proteasome were also separated using the same method (Fig. 2). The 20S proteasome was well focused by IEF, and proteins that did not interact with the 20S proteasome were well separated. These results suggest that a protein complex of approximately 700 kDa could be focused on keeping the interaction among subunits in agarose gel by native IEF. The usefulness of additives such as sorbitol and threhalose for improving the stability of the agarose gel and protein complex under IEF was investigated. No differences were observed with these additives versus without them. We concluded that the method presented here is suitable for the separation and analysis of protein complexes from cells.

Native 2D–PAGE analysis of fractions containing high-molecular-weight protein complexes separated by ultracentrifugation, gel chromatography, and ion exchange chromatography

We next analyzed high-molecular-weight fractions from yeast cells. Yeast high-molecular-weight proteins were precipitated by ultracentrifugation after the removal of microsomes and organelles. This fraction contained proteasomes, which are multisubunit complexes with a molecular weight of approximately 700 kDa. The sample was further separated by gel chromatography using a Bio-Gel A 1.5-m column to isolate high-molecular-weight complexes, including proteasomes. A typical chromatogram is shown in Fig. 3. We selected the fraction that eluted just after the fraction containing proteasome activity (Fig. 3). This fraction was further separated into six fractions by ion exchange chromatography, and each of the six fractions was then separated by ultracentrifugation through a density gradient to isolate high-molecular-weight complexes. Fraction 6 from the ion exchange column contained proteasome activity, and this activity was isolated in fraction 15 from the subsequent density gradient (designated as 6–15). This fraction contained mainly 20S proteasome. We did not analyze it further. The protein amount of each fraction of ultracentrifuge was analyzed semiquantitatively by the Bradford method. A protein peak was observed only for the fractions selected in Fig. 4 except for the fractions with low-molecular-weight proteins.

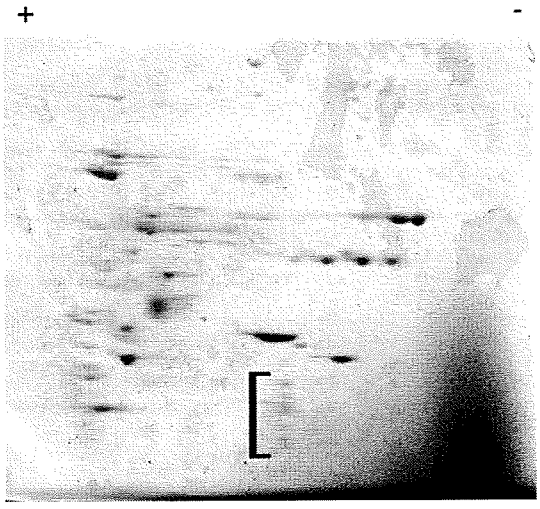


Fig. 2. Crude proteasome fraction from yeast extract prepared by ammonium sulfate precipitation, gel chromatography, and ion exchange chromatography [17]. Native IEF was performed in an agarose tube gel without adding denaturing reagents. The agarose solution was prepared by mixing 30 mg of agarose IEF and 1.05 g of threhalose in 1.95 ml of distilled water and then heating until dissolved. After cooling, 150 μl of carrier ampholite (ampholine pH 3.5–10.0) was added and mixed. The agarose tube gel was prepared as described in Materials and Methods, and electrophoresis was performed by applying voltage from 100 to 400 V in 100-V steps. The crude proteasome fraction (34 μg) contained many proteins in addition to the proteasome subunits. These proteins were separated from the 20S proteasome complex by native IEF.

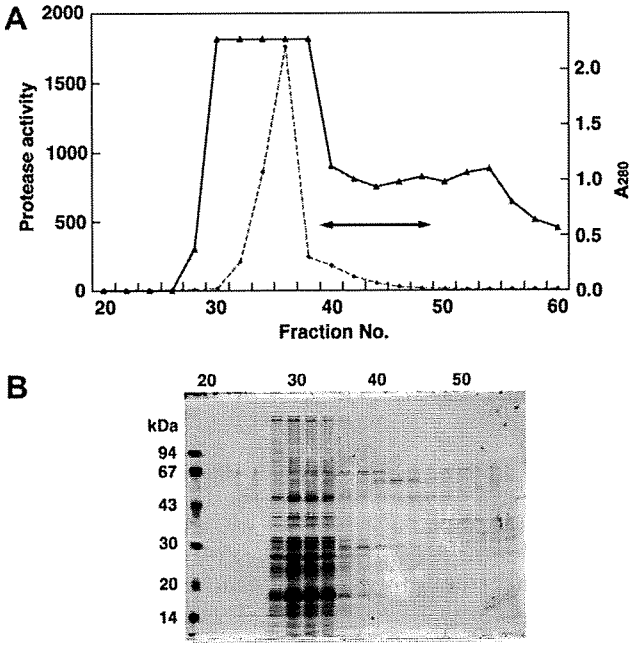


Fig. 3. Separation of the high-molecular-weight protein fraction from yeast by gel chromatography. (A) Protein precipitated by ultracentrifugation was separated by gel chromatography using a Bio-Gel A 1.5-m column (2.5 [i.d.] × 45 cm) at a flow rate of 0.25 ml/min, collecting 2.5-ml fractions. The fractions indicated by the arrows were pooled. The absorbance at a wavelength of 280 nm (solid line) and the chymotryptic activity of the proteasomes (dashed line) are shown. (B) SDS–PAGE of the fractions separated by gel chromatography.

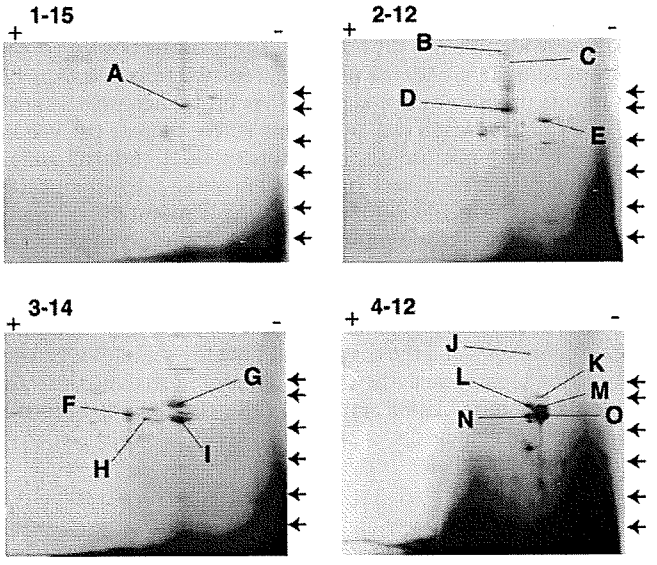


Fig. 4. Separation of the protein complexes in the fractions obtained by density gradient separation. The proteins in the pooled fractions from gel chromatography were separated by anion exchange chromatography. Protein complexes were recovered in six fractions and were then further separated by density gradient. The proteins in each fraction were analyzed by 2D–PAGE with IEF in native agarose gels. Native IEF was performed in an agarose tube gel without adding denaturing reagents. The agarose gel was composed of 1% agarose IEF, 12% sorbitol, 3% ampholine (pH 3.5–10.0), and 4% ampholine (pH 5.0–8.0). IEF separation was performed as described in Materials and Methods. The polyacrylamide gel was stained after PAGE in the second dimension. Fraction notation: the first number refers to the ion exchange fraction, and the second number refers to the density gradient fraction number (e.g., 1–15 is density gradient fraction 15 separated from ion exchange fraction 1). Native IEF analysis for 4–12 was performed in agarose gels (1% agarose IEF, 12% sorbitol, and 10% ampholine [pH 3.5–10.0]). Two spots, N and O, in 4–12 are positioned close together compared with the identical protein pair, H and I, in 3–14 (see spots A–O in Table 1). Arrows indicate the positions of the molecular weight markers (from the upper band, 94, 67, 43, 30, 20, and 14 kDa).

Table 1
Proteins identified in high-molecular-weight fractions from yeast.

Spot(s)	Gene name	Protein description	MW(kDa)	pI	Complex size (kDa)	Reference
A	YHR113W	Aspartyl aminopeptidase	54,168	6.79	680	[18]
B, C	*CTT1*/YGR088W	Cytosolic catalase	65,602	6.41	260	[21]
D, G	*ICL1*/YER065C	Isocitrate lyase	62,266	6.19	250	[22]
E	*ADE2*/YOR128C	Phosphoribosylaminoimidazole carboxylase	62,337	7.19	?	
F	*LAP4*/YKL103C	Aminopeptidase I	1680	5.91	620	[23]
H, N	*GDH1*/YOR375C	Glutamate dehydrogenase	49,461	5.37	300	[24]
I, O	*UGP1*/YKL035W	UDP-glucose pyrophosphorylase	55,854	7.26	450	[25]
J	*GDB1*/YPR184W	Glycogen debranching enzyme	174,845	5.61	170[a]	
L	*PUT2*/YHR037W	δ-1-Pyrroline-5-carboxylate dehydrogenase	61,157	5.69	370[b]	[26]
M	*GAD1*/YMR250W	Glutamate decarboxylase	65,989	6.61	?	

?, No report for complex formation and/or complex size.
[a] Monomeric.
[b] *Thermus thermophilus.*

Fractions 1 through 10 from the density gradient contained lower molecular weight proteins (<200 kDa) such as glyceraldehyde-3-phosphate dehydrogenase (Tdh3, tetramer, 114 kDa) and arginosuccinate synthase (Arg1, tetramer, 188 kDa). We analyzed fractions 12 through 15 from the density gradient of each ion exchange fraction using 2D-PAGE with native IEF (Fig. 4). Several protein spots separated by native IEF-PAGE were identified by mass fingerprinting using MALDI-TOF MS. We identified yeast aspartyl aminopeptidase (YHR113 W) from fraction 1–15. We had confirmed that yeast aspartyl aminopeptidase was a 680-kDa complex that had been prepared using a similar procedure [18]. The fractions of sucrose density gradient (12–15) analyzed by native IEF contained only high-molecular-weight protein complexes. As shown in Table 1, the proteins identified have been reported to form high-molecular-weight complexes. The protein complexes identified in the fractions were homo-multimeric complexes only. No hetero-multimeric complexes were identified in the fractions. The focusing position of each spot was consistent with the pI value predicted from the sequence. The results presented here suggest that protein complexes up to 700 kDa can be focused well by IEF in a native agarose gel. We also tried to separate microsomes using this method. Unfortunately, it did not work well because a large portion of the microsomes stuck on the top of the agarose gel. This method is useful for separating molecules and particles within a limited size range compared with the size range for free flow fractionation [19], ultracentrifugation, size exclusion chromatography, and free flow electrophoresis [20]. However, its advantages include the ability to focus and concentrate protein complexes and the ease of handling the samples when performing the separation in the second dimension, making this a useful method for the isolation and characterization of high-molecular-weight protein complexes.

References

[1] B. Alberts, The cell as a collection of protein machines: preparing the next generation of molecular biologists, Cell 92 (1998) 291–294.
[2] T.W. Nilsen, The spliceosome: the most complex macromolecular machine in the cell?, Bioessays 25 (2003) 1147–1149
[3] G. Brosch, A. Lusser, M. Goralik-Schramel, P. Loidl, Purification and characterization of a high molecular weight histone deacetylase complex (HD2) of maize embryos, Biochemistry 35 (1996) 15907–15914.
[4] C.G. Huber, W. Walcher, A.M. Timperio, S. Troiani, A. Porceddu, L. Zolla, Multidimensional proteomic analysis of photosynthetic membrane proteins by liquid extraction–ultracentrifugation–liquid chromatography–mass spectrometry, Proteomics 4 (2004) 3909–3920.
[5] O. Puig, F. Caspary, G. Rigaut, B. Rutz, E. Bouveret, E. Bragado-Nilsson, M. Wilm, B. Seraphin, The tandem affinity purification (TAP) method: a general procedure of protein complex purification, Methods 24 (2001) 218–229.
[6] A.C. Gingras, R. Aebersold, B. Raught, Advances in protein complex analysis using mass spectrometry, J. Physiol. 563 (2005) 11–21.
[7] Y. Ho, A. Gruhler, A. Heilbut, G.D. Bader, L. Moore, S.L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier, L. Yang, C. Wolting, I. Donaldson, S. Schandorff, J. Shewnarane, M. Vo, J. Taggart, M. Goudreault, B. Muskat, C. Alfarano, D. Dewar, Z. Lin, K. Michalickova, A.R. Willems, H. Sassi, P.A. Nielsen, K.J. Rasmussen, J.R. Andersen, L.E. Johansen, L.H. Hansen, H. Jespersen, A. Podtelejnikov, E. Nielsen, J. Crawford, V. Poulsen, B.D. Sorensen, J. Matthiesen, R.C. Hendrickson, F. Gleeson, T. Pawson, M.F. Moran, D. Durocher, M. Mann, C.W. Hogue, D. Figeys, M. Tyers, Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry, Nature 415 (2002) 180–183.
[8] L.V. Zhang, S.L. Wong, O.D. King, F.P. Roth, Predicting co-complexed protein pairs using genomic and proteomic data integration, BMC Bioinformatics 5 (2004) 38.
[9] T. Wilhelm, H.P. Nasheuer, S. Huang, Physical and functional modularity of the protein network in yeast, Mol. Cell. Proteomics 2 (2003) 292–298.
[10] T. Manabe, N. Yamaguchi, J. Mukai, O. Hamada, O. Tani, Detection of protein–protein interactions and a group of immunoglobulin G-associated minor proteins in human plasma by nondenaturing and denaturing two-dimensional gel electrophoresis, Proteomics 3 (2003) 832–846.
[11] T. Manabe, K. Tachi, K. Kojima, T. Okuyama, Two-dimensional electrophoresis of plasma proteins without denaturing agents, J. Biochem. 85 (1979) 649–659.
[12] T. Hirabayashi, Two-dimensional gel electrophoresis of chicken skeletal muscle proteins with agarose gels in the first dimension, Anal. Biochem. 117 (1981) 443–451.
[13] T. Hirabayashi, Agarose isoelectric focusing for the detection of many isoforms and high molecules in muscle protein analysis, Electrophoresis 21 (2000) 446–451.
[14] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, Nature 227 (1970) 680–685.
[15] H. Hirano, Microsequence analysis of winged bean seed proteins electroblotted from two-dimensional gel, J. Protein Chem. 8 (1989) 115–130.
[16] M. Groll, L. Ditzel, J. Lowe, D. Stock, M. Bochtler, H.D. Bartunik, R. Huber, Structure of 20S proteasome from yeast at 2.4 Å resolution, Nature 386 (1997) 463–471.
[17] Y. Iwafune, H. Kawasaki, H. Hirano, Electrophoretic analysis of phosphorylation of the yeast 20S proteasome, Electrophoresis 23 (2002) 329–338.
[18] R. Yokoyama, H. Kawasaki, H. Hirano, Identification of yeast aspartyl aminopeptidase gene by purifying and characterizing its product from yeast cells, FEBS J. 273 (2006) 192–198.
[19] J.C. Giddings, Field-flow fractionation: analysis of macromolecular, colloidal, and particulate materials, Science 260 (1993) 1456–1465.
[20] H. Eubel, C.P. Lee, J. Kuo, E.H. Meyer, N.L. Taylor, A.H. Millar, Free-flow electrophoresis for purification of plant mitochondria by surface charge, Plant J. 52 (2007) 583–594.
[21] A. Hartig, H. Ruis, Nucleotide sequence of the *Saccharomyces cerevisiae* CTT1 gene and deduced amino-acid sequence of yeast catalase T, Eur. J. Biochem. 160 (1986) 487–490.
[22] Y.S. Lopez-Boado, P. Herrero, M.T. Fernandez, R. Fernandez, F. Moreno, Purification of isocitrate lyase from *Saccharomyces cerevisiae*, Yeast 4 (1988) 41–46.
[23] W. Adachi, N.N. Suzuki, Y. Fujioka, K. Suzuki, Y. Ohsumi, F. Inagaki, Crystallization of *Saccharomyces cerevisiae* aminopeptidase 1, the major cargo protein of the Cvt pathway, Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 63 (2007) 200–203.
[24] A. DeLuna, A. Avendano, L. Riego, A. Gonzalez, NADP-glutamate dehydrogenase isoenzymes of *Saccharomyces cerevisiae*: purification, kinetic properties, and physiological roles, J. Biol. Chem. 276 (2001) 43775–43783.
[25] A. Roeben, J.M. Plitzko, R. Korner, U.M. Bottcher, K. Siegers, M. Hayer-Hartl, A. Bracher, Structural basis for subunit assembly in UDP-glucose pyrophosphorylase from *Saccharomyces cerevisiae*, J. Mol. Biol. 364 (2006) 551–560.
[26] E. Inagaki, N. Ohshima, H. Takahashi, C. Kuroishi, S. Yokoyama, T.H. Tahirov, Crystal structure of *Thermus thermophilus* δ1-pyrroline-5-carboxylate dehydrogenase, J. Mol. Biol. 362 (2006) 490–501.

3 2DICALを用いた疾患バイオマーカー探索

尾野雅哉[1,✉] 松原淳一[1] 根岸綾子[2] 山田哲司[1]
Masaya Ono Junichi Matsubara Ayako Negishi Tesshi Yamada
1 国立がんセンター研究所 化学療法部 2 東京医科歯科大学大学院 顎口腔外科学
✉ E-mail：masono@ncc.go.jp

筆者らは，ショットガンプロテオミクスの解析手法として，2DICAL (2 dimensional image converted analysis of liquid chromatography mass spectrometry) を開発した．このシステムは液体クロマトグラフィー質量分析計計測で得られるデータを，質量電荷比，保持時間，ピーク強度，サンプルの4つの要素から成るものとして捉え，多数検体のショットガンプロテオミクスデータを同位体標識することなく解析するものである．2DICALを用いた疾患バイオマーカー探索について詳述することにより，この新規プロテオーム解析手法の有用性を論じる．

Key words 2DICAL，ショットガンプロテオミクス，バイオマーカー

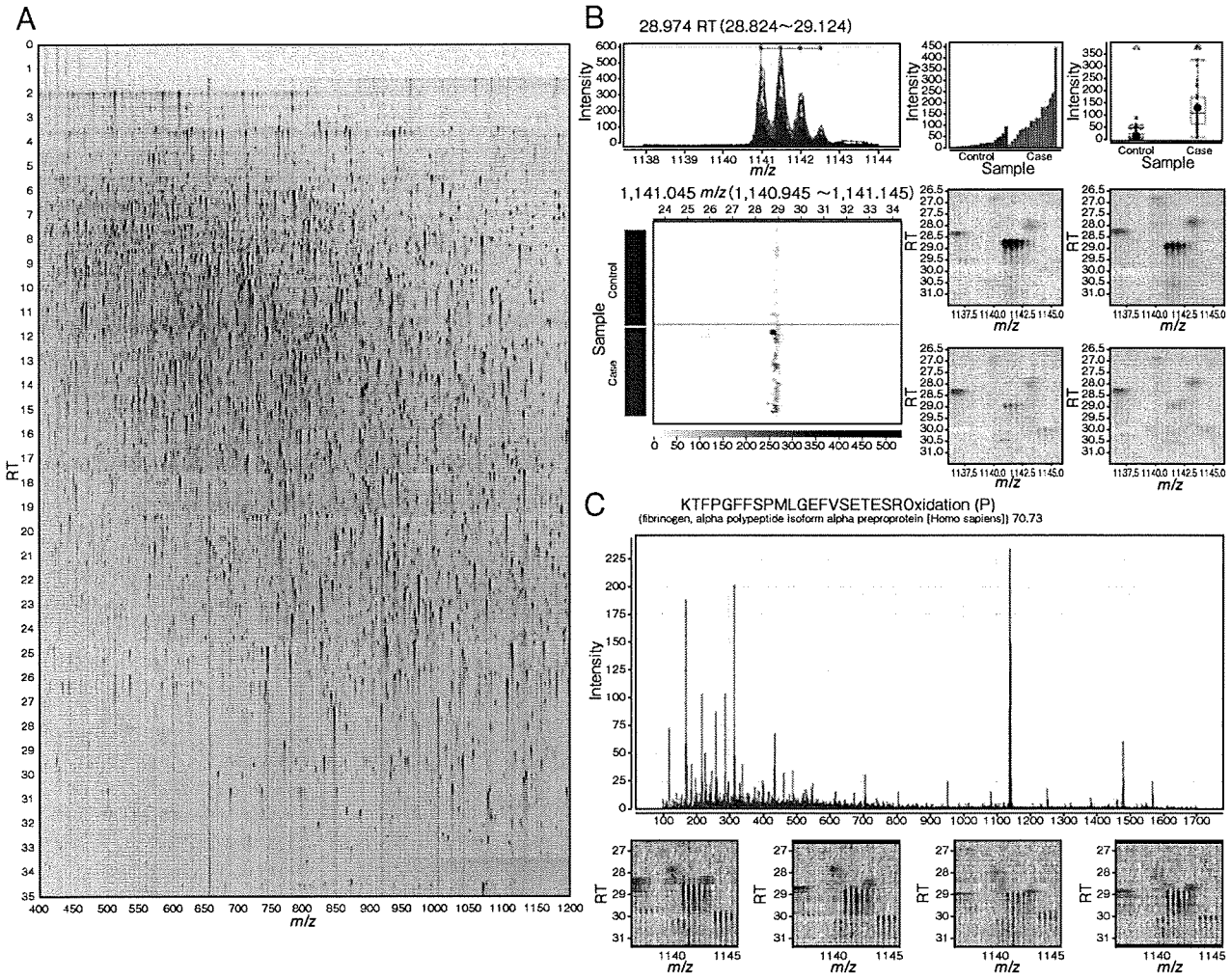
はじめに

筆者らは，ショットガンプロテオミクス[*1]の解析手法として，2DICAL (2 dimensional image converted analysis of liquid chromatography mass spectrometry) を開発した[1)]．2DICALは，多数検体の液体クロマトグラフィー質量分析計 (LC-MS) 計測で得られるデータを，質量電荷比 (m/z)，保持時間 (retention time；RT)，ピーク強度 (Intensity)，サンプル (Sample) の4つの要素から成るものとして捉え，様々な二次元画像に展開して解析するシステムである．

その根幹となるのは各スペクトラムを結合し，m/z，RTを2軸とし，ピーク強度を濃淡で表した二次元画像である (図1A)．この処理により，時間方向に分断されていた同一物質由来のピークが統合され，(m/z，RT) 座標で指定されるピークとして認識することが可能となった．さらに，サンプル間での質量電荷比，保持時間の再現性を確保するアルゴリズムを開発することにより，各サンプル間の同一ピーク同士の比較が正確に行えるようになり (図1B)，多数サンプルのショットガンプロテオミクスデータが同位体標識することなく解析可能となった．

本稿では2DICALの解析手法を詳述し，この技術を用いたバイオマーカー開発，細胞生物学への応用について解説する．

*1 ショットガンプロテオミクス
トリプシンなどのプロテアーゼで分解されたタンパク質を網羅的に解析する手法．液体クロマトグラフィー質量分析計がその解析に用いられる．



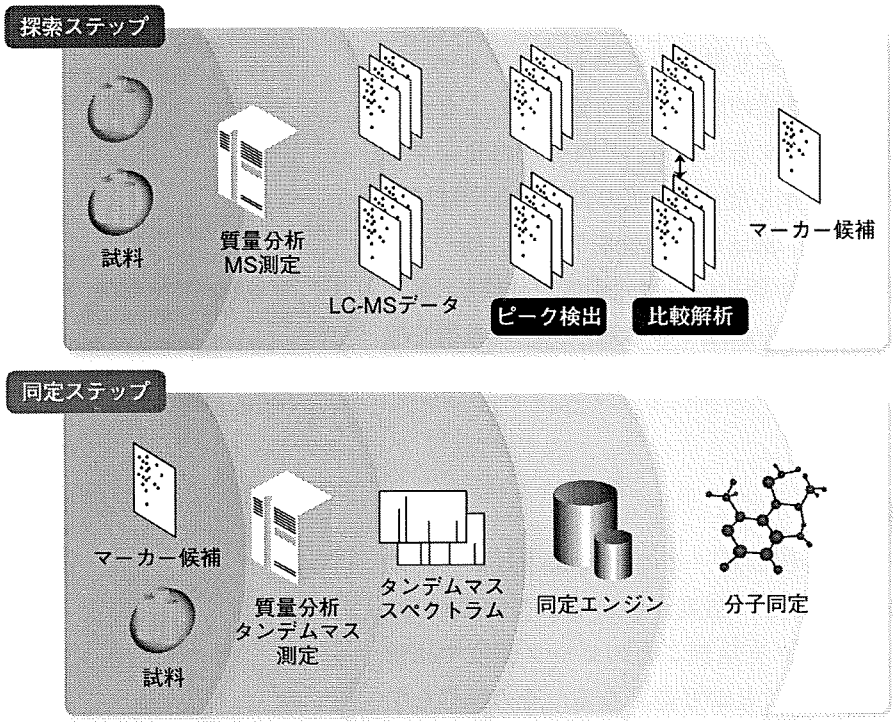
◆図1　2DICAL

時間ごとに作成されるスペクトラムをつなぎ合わせて質量電荷比（*m/z*）と保持時間（RT）の2軸を座標に持つ平面に描き出す（A）．この平面上の個々のピークを質量電荷比（*m/z*），保持時間（RT），ピーク強度（Intensity），サンプル（Sample）の4つの変数で表すことにより，多数サンプル間での同一ピーク同士の比較が正確に行えるようになった．（*m/z*，RT）=（1141.0，28.9）のピークを*m/z*，RT，Intensity，Sampleの様々な組み合わせの座標軸で示した二次元画像（B）と，そのピークを標的したタンデムマス測定によるピークの分子同定の結果（C）を示す．

I　2DICALの解析手法

■1. 2DICALの基礎概念

2DICALはLC-MSデータを二次元画像に変換して解析するシステムである．質量電荷比（*m/z*），保持時間（RT）を2軸とした二次元画像は，強度を濃淡に変換した各マススペクトラムを保持時間軸の方向に並べることで単純に描出できるが，複数サンプル間で比較検討するためには，質量電荷比，保持時間の再現性を確保しなければならない．そのため，2DICALに



◆図2 2DICALの解析の流れ
探索ステップ：比較する生体試料を用意し，LC-MS測定を行い，*m/z*，RT，Intensity，Sampleの項目でピークを数値化する．全ピークを探索し，試料間で有意に差のあるピークをマーカー候補分子として選択する．
同定ステップ：マーカー候補分子に対して*m/z*，RTをターゲットしたタンデムマス測定を行い，分子構造を決定する．

は質量電荷比，保持時間の再現性を得るためのキャリブレーションの工夫がなされている．また，2DICALではLC-MSデータを質量分析計によって*m/z*ごとに分画された液体クロマトグラフィー（LC）として捉えており，一般的なクロマトグラフィーと同様の手法でピーク検出を行っている．LC-MSデータが強度を持つ*m/z*，RT座標で指定されるピークとして認識されれば，複数サンプル間で各ピークの比較解析を行い，統計学的手法を用いて有用なピークの選択が可能となる．選択されたピークに対しては，そのピークが持つ質量電荷比，保持時間を標的してタンデムマス[*2]測定を行い，分子構造を決定する（図2）．

■2. 質量電荷比，保持時間の再現性

質量電荷比，保持時間の再現性の確保にあたって，筆者らは比較すべき対象の中で代表的なサンプル1つ（参照サンプル）に対して他のすべてのデータを一致させる手法を用いている．このため，血液同士の比較，培養細胞同士の比較のように，比較する対象のバックグラウンドが同一であることが必要となる．質量電荷比の再現性は現在の質量分析計ではほぼ保たれているが，参照サンプルの質量電荷比に合わせて，再度キャリブレーションを行っている．保持時間の再現性を得ることは質量電荷比ほど単純ではなく，いくつかの工夫が施されている．その1つとして，ダイナミックプロ

[*2] タンデムマス
試料溶液を液体クロマトグラフィー（LC）で分離し，溶出成分をイオン化して得られる第1段目の質量分析（MS）であるイオンを選択し，選択したイオンを解離（フラグメント）させて，第2段目の質量分析計（MS）で分析する手法．イオンのペプチド配列を同定することができる（LC-MS/MSとも言う）．

グラミング[*3]を用い，サンプル間で各スペクトラムの相同性から時間補正する技術が導入されている．この技術により液体クロマトグラフィーでは避けることのできない測定ごとの保持時間のばらつきを補正することが可能となった．しかし，サンプル数が増加するにつれ，それぞれのサンプルが持つ微妙な保持時間のゆらぎ[*4]が大きくなり，上記の方法のみでは同一ピークの認識が難しくなった．そこで，*m/z*の再現性を利用し，同一*m/z*における保持時間，サンプルの2軸を持つ新しい二次元画像を作成し，個々のサンプルで微秒に変動する同一ピークを認識できるピーク検出技術を開発した．これらの技術により同一ピークを*m/z*，RT座標で指定することが可能となった．

■3. 統計処理（マーカーの探索）

2DICALで*m/z*，RT座標を与えられるピークの総数は数万ピーク以上となるが，1つ1つのピークは症例ごとに強度が測定されているので，*m/z*，RT座標で特定される1つのピークを症例ごとのピーク強度で比較し，群分けした症例間で統計処理を行うことが可能となる．その操作をすべてのピークに繰り返し施行し，有意なピークを選択する．臨床検体を扱うために，一般にはU検定[*5]，ROC（receiver operating characteristic）[*5]などを用いているが，群間比較する既存のほとんどの統計解析が使用可能であるので，目的に応じて統計手法を選択することができる．

■4. ペプチド配列同定と翻訳後修飾

統計処理で有意な差が認められたピークの構造決定のために，*m/z*，RTで指定したタンデムマス測定を行う．液体クロマトグラフィーの保持時間の誤差と質量分析計の質量誤差を考慮に入れて標的ピークのタンデムマス測定を行い，その中で親イオン[*6]が標的ピークと一致することが確認されたもののタンデムマスデータより修飾構造まで含めたペプチド配列を決定する．

■5. まとめ

2DICALはLC-MSデータから得られるペプチドピークの強度を用いた比較解析から重要なピークを選択し，選択したピークに標的を絞って詳しい構造解析を行う新規プロテオーム解析手法である．2DICALは既存のショットガンプロテオミクスの解析手法と比較し，同位体標識法を用いる解析法[2), 3)]に比べ多数検体が処理できる点で優れており，ペプチド同定結果から解析する手法[4), 5)]に比べ定量性に優れていると言える．

*3 ダイナミックプログラミング
相同性のあるデータ間で最適な対応点を探索する手法の1つである．代表的な用法として，生物種間で相同性のあるDNA配列を一致させるために用いられている．

*4 保持時間のゆらぎ
液体クロマトグラフィーでの分離の再現性にはある程度の変動があり，ダイナミックプログラミングを用いても，数万ピークを多数例で検討すると補正できない微妙な変動が残る．この微妙な変動を「ゆらぎ」と表現した．

*5 U検定，ROC
統計解析手法の1つで，U検定は順位データの2群比較に適しており，ROCは感度，特異度を検討することに適している．

*6 親イオン
タンデムマス測定で解離されるイオン．

II 2DICALによる疾患バイオマーカーの開発

■1. 疾患バイオマーカーの開発

1) バイオマーカーの定義

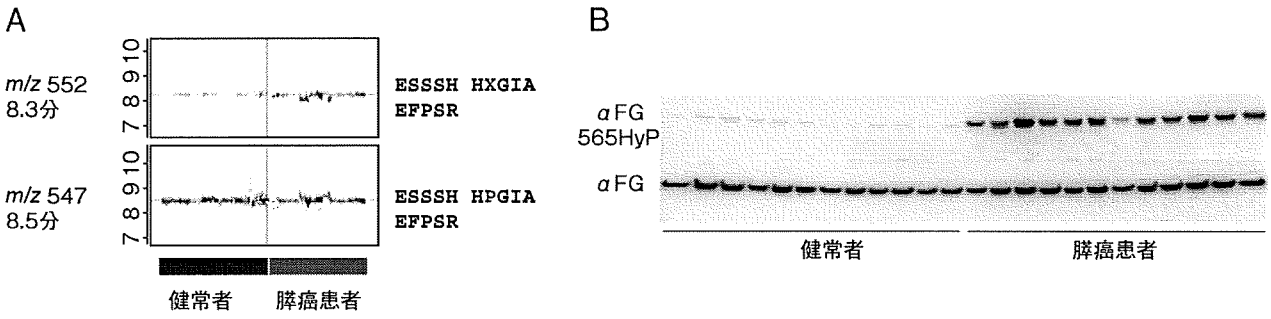
バイオマーカーは“a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention.” (NIH study group 1998) と定義されるように，生理的，病理的，薬理的な変化に対応する計測可能な客観的な指標となる性格を有するものである．生物学的な機構の詳細が不明なものでも，疾患の良い指標となるものであればバイオマーカーとなり，プロテオミクスで扱うスポットやピークもバイオマーカーとなりうる．

2) バイオマーカーの探索

生体に存在する物質は無数にあり，遺伝子（核酸），タンパク質，糖鎖，脂質などの生体物質として計測できるものであれば，どのようなものでもバイオマーカーとなる可能性がある．しかし，バイオマーカーとして意義を持つものはきわめて限られており，バイオマーカーの開発にあたっては，効率良くバイオマーカーを探索する方法を選択することが重要となる．候補となりうる可能性の高いバイオマーカーをなるべく絞り込むためには，臨床材料を用いたバイオマーカー探索においては，100例規模での比較実験が必要であると統計学的に推計されている．

3) バイオマーカーの検証および実用化

探索されたバイオマーカーはその段階では候補にすぎず，検証されて初めてバイオマーカーの地位を確立する．検証の際は探索に用いた方法にこだわらず，そのバイオマーカーを最も効率良く定量解析できる方法を選ぶことが重要である．抗体が存在するバイオマーカーであれば，その抗体を用いた既存の検査法を用いることで検証が可能である．抗体が存在しない場合には，特異抗体を作製したうえで，既存の検査法を用いることも可能である．しかし，そのような簡便な方法に置き換えられず，探索と同じ方法で検証しなければならない場合には，多数例で正確に定量解析が行えるシステムへの改良が必要となる．最終的には，臨床検査や機器開発の企業と連携した臨床治験で検証され，政府の認可を得て初めて，新規のバイオマーカーが既存の臨床検査法と同様に扱われるようになる[6]．



◆図3　2DICALで検出されたα-フィブリノゲンの水酸化プロリンペプチド断片(A)と特異抗体で認識された水酸化プロリンα-フィブリノゲン(B)（文献8)より改変）

A：2DICALにより検出された質量電荷比と保持時間がm/z 552，8.3分であるピークが，健常者と膵癌患者の間で有意に差があり，この構造は水酸化プロリン(X)を含むα-フィブリノゲンのペプチド断片であった．修飾のない同ペプチド(m/z 547，8.5分)は，健常者と膵癌患者の間で差が認められなかった．

B：水酸化プロリンα-フィブリノゲン(αFG565HyP)を特異的に認識する抗体で血漿のウェスタンブロットを行うと，健常者と膵癌患者の間で明らかな差を認めたが，α-フィブリノゲン(αFG)の量には差は認められなかった．

■2. 血漿膵癌腫瘍マーカーの開発

1) 腫瘍マーカーの探索

血漿膵癌腫瘍マーカー[*7]を探索するために，倫理要件を満たす膵癌患者血漿および健常者血漿を，国立がんセンター中央病院にてそれぞれ38症例，39症例，東京医科大学病院にてそれぞれ5症例，4症例集積し，計86症例を2DICALで解析した．コンカナバリンA(ConA)[*8]で血漿を前処理することにより，アルブミンなど血中に大量に存在するタンパク質を除去し，ConAに吸着する画分をトリプシンで消化し，質量分析測定用試料を作製した．LC-MS測定には，C18逆相カラムを用い，毎分200nlの超低流量の流速でアセトニトリル濃度勾配をかけ，質量分析計でスペクトラムデータを採取した．実験誤差を軽減するために，同一試料に対して3回の測定を行った．2DICALにより115,325ピークを検出し，膵癌患者群と健常者群間で有意差のあるピークをU検定でp＜0.0005，ピーク強度≧10，強度比≧2の条件で拾い出し，目視にてピークの有効性を確認した6ピークを選出し，バイオマーカー候補とした(図3A)．

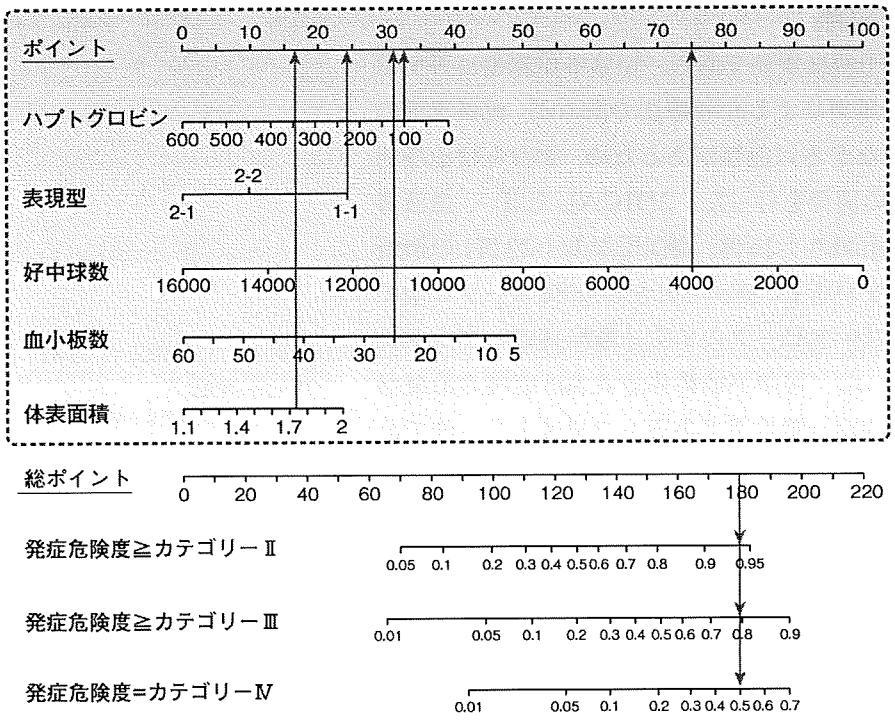
2) 腫瘍マーカーの構造決定

選出したバイオマーカー候補の構造決定のために，これらのピークを標的したタンデムマス測定を行った．質量電荷比，保持時間で(m/z 552，RT 8.3分)，(m/z 827，RT 8.3分)，(m/z 1141，RT 29.0分)で認識されるピークが，α-フィブリノゲンのペプチド断片であることが明らかとなった．これらのペプチドは予想されるペプチド質量より分子量が16Da重く，何らかの修飾を受けているものと予想された．精密質量分析計オービトラップ[*9]

*7 血漿膵癌腫瘍マーカー
膵癌は早期発見が難しく，非常に予後の悪い癌である．予後改善のために新規血漿マーカーなどの有効な診断法による早期発見が強く望まれている．

*8 コンカナバリンA(ConA)
レクチンの1つで，マンノースと強く反応する性格があり，N型糖鎖を持つ糖ペプチドを捕捉するためによく用いられる．

*9 精密質量分析計オービトラップ
フーリエ変換を用いて，分子振動から質量を測定する質量分析計．高精度，高解像度のデータが得られる．



◆図4 ノモグラムを用いた副作用予測モデル(文献9)より改変)

ある症例のデータを用いて具体的な使用法を示す．各項目の該当する値から最上段のポイントスケールまで垂線を引き点数を求め，その合計を総ポイントとする．その総ポイント(この症例では180点)から下へ垂線を引き，発症危険度スケールとの各交点が，それぞれのカテゴリーの副作用が出現する確率となる．この症例では，50%の確率でカテゴリーⅣの副作用が出現するという計算になる．

を用いて，この修飾が酸素1原子によるものであることを証明し，その酸素の付着部位がプロリン残基であることを解明した．

3) 腫瘍マーカーの検証

プロリン残基の酸素付加で安定した構造は水酸化プロリンであり，生体に主に存在するものは4位が水酸化された4-水酸化プロリンである．上記の結果を確証するために，4-水酸化プロリンに置換した合成ペプチドを作製し，これを特異的に認識する抗体(11A5)をGANP (germinal center associated nuclear protein)[*10]マウス[7]で作製した．11A5抗体は水酸化プロリンα-フィブリノゲンを認識し，膵癌患者血漿で水酸化プロリンα-フィブリノゲン量が上昇していることが確認された(図3B)．さらに，11A5抗体を用いて競合的ELISA (enzyme-linked immunosorbent assay)を作製し，多施設共同研で収集した686例の血液で血中水酸化プロリンα-フィブリノゲン量を測定したところ，健常者に比べ膵癌で有意に上昇していることが確認され，膵癌の早期段階から上昇することが確認された[8]．

*10GANP
胚中心のB細胞で発現上昇する核内因子として発見された．この遺伝子を改変したマウスは，胚中心B細胞において抗体可変領域に多くの体細胞突然変異が誘発され，通常のマウスに比べ親和性，特異性の高い抗体を作る．

■3. 膵癌治療毒性のマーカー開発

癌の臨床で化学療法の副作用予測は重要である．切除不能膵癌に対する標準的初回治療はゲムシタビン単剤による化学療法であるが，血液毒性の副作用は致命的な有害事象につながる可能性があるため，筆者らは治療前に血液毒性を予測できる診断法の開発を行った．2DICALを用い，血液毒性の両極端な症例群（強い副作用あり：25例，副作用なし：22例）の治療前血漿タンパク質プロファイルを解析し，検出された60,888ピークの中で最も2群間で差の大きかったペプチドピークのアミノ酸配列がハプトグロビン*11由来であることを同定した．切除不能膵癌患者305症例の治療前血液検体のハプトグロビン値を臨床一般検査であるネフロメトリー法で計測し，切除不能膵癌に対する化学療法副作用マーカーとして血中ハプトグロビン値が有用なマーカーであることを示した．さらに，血中ハプトグロビン値およびその表現型，治療前好中球数・血小板数，体表面積を用いた血液毒性予測モデルを構築した（図4）[9]．

III　2DICALの細胞生物学への応用

■1. 培養細胞への応用

2DICALは培養細胞のプロテオーム解析にも応用可能である．特に，石濱らが開発したSDC（sodium deoxycholate）を用いるトリプシン消化法[10]，*12により，タンパク質染色を行わずに直接2DICALでの解析が可能となった．この消化法を用いることにより，今まではゲルに流していた溶液から直接2DICALでの解析ができるため，既存のタンパク質染色より多くの情報を比較解析して有用なピークを探索し，そのピークのタンパク質同定を同時に行うことが可能となった．現在筆者らは，遺伝子導入やsiRNA（small interfering RNA）*13で変化させた細胞間のプロテオーム変化を上記の手法により検討している．

■2. 病理組織への応用

2DICALではトリプシン処理されたペプチドを用いるため，タンパク質の一次構造のみの情報で解析可能である．そのため，ホルマリン固定パラフィン包埋された病理組織標本の解析も可能である．筆者らは，重力落下回収法を採用したレーザーマイクロダイセクションシステム（LMD6000）を用いて舌癌のプロテオーム解析を行った[11]．また，SDCを用いた安価な解析手法も開発中である．

*11 ハプトグロビン
血中に豊富に存在するタンパク質で，感染や癌などの炎症で上昇すると言われているが，その働きの詳細は不明な部分が多い．

*12 SDCを用いるトリプシン消化法
界面活性剤はLC-MS測定で不利となるが，この物質を用いることにより界面活性剤のないLC-MS測定が可能となった．

*13 siRNA
特定のRNAの発現を抑制するための短い配列のRNA．

おわりに

2DICALで臨床的に有用なピークとして拾われてきたものの半数以上は，一般的なタンパク質同定システムでは同定されない．そのようなピークは，フィブリノゲンの水酸化プロリン化のように何らかの修飾を受けている可能性がある．筆者らは，この解決のためには修飾タンパク質（ペプチド）を効率良く同定できるシステムの開発が必要であると考え，新しいプロジェクトを立ち上げている．しかし，新しいシステムが修飾タンパク質（ペプチド）の構造を正確に同定しても，その構造を認識する抗体の作製が不可能な場合や，修飾構造が複雑なため同定さえ不可能な場合もありうる．その場合，質量分析計のピークのみがその物質を認識できるものとなり，質量分析計での定量値がその物質の増減を示す唯一の手段となる．MRM (multiple reaction monitoring) [12), *14]などの手法を発展させ，質量分析計での正確な定量法を確立し，質量分析計による臨床診断法を開発することも視野に入れて，今後の研究開発を進めていかなければならない．

2DICALは，長年，医学・生物学の分野で仕事をしてきた筆者が，初めてLC-MSデータを見たときにこのデータの持つ無限の可能性を直感し，いかに医学・生物学の分野でわかりやすく使えるかということを目標に開発してきたシステムである．基盤技術のさらなる向上とともに，2DICALが発展していくことを願っている．

*14 MRM
選別したイオンを解離させ，解離したイオンの一部を選択的に測定する手法．古くからの手法であるが，質量分析計による感度と定性，定量性を向上させる手段として，近年注目を集めている．

文献

1) Ono M, et al: Mol Cell Proteomics (2006) 5: 1338-1347
2) Gygi SP, et al: Nat Biotechnol (1999) 17: 994-999
3) DeSouza L, et al: J Proteome Res (2005) 4: 377-386
4) Liu H, et al: Anal Chem (2004) 76: 4193-4201
5) Ishihama Y, et al: Mol Cell Proteomics (2005) 4: 1265-1272
6) Ono M, et al: Cancer Frontier (2008) 10: 14-20
7) Sakaguchi N, et al: J Immunol (2005) 174: 4485-4494
8) Ono M, et al: J Biol Chem (2009) 284: 29041-29049
9) Matsubara J, et al: J Clin Oncol (2009) 27: 2261-2268
10) Masuda T, et al: J Proteome Res (2008) 7: 731-740
11) Negishi A, et al: Cancer Sci (2009) 100: 1605-1611
12) Finlay EM, et al: Biomed Environ Mass Spectrom (1986) 13: 633-639

For Beginners

1)「細胞工学別冊 最新プロテオミクス実験プロトコール」谷口寿章，秀潤社(2003)（質量分析入門書）
2)「実験医学増刊 分子レベルから迫る癌診断研究—臨床応用への挑戦」中村祐輔 監修，羊土社(2007)（バイオマーカー探索）
3) "LCMS解析手法2DICALを用いた糖鎖疾病マーカーの探索" 尾野雅哉，「遺伝子医学MOOK11 臨床糖鎖バイオマーカーの開発—糖鎖機能の解明とその応用」メディカルドゥ，pp.66-72 (2008)（2DICALの応用）

Snapshot Peptidomics of the Regulated Secretory Pathway*[S]

Kazuki Sasaki‡§, Yoshinori Satomi¶||, Toshifumi Takao¶, and Naoto Minamino‡**

Neurons and endocrine cells have the regulated secretory pathway (RSP) in which precursor proteins undergo proteolytic processing by prohormone convertase (PC) 1/3 or 2 to generate bioactive peptides. Although motifs for PC-mediated processing have been described ((R/K)X_n(R/K) where n = 0, 2, 4, or 6), actual processing sites cannot be predicted from amino acid sequences alone. We hypothesized that discovery of bioactive peptides would be facilitated by experimentally identifying signal peptide cleavage sites and processing sites. However, *in vivo* and *in vitro* peptide degradation, which is widely recognized in peptidomics, often hampers processing site determination. To obtain sequence information about peptides generated in the RSP on a large scale, we applied a brief exocytotic stimulus (2 min) to cultured endocrine cells and analyzed peptides released into supernatant using LC-MSMS. Of note, 387 of the 400 identified peptides arose from 19 precursor proteins known to be processed in the RSP, including nine peptide hormone and neuropeptide precursors, seven granin-like proteins, and three processing enzymes (PC1/3, PC2, and peptidyl-glycine α-amidating monooxygenase). In total, 373 peptides were informative enough to predict processing sites in that they have signal sequence cleavage sites, PC consensus sites, or monobasic cleavage sites. Several monobasic cleavage sites identified here were previously proved to be generated by PCs. Thus, our approach helps to predict processing sites of RSP precursor proteins and will expedite the identification of unknown bioactive peptides hidden in precursor sequences. ***Molecular & Cellular Proteomics 8:1638–1647, 2009.***

The generation of peptide hormones or neuropeptides involves the proteolytic processing of precursor proteins by specific proteases. In neurons and endocrine cells, most, if not all, of these bioactive peptides are generated within the RSP[1] in which the processing enzymes PC1/3 or PC2 cleave precursors at basic residues (1, 2). The PC-mediated cleavage most often occurs at consecutive basic residues, but not all basic residues serve as PC recognition sites (2). This is partly because the secondary structure of a precursor also affects the substrate recognition (3). Identification of processing sites is hence a prerequisite for locating unknown peptides hidden in a precursor sequence.

Peptidomics has been advocated to comprehensively study peptides cleaved off from precursor proteins by endogenous proteases (4–6). These naturally occurring peptides are beyond the reach of current proteomics and should be analyzed in their native forms. Unlike proteomics, peptidomics has the potential to uncover processing sites of precursor proteins. Most peptidomics studies, which target tissue peptidomes from brain or endocrine organs (7–11), have provided limited information about secretory peptides that could help to identify processing sites; they are too often blurred by subsequent actions of exopeptidases (cutting off a single amino acid or dipeptide from either end of a peptide).

In MS-based identification of bioactive peptides present in biological samples, their relative low abundance in a total pool of naturally occurring peptides should be considered. Once extracted from cultured cells or tissues, *bona fide* secretory peptides and nonsecretory peptides or peptide fragments caused by degradation of abundant cytosolic proteins cannot be discriminated, and therefore we need to analyze samples rich in secretory peptides to facilitate the identification of bioactive peptides. Several attempts have been made to isolate secretory proteins or peptides, such as subcellular fractionation for harvesting secretory granules (12, 13). With all these efforts, a limited number of secretory peptides have been identified, and many known bioactive peptides still escape analysis.

We took advantage of the fact that peptides processed in the RSP are enriched in secretory granules of neurons and endocrine cells and released on exocytosis. Here we applied a brief exocytotic stimulus (2 min) to cultured human endocrine cells and identified peptides released into supernatant using LC-MSMS on an LTQ-Orbitrap mass spectrometer. Nearly 97% of the identified peptides arose from precursor proteins known to be recruited to the RSP, such as peptide hormone precursors and granin-like secretory proteins. Our

From the ‡Department of Pharmacology, National Cardiovascular Center Research Institute, Fujishirodai, Suita, Osaka 565-8565, Japan and ¶Laboratory of Protein Profiling and Functional Proteomics, Institute for Protein Research, Osaka University, Yamadaoka, Suita, Osaka 565-0871, Japan

Received, January 27, 2009, and in revised form, March 25, 2009
Published, MCP Papers in Press, March 31, 2009, DOI 10.1074/mcp.M900044-MCP200

[1] The abbreviations used are: RSP, regulated secretory pathway; CgA, chromogranin A; CgB, chromogranin B; CT, calcitonin; CGRP, calcitonin gene-related peptide; GRP, gastrin-releasing peptide; PC, prohormone convertase; SgII, secretogranin II; SgIII, secretogranin III; SST, somatostatin; LTQ, linear trap quadrupole; IPI, International Protein Index.

approach was validated by the identification of previously known processing sites of peptide hormone precursors. In addition, a majority of the identified peptides retained cleavage sites that agree with consensus cleavage sites for PCs, which are informative enough to deduce the processing sites of RSP proteins. This peptidomics approach will expedite the identification of unknown bioactive peptides.

EXPERIMENTAL PROCEDURES

Peptide Preparation—Monolayer cultures of TT cells (14, 15) were rinsed three times with Hanks' medium (Invitrogen). Culture supernatants of the cells incubated for 2 min before and after stimulation with 10 μM forskolin plus 10 μM carbachol were harvested and rapidly extracted at 4 °C using an RP-1 solid phase extraction cartridge (GL Sciences) without centrifuging the supernatants. Bound substances were eluted in 60% ACN, 0.1% formic acid. After lyophilization of a small aliquot, samples were reconstituted in 10 μl of 2% ACN, 0.1% formic acid. In the one-dimensional analysis, solid phase-extracted analytes were subjected to LC-MSMS without gel filtration using an aliquot equivalent to 5×10^5 cells. Peptide fractions were obtained by HPLC on a gel filtration column equilibrated with 60% ACN, 0.1% TFA at a flow rate of 1.5 ml/min (G2000SWXL, 21.5 × 300 mm, Tosoh Corp.). For gel-filtrated fractions, an aliquot corresponding to 1.5×10^6 cells was used for LC-MSMS.

LC-MSMS—Nano-LC-MSMS experiments were performed with a Chorus nanoflow system (CS Analytics) connected to an LTQ-Orbitrap mass spectrometer (ThermoFisher Scientific) equipped with a nanoelectrospray emitter (MonoSpray C_{18} Nano, 100 μm × 50 mm, GL Sciences). Samples were dissolved in solvent A (2% ACN, 0.1% formic acid). The nanoflow system was run at a flow rate of 500 nl/min with a gradient from 5 to 45% solvent B (89% ACN, 0.1% formic acid) in 40 min and then to 95% B in 1 min. A protonated ion of polycyclodimethylsiloxane with *m/z* 445.120025 was used for internal calibration throughout. The mass spectrometer was operated in a data-dependent mode to automatically switch between MS and MSMS acquisitions. Survey full-scan spectra were acquired in the *m/z* range 400–1500 with five most intense ions (intensity threshold, 2e+05) sequentially isolated for MSMS in the linear ion trap using collision-induced dissociation with dynamic exclusion onward throughout the following scans. The resultant product ions were recorded in the Orbitrap.

Data Analysis and Peptide Identification—Peak picking, deisotoping, and deconvolution of MSMS spectra were preformed using Mascot Distiller (version 2.1.1.0) with the default parameters for Orbitrap. Peak lists were searched against IPI human (72,079 entries on July 2, 2008) using Mascot (version 2.2) with no enzyme specification. Pyroglutamination, C-terminal amidation, N-terminal acetylation, and methionine oxidation were simultaneously allowed as variable modifications. Peptide tolerance was set to 2 ppm, and MSMS tolerance was 25 millimass units. The significance threshold was the Mascot default setting of 5%. Each MSMS spectrum was checked manually to confirm or contradict the Mascot assignment. The false discovery rate for the identity threshold was in all cases 0% as estimated by using the Mascot decoy database function. The signals corresponding to intact calcitonin (CT), calcitonin gene-related peptide (CGRP), and somatostatin (SST) (with 1-ppm mass tolerance) underwent a mass shift of 116.01 Da after reductive alkylation with iodoacetamide and were sequenced as such by MSMS in a separate LC-MSMS analysis. Table I lists peptides that were identified with a score above the Mascot homology threshold. In the supplemental table, peptides with a score above the identity threshold (corresponding to an expectation value below 0.05) are listed and were considered identified.

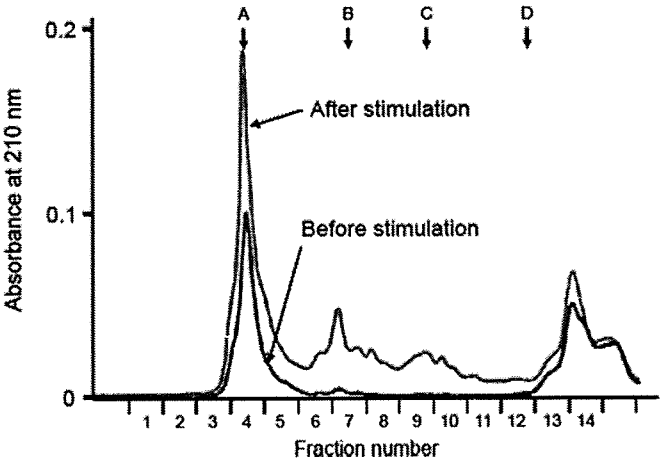


FIG. 1. **Gel filtration profiles of culture supernatant extracts from TT cells before (*black* trace) and after stimulation (*gray* trace).** *Arrows* indicate molecular mass markers: *A*, 66,500 Da; *B*, 4,271 Da; *C*, 1,673 Da; *D*, 556 Da.

RESULTS

Comprehensive Analysis of Peptides Released on Exocytosis—As a model system, we used the human medullary thyroid carcinoma cell line TT that stores peptide hormones including CT and CGRP in secretory granules (14, 15). A combination of forskolin and carbachol was used to induce exocytosis. Media from cells incubated for 2 min before and after stimulation were separately harvested and solid phase-extracted for peptide analysis. Total peptide amounts were assessed by gel filtration HPLC in which 1000–10,000-Da molecules are eluted in fractions labeled 7–10 (Fig. 1). This exocytotic stimulus elicited a 5.5-fold increase in secreted peptide amounts as assessed by the absorbance at 210 nm.

The solid phase-extracted samples were directly analyzed by LC-MSMS without gel filtration. We first examined a basal level secretion of peptides. In the medium conditioned by TT cells for 2 min, 36 peptides were identified from 13 precursors of which 30 peptides arose from nine secretory proteins including four peptide hormone precursors (CT/CGRP, gastrin-releasing peptide (GRP), and SST), four granin-like proteins (chromogranin A (CgA), chromogranin B (CgB), secretogranin III (SgIII), and VGF), and the processing enzyme PC2 (Fig. 2 and Table I). Because this cell line is known as a hyperproducer of CT and CGRP (14, 16), we tried to locate signals with mass values (within a mass tolerance of 2 ppm from a theoretical value) corresponding to bioactive CT (3415.58 Da) and CGRP (3786.96 Da). Signals from CGRP were observed, but no signals for CT were detected in the base peak chromatogram (Fig. 3).

In contrast, this stimulation facilitated identification of larger numbers of peptides, which resulted in 152 peptides being identified from 18 precursors (Fig. 2 and Table I). The six additional precursors all belonged to secretory proteins known to be processed in the RSP, including three peptide hormone precursors (pituitary adenylate cyclase-activating

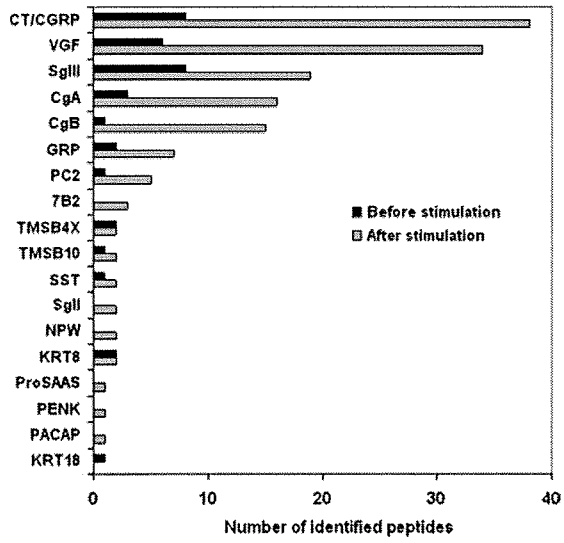


FIG. 2. **The numbers of peptides identified before and after stimulation sorted by precursor names.** CT and CGRP are grouped as they arise from alternatively spliced exons. Peptide sequences are indicated in Table I. *PACAP*, pituitary adenylate cyclase-activating polypeptide; *NPW*, neuropeptide W; *PENK*, proenkephalin A; *KRT*, cytokeratin; *TMSB4X*, thymosin β-4 X-linked; *TMSB10*, thymosin β-10.

polypeptide, neuropeptide W, and proenkephalin A) and three granin-like proteins (7B2, SgII, and pro-SAAS). In total, 146 of 152 peptides arose from the 15 RSP precursors. The remaining six peptides were derived from thymosins and cytokeratin 8. According to the stimulus-induced increase in total peptide amounts released to culture supernatant, more peptides were identified from the former nine precursor proteins (Figs. 1 and 2). Across the CT and CGRP precursor sequences, known major processing products (17) were identified, namely CT N-terminal propeptide (6217.04 Da), bioactive CT (3415.58 Da), katacalcin (2435.07 Da), CGRP N-terminal propeptide (6056.04 Da), and bioactive CGRP (3786.96 Da) (Fig. 3).

Investigation of Cleavage Sites through Identified Peptides—Regarding the processing of peptide hormone precursors, it has long been known that PC1/3 or PC2 cleaves the precursors at sites containing consecutive basic amino acids following N-terminal signal peptide cleavage (1, 2). Cameron *et al.* (18) recently studied the specificity of PCs and drew a conclusion that the PC-mediated cleavage occurs at sites containing pairs of basic amino acids separated by 0, 2, 4, or 6 residues. However, some peptide hormone precursors are processed at monobasic residues although at much less frequency (2). In any case, the resultant C-terminal basic residues are subsequently removed by carboxypeptidase E. If the exposed C-terminal residue is glycine, peptidyl-glycine α-amidating monooxygenase catalyzes peptide α-amidation, a common post-translational modification often required for a peptide to be fully bioactive (19).

Having confirmed that almost all the sequenced peptides arose from RSP precursors (Table I), we extracted 10 amino acids N- or C-terminally flanking the sequenced peptides to analyze their cleavage sites. Any glycine immediately followed by a basic residue(s) that creates an amidation site was also counted as a PC consensus site. Monobasic sites were defined as those that do not harbor basic residues except for P1 position and considered potential processing sites as well. PC consensus cleavage sites, signal sequence cleavage sites, and monobasic sites, referred to as informative cleavage sites in the present study, were found in 120, 26, and 29 peptides, respectively (72 peptides were counted twice under this definition). Overall 142 of 152 had such informative cleavage sites at either or both ends.

Identity of Major Peptides in the TT Secretopeptidome—In the LC-MSMS setting used throughout this study, any signal that transcended a given intensity threshold was automatically subjected to MSMS and ignored thereafter if it persisted within a precursor mass tolerance of 5 ppm with the aim of sequencing as many peptides as possible. The signal is not always subjected to MSMS at its maximum intensity in LC-MS profiles, and therefore this setting could return relatively low scores even for abundant peptides, which may not be included in Table I. To identify intense signals in LC-MSMS base peak chromatograms, we examined peptide peak intensities in all MS spectra. Table II provides the list of 35 peptides that were detected at the indicated monoisotopic *m/z* and charge state with a base peak intensity beyond 2e+06 (see also Fig. 3 and supplemental Fig. 1). Three peptide sequences had expectation values (above 0.05) that did not exceed the Mascot significance threshold. First, the 1278.67-Da peptide was qualified as a CGRP-derived peptide based on the observation that CGRP-derived peptides are most abundantly expressed in the TT secretopeptidome (Fig. 3 and supplemental Fig. 3). Second, the 5687.91-Da peptide also yielded suboptimum MSMS spectra but was identified as a PC2-derived peptide because of matches for eight consecutive b-ions (supplemental Fig. 3). Third, the 8559.61-Da peptide was qualified using MSMS spectral comparison with commercially available human ubiquitin (supplemental Fig. 2).

This list covered all the major processing products of CT and CGRP. Intact CGRP continued to be observed at multiple charged ions (+3 to +6) over a retention time of 10 min in the mass chromatogram, suggesting that it represents one of the most abundant peptides in this secretopeptidome (Fig. 3). The CGRP-derived peptide (ACDTATCVTHRLAGLLSRSGGVVKN) appeared to be generated from endoproteolytic cleavage of intact CGRP because the C-terminal cleaved half (1278.67 Da) was detected as well (Table II). Except for this peptide, all the N-terminal cleavage sites of CT/CGRP-derived peptides are known as major processing sites (17). Similar findings were obtained with C-terminal cleavage sites except for four non-basic sites. It remains to be clarified whether these non-basic cleavage sites point to the processing that actually occurred in the RSP.

TABLE I

Peptides identified before and after stimulation

Data were obtained without gel filtration chromatography and summarized from three runs using an identical LC-MSMS setting and peptides whose scores (column 7) exceeded homology thresholds (HT, column 8) in at least two runs are listed. For peptides identified in multiple runs, a higher score is listed. Mr(Calc) represents the theoretical monoisotopic molecular mass (Da) based on the peptide sequence. The score value beyond an identity threshold (IT, column 9) is indicated in bold. In column 1, "CT/CGRP" indicates that the peptides are shared by CT and CGRP precursors. If the peptide was identified at different charges, the charge states are also shown in column 4. Expectation values are indicated in column 10. The N- (N-term) and C-terminal (C-term) flanking 10 amino acids (columns 11 and 13) are shown and marked as follows: closed boxes with white letters, typical PC cleavage sites containing consecutive basic residues and sites having C-terminal amidation motifs; dark gray boxes, cleavage sites containing basic residues at P4, P6 or P8 position; pale gray boxes, sites having basic residues at P1 but not at P2, P4, P6 or P8 position. "Signal" indicates that the peptide flanks its signal sequence. "C-term" indicates that the peptide C-terminus is the end of the precursor protein. Ac-, N-terminal acetylation; -NH2, C-terminal amidation; <Q, pyroglutamic acid. Oxidized methionine residues are underlined in column 12. Sequences are based on the following IPI accession numbers: CgA, 00746813; CgB, 00006601, CGRP, 00027855; CT, 00000914; GRP, 00011722; KRT18, 00554788; KRT8, 00554648; PC2, 00029131; SgIII, 00292071; SST, 00000130; TMSB10, 00220827; TMSB4X, 00220828; VGF, 00069058; 7B2, 00008944; NPW, 00853190; PACAP, 00000027; PENK, 00000828; SgII, 00009362. PACAP, pituitary adenylate cyclase-activating polypeptide; NPW, neuropeptide W; PENK, proenkephalin A; KRT, cytokeratin; TMSB4X, thymosin β-4 X-linked; TMSB10, thymosin β-10.

Before stimulation

Precursor	m/z (obsd)	z	Also ID at	Mr(calc) (Da)	Mass error (Da)	Score	HT	IT	Expect. value	N-term	Peptide	C-term
CgA	581.8036	2		1161.5931	-0.0005	60	29	45	0.0016	SSMKLSFRAR	AYGFRGPGPQL	RRGWRPSSRE
CgA	725.3738	2		1448.7333	-0.0003	30	16	45	1.8	LSKEWEDSKR	WSKMDQLAKELT	AEKRLEGQEE
CgA	825.4139	2		1648.8130	0.0002	54		45	0.0078	LSKEWEDSKR	WSKMDQLAKELTAE	KRLEGQEEEE
CgB	587.9692	3		1760.8846	0.0012	51	20	47	0.022	ARVPKLDLKR	QYDRVAQLDQLLHY	RKKSAEFPDF
CGRP	877.8292	3		2630.4657	0.0000	74		43	4.40E-05	CDTATCVTHR	LAGLLSRSGGVVKNNFVPTNVGSKAF-NH2	GRRRRDLQA
CGRP	778.0571	3		2331.1488	0.0007	51	35	47	0.023	LAALVQDYVQ	MKASELEQEQEREGSRIIAQ	KRACDTATCV -
CGRP	961.0179	2		1920.0218	-0.0005	56	20	46	0.0065	THRLAGLLSR	SGGVVKNNFVPTNVGSKAF-NH2	GRRRRDLQA
CGRP	509.7875	2		1017.5607	-0.0003	35	21	46	0.71	RSGGVVKNNF	VPTNVGSKAF-NH2	GRRRRDLQA
CT	751.0047	3		2249.9906	0.0017	44	17	42	0.035	IGVGAPGKKR	DMSSDLERDHRPHVSMPQN	AN
CT	774.6831	3		2321.0277	-0.0003	50	16	43	0.011	IGVGAPGKKR	DMSSDLERDHRPHVSMPQNA	N
CT	812.6974	3		2435.0706	-0.0003	52	17	42	0.0054	IGVGAPGKKR	DMSSDLERDHRPHVSMPQNAN	C-term
CT/CGRP	881.7805	3		2642.3187	0.0009	122	25	48	1.90E-09	Signal	APFRSALESSPADPATLSEDEARLL	LAALVQDYVQ
GRP	800.9424	2		1599.8695	0.0009	41	20	45	0.15	Signal	VPLPAGGGTVLTKMYP	RGNHWAVGHL
GRP	586.3308	3		1755.9706	0.0001	76	24	44	3.50E-05	Signal	VPLPAGGGTVLTKMYPR	GNHWAVGHLM
KRT18	485.4749	5	4	2422.3405	-0.0021	40		44	0.12	NSMQTIQKTT	TRRIVDGKVVSETNDTKVLRH	C-term
KRT8	884.2651	5		4416.2885	0.0007	93		49	2.10E-06	TSPGLSYSLG	SSFGSGAGSSSFSRTSSSRAVVVKKIETRDGKLVSESSDVLPK	C-term
KRT8	742.1170	4		2964.4438	-0.0048	23	16	47	16	RVTQKSYKVS	TSGPRAFSSRSYTSGPGSRISSSSFSRVG	SSNFRGGLGG
PC2	633.3122	2		1264.6122	-0.0023	49	23	46	0.032	HKQQLERDPR	VKMALQQEGFD	RKKRGYRDIN
SgIII	735.3826	2		1468.7521	-0.0014	37	33	46	0.44	GSQDKSLHNR	ELSAERPLNEQIA	EAEEDKIKKT
SgIII	799.9040	2		1597.7947	-0.0012	48	30	46	0.035	GSQDKSLHNR	ELSAERPLNEQIAE	AEEDKIKKTY
SgIII	899.9442	2		1797.8744	-0.0005	67	22	46	0.00045	GSQDKSLHNR	ELSAERPLNEQIAEAE	EDKIKKTYPP
SgIII	1021.9787	2		2041.9439	-0.0011	101	22	45	1.50E-07	GSQDKSLHNR	ELSAERPLNEQIAEAEED	KIKKTYPPEN
SgIII	762.0485	3		2283.1230	0.0007	54	27	47	0.011	GSQDKSLHNR	ELSAERPLNEQIAEAEEDKI	KKTYPPENKP
SgIII	756.3841	2	3	1510.7528	0.0008	50	29	46	0.022	Signal	FPKPGGSQDKSLHN	RELSAERPLN
SgIII	787.4010	5		3931.9663	0.0021	29	16	49	5.4	Signal	FPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKI	KKTYPPENKP
SgIII	866.9151	2		1731.8163	-0.0007	45	22	45	0.06	KIEKERQSIR	SSPLDNKLNVEDVDST	KNRK
SST	622.7880	2		1243.5615	0.0000	66	22	41	0.0002	QDEMRLELQR	SANSNPAMAPRE	RKAGCKNFFW
TMSB10	823.2613	6	4, 5, 7	4933.5229	0.0012	148		49	8.50E-12	M	Ac-ADKPDMGEIASFDKAKLKKTETQEKNTLPTKETIEQEKRSEIS	C-term
TMSB4X	827.7546	6	4, 5, 7	4960.4862	-0.0020	95		50	1.40E-06	M	Ac-SDKPDMAEIEKFDKSKLKKTETQEKNPLPSKETIEQEKQAGES	C-term
TMSB4X	996.3035	5		4976.4811	-0.0002	77		49	9.90E-05	M	Ac-SDKPDMAEIEKFDKSKLKKTETQEKNPLPSKETIEQEKQAGES	C-term
VGF	917.7143	4		3666.8278	0.0004	68	23	48	0.00067	Signal	APPGRPEAQPPPLSSEHKEPVAGDAVPGPKDGSAPEV	RGARNSEPQD
VGF	988.7540	4	5, 6	3950.9875	-0.0006	90	23	49	3.90E-06	Signal	APPGRPEAQPPPLSSEHKEPVAGDAVPGPKDGSAPEVRGA	RNSEPQDEGE
VGF	742.1746	5		3705.8346	0.0020	32	18	48	2.5	YPGREAQARR	AQEEAEAEERRLQEQEELENYIEHVLRRRP	C-term
VGF	639.0110	3		1914.0112	0.0001	32	20	46	1.3	EVEEKRKRKK	NAPPEPVPPPRAAPAPTHV	RSPQPPPPAP
VGF	848.4221	4		3389.6600	-0.0007	109	20	48	4.00E-08	RDFSPSSAKR	<QQETAAAETETRTHTLTRVNLESPGPERVW	RASWGEFQAR
VGF	852.6792	4		3406.6865	0.0010	102		48	2.50E-07	RDFSPSSAKR	QQETAAAETETRTHTLTRVNLESPGPERVW	RASWGEFQAR

After stimulation

7B2	851.4489	2		1700.8846	-0.0012	44	25	46	0.083	MKGGERRKRR	SVNPYLQGQRLDNVV	AKKSVPHFSD
7B2	886.9685	2		1771.9217	0.0008	51	26	47	0.022	MKGGERRKRR	SVNPYLQGQRLDNVVA	KKSVPHFSDE
7B2	751.3260	2		1500.6369	0.0006	64	22	40	0.00022	QRLDNVVAKK	SVPHFSDEDKDPE	C-term
CgA	525.2614	2		1048.5090	-0.0007	69	32	45	0.0002	SSMKLSFRAR	AYGFRGPGPQ	LRRGWRPSSR
CgA	581.8039	2		1161.5931	0.0001	65	29	45	0.00055	SSMKLSFRAR	AYGFRGPGPQL	RRGWRPSSRE
CgA	464.7536	2		927.4927	-0.0001	40	37	46	0.19	MKLSFRARAY	GFRGPGPQL	RRGWRPSSRE
CgA	666.6731	3	2	1996.9966	0.0010	43	25	47	0.12	FRGPGPQLRR	GWRPSSREDSLEAGLPLQ	VRGYPEEKKE
CgA	699.6958	3		2096.0650	0.0004	54	26	47	0.012	FRGPGPQLRR	GWRPSSREDSLEAGLPLQV	RGYPEEKKEE
CgA	1302.9266	3		3905.7636	-0.0056	51	16	43	0.0092	AKERAHQQKK	HSGFEDELSEVLENQSSQAELKEAVEEPSSKDVME	KREDSKEAEK
CgA	699.3080	3		2094.9011	0.0010	59	18	41	0.00082	LAKELTAEKR	LEGQEEEEDNRDSSMKLS	FRARAYGFRG
CgA	748.3303	3		2241.9695	-0.0003	83	23	41	3.20E-06	LAKELTAEKR	LEGQEEEEDNRDSSMKLSF	RARAYGFRGP
CgA	824.0432	3		2469.1077	-0.0001	94	18	43	4.70E-07	LAKELTAEKR	LEGQEEEEDNRDSSMKLSFRA	RAYGFRGPGP
CgA	754.7653	5		3768.7914	-0.0013	37	19	47	0.51	LAKELTAEKR	LEGQEEEEDNRDSSMKLSFRARAYGFRGPGPQL	RRGWRPSSRE
CgA	727.3265	3	2	2178.9552	0.0023	109	24	42	1.10E-08	VPQGLFRGGK	SGELEQEEERLSKEWEDS	KRWSKMDQLA
CgA	674.8492	2		1347.6856	-0.0017	64	32	47	0.0011	LSKEWEDSKR	WSKMDQLAKEL	TAEKRLEGQE
CgA	725.3741	2		1448.7333	0.0003	63	26	44	0.00068	LSKEWEDSKR	WSKMDQLAKELT	AEKRLEGQEE
CgA	760.8932	2		1519.7704	0.0015	54	26	46	0.0093	LSKEWEDSKR	WSKMDQLAKELTA	EKRLEGQEEE
CgA	825.4145	2	3	1648.8130	0.0014	92	27	46	1.50E-06	LSKEWEDSKR	WSKMDQLAKELTAE	KRLEGQEEEE
CgA	546.2849	2		1090.5560	-0.0007	27	16	44	3.4	SMKLSFRARA	YGFRGPGPQL	RRGWRPSSRE
CgB	624.7640	2		1247.5128	0.0006	56	29	40	0.0013	HHRGRGGEPR	AYFMSDTREE	KRFLGEGHHR
CgB	448.6857	2		895.3573	-0.0005	15	14	41	20	VSMASLGEKR	DHHSTHY	RASEEEPEYG
CgB	689.3519	3		2065.0327	0.0010	78	22	47	4.60E-05	VLKTSRKDVK	DKETTENENTKFEVRLL	RDPADASEAH
CgB	599.6091	3		1795.8060	-0.0004	21	17	44	9.8	FMSDTREEKR	FLGEGHHRVQENQMD	KARRHPQGAW
CgB	665.9865	3		1994.9381	-0.0004	87	21	46	3.70E-06	FMSDTREEKR	FLGEGHHRVQENQMDKA	RRHPQGAWKE
CgB	653.2990	3		1956.8748	0.0004	58	19	43	0.0019	LEPGKGRHHR	GRGGEPRAYFMSDTREE	KRFLGEGHHR
CgB	689.3477	4		2753.3602	0.0013	86	23	48	8.10E-06	YSSHHTAEKR	KRLGELFNPYYDPLQWKSSHFE	RRDNMNDNFL
CgB	824.0616	3		2469.1641	-0.0012	35	19	46	0.65	YSSHHTAEKR	LGELFNPYYDPLQWKSSHFE	RRDNMNDNFL
CgB	569.8506	2		1137.6870	-0.0003	35	32	41	0.24	EDVNWGYEKR	NLARVPKLDL	KRQYDRVAQL

TABLE I—*continued*

CgB	731.1431	5		3650.6808	-0.0015	50		45	0.016	EDVNWGYEKR	NYPSLELDKMAHGYGEESEEERGLEPGKGRHH	RGRGGEPRAY
CgB	671.7838	2		1341.5513	0.0016	22	16	41	4.4	EYNYDWWEKK	PFSEDVNWGYE	KRNLARVPKL
CgB	731.3881	2		1460.7623	-0.0007	27	27	46	4.3	ARVPKLDLKR	QYDRVAQLDQLL	HYRKKSAEFP
CgB	872.9364	2		1743.8580	0.0003	61	32	46	0.0016	ARVPKLDLKR	<QYDRVAQLDQLLHY	RKKSAEFPDF
CgB	587.9688	3		1760.8846	0.0000	64	21	47	0.001	ARVPKLDLKR	QYDRVAQLDQLLHY	RKKSAEFPDF
CgB	1068.4681	3		3202.3844	-0.0019	100	21	40	5.80E-08	HHHGRSRPDR	SSQGGSLPSEEKGHPQEESEESNVSMASLGE	KRDHHSTHYR
CGRP	1186.6031	5	6	5927.9850	-0.0059	136		50	1.50E-10	Signal	APFRSALESSPADPATLSEDEARLLLAALVQDYVQMKASELEQEQEREGSRIIA	QKRACDTATC
CGRP	1010.3481	6	5	6056.0436	0.0014	156	24	50	1.60E-12	Signal	APFRSALESSPADPATLSEDEARLLLAALVQDYVQMKASELEQEQEREGSRIIAQ	KRACDTATCV
CGRP	734.3766	3		2200.1083	-0.0004	52	20	46	0.014	AALVQDYVQM	KASELEQEQEREGSRIIAQ	KRACDTATCV
CGRP	877.8294	3		2630.4657	0.0006	118		43	2.10E-09	CDTATCVTHR	LAGLLSRSGGVVKNNFVPTNVGSKAF-NH2	GRRRRDLQA
CGRP	797.4484	3		2389.3230	0.0002	81		44	1.20E-05	ATCVTHRLAG	LLSRSGGVVKNNFVPTNVGSKAF-NH2	GRRRRDLQA
CGRP	735.3707	3		2203.0902	0.0000	55	34	47	0.0087	LAALVQDYVQ	MKASELEQEQEREGSRIIA	QKRACDTATC
CGRP	778.0572	3		2331.1488	0.0010	100	33	47	3.30E-07	LAALVQDYVQ	MKASELEQEQEREGSRIIAQ	KRACDTATCV
CGRP	788.4104	2		1574.8053	0.0010	38	21	47	0.43	THRLAGLLSR	SGGVVKNNFVPTNVGS	KAFGRRRR
CGRP	961.0179	2		1920.0218	-0.0005	96	27	46	5.90E-07	THRLAGLLSR	SGGVVKNNFVPTNVGSKAF-NH2	GRRRRDLQA
CGRP	509.7875	2		1017.5607	-0.0003	38	22	46	0.3	RSGGVVKNNF	VPTNVGSKAF-NH2	GRRRRDLQA
CGRP	853.7657	3		2558.2758	-0.0005	40	18	47	0.28	LLLAALVQDY	VQMKASELEQEQEREGSRIIAQ	KRACDTATCV
CGRP	649.3277	4		2593.2805	0.0010	71	17	47	0.00025	RLLLAALVQD	YVQMKASELEQEQEREGSRIIA	QKRACDTATC
CGRP	908.1203	3	4	2721.3391	-0.0001	120	33	47	3.30E-09	RLLLAALVQD	YVQMKASELEQEQEREGSRIIAQ	KRACDTATCV
CT	1037.1800	6	5	6217.0396	-0.0033	143	23	49	2.50E-11	Signal	APFRSALESSPADPATLSEDEARLLLAALVQDYVQMKASELEQEQEREGSSLDSPRS	KRCGNLSTCM
CT	1247.6136	5		6233.0345	-0.0029	58		49	0.0077	Signal	APFRSALESSPADPATLSEDEARLLLAALVQDYVQMKASELEQEQEREGSSLDSPRS	KRCGNLSTCM
CT	602.7485	2		1203.4826	-0.0002	74	21	37	1.30E-05	IGVGAPGKKR	DMSSDLERDH	RPHVSMPQNA
CT	610.7458	2		1219.4775	-0.0005	54	21	36	0.00099	IGVGAPGKKR	DMSSDLERDH	RPHVSMPQNA
CT	532.2388	3		1593.6954	-0.0009	29	16	42	1.1	IGVGAPGKKR	DMSSDLERDHRPH	VSMPQNAN
CT	756.3361	3		2265.9855	0.0008	40	16	41	0.079	IGVGAPGKKR	DMSSDLERDHRPHVSMPQN	AN
CT	761.6672	3		2281.9804	-0.0007	24	15	41	2.6	IGVGAPGKKR	DMSSDLERDHRPHVSMPQN	AN
CT	780.0152	3		2337.0226	0.0013	49	16	42	0.011	IGVGAPGKKR	DMSSDLERDHRPHVSMPQNA	N
CT	774.6831	3		2321.0277	-0.0002	65	20	43	0.00033	IGVGAPGKKR	DMSSDLERDHRPHVSMPQNA	N
CT	812.6975	3	2, 4	2435.0706	0.0000	65	21	42	0.00031	IGVGAPGKKR	DMSSDLERDHRPHVSMPQNAN	C-term
CT	818.0293	3		2451.0655	0.0005	59	17	41	0.0009	IGVGAPGKKR	DMSSDLERDHRPHVSMPQNAN	C-term
CT	721.3836	2		1440.7514	0.0012	53	29	45	0.0086	LGTYTQDFNK	FHTFPQTAIGVGAP-NH2	GKKRDMSSDL
CT	831.7223	3	4	2492.1449	0.0002	134	27	45	6.40E-11	LAALVQDYVQ	MKASELEQEQEREGSSLDSPRS	KRCGNLSTCM
CT	490.7348	2		979.4545	0.0005	41	32	43	0.09	MSSDLERDHR	PHVSMPQNA	N
CT	547.7556	2		1093.4975	-0.0007	37	23	43	0.26	MSSDLERDHR	PHVSMPQNAN	C-term
CT	961.7856	3		2882.3352	-0.0003	134	22	45	7.70E-11	RLLLAALVQD	YVQMKASELEQEQEREGSSLDSPRS	KRCGNLSTCM
CT/CGRP	674.3307	2		1346.6466	0.0002	40	27	44	0.14	Signal	APFRSALESSPAD	PATLSEDEAR
CT/CGRP	881.7808	3		2642.3187	0.0020	106	23	48	8.60E-08	Signal	APFRSALESSPADPATLSEDEARLL	LAALVQDYVQ
CT/CGRP	919.4757	3		2755.4028	0.0025	28	22	48	5.1	Signal	APFRSALESSPADPATLSEDEARLLL	AALVQDYVQM
CT/CGRP	1118.5786	3		3352.7150	-0.0010	61	23	48	0.003	Signal	APFRSALESSPADPATLSEDEARLLLAALVQD	YVQMKASELE
CT/CGRP	560.3060	2		1118.5972	0.0002	39	27	46	0.24	TLSEDEARLL	LAALVQDYVQ	MKASELEQEQ
GRP	560.7768	2		1119.5396	-0.0005	28	21	45	2.8	GTVLTKMYPR	GNHWAVGHLM-NH2	GKKSTGESSS
GRP	670.8837	2		1339.7534	-0.0004	62	28	41	0.00039	Signal	VPLPAGGGTVLTKM	YPRGNHWAVG
GRP	752.4157	2		1502.8167	0.0001	64	22	43	0.0004	Signal	VPLPAGGGTVLTKMY	PRGNHWAVGH
GRP	800.9420	2	3	1599.8695	-0.0001	50	21	43	0.011	Signal	VPLPAGGGTVLTKMYP	RGNHWAVGHL
GRP	808.9396	2		1615.8644	0.0003	63		44	0.00074	Signal	VPLPAGGGTVLTKMYP	RGNHWAVGHL
GRP	878.9924	2	3	1755.9706	-0.0003	139	26	43	1.40E-11	Signal	VPLPAGGGTVLTKMYPR	GNHWAVGHLM
GRP	591.6626	3		1771.9655	0.0003	54	23	45	0.007	Signal	VPLPAGGGTVLTKMYPR	GNHWAVGHLM
KRT8	479.9398	3		1436.7987	-0.0010	60	27	44	0.0014	M	Ac-SIRVTQKSYKVS	TSGPRAFSSR
KRT8	513.6225	3		1537.8464	-0.0008	39	18	44	0.17	M	Ac-SIRVTQKSYKVST	SGPRAFSSRS
NPW	541.8009	2		1081.5880	-0.0007	31	22	45	1.2	VQELWETRRR	SSQAGIPVRAP	RSPRAPEPAL
NPW	413.5702	3		1237.6891	-0.0003	28	17	41	1.2	VQELWETRRR	SSQAGIPVRAPR	SPRAPEPALE
PACAP	774.3423	2		1546.6688	0.0012	31	18	42	0.75	GDDAEPLSKR	HSDGIFTDSYSRY	RKQMAVKKYL
PC2	715.7713	5		3573.8229	-0.0030	88	21	48	5.50E-06	Signal	ERPVFTNHFLVLELHKGGEDKARQVAAEHGFGV	RKLPFAEGLY
PC2	813.5659	7		5687.9079	0.0027	44		50	0.19	Signal	ERPVFTNHFLVLELHKGGEDKARQVAAEHGFGVRKLPFAEGLYHFYHNGLA	KAKRRRSLHH
PC2	496.5950	3		1486.7641	-0.0009	51	24	46	0.016	NGLAKAKRRR	SLHHKQQLERDP	RVKMALQQEG
PC2	633.3134	2		1264.6122	0.0000	69	25	45	0.00022	HKQQLERDPR	VKMALQQEGFD	RKKRGYRDIN
PC2	641.3111	2		1280.6071	0.0006	72	27	44	9.20E-05	HKQQLERDPR	VKMALQQEGFD	RKKRGYRDIN
PENK	693.8411	2		1385.6674	0.0002	62	28	46	0.0013	YGGFMRGLKR	SPQLEDEAKELQ	KRYGGFMRRV
ProSAAS	743.6491	4		2970.5635	0.0037	53	21	47	0.013	ETGAPRRFRR	SVPRGEAAGAVQELARALAHLLEAERQE	RARAEAQEAE
SgII	831.4189	6		4982.4661	0.0039	37		49	1	TDKLAPVSKR	FPVGPPKNDDTPNRQYWDEDLLMKVLEYLNQEKAEKGREHIA	KRAMENM
SgII	1088.2673	4		4349.0419	-0.0018	112		46	1.50E-08	EIINSNQVKR	VPGQGSSEDDLQEEEQIEQAIKEHLNQGSSQETDKLAPVS	KRFPVGPPKN
SgIII	643.3237	2		1284.6310	0.0019	49	29	46	0.029	GSQDKSLHNR	ELSAERPLNEQ	IAEAEEDKIK
SgIII	699.8649	2		1397.7150	0.0002	41	26	46	0.18	GSQDKSLHNR	ELSAERPLNEQI	AEAEEDKIKK
SgIII	735.3837	2		1468.7521	0.0007	44	41	44	0.066	GSQDKSLHNR	ELSAERPLNEQIA	EEDKIKKTYP
SgIII	799.9048	2		1597.7947	0.0003	49	36	46	0.03	GSQDKSLHNR	ELSAERPLNEQIAE	AEEDKIKKTY
SgIII	899.9447	2		1797.8744	0.0005	67	22	46	0.00047	GSQDKSLHNR	ELSAERPLNEQIAEAE	EDKIKKTYPP
SgIII	964.4659	2		1926.9170	0.0002	99	32	45	2.30E-07	GSQDKSLHNR	ELSAERPLNEQIAEAEE	DKIKKTYPPE
SgIII	1021.9788	2		2041.9439	-0.0009	98	21	45	2.80E-07	GSQDKSLHNR	ELSAERPLNEQIAEAEED	KIKKTYPPEN
SgIII	1142.5676	2	3	2283.1230	-0.0023	95	32	47	8.50E-07	GSQDKSLHNR	ELSAERPLNEQIAEAEEDKI	KKTYPPENKP
SgIII	466.7270	2		931.4400	-0.0005	32	30	45	1.1	Signal	FPKPGGSQD	KSLHNRELSA
SgIII	630.8329	2		1259.6510	0.0002	27	25	45	3.4	Signal	FPKPGGSQDKSL	HNRELSAERP
SgIII	504.5913	3	2	1510.7528	-0.0007	48	22	45	0.025	Signal	FPKPGGSQDKSLHN	RELSAERPLN
SgIII	862.6868	4		3446.7178	0.0004	22	19	48	24	Signal	FPKPGGSQDKSLHNRELSAERPLNEQIAEAE	EDKIKKTYPP
SgIII	923.7038	4		3690.7873	-0.0013	50	19	48	0.033	Signal	FPKPGGSQDKSLHNRELSAERPLNEQIAEAEED	KIKKTYPPEN
SgIII	787.4007	5	6	3931.9663	0.0008	59	18	49	0.0055	Signal	FPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKI	KKTYPPENKP
SgIII	661.1581	6		3960.9050	0.0000	71		48	0.00027	IEWLKKHDKK	GNKEDYDLSKMRDFINKQADAYVEKGILDKEEAE	AIKRIYSSL
SgIII	830.0125	5	6	4145.0262	-0.0002	95	23	48	1.40E-06	IEWLKKHDKK	GNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAI	KRIYSSL
SgIII	999.5105	5	6	4992.5178	-0.0018	177		50	1.10E-14	IEWLKKHDKK	GNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL	C-term
SgIII	591.8170	2		1181.6193	0.0001	47	23	44	0.026	TEAYLEAIRK	NIEWLKKHD	KKGNKEDYDL
SgIII	866.9154	2		1731.8163	-0.0001	52	27	45	0.014	KIEKERQSIR	SSPLDNKLNVEDVDST	KNRKLIDDYD
SST	480.2160	2		958.4178	-0.0004	31	26	42	0.82	QDEMRLELQR	SANSNPAMAP	RERKAGCKNF
SST	622.7879	2		1243.5615	-0.0002	74	20	42	3.30E-05	QDEMRLELQR	SANSNPAMAPRE	RKAGCKNFFW
TMSB10	987.7114	5	6, 7	4933.5229	-0.0022	131		49	3.80E-10	M	Ac-ADKPDMGEIASFDKAKLKKTETQEKNTLPTKETIEQEKRSEIS	C-term
TMSB10	825.9273	6		4949.5179	0.0021	47		49	0.089	M	Ac-ADKPDMGEIASFDKAKLKKTETQEKNTLPTKETIEQEKRSEIS	C-term
TMSB4X	1241.1269	4	5, 6, 7	4960.4862	-0.0077	100		49	5.10E-07	M	Ac-SDKPDMAEIEKFDKSKLKKTETQEKNPLPSKETIEQEKQAGES	C-term
TMSB4X	996.3027	5		4976.4811	-0.0041	93		49	2.20E-06	M	Ac-SDKPDMAEIEKFDKSKLKKTETQEKNPLPSKETIEQEKQAGES	C-term
VGF	892.9472	4		3567.7594	0.0003	47	23	48	0.074	Signal	APPGRPEAQPPPLSSEHKEPVAGDAVPGPKDGSAPE	VRGARNSEPQ
VGF	1223.2623	3	4, 5	3666.8278	-0.0027	72		48	0.00024	Signal	APPGRPEAQPPPLSSEHKEPVAGDAVPGPKDGSAPEV	RGARNSEPQD
VGF	776.9976	5		3879.9504	0.0014	23	18	49	22	Signal	APPGRPEAQPPPLSSEHKEPVAGDAVPGPKDGSAPEVRG	ARNSEPQDEG
VGF	988.7539	4	5, 6	3950.9875	-0.0012	119	23	49	5.40E-09	Signal	APPGRPEAQPPPLSSEHKEPVAGDAVPGPKDGSAPEVRGA	RNSEPQDEGE
VGF	779.8765	2		1557.7383	0.0002	53	23	45	0.0078	YPGREAQARR	AQEEAEAEERRLQ	EQEELENYIE
VGF	927.4661	4	5, 6	3705.8346	0.0009	112		48	2.40E-08	YPGREAQARR	AQEEAEAEERRLQEQEELENYIEHVLLRRP	C-term
VGF	542.5338	4		2166.1069	-0.0010	39	16	46	0.3	EAEAEERRLQ	EQEELENYIEHVLLRRP	C-term
VGF	807.0214	3	2	2418.0418	0.0005	101	28	41	4.80E-08	GARQRGLGGR	GLQEAAEERESAREEEEAEQE	RRGGEERVGE
VGF	507.0039	4	3	2023.9877	-0.0011	44	27	46	0.098	LQPPSALRRR	HYHHALPPSRHYPGREA	QARRAQEEAE
VGF	718.3561	3	4	2152.0463	0.0002	56	20	46	0.005	LQPPSALRRR	HYHHALPPSRHYPGREAQ	ARRAQEEAEA
VGF	742.0357	3	4	2223.0834	0.0018	63	21	47	0.0014	LQPPSALRRR	HYHHALPPSRHYPGREAQA	RRAQEEAEAE
VGF	537.6075	3		1609.8014	-0.0006	19	15	46	28	LQPPSALRRR	HYHHALPPSRHYP-NH2	GREAQARRAQ

TABLE I—*continued*

VGF	543.5572	4	3	2170.2011	-0.0015	69	21	44	0.00019	IEEVEEKRKR	KKNAPPEPVPPPRAAPAPTHV	RSPQPPPPAP
VGF	602.8200	4		2407.2495	0.0015	68	19	47	0.00043	QEEAEAEERR	LQEQEELENYIEHVLLRRP	C-term
VGF	909.6714	5	6	4543.3210	-0.0002	145		49	1.50E-11	PPAPSQFQAR	MPDSGPLPETHKFGEGVSSPKTHLGEALAPLSKAYQGVAAPFPK	ARRPESALLG
VGF	923.8788	5	6	4614.3581	-0.0002	118	22	49	7.60E-09	PPAPSQFQAR	MPDSGPLPETHKFGEGVSSPKTHLGEALAPLSKAYQGVAAPFPKA	RRPESALLGG
VGF	585.8179	2		1169.6193	0.0019	43	23	46	0.11	EVEEKRKRKK	NAPPEPVPPPR	AAPAPTHVRS
VGF	839.9493	2		1677.8838	0.0002	83	28	46	1.10E-05	EVEEKRKRKK	NAPPEPVPPPRAAPAPT	HVRSPQPPPP
VGF	958.0125	2	3	1914.0112	-0.0008	89	22	46	3.30E-06	EVEEKRKRKK	NAPPEPVPPPRAAPAPTHV	RSPQPPPPAP
VGF	958.9349	2		1915.8548	0.0005	35	18	43	0.35	GSAPEVRGAR	NSEPQDEGELFQGVDPR	ALAAVLLQAL
VGF	828.8768	2		1655.7387	0.0002	77	23	43	2.10E-05	RDFSPSSAKR	<QQETAAAETETRTHT	LTRVNLESPG
VGF	747.3683	3		2239.0829	0.0001	48	17	47	0.043	RDFSPSSAKR	<QQETAAAETETRTHTLTRVN	LESPGPERVW
VGF	848.4226	4	3	3389.6600	0.0011	93	19	48	1.70E-06	RDFSPSSAKR	<QQETAAAETETRTHTLTRVNLESPGPERVW	RASWGEFQAR
VGF	629.9729	3		1886.8970	-0.0002	23	14	46	9.3	RDFSPSSAKR	QQETAAAETETRTHTLT	RVNLESPGPE
VGF	969.4860	3		2965.4377	-0.0017	72	17	47	0.00021	RDFSPSSAKR	QQETAAAETETRTHTLTRVNLESPGPE	RVWRASWGEF
VGF	852.6789	4	3,5	3406.6865	-0.0002	118	22	48	5.70E-09	RDFSPSSAKR	QQETAAAETETRTHTLTRVNLESPGPERVW	RASWGEFQAR
VGF	572.2870	3		1713.8394	-0.0002	54	25	46	0.009	HYPGREAQAR	RAQEEAEAEERRLQ	EQEELENYIE
VGF	773.3947	5		3861.9357	0.0013	56		49	0.0099	HYPGREAQAR	RAQEEAEAEERRLQEQEELENYIEHVLLRRP	C-term
VGF	641.8454	4		2563.3506	0.0018	59	22	46	0.0032	AQEEAEAEER	RLQEQEELENYIEHVLLRRP	C-term
VGF	442.6910	2		883.3672	0.0003	32	25	41	0.41	DGEAGAEDKR	SQEETPGH	RRKEAEGTEE
VGF	727.3609	2		1452.7069	0.0003	17	16	44	32	DGEAGAEDKR	SQEETPGHRRKE	AEGTEEGGEE
VGF	698.0365	3		2091.0861	0.0016	40	17	46	0.24	ETAAAETETR	THTLTRVNLESPGPERVW	RASWGEFQAR
VGF	691.8575	2		1381.6990	0.0014	35	24	46	0.88	TETRTHTLTR	VNLESPGPERVW	RASWGEFQAR
VGF	577.6458	3		1729.9151	0.0003	37	21	45	0.44	RASWGEFQAR	VPERAPLPPPAPSQFQ	ARMPDSGPLP

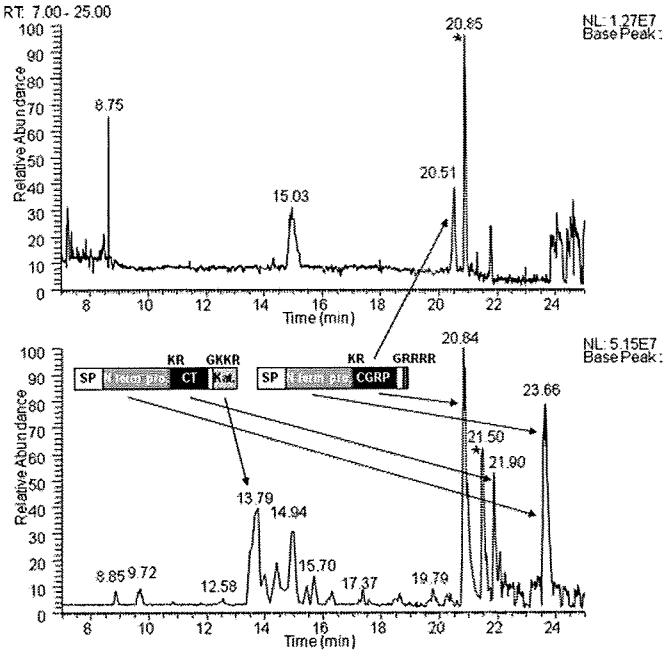


FIG. 3. **Representative base peak chromatograms of the secretopeptidome from unstimulated (*top*) and stimulated (*bottom*) cells.** Samples without gel filtration chromatography were analyzed. Major processing products of CT and CGRP precursors are illustrated along with *arrows* pointing at their peaks in the chromatogram. The base peaks marked with *asterisks* at 20.85 min (unstimulated) and 21.50 min (stimulated) are intact ubiquitin. *SP*, signal peptide; *N-term pro.*, N-terminal propeptide; *Kat.*, katacalcin. Base peaks at 8.75 and 15.03 min in the *top panel* were unrelated to peptide signals. *RT*, retention time; *NL*, normalized ion intensity.

With regard to the SST precursor, the 1243.56-Da peptide corresponds to the first 12 amino acids of SST-28. The dibasic sites RK (position 78–79) located downstream of this 12-residue peptide (Table II) is the known cleavage site for SST-14 (AGCKNFFWKTFTSC) (20). The single arginine (position 65) flanking the 1243.56-Da peptide and SST-28 is also an established processing site (20). At position −4 relative to this scissible bond (referred to as the P4 position), a single arginine (position 62) exists and forms a consensus PC cleavage site. As for the GRP precursor, the 1599.87-Da peptide is C-terminally flanked by an atypical single arginine, which is followed by intact neuromedin C (21). This arginine is known as an established processing site, although responsible enzymes remain to be identified. Thus, a majority of the cleavage sites of peptide hormone precursors were consistent with the previously identified processing sites. Except for precursor C termini, all the peptides derived from granin-like precursors (SgII, SgIII, and VGF) retained informative cleavage sites defined in this study. The 2677.41-Da VGF-derived amidated peptide was recently identified and designated NERP-1 (16).

Integrity of the Secretopeptidome Demonstrated by an In-depth Analysis—The secretopeptidome was separated into four fractions using gel filtration HPLC to perform an in-depth analysis. This analysis contributed to a substantial increase in sequenced peptides, and thus we were able to identify a total of 400 peptides from 23 precursors (Fig. 4 and the supplemental table). Some peptides arose from precursors not identified by the one-dimensional analysis; these included peptide hormone precursors neuromedin U and ghrelin, processing enzymes PC1 and peptidyl-glycine α-amidating monooxygenase, and the calcium-binding protein calnuc. The identification of PC1- and peptidyl-glycine α-amidating monooxygenase-derived peptides confirmed the integrity of this secretopeptidome as they are enzymes involved in the RSP proteolytic processing. Four calnuc-derived peptides had typical cleavage sites suggestive of the PC function (supplemental table). Recently calnuc was identified in a soluble fraction of bovine adrenal secretory granules (13). Altogether it is likely that nearly 99% of sequenced peptides were released upon exocytosis, which again demonstrates that this secretopeptidome is extremely rich in peptides stored in secretory granules.

We examined a total of 400 sequenced peptides to see whether they meet the criteria for the PC consensus sites (supplemental table). PC consensus sites were found in 299 peptides, and signal sequence cleavage sites were found in 43

TABLE II

Identity of the major peptides released on exocytosis (base peak intensities beyond 2e+06 in the Fig. 3 base peak chromatogram)

Values in brackets represent those obtained by peptides after reductive alkylation. For the ubiquitin MSMS spectrum, see supplemental Fig. 2. Note that the MSMS spectra of the 1278.67- and 5687.91-Da peptides (supplemental Fig. 3) did not meet the homology threshold criteria but were considered identified as described in the text. Grayscale boxes are defined in Table I legend. Oxidized methionine residues are underlined in column 10. TMSB4X, thymosin β-4 X-linked; TMSB10, thymosin β-10; RT, retention time; obsd, observed; Mr(Calc), theoretical monoisotopic molecular mass (Da) based on the peptide sequence; Expect., Expectation; N-term, the N-terminal flanking 10 amino acids; C-term, the C-terminal flanking 10 amino acids.

RT (min)	m/z (obsd)	z	Mr(Calc)	Mass error (ppm)	Base peak intensity	Score	Expect. value	N-term	Sequence	C-term	Pre-cursor	Validated by
8.85	622.7882	2	1243.5615	0.27	4.33E+06	74	3.30E-05	QDEMRLELQR	SANSNPAMAPRE	RKAGCKNFFW	SST	
9.72	504.5915	3	1510.7528	-0.09	4.74E+06	48	0.025	Signal	FPKPGGSQDKSLHN	RELSAERPLN	SgIII	
12.55	613.7739	4	2451.0655	0.40	2.78E+06	59	0.0009	IGVGAPGKKR	DMSSDLERDHRPHVSMPQNAN	C-term	CT	
13.51	639.0115	3	1914.0112	0.76	1.17E+07	89	0.0000033	EVEEKRKRKK	NAPPEPVPPPRAAPAPTHV	RSPQPPPPAP	VGF	
13.59	609.7757	4	2435.0706	1.26	1.29E+07	65	0.00031	IGVGAPGKKR	DMSSDLERDHRPHVSMPQNAN	C-term	CT	
13.68	751.0048	3	2249.9906	0.87	4.65E+06	45	0.03	IGVGAPGKKR	DMSSDLERDHRPHVSMPQN	AN	CT	
13.79	659.5054	6	3950.9875	0.31	1.96E+07	119	5.40E-09	Signal	APPGRPEAQPPPLSSEHKEPVAGDAVPGPKDGSAPEVRGA	RNSEPQDEGE	VGF	
14.01	774.6837	3	2321.0277	0.67	7.48E+06	65	0.00033	IGVGAPGKKR	DMSSDLERDHRPHVSMPQNA	N	CT	
14.42	734.3739	5	3666.8278	1.45	9.82E+06	72	0.00024	Signal	APPGRPEAQPPPLSSEHKEPVAGDAVPGPKDGSAPEV	RGARNSEPQD	VGF	
14.94	827.7560	6	4960.4862	1.23	1.58E+07	100	5.10E-07	M	SDKPDMAEIEKFDKSKLKKTETQEKNPLPSKETIEQEKQAGES	C-term	TMSB4X	
14.94	633.3134	2	1264.6122	0.03	2.66E+06	69	0.00022	HKQQLERDPR	VKMALQQEGFD	RKKRGYRDIN	PC2	
15.46	823.2621	6	4933.5229	1.22	5.12E+06	131	3.80E-10	M	ADKPDMGEIASFDKAKLKKTETQEKNTLPTKETIEQEKRSEIS	C-term	TMSB10	
15.70	640.3438	2	1278.6721	0.74	7.21E+06	28	0.86	LSRSGGVVKN	NFVPTNVGSKAF-NH2	GRRRDLQA	CGRP	manual
16.14	787.4012	5	3931.9663	0.84	2.37E+06	59	0.0055	Signal	FPKPGGSQDKSLHNRELSAERPLNEQIAEAEEDKI	KKTYPPENKP	SgII	
16.30	682.3453	5	3406.6865	1.06	4.07E+06	118	5.70E-09	RDFSPSSAKR	QQETAAAETETRTHTLTRVNLESPGPERVW	RASWGEFQAR	VGF	
17.37	848.4229	4	3389.6600	0.73	4.51E+06	93	1.70E-06	RDFSPSSAKR	<QQETAAAETETRTHTLTRVNLESPGPERVW	RASWGEFQAR	VGF	
17.60	762.0490	3	2283.1230	0.95	2.56E+06	95	8.50E-07	GSQDKSLHNR	ELSAERPLNEQIAEAEEDKI	KKTYPPENKP	SgII	
18.64	800.9424	2	1599.8695	0.46	3.70E+06	50	0.011	Signal	VPLPAGGGTVLTKMYP	RGNHWAVGHL	GRP	
19.65	632.5773	4	[2642.3381]	[-1.20]	2.39E+06	[98]	[2.50E-07]	EGSRIIAQKR	ACDTATCVTHRLAGLLSRSGGVVKN	NFVPTNVGSK	CGRP	
19.79	819.3663	2	[1752.7753]	[-0.79]	4.71E+06	[70]	[4.90E-05]	NPAMAPRERK	AGCKNFFWKTFTSC	C-term	SST	
19.91	881.7797	3	2642.3187	-0.55	2.47E+06	106	8.60E-08	Signal	APFRSALESSPADPATLSEDEARLL	LAALVQDYVQ	CT/CGRP	
20.25	577.2836	3	[1844.8873]	[-1.10]	3.57E+06	[98]	[1.20E-07]	EGSRIIAQKR	ACDTATCVTHRLAGLLS	RSGGVVKNNF	CGRP	
20.33	711.9964	8	5687.9079	0.89	2.13E+06	44	0.19	Signal	ERPVFTNHFLVELHKGGEDKARQVAAEHGFGVRKLPFAEGLYHFYHNGLA	KAKRRRSLHH	PC2	manual
20.37	1088.2665	4	4349.0419	-1.15	3.73E+06	112	1.50E-08	EIINSNQVKR	VPGQGSSEDDLQEEEQIEQAIKEHLNQGSSQETDKLAPVS	KRFPVGPPKN	SgII	
20.60	1144.8639	3	[3547.6323]	[-1.68]	2.29E+06	[62]	[0.00049]	SSLDSPRSKR	CGNLSTCMLGTYTQDFNKFHTFPQTAIGVGAP-NH2	GKKRDMSSDL	CT	
20.84	758.3990	5	[3902.9996]	[-1.10]	5.05E+07	[117]	[2.80E-09]	EGSRIIAQKR	ACDTATCVTHRLAGLLSRSGGVVKNNFVPTNVGSKAF-NH2	GRRRDLQA	CGRP	
21.16	830.0130	5	4145.0262	0.58	2.61E+06	95	1.40E-06	IEWLKKHDKK	GNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAI	KRIYSSL	SgII	
21.50	714.3080	12	8559.6167	-0.94	3.09E+07	8	2.20E+02		intact ubiquitin	C-term	Ubq	standard
21.52	742.1743	5	3705.8346	0.13	2.51E+07	112	2.40E-08	YPGREAQARR	AQEEAEAEERRLQEQEELENYIEHVLLRRP	C-term	VGF	
21.55	893.4781	3	2677.4147	-0.82	4.27E+06	69	0.00031	GVAAPFPKAR	RPESALLGGSEAGERLLQQGLAQVEA-NH2	GRRQAEATRQ	VGF	
21.90	1139.5341	3	[3531.6374]	[-1.57]	2.68E+07	[78]	[1.30E-05]	SSLDSPRSKR	CGNLSTCMLGTYTQDFNKFHTFPQTAIGVGAP-NH2	GKKRDMSSDL	CT	
21.93	1158.8690	3	[3589.6429]	[-1.88]	9.68E+06	[64]	[0.0003]	SSLDSPRSKR	CGNLSTCMLGTYTQDFNKFHTFPQTAIGVGAPG	KKRDMSSDLE	CT	
22.25	833.0936	6	4992.5178	0.02	7.83E+06	177	1.10E-14	IEWLKKHDKK	GNKEDYDLSKMRDFINKQADAYVEKGILDKEEAEAIKRIYSSL	C-term	SgII	
23.66	1037.1808	6	6217.0396	0.24	3.95E+07	143	2.50E-11	Signal	APFRSALESSPADPATLSEDEARLLLAALVQDYVQMKASELEQEQEREGSSLDSPRS	KRCGNLSTCM	CT	
23.69	1010.3493	6	6056.0436	1.41	3.91E+07	156	1.60E-12	Signal	APFRSALESSPADPATLSEDEARLLLAALVQDYVQMKASELEQEQEREGSRIIAQ	KRACDTATCV	CGRP	

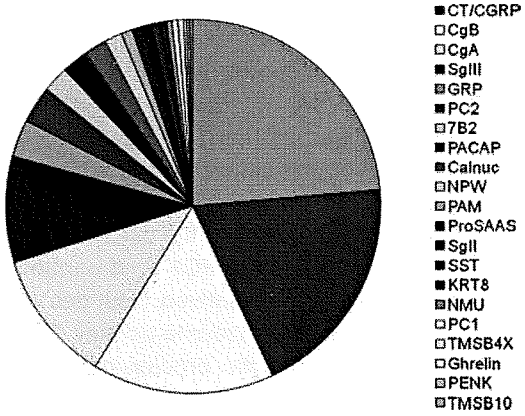


FIG. 4. **Pie representation of the 400 peptides (sorted by names of 23 precursors) secreted after stimulation.** *PACAP*, pituitary adenylate cyclase-activating polypeptide; *NPW*, neuropeptide W; *PENK*, proenkephalin A; *KRT*, cytokeratin; *TMSB4X*, thymosin β-4 X-linked; *TMSB10*, thymosin β-10; *PAM*, peptidyl-glycine α-amidating monooxygenase; *NMU*, neuromedin U.

peptides. Monobasic sites were found in 98 peptides. Overall a total of 373 peptides were shown to have informative cleavage sites to predict the precursor processing in the RSP. As an example, the sequences of CT-derived peptides listed in Tables I and II were mapped to the precursor sequence (Fig. 5). The map consists of the known major processing products and intermediate products sharing N or C termini, which correspond to the established PC1/3 or PC2 cleavage sites. Consistent with the reported length of its signal peptide (17), no peptide was identified in the first 25 amino acids. Overall this map is reminiscent of CT precursor processing previously elucidated in parafollicular cells of the thyroid gland (17).

DISCUSSION

The most outstanding finding in the present study is that an exocytotic stimulus applied to cultured endocrine cells is highly effective in identifying secretory peptides. Peptide profiles identified with this protocol strongly suggest that most peptides were released from secretory granules on exocytosis (Fig. 4). It should be noted that in a total of 400 identified peptides nearly 97% arose from previously known RSP precursor proteins. This non-invasive approach dispenses with time-consuming procedures such as subcellular fractionation and can be extended to different cell culture models. To the best of our knowledge, this is the first study ever to conduct a comprehensive analysis focused on peptides from the RSP proteins.

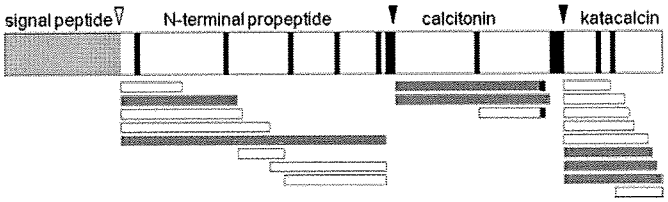


FIG. 5. **CT precursor processing deduced by a panel of identified peptides.** The established signal peptide cleavage site (*open arrowhead*) and known processing sites (*closed arrowheads*) are shown across the *top* of the CT precursor with basic residues (*thin black boxes*) and the signal peptide (*hatched box*) indicated. Sequences of identified peptides (*bars*) are detailed in Tables I and II. Major peptides, defined in Table II, are indicated by *gray bars*. *Black boxes* denote C-terminally amidated residues.

Most peptidomics studies have dealt with endocrine organs or brains, which are considered major sources of peptide hormones or neuropeptides (4, 7–11). It is now recognized that endogenous proteases must be inactivated before tissue extraction to prevent massive production of peptide fragments caused by degradation of abundant intracellular proteins, which hampers MS detection of endogenous peptides (7, 8). For tissue peptidome studies, microwave irradiation before or after decapitation has been proposed to prevent protease activation that is thought to occur immediately after sacrifice (7, 8). Despite these efforts, a limited number of secretory peptides were identified (4, 7–11).

Regarding MSMS of naturally occurring peptides, precursor mass acquisition with conventional mass spectrometers (mass accuracy, ~50 ppm) and subsequent filtering in the setting of "no enzyme" very often lead to ambiguous peptide identification (22). To cope with this issue, many peptidomics studies have used in-house databases containing a limited number of entries to identify peptides that would otherwise remain elusive (11, 23). On the other hand, we used the public IPI human database and took into account four variable modifications simultaneously for MSMS interpretation. Because of the mass accuracy of Orbitrap (~2 ppm), a false discovery rate using a decoy database was minimized to 0% for peptide matches above identity thresholds in the Mascot MSMS ion search. This MSMS identification scheme may inevitably miss many peptides, which could be considered identified if a specific database with limited entries was used as in previous peptidomics studies (11, 23). The peptides listed in the supplemental table are all beyond identity thresholds; accepting peptides beyond homology thresholds will allow more than 200 additional secretory peptides to enter the table, with only two different peptides from keratin 8 turning up (data not shown). These examples include neuromedin U 25 (3018.52 Da) and pancreastatin (5076.36 Da) with a Mascot expectation value of 0.063 and 0.43, respectively. In the supplemental table, they are not considered identified because an expectation value for accepting MSMS spectra is the Mascot default significant threshold value of 0.05. Nonetheless we used the stringent setting (described under "Experimental Procedures") to preclude misleading assignments and to demonstrate that this secretopeptidome shows little contamination by non-secretory components.

In the present study, we made every effort to prevent peptide degradation or chemical modifications (deamidation, methionine oxidation, and pyroglutamination) that may occur during sample preparation. As described under "Experimental Procedures," peptides released during 2 min were immediately subjected to solid phase extraction. Peptide extraction was performed at 4 °C and completed within 20 min after harvesting the supernatant. In addition, lyophilized samples were analyzed by LC-MSMS immediately after reconstitution. Even with these attentive procedures, several clusters of N- or C-terminally truncated peptides that share cleavage sites at the other end were sequenced as reported in previous peptidomics studies on tissue peptidomes (4, 7–11). However, most peptides (30 of 35) dominantly detected in LC-MSMS (Table II, Fig. 3, and supplemental Fig. 1) did not have cleavage sites suggestive of exopeptidase digestion aside from the five CT- or CGRP-derived peptides mentioned under "Results." These five peptides appeared to be N- or C-terminally truncated peptides of major processing products. They have not been reported as major processing products in previous biochemical studies to the best of our knowledge. However, the possibility that their C termini or N termini were generated by unknown endopeptidases (cutting within a peptide) cannot be excluded.

It was unexpected that thymosins and ubiquitin represented major peptides in the TT secretopeptidome. Given a previous report of its storage in adrenal secretory granules and exocytosis-induced secretion (24), ubiquitin may be localized in TT secretory granules and secreted upon exocytosis. The secretory nature of thymosin β-4 has also been reported (25); however, thymosins are not regarded as peptides localized in secretory granules. In any case, the successful identification of intact peptide forms indicates that a majority of peptides may not be affected by exopeptidase digestion.

Gel filtration-based separation caused an increase in the number of sequenced peptides among which several peptides suggestive of the PC-mediated cleavage were identified, such as those from the calcium-binding protein calnuc (supplemental table). Calnuc was identified in a soluble fraction of bovine adrenal secretory vesicles (13), and hence our finding suggests that it is a precursor to unknown bioactive peptides.

To identify unique cleavage sites of RSP precursor proteins, we examined N- and C-terminal flanking sequences of the 400 peptides identified. Overall 152 unique cleavage sites that match PC consensus sites were elucidated of which 105 cleavage sites were conserved consecutive dibasic residues. This finding appears to support the contention revealed by previous studies that the most often encountered PC cleavage sites are conserved paired dibasic sites (1, 2, 19). The

observation that a majority of cleavage sites were consistent with PC consensus sites ((R/K)X_n(R/K) where n = 0, 2, 4, or 6) should not be overestimated. For instance, the C-terminal cleavage site of the CGRP precursor (RLAGLLS ↓ RSG) matches this rule but is not known as a processing site. At present its abundance relative to the longer major product of intact CGRP was unavailable, and therefore it remains to be clarified whether the peptide would represent a major processing product secreted by TT cells. We should also consider that shorter peptides tend to be better ionized and readily detected in mass spectrometry schemes.

Conversely cleavages at non-consensus monobasic sites could represent *bona fide* initial endoproteolytic cleavage sites. In the present study, CgA-derived peptides (1048.51 and 1161.59 Da) sharing the N-terminal cleavage site (FRAR ↓ AYGF) were identified (supplemental table), suggesting that the single arginine (position 356) is a processing site. The C-terminal cleavage site of the 1914.01-Da peptide (VGF residues 463–481) had an arginine at the P10 position. Indeed they have both been shown to be an authentic recognition site for PC2 using PC2 knock-out mice (26). Another example is the N-terminal cleavage site (LSFR ↓ ARAY) of the CgA 1275.65- and 1388.73-Da peptides known as the processing site for CgA LF-19 peptide (27). Thus atypical monobasic sites should also be considered potential processing sites for a precursor whose processing remains largely unknown. In this context, the single arginine of VGF (position 212) in FQAR ↓ MPDS may represent a cleavage site as shown by six peptides (supplemental table). It is envisaged that further detailed analysis of the secretopeptidome could identify processing sites with higher confidence. Although the peptide repertoire from a cancer cell line does not necessarily reflect the *in vivo* processing pattern of RSP precursor proteins, this will not detract from the significance of our study. Indeed the peptides identified with our approach retain cleavage sites created in the RSP to a degree that allows the accurate prediction of processing sites in known peptide hormone precursors (Fig. 5). In summary, we showed that peptidomics has the potential to identify processing sites of precursors processed in the RSP. By dissecting the secretopeptidome we should have a clearer picture of the precursor processing that actually occurs in the RSP that would also facilitate the discovery of bioactive peptides.

Acknowledgment—We thank Gary S. Goldberg (University of Medicine and Dentistry of New Jersey) for critical reading of the manuscript.

* This work was supported in part by the Program for Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation and by grants-in-aid for scientific research from the Japanese Society for the Promotion of Science, Japan.

[S] The on-line version of this article (available at http://www.mcponline.org) contains supplemental material.

§ To whom correspondence may be addressed. Tel.: 81-6-6833-5004; Fax: 81-6-6835-5349; E-mail: ksasaki@ri.ncvc.go.jp.

‖ Present address: Pharmaceutical Research Division, Discovery Research Center, Takeda Pharmaceutical Co. Ltd., 17-85 Jusohonmachi 2-chome, Yodogawa-ku, Osaka 532-8686, Japan.

** To whom correspondence may be addressed. Tel.: 81-6-6833-5004; Fax: 81-6-6835-5349; E-mail: minamino@ri.ncvc.go.jp.

REFERENCES

1. Zhou, A., Webb, G., Zhu, X., and Steiner, D. F. (1999) Proteolytic processing in the secretory pathway. *J. Biol. Chem.* **274,** 20745–20748
2. Fricker, L. D. (2005) Neuropeptide-processing enzymes: applications for drug discovery. *AAPS J.* **7,** E449–455
3. Brakch, N., Rholam, M., Boussetta, H., and Cohen, P. (1993) Role of beta-turn in proteolytic processing of peptide hormone precursors at dibasic sites. *Biochemistry* **32,** 4925–4930
4. Clynen, E., Baggerman, G., Veelaert, D., Cerstiaens, A., Van der Horst, D., Harthoorn, L., Derua, R., Waelkens, E., De Loof, A., and Schoofs, L. (2001) Peptidomics of the pars intercerebralis-corpus cardiacum complex of the migratory locust, Locusta migratoria. *Eur. J. Biochem.* **268,** 1929–1939
5. Schrader, M., and Schulz-Knappe, P. (2001) Peptidomics technologies for human body fluids. *Trends Biotechnol.* **19,** S55–60
6. Sasaki, K., Sato, K., Akiyama, Y., Yanagihara, K., Oka, M., and Yamaguchi, K. (2002) Peptidomics-based approach reveals the secretion of the 29-residue COOH-terminal fragment of the putative tumor suppressor protein DMBT1 from pancreatic adenocarcinoma cell lines. *Cancer Res.* **62,** 4894–4898
7. Svensson, M., Sköld, K., Svenningsson, P., and Andren, P. E. (2003) Peptidomics-based discovery of novel neuropeptides. *J. Proteome Res.* **2,** 213–219
8. Che, F. Y., Lim, J., Pan, H., Biswas, R., and Fricker, L. D. (2005) Quantitative neuropeptidomics of microwave-irradiated mouse brain and pituitary. *Mol. Cell. Proteomics* **4,** 1391–1405
9. Boonen, K., Baggerman, G., D'Hertog, W., Husson, S. J., Overbergh, L., Mathieu, C., and Schoofs, L. (2007) Neuropeptides of the islets of Langerhans: a peptidomics study. *Gen. Comp. Endocrinol.* **152,** 231–241
10. Cape, S. S., Rehm, K. J., Ma, M., Marder, E., and Li, L. (2008) Mass spectral comparison of the neuropeptide complement of the stomatogastric ganglion and brain in the adult and embryonic lobster, Homarus americanus. *J. Neurochem.* **105,** 690–702
11. Bora, A., Annangudi, S. P., Millet, L. J., Rubakhin, S. S., Forbes, A. J., Kelleher, N. L., Gillette, M. U., and Sweedler, J. V. (2008) Neuropeptidomics of the supraoptic rat nucleus. *J. Proteome Res.* **7,** 4992–5003
12. Wegrzyn, J., Lee, J., Neveu, J. M., Lane, W. S., and Hook, V. (2007) Proteomics of neuroendocrine secretory vesicles reveal distinct functional systems for biosynthesis and exocytosis of peptide hormones and neurotransmitters. *J. Proteome Res.* **6,** 1652–1665
13. Brunner, Y., Couté, Y., Iezzi, M., Foti, M., Fukuda, M., Hochstrasser, D. F., Wollheim, C. B., and Sanchez, J. C. (2007) Proteomics analysis of insulin secretory granules. *Mol. Cell. Proteomics* **6,** 1007–1017
14. Gkonos, P. J., Born, W., Jones, B. N., Petermann, J. B., Keutmann, H. T., Birnbaum, R. S., Fischer, J. A., and Roos, B. A. (1986) Biosynthesis of calcitonin gene-related peptide and calcitonin by a human medullary thyroid carcinoma cell line. *J. Biol. Chem.* **261,** 14386–14391
15. Zabel, M., Seidel, J., Kaczmarek, A., Surdyk-Zasada, J., Grzeszkowiak, J., and Górny, A. (1994) Hybridocytochemical and immuno-ultrastructural study of calcitonin gene expression in cultured medullary carcinoma cells. *Histochemistry* **102,** 323–327
16. Yamaguchi, H., Sasaki, K., Satomi, Y., Shimbara, T., Kageyama, H., Mondal, M. S., Toshinai, K., Date, Y., González, L. J., Shioda, S., Takao, T., Nakazato, M., and Minamino, N. (2007) Peptidomic identification and biological characterization of neuroendocrine regulatory peptide-1 and -2. *J. Biol. Chem.* **282,** 26354–26360
17. Jacobs, J. W., Goodman, R. H., Chin, W. W., Dee, P. C., Habener, J. F., Bell, N. H., and Potts, J. T., Jr. (1981) Calcitonin messenger RNA encodes multiple polypeptides in a single precursor. *Science* **213,** 457–459
18. Cameron, A., Apletalina, E. V., and Lindberg, I. (2002) *The enzymology of PC1 and PC2.* In *The Enzymes* (Dalby, R. E., and Sigman, D. S., eds) 3rd Ed., pp. 291–328, Academic Press, NY
19. Eipper, B. A., Stoffers, D. A., and Mains, R. E. (1992) The biosynthesis of neuropeptides: peptide alpha-amidation. *Annu. Rev. Neurosci.* **15,** 57–85