11342.
[45] E. Clynen, G. Baggerman, D. Veelaert, A. Cerstiaens, D. Van der Horst, L. Harthoorn, R. Derua, E. Waelkens, A. De Loof, L. Schoofs, Eur. J. Biochem. 268 (2001) 1929.
[46] G. Baggerman, A. Cerstiaens, A. De Loof, L. Schoofs, J. Biol. Chem. 277 (2002) 40368.
[47] P. Escoubas, G. Corzo, B.J. Whiteley, M.L. Celerier, T. Nakajima, Rapid. Commun. Mass Spectrom. 16 (2002) 403.
[48] K. Dyason, W. Brandt, L. Prendini, F. Verdonck, J. Tytgat, J. du Plessis, G. Muller, J. van der Walt, Rapid. Commun. Mass Spectrom. 16 (2002) 768.