COMMENTARY

高性能マトリックス支援レーザー脱離イオン化飛行時間質量分析計を利用したペプチドマスフィンガープリント法によるタンパク質解析能力の検討

Analytical Capability of a High Performance Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometer for Peptide Mass Fingerprinting

佐藤貴弥

Takaya SATOH

日本電子株式会社 JEOL Ltd., Akishima, TOKYO, JAPAN

Protein identification by peptide mass fingerprinting was performed with a high performance matrix-assisted laser desorption/ionization time-of-flight mass spectrometer. The analytic capability was evaluated using two mixtures consisting of three and five tryptic digests with internal and external calibration methods. In both methods, the high mass resolving power contributed to a precise selection of monoisotopic peaks from mass spectra. In the former method, the high mass accuracy facilitated unambiguous protein identification from protein database searches. In the latter method, the high mass accuracy allowed the discrimination of the matched peptides of positive proteins from those of false-positive proteins through the difference in their mass error distributions. The mass error distribution of the matched peptides of positive proteins consisted of about 10 ppm systematic error with a random error of several ppm, whereas that of the matched peptides of false-positive proteins was random within the selected range of the peptide tolerance window.

(Received January 29, 2007; Accepted April 9, 2007)

1. 序 論

質量分析法とタンパク質の配列データベースを利用してタンパク質の同定を行う方法は、大きく三つに分けられる。ペプチドマスフィンガープリント (peptide mass fingerprinting: PMF) 法^{1)~3)}、プロダクトイオンマスフィンガープリント法^{4), 5)}、ペプチドシーケンスタグ法⁶⁾である。これらの方法の共通点は、タンパク質をアミノ酸残基特異性の高い酵素で消化しペプチドへの断片化を行うことである。断片化したペプチド (ペプチド断片) 群を質量分析計で測定し、その結果をデータベース検索用のソフトウェア (データベース検索エンジン) に入力することで、データベースから統計的に確からしいタンパク質の候補を出力として得ることができる。この出力をユーザーが判断し、タンパク質の同定に至る。さて上記三つの方法は、利用する質量分析法により二つに区分される。すなわち、PMF 法はタンデム質量分析を必要としない方法であるのに対し

て、プロダクトイオンマスフィンガープリント法、ペプチドシーケンスタグ法はタンデム質量分析を必要とする方法である。さらにプロダクトイオンマスフィンガープリント法がプロダクトイオンマスペクトルから得られた質量情報をデータベース検索に利用するのにに対し、ペプチドシーケンスタグ法はプロダクトイオンマスペクトルから部分的なアミノ酸配列情報を読み取り、データベース検索に利用するという違いがある。それぞれの方法については吉野らの解説⁷⁾に詳しい。

ここでは、本稿に関連する PMF 法について詳しく述べる。PMF 法では、タンパク質をトリプシンなどに代表されるアミノ酸残基特異性の高い消化酵素で断片化したペプチド断片群を測定する。各ペプチド断片はアミノ酸配列に由来するアミノ酸組成に基づく質量を有するため、質量分析計で測定するとペプチド断片群に由来する質量のリスト (観測質量リスト) が得られる。データベース検索エンジンは、データベース上のタンパク質を仮想的に断片化して得られるペプチド断片群の質量リスト (仮想質量リスト) と入力された観測質量リストを比較し、統計的に有意性の高いタンパク質を出力する。PMF 法では 1 種類のタンパク質の仮想質量リストと観測質量リストの比較が行われるため、酵素消化を行う前にタンパク質の単離が必要となる。

Correspondence to: Takaya SATOH, JEOL Ltd., Musashino 3-1-2, Akishima, Tokyo 196-8558, JAPAN,
e-mail: taksatoh@jeol.co.jp
佐藤貴弥, 日本電子株式会社, 〒196-8558 東京都昭島市武蔵野 3-1-2

プロテオミクスでは高いタンパク質の分離能をもつ2次元電気泳動とマトリックス支援レーザー脱離イオン化(matrix-assisted laser desorption/ionization: MALDI)―飛行時間(time-of-flight: TOF)質量分析計^{8),9)}の組み合わせがよく利用される。

統計的な手法が用いられる主要な理由としては、観測質量リストが仮想質量リストとはさまざまな要因により異なるためである。

第1の要因としては、測定から得られるペプチド断片の質量と仮想的なペプチド断片の質量の計算値が異なるためである。たとえ予想されたペプチド断片のみが観測されたとしても、質量分析計の測定誤差(質量誤差)により観測質量と計算質量はわずかにずれる。質量誤差の大きさは、質量分析計の種類、同種の質量分析計でもメーカーあるいは機種によっても異なる。さらに同じ装置であったとしても、測定条件、質量校正状況といった日々の実験環境にも影響される。データベース検索エンジンには質量誤差について許容範囲を設定するパラメータが存在する。この許容範囲内であれば観測値と計算値は一致した(マッチした)とみなされるため、許容範囲は実験環境に応じてユーザーが適切な値を設定する必要がある。

第2の要因としては、各ペプチド断片が同じ効率で観測されないことである。これは、前処理、イオン化効率、質量分析計の検出効率の質量依存性などが原因である。検出効率の違いにより、マスペクトル上に全く観測されないペプチド断片もあれば、マスペクトル上のシグナル強度(イオン強度)が低く妥当な m/z 値が得られないものもある。データベース検索エンジンには、イオン強度を加味できるものとそうでないものが存在するが、使用する装置の質量誤差の分布がイオン強度、特にシグナルノイズ比に対しどのように変化するかをよく検討し、利用の可否を判断すべきである。

第2の要因では想定されたペプチド断片が観測されない場合が多いが、第3の要因として逆に想定されないピークが観測される場合もある。例えば、タンパク質を断片化する際に生じる切れ残りや意図しない修飾などがこの原因である。また前述のように、PMF法では1種類のタンパク質の仮想質量リストと観測質量リストの比較が行われるためタンパク質の単離が必要である。分離能の高い2次元電気泳動がよく利用されるが、そこから切り出されたゲル片に含まれるタンパク質が単離されているという確証はなく、またたとえそうであったとしてもその後の処理の過程でコンタミネーションが起こる可能性も考えられる。

さて、PMF法を利用できるデータベース検索エンジンには、現在World Wide Web(WWW)上に無料公開されているものがいくつか存在する^{10)~13)}。それぞれのデータベース検索エンジンのアルゴリズムは異なるものの、統計的な有意さを示す指標値順にタンパク質が表示される(ヒットする)。ヒットしたタンパク質については、そのタンパク質の情報、マッチしたペプチド断片がタンパク質のアミノ酸配列をどの程度カバーしているか(シーケンスカ

バレッジ)、マッチしたペプチド断片それぞれについて質量誤差が提供される。データベース検索エンジンの検索結果には、統計的な有意さを示す指標値に加え、統計的に有意であることを示す閾値を表示するものもある。しかしながら、統計的な有意さを示す指標値は、検索条件、マスペクトルの質により大きく左右されるため、有意であることを示す閾値以上のタンパク質を陽性とするには注意が必要である。もちろん、統計的な有意さを示す指標値が高いほど現実的にも確からしいタンパク質であることは事実である。しかし、より確度の高い結果を得るためには、データベース検索により示される結果を他の情報も加味して総合的に判断し、擬陽性と陽性の区別を明確に行うこと、データベース検索による擬陰性を減らすことが重要である。そのためには、タンパク質の断片化処理の工夫、質量分析計の性能向上、データベース検索エンジンのアルゴリズム改良などが必要である。高い質量精度がデータベース検索結果に与える効果については、Clauserらの報告¹⁴⁾がある。また、フーリエ変換イオンサイクロトロン共鳴質量分析計を利用した高い質量精度でのPMFの検討についても報告されている¹⁵⁾。

我々は最近独自に開発したらせん軌道を利用したTOF質量分析計^{16),17)}をベースに、高い質量分解能と高質量精度を実現できるMALDI-TOF質量分析計を新規に開発した¹⁸⁾。通常TOF質量分析計の質量分解能、質量精度は飛行距離に比例する。この装置では、四つの階層状トロイダル電場を用いてらせん軌道を実現し、現在市販されているTOF質量分析計の飛行距離が数m程度であるところを、17mにまで延長した。それでもそのイオン軌道を50cm四方程度の限られた空間に納めることにも成功している。一般的なMALDI-TOF質量分析計では、遅延引き出し法^{19),20)}が利用されており、質量分解能の向上に大きく寄与している。しかしこの方法では、高い質量分解能が限られた m/z 領域でしか実現できない点や、質量精度に関して系統的な質量誤差を生じやすく m/z 500~3,000といったPMF法でよく利用される m/z 領域全体における質量精度の低下も招いている。本装置では、従来MALDI-TOF質量分析計の弱点とされていた問題点を改善し、 m/z 2,500にて質量分解能80,000(半値幅)を実現し、解析対象のマスペクトル上に観測した1点あるいは2点の内部標準ピークを利用して質量校正を行う方法(内部標準法)にて m/z 500~3,000の領域で平均2乗偏差(root mean square: rms) 2 ppm以内の質量精度が達成されている。

本稿では、この高性能MALDI-TOF質量分析計と一般的に最もよく利用されているデータベース検索エンジンであるMatrix Science社mascot¹⁰⁾を利用してPMF法によるタンパク質の同定能力について検討を行った。特に、(i)内部標準法と内部標準物質を利用しない質量校正法(外部標準法)の両方について高性能MALDI-TOF質量分析計の利用が陽性・擬陽性識別へ与える効果、(ii)3ないし5種類のタンパク質消化物の混合物解析の2点に注目

しながら、PMF 法を行う場合の一連の作業や結果の解釈について概説する。

2. サンプル調整

タンパク質消化物の標準試料には Michrom Biore-source 社より購入した, carbonic anhydrase (cow), serum albumin (cow), apotransferrin (cow), catalase (cow), glutamate dehydrogenase (cow) の 5 種類を用いた。1 種類の試料を用いた測定においては, それぞれのサンプルは脱イオン水にて 100 fmol/ μ L に濃度を調整した。また 3 種類のタンパク質消化物の混合物測定には, carbonic anhydrase, serum albumin, catalase の 3 種類の試料を各々 100 fmol/ μ L になるように濃度調整した (以下 Mix3 と呼ぶ)。5 種類のタンパク質消化物の混合物測定には, 上記すべての試料を各々 100 fmol/ μ L になるように濃度調整した (以下 Mix5 と呼ぶ)。

マトリックスには, 和光純薬より購入したプロテオーム研究用 α -シアノ-4-ヒドロキシケイ皮酸 (α -Cyano-4-hydroxycinnamic acid: α -CHCA) を用いた。 α -CHCA は, 10 mg/mL となるように濃度を調整した。作成したサン

プル溶液とマトリックス溶液を等量混合し, ステンレス製のターゲットプレートに 1 μ L 滴下風乾した。ターゲットプレートは, 384 ウェルのマイクロタイタープレート形式であり, 隣接するウェルの中心間距離は, 4.5 mm である。

3. 質量校正およびモノアイソトピックピークの選出

m/z 500~3,000 の領域で同位体イオンの分離が可能な質量分析計を用い, PMF 法によりタンパク質の同定を行う場合, データベース検索エンジンにはモノアイソトピックピークの m/z 値リストを入力する。サンプル測定の前に, 五つの標準ペプチド bradykinin fragment 1~7 (分子量 756.3997), angiotensin II (分子量 1,045.5), angiotensin I (分子量 1,296.5), P14R (分子量 1,532.9), Adrenocorticotrophic hormone fragment 18~39 (分子量 2,464.2) を用いて質量校正を行った。質量校正式は, eq. (1) を用いた¹⁸⁾。

$$\sqrt{m/z} = a + bT + cT^2 \quad (1)$$

ここで T は飛行時間であり, a, b, c は定数である。次に内部標準法にて質量校正する場合は, 2 点の内部標準ピークを利用し, 各ペプチド断片の m/z 値が計算値と一致する

Fig. 1. Mass spectra of the mixture of three standard tryptic digests of serum albumin, catalase, and carbonic anhydrase, called Mix3 in this study. The mass spectrum in the region m/z 1,282–1,294 (a) shows the difference between the chemical background peak (A) and the signal peak (B). The mass spectrum in the region m/z 1,479.5–1,480.0 (b) shows the peak separation of monoisotopic peaks of protonated FNSANDDNVTQVR (m/z 1,479.68226) and LGEYGFQNALIVR (m/z 1,479.79544).

ように式 (1) の a, b を変更した¹⁸⁾。2 点の内部標準物質は、1 種類のタンパク質消化物の場合には、各タンパク質に由来する m/z 1,000 および 2,000 前後のペプチド断片のプロトン化分子のモノアイソトピックピークを利用した。Mix3, Mix5 については、serum albumin の二つのペプチド断片 YLYEIR と RPCFSALTPDETYVVK のプロトン化分子のモノアイソトピックイオン（それぞれの m/z 値は 927.4934 と 1,881.9051）を内部標準物質として質量較正を行った。

モノアイソトピックピークの選出を行う場合には、まずノイズと有意なシグナルを識別しなければならない。ノイズには一般的にベースラインの電氣的なノイズと何らかのイオンに由来するノイズ（ケミカルバックグラウンド）が考えられる。両者の相対的な関係は、サンプルの調整方法、装置条件、測定方法などにより異なると考えられる。Fig. 1(a) に本装置で Mix3 を測定した m/z 1,282~1,294 のマスペクトルを示す。ケミカルバックグラウンドは、 m/z 1 間隔で m/z 0.2 程度の幅の中に検出器の時間応答に起因する m/z 0.01 程度の半値幅をもつピークの集合として観測された。ピークの数が多い場合はこの集合は幅広いピークを形成し（例えば Fig. 1(a) 中ピーク A）、少ない場合は非常に幅の狭いピークのまま観測された。有意なシグナルの場合は、 m/z 0.02~0.03 程度の半値幅をもつピークを形成したので（例えば Fig. 1(a) 中ピーク B）、ケミカルバックグラウンドとの識別は容易であった。データベース検索のための m/z 値リストは、このケミカルバックグラウンドの数倍以上のイオン強度をもつモノアイソトピックピークを選出し作成した。さらにマトリックスのみの測定を行いマトリックス由来のピークを調査することで、データベース検索のための m/z 値リストからそれらを排除した。

より精度の高い同定結果を得るためには、マスペクトルからモノアイソトピックピークを正確に読み取る必要がある。高い質量分解能がより正確にマスペクトルを読むことに貢献できる点として、近接したピークの分離が挙げられる。本装置で 3 種類のタンパク質混合物 Mix3 を測定した m/z 1,479.5~1,480.0 のマスペクトルを Fig. 1(b) に示す。イオン強度差があるもののよく分離された 2 本のピークが観測されている。この 2 本のピークは catalase 由来のペプチド断片 FNSANDDNVTQVR と albumin 由来のペプチド断片 LGEYGFQNALIVR のプロトン化分子のモノアイソトピックピークである。それぞれのモノアイソトピックイオンの m/z 値の計算値は、1,479.68226 と 1,479.79544 である。また、このときの質量分解能は、60,000（半値幅）であった。

このようにマスペクトルには、近接した m/z 値をもつ 2 本以上のピークが現れることがしばしばある。分離が不十分な場合は、イオン強度に差があれば低イオン強度のピーク判定が不可能になり、同等程度の場合両者の m/z 値が大ききずれることとなる。いずれにしても正確な情報が得られない。データベース検索に利用するのは、モノア

イソトピックピークの m/z 値のみであるが、マスペクトル上に観測される同一化合物由来のピークには m/z 500~3,000 の領域であれば数本から 5 本程度の同位体ピークが存在する。質量分解能が不十分である場合、イオン強度の強いイオンが観測されると、その同位体ピークと重なる部分の情報を失う可能性がある。モノアイソトピックピークの選出には、モノアイソトピックピーク同士だけでなく、同位体ピークとの分離も含むので、高い質量分解能はたいへん有効となる。

4. 内部標準法における質量誤差の許容範囲の検討

データベース検索エンジンは、WWW 上に公開されている Matrix Science 社の mascot search の peptide mass fingerprinting を利用した。まず複数種のタンパク質混合物の解析を行う前に、serum albumin の酵素消化物を用いて内部標準法における質量誤差の許容範囲の設定値について検討を行った。mascot search の peptide mass fingerprinting においては、peptide tol. (tolerance) 値が観測質量と仮想質量の許容範囲である。peptide tol. 値以外のデータベース検索条件については固定して検討を行ったので Table 1 にまとめる。peptide tol. 値を設定する際注意しなければならないのは、装置メーカーから提示される質量精度の仕様値をそのまま使用することはできない点である。質量精度の仕様値は「2 ppm rms 内部標準法」といった形式で表記されるが、規定方法はメーカーにより異なり、必ずしも PMF 法のように m/z 500~3,000 といった比較的広い m/z 範囲を利用するアプリケーションを考慮されているとは限らない。また仕様値の規定では、比較的イオン強度の高いピークを用い、統計的に「rms」で表現される場合が多く、PMF 法で利用するイオン強度の低いイオンの m/z 値のばらつき方などは反映されていない可能性が高い。そのため、使用装置、質量較正法に応じて適切な peptide tol. 値を把握することは非常に重要である。

まず peptide tol. 値を ± 20 ppm に設定してデータベース検索を行い、本装置におけるイオン強度と質量誤差の関係について調べた。質量較正に内部標準法を利用したので、データベース検索結果に示されるマッチしたペプチド断片の質量と質量誤差との関係に大きな系統誤差はなかつ

Table 1. Database Search Conditions for Mascot Search

Database	NBCInr
Taxonomy	All
Enzyme	Trypsin
Allowed up to 1 missed cleavage	
Fixed modification	Carboxymethyl (C)
Variable modification	Not selected
Protein mass	Not restricted
Mass values	MH ⁺ (Monoisotopic)

Fig. 2. Correlation between mass errors and the intensities of matched peptides in the mass spectrum. The dotted line indicates the intensity of the chemical background.

た。そこで、Fig. 2 にイオン強度とマッチしたペプチド断片の質量誤差をプロットした。Fig. 2 の横軸は対数表示である。ケミカルバックグラウンドのイオン強度は 0.2 程度であったので、Fig. 2 にあわせて表示した。この結果を見ると、イオン強度がケミカルバックグラウンドの 5 倍を超えたあたりから、質量誤差が ± 4 ppm 程度に収まっていることがわかる。

次に Table 2 に peptide tol. 値を変化させた場合の、データベース検索結果の serum albumin に関する情報をまとめた。mascot search では、統計的な有意さを示す指標値として Probability Based MOWSE Score (スコア) が出力される。WWW 上の mascot search の説明によれ

Table 2. Variation of the Mascot Search Results from the Internally Calibrated Data of the Tryptic Digest of Serum Albumin When Changing the Peptide Tolerance Value

Pep tol.	Score	Matched queries	Sequence coverage
± 3 ppm	261	32	49%
± 5 ppm	334	38	57%
± 8 ppm	393	42	61%
± 10 ppm	408	43	61%
± 20 ppm	408	43	61%
± 30 ppm	406	43	61%
± 50 ppm	403	43	61%
± 100 ppm	390	44	61%

ば、マッチしたペプチド断片数が変わらなければ、peptide tol. 値が小さいほどスコアが高くなるとある。マッチしたペプチド断片数がほぼ変化しない ± 10 ppm から ± 100 ppm の範囲では、peptide tol. 値が小さくなるとスコアに多少の増加傾向はあるものの顕著な変化ではないことがわかった。しかしながら、peptide tol. 値を ± 8 ppm, ± 5 ppm, ± 3 ppm と小さくしていくと、マッチするペプチド断片数が減少しスコアも減少した。これは、 ± 10 ppm の場合には、マッチしていたいくつかのペプチド断片について、質量誤差が許容範囲を超えたためである。

データベース検索による擬陽性を排除することを考えると、マッチしたペプチド断片の大多数が許容誤差範囲外にでない条件で peptide tol. 値をできるだけ小さくしたほう

Table 3. Mascot Search Results from the Internally Calibrated Data of the Mixture of Three Tryptic Digests of Serum Albumin, Catalase, and Carbonic Anhydrase

Hit No.	Accession	Taxonomy	Description	Score	M_r	Matched queries	Sequence coverage
1	gi 30794280	<i>Bos taurus</i>	Albumin	103	7,1309	28	45%
2	gi 74267962	<i>Bos taurus</i>	ALB protein	95	71,221	27	43%
3	gi 229299	<i>Bos taurus</i>	Catalase	58	57,668	16	27%
4	gi 46016008	<i>Bos taurus</i>	Chain A, crystal structure analysis of bovine carbonic anhydrase li	56	28,965	10	38%
5	gi 74223714	<i>Mus musculus</i>	Unnamed protein product	42	59,951	13	24%
6	gi 118582300	<i>Ovis aries</i>	Carbonic anhydrase 2 (carbonic anhydrase II) (carbonate dehydratase II) (CA-II)	39	29,193	8	26%
7	gi 120404891	<i>Mycobacterium vanbaalenii</i> PYR-1	Regulatory protein, ArsR	23	12,183	5	44%
8	gi 119930904	<i>Bos taurus</i>	PREDICTED: similar to ornithine carbamoyltransferase, mitochondrial precursor (OTCase) (ornithine transcarbamylase), partial	23	5,476	3	82%
9	gi 110739783	<i>Arabidopsis thaliana</i>	Hypothetical protein	23	2,584	2	80%
10	gi 86751568	<i>Rhodospseudomonas palustris</i> HaA2	Transcriptional modulator of MazE/toxin, MazF	22	11,816	4	32%

がよい。Fig. 2 および Table 2 の結果からは、本装置を利用する場合、peptide tol. 値は ± 8 から ± 10 ppm が妥当であると考えられる。そこで次章では ± 8 ppm を採用した。

5. 内部標準法による複数種のタンパク質混合物の解析

次に内部標準法により Mix3 および Mix5 について解析を行った。peptide tol. 値は前章の検討結果から得た ± 8 ppm とした。Tables 3, 4 に、Mix3 および Mix5 のデータベース検索結果をそれぞれ Hit No. 10 および 15 まで示す。Table 3 に示す Mix3 の同定結果では、上位 4 位までに *Bos taurus* 由来の serum albumin (Hit Nos. 1, 2), catalase (Hit No. 3), carbonic anhydrase (Hit No. 4) が候補として挙げられた。Hit Nos. 5, 6 は異なる種の catalase および carbonic anhydrase であった。また、Table 4 に示す 5 種類の混合物の同定結果では、Hit Nos. 1~5 および

No. 7 に、サンプルに含まれる *Bos taurus* 由来のタンパク質が候補として挙げられた。Hit Nos. 6, 8, 9, 10, 11 は、異なる種のサンプルに含まれるタンパク質であり、Hit No. 13 は Glutamate Dehydrogenase に関連するものであった。これらの結果から、本装置と mascot search の組み合わせによって 5 種類のタンパク質混合物であっても、解析可能であることが示された。しかしながら、陽性と擬陽性の区別については、単純にスコアだけをみて、ある閾値を設定することは不可能であることもわかる。マッチしたペプチド断片数、シーケンスカバレッジ、分類などを総合的に判断して決定する必要がある。そこで 1 種類のタンパク質解析に比較して、混合物解析が困難になる要素を次に考察した。

まずサンプルに含まれるタンパク質数が増えるほど、マスペクトルが複雑になる。1 種類の場合は、各タンパク質消化物のマスペクトルから 100 本程度のモノアイソ

Table 4. Mascot Search Results Obtained from the Internally Calibrated Data of the Mixture of Five Tryptic Digests of Serum Albumin, Catalase, Carbonic Anhydrase, Apotransferrin, and Glutamate Dehydrogenase

Hit No.	Accession	Taxonomy	Description	Score	M_r	Matched queries	Sequence coverage
1	gi 2501351	<i>Bos taurus</i>	Serotransferrin precursor (transferrin) (siderophilin) (Beta-1-metal-binding globulin)	101	79,907	29	39%
2	gi 231250	<i>Bos taurus</i>	Chain A, catalase (EC 1.11.1.6)	68	57,782	18	39%
3	gi 46016008	<i>Bos taurus</i>	Chain A, crystal structure analysis of bovine carbonic anhydrase II	66	28,965	12	54%
4	gi 74354891	<i>Bos taurus</i>	GLUD1 protein	49	61,998	18	33%
5	gi 30794280	<i>Bos taurus</i>	Albumin	42	71,309	22	35%
6	gi 118582300	<i>Ovis aries</i>	Carbonic anhydrase 2 (carbonic anhydrase II) (carbonate dehydratase II) (CA-II)	36	29,193	8	31%
7	gi 74267962	<i>Bos taurus</i>	ALB protein	36	71,221	21	33%
8	gi 114631609	<i>Pan troglodytes</i>	PREDICTED: similar to GDH isoform 1	34	51,634	13	28%
9	gi 74228849	<i>Mus musculus</i>	Unnamed protein product	31	60,048	11	27%
10	gi 20151189	<i>Homo sapiens</i>	Chain A, structure of human glutamate dehydrogenase-apo form	30	56,321	14	29%
11	gi 6753272	<i>Mus musculus</i>	Catalase	29	59,987	11	25%
12	gi 70921651	<i>Plasmodium chabaudi chabaudi</i>	Hypothetical protein PC401508.00.0	28	4,345	3	100%
13	gi 13096299	<i>Bos taurus</i>	Chain A, crystal structure of bovine liver glutamate dehydrogenase complexed with Gtp, Nadh, and L-glutamic acid	27	55,874	15	30%
14	gi 58261394	<i>Cryptococcus neoformans</i> var. <i>neoformans</i> JEC21	Hypothetical protein CNL05740	26	4,632	3	79%
15	gi 41189584	Bacteriophage 77	77ORF178	25	4,080	3	72%

トピックピークを選び出すことができたが、Mix3 および Mix5 の場合は、どちらもその数は 200~250 程度であった。この結果から、この値がイオン化法を含めた本装置のモノアイソトピックピーク選出数の上限であると考えられる。読み取れるモノアイソトピックピークの数が 250 程度までであるのに対し、有意なシグナルピークが増えた場合には、マスペクトルからモノアイソトピックピークを読み取る段階で見落としが増加することが予想される。Table 5 に 1 種類、3 種類、5 種類の試料を分析した場合にマッチしたペプチド断片数がどのように変化したかを示した。タンパク質ごとにばらつきはあるものの全体的にタンパク質が増えるに従ってマッチしたペプチド断片数は減少傾向にある。特に顕著に減少した serum albumin について調べると、タンパク質数が増えるに従って観測されなくなったペプチド断片ピークのほとんどは質量分離の問題ではなく、イオン強度がケミカルバックグラウンドのそれ以下にまで下がったためであった。これは、サブプレッション効果など MALDI 法に限界があることによるものと考えられる。

次にタンパク質のデータベース検索結果について考察する。Tables 3, 4 を見るとサンプルに含まれるタンパク質

以外に、異なる種の同名のタンパク質が候補として挙げられている。そのため、サンプルに含まれるタンパク質が多くなればなるほど、結果として示される統計的に有意なタンパク質のほとんどがサンプルに含まれるタンパク質もしくはそれと関連のあるものとなる。また、擬陽性について見てみると Table 3, 4 に共通する結果として、分子量 10,000 前後のタンパク質がスコア 20 前後で候補として挙がってくるのがわかる。これは、peptide tol. 値を狭くしていくと、質量誤差の許容範囲が制限されるためマッチするペプチド断片数が減少し、結果として仮想断片数の少ない分子量の小さなタンパク質が擬陽性として候補に挙がってくるためだと考えられる。実際のサンプルにこのような分子量 10,000 以下のタンパク質が含まれる場合、データベース検索からの結果からだけでは、陽性と擬陽性の判断は難しくなることを示唆している。

6. 外部標準法による複数種のタンパク質混合物の同定

MALDI-TOF 質量分析計を利用した場合、サンプルプレート表面上の凹凸やマトリックスの結晶化状態、質量分析計を構成する機械系、電源系の経時的な変化に伴い同じピークでも観測される m/z 値が変動する。そのため内部

Table 5. Variation in the Number of Matched Peptides with Difference of Included Protein Number

	Serum albumin	Carbonic anhydrase	Catalase	Apotransferrin	Glutamate dehydrogenase
Each protein	40	13	22	36	25
Mix3	28	10	16		
Mix5	22	12	18		

Table 6. Mascot Search Results Obtained from the Externally Calibrated Data of the Mixture of Three Tryptic Digests of Serum Albumin, Catalase, and Carbonic Anhydrase

Hit No.	Accession	Taxonomy	Description	Score	M_r	Matched queries	Sequence coverage
1	Mixture 1	<i>Bos taurus</i>	Hit No. 3 + Hit No. 5	235			
2	Mixture 2	<i>Bos taurus</i>	Hit No. 4 + Hit No. 5	226			
3	gi 30794280	<i>Bos taurus</i>	Albumin	146	71,309	36	54%
4	gi 74267962	<i>Bos taurus</i>	ALB protein	137	71,221	35	51%
5	gi 46016010	<i>Bos taurus</i>	Chain C, crystal structure analysis of the site specific mutant (Q253c) of bovine carbonic anhydrase II	81	29,315	13	54%
6	gi 86159730	<i>Anaeromyxobacter dehalogenans</i> 2CP-C	Tetratricopeptide repeat protein	68	442,617	68	17%
7	gi 68484613	<i>Candida albicans</i> SC5314	Putative UDP glucose-4-epimerase	63	75,898	19	27%
8	gi 3776990	Human immunodeficiency virus 1	Envelope protein	59	12,220	8	80%
9	gi 229299	<i>Bos taurus</i>	Catalase	56	57,668	16	27%
10	gi 39594855	<i>Caenorhabditis briggsae</i>	Hypothetical protein CBG17466	55	53,093	16	28%

標準法が有効になるのであるが、内部標準物質をサンプルと混合しない場合、あるいは内部標準物質が観測できない場合には、質量較正することができない。このような場合は、質量誤差は偶然誤差が大きくなるのではなく、系統誤差が生じることになる。外部標準法の場合、この系統誤差を含む程度に peptide tol. 値を大きくする必要がある。一般的には、(i) 測定の前に質量較正を行う、(ii) 質量較正後は質量軸が変化するパラメータを変更しない（質量軸が変化しないパラメータは一般的には検出器電圧やレーザー強度など）、(iii) 近接するウェルに標準物質を置き適宜質量較正を行うなど、実験環境をコントロールすれば系統誤差は、数十 ppm 以内に抑えることができる。

以下では、外部標準法により Mix3 の解析を行った結果を示す。今回は、隣接するウェルに滴下した、serum albumin のサンプルで質量較正を行い、パラメータを変更することなく、Mix3 のマスペクトルを測定した。peptide tol. 値を ± 50 ppm に設定し、データベース検索を行った結果を Table 6 に示す。検索結果の Hit Nos. 1, 2 は、serum albumin および carbonic anhydrase の混合物であると示された。Hit Nos. 3~5 は、serum albumin およ

び carbonic anhydrase であり、Hit No. 9 に catalase が示された。Hit Nos. 6, 7, 8, 10 は擬陽性であるが、スコア、シーケンスカバレッジからだけでは Hit No. 9 の catalase を識別することは難しい。しかしながら、外部標準法では、陽性であれば系統誤差が観測されると予想されるので、Hit Nos. 3~9 のタンパク質についてマッチしたペプチド断片の質量誤差を調査した。Fig. 3(a) にマッチしたペプチド断片の質量と質量誤差の関係を、Fig. 3(b) に -50 ppm から $+50$ ppm まで 10 ppm 間隔でマッチしたペプチド断片数を示した。ただし Fig. 3(b) の縦軸は、各タンパク質候補についてマッチしたペプチド断片数で規格化している。Fig. 3(a) をみると、Hit Nos. 3, 5, 9 は、いくつかの例外はあるものの質量誤差が $+10$ ppm 前後に分布している。それに対し、Hit Nos. 6, 7, 8 については、質量誤差が ± 50 ppm の範囲全体にばらついて分布している。Fig. 3(b) はそれをより明確に示している。すなわち、Hit Nos. 3, 5, 9 では、0 から $+20$ ppm の間にマッチしたペプチド断片の 80% 以上が分布しているのに対し、Hit Nos. 6~8 では ± 50 ppm の範囲全体に分布している。この結果から Hit Nos. 3, 5, 9 は陽性であり、また、Hit Nos. 6~8 が peptide tol. 値を広く設定したことによる擬陽性であることがわかる。また今回の実験では系統誤差が $+10$ ppm 程度であることがわかり、偶然誤差は系統誤差から ± 5 ppm 程度で分布していることがわかる。また、陽性のタンパク質中でマッチしたペプチド断片でも系統誤差から数十 ppm 以上の質量誤差をもつものは、peptide tol. 値を広げたことによるペプチド断片レベルの擬陽性であるといえる。

以上のように高い質量精度の装置を利用することは、外部標準法の場合においても陽性と擬陽性の区別に効果的であることがわかる。ただし、peptide tol. 値を広げるにより、より多くの擬陽性を生むことは事実である。

7. ま と め

本稿では、高性能 MALDI-TOF 質量分析計を利用した、PMF 法によるタンパク質同定能力の検討結果について述べた。質量分解能の向上による正確なマスペクトルからの情報の読み取り、あるいは質量精度の向上による偶然誤差の低減が、陽性と擬陽性の区別に効果的であることを示した。データベースと質量分析法を利用したタンパク質の同定では、サンプルの処理法、質量分析計の性能、データベース検索アルゴリズムの特徴だけではなく、それらの使用条件も最終的な同定結果に影響を与える。統計的な有意さの高いタンパク質が陽性である可能性はもちろん高いが、単純に 1 位のスコアのタンパク質のみを陽性とすることは危険である。より確度の高い結果を得るためには、データベース検索結果により示されるさまざまな情報から陽性と擬陽性の区別を行うことが重要である。

文 献

- 1) M. Mann, P. Højrup, and P. Roepstorff, *Biol. Mass Spec-*

Fig. 3. Mass error distribution of matched peptides of Hit Nos. 3-9. The solid and open points relate to matched peptides of the positive and false-positive proteins, respectively. The relationships between (a) mass errors and mass, and (b) the mass error distribution of matched peptides from -50 to $+50$ ppm in 10 ppm intervals are shown. The vertical axis of (b) is normalized by the total number of matched peptides of each protein.

- trom.*, **22**, 338 (1993).
- 2) J. R. Yates, III, S. Speicher, P. R. Griffin, and T. Hunkapiller, *Anal. Biochem.*, **214**, 397 (1993).
- 3) D. J. C. Pappin, P. Højrup, and A. J. Bleasby, *Curr. Biol.*, **3**, 327 (1993).
- 4) J. K. Eng, A. L. McCormack, and J. R. Yates, III, *J. Am. Soc. Mass Spectrom.*, **5**, 976 (1994).
- 5) D. N. Perkins, D. J. C. Pappin, D. M. Creasy, and J. S. Cottrell, *Electrophoresis*, **20**, 3551 (1999).
- 6) M. Mann and M. Wilm, *Anal. Chem.*, **66**, 4390 (1994).
- 7) 吉野健一, 大城紀子, 徳永千春, 米澤一仁, *J. Mass Spectrom. Soc. Jpn.*, **52**, 106 (2004).
- 8) K. Tanaka, Y. Ido, S. Akita, Y. Yoshida, and T. Yoshida, "Proceeding of the 2nd Japan-China Joint Symposium on Mass Spectrometry; September", (1987), p. 185.
- 9) M. Karas and F. Hillenkamp, *Anal. Chem.*, **60**, 2229 (1988).
- 10) URL: http://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=PMF.
- 11) URL: <http://www.expasy.org/tools/aldente/>.
- 12) URL: <http://prospector.ucsf.edu/prospector/4.0.7/html/msfit.htm>.
- 13) URL: <http://prowl.rockefeller.edu/prowl-cgi/profound.exe>.
- 14) K. R. Clauser, P. Baker, and A. L. Burlingame, *Anal. Chem.*, **71**, 2871 (1999).
- 15) M. Witt, J. Fuchser, and Gökhan Baykut, *J. Am. Soc. Mass Spectrom.*, **14**, 553 (2003).
- 16) T. Satoh, H. Tsuno, M. Iwanaga, and Y. Kammei, *J. Am. Soc. Mass Spectrom.*, **16**, 1969 (2005).
- 17) T. Satoh, H. Tsuno, M. Iwanaga, and Y. Kammei, *J. Mass Spectrom. Soc. Jpn.*, **54**, 11 (2006).
- 18) T. Satoh, T. Sato, and J. Tamura, *J. Am. Soc. Mass Spectrom.*, **18**, (2007) *in press*.
- 19) R. S. Brown and J. J. Lennon, *Anal. Chem.*, **67**, 1998 (1995).
- 20) M. L. Vestal, P. Juhasz, and S. A. Martin, *Rapid Commun. Mass Spectrom.*, **9**, 1044 (1995).

Keywords: MALDI, TOF mass spectrometer, Spiral ion trajectory, Peptide mass fingerprinting